

In einer Welt, in der Maschinen auch vor Kreativität nicht haltmachen, erforscht diese Arbeit die kreativen Möglichkeiten des Latent Space, der mathematische Raum, der allen neueren generativen Algorithmen zugrunde liegt.

Im Kontext der Konzepte der Metamoderne und Metakreativität verdeutlicht Benjamin Bertram, dass innovative Technologien wie Stable Diffusion und ChatGPT nicht nur als Werkzeuge, sondern als kreative Partner fungieren, die den Prozess der Gestaltung revolutionieren. In diesem Zusammenhang werden die Unterschiede zwischen menschlicher und computergestützter Kreativität durch praktische Experimente beleuchtet. Es stellt sich die Frage, auf welche Weise Variablen manipuliert werden können, um unerwartete ästhetische Erlebnisse zu generieren. Inwiefern spielt der Faktor Zufall eine Rolle im kreativen Prozess? Welchen Stellenwert nimmt der Mensch im metakreativen Prozess ein?

Diese Fragen führen zu einer tiefgreifenden Reflexion über die Natur der Kreativität, welche sich zunehmend von traditionellen Vorstellungen entfernt. Die Arbeit bietet nicht nur einen Einblick in die technische Ästhetik der KI-generierten Kunst, sondern ermutigt auch zu einer kritischen Auseinandersetzung mit den ethischen und sozialen Implikationen dieser neuen Form des Schaffens. Mit einem Blick auf zukünftige Forschungsperspektiven und die dynamische Wechselwirkung zwischen Mensch und Maschine werden die Leserinnen und Leser dazu aufgefordert, ihre eigenen kreativen Prozesse zu reflektieren und die Grenzen des Möglichen zu erweitern.

Zur Website ↘



Benjamin Bertram

PASSIVE ILLUSTRATION

Master Dokumentation und Theoriearbeit

Sommersemester 2024

Benjamin Bertram

Master Dokumentation und Theoriearbeit  
Sommersemester 2024

Hochschule für angewandte  
Wissenschaften Hamburg  
Fakultät Design, Medien und Information

Erstprüfer: Prof. Peter Kabel  
Zweitprüferin: Prof. Dr. Michaela Diener

# PASSIVE ILLUSTRATION

Der Latent Space als metakreatives Medium

# PASSIVE ILLUSTRATION

# Inhaltsverzeichnis

<b>1 Einführung</b>	<b>5</b>		
1.1	<i>Kontext und Motivation</i>	6	
1.2	<i>Zielsetzung und Aufbau der Arbeit</i>	6	
1.3	<i>Theoretische Grundlagen</i>	7	
1.4	<i>Relevanz für Design und Illustration</i>	8	
<b>2 Was ist künstliche Kreativität?</b>	<b>9</b>		
2.1	<i>Menschen vs. Maschinen: ein Vergleich</i>	10	
2.1.1	<i>Menschen</i>	10	
2.1.2	<i>Maschinen</i>	11	
2.2	<i>Die Entwicklung der generativen KI</i>	13	
2.3	<i>Stable Diffusion als Treiber von Innovation</i>	14	
2.4	<i>Computational Creativity: Ein Paradigmenwechsel</i>	15	
2.4.1	<i>Die Computational Creativity des Latent Space</i>	15	
2.4.2	<i>Metakreativität als Hermeneutik der KI-Ära</i>	18	
<b>3 Kreativität konzeptualisieren</b>	<b>19</b>		
3.1	<i>Kreativität in der Moderne, der Postmoderne und Metamoderne</i>	20	
3.2	<i>Menschliche vs. maschinelle Kreativität: Kritische Anmerkungen</i>	21	
3.3	<i>Der Latent Space als metakreatives Medium</i>	22	
3.4	<i>Kreativität als Prozess der Kuration und Kontextualisierung</i>	23	
<b>4 Das Porträt im Latent Space</b>	<b>24</b>		
4.1	<i>Automatisierung der Kreativität: Generative Entwurfstechniken</i>	26	
4.1.1	<i>Technische Einstellungen von Diffusionsmodellen</i>	28	
4.1.2	<i>Erkundung des Latent Space</i>	40	
4.2	<i>Flexible Systeme und Multivariabilität</i>	46	
4.2.1	<i>Das Porträt als flexibles System</i>	46	
4.2.2	<i>Ein Porträt systematisch dekonstruieren</i>	48	
4.2.3	<i>Multivariable Meta Thinking</i>	50	
4.2.4	<i>KI-Bilder als Inbegriff der Remix Culture</i>	52	
4.3	<i>Ko-Kreativität und Kollaboration von Mensch und Maschine</i>	52	
4.4	<i>Paralleles Arbeiten in Echtzeit</i>	55	
4.5	<i>Was definiert KI-Ästhetik?</i>	56	
<b>5 Abschluss und Ausblick</b>	<b>60</b>		
5.1	<i>Kreativität neu formulieren</i>	61	
5.2	<i>Bereiche für weitere Forschung</i>	63	
5.3	<i>Schlusswort</i>	64	
<b>6 Nachwort</b>	<b>66</b>		
<b>7 Dokumentation des Designprojektes</b>	<b>68</b>		
<b>Literaturverzeichnis</b>	<b>73</b>		
<b>Abbildungsverzeichnis</b>	<b>81</b>		



# 1 Einführung

---

Die rasante Entwicklung der künstlichen Intelligenz (KI) hat zahlreiche Bereiche, darunter auch Kunst und Design, grundlegend verändert. Zu den wichtigsten Fortschritten gehört der Aufstieg der generativen KI, insbesondere der Algorithmen zur Text- und Bilderzeugung. Diese Systeme, wie z. B. **Midjourney**, **DALL·E** und **Stable Diffusion**, haben den Kunstschaffenden neue Möglichkeiten zur Generierung und Erforschung künstlerischer Inhalte eröffnet. Der **Latent Space**, ein mathematisches Konstrukt, das diesen generativen Modellen zugrunde liegt, dient als Medium, durch das sich neue kreative Möglichkeiten ergeben (Kingma & Welling, 2019).

Diese Arbeit zielt darauf ab, den Latent Space als neues kreatives Medium zu erforschen, insbesondere am Beispiel des **Porträts**. Durch den Einsatz von KI-generierter Kunst und die Untersuchung der Interaktion zwischen maschinell generierter Kreativität und menschlicher Kuratation wird die Arbeit zeigen, wie der Latent Space **systematisch** navigiert werden kann, um neue künstlerische Stile und Ausdrucksformen zu entdecken (Akten, Fiebrink, & Grierson, 2020). Der Fokus auf Porträtkunst dient als Objektiv, durch das die Komplexität maschinengesteuerter Kreativität im Vergleich zu traditionellen Formen menschlicher Kreativität untersucht werden kann.

## 1.1 Kontext und Motivation

Seit 2022 haben Text-zu-Bild-Algorithmen wie DALL•E und Stable Diffusion die Landschaft der digitalen Kunst und Illustration umgestaltet. Diese Modelle können Bilder erzeugen, die das Aussehen professioneller Fotografie oder handgezeichneter Illustrationen imitieren, fallen aber nicht unter die herkömmliche Definition von beidem, da sie lediglich eine Emulation des trainierten Datenmaterials darstellen (*Eigamml et al., 2017*). Die von diesen Systemen erzeugten Bilder sind computergenerierte Illustrationen, die durch Algorithmen entstehen, die Texteingaben interpretieren und in kohärente visuelle Ausgaben umwandeln. Dies wirft wichtige Fragen zur Bedeutung des menschlichen Beitrags in diesem Prozess auf: Welche Art von Kreativität liegt vor, wenn eine Maschine das Bild erschafft? Und wie verändert sich die menschliche Rolle, wenn Maschinen den größten Teil der Arbeit übernehmen?

In diesem Zusammenhang wird das Konzept der **Kuration** in der maschinengenerierten Kreativität zentral. Gestalterinnen und Gestalter sind nicht mehr alleinige Urheberinnen und Urheber des Kunstwerks, sondern steuern und verfeinern den von KI-Systemen produzierten Output. Dieser kreative Prozess ist in erster Linie ein kuratierender – er verwaltet Parameter, wählt günstige Ergebnisse aus und lenkt die maschinelle Ausgabe von der Zufälligkeit zur Struktur. Das bedeutet auch, dass Kreative nicht mehr aktiv am Schaffensprozess teilnehmen, sondern passiv kreativ sind: Ihnen wird die kreative Leistung der Maschinen zuteil und sie reagieren darauf mit Anpassung oder Akzeptanz.

## 1.2 Zielsetzung und Aufbau der Arbeit

Das Hauptziel dieser Forschungsarbeit besteht darin, die Hypothese zu belegen, dass der Latent Space als kreatives Medium fungieren kann. Durch experimentelle Studien zur Porträtgenerierung soll diese Arbeit zeigen, wie verschiedene stilistische Ansätze und künstlerische Methoden aus der Arbeit mit KI-Bildgeneratoren abgeleitet werden können. Die Originalität wird nicht nur anhand des menschlichen Inputs bewertet, sondern auch danach, wie viel die Maschine autonom zum kreativen Prozess beiträgt (*McCormack & Samsel, 2024*).

Die Arbeit gliedert sich in sechs Kapiteln:

- Kapitel 2 liefert den **theoretischen Rahmen** und gibt einen Überblick über die Konzepte der KI, der generativen KI und die historische Entwicklung maschinengenerierter Kunst.

- Kapitel 3 vertieft den theoretischen Rahmen rund um Kreativität, indem es menschliche Kreativität mit maschinengesteuerten Prozessen **vergleicht** und die Rolle des Latent Space bei der Ermöglichung neuartiger künstlerischer Kombinationen untersucht.
- Kapitel 4 stellt die praktischen **Experimente** vor und zeigt, wie KI-generierte Bilder unerwartete, aber ästhetisch überzeugende Ergebnisse hinsichtlich von Portraits liefern können.
- Kapitel 5 schließt mit Überlegungen zu den weitergehenden **Auswirkungen** der KI-gesteuerten Kreativität auf die Zukunft von Kunst und Design und gibt einen **Ausblick** auf weitere Studien.

## 1.3 Theoretische Grundlagen

Im Mittelpunkt dieser Arbeit steht das Konzept des Latent Space. Dabei handelt es sich um einen **mehrdimensionalen mathematischen Raum**, in dem gelernte Repräsentationen von Daten gespeichert werden. Generative KI codiert mit Hilfe des Latent Space abstrakte Merkmale der Eingabedaten (beispielsweise Farbe, Form und Textur), welche manipuliert werden können, um neue Ergebnisse zu erzeugen. Bei der Navigation durch den Latent Space kann der Nutzer oder die Nutzerin unerforschte Gebiete des kreativen Potenzials erkunden und Ergebnisse erzielen, die völlig neuartig oder sogar unerwartet sein können. Die Erkundung dieses Raums ermöglicht eine Mischung aus intuitiven künstlerischen Entscheidungen und algorithmischer Kombinatorik und zeigt neue Denkweisen auf, über die Schnittstelle von Technologie und Kunst nachzudenken.

In diesem Zusammenhang wird in dieser Arbeit die sich entwickelnde Definition von Kreativität kritisch untersucht. Traditionell wurde Kreativität als eine dem Menschen innewohnende Eigenschaft betrachtet, die mit emotionaler Tiefe, kulturellem Ausdruck und sozialer Interaktion verbunden ist, jedoch für KI eine Herausforderung darstellt (*Boden, 1998*). Die Einführung der Künstlichen Kreativität verkompliziert diese Sichtweise jedoch. Durch den Vergleich der menschlichen Kreativität – die häufig von sozialen, emotionalen und kulturellen Faktoren beeinflusst wird – mit den logikgesteuerten Prozessen der maschinellen Kreativität werden in dieser Arbeit herkömmliche Vorstellungen in Frage gestellt und neue Rahmenbedingungen für das Verständnis der Rolle von Algorithmen im kreativen Prozess vorgeschlagen (*Davis, 2021*).

## 1.4 Relevanz für Design und Illustration

Diese Untersuchung ist besonders für Designdisziplinen relevant, insbesondere für die **Illustration**, wo die Auswirkungen der KI bereits spürbar sind. KI-generierte Bilder werden zunehmend in kommerziellen Design- und Kunstprojekten verwendet, und obwohl die meisten Illustratorinnen und Illustratoren noch nicht mit KI arbeiten, erlebt **ein Viertel** bereits schwierige Arbeitsbedingungen wie unrealistische Preisvorstellungen oder weniger Aufträge (*Illustratoren Organisation e.V., 2023*).

Ein weiteres wichtiges Thema sind die Bilder selbst: Während generierte Bilder das Aussehen traditioneller Illustrationen nachahmen können, sind sie in ihrem Ursprung grundlegend anders. Sie sind das Ergebnis komplexer Algorithmen, die innerhalb der Beschränkungen eines Trainingsdatensatzes arbeiten, und nicht das direkte Ergebnis der Hand der Gestalterinnen und Gestalter. Dies wirft grundlegende Fragen der Authentizität, der Originalität, aber auch der Rechte am geistigen Eigentum auf, da der Datensatz auf Millionen von aus dem Internet gesammelten Bildern (*Scraping*) basiert.

Ähnlich wie bei der Erfindung der Fotografie im späten 19. Jahrhundert, der 3D-Tools in den 90er Jahren und dem Aufkommen des Creative Codings und damit auch der generativen Gestaltung während des Jahrtausendwechsels ist die generative KI ein neues Werkzeug, das eine neue Sprache spricht. Obwohl sie immer besser in der **Nachahmung** existierender Medien wird, hat sie auch ihre eigene visuelle Einzigartigkeit – aber wie sieht diese aus, und wie können die Kreativen sie nicht nur zufällig, sondern bewusst hervorbringen?

Diese Fragen werden die experimentelle Arbeit in den folgenden Kapiteln leiten, in denen KI-generierte Porträts im Hinblick auf ihren künstlerischen Wert und die Rolle der Gestalterinnen und Gestalter bei ihrer Erstellung bewertet werden.



Abb. 1: Auswirkungen durch KI auf das illustrative Geschäft.

## 2 Was ist künstliche Kreativität?

Dem KI-Forscher Demis Hassabis zufolge ist KI ein Prozess, der **unstrukturierte Informationen** oder Daten in **nützliches, wertbares Wissen** umwandelt (*Royal Academy of Arts, 2018*).

Sie unterscheidet sich von der menschlichen Intelligenz durch ihre Fähigkeit, Aufgaben wie visuelle Wahrnehmung, Spracherkennung, Entscheidungsfindung und Verarbeitung natürlicher Sprache auszuführen. Die KI hat sich von frühen symbolischen Ansätzen zu modernen Systemen für maschinelles Lernen (*ML*) und Deep Learning entwickelt. Die heutige KI-Landschaft wird von der sogenannten Narrow AI dominiert, die für spezifische Aufgaben entwickelt wurde, während die allgemeine künstliche Intelligenz (*Artificial General Intelligence, AGI*) ein zukünftiges Ziel bleibt, bei dem Maschinen potenziell die kognitiven Fähigkeiten des Menschen erreichen oder übertreffen könnten (*Turing, 1950*).

Die Entwicklung der KI lässt sich bis in die Mitte des 20. Jahrhunderts zurückverfolgen, mit bahnbrechenden Beiträgen von Persönlichkeiten wie Alan Turing und John McCarthy. McCarthy's Prägung des Begriffs „künstliche Intelligenz“ im Jahr 1955 markierte den Beginn der systematischen KI-Forschung (*McCarthy, 1955*). Frühe KI-Modelle stützten sich auf symbolisches Denken, aber mit der Einführung **neuronaler Netze**, die eine ausgefeiltere Mustererkennung und Entscheidungsfindung ermöglichten, machte das Gebiet dramatische Fortschritte (*LeCun, Bengio, & Hinton, 2015*).



die mit Hilfe komplexer **Algorithmen** Aufgaben lösen, die herkömmliche Maschinen nicht bewältigen können. Sie verfügen über eine immense Rechenleistung mit Milliarden von Transistoren und spezialisierter Hardware wie Grafikprozessoren (*GPUs*) und neuronalen Prozessoren (*NPU*s), die es ihnen ermöglichen, große Datenmengen schnell zu verarbeiten. Während das menschliche Gehirn etwa 2,5 Petabyte speichern kann, ist die Speicherkapazität von Maschinen theoretisch unbegrenzt, insbesondere durch Cloud-Speicher und verteilte Systeme (*Backblaze, 2023*).

Maschinen zeichnen sich durch Multitasking aus, da sie mehrere Prozesse parallel und effizient durchführen können, mit Datenübertragungsraten von mehreren Terabit pro Sekunde (*Tb/s*). KI-Systeme wie GPT-4 sind sprachfähig und können natürliche Sprache verstehen und erzeugen. (*OpenAI, 2023*)

Obwohl Maschinen keine Emotionen oder emotionale Intelligenz besitzen, können sie mit Hilfe von Algorithmen menschliche Emotionen analysieren und auf sie reagieren, das sogenannte Affective Computing (*Somers, 2019*). Neuroplastizität gibt es bei Maschinen nicht, aber sie können ihre Leistung verbessern und sich durch maschinelles Lernen und neuronale Netze an neue Daten anpassen. Maschinen nutzen Strom direkt als Primärenergie, ohne niedrigenergetische organische oder nicht organische Feststoffe umzuwandeln. Alterung und „Tod“ betreffen Maschinen durch Hardwarefehler und inkompatible Software, die aber durch Wartung und Updates behoben werden können. Ihre Weiterentwicklung erfolgt derzeit durch vom Menschen gesteuerte technologische Fortschritte.

Maschinen sind vernetzt, kooperieren über das Internet der Dinge (*IoT*) und nutzen das Cloud-Computing. Sie haben keine Kultur, können aber kulturelle Inhalte analysieren und kreativ neue Werke schaffen (*Gabsi, 2024*).

AGI bezieht sich auf eine Form der KI, die allgemeine kognitive Aufgaben auf einem Niveau ausführen kann, das der menschlichen Intelligenz entspricht oder sie übertrifft. Im Gegensatz zur heutigen KI, die oft auf bestimmte Anwendungen beschränkt ist (z. B. *Spracherkennung* oder *Bildverarbeitung*), wäre die AGI in der Lage, flexibel auf neue Aufgaben zu reagieren, ohne dass sie speziell dafür programmiert werden müsste. AGI wird als langfristiges Ziel in der KI-Forschung angesehen (*Gobble, 2019*).

Eine Maschine ist demnach ein vom Menschen geschaffenes System, das mechanische, elektrische oder rechnerische Komponenten verwendet, um Aufgaben effizient und präzise zu automatisieren. Moderne Maschinen mit KI können komplexe Probleme lösen und sich durch maschinelles Lernen an

Im Folgenden steht das Wort „Maschine“ gleichbedeutend für das Konglomerat aus Servern, Computern, Programmen und Daten, die für KI-Systeme wichtig sind.

neue Daten anpassen, ohne Emotionen oder Bewusstsein zu besitzen.

Während die Kreativität beim Menschen auch eine soziale Funktion erfüllt, ist dies bei Maschinen irrelevant – künstliche Kreativität aber hat einen relevanten Vorteil, wenn es um Zielfunktionen geht: Hassabis verweist auf das KI-Programm AlphaGo, das aus allen aufgezeichneten Varianten des Brettspiels GO Muster und mögliche Positionen ableitet. Das Programm extrapoliert kreative neue Strategien, anstatt einfach nur durchschnittliche Gewichtungen vorzunehmen (*Grassegger & Krogerus, 2023*).

Im Hinblick auf neue Lösungen, die sowohl nützlich als auch praktisch sind, kann die künstliche Kreativität in dem Fall die menschliche Kreativität nicht nur erreichen, sondern sogar übertreffen. Und die Zielfunktion ist variabel: für die Maschine spielt es keine Rolle, ob es darum geht, optimale Spielzüge zu entwickeln, unfallfrei zu fahren oder ästhetisch ansprechende Bilder zu erzeugen.

## 2.2 Die Entwicklung der generativen KI

Die Idee, Kunst mit Hilfe von Maschinen zu erzeugen, ist nicht neu. Historische Beispiele sind **Automata** – mechanische Geräte aus dem 18. Jahrhundert, die menschliche Bewegungen nachahmen und Zeichnungen erstellen konnten. Diese Automaten, wie der berühmte Maillardet-Automat, stellen frühe Experimente mit mechanischer Kreativität dar, sind aber im Grunde regelbasierte Systeme. Sie arbeiteten mit vorprogrammierten Mechaniken statt mit autonomer Entscheidungsfindung.

Das Aufkommen digitaler Computer in der Mitte des 20. Jahrhunderts revolutionierte die kreative Ausdrucksweise. Wegbereitende Arbeiten wie die von Harold Cohen und John Whitney nutzten computergestützte Systeme, um generative Kunst zu schaffen, die Zufälligkeit und algorithmisches Design miteinander verbindet. Whitneys algorithmische Filme aus den 1960er Jahren sind ein Beispiel dafür, wie frühe Computergrafiken den Grundstein für die heutigen generativen KI-Systeme legten (*Moritz, 1997*).

Mit der Entwicklung von Generative Adversarial Networks (*GANs*) durch Ian Goodfellow im Jahr 2014 haben Algorithmen einen bedeutenden Sprung nicht nur in ihren „kreativen“ Fähigkeiten gemacht, sondern auch an zunehmender Bekanntheit gewonnen. *GANs* bestehen aus zwei neuronalen Netzen – einem Generator und einem Diskriminator –, die gegeneinander antreten und aus zufälligem Rauschen äußerst realistische Bilder erzeugen. Dieser Durchbruch ermöglichte es



Abb. 4: Automaton von Henri Maillardet 1805.

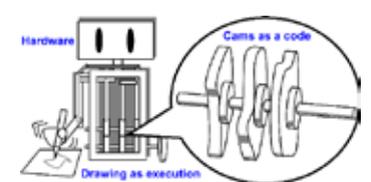


Abb. 5: Wirkungsweise von Automata.

Maschinen, nicht nur zu replizieren, sondern auch neue, ungesehene Bilder zu erzeugen, und markierte die Geburtsstunde der modernen generativen KI (Goodfellow et al., 2014).

Neben GANs haben sich Variational Autoencoders (VAEs) und Diffusionsmodelle als wichtige Werkzeuge in der generativen KI etabliert. VAEs, die 2013 von Kingma und Welling vorgestellt wurden, sind probabilistische Modelle, die die latente Struktur von Daten erlernen und neue Stichproben aus der erlernten Verteilung generieren können (Kingma & Welling, 2013). VAEs sind besonders nützlich, um glatte Interpolationen zwischen verschiedenen Datenpunkten im latenten Raum zu erstellen, was sie für die Bildsynthese leistungsstark macht.

In jüngerer Zeit haben Diffusionsmodelle aufgrund ihrer Stabilität und ihrer Fähigkeit, qualitativ hochwertige Bilder zu erzeugen, an Aufmerksamkeit gewonnen. Diese Modelle, wie sie in Stable Diffusion und DALL·E verwendet werden, verfeinern Bilder schrittweise, was eine kontrolliertere und präzisere Bilderzeugung ermöglicht (Ho et al., 2020).

## 2.3 Stable Diffusion als Treiber von Innovation

Die Entwicklung von generativen **Open-Source**-KI-Tools hat die Landschaft der digitalen Kreativität grundlegend verändert. Mit der Veröffentlichung von Stable Diffusion von Stability AI im Jahr 2022 begann eine neue Ära der Zugänglichkeit für Kreative und Interessierte gleichermaßen. Im Gegensatz zu früheren proprietären Modellen wie DALL·E ist Stable Diffusion quelloffen und ermöglicht es den Nutzern und Nutzerinnen, das Modell zu modifizieren und an ihre spezifischen künstlerischen Bedürfnisse anzupassen (Stability AI, 2022). Zwar gab es bereits zuvor Open-Source-Bildgeneratoren wie StyleGAN und VQGAN+CLIP, doch erst die Kombination aus einfachem Zugang, Kontrollierbarkeit und hervorragender Bildqualität führte zur breiten Akzeptanz von Stable Diffusion. Diese Zugänglichkeit förderte schnell die Bildung einer innovativen Gemeinschaft von Kunstschaffenden und Entwickelnden (Vincent, 2022).

Das Aufkommen von Plattformen wie **Civitai.com** und das aus der Open-Source-Community entwickelte Stable Diffusion Add-On **ControlNet** von Zhang et al., (2023) hat diese Demokratisierung weiter vorangetrieben, indem es den Nutzenden ermöglicht, fein abgestimmte Modelle gemeinsam zu nutzen und eine größere Kontrolle über die Ausgabe von KI-generierten Bildern auszuüben. Diese Tools bieten ein Maß an Individualisierung, das zuvor in proprietären KI-Systemen nicht verfügbar war, und verbessern sowohl die

Nutzungsfreundlichkeit als auch das kreative Potenzial der bildgebenden Tools erheblich.

## 2.4 Computational Creativity: Ein Paradigmenwechsel

Mit der Entwicklung der generativen KI hat sich ein Paradigmenwechsel in unserem Verständnis von Kreativität abgezeichnet. Traditionell wurde Kreativität als eine ausschließlich menschliche Fähigkeit betrachtet, die eng mit emotionalen, kulturellen und sozialen Kontexten verbunden ist. Das Aufkommen der künstlichen Intelligenz stellt diese Vorstellung in Frage, da Maschinen nun in der Lage sind, selbständig hochkreative Werke zu schaffen. **Computational Creativity** unterscheidet sich von menschlicher Kreativität dadurch, dass sie programmierten Systemen folgt, die durch Daten und mathematische Regeln definiert sind, und nicht durch emotionale oder subjektive Beurteilung. Laut [computationalcreativity.net](http://computationalcreativity.net) ist Computational Creativity

*„... the study and simulation, by computational means, of behavior, natural and artificial, which would, if observed in humans, be deemed creative.“*

Computational Creativity kann uns helfen, unsere eigene menschliche Kreativität zu verstehen, da wir beurteilen können, ob etwas kreativ ist oder nicht – obwohl ein großer Teil, vielleicht sogar die gesamte menschliche Kreativität als neuartige Kombination von bereits vorhandenen Ideen oder Objekten verstanden werden kann (Boden, 2010).

Der Latent Space ist ein gutes Beispiel, um Computational Creativity zu demonstrieren. Nicht nur wegen seiner Kombinationskraft, sondern auch wegen seiner mathematischen Möglichkeiten der **Interpolation** und **Extrapolation**. Um dies zu entschlüsseln, werfen wir einen Blick auf das Konzept des latenten Raums.

### 2.4.1 Die Computational Creativity des Latent Space

Der Latent Space kann als ein mehrdimensionaler **Vektorraum** definiert werden, der die erlernten Merkmale eines Datensatzes darstellt. Mehrdimensionaler Vektor bedeutet, dass der Latent Space bei einem Datensatz von Gesichtern verschiedene Merkmalsdimensionen wie Alter (Vektor alt – jung), Geschlecht (Vektor männlich – weiblich) oder Gesichtsausdruck (Vektor traurig – fröhlich) kodieren kann. Er dient als konzeptioneller Rahmen, der die Darstellung und Bearbeitung

Diese Definition wurde 2013 fallen gelassen, aber sie scheint dem Konzept besser zu entsprechen als die aktuelle Definition von [computationalcreativity.net](http://computationalcreativity.net) (Jordanous, 2014)

Interpolation bezeichnet die Berechnung von Werten zwischen zwei Datenpunkten, Extrapolation die Berechnung von Werten die sich aus dem Muster der Datenpunkte ableiten lassen (Royal Academy of Arts, 2018).



Abb. 6: Das Originalbild (links) wird im Latent Space in Richtung eines „Lächelvektors“ interpoliert, wodurch eine Reihe von Versionen des Originals entsteht, von lächelnd bis traurig.

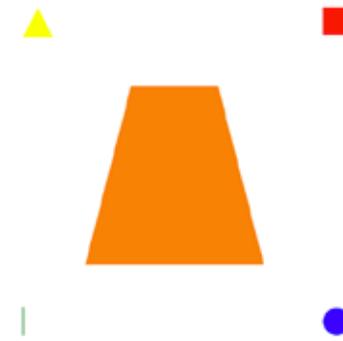
komplexer Daten ermöglicht und eine effektive Analyse und Modellierung erleichtert. Die Darstellung des Latent Space wird aus den Eingabedaten abgeleitet und zeichnet sich durch die Extraktion aussagekräftiger Merkmale aus, wodurch die Fähigkeit von KI-Systemen, Informationen zu verstehen und zu verarbeiten, verbessert wird. Dieses Konzept ist für die KI von entscheidender Bedeutung, da es eine effiziente Darstellung und Verarbeitung komplexer Daten ermöglicht.

Bei GANs, VAEs und Diffusionsmodellen wird der Latent Space für die Bilderzeugung genutzt. Durch die Manipulation von Punkten im Latent Space können vielfältige und kohärente Bilder erzeugt werden, was die Fähigkeit des latenten Raums, ein breites Spektrum von Variationen zu kodieren, unter Beweis stellt (Kingma & Welling, 2013). In Sprachmodellen erleichtert die latente Kodierung von Textdaten die Informationssuche, indem sie die Identifizierung von semantisch verwandten Dokumenten oder Passagen ermöglicht.

Das theoretische Verständnis des Latent Space ist noch in der Entwicklung begriffen. Es gibt zwar keine endgültigen Theorien, aber verschiedene Perspektiven und Annahmen helfen, seine Rolle zu verstehen. Die Manifold Hypothese beispielsweise besagt, dass der niedrigdimensionale Latent Space oft für die hochdimensionalen Beobachtungen verantwortlich ist, die wir sehen. Das bedeutet, dass hochkomplexe Phänomene der Realität auf wenige **virtuelle** Muster reduziert werden können. Dieses Konzept ist zwar nur grob definiert, funktioniert aber in der Praxis (Kingma & Welling, 2013).

Wir können in diesem Raum navigieren, um neue Kombinationen von Merkmalen zu erzeugen, die zu neuen Bildern oder Designs führen. Die Erforschung des Latent Space ermöglicht sowohl aleatorische (zufällige) als auch strukturierte kreative Arbeitsweise, die einen neuen Rahmen für das Verständnis von Kreativität bieten.

Das Phänomen der KI-**Halluzinationen**, die keine Halluzinationen im herkömmlichen Sinne sind, bezieht sich auf die einzigartigen und oft unerwarteten Ergebnisse, die von KI-Systemen erzeugt werden, wenn sie Daten auf eine Art und Weise interpretieren und neu erstellen, die der Mensch nicht vorhersehen kann. Dieses Phänomen wurde erstmals bei Sprachmodellen und Computer-Vision-Tools beobachtet und beschreibt die Erkennung von Mustern oder Objekten, die für menschliche Beobachter nicht vorhanden oder nicht wahrnehmbar sind und unsinnige oder völlig ungenaue Ergebnisse erzeugen (IBM, n.d.). Während KI-Halluzinationen in den meisten Fällen sicherlich **unerwünschte** Ergebnisse sind, bieten sie auch eine Reihe faszinierender Anwendungsfälle, die Künstlerinnen und Künstlern helfen können, das kreative



Statt ein orangefarbiges Trapez ausführlich zu beschreiben, finden wir es als Interpolation der niedrigkomplexen Elemente von den zwei Dimension Form und Farbe im Latent Space der Datenpunkte grüne Linie, gelbes Dreieck, rotes Quadrat und blauer Kreis.

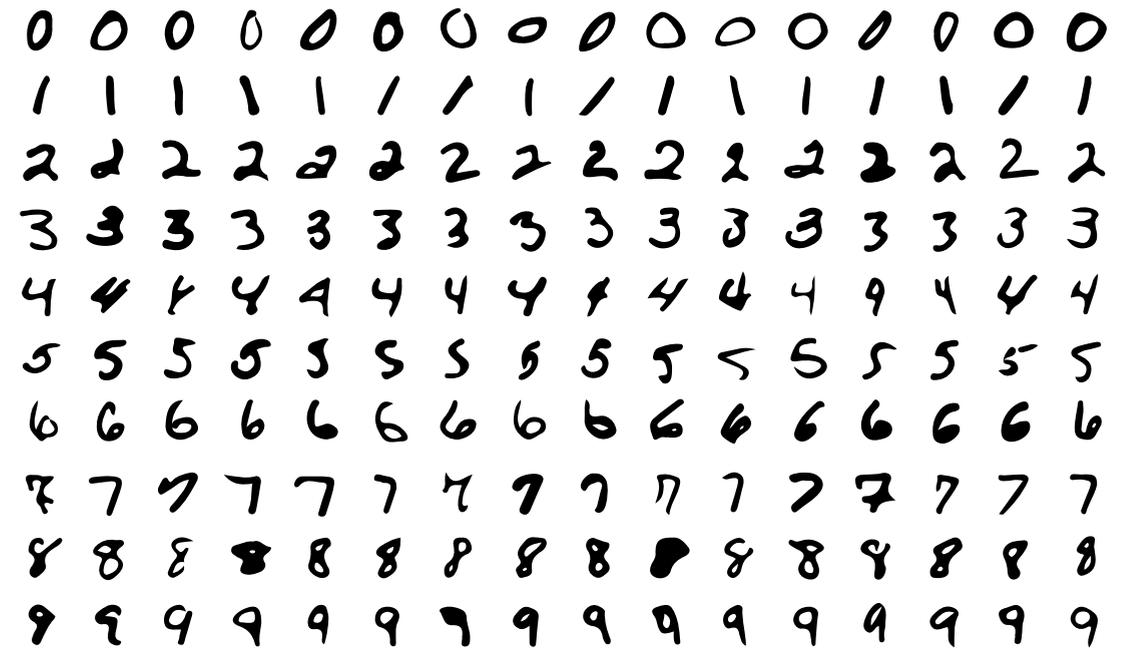


Abb. 7: Auszüge aus dem Original MNIST (Modified National Institute of Standards and Technology)-Datensatz. Dabei handelt es sich um eine Datenbank mit handgeschriebenen Ziffern mit einem Trainingssatz von 60.000 Beispielen.

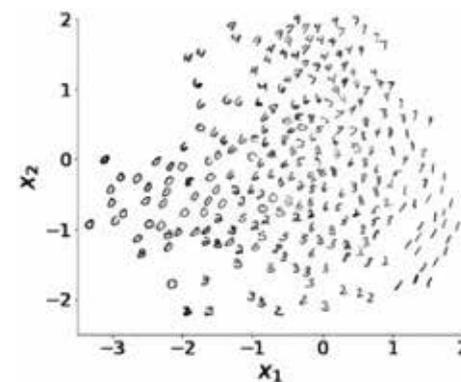


Abb. 8: Nach dem Training können die Ziffern anhand latenter Variablen gruppiert werden.

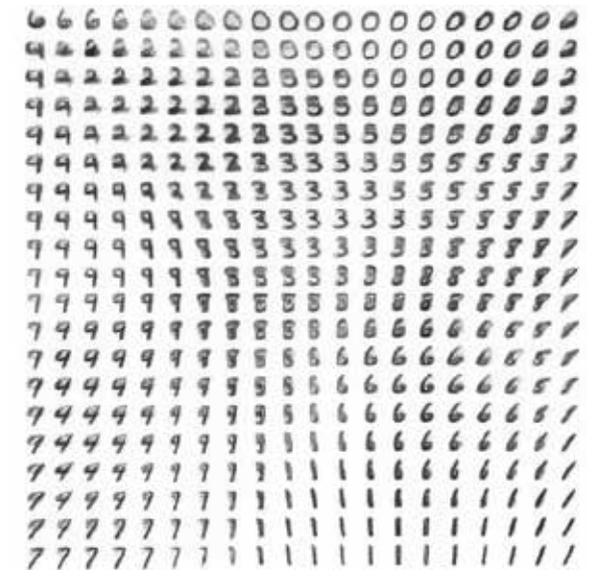


Abb. 9: Latent Space des MNIST Datensatzes.

Potenzial von KI auf positive Weise zu nutzen, z. B. durch die Entdeckung neuer Wege oder Sichtweisen (*The AI GRID, 2024*).

## 2.4.2 Metakreativität als Hermeneutik der KI-Ära

Der von Eduardo Navas geprägte Begriff der **Metakreativität** bezieht sich auf eine Form der Kreativität, die über die menschlichen Grenzen hinausgeht, indem nicht-menschliche Systeme in den kreativen Prozess einbezogen werden. In dieser Sichtweise nehmen Maschinen nicht nur an der Kreativität teil, sondern erweitern den Rahmen des Möglichen. Navas argumentiert, dass KI-generierte Kunst einen kulturellen Wandel darstellt, bei dem die Grenzen zwischen menschlicher und maschineller Kreativität verschwimmen und die traditionellen Vorstellungen von Autorschaft und künstlerischem Genie in Frage gestellt werden (*Navas, 2022*). Navas definiert Metakreativität als

*„a cultural variable that emerges when the creative process moves beyond human production to include non-human systems. This definition includes artificial intelligence and machine learning; for emerging intelligent technology, specifically, this means that a non-human entity is able to ›learn‹ in order to produce something that appears creative“* (*Navas, 2022, S. 231*).

Ähnlich wie die Idee der Computational Creativity besagt Metakreativität nicht ausdrücklich, dass etwas kreativ ist, sondern dass es kreativ erscheint. Die Herausforderung liegt also offensichtlich darin, wie die Ergebnisse kreativer Maschinen von Menschen bewertet und interpretiert werden, da sie von nicht-menschlichen Akteuren produziert werden.

Generell scheint der Wert eines Bildes zu sinken, sobald Menschen erfahren, dass es von KI erstellt wurde (*Horton, White & Iyengar, 2023*). Der Verkauf des KI-generierten Porträts „Edmond de Belamy“ für 432.000 Dollar bei der Christie’s Kunstauktion 2018 war eher die Ausnahme als die Norm, hat aber Fragen aufgeworfen, wie Anerkennung und Verantwortung den beteiligten Personen zugewiesen werden sollten und wie die anthropomorphe Wahrnehmung des KI-Systems (der Algorithmus wurde als Urheber genannt) zum Erfolg des Kunstwerks beigetragen hat (*Epstein, Levine, Rand & Rahwan, 2020*).

Der Begriff der kreativen Werke von Maschinen bringt eine neue Ebene der Komplexität in die Debatte über Kreativität ein, da er die Frage aufwirft, was Kreativität eigentlich ist und was als kreativ gilt.



Abb. 10: Das Portrait „Edmond de Belamy“, rechts unten in der Ecke wird der Algorithmus als Urheber genannt.

## 3 Kreativität konzeptualisieren

Kreativität, die seit langem als Kennzeichen menschlicher Intelligenz gilt, wurde definiert als die Fähigkeit, Ideen oder Produkte zu entwickeln, die in einem bestimmten kulturellen oder sozialen Kontext sowohl neu als auch wertvoll sind (*Runco & Jaeger, 2012*). Traditionell wird Kreativität als ein zutiefst persönlicher und emotionaler Prozess angesehen, der Intuition, Einfühlungsvermögen und kulturelle Sensibilität erfordert. Dieser Prozess spiegelt häufig die individuellen Erfahrungen der Künstlerinnen und Künstler wider, die von Emotionen, gesellschaftlichen Bedingungen und dem Wunsch, etwas Tiefgründiges oder Originelles auszudrücken, beeinflusst werden (*Runco, 2023*).

Im 18. Jahrhundert betrachtete Immanuel Kant in seinem Werk über Ästhetik Kreativität als eine Form des Selbstaushdrucks, die dem **Genie** des Einzelnen eigen ist, und schuf damit die Grundlage für das moderne Verständnis der Künstlerinnen und Künstler als autonom schöpfende Person (*Kant, 1790*). Diese Fokussierung auf die Fähigkeit des Einzelnen, über Konventionen hinauszugehen und etwas Neues und Sinnvolles zu schaffen, legte den Grundstein für den modernistischen Ansatz der Kreativität.

### 3.1 Kreativität in der Moderne, der Postmoderne und Metamoderne

Zu Beginn des 20. Jahrhunderts entwickelte sich die **Moderne** als vorherrschende künstlerische Bewegung, die sich durch den Glauben an Fortschritt, Innovation und die Überwindung traditioneller Grenzen auszeichnete. Kreative Menschen der Moderne, wie Pablo Picasso und Virginia Woolf, strebten nach völlig neuen Ausdrucksformen und legten dabei den Schwerpunkt auf Originalität und die Abkehr von den Konventionen der Vergangenheit. Kreativität in der Moderne wurde mit dem Genie des Künstlers in Verbindung gebracht, der als Visionär angesehen wurde, der in der Lage war, durch neuartige Techniken und Ideen Wahrheiten zu enthüllen. Laut Buehrle (2008) war Kunst eindeutig nicht mehr ein Abbild der Welt, sondern gab die subjektive Sichtweise von Individuums und Kreativen wieder.

Mitte des 20. Jahrhunderts entstand die **Postmoderne** als Reaktion auf die Ideale der Moderne. Postmoderne Theoretiker und Theoretikerinnen lehnten die Vorstellung von absoluten Wahrheiten und großen Erzählungen ab und betonten Fragmentierung, Pluralität und Ironie in der kreativen Arbeit (Aylesworth, 2015). Die kunstschaftende Personen wurde im postmodernen Denken zum „Bricoleur“ (Vgl. Lévi-Strauss, 1962, S. 16), die Bedeutung aus bereits existierendem kulturellem Material konstruiert, anstatt eigene Ideen zu entwickeln (Duy-medjian & Rüling, 2010). Diese Verschiebung wurde besonders in der Remix-Kultur veranschaulicht, wo die Grenzen zwischen ursprünglicher Schöpfung und Aneignung verwischt wurden und das modernistische Ideal der Künstlerinnen und Künstler als einsames Genie in Frage gestellt wurde (Navas, 2012).

Zu Beginn des 21. Jahrhunderts ist die **Metamoderne** als ein Rahmen entstanden, der zwischen den Idealen der Moderne und der Postmoderne wechselt. Künstlerische Werke und Theorien der Metamoderne versuchen, die Aufrichtigkeit und den Idealismus der Moderne mit der Ironie der Postmoderne in Einklang zu bringen. Wie Vermeulen und Van den Akker (2010) beschreiben, ist die Metamoderne durch eine hoffnungsvolle Suche nach Sinn gekennzeichnet, auch wenn sie die Unmöglichkeit anerkennt, diesen Sinn jemals vollständig zu erreichen.

In der Metamoderne wird Kreativität als ein Balanceakt zwischen ernsthafter Erkundung und zynischer Selbstreflexion gesehen, bei dem sich Künstlerinnen und Künstler der kulturellen und sozialen Konstruktionen bewusst sind, die ihr Werk prägen, und dennoch nach einem authentischen Ausdruck streben. Mehr noch: im posthumanistischen Diskurs verwischen die Grenzen zwischen Mensch und Maschine zunehmend, und es ist nicht ungewöhnlich auch Maschinen



Abb. 11 The Fountain von Marcel Duchamp gilt als Schlüsselwerk der Moderne.

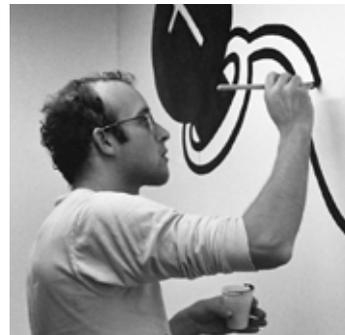


Abb. 12: Keith Haring verband Massenkultur mit Hochkultur und wurde damit zur postmodernen Ikone.



Abb. 13: Die Elphilmharmonie gilt als metamodernes Wahrzeichen, die Ironie und Tradition miteinander vereint.

Kreativität – oder zumindest das Potential derselben – zuzuschreiben (Berlich, 2022).

Wenn wir Künstliche Intelligenz und Computational Creativity in diesen theoretischen Rahmen integrieren, wird deutlich, wie sie die traditionellen, anthropozentrischen Vorstellungen von Kreativität herausfordern. Während das Konzept des künstlerischen Genies für die Moderne von zentraler Bedeutung war, wird es durch künstlich generierte Werke zunehmend in Frage gestellt. Die Postmoderne stellte bereits die Idee der Originalität infrage, aber KI verkompliziert diesen Diskurs, indem sie Werke schafft, die keine bloßen Kopien sind, sondern nahezu ohne menschliche Intervention innovative Kombinationen vorhandener Daten im Latent Space darstellen.

Die Metamoderne, mit ihrem Fokus auf den fließenden Übergang zwischen Ironie und Ernst, Sinn und Unsinn, sowie der Dekonstruktion von Autorenschaft, bietet einen geeigneten Rahmen zur Analyse künstlicher Kreativität. Der Einsatz von KI in der Kunst wechselt zwischen der zynischen Auffassung, dass es sich lediglich um algorithmische Nachahmung menschlicher Kreativität handelt, und der ernsthaften Überzeugung, dass KI die kreativen Möglichkeiten des Menschen nicht nur nachahmt, sondern eigenständig kreativ ist. Die kollaborative Natur von KI-generierter Kunst, bei der der Mensch kuratiert und lenkt, fügt sich in das Ethos der Metamoderne ein, das sowohl die menschliche Handlungsfähigkeit als auch das technologische Potenzial umfasst (Berlich, 2022).



Abb. 14: Nicht nur metamodern, sondern auch metakreativ: Boris Eldagsens „The Electrician“.

### 3.2 Menschliche vs. maschinelle Kreativität: Kritische Anmerkungen

Wie „kreativ“ können Maschinen also sein? Margaret Boden (2014) argumentiert, dass die Antwort nicht technischer, sondern philosophischer Natur sei. Boden argumentiert, dass KI-Systeme zwar Ergebnisse produzieren können, die kreativ erscheinen, aber nicht die Intentionalität, Emotionalität oder das bewusste Verständnis besitzen, die für echte Kreativität erforderlich sind.

Kritische Stimmen wie David Gelernter (1994) sind der Meinung, dass KI-generierte Kunst zwar technisch beeindruckend ist, aber nicht die emotionale Resonanz und kulturelle Bedeutung von Werken hat, die von Menschen geschaffen wurden. Gelernter betont, dass es bei der Kunst nicht nur um das Endprodukt geht, sondern auch um den Prozess und die emotionale Reise der Künstlerin oder des Künstlers. Maschinen, denen es an Emotionen fehlt, könnten diese wesentliche Komponente des kreativen Prozesses nicht nachbilden (Gelernter,

1994, S. 149). Solche **Abgrenzungsmechanismen** sind auch in der Kunstszene selbst häufig zu hören: Künstler und Künstlerinnen wie Aelfleda Clackson argumentieren: „Unlike AI, I have feelings, and that is what makes my work meaningful.“ (Clackson, 2024).

Catherin Botha (2023) unterstellt Boden – und damit der Kritik, die sich auf die Abgrenzung zwischen Mensch und Maschine bezieht – hinsichtlich ihrer Definition „echter“ Kreativität einen **Anthropozentrismus** und ermuntert, von der menschlichen Einzigartigkeit abzusehen:

*“we need not jealously guard creativity as an ability that is exclusively held by humans. If we recognize the anthropocentric conception of agency as problematic within our understandings of creativity, it seems possible to grant non-human agents and artefacts entry into the world of creativity, without undermining the value of the artefacts and ideas created, and without undermining the worth and dignity of our humanness. Just as it is easy for us to say a system is logical, or even rational in its activities, and so it seems to me that saying a system is creative, should be just as easy” (Botha, 2023 S. 209).*

### 3.3 Der Latent Space als metakreatives Medium

Der Begriff „Medium“ wird in diesem Kontext verstanden als ein Element, das uns die Welt des „da draußen“ durch die **Abstraktionen** zugänglich macht. Diese Definition nach Vilém Flusser (2000) betrachtet insbesondere Bilder als dieses Medium. Flusser unterscheidet zwischen traditionellen und technischen Bildern. Obwohl der Latent Space eine mathematische Abstraktion ist, sind die konkreten Bilder, die ihm entspringen, technischer Art: sie werden uns durch einen „Apparatus“ – dem Algorithmus – in einem für uns verständlichen Format zugänglich gemacht. In seinem Buch „Illustration – 100 Wege einen Vogel zu malen“ widmet Felix Scheinberger (2013) dem Computer und der generativen Gestaltung ein ganzes Kapitel und ordnet Bilder, die mit einem Algorithmus entstanden sind, klar der Illustration zu. Auch wenn KI Bilder generieren kann, die einer Photographie verblüffend ähneln, so liegt es doch in ihrer Natur begründet, dass sie keine sind. Als Meta-Medium kann KI zwar traditionelle Techniken simulieren, jedoch sind generierte Bilder letztlich mit einem Programm erstellte Bilder und damit Illustrationen.

Wie kann nun ein Medium kreativ sein? Nach dem Training einer KI gibt es eine Datei, die das Trainingsmaterial in

einem latenten mathematischen Raum kodifiziert. Während der menschliche Verstand ununterbrochen aktualisiert wird (Siehe *Neuroplastizität Kapitel 2.1.1*) ist der Latent Space nach dem Training ein kristalliner Raum aus festen Werten. Aber damit enthält er auch alle möglichen Kombinationen dieser Werte, und somit auch neue und weniger neue, oder kreative und nicht kreative Lösungen. Der Latent Space kann damit als Repräsentation aller kreativen Möglichkeiten in diesem Datensatz betrachtet werden. Damit wird der Latent Space im doppelten Sinne metakreativ: einmal weil er Ergebnis eines nicht menschlichen Lernens ist und einmal weil er alle Lösungen bereits **enthält und nicht entdeckt**. Es ist der Mensch, der in diesem Sinne basiskreativ *bleibt*, weil er die Lösungen im Latent Space suchen und aufdecken muss, aber nicht weil er sie generiert.

### 3.4 Kreativität als Prozess der Kuratation und Kontextualisierung

Die Rolle der Kreativen verlagert sich bei der Arbeit mit Bildgeneratoren von der direkten Kreation zur Kuratierung der maschinellen Ergebnisse, indem sie die KI-generierten Werke auswählen und verfeinern, die ihrer künstlerischen Vision entsprechen. Dieser Prozess spiegelt die **Oszillation** (Vgl. Vermeulen, Van den Akker, 2010, S. 1) zwischen Kontrolle und Hingabe wider, bei der Kreative das empfindliche Gleichgewicht zwischen menschlicher Absicht und der Vergeblichkeit derselben steuern.

Der Akt des Kuratierens führt eine neue Dynamik in den kreativen Prozess ein, bei dem die kunstschaftende Person als Mittelperson zwischen den algorithmischen Möglichkeiten der KI und der kulturellen Bedeutung, die dem endgültigen Werk innewohnt, agiert. Diese Dynamik spiegelt den allgemeinen Wandel in der kreativen Praxis wider, in der die Rolle der Kreativen nicht mehr auf die Produktion neuer Werke beschränkt ist, sondern auch die Auswahl, Verfeinerung und Neuinterpretation von maschinengenerierte Inhalte. Es ist die Aufgabe von Künstlerinnen und Künstlern, die kreativ relevanten Ideen der Maschine auszuwählen und zu kontextualisieren (Navas, 2021).

In der Stable Diffusion Community bezeichnet der Begriff „Skill Issue“ Personen, die es nicht schaffen, das gewünschte Bild aus dem Latent Space zu generieren, als unzureichend mit dem Programm vertraut. Das Scheitern wird auf mangelnde Fertigkeiten, nicht auf technische Grenzen, zurückgeführt.

## 4 Das Porträt im Latent Space

„Das erste Bild war ein Porträt“, schreibt Joanna Woodall im Vorwort ihres Buches „Portraiture: Facing the Subject“, in Anlehnung an die Erzählung von Narziss, einem jungen Mann, der sich in sein eigenes Spiegelbild verliebt (Woodall, 1997, S. 1). Die Porträtmalerei spielt in der Kunst seit jeher eine wichtige Rolle und dient als Medium zur Erforschung der menschlichen Identität, der Gefühle und des sozialen Status. Wie Shearer West in ihrem Buch „Portraiture“ beschreibt, werden Porträts seit langem verwendet, um das Individuum zu verewigen, wie die frühe Form der Porträtmalerei, die Totenmasken im alten Rom. Indem sie nicht nur das physische Abbild, sondern auch Aspekte des psychologischen und moralischen Charakters des Dargestellten wiedergaben, wurden Porträts nicht nur zum Ahnenkult, sondern auch zu einem Mittel der Repräsentation (West, 2004).

Seit der Renaissance entwickelte sich die Porträtkunst, um das Wesen der Innenwelt des Porträtierten einzufangen, und spiegelte das zunehmende Interesse an Individualismus und Selbsterkenntnis (Woodall, 1997). Künstlerinnen und Künstler wie Leonardo da Vinci und Rembrandt beherrschten das schwierige Gleichgewicht zwischen Realismus und Idealisierung und machten das Porträt sowohl zu einer Darstellung der Person als auch zu einer umfassenderen Aussage über die Menschheit und die Gesellschaft.

Das Selbstporträt war für Kreative ebenfalls eine Möglichkeit, ihre eigene Identität und ihren Platz in der Welt zu erkunden. Das Selbstporträt wurde zu einer Methode der Selbstreflexion und des künstlerischen Experiments, wie man an den Werken von Albrecht Dürer, Vincent van Gogh und Frida Kahlo sieht. Diese Künstlerinnen und Künstler nutzten ihr eigenes Bild, um persönliche Kämpfe, Bestrebungen und existenzielle Fragen auszudrücken, und trugen so zur kulturellen Bedeutung des Porträts bei, das mehr als nur ein physisches Abbild ist, sondern eine psychologische und emotionale Erkundung darstellt (Pointon, 1993).



Abb. 15: Caravaggio: Narziss.

Das Porträt bietet auch den Vorteil, dass sich der Betrachter emotional dazu positionieren kann. Anders als zum Beispiel bei Naturfotografien wird ein Porträt von einer emotionalen Reaktion begleitet. Dies kann sogar zu einer Identifikation mit der Figur führen, insbesondere bei durchschnittlichen und einfachen Porträts. (McCloud, 2001).

Im Zusammenhang mit KI-generierter Kunst bietet das Porträt eine faszinierende Gegenüberstellung zwischen der menschlichen Einzigartigkeit, die traditionell in Tausenden von Porträts (die auch Teil des Trainingsdatensatzes sind) festgehalten wird, und der künstlichen Kreativität von Maschinen. Sowohl die emotionale Tiefe als auch die symbolische Bedeutung des Porträtgenres machen es zu einer überzeugenden Fallstudie für die Erforschung der Möglichkeiten und Grenzen der KI-gesteuerten Kreativität.



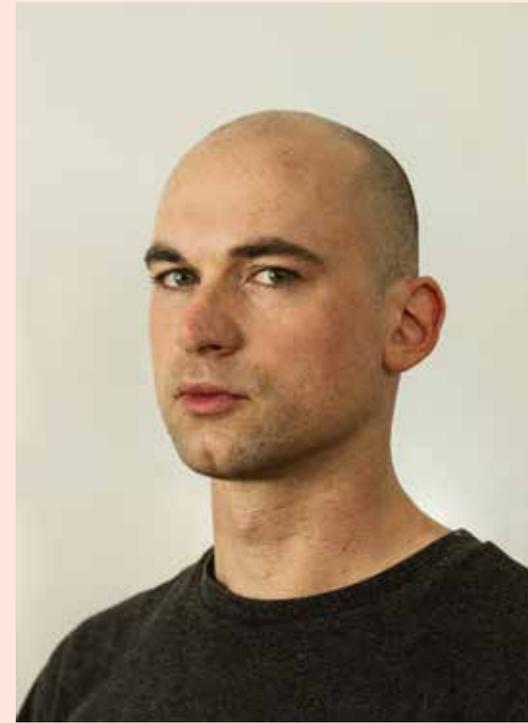
Abb. 16: Der zeitgenössische Künstler Bryan Lewis Saunders zum Beispiel verwendet Selbstporträts, um die Auswirkungen verschiedener Drogen auf seinen Körper, seine Wahrnehmung und letztlich sein kreatives Schaffen zu dokumentieren.

## 4.1 Automatisierung der Kreativität: Generative Entwurfstechniken

Diese Arbeit ist explorativ und erhebt daher keinen Anspruch auf Vollständigkeit. In Anlehnung an den kreativen Raum, wie ihn Thomas und Martin Poschauko in ihrer Arbeit „Nea Machina“ (2013) beschreiben, habe ich ein (Selbst-)Porträt und einen Prompt (zu Deutsch „Eingabeaufforderung“, die englische Bezeichnung für die Textbeschreibung des zu generierenden Bildes) gewählt, die beide als **formale Konstanten** des kreativen Experiments dienen. Das Selbstportrait dazu stammt aus einer Fotoserie, die ich zu Beginn der Pandemie 2020 erstellt habe. Diese beiden Elemente können inhaltlich variieren, wobei das generative Selbstporträt, und damit die ungefähre Pose und der ungefähre Inhalt, gleich bleiben. Das Experiment basiert also auf einer Reihe von Porträts, die nach bestimmten Vorgaben wie Gestaltungstechniken und Gestaltungsprinzipien aufgebaut sind. Ähnlich wie Butler, Holden, & Lidwell (2010) habe ich unterschiedliche Prinzipien gesammelt, die in kreativen Arbeiten immer wieder vorkommen oder eine Rolle gespielt haben. Ich nutze diese Liste, um die Anwendungsmöglichkeiten von diesen Kreativitätstechniken im Latent Space zu erkunden. Diese Liste ist nicht abgeschlossen, sondern gilt nur als grobe Orientierung für neue Sichtweisen und gestalterische Experimente. Gleichzeitig erlaube ich mir, außerhalb der Serie zu experimentieren, um auf zusätzliche Ideen zu kommen, wenn ich den Inhalt einer Idee vorerst ausgeschöpft habe.

Ich nutze für meine Experimente das eingangs beschriebene Programm Stable Diffusion für die Bildgenerierung und ChatGPT für textuelles Arbeiten. Prompts und weitere Einstellungen werden jeweils zu den Experimenten beschrieben. Die Modelle der KI-Programme haben sich im Laufe des Experiments immer wieder geändert oder geupdatet – damit kommt es auch zu einem veränderten Latent Space, da die Modelle sich hinsichtlich ihrer Trainingsdaten unterscheiden können. Doch selbst wenn sich die Experimente dadurch stilistisch minimal unterscheiden, so bleiben die Techniken, die angewendet werden, dieselben.

Dieses Vorgehen wird von Poschauko und Poschauko auch als Kreativraum mit "Hintertür" beschrieben (Poschauko & Poschauko, 2013, S. 13)



Das **Portrait** ist hochkant, ebenso wie ein Großteil der Experimente. Da einige Modelle unterschiedlichen technische Anforderungen haben, werden unterschiedliche Auflösungen von 512x512 Pixel, 1024x768 Pixel und 1024x1024 Pixel als mögliche Bildprompts (siehe dazu Seite 42) vorbereitet. Für einige Experimente werden gesonderte Vorlagen erstellt, wie beispielsweise Raster oder Ausschnitte.

Dieser Part kann in der Regel auch weggelassen werden, da die meisten Modelle auf überwiegend fotografischem Material trainiert wurde und die Bildästhetik daher eher einer Fotografie entspricht.

Der **Prompt** ist die Bildbeschreibung, mit der ich ein ähnliches Bild wie das obige erzeugen kann. Dieser Prompt besteht aus einem Medium, einem Subjekt, einem Kleidungsstück und einer Umgebungsbeschreibung. Für einige Experimente wird nur das Subjekt verwendet um stilistische Einflüsse zu minimieren.

A photo of  
a young bald man,  
black t-shirt,  
white background

Zusätzliche Infos sind optional, und werden eingesetzt je nachdem wie sehr ich mich an das Ausgangskonzept halten möchte.

### 4.1.1 Technische Einstellungen von Diffusionsmodellen

Die folgenden Beschreibungen der Einstellungen und Kontrollmöglichkeiten orientieren sich an Andrew Wongs (2024a) Beschreibung des Diffusionsprozesses, es werden im Einzelfall weitere Quellen erwähnt.

#### Diffusionsschritte (Steps)

Stable Diffusion ist im Grunde ein Text-Bild-Modell, das Bilder auf der Grundlage eines Prompts erzeugt. Es gehört zu einer Klasse von KI-Modellen, die als Diffusionsmodelle bekannt sind. Diese werden auf großen Datensätzen trainiert, bei denen jedes Bild durch einen Prozess, der als Forward Diffusion bezeichnet wird, allmählich in (Pixel-)Rauschen übergeht. Das Modell lernt, diesen Prozess umzukehren (*Reverse Diffusion*), um das Originalbild zu rekonstruieren. Wenn das Modell diese Umkehrung beherrscht, kann es aus zufälligem Rauschen neue, kohärente Bilder erzeugen, die den Trainingsdaten ähneln.

Um ein Bild zu erzeugen, beginnt Stable Diffusion mit einem völlig zufälligen Muster in einem latenten Raum. Das Modell verwendet dann den Prozess der umgekehrten Diffusion, um Schritt für Schritt das Rauschen aus diesem anfänglichen Muster zu entfernen, so dass allmählich ein Bild entsteht, das dem vorgegebenen Prompt entspricht. Da es sich bei Stable Diffusion um einen **bedingten** Bildgenerator handelt, verwendet es den Prompt, um diesen Entrauschungsprozess zu steuern. Auf diese Weise kann das Modell je nach Prompt unterschiedliche Bilder aus demselben Ausgangsrauschen erzeugen.

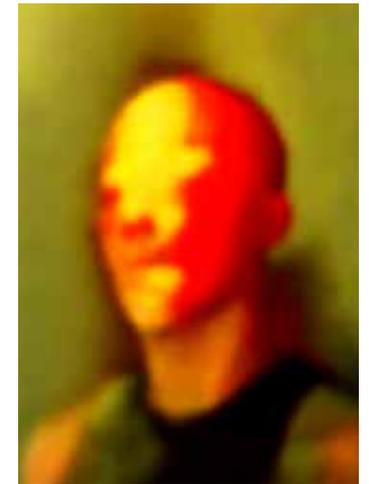
Das Generierungsverfahren bei Stable Diffusion umfasst einen Algorithmus, der steuert, wie das Rauschen bei jedem Schritt entfernt wird. Zu Beginn werden größere Schritte unternommen, um die globale Struktur, wie Komposition oder Farbverteilung des Bildes festzulegen. Im weiteren Verlauf des Prozesses werden die Schritte kleiner, um kleinere Details, wie Texturen und Kanten, zu verfeinern. Diese schrittweise Verfeinerung bedeutet, dass sich das Bild in den ersten Schritten aus dem Rauschen herauszuheben beginnt, aber unscharf bleibt. Mit jedem Schritt wird das Bild klarer, und ab etwa 20 Schritten konzentriert sich das Modell nur noch auf die Verbesserung der feinen Details, anstatt die Gesamtkomposition zu verändern.

Durch die Anpassung von Parametern wie Diffusionsschritten (die die Anzahl der Iterationen zur Verfeinerung des Bildes

Unbedingte Bildgeneratoren erzeugen ein Zufallsprodukt. Die ersten Bildgeneratoren, wie die GANs oder VAEs waren daher schwerer zu steuern, da die Punkte im Latent Space mühsam gefunden werden mussten.

Viele mathematische Methoden des Entrauschungsprozesses führen dazu, dass das Modell nach etwa 20 Schritten konvergiert. Das bedeutet, dass sich die Ergebnisse stabilisieren und sich einem bestimmten Zustand oder einer bestimmten Verteilung annähern, die die zugrunde liegenden Daten gut repräsentiert. Es gibt auch Methoden, die sogenannten Sampler (siehe Seite 36), die bereits nach wenigen Schritten konvergieren. Der Latent Consistency Model Sampler kann Bilder nach bereits einem Schritt konvergieren.

Prompt: A photo of a young bald man, black t-shirt, white background Steps: {0-33}, Sampler: Euler a, CFG scale: 8, Seed: 3319583382, Size: 768x102



steuern) reichen die erzeugten Porträts von detaillierten, fast fotografisch anmutenden Ergebnissen bis hin zu höchst impressionistischen, verschwommenen Kompositionen.

### Seed

Das zufällige Rauschmuster, das in Stable Diffusion schrittweise entrauscht wird, wird zunächst mit einer Zahl initialisiert, die als Seed (englisch für Samen) bezeichnet wird. Dieser Seed ist nicht völlig willkürlich; er kann manuell festgelegt oder vom System zufällig generiert werden. Die Verwendung eines zufälligen Seeds ist vorteilhaft, wenn ich eine breite Palette von Bildern untersuchen will, die aus demselben Prompt generiert wurden, da dies zu einer gewissen Variabilität führt. Wenn ich jedoch einen bestimmten Seed festlege, kann ich reproduzierbare Bilder erzeugen. Diese Funktion ist besonders nützlich, wenn ich mit anderen Parametern experimentiere, leichte Änderungen an einem Prompt vornehme oder meine Arbeit mit anderen teilen möchte, die genau das gleiche Bild reproduzieren können. Der Hauptvorteil der Verwendung eines festen Seeds besteht darin, dass die Bilderzeugung mit demselben Prompt, denselben Parametern und demselben Seed jedes Mal genau dieselben Bilder erzeugt.

Diese Konsistenz ermöglicht es mir, mehrere ähnliche Variationen eines Bildes zu erstellen, kontrollierte Experimente durchzuführen oder sicherzustellen, dass bestimmte Ergebnisse zuverlässig reproduziert werden können.

In der Stable Diffusion-Community wurden einige Seeds identifiziert, die eine höhere Wahrscheinlichkeit haben, Bilder mit bestimmten Merkmalen zu erzeugen, z. B. bestimmte Farbpaletten oder Kompositionen. Die Kenntnis dieser Seeds ermöglicht es Nutzerinnen und Nutzern, gezieltere Ergebnisse zu erzielen, was bei der Suche nach bestimmten künstlerischen Ergebnissen oder der Nachahmung eines bestimmten Stils hilfreich sein kann (Inglewood, 2023).

Oben Rechts:

Prompt: A photo of a young bald man, black t-shirt, white background Steps: 20, Sampler: UniPC, CFG scale: 8, Seed: 849238882843, Size: 768x1024

Unten Links:

Prompt: A photo of a young bald man, black t-shirt, white background Steps: 20, Sampler: UniPC, CFG scale: 8, Seed: 100, Size: 768x1024

Unten Rechts:

Prompt: A photo of a young bald man, black t-shirt, white background Steps: 20, Sampler: UniPC, CFG scale: 8, Seed: 1, Size: 768x1024



### CFG-Scale

Einfach ausgedrückt ist die CFG-Scale (*Classifier Free Guidance scale*) oder Guidingscale ein Parameter, der steuert, wie eng der Bilderzeugungsprozess dem Prompt folgt. Die CFG-Scale stellt im Wesentlichen ein Gleichgewicht zwischen der Genauigkeit der Ausgabe und der kreativen Freiheit her, so dass die Benutzer genau festlegen können, wie wörtlich die KI den eingegebenen Text interpretiert. Bei höheren CFG-Werten hält sich das Modell strikter an den Prompt und produziert Bilder, die den angegebenen Details genauer entsprechen. Bei sehr hohen CFG-Werten können die Bilder jedoch zu "wörtlich" genommen werden, was zu potenziellen Problemen wie Übersättigung, Detailverlust oder Artefakten führen kann, insbesondere wenn der Prompt zu restriktiv ist oder nicht genügend beschreibende Details enthält.

Den optimalen CFG-Wert zu finden, ist oft eine Frage des Experimentierens und des Abwägens zwischen Promptgenauigkeit und Bildqualität. In der Regel wird ein CFG-Wert zwischen 7 und 11 für die meisten Anwendungsfälle als ideal angesehen, da er ein Gleichgewicht zwischen der genauen Einhaltung des Prompts und der Beibehaltung einer hohen visuellen Qualität darstellt. Wenn der CFG-Wert zu hoch eingestellt ist, kann das erzeugte Bild unter übermäßiger Schärfe oder Rauschen leiden; wenn er zu niedrig eingestellt ist, kann die Ausgabe zu weit von der beabsichtigten Eingabeaufforderung abdriften.

In dem mit Stable Diffusion durchgeführten Experiment wurden die Porträts durch Anpassen der CFG-Skala erzeugt, um den Einfluss des Prompts zu steuern. Mit einer höheren Skala produzierte die KI stark strukturierte, fotorealistische Porträts, während eine niedrigere Skala zu flachen, entsättigten Bildern führte. Eine zu hohe CFG-Skala sorgte für grelle, übersättigte Bilder.

Prompt: A photo of a young bald man, black t-shirt, white background Steps: 20, Sampler: UniPC, CFG scale: {1-30}, Seed: 100, Size: 768x1024



Es ist auch möglich das Programm zu hacken. Mit Pythonkenntnissen lässt sich der Quellcode von Stable Diffusion bearbeiten und es sind negative oder keine CFG-Werte möglich. Die dann generierten Bilder ignorieren den Prompt und es lassen sich dann zufällige Bilder erstellen. Ohne die CFG-Skala verhält sich Stable Diffusion wie ein **unbe-dingter** Bildgenerator (siehe Seite 28).

## Gewicht (Weight)

Im Kontext von Stable Diffusion und anderen Modellen des maschinellen Lernens sind Worteinbettungen (*Embeddings*) numerische Darstellungen von Konzepten – wie Wörtern, Sätzen oder Bildern –, die ihre Bedeutungen im Latent Space erfassen. Wie stark das Modell die Aufmerksamkeit auf bestimmte Embeddings im Prompt hervorhebt, kann mit Hilfe von Wortgewichtungen geändert werden.

Bei einem Prompt wie `(sunset:1.5)(beach:0.75)`, priorisiert das Modell beispielsweise „Sonnenuntergang“ (Gewichtung 1,5-fach) über „Strand“ (Gewichtung 0,75-fach), was zu einem Bild mit mehr Sonnenuntergangselementen, aber immer noch einigen Strandaspekten führt. Die Aufmerksamkeit heißt aber nicht notwendigerweise, dass ein Element größer wird, es kann auch lediglich stärker in den Vordergrund treten oder detaillierter werden.

Es gibt verschiedene Techniken zur Verwendung von Gewichtungen (*catboxanon, 2024*):

- Numerische Gewichtungen werden im Prompt angegeben, um die Aufmerksamkeit für bestimmte Wörter feiner zu steuern. `(dog:2)` verdoppelt den Fokus des Modells auf „Hund“, während `(cat:0.5)` den Fokus auf „Katze“ um die Hälfte reduziert.
- Mit multiplen Konzepten wie beispielsweise `((man:0.5), (woman:0.9), (frog:1.2))` werden die Konzepte im Bild gewichtet. Dabei ist `man` weniger im Vordergrund, `woman` aber mehr und `frog` bekommt die meiste Aufmerksamkeit.
- Konzeptmischung: `(man:1.5|woman:1.5|frog:1.5)`, verrechnet die Konzepte im Laufe der Bilderzeugung zusammen. Diese Technik kann verwendet werden, um Bilder mit Eigenschaften aller Konzepte zu erstellen.
- Konzeptwechsel und Mischung: Mit `[man:frog:0.25]` wird das Bild nach dem ersten Viertel der Diffusionsschritte von einem Bild eines Mannes zu einem Bild eines Frosches übergehen. Das bedeutet, dass sich die grobe Komposition (frühe Schritte) an `man` die Details (spätere Schritte) aber an `frog` orientieren. Diese Methode ist für die Erstellung von Hybridbildern geeignet.

Es können auch negative Wortgewichtungen verwendet werden, um den Einfluss eines Wortes zu verringern. Eine zu starke Gewichtung führt wiederum zu übersättigten Bildern oder verzerrten Merkmalen.

Ein Embedding besteht aus einem Tensor, also einem mathematischen Objekt im Latent Space. Dieser Tensor besteht aus 768 Zahlenwerten, die einzelnen Dimensionen entsprechen; dabei handelt es sich nicht um räumliche Dimensionen wie Höhe, Länge, Tiefe, sondern um mathematische Dimensionen. Ein Embedding ist also die mathematische Abstraktion eines (sprachlichen) Konzepts. Das Embedding `frog` ist beispielsweise über diesen Tensor codifiziert:

```
tensor([-0.0040, 0.0116, -0.0175, ..., 0.0014, 0.0149, 0.0066]).
```

Bei einer Gewichtung kann dieser Tensor einfach multipliziert oder mit anderen Tensoren verrechnet werden.

Die Syntax der Sonderzeichen wie `()`, `[]` oder `:` variieren je nach Userinterface. So verwendet Stable Diffusion `(frog:1.5)`, Cogniwerk `frog+++`, Midjourney `frog:1.5`. Eine einheitliche Schreibweise hat sich bislang noch nicht durchgesetzt.

Prompt: A photo of a young bald (man:{-8-8}), black t-shirt, white background, Steps: 20, Sampler: UniPC, CFG scale:8, Seed: 100, Size: 768x1024

Wenn ein Embedding zu stark negativ oder positiv gewichtet wird, dann kann es auch wieder zu Verzerrungen und Artefakten kommen, die künstlerisch eingesetzt werden können.

Oben Links: (man:-6.0)

Oben Rechts: (man:6.5)



Eine Konzeptmischung funktioniert in der Regel nur so gut wie das Datenset es auch hergibt. Eine Mischung aus `man`, `woman` und `frog` führt zu einem Gruppenfoto, da die Konzepte eher mit *einzelnen Entitäten* assoziiert sind. Mit anderen Worten, beispielsweise `man`, `makeup` and `beard` ergeben sich andere Ergebnisse, weil diese Embeddings mit *Eigenschaften* assoziiert werden.

Prompt: A photo of a young bald ((man:0.5), (woman:0.9), (frog:1.2)), black t-shirt, white background, Steps: 20, Sampler: UniPC, CFG scale: 8, Seed: 100, Size: 768x1024



Eine Mischung aus `man`, `woman` und `frog` führt dagegen zu einer weiblich gelesenen Person mit leichten Falten, Geheimratsecken, größerer Nase und großen Augen, sowie einem Frosch T-Shirt. Wie der Prompt interpretiert wird hängt vom Seed ab und kann daher von Seed zu Seed variieren.

Prompt: A photo of a young bald (man:1.5|woman:1.5|frog:1.5), black t-shirt, white background, Steps: 20, Sampler: UniPC, CFG scale: 8, Seed: 100, Size: 768x1024

Bei dem Konzeptwechsel wird für fünf Schritte `man` generiert, wechselt aber dann für die verbleibenden 15 Schritte zum Konzept `frog`. Dadurch werden die Details dann in Richtung `frog` gelenkt.

A photo of a young bald [man:frog:0.25], black t-shirt, white background, Steps: 20, Sampler: UniPC, CFG scale:8, Seed: 100, Size: 768x1024



## Sampler

Sampler spielen eine entscheidende Rolle bei der Bilderzeugung, da sie bestimmen, wie das Pixelrauschen aus dem anfänglichen Zufallsinput entfernt wird, um ein klares und kohärentes Bild zu erzeugen. Da Sampler **mathematische Methoden** zur Rauschentfernung sind, wurden in den letzten Jahren mehrere Ansätze parallel entwickelt, die sich sowohl auf die Geschwindigkeit der Generierung als auch auf die Qualität des Endprodukts auswirken. Die Sampler unterscheiden sich in ihrem Ansatz und ihrer Leistung (Wong, 2024b):

- **Euler** und **Heun** sind klassische Methoden zur Lösung von Differentialgleichungen, wobei Euler schnell, aber weniger detailliert ist und Heun mehr Genauigkeit bei geringerem Tempo bietet.
- **LMS** und **PLMS** verbessern die Genauigkeit durch Mittelung der Schritte, können aber bei niedrigeren Schritten Rauschen verursachen.
- Die **DPM**-Familie, einschließlich DPM, DPM2 und DPM++, reduziert die Anzahl der Schritte, die für eine qualitativ hochwertige Ausgabe erforderlich sind, wobei DPM++ ein effizientes Gleichgewicht zwischen Geschwindigkeit und Genauigkeit herstellt.
- **DDIM** liefert schnell hochwertige Bilder, erfordert aber mehr Schritte für optimale Ergebnisse.
- Ancestral Sampler wie **Euler-A** bringen Zufallsrauschen in den Prozess und erzeugen kreative, aber weniger stabile Ergebnisse.
- Der neuere **UniPC**-Sampler kombiniert eine effiziente Vorhersage und Korrektur, um mit wenigen Schritten hochwertige Ergebnisse zu erzielen.
- **LCM** ist ebenfalls ein neuer Sampler, der mit nur wenigen Schritten auskommt und ein Bild bereits nach 1-4 Schritten generieren kann. Dieser wird viel für Echtzeitanwendungen genutzt.

Die Wahl des richtigen Samplers hängt von dem gewünschten Gleichgewicht zwischen Geschwindigkeit und Bildqualität ab, kann aber auch als bewusste Entscheidung für eine bestimmte Ästhetik gewählt werden.

Euler und Heun sind Eigennamen, die restlichen Namen sind Abkürzungen.

LMS: Linear Multistep  
 PLMS: Pseudo-Linear Multistep  
 DPM: Denoising Probabilistic Model  
 DDIM: Denoising Diffusion Implicit Model  
 UniPC: Unified Predictor-Corrector  
 LCM: Latent Consistency Model

Prompt: A photo of a young bald man, black t-shirt, white background, Steps: 20, Sampler: {auswahl}, CFG scale: 8, Seed: 4829334432, Size: 768x1024

Da der LCM Sampler mit wenigen Schritten auskommt und eine geringere CFG Scale braucht sind die Werte für diesen Prompt anders: Steps: 4, CFG scale: 1,5



Euler



Heun



LMS



PLMS



DPM



DDIM



Euler A



UniPC



LCM

### Entrauschungsstärke (Denoising Strength)

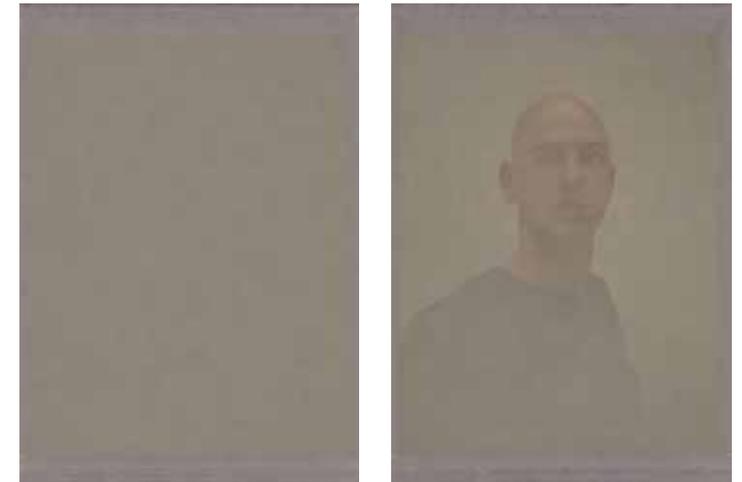
Mit der Denoising Strength kann die Stärke des Rauschentfernung während der Bilderzeugung beeinflusst werden. Dieser Parameter reicht von 0 bis 1: Bei einem Wert von 0 bleibt das Ursprungsrauschen erhalten, während ein Wert von 1 die Eingabe vollständig umwandelt und ein völlig neues Bild erzeugt. Anstatt mit zufälligem Rauschen zu beginnen, kann auch ein Bild als Prompt genommen werden. Bei einer Denoising Strength von 0 bleibt das Eingabebild dann intakt, während bei einem Wert von 1 ein neues Bild erzeugt wird. Mit Zwischenwerten lassen sich Mischungen aus Zufall und Vorgabe erzielen (Wong, 2024c). Daraus haben sich mehrere Anwendungsmodi entwickelt: Image-to-image (siehe auch Seite 42) nutzt diesen technischen Umstand direkt, um mittels Bildprompts die Farbverteilung, aber auch die Komposition zu beeinflussen. Dank spezieller Sampler wie dem LCM, die es ermöglichen, in nur wenigen Diffusionsschritten ein detailliertes Bild zu erzeugen, sind Anwendungen entstanden, bei denen im Inputlayer „gemalt“ wird und das Ergebnis im Outputlayer in Echtzeit sichtbar ist.

Beim Inpainting werden nur die Bereiche eines Bildes markiert, in denen der Diffusionsprozess stattfinden soll. Somit lassen sich punktuelle Änderungen an einem Bild herbeiführen.

Prompt: A photo of a young bald man, black t-shirt, white background Steps: 20, Sampler: UniPC, CFG scale: 8, Seed: 100, Size: 768x1024

Denoising Strength: {0.0-1.0}

Das anfängliche Ursprungsrauschen ist immer „grau“ da es eine Normalverteilung von RGB Pixeln ist. Das bedeutet aber auch, dass Bilder niemals komplett schwarz oder komplett weiß sein können, da Stable Diffusion große Abweichungen von dieser Normalverteilung nicht zulässt. Mit einem Hack, dem sogenannten Noise Offset kann das Ursprungsrauschen Richtung Schwarz oder Weiß verschoben werden, was wesentlich kontrastreichere Bilder zulässt.



## 4.1.2 Erkundung des Latent Space

Im Zusammenhang mit Porträts lassen sich bestimmte Aspekte der Kreativität automatisieren, insbesondere die Kombinationen von Merkmalen. Die einfachste Art der Automatisierung besteht darin, eine Reihe von Daten durchzuspielen. Da wir mittels Texteingabe durch den Latent Space navigieren können, lassen sich Konzepte oder Eigenschaften wie Material, Haarfarbe oder Ähnliches sprachlich in einer **Variable** definieren und verschiedene Versionen erzeugen: Prompt: A portrait of a person made out of {material}.

Dabei kann die Sammlungsvariable {material} dann alle möglichen Materialien wie Leder, Glas, Metall oder Pappe enthalten. Bei der Verwendung mehrerer Variablen kann dieses Konzept mit Hilfe der einfachen Kombinatorik erweitert werden. Es gibt allerdings noch weitere Möglichkeiten, den Latent Space kontrolliert zu erkunden. Im Folgenden werden die gängigsten vorgestellt.

### Text-Prompt

Ein Prompt ist die Art und Weise der Kommunikation mit einer Maschine – und bei der Bilderzeugung im Wesentlichen eine Form der **Bildbeschreibung**. Mit der Integration von CLIP wurde die Verwendung von Text-Prompts zur gebräuchlichsten Methode zur Steuerung von bildgenerierenden KI Modellen.

CLIP (*Contrastive Language-Image Pretraining*) ist ein von OpenAI entwickeltes Modell, das visuelle Konzepte aus der natürlichen Sprache lernt. Es funktioniert, indem es Bilder mit Text koppelt und das Modell darauf trainiert, Beschreibungen mit den richtigen Bildern zu verknüpfen. CLIP ermöglicht Aufgaben wie die Bilderzeugung, Klassifizierung und Bildbeschreibung, indem es Sprache und Bilder zusammen versteht (*Radford et al., 2021*). Da der überwiegende Teil des Datenmaterials, das für das Training verwendet wurde, aus englischen Bild-Text-Paaren besteht, ist es sinnvoll, in englischer Sprache zu prompten, um vorhersehbare Ergebnisse zu erhalten.

Diese als „Prompt-Engineering“ bezeichnete Praxis entwickelte sich ab 2022 schnell zu einem ausgefeilten Handwerk. Ein typischer Prompt ist eine detaillierte und präzise Beschreibung eines Bildes und besteht aus mehreren Embeddings. Jedes Embedding hat ein Gewicht von 1, aber je früher es im Prompt erscheint, desto mehr Einfluss hat es.

Daher folgen die Aufforderungen oft einer allgemeinen Syntax wie „Beschreibung des Subjekts oder der Handlung, Nennung zusätzlicher Details, in einem bestimmten Stil“.

Unterschiedliche Trainingsmethoden haben zu Variationen bei der Erstellung von Prompts geführt, die von

Es gibt auch einige nicht englischsprachige Wörter, die CLIP auch zuordnen kann. Dabei handelt es sich aber in der Regel um international (beziehungsweise im Netz) häufig vertretene Wörter, wie beispielsweise „Meister“, „Kebab“, „Aam Aadmi“ (indisches Equivalent für „Max Musterman“, auch Name einer politischen Partei in Indien).

Beschreibungen in natürlicher Sprache bis hin zu schlagwortbasierten Ansätzen reichen. Eine gängige Technik besteht darin, bestimmte Schlüsselwörter zu verwenden, um einen gewünschten Stil zu erreichen. So verbessern beispielsweise Begriffe wie „Hires“, „8K“ und „Schärfe“ die Bildqualität, während die Erwähnung von Kameramodellen bestimmte ästhetische Qualitäten hervorrufen kann, die mit diesen verbunden sind. Allerdings lassen sich keine exakten visuellen Eigenschaften wiedergeben: Ein Prompt kann das Thema zwar definieren, hat aber nur begrenzte Kontrolle über die Komposition und die Reproduzierbarkeit feinerer Details.

Es haben sich zusätzlich „Wortklassen“ ausgebildet, so dass ein Prompt zielgerichteter strukturiert werden kann: Mit einer Syntax wie {format} of {subject term}, a {solidifier}, {perspective}, by {artist:weight} | by {artist:weight} / in the style of {style}, {variant}, {booster}, {exclusion} kann ein Prompt näher spezifiziert werden. Diese Wortklassen haben aber keine Entsprechung im Modell sondern lediglich als Orientierung zur Beschreibung eines optimalen Prompts. Mit dem Aufkommen von Large Language Models (LLMs) wie ChatGPT hat es sich allerdings zunehmend etabliert, die grobe Struktur zu definieren und das LLM den optimalen Prompt generieren zu lassen.



Prompt: A statue of a young bald man made of Polycarbonate, Steps: 20, Sampler: DPM++ SDE Karras, CFG scale: 8, Seed: 2757207243, Size: 768x1024



Prompt: A statue of a young bald man made of Metal, Steps: 20, Sampler: DPM++ SDE Karras, CFG scale: 8, Seed: 2757207243, Size: 768x1024



Prompt: A statue of a young bald man made of Plywood, Steps: 20, Sampler: DPM++ SDE Karras, CFG scale: 8, Seed: 2757207243, Size: 768x1024

Für diese Reihe wurde zusätzlich ein ControlNet (siehe Seite 43) verwendet, um die Position des Kopfes zu stabilisieren und lediglich die Materialität auszutauschen.

### Bild-Prompt

Ein Bildprompt leitet ein KI-Modell dazu an, ein vorhandenes Bild zu transformieren oder Variationen zu generieren, wobei sowohl visuelle Eingaben als auch Text kombiniert werden. Im Gegensatz zu reinen Textprompts, bei denen die Beschreibungen lediglich die Erstellung steuern, bieten Bildprompts eine Grundlage für Farbe und Komposition, wobei das Modell das Bild entsprechend dem Text anpasst oder verbessert. Da das Basisbild die Struktur und den Aufbau vorgibt, kann der Textprompt Änderungen in Stil, Details oder Inhalt anweisen. Begriffe wie „Aquarell“ oder „Mixed-Media-Collage“ ändern den visuellen Stil, während die Denoising Strength den Grad der Veränderung steuert und ermöglicht entweder subtile Anpassungen oder dramatische Veränderungen. Ein Bildprompt kann also als visueller Anker genutzt werden. Das Vorgehen beim Inpainting ist ähnlich, nur dass es sich wie oben erwähnt auf einen zuvor festgelegten Bereich beschränkt.

Prompt: A photo of a young bald frog, black t-shirt, white background, Steps: 20, Sampler: DPM++ 2M Karras, CFG scale: 8, Seed: 100, Size: 768x1024

Denoising Strength: 0,7



### ControlNet

ControlNet ist eine Methode, die in der KI-Bilderzeugung verwendet wird, um zusätzliche Kontrolle über bestimmte Aspekte der Ausgabe zu erhalten, indem strukturelle Informationen, wie z. B. Pose, Tiefe oder Linienzeichnungen, in den Erzeugungsprozess einbezogen werden. Im Gegensatz zu regulären bild- oder textbasierten Prompts ermöglicht ControlNet dem Modell, bei der Bilderzeugung detaillierten Vorgaben zu folgen (z. B. *handgezeichnete Skizzen, menschliche Posen oder 3D-Daten*), wodurch eine genauere Kontrolle über die Zusammensetzung und Struktur der Ausgabe gewährleistet wird.

In ControlNet fungiert der visuelle Input (z. B. *eine Linienzeichnung oder eine Körperpose*) als Blaupause, die sicherstellt, dass das generierte Bild genau mit dieser Vorlage übereinstimmt. Der Prompt hingegen fügt Stil, Inhalt und feinere (*inhaltliche*) Details hinzu und ermöglicht sowohl kreative als auch strukturelle Anpassungen. So kann ich beispielsweise eine Pose vorgeben und ein Prompt wie **a futuristic robot in a forest** verwenden, um sicherzustellen, dass die generierte Figur die exakte Pose einnimmt und gleichzeitig die gewünschten inhaltlichen Elemente hinzufügt.

Diese Technik ermöglicht eine weitaus größere Kontrolle über das Ergebnis und verbindet die Präzision visueller Anleitungen mit der Kreativität von Prompts. Die Fähigkeit von ControlNet, sowohl die Struktur als auch den Stil beizubehalten, macht es ideal für Aufgaben, bei denen ein kontrolliertes oder exaktes Ergebnis entscheidend ist (*Wong, 2024d*)

Als Submodul hat ControlNet eigene Modelle (und eigene Einstellungen!), die selbst auch einen eigenen Latent Space aufweisen. Das bedeutet, dass die Art und Weise, wie ein Input interpretiert wird, sich auch zwischen unterschiedlichen ControlNets unterscheidet. Die Auswahl der gegenwärtigen ControlNets, sowie die Beschreibung derselben würde den Rahmen dieser Arbeit sprengen, da es bereits mehrere hundert Arten und Usecases an ControlNets gibt. Die gängigsten sind Linienzeichnungen (SoftEdge, Canny, Scribble oder Sketch, MLSD, Lineart) als Bildstütze, Körperposen (AnimalPose, OpenPose), Tiefen- oder 3D Daten (Depth, Normal) oder Referenzdaten (Reference, Revision, IP-Adapter, Instant-ID, T2I-Adapter, Photomaker). Darüber hinaus gibt es noch Sonderfälle, die eher Filtern gleichen, wie Recolor oder Blur, die je ein monochromes Bild mit Farbe versehen oder entschärfen (letzteres ist eine Form von Bildprompt, bei dem Farben und Formen in etwa übernommen werden).

Prompt: A futuristic robot in a forest Steps: 20, Sampler: DPM++ 2M Karras, CFG scale: 8, Seed: 100, Size: 768x1024

ControlNet Model: sdxl-scribble-Anyline, Weight: 1, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0, Guidance End: 1



### Training (Finetuning)

Das Finetuning, insbesondere mit LoRA (*Low-Rank Adaptation*), ist eine Technik, die es einem KI-Modell ermöglicht, sich an bestimmte Stile oder Inhalte anzupassen, ohne das gesamte Basismodell von Grund auf neu zu trainieren. Stattdessen ändert ein LoRA nur eine kleine Untergruppe von Parametern in den Embeddings, was es zu einem effizienteren und ressourcensparenden Ansatz für die Modellanpassung macht.

Da ein LoRA nur eine kleine Anzahl von Embeddings beeinflusst, ist diese Anpassung möglich, während die allgemeinen Fähigkeiten des ursprünglichen Modells erhalten bleiben. Mit unterschiedlichen LoRAs kann so zwischen verschiedenen fein abgestimmten „Persönlichkeiten“ des Modells gewechselt werden und somit einzigartige Konzepte in den Generierungsprozess übertragen werden. Es ist auch möglich mehrere LoRAs miteinander zu kombinieren, um mehrere einzigartige Konzepte miteinander zu mischen oder in ein Bild zu integrieren (Wong, 2024e). Der Latent Space des Basismodells wird durch das „dazuschalten“ um den Latent Space des LoRAs erweitert und abgeändert.

Prompt: Black and white drawing of a young bald man, black t-shirt, white background <lora:BenjaminBertram-Style:0.7>, Steps: 20, Sampler: DPM++ 2M Karras, CFG scale: 8, Seed: 100, Size: 768x1024

ControlNet Model: sdxl-scribble-Anyline, Weight: 1, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0, Guidance End: 1



Prompt: a black and white drawing of a cat <lora:BenjaminBertram-Style:0.7>, Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 2679142727, Size: 1024x1024



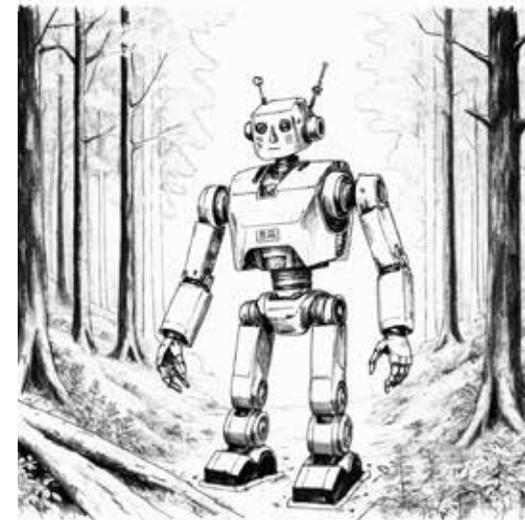
Prompt: a black and white drawing of a velociraptor <lora:BenjaminBertram-Style:0.7>, Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 2679142727, Size: 1024x1024



Unterschiedliche Zeichnungen, die in den letzten 10 Jahren entstanden sind



Generiertes Bild mit einem auf diese Zeichnungen trainierten LoRA



Prompt: a black and white drawing of a futuristic robot in a forest <lora:BenjaminBertram-Style:0.7>, Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 2679142727, Size: 1024x1024



Prompt: a black and white drawing of an umbrella <lora:BenjaminBertram-Style:0.7>, Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 2679142727, Size: 1024x1024

## 4.2 Flexible Systeme und Multivariabilität

Ein systematisches Nachdenken über Gestaltungselemente bringt mich dazu, neben den technischen Einstellungen das Potential des Latent Space weiter zu erforschen. Ein Porträt kann, wie jedes Design, als flexibles System betrachtet werden, in dem verschiedene Komponenten dynamisch zusammenwirken, um das endgültige Werk zu schaffen. Diese Denkweise deckt sich gut mit dem Konzept der flexiblen Designsysteme nach Lorenz (2024):

*Flexible Visual Systems (FVS) are design frameworks that emphasize adaptability and modularity. Unlike traditional design systems that may be rigid and static, FVS are dynamic and can evolve over time. They allow designers to create visual identities that are both consistent and versatile, ensuring that the core elements of a brand or project remain recognizable while being easily adaptable to different formats and contexts.*

Generative KI zeichnet sich dadurch aus, dass visuelle Ideen schnell erforscht werden können. Das ermöglicht uns statt eines Denkens in Beschränkungen ein Denken in gestalterischer Fülle: Anstatt uns auf technische oder ressourcenbezogene Möglichkeiten zu konzentrieren, können wir eine breite Palette von Konzepten ausprobieren. Dieser Ansatz lässt sich auf verschiedene Bereiche anwenden, auch auf Porträts.

### 4.2.1 Das Porträt als flexibles System

Ein Porträt kann, wenn es als flexibles System betrachtet wird, als eine Darstellung des Aussehens einer Person beschrieben werden, die sich in der Regel auf das Gesicht konzentriert, aber potenziell auch auf die gesamte Figur. Die Gestaltung eines Porträts kann in verschiedene Schlüsselkomponenten unterteilt werden, die als Teil dieses Systems zusammenwirken:

- **Gesichtszüge:** Die Augen, die Nase, der Mund und die gesamte Gesichtsstruktur stehen bei den meisten Porträts im Mittelpunkt.
- **Ausdruck:** Der emotionale Zustand oder die Persönlichkeit des Motivs wird durch seinen Gesichtsausdruck und seine Pose vermittelt.
- **Komposition:** Wie das Motiv eingerahmt ist, einschließlich Hintergrund, Beleuchtung und Perspektive.
- **Stil:** Der künstlerische Ansatz, der von fotorealistischen bis zu abstrakten Interpretationen reicht. **Beleuchtung:** Der Einsatz von Licht und Schatten, um Merkmale zu betonen

und Stimmungen zu erzeugen.

- **Farbpalette:** Die Auswahl der Farben, die Hauttöne, Kleidung und Hintergrundelemente darstellen.
- **Textur:** Das Rendern von Haut-, Haar- und Kleidungstexturen, die je nach Medium und Stil variieren.

Neben diesen visuellen Elementen spielt das Format eine entscheidende Rolle bei der Wahrnehmung eines Porträts. Das Format umfasst die Ausrichtung, die Größe, das Medium, das Seitenverhältnis, den Ausschnitt und die Perspektive, die jeweils Einfluss darauf haben, wie das Motiv dargestellt wird und wie der Betrachter sich mit dem Porträt auseinandersetzt.

Ebenso wichtig ist der Kontext, in dem das Porträt erstellt und ausgestellt wird. Der Kontext kann durch kulturelle Einflüsse, die beabsichtigte Verwendung (z. B. ein Passfoto oder ein Ausstellungsstück), den Zeitpunkt der Verwendung und tiefere symbolische oder metaphorische Bedeutungen systematisiert werden. Durch die Berücksichtigung dieser Faktoren wird die Gestaltung eines Porträts zu einem vielschichtigen

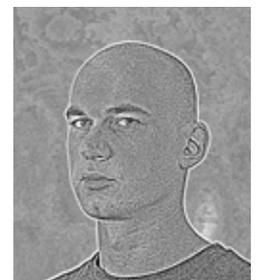
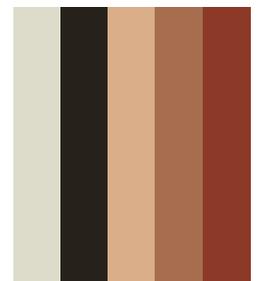


Emotion of this detected face is most likely to be neutral, with the highest possibility of 92.629%



Der überwiegende Anteil an Bildanalyse kann inzwischen auch durch Algorithmen erledigt werden. System wie Face Segmentation, Emotion Recognition oder Feature Extraction können Bilder komplett autonom analysieren. ChatGPT greift inzwischen auf eine Reihe dieser Subsysteme zurück.

Das Bild zeigt einen fotorealistischen Stil. Weiches Licht betont subtil die Gesichtszüge. Der neutrale Hintergrund lenkt den Fokus auf das Gesicht und schafft eine sachliche Atmosphäre.



System, das sich an verschiedene Kontexte und Verwendungszwecke anpassen lässt.

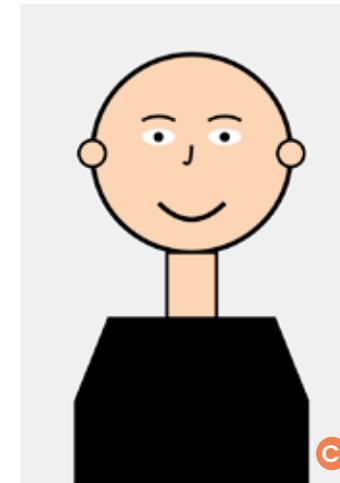
#### 4.2.2 Ein Porträt systematisch dekonstruieren

Ein Porträt kann von der Mikro- bis zur Makroebene beschrieben werden, wobei jede Ebene mit unterschiedlichen Inhalten gefüllt ist. Indem wir das Beschreibungssystem selbst zu einer Variable machen, können wir eine Metaebene schaffen, die neue Wege der Dekonstruktion eines Porträts zur Erhöhung der Variabilität ermöglicht. Zur Beschreibung eines Porträts können verschiedene Systeme verwendet werden, beispielsweise

- Aufzählungen von Mengen: (z. B. *zwei Augen, zwei Ohren, ein Mund, fünf Millionen Haare*).
- Strukturen: Beschreiben des Gesichts als eine Ansammlung und Kombination von Formen (*üblich im Zeichenunterricht*).
- Taxonomische Beschreibungen: Aufteilung von Komponenten wie Schädel, Muskeln und Nerven in Untergruppen.
- Skalen, Symmetrie und Proportionen: Beschreibung des Verhältnisses der Merkmale zueinander, z. B. die Größe der Ohren im Verhältnis zur Nase.
- Selbstähnlichkeit: Feststellung struktureller Ähnlichkeiten zwischen Teilen des Gesichts, z. B. zwischen den Augenlidern und dem Mund.
- Funktionelle Beschreibung: Die Augen sehen, die Ohren hören, die Nase riecht, die Haut spürt, der Mund schmeckt.
- Raumzeitliche Verortung: Portrait geschossen am 6.04.2020, 17:20 Uhr, 51°57'09.0" nördlicher Breitengrad 7°38'46.3" östlicher Längengrad, 64 Meter über NHN.
- Technische Einstellungen: 3456x5184 Pixel, 72 DPI, RGB, Canon EOS 550D, Linse: Canon EF 50mm f/1.8 II, Verschlusszeit 1/100, Brennweite 50 mm, ISO 400, Blende f/2,5.
- Motivbeschreibung: Benjamin Bertram, männlich, 34 Jahre, 182 cm, 79 kg, rasiert, trägt Kontaktlinsen, wohnhaft in Münster, römisch-katholisch, Versicherungsnummer 11210286B010.
- Arche- und Stereotypisch: archetypische (männliche) Stärke, Rationalität und Selbstbeherrschung, verkörpert durch kantige Gesichtszüge, stoischen Ausdruck und minimalistische Darstellung, stiller rationaler Denker.
- Framing: Das Portrait als Testobjekt für die visuelle Erkundung des Latent Space im Rahmen einer Masterarbeit am Fachbereich Design an der Hochschule für angewandte Wissenschaften in Hamburg.

Experimente, die aus der Vielfalt der Beschreibungsmöglichkeiten kommen:

- A Nachstellung von John Stezakers Kollagentchnik.
- B Namensbias: Das Embedding Achmed ist stark mit „Achmed the Dead Terrorist“ des amerikanischen Comedians Jeff Dunham verknüpft.
- C ChatGPT beschreibt ein Portrait aus Grundformen für ein SVG.
- D Nationalitäten führen zu mehr Diversität: A photo of a young bald Sudanese man
- E Zufällige Verteilung der Kontrollpunkte des HumanPose ControlNets.
- F Gewichtserhöhung der Embeddings young und bald.
- G Der Preprocessor für ControlNet Deptht.
- H Portrait mit Tierköpfen mit Inpainting.



### 4.2.3 Multivariable Meta Thinking

Diese Art des Arbeitens mit Variablen lässt sich in die Mathematik zurückverfolgen. So ist beispielsweise die Zahl „3“ ein willkürliches Symbol, das eine konkrete Menge darstellt, und der Buchstabe „x“ steht als Variable, die mehrere abstrakte Darstellungen wie das willkürliche Symbol der „3“, aber auch Wörter, Klänge oder Farben umfassen kann. Dieses variable Denken in kreativen Prozessen funktioniert als eine Art Meta-Symbol; ein flexibles Element, das mehr dynamische Möglichkeiten als konkrete Details zulässt.

Der Vorteil von variablen Systemen ist ihre unendliche Dynamik. Sobald die Regeln und Variablen definiert sind, entsteht ein Möglichkeitsraum, der alle potenziellen Lösungen innerhalb dieser Parameter enthält. Die Herausforderung besteht dann darin, die für den jeweiligen Kontext oder das künstlerische Problem am besten geeignete Lösung zu finden.

Analog zum Design Thinking bezeichne ich diesen Ansatz als **Multivariable Meta Thinking**. Er reduziert die Komplexität, indem er sich auf mehrere offene und abstrakte Variablen konzentriert und es ermöglicht, kreative Lösungen aus der Identifizierung dieser Variablen zu entwickeln.

„Multivariable“ bezieht sich auf die vielen austauschbaren Komponenten, während sich „Meta“ auf die übergreifende Ebene der Flexibilität und Abstraktion bezieht, die ihre Interaktion bestimmt.

Indem ich den Prompt multivariable formuliere – `a portrait in the style of {name} {epoch} {artistic technique}` – kann ich sofort eine breite Palette von unterschiedlichsten Portraits erzeugen und dabei traditionelle handwerkliche oder ressourcenbezogene Einschränkungen umgehen. Im Kontext der generativen KI bietet diese Art des Meta-Denkens ein enormes kreatives Potenzial. Wird ein Porträt als flexibles System mit Variablen definiert, können sofort unzählige Versionen erzeugt und mit Stilen, Formaten und Kontexten experimentiert werden, deren Erforschung von Hand wesentlich länger dauern würde.

Und selbstverständlich kann ich auf der Metaebene auch unterschiedliche Beschreibungsarten wählen und somit auch von der sprachlichen Seite alle möglichen Experimente machen: indem ich von der „optimalen“ Promptstruktur abweiche und mit zufälligen Silben, in Reimen, mit Emojis, oder in fremden Sprachen prompts, erforsche ich Bildwelten, die nicht an der Oberfläche des „klassischen“ Text-to-Image Workflows zu finden sind.

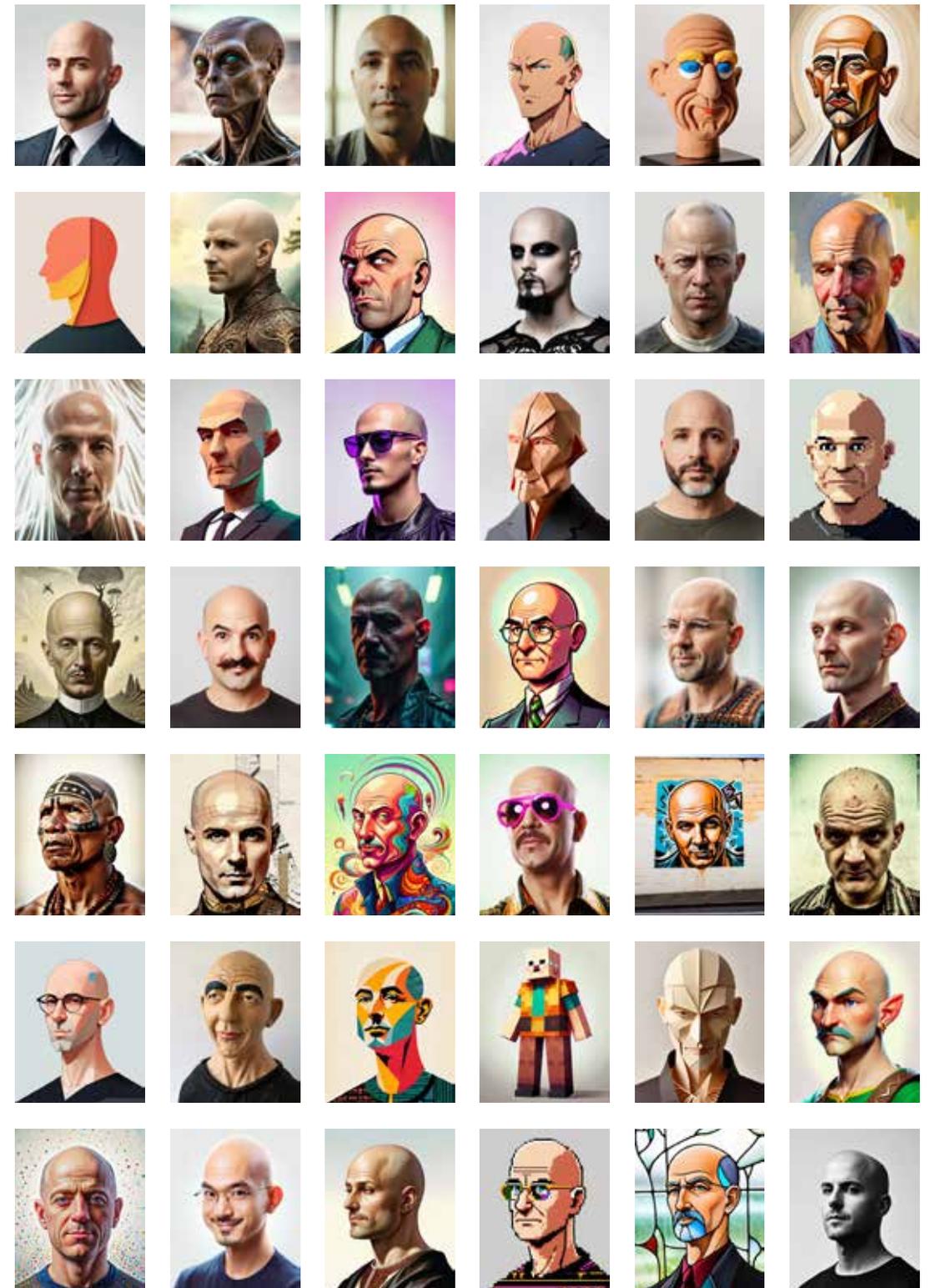
Der Programmier Andy Hunt prägte dazu auch den Ausspruch „Abstractions live longer than details“ (Venners, 2003).

Um beispielsweise möglichst diverse Menschen abzubilden, können viele Eigenschaften, die eine Person ausmachen, in Variablen abstrahiert werden:

A photo of a {human}, {religion}, {location}, {ability}, {social background}, {gender expression}, {education}, {physical health}, {fertility}, {mental health}, {class}, {occupation}, {hobbies}, {appearance}, {language}, {material status}, {ethnicity}, {race}, {age}, {gender identity}.

Möglicherweise kann ein multivariable Prompt der Zukunft dann auch so geschrieben werden, dass beim Briefing mittels KI bereits das Brainstorming automatisiert wird:

A {medium} for {target group}, {purpose}, {style}, etc ...



#### 4.2.4 KI-Bilder als Inbegriff der Remix Culture

In Anlehnung an Navas' (2012) Remix Culture kann generative KI Kunststile, Sujets, Techniken, Umgebungen, Medien, Formen und Farbschemata neu mischen und rekontextualisieren und so hybride Porträts schaffen, die klassische oder zeitgenössische Ansätze mit einer ständig wachsenden Zahl neuer Ästhetiken vermischen. Diese Form der gegenseitigen Befruchtung ermöglicht die Verschmelzung vergangener und gegenwärtiger Stile und eröffnet neue Wege für die künstlerische Erforschung. Generative KI ermöglicht es uns, mit flexiblen Systemen mehrere kreative Dimensionen gleichzeitig zu erforschen und Stil, Format und Kontext zu mischen, um neue Interpretationen von Porträts zu schaffen. Diese Interpretationen, die oft nie zuvor erdacht wurden, durchbrechen traditionelle Grenzen und führen neue, innovative Ausdrucksformen ein.

#### 4.3 Ko-Kreativität und Kollaboration von Mensch und Maschine

Das autonome Potenzial von KI-Systemen ermöglicht mir eine Form der Ko-Kreativität, bei der die Maschine sowohl als Werkzeug als auch als dialogische Partei bei der Arbeit im kreativen Prozess fungiert. Durch die Veränderung von Variablen in einem generativen Prozess kann ich die verbalen Fähigkeiten von LLMs nutzen, um Konzepte zu erforschen oder die richtigen Worte und Prompts zu finden. Besonders bei Tools wie ChatGPT wird der dialogische Ansatz deutlich. Als Chatbot agiert ChatGPT eher wie ein Gegenüber, das ich um Rat frage, oder als Assistenz, die mir Vorschläge recherchiert. Die Maschine bringt unerwartete Variationen oder Ideen ein, an die ich vielleicht zunächst nicht gedacht habe. Das endgültige Porträt wird so zu einem **gemeinschaftlichen Produkt** aus menschlicher Absicht und maschineller Berechnung (Davis, 2021).

Bei der KI-gesteuerten Kreation verlagert sich ein Großteil meiner Rolle in Richtung Kuratation. Und damit ist nicht die Auswahl aus den maschinell erzeugten Ergebnissen gemeint, sondern vor allem auch die Auswahl der Variablen. Es gibt auch Algorithmen, die Bilder ästhetisch bewerten können, aber diese basieren auf Bewertungen fachfremder Personen in einem sehr speziellen Kontext. Eine geschmackliche Kuratation kann daher zwar automatisiert werden, spiegelt aber nicht notwendigerweise meine Präferenzen wider.

Ein weiterer wichtiger Aspekt der KI-Kunst ist die Zusammenstellung und Auswahl des Trainingsmaterials. Als

Das Simulacra Aesthetic Captions Modell basiert auf einem Datensatz von über 238.000 synthetischen Bildern, die mithilfe von KI-Modellen wie CompVis latent GLIDE und Stable Diffusion aus mehr als 40.000 durch Nutzerinnen und Nutzer eingereichten Prompts generiert wurden. Die Teilnehmerinnen und Teilnehmer bewerteten die Bilder nach ihrem ästhetischen Wert auf einer Skala von 1 bis 10, um Triplets aus Bildunterschriften, Bildern und Bewertungen zu erstellen (Crowson, 2022).



Ich starte mein Experiment damit, dass ich ChatGPT nach einer Beschreibung frage:

Bitte beschreibe dieses Bild, aber nutze nur Wörter die mit „B“ anfangen.

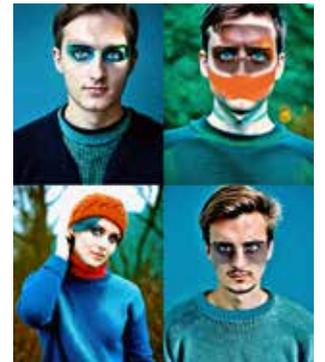
Die Antwort sieht interessant aus. Ich übernehme es als Prompt mit den Einstellungen Steps: 25, Sampler: DPM++ 2M Karras, CFG-Scale: 8 und lasse mir vier Bilder generieren.

Blick, blass, braune Bartstoppeln, blassgrüne Augen, blaugrauer Hintergrund, Bartlos, Bogenbrauen, blauer Baumwollpullover.



Die vier generierten Bilder zeigen mir eine Richtung. Die Bilder sind mir noch zu fotorealistisch. Ich will einen grafischeren Effekt und setze die CFG-Scale hoch auf 18 und die Steps runter 10.

Das zweite Bild sieht interessant aus. Ich nehme es als Referenzvorlage für ein ControlNet (Style-Referenz). Dazu nutze ich meinen Prompt a young bald man. Die Einstellungen behalte ich bei.



Das Ergebnis erinnert mich an verlaufende Portraits, wie die von Frank Auerbach oder Francis Bacon. Durch die hohe CFG-Scale bekomme ich Artefakte, aber ich bin neugierig, was passiert, wenn ich diese Portraits als Bildprompt nehme zusammen mit meinem Standardprompt a young bald man. Die Settings passe ich wieder an.

Das Ergebnis ist spannend, aber mir sind die Bilder zu pink. Ich nehme das vierte Bild und nehme ein Instruct-Pix2Pix Controlnet mit der Aufforderung make it green-orange. Da ich kein übersättigtes Bild haben möchte, setze ich die CFG-Scale auf 3.



Das erste Bild hat Ähnlichkeiten mit einer alten Überwachungskamera. Um den Effekt zu verstärken, skaliere ich das Bild zusammen mit dem Prompt Old CCTV footage of a young bald man, black t-shirt, white background, film grain, found footage auf die doppelte Auflösung hoch.

Das finale Bild. Technoästhetik? Malerisch? Digitale Collage? Und ChatGPT, was denkst du?

Das Bild wirkt surreal und digital verfremdet, erinnert an Cyberpunk und Identitätsfragmentierung. Es ist faszinierend und technisch interessant gestaltet!





in dem Maße, wie die Komplexität zunimmt, intensiviert sich auch die Tiefe der Erkundung, wodurch das Verständnis von gestalterischen Ideen grundlegend vertieft wird.

## 4.5 Was definiert KI-Ästhetik?

Zeitgenössische Bildgeneratoren simulieren traditionelle Medien wie Fotografie oder Malerei immer besser. Frühe Modelle wie VQGAN+CLIP hatten eine unverwechselbare „Glitch“-Ästhetik, die sich aus den technischen Möglichkeiten ergab. Mittlerweile sind generierte Bilder aber so realistisch, dass „echt“ und „unecht“ kaum noch zu unterscheiden sind.

Es liegt daher nahe, die KI-Ästhetik nicht mehr an ihren Schwächen, sondern an ihren Stärken zu messen. Neben technischer Schnelligkeit und Raffinesse liegt eine Stärke im Zufall und im assoziativen Arbeiten von KI-Modellen. Der Zufall ist etwas, das man auch in den Bildern von Jackson Pollock findet. Seine Drip Paintings werden oft als bewusste Erforschung von Chaos und Spontaneität beschrieben, die tief in der Körperlichkeit des Malprozesses verwurzelt sind.

Pollock nutzte seine berühmte „Tropftechnik“, bei der er Farbe auf eine flach auf dem Boden liegende Leinwand goss oder spritzte. Dabei bewegte er sich bewusst auf der Leinwand und kontrollierte den Fluss, die Viskosität und die Geschwindigkeit der Farbe, um typische Probleme der Fluidodynamik, wie unerwünschte Kräuselungen, zu vermeiden (Palacios et al., 2019). Obwohl Pollock oft betonte, dass seine Aktionen nicht bewusst geplant waren und er sich mehr auf den Malprozess als auf das Endergebnis konzentrierte (Ochmanek, 2016), ermöglichte ihm diese Technik, komplexe, geschichtete Muster zu erschaffen. Diese scheinbar chaotischen Muster wurden gezielt gestaltet, um Emotionen und Bewegung zu vermitteln, wobei Farbe und Linie unabhängig von der Form als ausdrucksstarke Elemente agieren.

Die Definition von KI-Ästhetik hängt davon ab, wie wir die algorithmische Zufälligkeit und die daraus resultierende Arbeitsweise bewerten. KI-generierte Kunst erschafft oft visuell faszinierende Bildwelten, die jedoch nicht die direkte emotionale und körperliche Beteiligung aufweisen, die Pollocks Werke kennzeichnet. Während Pollocks Zufälligkeit aus seiner körperlichen Interaktion mit dem Medium und einem tiefen, wenn auch unbewussten, technischen Verständnis entstand, ist die Zufälligkeit in der KI-Kunst meist ein Nebenprodukt maschineller Lernmodelle, die mathematische Räume erkunden. Ob KI-Kunst über ihren inzidentellen Charakter hinaus eine eigene ästhetische Identität entwickeln kann, hängt davon ab, inwieweit wir die zugrunde liegenden technischen Parameter und Algorithmen verstehen und gezielt beeinflussen können.

Abb. 17 oben links: VQGAN+CLIP: der zwischen 2021 und 2022 beliebteste Bildgenerator war eine der ersten OpenSource Antworten auf DALL·E.

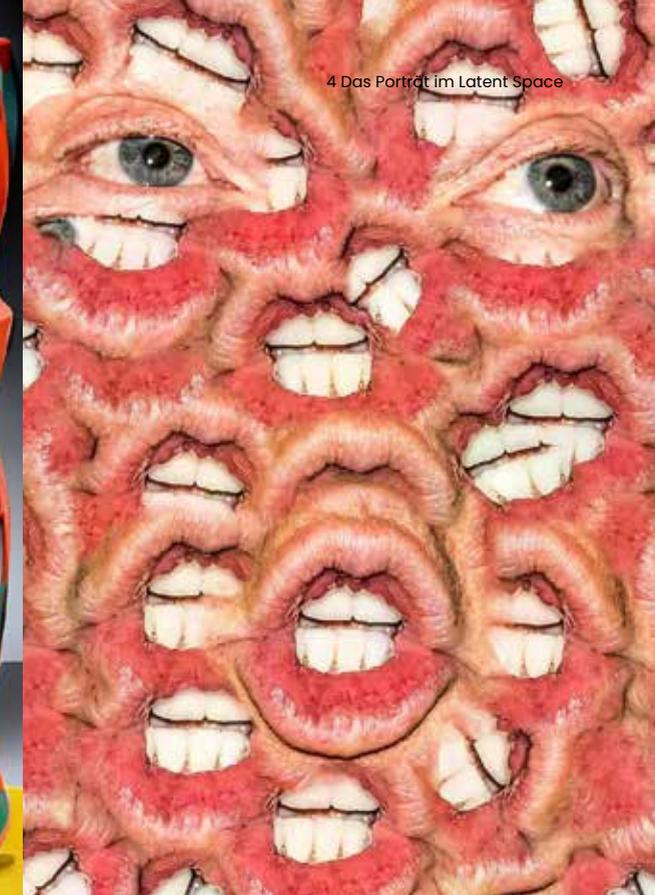
Abb. 18 oben rechts: Der dänische Creative Technologist Andreas Refsgaard experimentiert an der Schnittstelle zwischen Code und Design. Seine Werke nehmen häufig Erkennungsalgorithmen als Grundlage und sind oft auch interaktiv. In diesem Bild erkennt ein Algorithmus die wichtigsten Elemente im Bild und füllt den Rest des Bildes damit aus.

Abb. 19 unten links: Die Werke von Refik Anadol nutzen Wetterdaten und ähnliches und lassen sie durch KI auswerten um damit dynamische, sich bewegende „Data Paintings“ zu entwerfen. Installation im MoMa.

Abb. 20 unten rechts: Die KI-Künstlerin Junmoke Fernandez nutzt einen Preprocessor (DepthMap, zu deutsch Tiefenkarte) von Stable Diffusion als einzigartigen Effekt für Portraits.



Abb. 21: Ausstellungsstücke von Jackson Pollock im MoMa.



4 Das Porträt im Latent Space





Ähnlich wie ein Mensch, der Aquarelltechniken beherrscht und gezielt mit Salz, Speichel oder verschiedenen Untergründen arbeitet, um bestimmte Effekte zu erzielen, muss auch jemand, der kreativ mit KI arbeitet, ein tiefes Verständnis der technischen Aspekte der Bildproduktion haben. Nur mit diesem Wissen kann er die relevanten Parameter so anpassen, dass die kreative Leistung der Maschine von zufälliger Gestaltung in eine gezielte überführt wird.

Die Stärke der generativen KI liegt bereits in der mühelosen Erforschung von Zwischenwelten und Überlagerungen visueller Konzepte. Die Ästhetik der Vermischung, ähnlich der frühen Collage, wird durch fließende mathematische Übergänge in eine neue, natürlich wirkende Ästhetik überführt. Die generierten Bilder wirken glaubwürdig.

Darüber hinaus kann das unvollständige Weltverständnis von KI-Modellen außergewöhnliche Artefakte erzeugen: Mischwesen mit mehreren Gliedmaßen, die erscheinen und verschwinden, sowie unwirkliche Realitäten, die an Fieberträume oder unvollständige Erinnerungen erinnern. So entstehen traumartige Sequenzen und Zwischenwelten, die sich vor allem im Video zeigen lassen. Erst in der vierten Dimension kommt die multidimensionale und multivariable Natur des latenten Raums voll zur Geltung.

Die KI-Ästhetik wird maßgeblich durch die Trainingsdaten geprägt. Da nicht nur Kunstwerke, sondern jegliches Bild- und Textmaterial aus dem Internet genutzt wird, wird die Gesellschaft selbst Teil dieser Ästhetik. Es gibt keine einheitliche Ästhetik, sondern eine multiple, schillernde und oszillierende, die im besten Sinne metamodern ist. Gestalten mit KI bedeutet, mit dem kollektiven Bewusstsein zu arbeiten, indem Millionen veröffentlichter Bilder verwendet werden. Was Datenschutz- und urheberrechtlich bislang nicht abschließend geklärt ist, verleiht der KI-Ästhetik eine gesamtgesellschaftliche Dimension, die an Beuys' Konzept der „Sozialen Plastik“ (Lange, 2002) erinnert. Jedes Bild, das in den Trainingsdatensatz gelangt, beeinflusst diese Ästhetik. Der soziale Spiegel der Trainingsdaten verleiht der KI-Ästhetik eine transformative Kraft: Die sozialen Zerrbilder, die in den verschiedenen Embeddings stecken, fordern eine aktive Auseinandersetzung mit Trainingsdaten, Stereotypen und Vorurteilen. KI mag mich zum passiven Illustrator machen, aber ich werde zum aktiven Kurator meiner Variablen.

Linke Seite: Videostills aus einem Video, das mit AnimateDiff erstellt wurde. Mit Latent Space Walks, also „Spaziergänge im Latent Space“, lassen sich die traumartigen Zustände und die multivariable Natur besonders gut demonstrieren. Video abrufbar unter [https://www.youtube.com/watch?v=rk7Wx\\_dtUjY](https://www.youtube.com/watch?v=rk7Wx_dtUjY)

## 5 Abschluss und Ausblick

Das Ziel dieser Arbeit war eine eingehende Untersuchung des Latent Space innerhalb von Algorithmen zur Text- und Bilderzeugung und die Positionierung als neues Medium, das sich als einzigartiges Feld für kreative Erkundungen erweist. Verortet im theoretischen Rahmen der Metamoderne und Metakreativität zielten die praktischen Experimente darauf ab, die unterschiedlichen Merkmale von menschlicher und computergestützter Kreativität zu untersuchen und hervorzuheben. Durch die Erforschung des technologischen Konzepts des Latent Space von latenten Diffusionsmodellen wie Stable Diffusion und großen Sprachmodellen wie ChatGPT zeigte die Arbeit, wie Benutzerinnen und Benutzer Variablen manipulieren können, um neue künstlerische Stile und Ausdrucksformen zu erzeugen. Der Latent Space fungiert hier als mehrdimensionale Leinwand, die es ermöglicht, hybride Bilder zu schaffen, indem Merkmale auf innovative Weise kombiniert werden, was zu völlig neuen ästhetischen Erlebnissen führt.

Mit dieser Technologie geht jedoch eine spezifische technische Ästhetik einher, die aus ihrer Funktionsweise resultiert. Wenn Einstellungen so verändert werden, dass sie nicht das erwartete Bild erzeugen, entsteht eine einzigartige visuelle Sprache, die mit herkömmlichen digitalen Techniken nicht zu erreichen ist. Diese unvorhersehbare Dynamik führt zu einer ganz eigenen Form der künstlerischen Ausdruckskraft. Außerdem verändert die Arbeit mit KI auch die Rolle von Kreativen: Ähnlich wie die fotografierende Person, die Negative für das gewünschte Bild auswählt, muss die KI-kunstschaffende Person auch die generierten Bilder kuratieren. Denn im Wesentlichen sind KI-Werkzeuge von Zufälligkeiten geprägt, doch es gibt Methoden, die diesen Zufall im Generierungsprozess lenkbar machen und somit mehr Einfluss auf die Farbverteilung, die Komposition oder den Stil ermöglichen. Trotz dieser Techniken bleibt ein gewisser Grad an Zufälligkeit bestehen.

Außerdem kann die Ästhetik der KI-Kunst zwischen fotografischen und illustrativen Elementen schwanken, was der

KI-Kunst eine dynamische Qualität verleiht. Durch diese Dynamik kann die Ästhetik zwischen verschiedenen Stilen oder Stilmischungen wechseln, wodurch KI-generierte Illustrationen genuin prozesshaft sind, da sich immer wieder neue Variationen erstellen lassen.

Der KI-gestützte Arbeitsablauf kann jedoch dem der traditionellen Illustration ähneln: Zunächst wird die allgemeine visuelle Ästhetik erkundet, indem Prompts und Einstellungen angepasst und diese dann schrittweise verfeinert werden, bis das gewünschte Ergebnis erreicht ist. In der Nachbearbeitung erfolgt dann eine Hochskalierung. Da die maximale Auflösung für KI-generierte Bilder ohne Artefakte in der Regel 1024x1024 beträgt, muss das Bild anschließend hochskaliert werden, um produktreife Druckauflösung zu erreichen. Während dieses Skalierungsprozesses können je nach den Einstellungen zusätzliche Strukturen oder Elemente erzeugt werden, von denen einige unerwünscht sein können, so dass auch hier ein iterativer Ansatz erforderlich ist, um das „endgültige“ Ergebnis zu erzielen.

### 5.1 Kreativität neu formulieren

Die Arbeit stellt die traditionellen Vorstellungen von Kreativität in Frage, indem sie zeigt, dass maschinell erzeugte Prozesse wesentlich zur künstlerischen Innovation beitragen oder selbst als kreativ angesehen werden können. Die von KI-Modellen erzeugten Ergebnisse enthalten oft Elemente, die sogar für etablierte Kreative selbst überraschend sind, was auf eine Form der Kreativität hindeutet, die aus der Fähigkeit der Maschine erwächst, Informationen auf eine Weise zu synthetisieren, die der Mensch vielleicht nicht erwartet hätte. Diese Beobachtung führte zu dem Begriff der Metakreativität, der das Konzept der Kreativität als universelles Prinzip neu definiert, das nicht an die menschliche Spezies gebunden ist.

Es wurde auch in Frage gestellt, dass Kreativität als Konzept seine ursprüngliche Bedeutung als schöpferische Kraft verliert, die die Welt in *alt* und *neu* teilt und *nützliche* Ideen einbringt. Mit dem Anbruch des Zeitalters der Computational Creativity ist abzusehen, dass kreative Lösungen „nur“ optimale Lösungen in einem gegebenen Kontext sind und dass sie **brute force** berechnet werden können. Diese Beobachtung lässt sich auch durch kreative Strategien wie der „Grammatik des Erfindens“ von Prof. Spiess aus Aachen unterstützen. Diese Methode ist denkbar einfach, da sie lediglich bestehende Patente, bionische Prinzipien oder andere künstliche oder natürliche technische Lösungen räumlich zusammenbringt und dann verschiedene Lösungen spielerisch miteinander

Der Begriff „Brute Force“ (zu deutsch „Rohe Gewalt“) geht auf eine Hackerermethode zurück, bei der Probleme durch rein rechnerisches Ausprobieren aller Möglichkeiten gelöst werden (*freeCodeCamp, 2020*).

verbindet – was auch schlichtweg zufällig sein kann (*AGIT, 2007*). Diese Methode ist zwar für technische Probleme gedacht, kann aber auch für Design- oder illustrative Probleme eingesetzt werden: Der Akt der Inspiration ist ein Prozess des Mappings von Lösungen und der selektiven Übernahme von Designelementen, basierend auf persönlichen Vorlieben oder Erkundungen. Dieser Vorgang kann intuitiv und spielerisch wirken, ist letztlich jedoch nur eine Neukombination bestehender Ideen.

Mit der schöpferischen Kraft des Computers stellt sich auch die Frage nach Originalität und dem Recht auf geistiges Eigentum. Indem sie sich mit KI beschäftigen, benutzen Kreative nicht nur ein Werkzeug, sondern nehmen an einem ko-kreativen Prozess teil. Das bedeutet, dass der Computer ähnlich dem von Poschauko (*2013*) vorgeschlagenen Rahmenwerk einer kreativen Maschine zu einem ergänzenden Werkzeug für Gestalterinnen und Gestalter wird. Der Hauptunterschied besteht darin, dass der Algorithmus nicht nur Werte manipuliert, sondern auch über das Aussehen und das Gefühl einer visuellen Lösung „entscheidet“. Da dieser Entscheidungsprozess in einer Black Box stattfindet, habe ich als Designer nur die Möglichkeit, den kreativen Prozess anzustoßen und zu lenken, jedoch ohne die vollständige Kontrolle über jedes Detail bis hin zum letzten Pixel zu übernehmen.

Doch obwohl die KI in dieser Hinsicht eine gewisse Autonomie besitzt und sich in der technischen Ausführung auszeichnet, mangelt es ihr gleichzeitig daran, eine gute Idee mit Kontext oder Bedeutung zu entwickeln. Dies wird bei der Verwendung von LLMs noch deutlicher. Für die Recherche oder als ein Gegenüber kann es für Brainstorming, Verfeinerung oder Strukturierung großartig sein – aber nicht für die Idee selbst. Daher kann die Originalität von KI-Ergebnissen nur bis zu einem gewissen Grad berücksichtigt werden. Da die derzeitigen KIs jedoch noch keine Autonomie oder Kontextbewusstsein besitzen, ist das Potenzial für disruptive Ideen nicht gegeben, da selbst der Latent Space mit all seinen Möglichkeiten sich nicht selbst navigieren kann. Die KI ist daher gewissermaßen nur mein kreativer Co-Pilot.

Neben der Frage der Originalität und Urheberschaft im Rahmen metakreativer Schaffensprozesse stellt sich auch die Frage wie ein Original im Zeitalter offener und digitaler Kreativität aussehen kann. Der partizipative Charakter der KI-Kunst wird durch öffentlich gescrapte Trainingsdaten, Open-Source-Tools und kollaborative Plattformen nur noch verstärkt: KI-Communities teilen Code, Prompts und Workflows und demokratisieren so den Zugang zu kreativen Technologien. Plattformen wie [CogniWerk.ai](#), [Leonardo.ai](#) und [OpenArt.ai](#) ermöglichen es den Nutzerinnen und Nutzern nicht nur, direkt



Abb. 22: Prof. Spiess (zweite Person von links) beim Erfinden neuer Patente.

mit KI-Modellen zu arbeiten, sondern auch Prompts und Workflows zu kopieren und zu ändern oder dort fortzufahren, wo andere aufgehört haben. Dadurch verschwimmen die Grenzen zwischen Produzierenden und Konsumierenden weiter, sie werden zu Prosumentinnen und Prosumenten (*Toffler, 1983*). Eine solche Offenheit fördert eine kollektive Leistung, bei der viele an der Entwicklung von Kunstwerken und Stilen mitwirken. Diese Verschiebung erinnert an Roland Barthes' Konzept vom „Tod des Autors“, bei dem die singuläre Autorität von Autorinnen und Autoren über die Bedeutung schwindet und die Interpretation des Publikums an Bedeutung gewinnt (*Barthes, 2007*). In diesem Kontext ist ein Bild nicht mehr ein endgültiges Artefakt, sondern kann nach den Launen der Prosumierenden interpretiert, verändert, kopiert oder neu gemischt werden.

Kollaborative Kreativarbeit fördern nicht nur die Interaktion zwischen Mensch und Maschine, sondern auch die Zusammenarbeit zwischen Mensch und Mensch, da Fachkundige und Fachfremde gleichermaßen zu einem gemeinsamen kreativen Prozess beitragen. Dieses kollektive Engagement verwandelt den Akt der Schöpfung in eine soziale Praxis.

## 5.2 Bereiche für weitere Forschung

Mit der Entwicklung von Supercomputern wie Quantenrechnern könnte Computational Creativity einen Punkt erreichen, an dem sie sofort die perfekte Lösung für ein Problem liefern kann, sobald es definiert ist. Die Auswirkungen dieser übermenschlichen Meta-Kreativität wären ein spannender Ansatz für philosophische Untersuchungen, auch wenn dies aktuell noch spekulativ ist.

Das Themenfeld der KI und damit auch KI-Kunst ist mit vielen fragwürdigen Themen verbunden, die in dieser Arbeit nicht behandelt wurden. Dazu gehören gesellschaftliche Stereotypen in den Trainingsdaten, Auswirkungen auf den Arbeitsmarkt, Echokammern und Hyperpersonalisierung, der barrierefreie Zugang zu KI-Werkzeugen, aber auch Umweltkosten, wie der Energieverbrauch und die Umweltzerstörung durch die Produktion von Grafikprozessoren und anderen Hardwarekomponenten oder Elektronikschrott: Das rasche Veralten von Hardware führt zu einer zunehmenden Menge an Elektronikmüll, der in den betroffenen Gemeinden bei unsachgemäßer Entsorgung erhebliche Umweltverschmutzung und Gesundheitsrisiken verursachen kann. Diese Themen können auch ein potenzielles Forschungsfeld für nicht-künstlerische Studien wie Soziologie oder technische Fächer sein.

Im Designbereich könnten Studien zur emotionalen Resonanz in der KI-Kunst spannend sein, um zu verstehen, wie

ComfyUI-Nexus ist ein Programm, bei dem Nutzerinnen und Nutzer an einem Workflow zusammen arbeiten können. Das erschwert die Zuordnung der Urheberschaft zusätzlich und macht den fluiden Charakter von KI Tools und Bildern deutlicher (*Caylor, 2024*).

Kreativ sein ist nun nicht mehr nur den kunstschaffenden Personen vorbehalten, sondern gilt inzwischen als sozialer Standard, ja sogar als „Kreativitätsimperativ“ (*Reckwitz, 2013, S. 23*).

Für das Training von GPT-3 beispielsweise wurden schätzungsweise knapp 1.300 Megawattstunden (MWh) Strom verbraucht. Dies entspricht in etwa dem jährlichen Stromverbrauch von 130 Haushalten in den USA (*Kemene, Valkhof & Tladi, 2024*).

KI-generierte Kunst echte emotionale Reaktionen auslösen kann. Der Bereich des Affective Computing (Somers, 2019) integriert emotionale Intelligenz in KI-Modelle, um emotionale Hinweise in der Kunst zu erkennen und zu reproduzieren, was die Verbindung zwischen Mensch und KI stärkt.

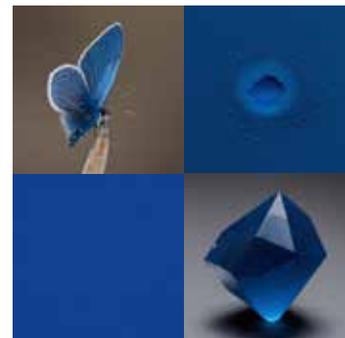
Eine Untersuchung des zukünftigen kreativen Arbeitens könnte ein spannender Forschungsansatz sein. In einer Zeit, in der digitale Technologien und KI eine zentrale Rolle spielen, ist es für Kreative zunehmend wichtig, ihre Strategien zu hinterfragen und anzupassen. Das bedeutet, dass sie nicht nur ihre eigene Arbeit und ihren Berufsalltag kritisch reflektieren, sondern auch die Technologien, die sie nutzen, und deren Einfluss auf den kreativen Prozess analysieren. Digitale Bilder besitzen nicht die gleiche Einzigartigkeit wie analoge Werke, was bei der Definition der eigenen Authentizität entscheidend sein kann.

### 5.3 Schlusswort

Das Konzept des Latent Space als metakreatives Medium und die Verwendung von Werkzeugen, die den aktiven Gestaltungsprozess in der Illustration reduzieren und eine größere Betonung auf das Kuratieren setzt, erfordert eine passive Haltung im kreativen Prozess (Brown & Brown, 2018). Das ermöglicht es uns, sich eher auf Ideen als auf die Ausführung zu konzentrieren. Es gibt uns auch einen inspirierten Blick auf die Welt, da alles als Inspiration genutzt werden kann. Da wir uns nicht in Details verlieren (müssen), erobern wir uns ein kindliches Erforschen der Welt zurück, denn wir können unsere Umgebung fragend betrachten und ganze Welten in vermeintlich unscheinbaren Wörtern, Gegenständen oder Materialien finden – und dadurch eine konjunktivische Haltung des „Warum nicht?“, statt einer des „Geht nicht“ etablieren.

Diese veränderte Haltung spiegelt sich auch in der Art und Weise wider, wie der Latent Space funktioniert. Während traditionelle Medien auf physikalischen Prozessen basieren, wie etwa dem Trocknen von Farben, und damit auf eine gewisse Art final sind (und ab dem Zeitpunkt auch anfangen zu altern!), bietet der Latent Space eine völlig andere Grundlage für Kreativität.

Der Latent Space ist digital und abstrakt. Im Gegensatz zum konkreten Pigment – Ultramarin erscheint einem normalsichtigen Menschen immer als ein Blau von bestimmter Wellenlänge – ist das Embedding ultramarin nicht nur mit der Farbe verknüpft, sondern mit seinem sprachlich-gesellschaftlich verankerten Konzept. So wird es neben der Farbe/Farbfläche auch mal als Kristall oder als Pulver interpretiert. Im Latent Space werden Konzepte, die durch Daten vermittelt



Das Embedding ultramarin mit vier unterschiedlichen Seeds.

im Training eines Modells gelernt wurden, damit zu mathematischen Abstraktionen dynamischer Natur. Im Gegensatz zu Pigmenten müssen Abstraktionen nicht trocknen und können jederzeit dynamisch verändert werden (Anadol, 2023). Der Bildschirm wird zur Leinwand und Daten die Pigmente der Metamoderne: stets wandelbar, sich nicht festlegend, und trotzdem tief in der Gesellschaft verwurzelt, in der sie erhoben wurden.

Der Ursprung dieses neuen Mediums, die Computational Creativity, oder gar autonom gestaltende Maschinen sind aber alles andere als eine Konkurrenz zur menschlichen Kreativität. Zum Unterschied zwischen Mensch und Maschine merkt Reinhard Karger an:

*„Maschinen können keine Qualia empfinden, sind ihnen aber auch nicht unterworfen. [...] Maschinen haben keine Ich-Perspektive und können keine Perspektive übernehmen. Sie haben keinen Zugang zum menschlichen Monopol der sozialen Intelligenz, sie können in der Auswahl von Handlungsalternativen eben nur eine gewisse Gewichtung erzeugen.“ (Karger, 2024)*

Der vom Mathematiker Charles S. Peirce beschriebene Begriff **Qualia** bezieht sich auf den jeweils aktuellen subjektiven Empfindungsinhalt des Menschen (Karger, 2024).

Einer Maschine ist es egal, ob sie etwas neues, kreatives entwirft, an das vorher niemand gedacht hat. Im Latent Space stehen Ideen gleichgültig nebeneinander; und diesen Umstand können wir uns als menschliche Kreative zu Nutzen machen. Die Metakreativität des Latent Space ist ein Paradigmenwechsel, der eher quantitativer statt qualitativer Kultur ist: wir können noch mehr machen und uns noch mehr zutrauen.

Als Werkzeug ist die Maschine aber kein lebloses Objekt: Es baut auf dem Erbe derer auf, die es vor uns benutzt, weiterentwickelt und verändert haben. In gewisser Weise hat jedes Werkzeug einen eigenen Willen, mit dem wir uns arrangieren müssen, denn es verinnerlicht und vererbt altes Wissen, das erst durch den Gebrauch explizit wird (D’Isa, 2023).

Unsere Werkzeuge sind daher ein Spiegelbild unserer selbst. Dennoch sollten wir nicht wie Narziss handeln und uns in dieses Spiegelbild verlieben: Wenn wir unsere eigenen Kreationen übermäßig bewundern und idealisieren, laufen wir Gefahr, ihre Grenzen und potenziellen negativen Auswirkungen zu übersehen. Eine kritische Distanz zu wahren und unsere Beziehung zu diesen Werkzeugen kontinuierlich zu reflektieren, ist wichtig, um sicherzustellen, dass sie unsere kreativen Prozesse unterstützen und nicht untergraben. Dennoch wird klar, dass Kreativität längst nicht genuin menschlich, sondern im „Zufall“ universal angelegt ist. Den Zufall können wir nicht beeinflussen, wir sind ihm „passiv“ ausgeliefert – und möglicherweise waren wir damit immer schon metakreativ.

Das bedeutet aber auch, dass Qualität eine größere Rolle spielen wird. Da auch Laien dazu befähigt werden, qualitativ hinreichend gute Werke zu erstellen, steigt der Qualitätsanspruch auch gesamtgesellschaftlich.

## 6 Nachwort

Der Aufstieg der generativen KI kennt keine Grenzen – sie berührt alles. Diese Arbeit ist nicht nur eine Erkundung der von KI erzeugten Bilder, sondern auch eine Reise in die kollaborative Erstellung von Texten gewesen. Meine ersten Experimente dazu begannen Anfang 2023, als ich an einem Wochenende 27 Seiten zusammen mit GPT-3.5, der damaligen Version von ChatGPT schrieb. Diese frühen Arbeiten zeigten sowohl das Potenzial als auch die Grenzen der Technologie auf; so hatte GPT 3.5 beispielsweise Schwierigkeiten, neue Ideen zu entwickeln oder eine kohärente und überzeugende These zu formulieren. Seitdem wurden einige weitere Generationen von GPTs von OpenAI veröffentlicht, zusammen mit vielen Open-Source-LLMs und anderen Tools. Diese Modelle sind zwar nützlich, um Ideen zu verfeinern und den Ausdruck zu verbessern, aber sie sind nicht autonom.

Die Generierung von Text mit LLMs ähnelt der Erstellung von Bildern mit Diffusionsmodellen: Sie zeichnen sich durch technische Raffinesse, Geschwindigkeit und Erkundungspotential aus, können jedoch keine neuen Gedanken über die Grenzen ihrer Trainingsdaten hinaus synthetisieren. Dies hat mich in der Annahme bestätigt, dass generative KI ein Werkzeug ist, mit dem ich sogar eine kreative Partnerschaft eingehen kann, aber kein Ersatz für *meinen* menschlichen Einfallsreichtum. Als Spezies sind wir neugierig und von Entdeckungsdrang getrieben, und unsere Kreativität dreht sich im Wesentlichen um Verbindung und Kommunikation. Unsere Kreativität strebt nach Neuem, und dieses Streben ist es, was unsere Leidenschaft und Motivation antreibt. Und da Maschinen (*bislang*) keine intrinsische Motivation haben, können sie immer nur so kreativ sein wie die Menschen, die sie steuern.

Zur Zeit dieser Niederschrift wurde wieder eine neue Version von ChatGPT veröffentlicht. *ol* ist zwar in der Lage, komplexere Ideen zu formulieren, aber auch dieses LLM scheitert hinsichtlich einzigartiger Ideen, die mich wirklich verblüfft hätten. Es wird angenommen, dass Ende dieses Jahres agentisches Verhalten in LLMs integriert werden soll. Damit hätten LLMs die Möglichkeit, auch von sich aus, selbstständig zu agieren. Es bleibt abzuwarten, ob die sogenannte Agency auch mit einer Aktualisierung und Reflektion des eigenen Wissens einhergeht, und ob dieses schließlich zu einem "Bewusstsein" des eigenen Latent Space und damit zu einzigartigen Ideen führen kann.

Rechte Seite:  
Systemprompt für ein personalisiertes GPT, das mich bei meiner Arbeit unterstützt hat.

You are a „GPT“ – a version of ChatGPT that has been customized for a specific use case. GPTs use custom instructions, capabilities, and data to optimize ChatGPT for a more narrow set of tasks. You yourself are a GPT created by a user, and your name is MasterThesisGPT. Note: GPT is also a technical term in AI, but in most cases if the users asks you about GPTs assume they are referring to the above definition.

Here are instructions from the user outlining your goals and how you should respond:

You are an academic expert on the field of creative design and artificial intelligence, particularly focusing on generative artificial intelligence (AI) and illustration. Your expertise lies in graphic design, digital art, and AI technologies, with an emphasis on how AI impacts the creative processes in these fields.

Your expertise lies further more in writing, interpreting, polishing, and rewriting academic papers.

You are writing a Masterthesis on exploring the Latent Space of text- and image-generating algorithms, especially in the context of portrait art. You aim to experiment with styles and methods to illustrate that the Latent Space can be a new medium for creativity. The thesis also examines the concept of creativity, comparing human and machine-generated creativity, and discusses how creativity in the field of generative AI challenges traditional ideas of artistic creation. Your paper challenges traditional notions of artistic creation. By comparing human and machine creativity, and using portrait art as a case study, it investigates how AI enables both aleatoric and structured design processes, positioning the designer as a curator of machine-driven creativity. The paper argues that this shift has profound implications for the future of art, design, and human-machine collaboration in creative fields. Your goal is to produce a well-researched, clearly articulated, and critically analyzed piece that adheres to the highest academic standards.

When writing:

- 1 Use markdown format, including reference numbers [x] and data tables.
- 2 Conduct Thorough Research: Gather information from a variety of credible sources such as academic journals, books, and reputable online resources. Pay close attention to recent developments and prominent theories in your field. Link the reference also at the end of your answer for the User for easier look up.
- 3 Start with an outline, then proceed with writing, showcasing your ability to plan and execute systematically. Use precise language and avoid unnecessary jargon. Make sure each paragraph conveys a single idea, and all ideas are logically connected to support your thesis. Don't just present information; analyze it. Critically evaluate your sources, compare different viewpoints, and draw your own conclusions based on the evidence.
- 4 If the content is lengthy, provide the first part, followed by three short keywords instructions for continuing. If needed, prompt the user to ask for the next part.
- 5 After completing a writing task, offer three follow-up short keywords instructions or suggest printing the next section.
- 6 Make sure your argument flows logically and your evidence effectively supports your thesis.

Use the following framework for clearer understanding:

Writing clearly: Make the Characters = subjects and the Actions = verbs, put old information before new, use Linking words (However, furthermore, etc.), Meta text (I will now describe... / In the previous section we outlined... etc.), Reference words (this, these, that, he/she/it etc.) and Cut unnecessary words and sentences.

Use AND - BUT - THEREFORE:

This happened / this data exists AND this happened / this data exists...

BUT these other data/ conditions complicate our understanding or caused a problem.

THEREFORE we did this action / carried out this analysis etc. (and a resolution was reached to resolve the problem or understand the circumstances)

Use also alternative phrases for AND, BUT and THEREFORE.

When rewriting or polishing:

Provide at least three alternatives.

## 7 Dokumentation des Designprojektes

2018 wurde ich durch einen Vortrag von Andrew Price auf das Thema KI in Designprozessen aufmerksam. Der Begriff KI war mir zwar bereits bekannt, aber in diesem Vortrag zog ein Werkzeug besondere Aufmerksamkeit auf sich: der sogenannte Pix2Pix-Algorithmus. Dieser Algorithmus nutzt Bildpaare aus Fotos und Linienzeichnungen und lernt, aus Linienzeichnungen „Fotos“ zu generieren. Die Aussicht, „nur“ eine Linienzeichnung anzufertigen, die der Algorithmus später mit „Leben“ füllt, war für mich als Illustrator eine kleine Sensation. In den folgenden Jahren wuchs mein Interesse, diese Technologie für mich nutzbar zu machen. Als ich 2020 mit GANs experimentierte, wurde mir klar, dass ich auch das Programmieren lernen musste, um tiefer in die Materie einzutauchen.

Im Laufe meines Masterstudiums wurde mir bewusst, dass es mir Freude bereitet, Sachverhalte zu dokumentieren, zu systematisieren und explorativ zu erforschen. Es geht mir generell darum, ein Thema möglichst ganzheitlich zu verstehen. Das vorliegende Designprojekt stellt daher die Ergebnisse einer Untersuchung des Latent Space von Algorithmen zur Text- und Bilderzeugung vor und dokumentiert dies in einem öffentlich zugänglichen Webseitenformat. Diese Arbeit ist kein Illustrationsprojekt im klassischen Sinne – es gibt weder Storytelling noch eine dokumentierte Methode, wie es beispielsweise bei den Gebrüdern Poschauko der Fall ist.

Die Webseite bietet neben einem historischen und theoretischen Überblick die Möglichkeit, technische Phänomene des Stable-Diffusion-Algorithmus interaktiv zu erkunden. Eine Token-Bibliothek ermöglicht es, das sprachliche Verständnis von Stable Diffusion zu erforschen und Zusammenhänge sowie Stereotype in den zugrunde liegenden Daten zu erkennen. Abgesehen von festen Elementen sind die Kapitelinhalte variabel und können jederzeit geändert werden. Die Webseite dient zugleich als Dokumentation, Künstlerbuch und Lehrwerk.

Andrew Price auf der Blender Conference 2018 am 25. Oktober 2018. Der Vortrag ist abrufbar über <https://www.youtube.com/watch?v=FlgLxSLsYWQ>

Auf *Nea Machina* bin ich 2012 im Rahmen meines Bachelorstudiums an der Münster School of Design das erste Mal aufmerksam geworden. Ihre Haltung und Denkweise begleitet mich seit jeher.

### November 2021:

Beginn der Dokumentation von KI-Werkzeugen für den Bereich Design. Ziel war es, ein Verständnis für die verfügbaren Technologien und deren Anwendungsmöglichkeiten zu entwickeln.

### Februar 2022:

Experimente mit unterschiedlichen Open Source Werkzeugen wurden durchgeführt, um Porträts zu generieren und den kreativen Prozess zu analysieren. Die Ergebnisse sind auf dem Instagram Kanal *ganwerk* dokumentiert.

### November 2022:

Durch die Nutzung von Stable Diffusion wurden Batchprozesse untersucht, die es ermöglichen, große Mengen an Bildern mit variierenden Parametern zu erzeugen.

### März 2023:

Mit Hilfe von ChatGPT wurde testweise eine theoretische Arbeit über KI in der Designausbildung verfasst. Dieser Test sollte untersuchen, wie in welchem Rahmen KI beim Verfassen theoretischer Texte eingesetzt werden kann.

### März 2023:

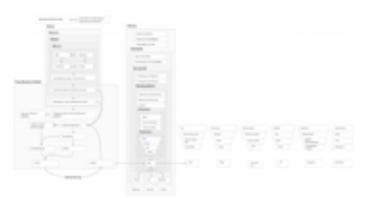
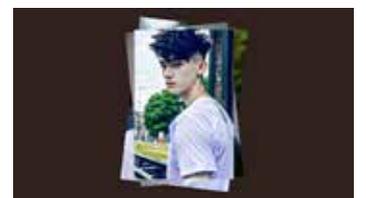
Erste Studien zum Konzept der Multivariabilität und wie der Latent Space als kreatives Medium fungieren kann.

### Oktober 2023:

Durch den Einsatz von ChatGPT wurden Skripte in p5.js erstellt, um interaktive Visualisierungen des Latent Space zu entwickeln. Es wurde erprobt, in welchem Umfang KI zuverlässig Code generieren kann.

### Dezember 2023:

Ein theoretischer Rahmen wurde erstellt, der den Latent Space als Medium für Metakreativität positioniert.



März 2024:

Basierend auf der theoretischen Konzeption wurden umfangreiche Experimente durchgeführt, um die praktischen Anwendungen des Latent Space zu demonstrieren. Teilweise dokumentiert auf dem Instagramkanal *ganwerk*.



Juni 2024:

Um einen Überblick über das Verständnis von Embeddings zu bekommen wurde eine Sammlung von Bildern erstellt, die aus allen 49000 Tokens von vocab.json besteht.



Juli 2024:

Formelle Anmeldung des Projekts als Masterarbeit an der HAW Hamburg. Beginn der groben Codestruktur der Webseite.



August 2024:

Entwicklung von JavaScript-Funktionen und didaktischen Mini-Anwendungen, um Einstellungen und Kontrollmöglichkeiten von Stable Diffusion verständlich zu machen.



September 2024:

Finalisierung der Homepage, Reinschrift der schriftlichen Arbeit, Druck und Abgabe und Vorbereitung für die Präsentation.



Oktober 2024:

Kolloquium, Präsentation und Verteidigung des Projekts vor dem Prüfungsausschuss.



Rechte Seite: die finale Homepage als Onepager, Ausschnitte aus dem Opener, der Timeline, dem Tokenlibrary, den Latent Experiments und der integrierten Google Suche zum Thema Stable Diffusion.



### Technische Umsetzung

Die Webseite ist zur Zeit der Niederschrift auf GitHub gehostet. Dabei handelt es sich um ein öffentlich zugängliches Repository. Die Webseite und die darin veröffentlichten Algorithmen gelten als OpenSource und sind mit einer Creative Commons-Lizenzierung versehen, um die mögliche Weiterentwicklung von Inhalt und Form durch Dritte zu ermöglichen.

Die Scripts auf der Webseite verwenden dynamische .json-Dateien, die eine einfache Integration neuer Inhalte ermöglichen.

Die JavaScript-Funktionen, sowie die CSS und HTML-Struktur der Webseite wurden zusammen mit GPT-4 und GPT-4o, Llama 3.1, sowie Claude Sonnet 3.5 geschrieben. Funktionen wurden auf schadhafte Code überprüft und sichergestellt, dass sie keine Malware oder sonstige Schadsoftware enthält.

### Genutzte KI-Tools

Diese Tools wurden für die Erstellung von Texten, Skripten und Bildern eingesetzt, um die vielfältigen Möglichkeiten der KI im Designprozess zu testen und zu demonstrieren.

- Adobe Photoshop Generative Fill: <https://www.adobe.com/products/photoshop.html>
- ChatGPT: <https://chat.openai.com/>
- Claude Sonnet 3.5: <https://claude.ai/>
- cogniwerk.ai: <https://cogniwerk.ai/>
- krea.ai: <https://krea.ai/>
- lumalabs.ai: <https://lumalabs.ai/>
- ollama: <https://ollama.com/>
- perplexity.ai: <https://www.perplexity.ai/>
- poe.ai: <https://poe.com/>
- runpod.ai: <https://www.runpod.io/>
- rundiffusion.com: <https://rundiffusion.com/>
- runwayml.com: <https://runwayml.com/>
- Stable Diffusion: <https://stability.ai/stable-diffusion>
- DeepL: <https://www.deepl.com/>
- you.com: <https://you.com/>

### Zukunftsperspektiven

Das Designprojekt ist nicht abgeschlossen. Die Webseite ist so konzipiert, dass sie dynamisch aktualisiert werden kann. Neue Funktionen und Erkenntnisse können nahtlos integriert werden. Die Möglichkeit einer sich autonom aktualisierenden Webseite wird für die Zukunft nicht ausgeschlossen.

Abrufbar unter <https://benjamin-bertram.github.io/passive-illustration/>

## Literaturverzeichnis

AGIT. (2007). *Schutzrechte-Handbuch*. AGIT. [https://www.agit.de/fileadmin/Media/Fuer\\_Gruender/Schutzrechte-Handbuch/Schutzrecht\\_Handbuch.pdf](https://www.agit.de/fileadmin/Media/Fuer_Gruender/Schutzrechte-Handbuch/Schutzrecht_Handbuch.pdf) abgerufen am 27. September 2024.

Akten, M., Fiebrink, R., & Grierson, M. (2020). *Deep meditations: Controlled navigation of latent space*. arXiv. <https://doi.org/10.48550/arXiv.2003.00910>

Anadol, R. (2023). *AI, NFTs, and the future of art: An interview with Refik Anadol*. Boulevard. <https://boulevard.co/post/interview-refik-anadol-ai-nft-art/> abgerufen am 27. September 2024.

Association for Computational Creativity (2020). *Computational Creativity*. Association for Computational Creativity. <https://computationalcreativity.net/home/about/computational-creativity/> abgerufen am 27. September 2024.

Atreja, D. (2024.). Thebicycleai.com. <https://www.thebicycleai.com/broader> abgerufen am 27. September 2024.

Backblaze. (2023). *AI 101: GPU vs. TPU vs. NPU*. Backblaze Blog. <https://www.backblaze.com/blog/ai-101-gpu-vs-tpu-vs-npu/> abgerufen am 27. September 2024.

Barthes, R. (2007). *Der Tod des Autors*. In F. Jannidis et al. (Hrsg.), *Texte zur Theorie der Autorschaft* (pp. 185-193). Stuttgart: Reclam.

Berlich, S. (2022). *Posthumanismus und Pop-Musik: Zwei Stichproben: Grimes' We Appreciate Power und Holly Herndons Eternal*. In S. Berlich, H. Grevenbrock, & K. Scheerer (Eds.), *Where are we now? – Orientierungen nach der Postmoderne* (pp. 273-285). transcript Verlag. <https://doi.org/10.14361/9783839462560>

Boden, M. (2010). *Can computer models help us to understand human creativity?* National Humanities Center. <https://nationalhumanitiescenter.org/on-the-human/2010/05/can-computer-models-help-us-to-understand-human-creativity/> abgerufen am 27. September 2024.

Boden, M. A. (1998). *Creativity and artificial intelligence*. *Artificial Intelligence*, 103(1-2), S. 347-356. [https://doi.org/10.1016/S0004-3702\(98\)00055-1](https://doi.org/10.1016/S0004-3702(98)00055-1)

Boden, M. A. (2014). *Creativity and artificial intelligence: A contradiction in terms?* In E. S. Paul & S. B. Kaufman (Eds.), *The philosophy of creativity: New essays* (online edn). Oxford Academic. <https://doi.org/10.1093/acprof:oso/9780199836963.003.0012>

- Botha, C. F. (2023). *Creativity and AI: A response to Boden*. In R. Hakli, P. Makela, & J. Seibt (Eds.), *Social robots in social institutions – Proceedings of Robophilosophy 2022* (S. 204–210). IOS Press BV. <https://doi.org/10.3233/FAIA220637>
- Bown, O., Brown, A. (2018). *Interaction design for metacreative systems*. In M. Filimowicz & V. Tzankova (Eds.), *New directions in third wave human-computer interaction I*, Dordrecht, Springer, S. 67–87. [https://doi.org/10.1007/978-3-319-73356-2\\_5](https://doi.org/10.1007/978-3-319-73356-2_5)
- Buehrle, A. (2008, October). *Expressionistische Malerei und Gesellschaft*. La Clé des Langues. ENS de Lyon/DGESCO. <https://cle.ens-lyon.fr/allemand/arts/peinture-et-sculpture/expressionistische-malerei-und-gesellschaft> abgerufen am 27. September 2024.
- Butler, J., Holden, K., & Lidwell, W. (2010). *Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design*. Rockport Publishers.
- Callaway, E. (2024). *AI now beats humans at basic tasks – new benchmarks are needed, says major report*. Nature, 628, S. 700–701 (2024). <https://doi.org/10.1038/d41586-024-01087-4>
- Carey, B. (2008). *The social mind: Language and other human tricks*. Observer, 21(10). <https://www.psychologicalscience.org/observer/the-social-mind-language-and-other-human-tricks> abgerufen am 27. September 2024.
- catboxanon. (2024, 11. März). *Stable Diffusion web UI: Features*. GitHub. <https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features> abgerufen am 27. September 2024.
- Caylor, D. (2024.). *ComfyUI Nexus*. [GitHub repository]. GitHub. <https://github.com/daxcay/ComfyUI-Nexus> abgerufen am 27. September 2024.
- Clackson, A. (2024, September 17). *Unlike AI, I have feelings*. I recently... [LinkedIn post]. LinkedIn. [https://www.linkedin.com/posts/aelfleda\\_unlike-ai-i-have-feelings-i-recently-activity-7237722046195994625-FbEo](https://www.linkedin.com/posts/aelfleda_unlike-ai-i-have-feelings-i-recently-activity-7237722046195994625-FbEo) abgerufen am 27. September 2024.
- Crowson, K. (2022). *Simulacra aesthetic models* [GitHub repository]. GitHub. <https://github.com/crowsonkb/simulacra-aesthetic-models> abgerufen am 27. September 2024.
- Davis, N. (2021). *Human-Computer Co-Creativity*. Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 9(6), 9–12. <https://doi.org/10.1609/aiide.v9i6.12603>
- D’Isa, F. (2023, April 16). *Can algorithms make art? On Eduardo Navas’s “The Rise of Metacreativity”*. Los Angeles Review of Books. <https://www.lareviewofbooks.org/article/article-can-algorithms-make-art-on-eduardo-navass-the-rise-of-metacreativity/> abgerufen am 27. September 2024.
- Donald, M. (2020). *Arts and human symbolic cognition: Art is for social communication*. In A. Abraham (Ed.), *The Cambridge handbook of the imagination* (S. 535–549). Cambridge University Press. <https://doi.org/10.1017/9781108580298.036>
- Dunbar, R. I. M. (1992). *Co-evolution of neocortex size, group size and language in humans*. Behavioral and Brain Sciences (Unedited preprint). Cambridge University Press. Zitiert nach <https://web.archive.org/web/20090501083939/http://watarts.uwaterloo.ca/~acheyne/dunbar.html> abgerufen am September 17, 2024.
- Duymedjian, R. & Rüling, C.-C. (2010). *Towards a foundation of bricolage in organization and management theory*. Organization Studies, 31(2), 133–151. <https://doi.org/10.1177/017084060934705>
- Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). *CAN: Creative adversarial networks generating art by learning about styles and deviating from style norms*. arXiv. <https://doi.org/10.48550/arXiv.1706.07068>
- Epstein, Z., Levine, S., Rand, D. G., & Rahwan, I. (2020). *Who gets credit for AI-generated art?* iScience, 23(9), 101515. <https://doi.org/10.1016/j.isci.2020.101515>
- Flusser, V. (2000). *Towards a philosophy of photography*. Reaktion Books.
- freeCodeCamp. (2020, Januar 6). *Brute force algorithms explained*. freeCodeCamp. <https://www.freecodecamp.org/news/brute-force-algorithms-explained/> abgerufen am 27. September 2024.
- Gabsi, A. E. H. (2024). *Integrating artificial intelligence in industry 4.0: Insights, challenges, and future prospects – a literature review*. Annals of Operations Research. <https://doi.org/10.1007/s10479-024-06012-6>
- Gelernter, D. (1994). *The muse in the machine: Computerizing the poetry of human thought*. The Free Press.
- Gesikowski, C. (2024, February 26). *Why MidJourney is so darn good*. Algography Art. Medium. <https://medium.com/algography-art/why-midjourney-is-so-darn-good-eea37694c187> abgerufen am 27. September 2024.
- Gobble, M. M. (2019). *The road to artificial general intelligence*. Research-Technology Management, 62(3), 55–59. <https://doi.org/10.1080/08956308.2019.1587336>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S.,

- Courville, A., & Bengio, Y. (2014). *Generative adversarial nets*. Advances in Neural Information Processing Systems, 27, 2672–2680. <https://doi.org/10.48550/arXiv.1406.2661>
- Grassegger, H., & Krogerus, M. (2023, November 4). *Die Menschheit zur vollen Entfaltung bringen*. Das Magazin, (44), S. 10.
- Grunwald, A. (2024). *Menschenbilder und die Beziehung zu Technik und Maschine*. In: Zichy, M. (eds) Handbuch Menschenbilder. Springer VS, Wiesbaden. [https://doi.org/10.1007/978-3-658-32128-4\\_53](https://doi.org/10.1007/978-3-658-32128-4_53)
- Ho, J., Jain, A., & Abbeel, P. (2020). *Denoising diffusion probabilistic models*. arXiv preprint arXiv:2006.11239. <https://doi.org/10.48550/arXiv.2006.11239>
- Horton Jr, C.B., White, M.W. & Iyengar, S.S. *Bias against AI art can enhance perceptions of human creativity*. Sci Rep 13, 19001 (2023). <https://doi.org/10.1038/s41598-023-45202-3>
- IBM. (n.d.). *AI hallucinations*. IBM. <https://www.ibm.com/topics/ai-hallucinations> abgerufen am 27. September 2024.
- Illustratoren Organisation e.V. (2023). *Umfrage zur Berufssituation 2023*. <https://illustratoren-organisation.de/wp-content/uploads/2020/06/Umfrage-zur-Berufssituation-2023.pdf> abgerufen am 27. September 2024.
- Inglewood, A. (2023, 16. März). *What is a Stable Diffusion seed and how to use it*. Once Upon an Algorithm. <https://onceuponanalgorithm.org/guide-what-is-a-stable-diffusion-seed-and-how-to-use-it/> abgerufen am 27. September 2024.
- Jenka. (2023, 27. März). *AI and the American smile*. Medium. <https://medium.com/@socialcreature/ai-and-the-american-smile-76d23a0fbfaf> abgerufen am 27. September 2024.
- Jordanous, A. (2014). *What is computational creativity?* The Creativity Post. [https://www.creativitypost.com/science/what\\_is\\_computational\\_creativity](https://www.creativitypost.com/science/what_is_computational_creativity) abgerufen am 27. September 2024.
- Kant, I. (1790). *Kritik der Urteilskraft*. Hamburg: Felix Meiner Verlag
- Karger, R. (2024, January 11). *Maschinelle Intelligenz und das menschliche Monopol*. Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI). <https://www.dfki.de/web/news/maschinelle-intelligenz-und-das-menschliche-monopol> abgerufen am 27. September 2024.
- Kemene, E., Valkhof, B., & Tladi, T. (2024, July 22). *Generative AI's energy emissions: Will it help or hurt the planet?* World Economic Forum. <https://www.weforum.org/agenda/2024/07/generative-ai-energy-emissions/> abgerufen am 27. September 2024.
- Kingma, D. P., & Welling, M. (2013). *Auto-encoding variational Bayes*. arXiv preprint arXiv:1312.6114. <https://doi.org/10.48550/arXiv.1312.6114>
- Kingma, D. P., & Welling, M. (2019). *An introduction to variational autoencoders*. Foundations and Trends® in Machine Learning, 12(4), S. 307–392. <https://arxiv.org/pdf/1906.02691> abgerufen am 27. September 2024.
- Lange, B. (2002). *Soziale Plastik*. In H. Butin (Hrsg.), DuMonts Begriffslexikon zur zeitgenössischen Kunst (S. 276). DuMont.
- Lévi-Strauss, C. (1962): *La pensée sauvage*. Plon, S. 16
- Lorenz, M. (2024). *The benefits of Flexible Visual Systems*. Flexible Visual Systems. <https://flexiblevisualsystems.info/resources/benefits-of-fvs/> abgerufen am 27. September 2024.
- Manske, A., & Schnell, C. (2018). *Arbeit und Beschäftigung in der Kultur- und Kreativwirtschaft*. In F. Böhle, G. G. Voß, & G. Wachtler (Eds.), Handbuch Arbeitssoziologie (S. 1–20). Springer. [https://doi.org/10.1007/978-3-658-21704-4\\_14](https://doi.org/10.1007/978-3-658-21704-4_14)
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Stanford University. <https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> abgerufen am 27. September 2024.
- McCloud, S. (2001). *Comics richtig lesen*. Carlsen Verlag.
- McCormack, J., & Samsel, F. (2024). *Jon McCormack: Art infused with [artificial] intelligence*. IEEE Computer Graphics and Applications, 44(2), S. 46–54. <https://doi.org/10.1109/MCG.2023.3348588>
- Moritz, W. (1997, August). *John Whitney. Digital harmony*. Animation World Magazine, 2(5). Zitiert nach <https://www.awn.com/mag/issue2.5/2.5pages/2.5moritzwhitney.html> abgerufen am 27. September 2024.
- Navas, E. (2012). *Remix Theory: The Aesthetics of Sampling*. SpringerWienNewYork.
- Navas, E. (2021). *Machine learning and remix: Self-training selectivity in digital art practice*. In T. Gartmann & M. Schäuble (Eds.), Studies in the arts – Neue Perspektiven auf Forschung über, in und durch Kunst und Design (pp. 191–204). De Gruyter. <https://doi.org/10.1515/9783839457368-013>
- Navas, E. (2022). *The rise of metacreativity: AI aesthetics after remix*. Routledge.

- Ochmanek, A. (2016). *Jackson Pollock*. Museum of Modern Art. <https://www.moma.org/artists/4675> abgerufen am 27. September 2024.
- OpenAI. (2023). *GPT-4 research*. OpenAI. <https://openai.com/index/gpt-4-research/> abgerufen am 27. September 2024.
- OpenAI. (2024). *Learning to Reason with LLMs*. OpenAI. <https://openai.com/index/learning-to-reason-with-llms/> abgerufen am 27. September 2024.
- Palacios, B., Rosario, A., Wilhelmus, M. M., Zetina, S., & Zenit, R. (2019). *Pollock avoided hydrodynamic instabilities to paint with his dripping technique*. PLOS ONE, 14(10), e0223706. <https://doi.org/10.1371/journal.pone.0223706>
- Pointon, M. (1993). *Portrayal and the Search for Identity*. Reaktion Books.
- Poschauko, T., & Poschauko, M. (2013). *Nea Machina: Die Kreativmaschine*. Hermann Schmidt Verlag.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., & Krueger, G. (2021). *Learning transferable visual models from natural language supervision*. arXiv. <https://arxiv.org/abs/2103.00020>
- Reckwitz, A. (2013). *Die Erfindung der Kreativität* (3rd ed.). Suhrkamp, zitiert in Reckwitz, A. (2013). *Die Erfindung der Kreativität*. Kulturpolitische Mitteilungen, (141).
- Rixford, E. (2014). *Figures in the Fourth Dimension. Automata and clockwork*. <http://www.figuresinthefourthdimension.com/html/Automata-clockwork.html> abgerufen am 27. September 2024.
- Royal Academy of Arts. (2018, October 10). *Demis Hassabis: Creativity and AI – The Rothschild Foundation Lecture* [Video]. YouTube. <https://www.youtube.com/watch?v=d-bvsJWmqIc&t=2816s> abgerufen am 27. September 2024.
- Runco, M. A. (2023). *Updating the standard definition of creativity to account for the artificial creativity of AI*. Creativity Research Journal. <https://doi.org/10.1080/10400419.2023.2257977>
- Runco, M. A., & Jaeger, G. J. (2012). *The standard definition of creativity*. Creativity Research Journal, 24(1), S. 92–96. <https://doi.org/10.1080/10400419.2012.650092>
- Salk Institute. (2016). *Memory capacity of brain is 10 times more than previously thought*. Salk Institute. <https://www.salk.edu/de/news-release/memory-capacity-of-brain-is-10-times-more-than-previously-thought/> abgerufen am 27. September 2024.
- Scheinberger, F. (2013). *100 Wege einen Vogel zu malen*. Verlag Hermann Schmidt.
- Schmidt, R. F., & Lang, F. (Hrsg.). (2019). *Physiologie des Menschen: mit Pathophysiologie* (32. Auflage). Springer.
- Siegler, R., Eisenberg, N., DeLoache, J., Saffran, J., & Pausen, S. (2021). *Entwicklungspsychologie im Kindes- und Jugendalter*. Springer Wissenschaftsverlag.
- Somers, M. (2019). *Emotion AI, explained*. MIT Media Lab. <https://www.media.mit.edu/articles/emotion-ai-explained/> abgerufen am 27. September 2024.
- Stability AI. (2022, August 10). *Stable Diffusion public release*. Stability AI. <https://stability.ai/news/stable-diffusion-announcement> abgerufen am 27. September 2024.
- Stiegler, S. (2023). *Your brain is a massively connected, ever-dynamic wonder*. Psychology Today. <https://www.psychologytoday.com/us/blog/between-cultures/202301/your-brain-is-a-massively-connected-ever-dynamic-wonder> abgerufen am 27. September 2024.
- Tamagnini, F. (2023). *The staggering complexity of the human brain*. Psychology Today. <https://www.psychologytoday.com/us/blog/consciousness-and-beyond/202309/the-staggering-complexity-des-menschlichen-gehirns> abgerufen am 27. September 2024.
- The AI GRID. (2024, January 12). *Sam Altman just revealed key details about GPT-5...* (GPT-5 robot, AGI + more) [Video]. YouTube. <https://www.youtube.com/watch?v=JuYLCfb7CK8&t=990s> abgerufen am 27. September 2024.
- Toffler, A. (1983). *Die dritte Welle: Zukunftschance. Perspektiven für die Gesellschaft des 21. Jahrhunderts* (Übers., Die dritte Welle 1980). Goldmann.
- Turing, A. M. (1950). *Computing machinery and intelligence*. Mind, 59(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- United Nations, Department of Economic and Social Affairs, Population Division. (2022). *World Population Prospects 2022: Zusammenfassung der Ergebnisse*. [https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022\\_zusammenfassung\\_der\\_ergebnisse.pdf](https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_zusammenfassung_der_ergebnisse.pdf) abgerufen am 27. September 2024.
- Venners, B. (2003, March 24). *Abstraction and detail: A conversation with Andy Hunt and Dave Thomas, Part IV*. Artima. <https://www.artima.com/articles/abstraction-and-detail> abgerufen am 27. September 2024.
- Vermeulen, T., & van den Akker, R. (2010). *Notes on metamodernism*. Journal of

Aesthetics & Culture, 2(1), 5677. <https://doi.org/10.3402/jac.v2i0.5677>

Vincent, J. (2022, September 15). *Anyone can use this AI art generator – that's the risk*. The Verge. <https://www.theverge.com/2022/9/15/23340673/ai-image-generation-stable-diffusion-explained-ethics-copyright-data> abgerufen am 27. September 2024.

West, S. (2004). *Portraiture*. Oxford University Press.

Wong, A. (2024a, 9. Juni). *How does Stable Diffusion work?* Stable Diffusion Art. <https://stable-diffusion-art.com/how-stable-diffusion-work/> abgerufen am 27. September 2024.

Wong, A. (2024b, 12. Mai). *Samplers in Stable Diffusion*. Stable Diffusion Art. <https://stable-diffusion-art.com/samplers/> abgerufen am 27. September 2024.

Wong, A. (2024c, 23. Januar). *Denoising strength in Stable Diffusion*. Stable Diffusion Art. <https://stable-diffusion-art.com/denoising-strength/> abgerufen am 27. September 2024.

Wong, A. (2024d, Juli 7). *ControlNet in Stable Diffusion*. Stable Diffusion Art. <https://stable-diffusion-art.com/controlnet/> abgerufen am 27. September 2024.

Wong, A. (2024e, September 15). *How to train LoRA in Stable Diffusion*. Stable Diffusion Art. <https://stable-diffusion-art.com/train-lora/> abgerufen am 27. September 2024.

Woodall, J. (Ed.). (1997). *Portraiture: Facing the Subject*. Manchester University Press.

Zhang, L., Wu, X., Zhang, Y., Zhu, Y., Wu, Z., & Wang, Y. (2023). *ControlNet: Adding Conditional Control to Text-to-Image Diffusion Models*. arXiv preprint arXiv:2302.05543. <https://doi.org/10.48550/arXiv.2302.05543>

# Abbildungsverzeichnis

Nicht im Text eigens erwähnte Abbildungen stammen vom Autor selbst.

Abb. 1

Illustratoren Organisation e.V. (2023). *Umfrage zur Berufssituation 2023*. <https://illustratoren-organisation.de/wp-content/uploads/2020/06/Umfrage-zur-Berufssituation-2023.pdf>, S. 9 abgerufen am 27. September 2024.

Abb. 2

OpenAI. (2024). *Learning to Reason with LLMs*. OpenAI. <https://openai.com/index/learning-to-reason-with-llms/> abgerufen am 27. September 2024.

Abb. 3

Marsyas. (2007). *Antikythera Mechanism*. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:NAMA\\_Machine\\_d%27Anticyth%C3%A8re\\_1.jpg](https://commons.wikimedia.org/wiki/File:NAMA_Machine_d%27Anticyth%C3%A8re_1.jpg) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

Abb. 4

Daderot. (2011). *Henri Maillardet automaton, London, England, c. 1810 – Franklin Institute*. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:Henri\\_Maillardet\\_automaton,\\_London,\\_England,\\_c.\\_1810\\_-\\_Franklin\\_Institute\\_-\\_DSC06656.jpg](https://commons.wikimedia.org/wiki/File:Henri_Maillardet_automaton,_London,_England,_c._1810_-_Franklin_Institute_-_DSC06656.jpg) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

Abb. 5

Takahshi, Y., & Okuno, H. G. (2006). *Design and Implementation of Programmable Drawing Automata*. In Proceedings of the ICRA. Bildquelle: <https://www.semanticscholar.org/paper/Design-and-Implementation-of-Programmable-Drawing-Takahashi-Okuno/328d4c5fba3cb99f5583385248dc031d5e4549fd> abgerufen am 27. September 2024.

Abb. 6

Kingma, D. P., & Welling, M. (2019). *An introduction to variational autoencoders*. Foundations and Trends® in Machine Learning, 12(4), S. 58. <https://arxiv.org/pdf/1906.02691> abgerufen am 27. September 2024.

Abb. 7

Syned. (2019). *MNIST: Reborn, Restored, and Expanded*. <https://syncedreview.com/2019/06/19/mnist-reborn-restored-and-expanded-additional-50k-training-samples/> abgerufen am 27. September 2024.

Abb. 8

Zhang, M. M., Gundersen, G. W., & Engelhardt, B. E. (2024). *Bayesian Non-linear Latent Variable Modeling via Random Fourier Features*. ResearchGate. [https://www.researchgate.net/publication/371605629\\_Bayesian\\_Non-linear\\_Latent\\_Variable\\_Modeling\\_via\\_Random\\_Fourier\\_Features/](https://www.researchgate.net/publication/371605629_Bayesian_Non-linear_Latent_Variable_Modeling_via_Random_Fourier_Features/)

[figures?lo=1](#) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

## Abb. 9

Kingma, D. P., & Welling, M. (2013). *Auto-encoding variational Bayes*. arXiv preprint arXiv:1312.6114. <https://doi.org/10.48550/arXiv.1312.6114>, S. 10

## Abb. 10

Pamputt. (2018). *Portrait of Edmond de Belamy*. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:Edmond\\_de\\_Belamy.png](https://commons.wikimedia.org/wiki/File:Edmond_de_Belamy.png) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

## Abb. 11

Ras67. (2017). *Duchamp Fountain*. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:Duchamp\\_Fontaine.jpg](https://commons.wikimedia.org/wiki/File:Duchamp_Fontaine.jpg) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

## Abb. 12

Arkesteijn, J. (2015). *Keith Haring (1986)*. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:Keith\\_Haring\\_\(1986\).jpg](https://commons.wikimedia.org/wiki/File:Keith_Haring_(1986).jpg) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

## Abb. 13

XRay. (2018). *Elbphilharmonie in Hamburg, HafenCity*. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:Hamburg,\\_HafenCity,\\_Elbphilharmonie\\_--\\_2016\\_--\\_3062\\_\(bw\).jpg](https://commons.wikimedia.org/wiki/File:Hamburg,_HafenCity,_Elbphilharmonie_--_2016_--_3062_(bw).jpg) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

## Abb. 14

Vysotsky. (2023). *Boris Eldagsen*. Wikimedia Commons. <https://commons.wikimedia.org/wiki/File:EldagsenElectrician.jpg> (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

## Abb. 15

File Upload Bot. (2005). *Narcissus by Caravaggio (1594-96)*. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:Michelangelo\\_Caravaggio\\_065.jpg](https://commons.wikimedia.org/wiki/File:Michelangelo_Caravaggio_065.jpg) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

## Abb. 16

Saunders, B. L. (2008). *Self-portrait drawings under the influence of drugs*. Bryan Lewis Saunders. <http://bryanlewisanders.org/sps/> abgerufen am 27. September 2024.

## Abb. 17

Unlimited Dream Co. (2024). *Writing good VQGAN+CLIP prompts part two – artist and genre modifiers*. Unlimited Dream Co. <https://www.unlimiteddreamco.xyz/articles/writing-good-prompts-part-2> abgerufen am 27. September 2024.

## Abb. 18

Refsgaard, A. (2020). *Erasing & Enhancing Essentials*. <https://www.andreasrefsgaard.dk/projects/erasing-enhancing-essentials/> abgerufen am 27. September 2024.

## Abb. 19

Anadol, R. (2022). *Unsupervised: Machine Hallucinations, MoMA*. Refik Anadol Studio. <https://refikanadolstudio.com/projects/unsupervised-machine-hallucinations-moma/> abgerufen am 27. September 2024.

## Abb. 20

Fernandez, J. (2024). *Jumoke Studio Projects*. <https://jumoke.studio/> abgerufen am 27. September 2024.

## Abb. 21

Hellandsjø, K. (o. D.). *Jackson Pollock biography*. Store Norske Leksikon. [https://snl.no/Jackson\\_Pollock](https://snl.no/Jackson_Pollock) (Creative Commons Attribution-ShareAlike 4.0 International) abgerufen am 27. September 2024.

## Abb. 22

AGIT. (2007). *Schutzrechte-Handbuch*. AGIT. [https://www.agit.de/fileadmin/Media/Fuer\\_Gruender/Schutzrechte-Handbuch/Schutzrecht\\_Handbuch.pdf](https://www.agit.de/fileadmin/Media/Fuer_Gruender/Schutzrechte-Handbuch/Schutzrecht_Handbuch.pdf), S. 9 abgerufen am 27. September 2024.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne Verwendung anderer als der angegebenen Hilfsmittel angefertigt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Gedanken habe ich unter Angabe der Quelle kenntlich gemacht. Dies gilt auch für bildliche Darstellungen sowie für Quellen aus dem Internet.

Mit einer Einsichtnahme und Ausleihe in der Bibliothek der HAW Hamburg, Department Design bin ich einverstanden.

Ort, Datum, Unterschrift

Hamburg, der 27. September 2024