



TECHNISCHE HOCHSCHULE
OSTWESTFALEN-LIPPE
UNIVERSITY OF
APPLIED SCIENCES
AND ARTS



EVERYTHING MACHINE

*Ein explorativer Artistic Research-Prozess zur visuellen
Entwicklung der digitalen Künstleridentität 'Kepler'
mittels multimodaler KI-Systeme*

Jennifer Meier

Technische Hochschule Ostwestfalen-Lippe
Fachbereich Medien und Kultur, Master Medienproduktion

Artistic Research

Prof. Dr. Aristotelis Hadjakos

Februar 2026

Abstract

Die vorliegende Arbeit untersucht im Rahmen einer Practice-Led Research mit autoethnografischem Ansatz, wie der Einsatz multimodaler KI-Systeme den kreativen Prozess bei der Entwicklung einer digitalen Künstleridentität verändert. Am Beispiel von Kepler – einer maskierten, voxelisierten Musiker-Persona – werden fünf Experimente mit KI-Tools aus den Bereichen Bildgenerierung, Musikproduktion, Webentwicklung und Texterstellung dokumentiert. Die Analyse identifiziert vier zentrale Prozessveränderungen: (1) Flow entsteht unvorhersehbar – der Website-Build mit KI gelang produktiv, während die Bildgenerierung für Kepler an der Nischen-Ästhetik scheiterte; (2) KI beschleunigt und begrenzt gleichzeitig – Exploration und Prototyping werden schneller, aber die Kontrolle über ästhetische Details geht verloren; (3) die Autorschaft verschiebt sich fundamental – von der Ausführenden zur Kuratorin, Dirigentin, Lernenden; (4) der emotionale Zyklus zwischen KI-Enthusiasmus und KI-Burnout erweist sich als reale Belastung im kreativen Prozess – der ständige Zwang, neue Tools zu evaluieren und mit Limitierungen zu kämpfen, kostet kreative Energie.

Keywords: Artistic Research, generative KI, Mensch-KI-Kollaboration, digitale Identität, Autoethnografie, kreativer Prozess, Practice-Led Research

Inhaltsverzeichnis

Einleitung	5
Hintergrund und Motivation	5
Forschungsfrage und Überblick	5
Stand der Forschung	6
Artistic Research als Erkenntnisform	6
Kreativität und Mensch-KI-Kollaboration	6
Artistic Research mit generativer KI	7
Forschungslücke	8
Methodologie	8
Methodischer Rahmen: Practice-Led Research	8
Konkrete Methoden	8
Ethische Überlegungen und KI-Offenlegung	9
Meine Forschung: Der Weg zu Kepler	9
Ausgangspunkt und erste Schritte	9
Entscheidungen und Abzweigungen	9
Die Wahl des visuellen Stils	9
Die Website als Forschungsartefakt	10
Von der Intention zum Experiment	10
Experimente und Ergebnisse	10
Experiment 1: Visuelle Konsistenz (ComfyUI + Flux)	10
Experiment 2: Generierte Welten (Marble / World Labs)	11
Experiment 3: Musik für ein Alter Ego (Suno AI)	11
Experiment 4: KI-Dialog als Reflexionswerkzeug	11
Experiment 5: Automatisierte Bildpipeline (fal.ai API + Claude Code)	12
Experiment 6: Video-Generierung (fal.ai API + Claude Code)	29
Zentrale Beobachtungen aus dem Prozess	37
1. Unvorhersehbarer Flow	37
2. Speed vs. Control	38
3. Verschiebung der Autorschaft	38
4. KI-Burnout und KI-Enthusiasmus	39
Rückbezug auf die Forschungsfrage	39

Diskussion und Interpretation	40
Zusammenfassung der Ergebnisse	40
Theoretische Implikationen	41
Praktische Implikationen	41
Stärken und Limitierungen	42
Stärken	42
Limitierungen	42
Practice-Based Intention, Practice-Led Realität	42
Ausblick	43
Literaturverzeichnis	43
Anhang	45
Anhang A: Verzeichnis der Journal-Einträge	45
Anhang B: Journal-Einträge und Reflexionsinterviews	46
Journal-Eintrag 1: Referenzen für Kepler sammeln (15. November 2025)	46
Journal-Eintrag 2: ComfyUI Consistent Character Workflow (16. November 2025)	49
Journal-Eintrag 3: Marble von Worldlabs (18. November 2025)	54
Journal-Eintrag 4: Website mit Cursor aufsetzen (19. November 2025)	60
Journal-Eintrag 5: KI wirklich verstehen – Teil 1 (11. Januar 2026)	65
Journal-Eintrag 6: Suno-Experiment (12. Januar 2026)	72
Journal-Eintrag 7: Kepler-Recherche, zweiter Versuch (25. Januar 2026)	77
Eidesstattliche Erklärung	86

Einleitung

Hintergrund und Motivation

Die vorliegende Arbeit versteht Artistic Research im Sinne von Henk Borgdorff (2012) als künstlerische Praxis, die sich als Forschung qualifiziert: Man macht etwas, findet dabei etwas heraus und dokumentiert es so, dass andere davon lernen können. Zentral ist die Absicht, über die rein persönliche Entwicklung hinaus Wissen zu erzeugen – durch die Praxis selbst, nicht nur über sie.

Diese Arbeit entsteht in einem Moment, in dem generative KI nahezu alle kreativen Domänen erfasst hat: DALL-E, Midjourney und Stable Diffusion für Bilder; ChatGPT und Claude für Text; Suno und Udio für Musik; Runway und Sora für Video. Eine einzelne Person kann nun Visuals, Texte, Musik und Videos in einer Qualität erstellen, die früher ganze Teams erforderte. Gleichzeitig wirft diese Entwicklung Fragen auf, die nicht abstrakt-philosophisch sind, sondern die konkrete Praxis von Künstler*innen und Medienproduzent*innen betreffen: Was bedeutet Autorschaft, wenn Maschinen mitschreiben? Was ist Kreativität, wenn Algorithmen Bilder generieren?

Im Kontext meiner eigenen Praxis traf diese Frage auf ein konkretes Projekt: „Kepler“ – das musikalische Alter Ego meines Kollaborateurs Gavin, eine maskierte, voxelisierte Musiker-Persona. Kepler steht in einer Transformation: von einer 2D-Figur mit Illustrationen hin zu einer 3D-Präsenz mit Covern, Reels und einer eigenen visuellen Welt. Ich unterstützte diesen Übergang im visuellen Bereich, und genau dieser Prozess wurde zum Ausgangspunkt meiner Forschung: Kann generative KI die visuelle Konsistenz liefern, die Kepler braucht? Und was passiert dabei mit meinem kreativen Prozess, wenn ich nicht mehr nur Werkzeuge benutze, sondern mit ihnen zusammenarbeite?

Das Projekt „Everything Machine“ dokumentiert diesen Prozess transparent auf einer öffentlichen [Website](#), der vollständige Quellcode ist in einem [GitHub-Repository](#) einsehbar. Der Name verweist auf die Idee einer Maschine, die „alles“ produzieren kann – und fragt zugleich, was dieses „alles“ wert ist. Denn je mehr ich mit generativer KI arbeitete, desto deutlicher wurde: Sie kann vieles schnell, aber nicht alles gut. Die „Everything Machine“ ist zugleich mächtig und begrenzt, und genau diese Spannung macht sie zum Gegenstand künstlerischer Forschung.

Forschungsfrage und Überblick

Die zentrale Forschungsfrage lautet:

Wie verändert der Einsatz multimodaler KI-Systeme (Text, Bild, Audio) den kreativen Prozess bei der Entwicklung einer digitalen Künstleridentität?

Im Laufe der Praxis veränderte sich nicht nur der kreative Prozess selbst, sondern auch die Art, ihn zu dokumentieren: Die Website nahm einen größeren Raum ein als geplant, das KI-gestützte Reflexionsinterview etablierte sich als Methode, wurde dann als zu knapp kritisiert. Diese Verschiebungen bestätigten, dass die Prozessfrage die richtige war: Nicht das Ergebnis, sondern die Veränderung des kreativen Prozesses durch KI erwies sich als zentrale Erkenntnis.

Zur Beantwortung dieser Frage verbindet die Arbeit:

1. **Praktische Experimente:** Iterative Entwicklung von Kepler durch verschiedene KI-Tools, beobachtet auf Prozessveränderungen
2. **Theoretische Reflexion:** Einordnung in Diskurse zu Kreativität, Autorschaft und Human-AI Interaction. Die strukturierten Reflexionsinterviews orientieren sich dabei an fünf Subfragen: Wie

verändert sich der Workflow? Wie verschiebt sich die Autorschaft? Welche Rolle spielt Iteration?

Wie wird mit Scheitern umgegangen? Und wie entwickelt sich die Ästhetik im Prozess?

3. Methodische Transparenz: Öffentliche Dokumentation des gesamten Prozesses

Die Arbeit gliedert sich in fünf Abschnitte: Nach der Einleitung folgt eine Darstellung des Forschungsstands (Abschnitt 2), dann die methodische Beschreibung (Abschnitt 3), die Darstellung meiner Forschung (Abschnitt 4) und schließlich die Diskussion (Abschnitt 5).

Stand der Forschung

Artistic Research als Erkenntnisform

Die theoretische Grundlage für dieses Projekt bildet das Feld des Artistic Research. Borgdorff (2012) unterscheidet zwischen Forschung *über*, *für* und *in* den Künsten. Die vorliegende Arbeit fällt in die dritte Kategorie: Sie forscht nicht über Kunst von außen, sondern aus der Praxis heraus. Artistic Research positioniert künstlerische Praxis nicht als Illustration wissenschaftlicher Erkenntnisse, sondern als eigenständige Erkenntnisform.

Candy unterscheidet in ihrem einflussreichen Report am Creativity & Cognition Studios der University of Technology Sydney zwischen *practice-based* und *practice-led* Research. Practice-based Research bedeutet, dass die kreative Praxis zentral für die Forschung ist und das Kunstwerk selbst Teil des Forschungsergebnisses wird; practice-led Research hingegen nutzt die Praxis, um neues Wissen über Praxis zu generieren, ohne dass das Kunstwerk notwendigerweise Ergebnis ist. Die vorliegende Arbeit ist *practice-led*: Die Praxis (KI-Experimente, Website-Bau, Musikgenerierung) führt zu neuem Wissen über den kreativen Prozess. Die Artefakte belegen die Erkenntnisse, sind aber nicht selbst das primäre Forschungsergebnis.

Kreativität und Mensch-KI-Kollaboration

Boden (2004) unterscheidet drei Formen von Kreativität: explorative, kombinatorische und transformative. Generative KI-Systeme operieren primär explorativ und kombinatorisch — sie erkunden einen „conceptual space“ und kombinieren erlernte Muster auf neue Weise. Ob sie zu transformativer Kreativität fähig sind, bleibt umstritten. Für diese Arbeit ist die Frage weniger relevant als die praktische Beobachtung: In der Kollaboration mit KI entstehen Ergebnisse, die weder ich allein noch das System allein hervorgebracht hätte.

Galanter (2003) definiert generative Kunst als jede Praxis, bei der ein autonomes System zum Produktionsprozess beiträgt — vom Computerprogramm über physikalische Prozesse bis zu Regelwerken. KI-gestützte Kunstproduktion steht damit in einer langen Tradition. Gleichzeitig ist der aktuelle Moment qualitativ anders: Multimodale KI-Systeme operieren in Text, Bild und Audio gleichzeitig, mit historisch beispieloser Geschwindigkeit und Zugänglichkeit. Galanter ordnet generative Systeme auf einer Ordnung-Chaos-Skala an, wobei „effektive Komplexität“ — die Zone zwischen perfekter Ordnung und totalem Chaos — den ästhetisch interessantesten Bereich markiert. Dieses Konzept wird in den Experimenten dieser Arbeit mehrfach aufgegriffen.

Deterding et al. (2017) beschreiben ein Spektrum kreativer Mensch-Computer-Interaktion: von „Mensch als Schöpfer, Computer als Werkzeug“ über „Mixed-Initiative“ — beide Seiten kollaborieren in einem engen Feedback-Loop — bis „Computer als Schöpfer, Mensch als Publikum“. Die Entwicklung von Kepler findet genau in diesem Mixed-Initiative-Raum statt, in dem weder ich noch das

KI-System die volle Kontrolle hat.

Franceschelli und Musolesi (2024) identifizieren in ihrem systematischen Überblick eine zentrale Unterscheidung: Kreativität *des* Systems versus Kreativität *mit* dem System. In dieser Arbeit geht es um Letzteres — nicht ob die KI kreativ „ist“, sondern wie sich der kreative Prozess in der Kollaboration verändert.

Mazzone und Elgammal (2019) argumentieren, dass KI-Systeme sowohl als Werkzeug als auch als kreativer Partner fungieren können, betonen aber die Grenzen rein generativer Ansätze: Ohne menschliche Intention, Kontextwissen und ästhetisches Urteil bleibt der Output generisch. Diese Beobachtung deckt sich mit dem „Garbage in, garbage out“-Prinzip, das sich durch meine Experimente zieht: Domänenwissen erweist sich als Voraussetzung für ästhetisch relevante Ergebnisse.

Suh et al. (2021) untersuchen, wie generative KI in kollaborativen Kompositionenprozessen wirkt. Ihre Ergebnisse zeigen, dass KI nicht nur als technisches Werkzeug, sondern als „sozialer Klebstoff“ fungieren kann, der Kommunikation und kreative Aushandlung zwischen menschlichen Beteiligten verändert. Dieser Befund resoniert mit meinem Suno-Experiment, bei dem die KI-generierten Beschreibungen — obwohl die Musik selbst enttäuschte — ein gemeinsames ästhetisches Vokabular lieferten, das für die Weiterarbeit produktiv wurde.

Zwei bekannte Phänomene treten beim praktischen Arbeiten mit generativer KI wiederholt auf. Das *Uncanny Valley* (Mori et al., 2012) beschreibt das Unbehagen, das entsteht, wenn künstliche Darstellungen menschenähnlich, aber nicht perfekt realistisch sind. In der Arbeit mit Kepler zeigt sich dieses Phänomen als konkretes Hindernis — Bildmodelle halluzinieren Gesichtszüge in seine Maske, weil sie die Abwesenheit eines Gesichts nicht akzeptieren. Eng damit verbunden ist Coleridges Konzept der *Suspension of Disbelief* (Coleridge, 1817/1984): die Bereitschaft des Publikums, sich auf eine fiktionale Welt einzulassen. Für Kepler als virtuelle Persona stellt sich die Frage, ob Recipient*innen ihn als „authentischen“ Künstler akzeptieren — eine Frage, die im kreativen Prozess ständig mitverhandelt wird und durch KI-Artefakte immer wieder gestört wird.

Artistic Research mit generativer KI

Neben der theoretisch-empirischen Forschung gibt es eine wachsende Zahl von Artist-Practitioner-Projekten, die generative KI aus der Praxis heraus untersuchen. Monin und Sadokierski (2025) beschreiben Prompting als „Thinking-With“ — einen Denkprozess, bei dem die Formulierung von Prompts selbst zur Erkenntnismethode wird. Ihr Projekt zur Visualisierung eines ausgestorbenen Zwerp-Emus verbindet Collagepraktiken mit KI-Bildgenerierung und zeigt, wie die iterative Auseinandersetzung mit KI-generierten Bildern das Verständnis des Forschungsgegenstands vertieft.

Hansen (2023) entwickelt ein Modell für die sich verändernde Rolle von Grafikdesigner*innen im KI-Zeitalter. Gestützt auf die Trias von Deduktion, Induktion und Abduktion argumentiert er, dass deduktive und induktive Designaufgaben effektiv an KI delegierbar sind, während abduktive Aufgaben weiterhin menschliche Kompetenz erfordern — die Kernkompetenz verschiebt sich von der Ausführung zur Kuration und kritischen Bewertung, eine Beobachtung, die sich mit meinem Befund der Autorschaftsverschiebung deckt. Torres Núñez del Prado (2025) untersucht in „Encoding Culture“, wie kulturelle Kodierungen in KI-Trainingsdaten eingebettet sind und wie Künstler*innen diese sichtbar machen und unterlaufen können — eine Perspektive, die für Keplers Konfrontation mit normativen Gesichtsmodellen direkt relevant ist.

Diese Projekte teilen den reflexiven Ansatz der vorliegenden Arbeit, arbeiten aber primär in einer einzelnen Modalität (Bild oder Text). Die vorliegende Arbeit unterscheidet sich durch ihren multi-

modalen Ansatz: Die Experimente erstrecken sich über Bild, Audio, 3D und Code und ermöglichen so eine domänenübergreifende Analyse der Prozessveränderungen.

Forschungslücke

Während autoethnografische Studien zu KI-gestützten kreativen Prozessen in einzelnen Domänen existieren – insbesondere in der Musikkomposition und der visuellen Kunst – fehlen Tiefenstudien, die den iterativen Entstehungsprozess *multimodaler* hybrider Kunstwerke aus der Perspektive der beteiligten Künstler*innen dokumentieren. Die oben genannten Arbeiten von Franceschelli und Musolesi (2024), Mazzone und Elgammal (2019) sowie die Artist-Practitioner-Quellen zeigen zwar das wachsende Feld, aber keine dieser Arbeiten verbindet systematische autoethnografische Reflexion mit multimodaler KI-Praxis – über Text, Bild und Audio hinweg – im Kontext der Entwicklung einer konkreten digitalen Künstleridentität. Diese Arbeit adressiert diese Lücke.

Methodologie

Methodischer Rahmen: Practice-Led Research

Das Projekt folgt dem Paradigma des Practice-Led Research, bei dem die künstlerische Praxis als Erkenntnisquelle dient und Wissen über den kreativen Prozess generiert. Die Entwicklung von Kepler ist nicht Gegenstand einer externen Untersuchung, sondern der Ort, an dem Forschung stattfindet.

Konkrete Methoden

Die Forschung verbindet drei methodische Stränge: autoethnografisches Journaling mit iterativer Experimentierung, eine öffentliche Website als Forschungsartefakt und einen KI-gestützten Reflexionsprozess.

Regelmäßige Journal-Einträge dokumentieren den kreativen Prozess nach einer strukturierten Vorlage (Intention, Prozess, Ergebnis, Reflexion, nächste Schritte). Systematische Experimente mit verschiedenen KI-Tools – Stable Diffusion, DALL-E, ChatGPT, Claude, Suno, Udio – folgen einem Zyklus von Hypothese, Experiment, Dokumentation und Anpassung. Beide Stränge greifen ineinander: Jedes Experiment mündet in einen Journal-Eintrag, jeder Eintrag informiert das nächste Experiment.

Die Website „Everything Machine“ macht den gesamten Prozess transparent zugänglich und ist selbst Teil des künstlerischen Outputs. Sie wurde bewusst ohne CMS, sondern als maßgeschneiderte Webanwendung entwickelt, deren technische Architektur die methodischen Prinzipien spiegelt: Transparenz durch Versionierung (Git), Iterativität durch modularen Aufbau, Selbstreflexion durch ein integriertes Journal-System.

Ein zentrales methodisches Element ist der KI-gestützte Reflexionsprozess: Nach jedem Experiment schreibe ich einen Journal-Eintrag und führe anschließend eine strukturierte Konversation mit Claude, die den Eintrag liest, Rückfragen stellt und die Reflexion entlang fester Fragen verdichtet. Diese Fragen basieren auf autoethnografischer Praxis (Ellis et al., 2011) und Schöns „Reflection-in-Action“ (Schön, 1983) und decken Kontext, Prozess, Erkenntnisse und Weiterentwicklung ab. Die KI wird dabei zum Sparringspartner für die Selbstreflexion, nicht zum Autor: Sie strukturiert und verdichtet, aber die Erkenntnisse kommen aus meiner Erfahrung.

Ethische Überlegungen und KI-Offenlegung

Da es sich um ein autoethnografisches Projekt handelt, werden keine personenbezogenen Daten Dritter erhoben. Die verwendeten KI-Modelle wurden auf Datensätzen trainiert, deren Zusammensetzung nicht vollständig transparent ist und deren urheberrechtlicher Status Gegenstand laufender Debatten bleibt. Dieses Spannungsfeld wird als Teil der Forschung reflektiert, nicht verschwiegen.

Generative KI kam in dieser Arbeit auf vier Ebenen zum Einsatz: (1) als **Experimentgrundlage** – Stable Diffusion, DALL-E, Suno AI und Marble bildeten den Forschungsgegenstand; (2) als **Reflexionspartner** – nach jedem Experiment wurde eine strukturierte Reflexion im Dialog mit Claude durchgeführt (vollständige Transkripte in Anhang B); (3) als **Schreibassistent** – Claude unterstützte bei Strukturierung und Formulierung dieser Dokumentation; (4) als **Entwicklungstool** – die Projektwebsite wurde mit Cursor gebaut (selbst Gegenstand von Experiment 4). Argumentation, Quellenauswahl und alle inhaltlichen Entscheidungen liegen bei der Autorin. Der gesamte Prozess ist auf der Website und im GitHub-Repository öffentlich dokumentiert.

Meine Forschung: Der Weg zu Kepler

Ausgangspunkt und erste Schritte

Am Anfang dieser Forschung stand kein theoretisches Konzept, sondern ein ganz praktisches Problem: Kepler brauchte Bilder. Videos. Eine Welt. Kepler existierte bereits als musikalische Identität – als Name, als Idee, als Figur mit Maske und voxelisiertem Stil. Aber visuell war er fragmentarisch: ein paar Fotos, ein paar Renders aus Unreal Engine, zusammengehalten von einer ästhetischen Intuition, die ich teilte mit Gavin, meinem musikalischen Kollaborateur. Was fehlte, war Konsistenz. Und die Frage, ob generative KI diese Konsistenz liefern konnte, wurde zum Ausgangspunkt meiner Forschung.

Der erste Schritt kam ohne KI aus. Gavin und ich sammelten Referenzen: Pinterest-Boards für Cover-Inspiration, Kleidung, Song-spezifische Vibes. Ein Google-Drive-Ordner für geteilte Ressourcen: Fotos, existierende Artworks, Konzeptideen. Kein KI-Tool kam zum Einsatz. Was mir dabei bewusst wurde: Der Prozess des Sammelns, Sortierens und Kuratierens ist ein zutiefst menschlicher. Selbst wenn einige der gesammelten Bilder selbst KI-generiert gewesen sein mögen, die Auswahl, die Kombination, das Gespür für das Richtige, das bleibt bei uns. Diese Beobachtung wurde zum stillen Leitmotiv der gesamten Forschung: Wo genau verläuft die Grenze zwischen menschlicher Kuration und maschineller Generierung?

Entscheidungen und Abzweigungen

Die Wahl des visuellen Stils

Kepler trägt eine Maske. Er hat kein sichtbares Gesicht. Sein Stil ist voxelisiert, blockhaft, bewusst abstrahiert. Das ist keine kosmetische Entscheidung, sondern eine identitätsstiftende: Die Maske ist das, was Kepler von einem gewöhnlichen Avatar unterscheidet. Sie schafft Distanz und Projektionsfläche zugleich.

Genau diese Designentscheidung erwies sich als fundamentale Herausforderung für die Arbeit mit generativer KI. Bildgenerierungsmodelle wie Flux oder Stable Diffusion sind auf Trainingsdaten auf-

gebaut, die überwiegend menschliche Gesichter enthalten. Das Modell "will" ein Gesicht sehen, wo keines sein soll; es projiziert hartnäckig Gesichtszüge in Keplers Maske, selbst wenn der Prompt explizit dagegen formuliert ist. Keplers Ästhetik liegt in einer Nische, die das Modell nicht versteht, weil sie in den Trainingsdaten nicht oder kaum repräsentiert ist. Was bei einer realistischen, Anime- oder Cartoon-Figur funktioniert hätte, scheiterte an Keplers bewusster Andersartigkeit.

Die Website als Forschungsartefakt

Eine zentrale Entscheidung war, die Dokumentationswebsite nicht mit einem fertigen CMS zu erstellen, sondern von Grund auf mit Cursor zu bauen – einem KI-gestützten Code-Editor. Das war keine rein pragmatische Wahl. Die Website ist selbst Teil des künstlerischen Outputs: Ihr dunkles Theme, die Pixel-Headings in "Press Start 2P"-Font, der animierte Matrix-Gradient, der integrierte Musikplayer – all das transportiert Keplers digitale Identität. Die Website ist nicht Rahmen für die Forschung, sie ist Forschung.

Der Bauprozess mit Cursor offenbarte eine weitere Facette der Mensch-KI-Kollaboration. Ich fühlte mich, wie ich in meinem Journal notierte, "wie eine Managerin mit Gesamtvision, die einen Junior Developer dirigiert". Die KI war schnell bei repetitiven Aufgaben, aber bei komplexen ästhetischen Entscheidungen – dem richtigen Abstand, dem richtigen Farbton, der richtigen Interaktion – war ich diejenige, die führte. Mein Vorwissen in Webentwicklung war dabei entscheidend: "Garbage in, garbage out", notierte ich. Je spezifischer die Anforderungen und je größer die Codebase, desto schwieriger wird es für jemanden ohne Vorwissen, die Qualität zu steuern.

Von der Intention zum Experiment

Die Forschungsfrage lenkte die Experimentwahl nicht wie ein Skript, sondern wie ein Kompass. Ich wollte verschiedene Modalitäten testen: Bild, 3D, Code, Audio. Nicht um Vollständigkeit zu demonstrieren, sondern weil Kepler multimodal existiert und weil ich verstehen wollte, ob die Muster der Kollaboration über Domänen hinweg stabil bleiben. Die Methodik des Practice-Led Research (Candy, 2006) legitimierte dabei einen explorativen Ansatz: Die Praxis selbst generiert die Erkenntnisse, nicht ein vorab definiertes Versuchsdesign.

Experimente und Ergebnisse

Experiment 1: Visuelle Konsistenz (ComfyUI + Flux)

Mein erster ernsthafter Versuch, Kepler visuell konsistent mit KI zu generieren, nutzte ComfyUI mit dem Flux-Modell, ergänzt durch IP-Adapter, ControlNet und Face Detailer. Die Installation auf meinem M1 MacBook Pro dauerte Stunden und verschlang viele Gigabyte, schon das ein Hinweis darauf, dass es sich nicht um die "Zaubertools" handelt, als die generative KI oft dargestellt wird.

Die Ergebnisse waren gemischt. Wenn ich ein Foto als Referenz nutzte, auf dem Kepler einen Basketball hielt, wurde der Basketball im generierten Bild beibehalten, obwohl er dort nicht hingehörte. Das Modell hat kein semantisches Verständnis davon, welche Elemente zusammengehören. Gravierender war das Gesichtsproblem: Face Detailer, eigentlich zur Verbesserung von Gesichtsdetails entwickelt, interpretierte Gesichter in Keplers Maske hinein – eine Art Face Hallucination, die selbst durch negative Prompts nicht vollständig unterdrückt werden konnte. Es handelt sich um das, was Mori (2012) als Uncanny Valley beschreibt, allerdings in einer unerwarteten Richtung: nicht ein fast-menschliches Gesicht, das Unbehagen auslöst, sondern ein Nicht-Gesicht, das vom Modell zwanghaft vervollständigt wird. In Galanters (2003) Terminologie driften die Halluzinationen in den

Bereich der Unordnung ab — weg von der “effektiven Komplexität”, die Keplers Ästhetik ausmacht. Die Erkenntnis war ernüchternd: Art Directability – die Fähigkeit, die KI gezielt zu steuern – war nicht gegeben. Schöns (1983) “Reflection-in-Action” zeigte sich hier in seiner elementarsten Form: Die Situation sprach zurück, und zwar anders als erwartet. Meine Reaktion war nicht, den Prozess aufzugeben, sondern die Parameter zu variieren, den Workflow anzupassen, neue Referenzbilder zu testen. Die Modelle sind auf Mainstream-Ästhetiken trainiert. Keplers Nische fällt durch das Raster.

Experiment 2: Generierte Welten {Marble / World Labs}

Nachdem Kepler als Charakter in der Bildgenerierung an Grenzen gestoßen war, verlagerte ich den Fokus auf Hintergrundwelten. Marble, das erste kommerzielle Produkt von Fei-Fei Lis World Labs, ermöglicht die browserbasierte Erstellung von 3D-Welten durch Gaussian Splats. Ich testete vier verschiedene Ansätze: ein Foto aus einem sowjetisch inspirierten Café, ein Foto von Kepler, ein Pinterest-Bild mit Prompt-Beschreibung und eine reine Text-Prompt, die Claude aus mehreren Pinterest-Bildern synthetisiert hatte.

Der aufschlussreichste Kontrast lag zwischen den letzten beiden Ansätzen. Das direkte Foto ergab eine technisch korrekte, aber kreativ uninteressante Rekonstruktion: Das Modell orientierte sich zu nah am Input. Die Claude-generierte Prompt hingegen produzierte eine Welt, die sich weniger wie eine Rekonstruktion und mehr wie eine Interpretation anfühlte. Vibes statt Referenzbilder: ein Workaround, der ästhetisch produktiver war als der intendierte Workflow.

Bei Personen und Spiegelungen zeigte sich erneut das Uncanny Valley: verzerrte Proportionen, gruselige Doppelgänger. Die Frage nach Coleridges (1817/1984) “willing suspension of disbelief” stellte sich hier unmittelbar: Kann man in eine generierte Welt eintauchen, wenn die Figuren darin Unbehagen auslösen? In der Reflexion entstand die Idee, genau diese “Fehler” als ästhetisches Mittel einzusetzen, eine Perspektive, die sich durch alle folgenden Experimente ziehen sollte.

Experiment 3: Musik Für ein Alter Ego {Suno AI}

Im Januar 2026 testete ich Suno AI für die Generierung von Musik – minimal, loopbar, sowjetisch-dystopisch. Zwei Prompt-Ansätze: “Lo-fi” produzierte zu kommerziell klingende Ergebnisse; “Pixel Game Soundtrack” ergab etwas, das eher nach Hero-Soundtrack eines Indie-Games klang als nach atmosphärischer Hintergrundmusik.

Das Kernphänomen dieses Experiments war die Diskrepanz zwischen Beschreibung und Output. Suno generiert zu jedem Track eine detaillierte Beschreibung: “Bitcrushed chiptune pads over a sparse, detuned synth choir; slow, mechanical pulse with occasional metallic hits and distant wind FX.” Diese Beschreibung trifft exakt das, was ich mir vorstellte. Der tatsächliche Sound ging in eine andere Richtung: zu poliert, zu kommerziell, Stock-Music-Qualität. In Galanters (2003) Terminologie produziert Suno Output mit hoher Ordnung, aber niedriger effektiver Komplexität — zu glatt und vorhersehbar für die angestrebte Lo-Fi-Pixel-Ästhetik. Es war dasselbe Muster wie bei ComfyUI: Generalistische Modelle defaulten auf Mainstream-Ästhetik. Die Nische, die ich suchte, erfordert entweder viele Iterationen, spezialisierte Fine-Tunes oder menschliche Expertise.

Experiment 4: KI-Dialog als Reflexionswerkzeug

Ein unkonventionelleres Experiment: Auf einer 30-minütigen Autofahrt führte ich ein Voice-Gespräch mit Claude über KI-Grundlagen – von Shannons Informationstheorie über die Dartmouth Conference bis zur Generierung von 3D-Welten.

rence bis zur Transformer-Architektur. Das war kein Experiment im engeren Sinne, sondern eine Erkundung der KI als personalisiertem Lernpartner.

Die wichtigste Erkenntnis betraf die Fluency Illusion: Das Gespräch fühlte sich an wie tiefes Verstehen, aber ob ich die Konzepte tatsächlich internalisiert hatte, war unklar. Claude selbst wies mich darauf hin, dass es einen Unterschied gibt zwischen deklarativen Wissen ("wissen, dass") und prozeduralem Wissen ("wissen, wie"). Die Theorie wird erst real, wenn ich sie anwende, eine Beobachtung, die auf Schöns (1983) Konzept des "Reflection-in-Action" zurückführt: Erkenntnis entsteht nicht im Hören oder Lesen, sondern im Tun.

Experiment 8: Automatisierte Bildpipeline (Fal.ai API + Claude Code)

Nach vier Einzelexperimenten stellte sich eine methodische Frage: Was passiert, wenn ich den Prozess systematisiere? Gemeinsam mit Claude Code baute ich eine automatisierte Pipeline, die über die fal.ai API neun verschiedene Bildgenerierungsmodelle mit denselben Prompts und Referenzbildern ansteuerte. Die Kollaboration war dreistufig: Ich definierte die ästhetische Vision und die Evaluationskriterien, Claude Code schrieb den Pipeline-Code und orchestrierte die API-Aufrufe, die Bildmodelle generierten die Outputs – eine neue Konstellation, in der ich als Auftraggeberin eines KI-Agenten fungierte, der seinerseits KI-Modelle steuert.

Die Pipeline durchlief sieben Phasen, die den Möglichkeitsraum zwischen neun Modellen, zwei Input-Typen und drei Signature Scenes systematisch absteckten. Insgesamt entstanden 134 Bilder in 24 Output-Ordnern, evaluiert nach "Gefällt mir", "Halluzinationen" und Kommentar.

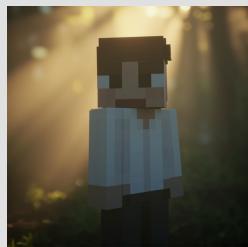
Phase 1: Modell-Screening (5 Modelle, A-Pose) Als Referenzbild diente Keplers A-Pose-Render – ein quadratischer 1024×1024 Render mit seitlich ausgestreckten Armen. Der Base-Prompt für alle fünf Modelle lautete:

"A voxel character made of small 3D cubes, faceless geometric humanoid figure, white shirt and dark pants, standing in a dramatic cinematic environment, volumetric lighting, no face, blocky pixel art style 3D character"

Ergänzt durch drei Szenen-Varianten: ...in a neon-lit cyberpunk city at night'',...in a foggy forest with golden light rays" und "...on a concert stage with dramatic spotlights". Fünf Modelle wurden getestet:



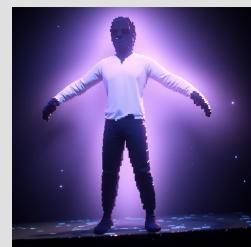
Abbildung 1: Input: Kepler A-Pose (1024×1024)



MiniMax Subject Reference



Z-Image Turbo ControlNet



FLUX Depth LoRA



FLUX Canny LoRA

Abbildung 2: Phase 1: Modell-Screening – MiniMax, Z-Image, FLUX Depth, FLUX Canny. Bilder in der Bild-Galerie vergrößerbar.

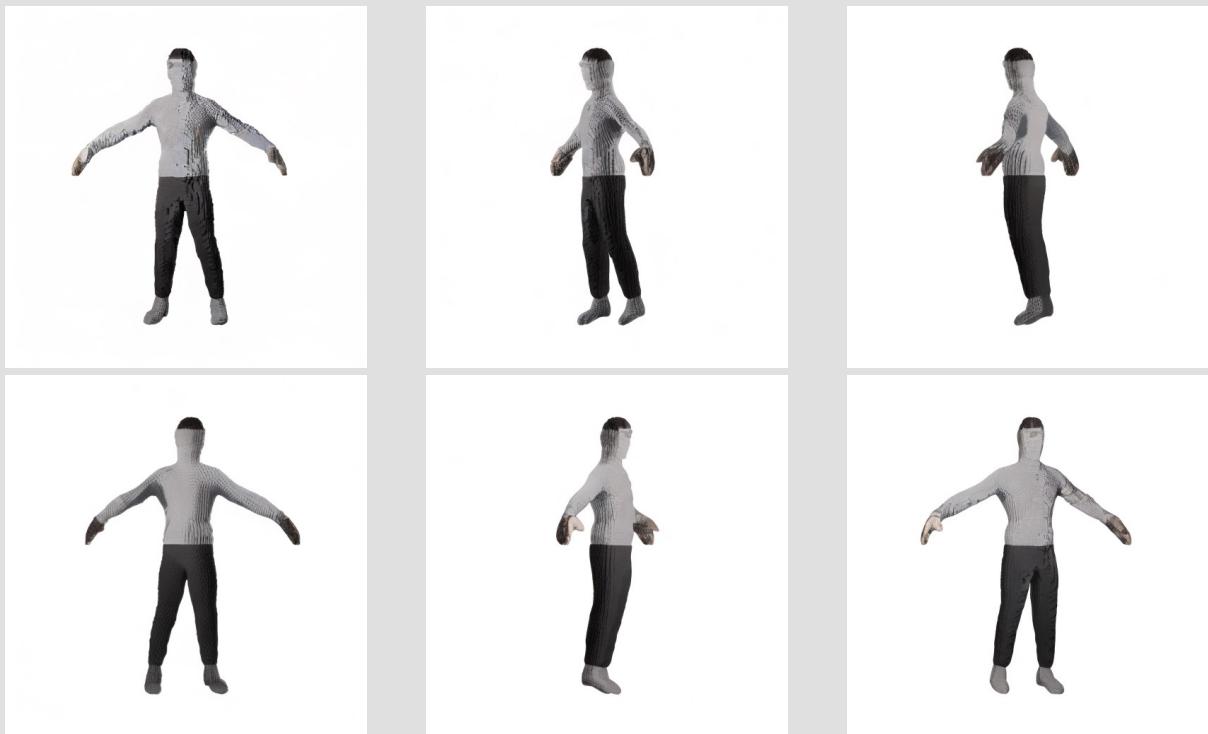


Abbildung 3: Phase 1: Era3D Multiview – nutzlos im Vergleich zu eigenen Renders. Bild-Galerie

Tabelle 1: Phase 1: Modell-Screening (A-Pose, quadratisch)

Modell	Bewertung	Halluz.	Kommentar
MiniMax Subject	nein	ja	Minecraft-Stil, Gesichtszüge
Z-Image ControlNet	bedingt	nein	A-Pose erhalten, clean
FLUX Depth LoRA	nein	-	Körper wird humanoid
FLUX Canny LoRA	nein	ja	Volles Gesicht halluziniert
Era3D	nein	-	Nutzlos – eigene Renders besser

Verworfene Modelle. Keines der fünf Modelle lieferte überzeugende Ergebnisse. MiniMax Subject Reference interpretierte Keplers Voxel-Ästhetik als Minecraft-Welt und halluzinierte menschliche Gesichtszüge, obwohl der Prompt “faceless” explizit vorgab. FLUX Depth LoRA bewahrte zwar die Körperhaltung über eine Tiefenkarte, glättete aber die blockige Geometrie zu einem humanoid wirkenden Körper. FLUX Canny LoRA produzierte das gravierendste Problem: Ein vollständiges menschliches Gesicht erschien auf Keplers faceless Kopf – dasselbe Halluzinationsphänomen wie in Experiment 1. Era3D generierte Multiview-Ansichten, die in besserer Qualität direkt in Unreal Engine gerendert werden können. Einzig Z-Image Turbo ControlNet verhinderte Halluzinationen und bewahrte die A-Pose, blieb aber insgesamt zu limitiert für Szenen-Kontexte.

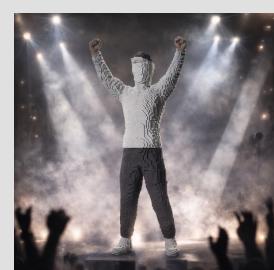
Methodisches Ergebnis. Depth-basierte Steuerung erwies sich als überlegen gegenüber Canny-Edges, und die Halluzinationsanfälligkeit variierte erheblich zwischen Modellen, die nominell dieselbe Aufgabe lösten. Für Phase 2 recherchierte ich gezielt leistungsstärkere Modelle – insbesondere FLUX Kontext Pro, GPT-Image-1.5, SeedDream und NanoBanana Pro –, die sich in Benchmarks als State-of-the-Art für Charakter-Konsistenz erwiesen hatten.

Phase 2: Top-Modelle in Szenen (A-Pose) Die vier vielversprechendsten Modelle wurden in Szenen-Kontexte eingeführt, jeweils drei Varianten mit demselben A-Pose-Input. Der Prompt folgte dem Schema:

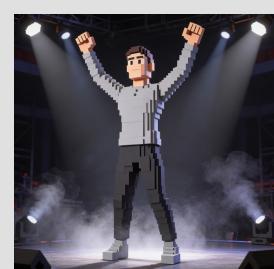
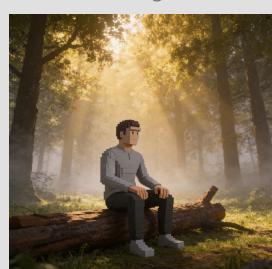
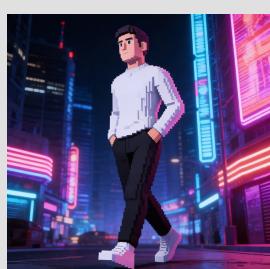
"Place this voxel character in [Szene], [Aktion], cinematic lighting"



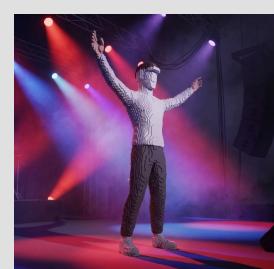
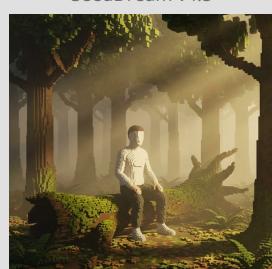
FLUX Kontext Pro



GPT-Image-1.5



SeedDream v4.5



NanoBanana Pro

Abbildung 4: Phase 2: Top-Modelle in Szenen (A-Pose). Bild-Galerie

Tabelle 2: Phase 2: Top-Modelle in Szenen (A-Pose)

Modell	Bewertung	Halluz.	Kommentar
FLUX Kontext Pro	bedingt	-	Wird besser
GPT-Image-1.5	bedingt	teilweise	Solide, aber Halluzinationen
SeedDream v4.5	nein	ja	Volles Gesicht halluziniert
NanoBanana Pro	gut	teilweise	Beste Phase-2-Ergebnisse

SeedDream schied wegen Gesichtshalluzinationen aus. Drei Finalisten blieben: FLUX Kontext Pro, GPT-Image-1.5 und NanoBanana Pro.

Phase 2b: Der Posed-Input-Durchbruch Der entscheidende Durchbruch kam mit dem Wechsel zum “Posed Input” – einem Render, in dem Kepler natürlich steht. Die A-Pose führte zu einem Overfitting-Effekt: Das Modell übernahm die starre Haltung wörtlich.

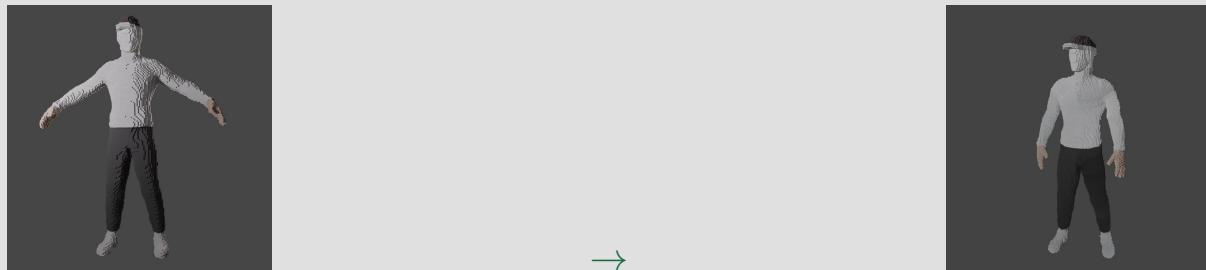


Abbildung 5: Input-Wechsel: A-Pose → Posed – natürliche Haltung statt T-Pose

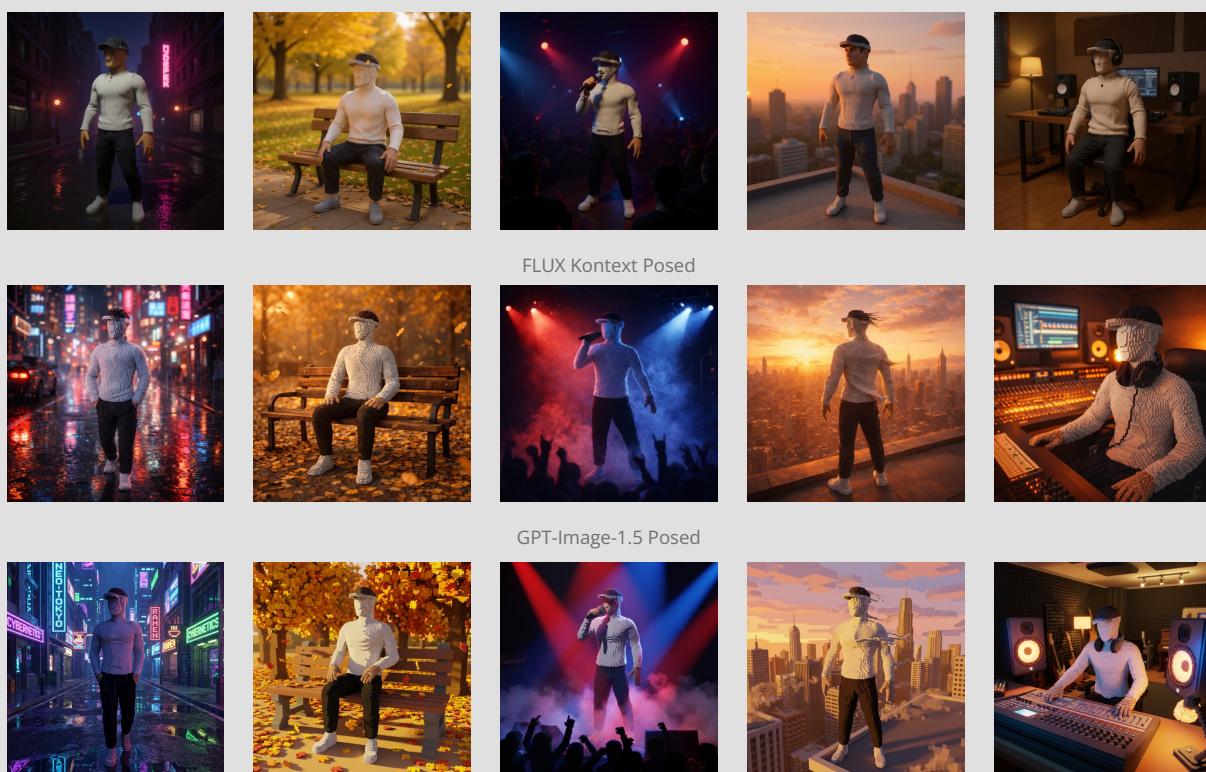


Abbildung 6: Phase 2b: Drei Finalisten mit Posed Input (je 5 Varianten). Bild-Galerie

Tabelle 3: Phase 2b: Posed Input (3 Finalisten)

Modell	Bewertung	Halluz.	Kommentar
FLUX Kontext Posed	nein	ja	Gesichter halluziniert
GPT-Image-1.5 Posed	ja	teilweise	Solideste Ergebnisse
NanoBanana Posed	ja	nein	Stärkste Voxel-Ästhetik

FLUX Kontext schied aus – Posed Input verstärkte die Gesichtshalluzinationen. Zwei Finalisten mit komplementären Stärken: GPT-Image-1.5 (atmosphärischer) und NanoBanana Pro (stilistisch konsistenter).

Phase 3a: Signature Scenes v1 – Ratio-Bug Ab Phase 3a wurden GPT-Image-1.5 und NanoBanana Pro parallel in "Signature Scenes" getestet: Studio, Retro Car Bridge, Spiral Staircase, Pool Floaty. Die vier Szenen-Prompts waren deutlich detaillierter als die vorherigen Phasen, z.,B. für die Studio-Szene:

"Place this voxel character in a professional recording studio, sitting at a large mixing console with both hands on the faders, studio monitors glowing, headphones on, warm amber lighting, acoustic foam panels on walls, DAW software visible on screens behind him, intimate atmosphere"

Für NanoBanana wurde der Prompt umformuliert: *"Show this voxel character in a cozy recording studio... IMPORTANT: The character must remain completely faceless."* Ein Ratio-Bug in der Pipeline betraf GPT – alle Outputs kamen quadratisch statt im gewünschten Format. NanoBanana Pro lieferte korrekte Portrait- und Landscape-Formate.



GPT-1.5 "Portrait" (13) – quadratisch durch Ratio-Bug



GPT-1.5 "Landscape" (13) – ebenfalls quadratisch



NanoBanana Portrait (14) – korrektes Format



NanoBanana Landscape (14) – korrektes Format

Abbildung 7: Phase 3a: GPT-1.5 mit Ratio-Bug (oben, quadratisch), NanoBanana korrekt (unten). [Bild-Galerie](#)

Tabelle 4: Phase 3a: Signature Scenes v1

Modell	Format	Bew.	Halluz.	Kommentar
GPT-1.5	portrait (sq.)	gemischt	-	Buggy 1:1 Ratio
GPT-1.5	landscape (sq.)	gemischt	-	Auch 1:1 leider
NanoBanana	portrait	gemischt	teilweise	Halb cool, halb Halluz.
NanoBanana	landscape	gemischt	teilweise	Ohne Halluz. stark



Gestretchte Inputs: Portrait (links), Landscape (rechts)

Abbildung 8: Phase 3b: Fehlerhafte Referenzbilder (gestretcht statt gepaddet)

Tabelle 5: Phase 3b: Gestretchte Inputs

Ordner	Modell	Format	Bew.	Halluz.	Kommentar
15/16	GPT-1.5	portrait	gemischt	-	Gestretzte Referenz
15/16	GPT-1.5	landscape	gemischt	teilweise	Stretching + Cap fehlt
15/16	NanoBanana	portrait	gemischt	nein	Stretching, keine Halluz.
15/16	NanoBanana	landscape	gemischt	teilweise	Weniger Stretching, gut
17/18	GPT-1.5	portrait	gemischt	teilweise	1:1 + leichte Halluz.
17/18	GPT-1.5	landscape	gemischt	teilweise	Dasselbe
17/18	NanoBanana	portrait	gemischt	-	-
17/18	NanoBanana	landscape	stark	-	-

Phase 3b: Gestretchte Inputs Der Ratio-Bug wurde behoben, Tokyo Rain ersetzte Bridge. Doch ein neuer Automatisierungsfehler trat auf: Das quadratische Referenzbild wurde naiv auf Portrait- bzw. Landscape-Format skaliert statt korrekt gepaddet – Kepler erschien verzerrt. Die Outputs zeigen deutlich sichtbare Proportionsverzerrungen: ein zu schmaler Körper im Portrait-Format, ein zu breiter im Landscape-Format. Trotzdem lieferte NanoBanana Pro selbst mit verzerrtem Input überzeugende Voxel-Ästhetik – ein Hinweis auf die Robustheit des Modells.

In der zweiten Iteration (Ordner 17/18) wurden die Prompts verfeinert: Voxel-Treppen statt Marmor, präzisere Szenenbeschreibungen. Das Stretching blieb jedoch bestehen, da der Fehler in der Input-Vorbereitung lag, nicht im Prompting. GPT-Image-1.5 produzierte erneut quadratische Outputs (1024×1024) trotz Portrait-Anfrage – derselbe API-Bug wie in Phase 3a. NanoBanana Pro hingegen lieferte korrekte Formate und die stärksten Landscape-Ergebnisse der gesamten Pipeline. Die Phasen 3a–3b zeigen Schöns (1983) “Situation, die zurückspricht” im Zeitraffer: Jede Iteration löste ein Problem und offenbarte das nächste. Die Bewertung erfolgte in diesen Phasen noch pauschal pro Modell und Format, da die Outputs insgesamt zu fehlerbehaftet waren, um sinnvoll zwischen einzelnen Szenen zu differenzieren.



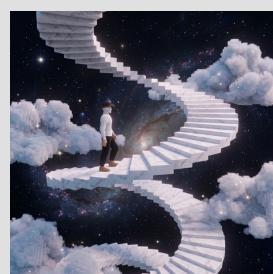
GPT-1.5 Portrait (15)



NanoBanana Portrait (16)



GPT-1.5 Portrait (17)



NanoBanana Landscape (18)

Abbildung 9:

Phase 3b: Outputs mit gestrechten Referenzbildern. [Bild-Galerie](#)



Korrekt gepaddete Inputs: Portrait (links), Landscape (rechts)

Abbildung 10: Phase 3c: Korrekt gepaddete Referenzbilder

Tabelle 6: Phase 3c: Korrigierte Pipeline

Modell	Format	Bew.	Halluz.	Kommentar
GPT-1.5	portrait	gemischt	-	-
GPT-1.5	landscape	gemischt	-	-
NanoBanana	portrait	gemischt	-	-
NanoBanana	landscape	stark	-	-

Phase 3c: Korrigierte Pipeline Mit korrekt gepaddeten Inputs – schwarze Balken statt Stretching – lieferte die Pipeline erstmals konsistente Ergebnisse. Night Drive ersetzte Tokyo Rain als dritte Szene, die Szenen-Auswahl stabilisierte sich auf das finale Set: Night Drive, Spiral Staircase, Pool Floaty und Studio. Die finalen Prompts waren nun hochspezifisch, z.B. für Night Drive:

"This voxel character is cruising down a night road in a sleek retro 1980s sports car with pop-up headlights. Behind him a glowing city skyline fades into the distance. Ahead, dark mountains rise on the horizon. Full moon in the sky, stars visible. Dashboard glowing softly, headlights cutting through mist. [...] IMPORTANT: The character must remain completely faceless — no eyes, no nose, no mouth, no sunglasses."

Der Unterschied war unmittelbar sichtbar: Keplers Proportionen stimmten, die Voxel-Geometrie blieb erhalten, und beide Modelle produzierten erstmals Outputs, die ohne Einschränkung als Promo-Material verwendbar wären. NanoBanana Pro dominierte im Landscape-Format mit durchgehend starken Bewertungen, während GPT-Image-1.5 im Portrait-Format atmosphärischere Szenen erzeugte. Die Korrektur eines einzigen Pipeline-Parameters – Padding statt Stretching – machte den Unterschied zwischen unbrauchbaren und überzeugenden Ergebnissen. Ab dieser Phase wurde jede Szene einzeln bewertet, da die Qualität nun hoch genug war, um szenenbezogene Unterschiede sinnvoll zu differenzieren.



GPT-1.5 Portrait (19)



NanoBanana Portrait (20)



GPT-1.5 Landscape (19)



NanoBanana Landscape (20)

Abbildung 11: Phase 3c: Outputs mit korrekt gepaddeten Inputs. Bild-Galerie

Phase 3d: Posed vs. A-Pose Vergleich Dieselben drei Signature Scenes – Night Drive, Spiral Staircase, Pool Floaty – wurden mit Posed (Ordner 21/22) und A-Pose (Ordner 23/24) Input parallel generiert.



Abbildung 12: Input-Vergleich Phase 3d: Posed (links) vs. A-Pose (rechts)

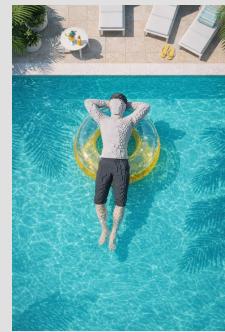
Das Ergebnis war eindeutig: Posed Input war A-Pose in fast allen Szenen überlegen. GPT-Image-1.5 produzierte die atmosphärischsten Ergebnisse, halluzinierte aber häufiger Details. NanoBanana Pro war konsistenter, aber weniger ausdrucksstark.

Tabelle 7: Phase 3d: Posed vs. A-Pose Vergleich (detailliert)

Ordner	Modell	Input	Format	Szene	Bew.	Halluz.
21	GPT-1.5	posed	portrait	night_drive	ja	ja
21	GPT-1.5	posed	portrait	staircase	ja	nein
21	GPT-1.5	posed	portrait	pool	ja	nein
21	GPT-1.5	posed	landscape	night_drive	ja	ja
21	GPT-1.5	posed	landscape	staircase	ja	nein
21	GPT-1.5	posed	landscape	pool	neutral	ja
22	NanoBanana	posed	portrait	night_drive	ja	nein
22	NanoBanana	posed	portrait	staircase	nein	nein
22	NanoBanana	posed	portrait	pool	ja	nein
22	NanoBanana	posed	landscape	night_drive	ja	ja
22	NanoBanana	posed	landscape	staircase	nein	ja
22	NanoBanana	posed	landscape	pool	ja	nein
23	GPT-1.5	a-pose	portrait	night_drive	nein	ja
23	GPT-1.5	a-pose	portrait	staircase	nein	ja
23	GPT-1.5	a-pose	portrait	pool	neutral	ja
23	GPT-1.5	a-pose	landscape	night_drive	ja	ja
23	GPT-1.5	a-pose	landscape	staircase	nein	nein
23	GPT-1.5	a-pose	landscape	pool	ja	ja
24	NanoBanana	a-pose	portrait	night_drive	ja	nein
24	NanoBanana	a-pose	portrait	staircase	nein	ja
24	NanoBanana	a-pose	portrait	pool	ja	nein
24	NanoBanana	a-pose	landscape	night_drive	neutral	ja
24	NanoBanana	a-pose	landscape	staircase	nein	ja
24	NanoBanana	a-pose	landscape	pool	neutral	ja



GPT-1.5 Posed Portrait (21)



GPT-1.5 A-Pose Portrait (23)



NanoBanana Posed Portrait (22)



NanoBanana A-Pose Portrait (24)

Abbildung 13: Phase 3d: Portrait-Vergleich – Posed (21/22) vs. A-Pose (23/24). [Bild-Galerie](#)



NanoBanana Posed Landscape (22)

NanoBanana A-Pose Landscape (24)

Abbildung 14: Phase 3d: Landscape-Vergleich – Posed (21/22) vs. A-Pose (23/24). [Bild-Galerie](#)

Gesamtbewertung Die folgende Tabelle fasst alle 52 Einzelbewertungen aus der Pipeline zusammen:

Tabelle 8: Gesamtbewertung: Alle Evaluationen der fal.ai-Pipeline

Nr.	Modell	Input	Format	Bew.	Halluz.	Kommentar
01	MiniMax	a-pose	sq.	nein	ja	Minecraft-Stil, Gesichtszüge
02	Z-Image	a-pose	sq.	bedingt	nein	A-Pose erhalten, clean
03	FLUX Depth	a-pose	sq.	nein	-	Körper humanoid geglättet
04	FLUX Canny	a-pose	sq.	nein	ja	Volles Gesicht halluziniert
05	Era3D	a-pose	sq.	nein	-	Nutzlos, eigene Renders besser
06	FLUX Kontext	a-pose	sq.	bedingt	-	Wird besser
07	GPT-1.5	a-pose	sq.	bedingt	teilweise	Solide, leichte Halluz.
08	SeedDream	a-pose	sq.	nein	ja	Volles Gesicht halluziniert
09	NanoBanana	a-pose	sq.	gut	teilweise	Beste Phase-2-Ergebnisse
10	FLUX Kont.	posed	sq.	nein	ja	Gesichter halluziniert
11	GPT-1.5	posed	sq.	ja	teilweise	Solideste Ergebnisse
12	NanoBanana	posed	sq.	ja	nein	Stärkste Voxel-Ästhetik
13	GPT-1.5	posed	portrait (sq.)	gemischt	teilweise	1:1 Ratio-Bug
13	GPT-1.5	posed	landsc. (sq.)	gemischt	teilweise	Auch 1:1
14	NanoBanana	posed	portrait (sq.)	gemischt	ja	Halb cool, halb Halluz.
14	NanoBanana	posed	landsc. (sq.)	gemischt	nein	Ohne Halluz. stark
15	GPT-1.5	posed	portrait	gemischt	teilweise	Gestreckte Referenz
15	GPT-1.5	posed	landsc.	gemischt	ja	Stretching + Cap fehlt
16	NanoBanana	posed	portrait	gemischt	nein	Stretching, keine Halluz.
16	NanoBanana	posed	landsc.	gemischt	nein	Weniger Stretching
17	GPT-1.5	posed	portrait (sq.)	gemischt	ja	1:1 Bug + Halluz.
17	GPT-1.5	posed	landsc. (sq.)	gemischt	ja	Dasselbe
18	NanoBanana	posed	portrait	gemischt	nein	-
18	NanoBanana	posed	landsc.	stark	nein	-
19	GPT-1.5	posed	portrait	gemischt	teilweise	-
19	GPT-1.5	posed	landsc.	gemischt	teilweise	-
20	NanoBanana	posed	portrait	gemischt	nein	-
20	NanoBanana	posed	landsc.	stark	nein	-
21	GPT-1.5	posed	portrait	ja	ja	ND: Scheibe fehlt
21	GPT-1.5	posed	portrait	ja	nein	SS
21	GPT-1.5	posed	portrait	ja	nein	PF
21	GPT-1.5	posed	landsc.	ja	ja	ND: Teil d. Autos fehlt
21	GPT-1.5	posed	landsc.	ja	nein	SS
21	GPT-1.5	posed	landsc.	neutral	ja	PF: Cap/Haare fehlen
22	NanoBanana	posed	portrait	ja	nein	ND
22	NanoBanana	posed	portrait	nein	nein	SS: Stufen nicht schön
22	NanoBanana	posed	portrait	ja	nein	PF: Sehr cool
22	NanoBanana	posed	landsc.	ja	ja	ND: Cockpit falsch
22	NanoBanana	posed	landsc.	nein	ja	SS: Doppelte Treppe
22	NanoBanana	posed	landsc.	ja	nein	PF: Transparent, gefällt
23	GPT-1.5	a-pose	portrait	nein	ja	ND: Scheinwerfer falsch
23	GPT-1.5	a-pose	portrait	nein	ja	SS: A-Pose overfitted
23	GPT-1.5	a-pose	portrait	neutral	ja	PF: Cap fehlt
23	GPT-1.5	a-pose	landsc.	ja	ja	ND: Cap fehlt
23	GPT-1.5	a-pose	landsc.	nein	nein	SS: Pose zu starr
23	GPT-1.5	a-pose	landsc.	ja	ja	PF: Pool Floaty falsch
24	NanoBanana	a-pose	portrait	ja	nein	ND
24	NanoBanana	a-pose	portrait	nein	ja	SS: Treppe nicht schön
24	NanoBanana	a-pose	portrait	ja	nein	PF: Super
24	NanoBanana	a-pose	landsc.	neutral	ja	ND: Halluz. stören
24	NanoBanana	a-pose	landsc.	nein	ja	SS: Treppen nicht schön
24	NanoBanana	a-pose	landsc.	neutral	ja	PF: Pool Floaty falsch

Zusammenfassung und theoretische Einordnung Theoretisch markiert dieses Experiment eine Verschiebung auf Deterdings (2017) Spektrum: Meine Rolle war nicht mehr die der Experimentierenden, die direkt mit einem Tool interagiert, sondern die einer Orchestratorin – ich definierte das "Was" und "Warum", während Claude Code das "Wie" implementierte. Schöns (1983) Reflection-in-Action fand im Zeitraffer statt: Wo ich in Experiment 1 stundenlang an ComfyUI-Parametern drehte, durchlief die Pipeline denselben explorativen Prozess in Minuten. Galanters (2003) Konzept des systematischen Mappings eines generativen Raums wurde hier buchstäblich umgesetzt – sieben Phasen, die den Möglichkeitsraum zwischen neun Modellen, zwei Input-Typen und drei Szenen systematisch absteckten.

Das Fazit war ambivalent. Einerseits gelang hier zum ersten Mal eine visuell überzeugende Kepler-Generierung ohne Fine-Tuning – die besten Bilder aus den Posed-Phasen waren die ersten, die ich mir als Album-Cover vorstellen konnte. Andererseits blieben die Halluzinationen: fehlende Caps, falsche Cockpit-Details, gelegentlich angedeutete Gesichtszüge unter der Maske. Die Pipeline machte den Prozess skalierbar, aber nicht kontrollierbar. Und eine neue Frage drängte sich auf: Wenn ich nicht mehr selbst prompte, sondern einen KI-Agenten beauftrage, der für mich promptet – wo endet dann meine Autorschaft?

Experiment 6: Video-Generierung (Fol.ai API + Claude Code)

Die automatisierte Bildpipeline aus Experiment 5 warf eine naheliegende Frage auf: Wenn statische Bilder von Kepler überzeugend gelingen – lässt sich derselbe Ansatz auf Bewegtbild übertragen? Image-to-Video-Modelle versprechen, aus Einzelbildern kurze Clips zu generieren. Doch wo Bildgenerierung mit Halluzinationen kämpft, potenziert Video die Herausforderung: Jedes halluzinierte Detail muss über mehrere Sekunden konsistent bleiben.

Setup und Methodik Sieben Modelle wurden systematisch verglichen: PixVerse v5.5, LTX-2, Wan 2.6, Seedance 1.0 Pro, Veo 3.1, Kling O3 und Kling 3.0 V3. Die Kosten reichten von \$0.20 (PixVerse) bis \$0.84 (Kling) pro Video. Als Testbilder dienten vier Szenen aus den besten Ergebnissen der Bildpipeline – zwei bewusst unterschiedliche Herausforderungen:

- **Studio** (Landscape, 1376×768): Kepler am Mischpult – kontrollierte Innenszene, wenig Bewegung
- **Pool** (Portrait, 768×1376): Kepler auf dem Pool-Ring, Vogelperspektive – Wasser, ungewöhnlicher Blickwinkel
- **Night Drive** (Portrait, 768×1376): Kepler im Retro-Auto auf Wüstenhighway – Nachtszene, Fahrtbewegung
- **Spiral Staircase** (Portrait, 768×1376): Kepler auf kosmischer Wendeltreppe – Prompt forderte blaues Leuchten statt Marmor

Die Prompts waren bewusst einfach gehalten: kurze Bewegungsbeschreibung plus "Camera holds still." Die vier verwendeten Video-Prompts:

Studio: *"The voxel character nods gently to the beat, fingers resting on the mixing console. Monitor screens glow steadily. Camera holds still."*

Pool: *"The voxel character drifts slowly on the pool ring. Water ripples gently around the float. Camera holds still from above."*

Night Drive: *"The voxel character drives steadily down the desert highway. Headlights illuminate the road"*

ahead. Camera holds still from behind."

Spiral Staircase: *"The white marble staircase begins to glow with a deep blue light. The voxel character slowly ascends the steps. Camera holds still."*

Jedes Modell erhielt identische Inputs. Insgesamt entstanden 28 Videos (7 Modelle × 4 Szenen) bei geschätzten Gesamtkosten von ca. \$16.



PixVerse v5.5



LTX-2



Wan 2.6



Seedance Pro



Veo 3.1



Kling O3



Kling 3.0 V3

Abbildung 15: Mittlere Frames, Studio-Szene (Landscape, 7 Modelle). Poster-Frames verlinken auf die Video-Galerie.

Modell	Bew.	Halluz.	Kommentar
PixVerse v5.5	okay	leicht	Kamerabewegung gefällt, aber Bewegung unpassend. Schwarzer Kasten entsteht.
LTX-2	gut	leicht	Schwer zu beurteilen ob halluziniert (Audio).
Wan 2.6	gut	nein	Bild wirkt leicht überzeichnet, Audio wurde trotzdem generiert.
Seedance Pro	okay	nein	Bewegt sich zwar nicht, aber okay.
Veo 3.1	schlecht	leicht	Slider bewegen sich. Bewegung gefällt nicht.
Kling O3	gut	leicht	Audio Meter unscharf. Bewegung gefällt sehr.
Kling 3.0 V3	gut	leicht	Audio Meter schwammig. Bewegung ruckartig aber cool.



Veo 3.1

Kling O3

Kling 3.0 V3



LTX-2 (Landscape-Output)

Abbildung 16: Mittlere Frames, Pool-Szene (Portrait, 7 Modelle). LTX-2 ändert das Seitenverhältnis zu Landscape. Poster-Frames verlinken auf die [Video-Galerie](#).

Modell	Bew.	Halluz.	Kommentar
PixVerse v5.5	schlecht	stark	Voxel-Gesicht verändert sich. Bewegung zu schnell.
LTX-2	schlecht	stark	Aspect Ratio geändert, Kamera bewegt sich zu sehr.
Wan 2.6	schlecht	stark	Gesicht entsteht mit der Zeit. Pool-Floaty-Bewegung gelungen.
Seedance Pro	schlecht	stark	Gesicht entsteht, Zehen zeichnen sich in Schuhen ab.
Vevo 3.1	gut	leicht	Voxel-Gesicht verändert sich leicht. Floaty-Bewegung cool.
Kling O3	gut	leicht	Ripples an falscher Stelle. Wassersimulation top.
Kling 3.0 V3	gut	nein	Das beste!



PixVerse v5.5



Wan 2.6



Seedance Pro



Veo 3.1



Kling O3



Kling 3.0 V3



LTX-2 (Landscape-Output)

Abbildung 17: Mittlere Frames, Night-Drive-Szene (Portrait, 7 Modelle). LTX-2 ändert das Seitenverhältnis zu Landscape. Poster-Frames verlinken auf die [Video-Galerie](#).

Modell	Bew.	Halluz.	Kommentar
PixVerse v5.5	schlecht	stark	Gesicht wird zu Lego-Figur. Autofahrt unnatürlich.
LTX-2	schlecht	stark	Auto fährt rückwärts, Licht hinten, Aspect Ratio geändert.
Wan 2.6	schlecht	leicht	Gesicht wird dazu generiert, Animation sonst gut.
Seedance Pro	gut	nein	Überraschenderweise gut.
Veo 3.1	okay	leicht	Funktioniert gut, fährt aber rückwärts – könnte man umkehren.
Kling O3	schlecht	stark	Okay, aber halluziniert am Rand und zoomt zu sehr heraus.
Kling 3.0 V3	okay	leicht	Etwas besser als Kling O3.



PixVerse v5.5



Wan 2.6



Seedance Pro



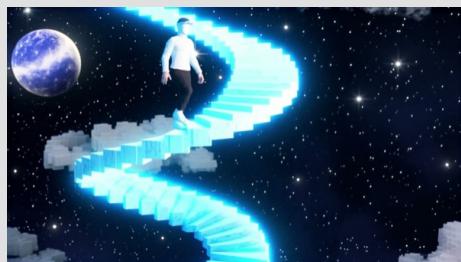
Veo 3.1



Kling O3



Kling 3.0 V3



LTX-2 (Landscape-Output)

Abbildung 18: Mittlere Frames, Spiral-Staircase-Szene (Portrait, 7 Modelle). LTX-2 ändert das Seitenverhältnis zu Landscape. Poster-Frames verlinken auf die [Video-Galerie](#).

Modell	Bew.	Halluz.	Kommentar
PixVerse v5.5	okay	leicht	Stufen etwas unnatürlich, aber interessanter Glow.
LTX-2	schlecht	stark	Er läuft fast von der Stufe, Aspect Ratio ändert sich.
Wan 2.6	okay	leicht	Bewegung leicht unnatürlich.
Seedance Pro	okay	leicht	Kopf verändert sich. Treppenanimation und Glow in Ordnung.
Veo 3.1	okay	leicht	Er überspringt Stufen, Animation wirkt aber relativ okay.
Kling O3	gut	leicht	Eine Stufe funktioniert nicht, aber Bewegung okay. Leuchten gefällt.
Kling 3.0 V3	schlecht	leicht	Stufen funktionieren nicht, Leuchten komisch, Kamerabewegung unerwartet.

Ergebnisse

Modell	Studio	Pool	Night Dr.	Staircase	Kosten	Gut	Schl.
Kling O3	gut	gut	schlecht	gut	\$0.84	3	1
Kling 3.0 V3	gut	gut	okay	schlecht	\$0.84	2	1
Veо 3.1	schlecht	gut	okay	okay	\$0.80	1	1
Seedance Pro	okay	schlecht	gut	okay	\$0.62	1	1
Wan 2.6	gut	schlecht	schlecht	okay	\$0.50	1	2
PixVerse v5.5	okay	schlecht	schlecht	okay	\$0.20	0	2
LTX-2	gut	schlecht	schlecht	schlecht	\$0.36	1	3

Analyse und Reflexion Die Ergebnisse zeigen ein klares Muster: **Preis korreliert mit Qualität, aber nicht linear**. Kling O3 (\$0.84) war mit drei von vier guten Ergebnissen der zuverlässigste Kandidat. LTX-2 (\$0.36) scheiterte an drei von vier Szenen, primär durch ein technisches Problem – das Modell erzwingt ein 16:9-Seitenverhältnis und beschneidet Portrait-Inputs.

Das auffälligste Phänomen war die **Gesichts-Halluzination bei Portrait-Szenen**. Keplers bewusst gesichtsloser Voxel-Kopf wurde von mehreren Modellen (Wan, Seedance, PixVerse) mit realistischen Gesichtszügen versehen – ein Problem, das bei der Landscape-Studio-Szene nicht auftrat. Die Modelle scheinen auf menschliche Gesichter als Default trainiert zu sein und können die bewusste Abwesenheit eines Gesichts nicht beibehalten. Dieses Phänomen spiegelt die Erfahrung aus Experiment 1 wider: Auch ComfyUI halluzinierte Gesichter in Keplers Maske.

Überraschend war die **Inkonsistenz innerhalb einzelner Modelle**. Seedance Pro war bei Studio und Pool schlecht bis mittelmäßig, lieferte aber beim Night Drive das überzeugendste Ergebnis aller Modelle – ein Befund, der gegen simple Rankings spricht. Veо 3.1 scheiterte beim Studio, glänzte aber beim Pool. Diese Szenenabhängigkeit zeigt, dass kein Modell universal überlegen ist – die Wahl muss pro Szene getroffen werden.

Methodisch erweiterte dieses Experiment die Orchestrator-Rolle aus Experiment 5: Dieselbe Pipeline, dasselbe Interview-basierte Evaluationsformat, nun auf Video angewandt. Die Bewertungskriterien – „Gefällt mir“, „Halluzination“, „Kommentar“ – erwiesen sich als robust über Medientypen hinweg. Die zentrale Erkenntnis für die Praxis: Video-Generierung mit Kepler ist möglich, aber nur mit den teuersten Modellen zuverlässig. Der kreative Prozess bleibt ein iteratives Auswählen aus einem unvorhersehbaren Möglichkeitsraum – Galanters (2003) generativer Ansatz, nun in Bewegtbild übersetzt.

Zentrale Beobachtungen aus dem Prozess

Aus der vergleichenden Analyse der Journal-Einträge und der systematischen Reflexion nach dem autoethnografischen Modell (Ellis et al., 2011) lassen sich vier wiederkehrende Muster identifizieren, die beschreiben, wie KI meinen kreativen Prozess verändert hat. Sie waren nicht vorab geplant, sondern sind durch die Praxis selbst hervorgebracht worden.

1. Unvorhersehbarer Flow

Flow entstand dort, wo ich ihn nicht geplant hatte, und blieb aus, wo ich ihn erwartet hatte. Der Website-Build mit Cursor funktionierte überraschend produktiv – „Oh mein Gott, das war einfach

sehr krass! Das hat das Aufsetzen der Website einfach tausendfach schneller gemacht", notierte ich im Journal (Eintrag 4). Die KI übernahm repetitive Aufgaben, während ich die ästhetische Vision steuerte. Umgekehrt scheiterte die Bildgenerierung für Kepler trotz intensiver Bemühungen: ComfyUI halluzinierte Gesichter in Keplers Maske – das Modell konnte die Abwesenheit eines Gesichts nicht akzeptieren, ein konkreter Fall des Uncanny Valley (Mori et al., 2012). Suno produzierte zu kommerzielle Ergebnisse, weil das Modell auf Mainstream-Ästhetik trainiert ist. Die Erkenntnis: KI-gestützte Kreativität ist nicht planbar. Flow entstand dort, wo die Interaktion tatsächlich "mixed-initiative" (Deterding et al., 2017) war — beim Website-Build mit engem Feedback-Loop, in dem beide Seiten beitrugen. Er blieb aus, wo die Interaktion einseitig wurde — bei Suno beschrieb ich, die KI generierte, ohne echte Rückkopplung. Die produktivsten Momente entstanden in Domänen und Konstellationen, die ich vorher nicht als Kernbereich identifiziert hatte.

2. Speed vs. Control

KI beschleunigt und begrenzt den kreativen Prozess gleichzeitig. Exploration und Prototyping werden drastisch schneller – beim Website-Build, bei der Generierung von Marble-Welten, im KI-Dialog als Lernwerkzeug. Aber die Kontrolle über ästhetische Details geht verloren. „Prompt rein, Ergebnis in Sekunden. Aber genau diese Geschwindigkeit hat auch gezeigt, wie wenig Kontrolle ich über das Ergebnis hatte“, reflektierte ich nach dem Suno-Experiment (Eintrag 6). Das Suno-Experiment illustriert das: Das Modell generiert zu jedem Track eine detaillierte Beschreibung – „bitcrushed chiptune pads over a sparse, detuned synth choir“ – die exakt das trifft, was ich mir vorstellte. Der tatsächliche Sound ging in eine andere Richtung: zu poliert, zu kommerziell. Bodens (2004) Konzept der explorativen Kreativität hilft hier: Das Modell erkundet den konzeptuellen Raum von „sowjetisch-dystopischer Pixel-Musik“ auf der Beschreibungsebene adäquat, scheitert aber an der Übersetzung in klangliche Materie. Die Diskrepanz zwischen Beschreibung und Output ist ein konkretes Symptom des Speed-vs-Control-Tradeoffs: Generierung ist schnell, aber Feinsteuerung ist nicht gegeben. Die automatisierte Pipeline aus Experiment 5 illustriert beide Seiten: 24 evaluierbare Bilder entstanden in Minuten, aber Details wie Keplers Cap oder die Cockpit-Perspektive blieben unkontrollierbar.

3. Verschiebung der Autorschaft

Über die Experimente hinweg verschob sich meine Rolle ständig. Beim Referenzen-Sammeln war ich Kuratorin, ein rein menschlicher Akt. Bei ComfyUI war ich Lernende und Experimentierende, orientiert am Workflow einer anderen Person. Beim Website-Build war ich Dirigentin: „Ich fühlte mich wie ein Manager mit Gesamtvision, der einen Junior Developer anleitet. Irgendwann wurde es eher Kollaboration, wenn wir zusammen nicht weiterkamen“ (Eintrag 4). Bei Suno war ich enttäuschte Kuratorin, die nur auswählen, nicht gestalten konnte. Im KI-Dialog war ich die Fragende.

Die wechselnden Rollen lassen sich als Wanderung auf Deterdings (2017) Spektrum lesen — mal näher am „Mensch als Schöpfer“ (Referenzen sammeln), mal in der Mixed-Initiative-Zone (Website-Build), mal näher am „Computer als Schöpfer“ (Suno). Diese Rollenvielfalt ist nicht Unschärfe, sondern Qualität. Ellis, Adams und Bochner (2011) beschreiben autoethnografische Forschung als Methode, die persönliche Erfahrung mit kultureller Analyse verbindet. Die wechselnden Rollen – Kuratorin, Dirigentin, Kollaborateurin, Lernende – sind Ausdruck einer Autorschaft, die sich fundamental verschoben hat: nicht mehr alleinige Urheberschaft, nicht vollständige Delegation. Und: „Garbage in, garbage out“: Domänenwissen ist die Voraussetzung für produktive Kollaboration, nicht KI-Kompetenz. Wer nicht weiß, was gut ist, kann nicht kuratieren. In Experiment 5 kam eine weite-

re Rolle hinzu: die der Orchestratorin, die nicht mehr direkt promptet, sondern einen KI-Agenten beauftragt, der seinerseits Bildmodelle steuert – eine dreistufige Kette, die die Frage nach der Autorschaft weiter verkompliziert.

4. KI-Burnout und KI-Enthusiasmus

Die emotionale Dimension der KI-Arbeit erwies sich als eigenständige Prozessveränderung, die ich nicht erwartet hatte. Es gibt einen Zyklus: Begeisterung, wenn etwas unerwartet gut funktioniert – der Website-Build, die ersten Marble-Welten, ein Claude-Dialog, der mich wirklich weiterbringt. Und dann Frustration und Erschöpfung, wie nach dem ComfyUI-Experiment: „Ich wollte ja genau durch die Verwendung von KI neue Möglichkeiten schaffen, die intuitiv sind und Spaß machen. Aber es sind halt keine magischen Tools, so wie es sich anfühlt“ (Eintrag 2).

Dieses Pendeln ist nicht Nebensache, sondern reale Belastung im kreativen Prozess. Der ständige Zwang, neue Tools zu evaluieren, Workflows umzubauen, mit Limitierungen zu kämpfen, kostet kreative Energie, die eigentlich in die Kunst fließen sollte. Man muss sich aktiv davor schützen, in diesem Zyklus aufgerieben zu werden. Die Versprechen der Technologie, schneller, besser, alles möglich, stehen in Spannung zur Realität einer Praxis, die oft mühsamer ist als die analoge Alternative. Der systematische Pipeline-Ansatz in Experiment 5 wirkte dem entgegen: Sieben strukturierte Phasen mit klaren Evaluationskriterien reduzierten die Frustration gegenüber den Einzelversuchen erheblich.

Rückbezug auf die Forschungsfrage

Am Ausgangspunkt dieser Arbeit stand eine praktische Frage: Kann generative KI die visuelle Konsistenz liefern, die Kepler braucht? Die Antwort ist: nicht zuverlässig. Bildgenerierungsmodelle scheitern an Keplers Nischen-Ästhetik — sie halluzinieren Gesichter in seine Maske, drifteten in Mainstream-Ästhetiken ab, und die Kontrolle über Details bleibt begrenzt. Aber genau dieses Scheitern wurde produktiv: Es offenbarte, wie fundamental sich der kreative Prozess durch die Kollaboration mit KI verändert — und damit die zweite, tiefere Frage der Einleitung: Was passiert, wenn ich nicht mehr nur Werkzeuge benutze, sondern mit ihnen zusammenarbeite?

Die zentrale Forschungsfrage — wie verändert der Einsatz multimodaler KI-Systeme den kreativen Prozess bei der Entwicklung einer digitalen Künstleridentität? — lässt sich anhand der vier Beobachtungen und der fünf Subfragen beantworten, die die strukturierten Reflexionsinterviews geleitet haben:

Wie verändert sich der Workflow? KI-Tools verschieben den kreativen Prozess von der linearen Ausführung zur iterativen Navigation. Flow entsteht unvorhersehbar (Beobachtung 1), und die Geschwindigkeit der Generierung steht in permanenter Spannung zur fehlenden Kontrolle über Details (Beobachtung 2). Der Workflow wird nicht einfach schneller — er wird anders: explorativer, weniger planbar, abhängiger von der Fähigkeit, Ergebnisse zu kuratieren statt zu produzieren.

Wie verschiebt sich die Autorschaft? Von der alleinigen Urheberschaft zu einem Spektrum wechselnder Rollen — Kuratorin, Dirigentin, Lernende, Orchestratorin. Domänenwissen, nicht KI-Kompetenz, erweist sich als entscheidende Voraussetzung (Beobachtung 3). Wer nicht weiß, was gut ist, kann nicht kuratieren.

Welche Rolle spielt Iteration? Iteration ist nicht Optimierung, sondern Erkenntnismethode. Jeder gescheiterte Prompt, jede unerwartete Halluzination wurde zum Ausgangspunkt für reflektierte Anpassung — Schöns (1983) „Situation, die zurückspricht“ in digitaler Form. Die automatisierte

Pipeline in Experiment 5 zeigt, dass systematische Iteration den explorativen Raum erweitert, ohne die zentrale Spannung zwischen Kontrolle und Generierung aufzulösen.

Wie wird mit Scheitern umgegangen? Scheitern ist Dauerzustand und produktive Kraft zugleich. Der emotionale Zyklus zwischen Enthusiasmus und Burnout (Beobachtung 4) zeigt, dass der Umgang mit Scheitern nicht nur eine technische, sondern eine psychologische Herausforderung ist. ComfyUI scheiterte an Keplers Gesicht, Suno an seiner Klangwelt — aber beide Erfahrungen schärften das Verständnis dafür, wo die Grenzen aktueller Modelle liegen und wie man mit ihnen arbeiten kann.

Wie entwickelt sich die Ästhetik im Prozess? Die Ästhetik entsteht nicht trotz, sondern durch die Einschränkungen der KI. Keplers Nischen-Ästhetik ließ sich nicht direkt generieren, aber die Auseinandersetzung mit den Grenzen der Modelle — Halluzinationen, Mainstream-Drift, fehlende Feinsteuerung — schärfe das ästhetische Urteil und führte zur bewussten Integration von „Fehlern“ als Gestaltungsmittel. In Galanters (2003) Terminologie: Die Suche nach „effektiver Komplexität“ wurde zum ästhetischen Kompass.

Diese Veränderungen sind nicht linear und nicht planbar; sie entstehen im iterativen Prozess und kosten mehr Energie, als die Versprechen der Technologie vermuten lassen. Die folgende Diskussion ordnet diese Befunde theoretisch ein.

Diskussion und Interpretation

Zusammenfassung der Ergebnisse

Die vorliegende Arbeit ging der Frage nach, wie der Einsatz multimodaler KI-Systeme den kreativen Prozess bei der Entwicklung des virtuellen Alter Egos Kepler verändert. Die Practice-Led Research, dokumentiert über fünf Experimente mit unterschiedlichen KI-Tools, identifiziert vier zentrale Prozessveränderungen.

Erstens entsteht *Flow unvorhersehbar*: Der Website-Build gelang produktiv, die Bildgenerierung scheiterte an Keplers Nischen-Ästhetik; Phänomene wie Face Hallucination und Uncanny Valley (Mori et al., 2012) erwiesen sich dabei als konkrete Hindernisse. Zweitens zeigt sich ein durchgängiger *Speed-vs-Control-Tradeoff*: KI beschleunigt Exploration und Prototyping, aber die Kontrolle über ästhetische Details geht verloren; das Modell „versteht“ Nischen sprachlich, kann sie aber medial nicht umsetzen. Drittens hat sich *meine Rolle fundamental verschoben*: von der Ausführenden zur Kuratorin, Dirigentin, Lernenden, wobei Domänenwissen sich als entscheidende Voraussetzung erwies, nicht KI-Kompetenz. Viertens erweist sich der *emotionale Zyklus zwischen KI-Enthusiasmus und KI-Burnout* als eigenständige Prozessveränderung: Der ständige Zwang, neue Tools zu evaluieren und mit Limitierungen zu kämpfen, kostet kreative Energie.

Diese vier Beobachtungen sind keine isolierten Phänomene. Sie verbindet ein gemeinsames Prinzip: Die produktivsten Momente entstanden dort, wo Intention und KI-Eigenlogik in Spannung zueinander traten, nicht wo sie sich deckten. Experiment 5 belegt zudem die Skalierbarkeit des Ansatzes: Eine automatisierte Pipeline kann den explorativen Raum systematisch abstecken, ohne dass die zentrale Spannung zwischen Kontrolle und Generierung aufgelöst wird.

Theoretische Implikationen

Die Ergebnisse legen nahe, dass bestehende Kreativitätstheorien im Kontext der Mensch-KI-Kollaboration neu betrachtet werden sollten. Boden (2004) unterscheidet kombinatorische, explorative und transformitive Kreativität. Meine Praxis zeigt, dass generative KI primär als Katalysator für *explorative* Kreativität fungiert: Die Systeme eröffnen einen konzeptuellen Raum, den ich als Künstlerin navigiere, kuratiere und bewerte. Transformative Kreativität – das Sprengen der Regeln eines konzeptuellen Raums – ging hingegen durchgehend von meinen Entscheidungen aus, nicht von den Modellen. Die KI generiert Variationen; die Künstlerin entscheidet, welche Variation die Grenzen verschiebt. Der unvorhersehbare Flow, den ich beschreibe, ist letztlich eine Konsequenz dieser Asymmetrie: Exploration ist schnell, aber Transformation erfordert menschliches Urteil.

Galanter (2003) zeigt, dass generative Kunst als Methode verstanden werden muss — es geht um das *Wie* der Produktion, nicht um das *Was* des Ergebnisses. Diese Perspektive verschiebt die Bewertung: Die Frage ist nicht, ob KI-generierte Bilder "authentisch" sind, sondern wie sich der kreative Prozess verändert, wenn ein autonomes System beteiligt ist. Der Speed-vs-Control-Tradeoff lässt sich als Navigation auf Galanters Ordnung-Chaos-Skala beschreiben: KI-Tools beschleunigen die Exploration des konzeptuellen Raums, aber die Kontrolle über die Zone effektiver Komplexität — das ästhetisch Interessante — bleibt beim Menschen.

Borgdorff (2012) argumentiert, dass künstlerische Forschung Wissen freilegt, das sich nur durch die Praxis selbst zeigt. Das bestätigt sich hier konkret: Mein Vorwissen über Keplers Ästhetik – die dystopische Klangwelt, die voxelisierte Maskierung, die Spannung zwischen Lo-Fi und Hyperdigital – ließ sich als Prompt-Anweisung an KI-Systeme nur begrenzt übertragen. "Garbage in, garbage out" erwies sich als Grundregel: Ohne Domänenwissen produziert die Kollaboration generische Ergebnisse. Dieses Wissen zeigte sich erst *im* iterativen Prozess, nicht vor ihm.

Besonders aufschlussreich ist die Verbindung zu Schön (1983): Reflection-in-Action, das spontane Reagieren auf unerwartete Situationen in der Praxis, beschreibt präzise den Modus, in dem ich auf KI-Halluzinationen, unerwartete Outputs und produktive Fehler reagierte. Das Scheitern eines Prompts war kein Abbruchkriterium, sondern Auslöser für reflektierte Anpassung. Diese Beobachtung erweitert Schöns Konzept um eine neue Dimension: In der Mensch-KI-Kollaboration ist die "Situation, die zurückspricht" kein physisches Material, sondern ein probabilistisches System mit eigener Handlungslogik. Der emotionale Zyklus zwischen Enthusiasmus und Burnout lässt sich als Konsequenz dieser permanenten Reflection-in-Action lesen: Die Situation spricht ständig zurück, und das kostet Kraft.

Deterding et al. (2017) fragen, wie Mixed-Initiative Creative Interfaces "upskilling without deskilling" ermöglichen können. Diese Frage resoniert mit meiner vierten Beobachtung: Der KI-Burnout-Zyklus ist möglicherweise ein Symptom dafür, dass aktuelle KI-Tools die Mixed-Initiative-Balance noch nicht gefunden haben — sie sind entweder zu autonom (Suno: keine Eingriffsmöglichkeit) oder zu abhängig von Expertenwissen (ComfyUI: hohe technische Hürde). Die produktivsten Momente meiner Forschung entstanden dort, wo die Balance gelang: beim Website-Build, wo die KI repetitive Aufgaben übernahm und ich die ästhetische Vision steuerte.

Praktische Implikationen

Für Künstler*innen und Medienproduzent*innen, die generative KI in ihre Praxis integrieren wollen, ergeben sich aus dieser Arbeit konkrete Erkenntnisse. Die wichtigste: Domänenwissen ist keine optionale Zugabe, sondern Voraussetzung für ästhetisch relevante Ergebnisse. Wer ohne klare

ästhetische Vision promptet, erhält den kleinsten gemeinsamen Nenner, die Mainstream-Ästhetik, auf die Generalist-Modelle trainiert sind. Keplers Nischen-Ästhetik ließ sich nur dort realisieren, wo ich spezifisches Vokabular, präzise Referenzen und ein geschultes Urteilsvermögen einbrachte.

Darüber hinaus zeigt sich ein durchgängiger Speed-versus-Control-Trade-off: KI-Tools beschleunigen die Produktion erheblich, reduzieren aber gleichzeitig die Kontrolle über ästhetische Details. Für die Medienproduktion bedeutet das, dass KI-Werkzeuge sich besonders für Ideation, Prototyping und Exploration eignen, während die finale ästhetische Entscheidung bei der Künstlerin verbleiben muss. Die Rolle verschiebt sich von der Ausführenden zur Kuratorin, eine Kompetenz, die in der Ausbildung von Medienproduzent*innen künftig stärker adressiert werden sollte.

Stärken und Limitierungen

Stärken

Die methodische Kombination aus Practice-Led Research und Autoethnografie ermöglichte eine Dokumentationstiefe, die rein empirische Ansätze nicht erreichen. Durch die Veröffentlichung des gesamten Prozesses auf der Everything-Machine-Website stelle ich eine Transparenz her, die in künstlerischer Forschung selten ist: Jedes Experiment, jede Reflexion, jeder Fehlschlag ist nachvollziehbar. Die multimodale Perspektive – Bild, Audio, Code, Text – bildet zudem die Komplexität zeitgenössischer Medienproduktion ab, anstatt sie auf ein einzelnes Medium zu reduzieren.

Limitierungen

Als autoethnografische Einzelfallstudie ist die Generalisierbarkeit der Ergebnisse eingeschränkt. Was für meine spezifische Praxis mit Kepler gilt, muss für andere künstlerische Kontexte nicht zu treffen. Die subjektive Perspektive, die in der Artistic Research als Stärke gilt (Ellis et al., 2011), bleibt in anderen Forschungsparadigmen eine methodische Schwäche. Eine weitere Limitierung liegt im technologischen Wandel: Die Tools, mit denen ich gearbeitet habe, werden sich schneller weiterentwickeln als akademische Publikationszyklen erlauben. Die Ergebnisse beschreiben den Stand einer Technologie in Bewegung. Schließlich muss eine Meta-Ebene benannt werden: Der Schreibprozess dieser Arbeit selbst wurde durch KI-Assistenz unterstützt, eine Verschränkung, die ich als konsequente Fortsetzung meiner Methodik verstehе, die aber offengelegt werden muss.

Practice-Based Intention, Practice-Led Realität

Ein ehrlicher Rückblick muss eine Verschiebung benennen, die sich erst im Laufe der Arbeit zeigte. Zu Beginn war das Projekt als practice-based intendiert: Die Artefakte – Keplers generierte Bilder, die Website, die Musik – sollten selbst Teil des Forschungsergebnisses sein. Im Verlauf hat sich die Arbeit aber als practice-led erwiesen. Der eigentliche Beitrag sind die vier Beobachtungen über Prozessveränderungen, nicht die Artefakte. Die generierten Bilder und Clips sind Lernartefakte, die den Weg dokumentieren, nicht „Kunst“ im Sinne eines eigenständigen Werks.

Paradoxalement war die Website das Artefakt mit dem meisten Flow – obwohl als Dokumentations-tool konzipiert, wurde sie zum vielleicht künstlerischsten Element des Projekts. Die KI-generierten Bilder hingegen fühlen sich wie Zwischenergebnisse an: ästhetisch teils interessant, aber nicht „meine Kunst“. Die Unsicherheit, ob das hier Kunst oder Technikerprobung ist, ist selbst eine Erkenntnis – sie spiegelt das Spannungsfeld, in dem sich viele Kreative befinden, die zum ersten Mal mit generativer KI arbeiten.

Nebenbei entstand ein wiederverwendbarer Workflow: Die Kombination aus Claude Code als Orchestrator und der fal.ai API als Modell-Backend erwies sich als reproduzierbare Methode für systematisches Modell-Screening. Dieser methodische Beitrag war nicht geplant, entspricht aber dem, was Frayling (1993) als "research for art" beschreibt – Forschung, die Werkzeuge und Methoden für die künstlerische Praxis bereitstellt.

Ausblick

Die Arbeit wirft Fragen auf, die über ihren Rahmen hinausgehen. Wie verändert sich die Kollaboration, wenn KI-Modelle präziser und kontrollierbarer werden? Wird der Speed-vs-Control-Tradeoff sich auflösen, wenn Modelle Nischen-Ästhetiken besser umsetzen können? Und vor allem: Wie nachhaltig ist kreatives Arbeiten mit KI, wenn der Burnout-Zyklus anhält, wenn Tools sich schneller ändern als man sie lernen kann?

Für Kepler als Projekt sehe ich die Arbeit als Fundament, nicht als Abschluss. Das Alter Ego existiert nun in einem dokumentierten Prozessraum, der weiter bespielt werden kann, mit zukünftigen Tools, in anderen Medien, mit neuen Kollaborationsmodellen. Die Forschungsfrage bleibt dabei offen, weil die Technologie, auf die sie sich bezieht, selbst in Bewegung ist. Artistic Research, die sich mit generativer KI beschäftigt, muss diesen Zustand der Vorläufigkeit nicht als Schwäche begreifen, sondern als methodische Grundbedingung akzeptieren.

Literaturverzeichnis

- Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd Aufl., S. 344). Routledge.
- Borgdorff, H. (2012). *The Conflict of the Faculties: Perspectives on Artistic Research and Academia*. Leiden University Press. <https://doi.org/10.5117/9789400600997>
- Candy, L. (2006). *Practice Based Research: A Guide* (CCS Report 1.0; S. 1–19). Creativity & Cognition Studios, University of Technology Sydney. <https://www.creativityandcognition.com/resources/PBR%20Guide-1.1-2006.pdf>
- Coleridge, S. T. (1984). *Biographia Literaria* (J. Engell & W. J. Bate, Hrsg.; Bd. 7). Princeton University Press. (Ursprünglich erschienen 1817)
- Deterding, S., Hook, J., Fiebrink, R., Gillies, M., Gow, J., Akten, M., Smith, G., Liapis, A., & Compton, K. (2017). Mixed-Initiative Creative Interfaces. *CHI EA '17: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 628–635. <https://doi.org/10.1145/3027063.3027072>
- Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An Overview. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 12(1), Art. 10. <http://www.qualitative-research.net/index.php/fqs/article/view/1589>
- Franceschelli, G., & Musolesi, M. (2024). Creativity and Machine Learning: A Survey. *ACM Computing Surveys*, 56(4), 1–41. <https://doi.org/10.1145/3664595>
- Frayling, C. (1993). Research in Art and Design. *Royal College of Art Research Papers*, 1(1), 1–5.
- Galanter, P. (2003). What is Generative Art? Complexity Theory as a Context for Art Theory. *Generative Art 2003: Proceedings of the International Conference*.
- Hansen, S. M. (2023). A Model for Understanding the Evolving Role of Graphic Designers in the Era of Artificial Intelligence. *HUB – Journal of Research in Art, Design and Society*. <https://doi.org/10.22501/hub.2156238>
- Mazzone, M., & Elgammal, A. (2019). Art, Creativity, and the Potential of Artificial Intelligence. *Arts*,

8(1), 26. <https://doi.org/10.3390/arts8010026>

- Monin, M., & Sadokierski, Z. (2025). Prompting as Thinking-With: Using Generative AI to Visualise an Extinct Dwarf Emu. *ArteActa*, 13(25). <https://doi.org/10.22501/artact.3790000>
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action* (S. 374). Basic Books.
- Suh, M., Youngblom, E., Terry, M., & Cai, C. J. (2021). AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–11. <https://doi.org/10.1145/3411764.3445219>
- Torres Núñez del Prado, P. (2025). Encoding Culture. *PARSE Journal*. <https://parsejournal.com/article/encoding-culture/>

Anhang

Anhang A: Verzeichnis der Journal-Einträge

Die folgende Tabelle gibt einen Überblick über die im Rahmen dieser Arbeit erstellten Journal-Einträge. Jeder Eintrag dokumentiert ein Experiment oder eine Arbeitsphase und wurde im Anschluss einer strukturierten KI-gestützten Reflexion unterzogen (siehe Anhang B).

Nr.	Datum	Titel	Tools	Modalität
1	15. November 2025	Referenzen für Kepler sammeln	Pinterest, Google Drive	—
2	16. November 2025	ComfyUI Consistent Character Workflow	ComfyUI, Flux, IP-Adapter	Bild
3	18. November 2025	Marble von Worldlabs	Marble, Claude	3D
4	19. November 2025	Website mit Cursor aufsetzen	Cursor, Claude	Code
5	11. Januar 2026	KI wirklich verstehen – Teil 1	Claude (Voice)	Dialog
6	12. Januar 2026	Suno-Experiment	Suno AI	Audio
7	25. Januar 2026	Kepler-Recherche, zweiter Versuch	Claude, Semantic Scholar, GitHub	Recherche

Die vollständigen Journal-Einträge mit Abbildungen sind auf der Projektwebsite einsehbar: <https://jnfmeier.github.io/everything-machine/>

Anhang B: Journal-Einträge und Reflexionsinterviews

Die folgenden Texte und Transkripte dokumentieren die Journal-Einträge und die strukturierten Reflexionen, die nach jedem Eintrag im Dialog mit Claude (Anthropic) entstanden sind. Eingebettete Audio- und Videomedien wurden für die Druckversion entfernt; die vollständigen Einträge finden sich auf der Projektwebsite.

Journal-Eintrag 1: Referenzen für Kepler sammeln (18. November 2023)

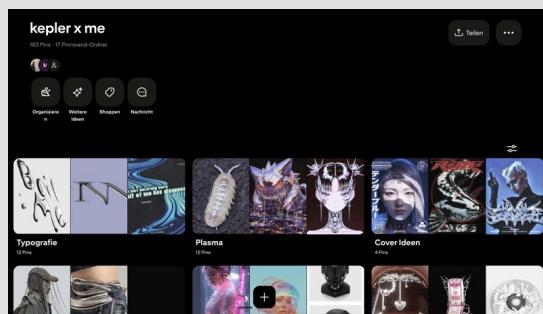
Der Start Im Rahmen der Artistic Research-Vorlesung bei Prof. Dr. Aristotelis Hadjakos an der TH OWL bin ich zum ersten Mal mit dem Konzept der Artistic Research in Berührung gekommen. Borgdorffs Idee – Forschung *in* und *durch* die Kunst, bei der die künstlerische Praxis selbst die Forschungsmethode ist – klang faszinierend, aber auch einschüchternd. Was ist AR überhaupt? Ist das, was ich mache, Kunst? Ich war unsicher, aber das Interesse war da.

Kepler ist Gavins Musikprojekt – eine digitale Künstleridentität, ein Charakter mit Voxel-Maske, der zwischen Mensch und Avatar existiert. KI-Tools interessieren mich gerade, und die Idee lag nahe: Was passiert, wenn ich generative KI-Systeme nutze, um Kepler weiterzuentwickeln? Wie verändert das den kreativen Prozess?

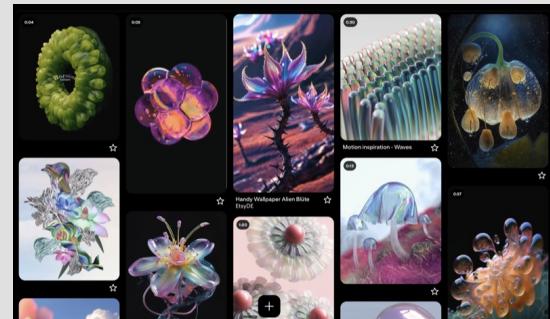
Hadjakos hat betont, dass Artistic Research frei sein soll – der Versuch, sich von starren Strukturen zu lösen und erstmal zu machen. Also haben wir heute angefangen: Referenzen sammeln, bevor ich mit der tatsächlichen Bildgenerierung beginne. Ich wollte verstehen, welche visuellen Stile und Darstellungen Gavin gefallen und gut funktionieren könnten.

Pinterest-Recherche Ich habe zusammen mit Gavin Pinterest durchforstet und verschiedene Boards angelegt – von Cover-Inspiration über Kleidung bis hin zu Vibes für spezifische Songs. Es ist erstaunlich, wie viele coole “Ästhetiken” sich auf Pinterest wiederfinden. Manche Looks sind hyperrealistisch, andere eher stilisiert. Wir haben versucht, eine gute Mischung zu finden, die zu der Vision von Kepler passt. Dabei sind uns bestimmte wiederkehrende Muster aufgefallen: dunkel, strukturiert, maskenhaft – eine visuelle Sprache, die schon durch die Auswahl entsteht, nicht erst durch Generierung.

Mein Pinterest-Board: Kepler x Me



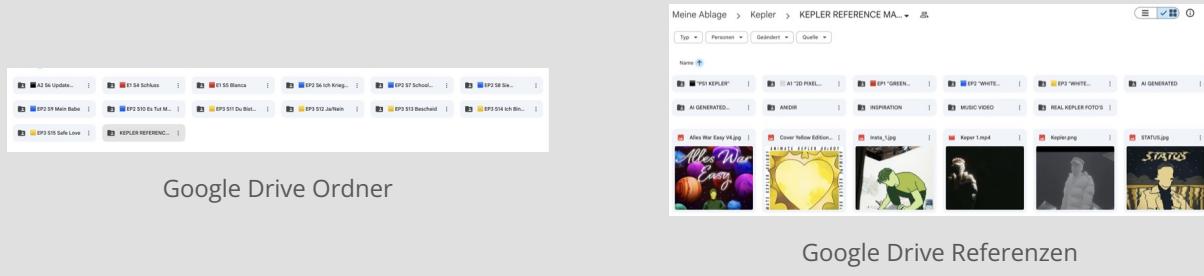
Pinterest Board Übersicht



Pinterest Song Bescheid

Google Ordner Parallel dazu habe ich einen Google Drive-Ordner eingerichtet, um Daten mit Gavin zu teilen. Wir beide arbeiten an Kepler-bezogenen Projekten, und es macht Sinn, unsere Ressourcen zu bündeln. Ich habe die gesammelten Referenzbilder, bereits erstellte Cover und 3D-

Daten hochgeladen. Gavin hat ebenfalls Material beigesteuert – Fotos, existierende Medien die von Anderen für Kepler erstellt worden sind (Cover, Artwork) und Konzeptideen, die ich noch nicht hatte. Der Ordner wurde zum zentralen Archiv für alles, was es zu Kepler bisher gibt: was wir selbst gemacht haben, was andere beigetragen haben, und was wir als Referenz nutzen wollen.



Reflexionsinterview Kontext

Claude Welche Tools hast du in dieser Session verwendet?

Jennifer Pinterest, Google Drive und keine KI-Tools so far.

Claude Was wolltest du erreichen?

Jennifer Alle drei (Referenzen sammeln, Zusammenarbeit starten, Inspiration finden).

Die 8 Forschungsfragen

Claude Hat sich dein Arbeitsablauf verändert? Was war neu, was ist weggefallen? (*Workflow*)

Jennifer Es war ganz wie gewohnt, wir haben einfach angefangen zu kollaborieren, in diesem Fall digital, und unsere Ressourcen zu sammeln.

Claude Wie hast du deine Rolle in diesem Prozess erlebt? (*Autorschaft*)

Jennifer Kollaboration mit Gavin. Wir haben parallel gearbeitet, jeder an seinem Rechner. Ich habe die Struktur vorgegeben, also Ordner und Boards, er hat Material beigesteuert. Keine KI involviert, rein menschliche Kuration und Abstimmung.

Claude Wie hat die schnelle KI-Ausgabe deinen Prozess beeinflusst? (*Iteration*)

Jennifer Keine KI verwendet, daher keine KI-Iteration. Aber der Prozess selbst war iterativ: Wir haben gesammelt, sortiert, verworfen, neu gesucht. Der Unterschied: Es war langsamer, aber alles war bewusst gewählt.

Claude Was ist schiefgelaufen und was hast du daraus gelernt? (*Scheitern*)

Jennifer Es war relativ reibungslos auf meiner Seite, aber Gavin wurde etwas ungeduldig, da ich aus Erfahrung mehr Struktur wollte (es fehlen immer noch ein paar Sachen, die er hochladen wollte, naja...).

Claude Hat sich durch die Arbeit eine eigene visuelle/stilistische Sprache entwickelt? (*Ästhetik*)

Jennifer Noch nicht direkt, das war Vorarbeit. Aber beim Sammeln sind mir bestimmte wiederkehrende Ästhetiken aufgefallen, die zu Kepler passen: dunkel, strukturiert, maskenhaft. Die visuelle Sprache entsteht hier durch Kuration, nicht durch Generierung.

Synthese

Claude Du hast erwähnt, dass du dir Gedanken über Urheberrechte machst – wenn man fremde Bilder als Stilreferenz nutzt. Hat das Sammeln der Referenzen dein Verhältnis zu 'originalem' versus 'inspiriertem' Schaffen verändert?

Jennifer Das Sammeln von Referenzen hat mir gezeigt, dass der kreative Prozess schon beim Kuratieren beginnt, nicht erst beim Generieren. Ob die Bilder, die mich inspirieren, selbst KI-generiert sind, spielt für die Inspiration keine Rolle. Aber für die Weiterverwendung als Trainingsdaten schon.

Journal-Eintrag 2: ComfyUI Consistent Character Workflow (16. November 2023)

Heute habe ich mich zum ersten Mal richtig mit dem ComfyUI Consistent Character Workflow beschäftigt. Das Ziel war klar: Ich wollte herausfinden, ob ich damit konsistente Darstellungen von Kepler über verschiedene Bilder hinweg erzeugen kann, ohne jedes Mal von vorne anfangen zu müssen.

Zuerst habe ich versucht, bestehende Bilder als Referenzen zu verwenden. Das hat teilweise funktioniert, aber es gab einige Probleme, die schnell klar wurden. Wenn ich zum Beispiel ein Bild hochgeladen habe, auf dem Kepler einen Basketball hält, und dann in den Prompts eine Maske hinzugefügt habe, wurde die Maske zwar generiert, aber der Basketball blieb einfach drin. Das Modell hat offenbar nicht verstanden, dass bestimmte Elemente entfernt werden sollten, wenn neue hinzukommen. Ähnlich war es mit Gesichtern – wenn ich ein Bild mit einem bestimmten Gesichtsausdruck als Referenz genommen habe, wurden bei der Generierung manchmal zusätzliche Gesichter eingefügt, als Platzhalter für Ergebnisse, die das System noch nicht richtig interpretieren konnte.

Nach ein paar frustrierenden Versuchen habe ich beschlossen, eine eigene Referenz zu erstellen. Ich habe Kepler auf einem neutralen grauen Hintergrund fotografiert, ohne störende Elemente, ohne Basketball, ohne komplexe Hintergründe. Nur er, klar und deutlich. Dann habe ich die Prompts leicht angepasst, um zu sehen, wie sich das auf die Ergebnisse auswirkt.



Abbildung 19: Referenzbild von Kepler auf neutralem Hintergrund

Die Ergebnisse waren gemischt. Einige Bilder sahen wirklich gut aus – der Charakter war konsistent, die Details stimmten, und die Maske wurde korrekt dargestellt. Andere waren weniger überzeugend. Manchmal verlor der Workflow die Konsistenz komplett, besonders bei unterschiedlichen Posen oder wenn ich versucht habe, komplexere Szenen zu generieren. Es scheint, als ob das System am besten funktioniert, wenn die Referenzbilder möglichst klar und unkompliziert sind.



Generiertes Kepler-Bild 1



Generiertes Kepler-Bild 2



Generiertes Kepler-Bild 3



Generiertes Kepler-Bild 4



Generiertes Kepler-Bild 5

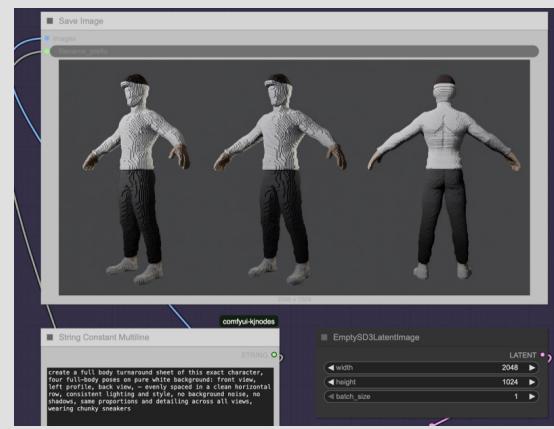


Generiertes Kepler-Bild 6

ComfyUI und Flux Der ComfyUI Consistent Character Workflow basiert auf mehreren Technologien, die zusammenarbeiten, um konsistente Charakterdarstellungen zu ermöglichen. ComfyUI selbst ist eine grafische Benutzeroberfläche für Stable Diffusion, die es ermöglicht, komplexe Bildgenerierungsprozesse durch modulare Workflows zu steuern. Statt Code zu schreiben, verbindet man Nodes miteinander, die verschiedene Funktionen repräsentieren – von der Bildladung über die Prompt-Verarbeitung bis hin zur finalen Ausgabe.



ComfyUI Workflow Screenshot 1



ComfyUI Workflow Screenshot 2

Der Consistent Character Workflow nutzt dabei hauptsächlich das Flux-Modell, ein von Black Forest Labs entwickeltes Bildgenerierungsmodell, das besonders gut darin ist, konsistente Darstellungen zu erzeugen. Flux funktioniert ähnlich wie Stable Diffusion, verwendet aber eine andere Architektur, die es ermöglicht, feinere Details und bessere Konsistenz über mehrere Bilder hinweg

zu erreichen.

IP-Adapter und ControlNet Ein wichtiger Teil des Workflows ist die Verwendung von IP-Adapter oder ähnlichen Technologien, die es ermöglichen, Referenzbilder in den Generierungsprozess einzubinden. Diese Adapter analysieren das Referenzbild und extrahieren bestimmte Merkmale – Gesichtszüge, Kleidung, Stil – und versuchen dann, diese Merkmale in neuen Generierungen beizubehalten. Das funktioniert durch sogenannte Attention-Mechanismen, bei denen das Modell lernt, welche Teile des Referenzbilds wichtig sind und welche ignoriert werden können.

ControlNet ist ein weiteres wichtiges Werkzeug in diesem Workflow. Es ermöglicht es, bestimmte Aspekte der Generierung zu kontrollieren – zum Beispiel die Pose eines Charakters oder die Tiefeinstuktur einer Szene. ControlNet funktioniert, indem es zusätzliche Informationen aus einem Kontrollbild extrahiert und diese Informationen dann in den Generierungsprozess einfließen lässt. Wenn ich zum Beispiel ein Bild mit einer bestimmten Pose habe, kann ControlNet diese Pose erkennen und sicherstellen, dass der generierte Charakter in derselben Pose erscheint.

Das Problem, das ich mit dem Basketball hatte, liegt wahrscheinlich daran, dass das System nicht wirklich „versteht“, was es generiert. Es erkennt Muster und Merkmale, aber es hat kein semantisches Verständnis davon, dass ein Basketball und eine Maske möglicherweise nicht gleichzeitig vorhanden sein sollten. Das Modell sieht einfach verschiedene visuelle Elemente und versucht, sie zu kombinieren, ohne die logischen Zusammenhänge zu berücksichtigen.

Face Detailer — und das Problem Face Detailer ist ein zusätzliches Tool, das speziell darauf ausgelegt ist, Gesichter zu verbessern. Es funktioniert, indem es das generierte Bild analysiert, Gesichter erkennt und dann diese Bereiche mit einem speziell trainierten Modell nachbearbeitet. Das kann helfen, Details zu schärfen und realistischere Gesichtszüge zu erzeugen, aber es kann auch zu Problemen führen, wenn das System Gesichter erkennt, wo keine sein sollten, oder wenn es versucht, Gesichter zu „reparieren“, die eigentlich korrekt waren.

Der gesamte Workflow ist im Grunde eine Kette von Transformationen. Zuerst wird das Referenzbild analysiert und seine Merkmale werden extrahiert. Dann wird ein Text-Prompt verarbeitet, der beschreibt, was generiert werden soll. Diese Informationen werden kombiniert und durch das Flux-Modell geleitet, das Schritt für Schritt ein neues Bild generiert. ControlNet kann dabei zusätzliche Kontrolle bieten, und Face Detailer kann am Ende noch einmal über die Gesichter gehen, um sie zu verbessern.

Das Problem ist, dass jeder dieser Schritte Fehler einführen kann. Wenn die Referenz nicht klar genug ist, werden die extrahierten Merkmale ungenau. Wenn der Prompt zu komplex ist, kann das Modell verschiedene Interpretationen haben. Und wenn ControlNet oder Face Detailer falsch interpretieren, was sie sehen sollen, entstehen Artefakte wie zusätzliche Gesichter oder nicht entfernte Objekte.

Die Lösung, die ich gefunden habe – eine klare Referenz auf neutralem Hintergrund – funktioniert, weil sie dem System weniger Raum für Fehlinterpretationen lässt. Wenn nur der Charakter zu sehen ist, ohne störende Elemente, kann das Modell sich auf die wichtigen Merkmale konzentrieren. Aber selbst dann ist die Konsistenz nicht garantiert, besonders wenn man versucht, den Charakter in verschiedenen Posen oder Kontexten zu zeigen.

Reflexionsinterview Kontext

Claude Welche Tools hast du in dieser Session verwendet?

Jennifer ComfyUI mit einem Network von Mickmumpitz (Füge einen Link zu seinem YT Channel hinzu).

Claude Was wolltest du erreichen?

Jennifer Ich wollte schauen, ob ich Album-Cover- oder Konzeptbilder erstellen konnte von Kepler. Mir war aber wichtig, dass eine Konsistenz besteht und es ästhetisch aussieht. Ich war auch etwas von dem ComfyUI Workflow und dem Konzept überfordert, deswegen wollte ich das als Plattform nutzen, um das Ganze zu lernen.

Die 8 Forschungsfragen

Claude Hat sich dein Arbeitsablauf verändert? (Workflow)

Jennifer Es ist ein komplett neuer Workflow mit ComfyUI zu arbeiten. Ich glaube, ich muss wiederkehren, um das Ganze besser zu verstehen, und vielleicht nicht mit Kepler als Charakter. Denn ich habe gleich gesehen, dass er besonders schwierig ist. Er ist jetzt kein Mensch per se, und sein Stil könnte schwierig zu replizieren sein, da er kein Gesicht hat, eine Maske trägt und in der Version, in der ich sie ihnen zeigen möchte, auch noch voxelized ist.

Claude Wie hast du deine Rolle erlebt? (Autorschaft)

Jennifer Ich war Lernende und experimentierte eher als ich das Tool wirklich verwendete. Ich hatte mich ja an dem Netzwerk einer anderen Person orientiert.

Claude Wie hat die schnelle KI-Ausgabe deinen Prozess beeinflusst? (Iteration)

Jennifer Es war gar nicht so schnell, da ich es tatsächlich lokal auf meinem M1-MacBook Pro Max 64 GB ausprobiert habe. Auch der Prozess, das Ganze zu installieren, hat länger gedauert und hat viele Gigabyte benötigt. Also war es etwas frustrierend.

Claude Was ist schiefgelaufen? (Scheitern)

Jennifer Es war sehr inkonsistent. Die Prompts haben nicht das erfüllt, was sie erfüllen sollten. Der Basketball wurde nicht entfernt, als ich ein Bild mit Basketball eingefügt habe. Besonders frustrierend war, dass Gesichter in Keplers Maske rein interpretiert worden sind. Auch als ich negative Prompts dafür eingefügt habe. Ich glaube, das wird ein größeres Problem.

Claude Hat sich eine eigene visuelle Sprache entwickelt? (Ästhetik)

Jennifer Nein, denn so weit bin ich gar nicht erst gekommen in diesem Anlauf. Denn es haben Sachen ja auch nicht funktioniert, und das, was herauskam, sah schon sehr generisch KI-mäßig aus.

Synthese

Claude Du hast erwähnt, dass das System Gesichter in Keplers Maske interpretiert hat – obwohl er bewusst kein Gesicht haben soll. Wie fühlt es sich an, wenn die KI deine künstlerische Intention so grundlegend 'missversteht'?

Jennifer Es ist einfach frustrierend. Am Ende des Tages ist es ein technisches Problem.

Ich wollte ja genau durch die Verwendung von KI neue Möglichkeiten schaffen, die intuitiv sind und Spaß machen und auch Zeit einsparen, um Dinge zu kreieren, ohne jetzt mir ganz die Arbeit abzunehmen. Aber es zeigt halt, dass wir noch nicht ganz da sind und die Art Directability noch nicht da ist. Und das würde ich mir eben wünschen. Aber es sind halt keine magischen Tools, so wie es sich anfühlt.

Für viele, wenn sie das erste Mal ChatGPT benutzen, sind es einfach nur Werkzeuge, die mehr oder weniger gut für den Anwendungsfall funktionieren. Es hätte zufällig gut funktionieren können. Es würde viel besser funktionieren, wenn es sich um eine realistische, oder Anime-style, oder Cartoon-style Figur handeln würde. Aber dem ist nicht so. Kepler hat eben kein sichtbares Gesicht, aber viele der Trainingsdaten enthalten Gesichter und deswegen bleiben sie hartnäckig dort.

Journal-Eintrag 3: Marble von Worldlabs (18. November 2029)

Heute habe ich Marble von Worldlabs getestet, ein neues Tool, das unter marble.worldlabs.ai verfügbar ist. Ich hatte davon gehört, dass es sich um ein Tool zur Erstellung von 3D-Welten handelt, und war neugierig, wie es sich von anderen Tools unterscheidet, die ich bisher verwendet habe.

Marble ist im Grunde eine browserbasierte Plattform, die es ermöglicht, interaktive 3D-Welten zu erstellen und zu erkunden. Das Besondere daran ist, dass man keine komplexe Software installieren muss – alles läuft im Browser, was den Einstieg deutlich erleichtert. Die Oberfläche ist relativ intuitiv gestaltet, auch wenn es am Anfang etwas dauert, bis man sich zurechtfindet.

Ich habe vier verschiedene Welten erstellt, um zu testen, wie das Tool mit unterschiedlichen Input-Strategien umgeht – von konkreten Fotos bis zu abstrakten Vibes.

Welt 1 Die erste Welt war meine erste Welt, die ich erstellt habe. Sie basiert auf einem Bild von meiner Freundin und mir in dem Badezimmer eines Soviet-inspirierten Cafés, mit alten Artefakten aus der Soviet-Zeit. Die Welt ist sehr creepy und verzerrt – man kann nur noch eine Person erkennen, die fast gruselige Proportionen hat. Das sehr interessante ist, dass das System das Bild, welches ein Spiegelbild ist, als Spiegel erkannt hat und quasi die gespiegelten Personen in den Raum gebracht hat. Es hat viele Artefakte übernommen, was sehr interessant ist.



Originalfoto: Soviet-inspiriertes Café

Badezimmer



Generierte 3D-Welt basierend auf dem Foto

Welt 2: Kepler in Spiegelscherben Für die zweite Welt habe ich ein Foto von Kepler verwendet – ohne zusätzliche Prompt-Beschreibung, nur das Bild als Input.



Originalfoto: Kepler



Generierte 3D-Welt: Kepler in Spiegelscherben

Das Ergebnis ist ein bisschen Uncanny Valley – Kepler steht da mit seltsamen Proportionen. Interessant ist, dass der Hintergrund einfach schwarz ist, wo das Modell nichts sehen kann. Das passt aber zur düsteren, fragmentierten Welt. In der Unschärfe als CG-Hintergrund könnte das funktionieren.

Welt 3: Unterwasser-Pool Hier habe ich ein Pinterest-Bild aus dem Board für den Song verwendet und eine kurze Prompt-Beschreibung hinzugefügt.

Prompt: "Underwater pool view with inflatable rings floating above. Summer Feelings Refractions."



Originalfoto: Pool



Generierte 3D-Welt: Unterwasser-Pool

Mir fällt auf, dass sich das Modell sehr nah am Input-Bild orientiert. Das ist nicht unbedingt schlecht, aber es limitiert die kreative Interpretation. Für abstraktere Ergebnisse brauche ich einen anderen Ansatz.

Welt 4: Botanischer Garten (Claude-generierte Prompt) Da ich als Free-User nur ein einziges Bild hochladen kann, habe ich einen Workaround versucht: Mehrere Pinterest-Bilder in den Claude-Chat hochgeladen und mir eine Marble-Prompt generieren lassen.



Claude-Input: Pinterest-Bilder 1



Claude-Input: Pinterest-Bilder 2



Claude-Input: Pinterest-Bilder 3

Claude-generierte Prompt: "Translucent glass botanical sculpture, iridescent resin figure with embedded flowers, ethereal bioluminescent flora, soft pastel gradients of lavender, mint, coral and gold, delicate organic forms suspended in crystalline transparency, surreal dreamlike atmosphere, intricate internal structures visible through glossy surfaces, dewdrops and light refractions, alien botanical specimens with thorny stems and luminous petals, macro photography aesthetic, soft bokeh backgrounds, hyperdetailed textures mixing smooth glass with organic growth, otherworldly garden specimens caught between organic and synthetic, studio lighting with rim light highlights, volumetric light rays, 8k cinematic quality"



Generierte 3D-Welt: Botanischer Garten 1



Generierte 3D-Welt: Botanischer Garten 2

Dieser Ansatz – Vibes statt konkretes Referenzbild – hat interessantere Ergebnisse gebracht. Die Welt fühlt sich weniger wie eine Rekonstruktion und mehr wie eine Interpretation an.

Fazit zu den Experimenten Marble orientiert sich stark am Input-Bild. Für kreativere Ergebnisse funktioniert es besser, nur Text-Prompts zu verwenden oder abstrakte Inputs. Der Workaround über Claude zur Prompt-Generierung aus mehreren Bildern ist eine interessante Methode, um die Einbild-Limitierung zu umgehen.

Die Technologie dahinter Marble ist das erste kommerzielle Produkt von [World Labs](#), einem Startup gegründet von der KI-Pionierin Fei-Fei Li (bekannt für ImageNet). Im Kern ist Marble ein sogenanntes "World Model" – ein KI-System, das aus Text, Bildern oder Videos persistente 3D-Umgebungen generiert.

Was Marble von anderen Tools unterscheidet: - **Persistente Welten:** Anders als Systeme wie Decart oder Odyssey, die Welten on-the-fly generieren, erstellt Marble herunterladbare 3D-Umgebungen - **Multimodaler Input:** Text, Bilder, Videos oder grobe 3D-Layouts - **Exportformate:** Gaussian Splats, Meshes oder Videos - **VR-ready:** Native Unterstützung für Vision Pro und Quest 3

Chisel Editor: Ein experimenteller 3D-Editor, der strukturierte Layouts (Wände, Boxen, Ebenen) mit Style-Prompts kombiniert – ähnlich wie HTML die Struktur und CSS den Stil einer Website definieren.

Timeline: - September 2025: Limited Beta - 12. November 2025: Public Release mit Freemium-Modell

Quellen: [World Labs Blog](#), [TechCrunch](#)

Reflexionsinterview [Kontext](#)

Claude Welche Tools hast du verwendet?

Jennifer Marble (marble.worldlabs.ai).

Claude Was wolltest du erreichen?

Jennifer Ich wollte zum einen das Tool testen, ich hatte es gerade gesehen und da die Gesichter ja nicht funktioniert haben, also Kepler als Charakter nicht funktioniert hat im Comp-V-Wire-Workflow, hatte mich interessiert, ob wir vielleicht Hintergrundwelten generieren könnten.

Die 8 Forschungsfragen

Claude Hat sich dein Arbeitsablauf verändert? (*Workflow*)

Jennifer Ich habe hier mit Marble erstmal experimentiert. Ich kannte bereits Gaussian Splats als Konzept und ich fand es interessant, wie schnell das jetzt ging, aus einem einzigen Bild einen übereinstimmenden Splat zu erzeugen.

Ich war aber jetzt nicht so überzeugt davon, dass ich es sicher verwenden werde.

Claude Wie hast du deine Rolle erlebt? (*Autorschaft*)

Jennifer Einfach nur jemand, der experimentiert und ein Produkt ausprobiert.

Claude Wie hat die KI-Ausgabe deinen Prozess beeinflusst? (*Iteration*)

Jennifer Das Ausprobieren machte Spaß, es ging schnell, und das könnte für die Zukunft interessant sein.

Aber mich interessieren dann eher die generativen Möglichkeiten. Also, wirklich, sich komplett Welten ausdenken anstatt existierende Welten 3D zu machen. Obwohl ich, wenn ich jetzt darüber nachdenke, mir gar nicht mehr so sicher bin. Es kann schon interessante Use Cases haben...

Claude Was ist schiefgelaufen? (*Scheitern*)

Jennifer Nicht unbedingt, außer dass das freiere, letzte Ergebnis nicht hundertprozentig überzeugend ist, aber schon mal interessant.

Und dass, als ich Personen drin hatte mit Spiegelungen und so, dass die Personen gruselig und Uncanny-Valley-mäßig verzerrt waren.

Aber ich hatte jetzt gar keine Ansprüche an das Tool und würde Input-Bilder ohne Personen oder komplexe Spiegelungen nehmen das nächste Mal.

Claude Hat sich eine eigene visuelle Sprache entwickelt? (*Ästhetik*)

Jennifer Noch nicht, also ich habe nur im Hinterkopf ein neues Tool, das ich verwenden könnte.

Synthese

Claude Im Eintrag steht, dass das System das Spiegelbild als Spiegel erkannt und die gespiegelten Personen in den 3D-Raum gebracht hat – faszinierend und gruselig zugleich. Könntest du dir vorstellen, solche 'Fehler' absichtlich als ästhetisches Mittel zu nutzen?

Jennifer Jetzt, wo du es sagst, ja, also man könnte tatsächlich wirklich damit visuell spielen. Ich weiß nur nicht, wie einfach es ist, das wirklich jetzt im 3D-Raum umzusetzen.

Klar, wenn ich ein reales Bild von ihm nehme und dann ein Gaussian-Splat erzeuge, könnte das interessant sein. Aber ich bin ja eher in der 3D/Computergrafik-Welt verankert und arbeite nicht viel mit Fotos. Aber es ist auf jeden Fall ein cooler Ansatz. Wobei, da fällt mir gerade ein, ich könnte auch 3D gerenderte Bilder in ein Gaussian Splat Generator eingeben.

Journal-Eintrag 4: Website mit Cursor aufsetzen (19. November 2023)

Für mein Artistic Research-Projekt "Everything Machine (Kepler)" brauchte ich eine Website, die sowohl den wissenschaftlichen Anspruch erfüllt als auch die künstlerische Identität von Kepler widerspiegelt. Statt mit einem fertigen CMS zu arbeiten, habe ich mich entschieden, die Seite von Grund auf mit **Cursor** zu bauen – einem AI-gestützten Code-Editor, der mir dabei hilft, schneller zu entwickeln und gleichzeitig den Code sauber zu halten.

Technical Setup: Cursor und GitHub Cursor ist im Grunde ein Fork von VS Code, der mit KI-Features erweitert wurde. Man kann direkt im Editor mit der KI chatten, Code generieren lassen oder bestehenden Code refactoren. Für mich war das perfekt, weil ich so nicht ständig zwischen Browser und Dokumentation wechseln musste – die KI kennt den Kontext meines Projekts und kann direkt helfen.

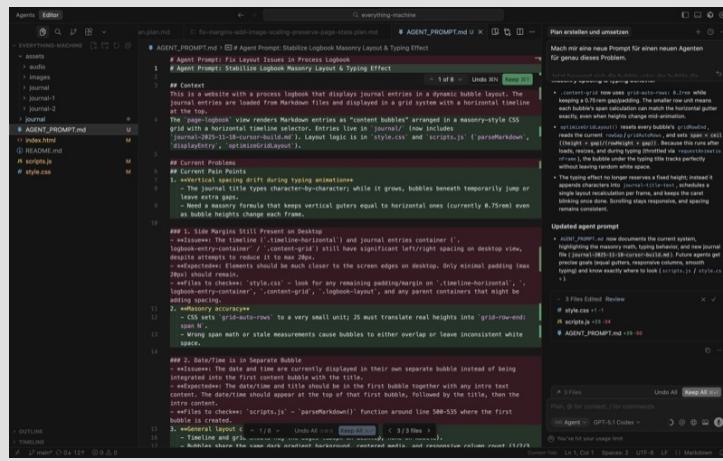


Abbildung 20: Cursor Oberfläche. Links die Liste der Dateien im GitHub Repository, in der Mitte eine Datei mit Changes, die die KI gemacht hat (die accepted werden können – in diesem Fall ein Prompt um das Layout zu fixen), rechts der Agent View, der ein Modell (in diesem Fall GPT 5.1 Codex) nutzt, um Prompts umzusetzen.

Das Projekt ist komplett in Git versioniert. Jede größere Änderung wurde als Commit festgehalten, was es mir ermöglicht, den Entwicklungsprozess nachzuvollziehen. Die Struktur ist simpel: `index.html` für die Hauptseite, `scripts.js` für die Logik, `style.css` für das Styling, und ein `journal/` Ordner für alle Markdown-Einträge. Keine komplexen Build-Tools, keine Dependencies außer Tailwind CSS über CDN – alles läuft direkt im Browser.

Design: Research Catalogue inspiriert, aber dunkel Als Inspiration diente mir die [Research Catalogue](#) – eine Plattform für Artistic Research, die ich während meines Studiums kennengelernt habe. Die hat diesen cleanen, wissenschaftlichen Look mit klarer Navigation und guter Lesbarkeit. Ich habe mir das Konzept geliehen, aber es für Kepler angepasst.

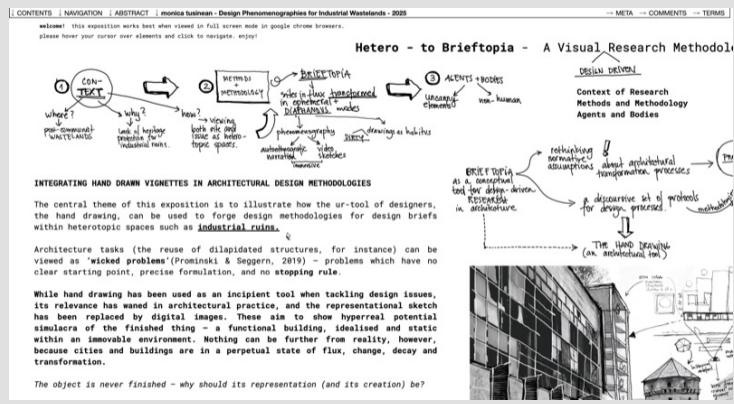


Abbildung 21: Research Catalogue Layout-Beispiel. Ein cooler Artikel auf researchcatalogue.net, der das Layout demonstriert – cleanes Design mit klarer Navigation und guter Lesbarkeit.

Ein besonders gutes Beispiel für das Research Catalogue Design ist dieser [Design Phenomenographies for Industrial Wastelands](#), der zeigt, wie multimediale Inhalte in einem wissenschaftlichen Kontext präsentiert werden können.

Statt hellem Hintergrund gibt es einen dunklen Theme (#111), der besser zu Keplers digitaler Identität passt. Die Navigation funktioniert ähnlich wie bei Research Catalogue – ein Dropdown-Menü oben links mit den Hauptsektionen. Links werden als “Pills” dargestellt (abgerundete Buttons), was moderner wirkt als normale Textlinks.

Das Besondere sind die Pixel-Headings. Kepler hat diese Voxel-Ästhetik, und um das visuell zu transportieren, verwende ich die “Press Start 2P” Font von Google Fonts für alle Überschriften. Das gibt der Seite sofort diesen retro-digitalen Vibe, ohne dass es zu verspielt wirkt. Die Hauptüberschriften im Journal haben sogar einen animierten Gradient-Effekt, der von grün zu weiß übergeht – eine Referenz an die Matrix-Ästhetik, die zu Keplers digitaler Existenz passt.

Musikplayer: Medien-Enhancement im Header Ein wichtiger Teil des Projekts ist die Musik von Kepler. Statt die Songs nur als statische Links einzubinden, habe ich einen Mini-Player in die Navigation integriert. Der bleibt immer sichtbar, auch wenn man durch die Seiten navigiert, und spielt die Songs aus einer Playlist ab.

Technisch funktioniert das so: In `scripts.js` gibt es ein `playlist` Array, in dem alle Songs mit Metadaten (Titel, Album, Cover, Spotify-Link) gespeichert sind. Der Player nutzt das native HTML5 `<audio>` Element und zeigt den aktuellen Track in einem Marquee-Text an. Über einen Playlist-Button kann man alle verfügbaren Songs sehen und direkt wechseln.

Die Implementierung war relativ simpel – keine externe Library nötig. Der Player lädt die Metadaten (Dauer) asynchron, damit die Seite schnell bleibt. Die Playlist wird dynamisch generiert, sodass ich einfach neue Songs zum Array hinzufügen kann, ohne HTML ändern zu müssen.

Titelseite: Abstract als Basis Die erste Seite (“Projektübersicht”) basiert direkt auf meinem Abstract. Der Text steht prominent oben in einem abgesetzten Container, gefolgt von Einleitung, Forschungsfragen und Methodologie. Das Layout ist zentriert und auf Lesbarkeit optimiert – max-width von 4xl, damit die Zeilenlänge nicht zu lang wird.

Die Struktur ist bewusst einfach gehalten: keine komplexen Animationen, keine überladenen Effekte. Der Fokus liegt auf dem Inhalt, was für eine wissenschaftliche Arbeit wichtig ist. Die Bilder

sind in einem Grid angeordnet, mit Fallback-Placeholders falls Dateien fehlen.

Logbuch: Pixel-Effekt und horizontale Timeline Das Herzstück der Website ist das Prozess-Logbuch. Hier werden alle Journal-Einträge aus dem `journal/` Ordner automatisch geladen und angezeigt. Jeder Eintrag ist eine Markdown-Datei, die beim Laden geparsst und in ein Grid-Layout umgewandelt wird.

Die Timeline oben ist horizontal scrollbar und zeigt alle verfügbaren Einträge mit Datum und Titel. Klickt man auf einen Eintrag, wird er im Hauptbereich angezeigt. Die Einträge selbst werden in "Bubbles" dargestellt – jeder Abschnitt (überschrieben mit `##` oder `###`) wird zu einer eigenen Karte im Grid.

Der Pixel-Effekt kommt beim Titel zum Tragen: Wenn ein Eintrag geladen wird, erscheint der Titel mit einem Typing-Effekt, Buchstabe für Buchstabe, als würde jemand tippen. Dazu gibt es einen blinkenden Cursor. Die Font ist "Press Start 2P", und der Titel hat einen animierten Gradient, der von grün zu weiß übergeht.

Technisch wird das so umgesetzt: Der Markdown-Parser (`parseMarkdown` Funktion) teilt jeden Eintrag in Abschnitte auf. Überschriften werden zu Bubble-Titeln, der Rest wird als HTML formatiert. Bilder, Links und spezielle Komponenten (wie die `WORLD_INFO` Dropdowns für Marble-Welten) werden dabei berücksichtigt. Das Grid passt sich automatisch an die Bildschirmgröße an – 3 Spalten auf Desktop, 2 auf Tablet, 1 auf Mobile.

Git-Versionierung und Projektstruktur Jeder größere Entwicklungsschritt wurde in Git committet. Die Commits zeigen den Prozess: Zuerst die Grundstruktur, dann der Player, dann das Journal-System, dann die Timeline, dann die Bubble-LAYOUTS. Das macht es einfach, zu sehen, wie sich die Website entwickelt hat.

Die Projektstruktur ist bewusst flach gehalten: `- index.html` – Haupt-HTML - `scripts.js` – Alle JavaScript-Logik (Player, Journal-Loader, Navigation) - `style.css` – Alle Styles (Dark Theme, Grid-LAYOUTS, Animationen) - `journal/*.md` – Markdown-Einträge, die automatisch geladen werden - `assets/` – Bilder, Audio, Journal-Assets

Keine Node-Modules, keine Build-Steps – alles läuft direkt im Browser. Das macht die Entwicklung schnell und das Deployment einfach. Für lokales Testen reicht ein einfacher HTTP-Server (z.B. `python -m http.server 8001`).

Fazit Die Website ist jetzt in Version 1.0 – Player funktioniert, Journal lädt automatisch, Design ist konsistent. Cursor hat den Prozess deutlich beschleunigt, besonders bei repetitiven Aufgaben wie dem Markdown-Parsing oder dem Grid-Layout. Die Git-Versionierung gibt mir Sicherheit, dass ich jederzeit zurückkehren kann, wenn etwas schiefgeht.

Was mir besonders geholfen hat: Ich habe bereits Erfahrung mit Web-Entwicklung, und das macht einen großen Unterschied, wenn man mit KI programmiert. Man kann die Vorschläge der KI besser einschätzen, weiß, was funktioniert und was nicht, und kann gezielt nachfragen oder anpassen. Cursor ist ein mächtiges Tool, aber es ersetzt nicht das Verständnis für die Grundlagen – es beschleunigt den Workflow enorm, wenn man weiß, was man will. Die Basis steht, und das war das Ziel.

Reflexionsinterview Kontext

Claude Welche Tools hast du verwendet?

Jennifer Cursor, Version Control via Github, und größtenteils die Modelle von Anthropic.

Claude Was wolltest du erreichen?

Jennifer Ich wollte sowohl die Website erstellen als auch Cursor ausprobieren, da es sich interessant für mich anhörte und ich bereits Webentwicklungserfahrung habe, aber nur sehr limitiert und es würde sehr lange dauern, wenn ich es ganz alleine mache.

Die 3 Forschungsfragen

Claude Hat sich dein Arbeitsablauf verändert? (*Workflow*)

Jennifer Oh mein Gott, das war einfach sehr krass! Das hat das Aufsetzen der Website einfach tausendfach schneller gemacht. Also, hier war der größte Unterschied zu merken, wo ich konnte mein bestehendes Wissen einbringen. Aber musste bei Kleinigkeiten nicht selber googeln und bug fixen, sondern Cursor. Die KI-Modelle haben das für mich erledigt. Ich konnte schnell experimentieren und ich hatte zwar Probleme beim Debugging, aber ich wäre gar nicht so weit gekommen ohne Cursor und Claude. Ich konnte quasi meine kreative Vision umsetzen.

Claude Wie hast du deine Rolle erlebt? (*Autorschaft*)

Jennifer I felt like a manager with the overall vision asking a junior developer to code for me. At some point, I felt like I was collaborating a bit when it couldn't figure out where it was going wrong. And I was reviewing a lot what I was doing. At the same time, I was learning because when I was doing things I couldn't have done myself, I asked it, or I could see its thinking process too. So I also learned a lot.

Claude Wie hat die schnelle KI-Ausgabe deinen Prozess beeinflusst? (*Iteration*)

Jennifer Die KI-Ausgabe war sofort da. Code, Layouts, Fixes. Das hat meinen Workflow massiv beschleunigt, weil ich nicht mehr googeln und Stack Overflow durchforsten musste. Aber es hat auch dazu geführt, dass ich manchmal zu schnell akzeptiert habe, ohne den Code richtig zu verstehen. Das Tempo verleiht zur Oberflächlichkeit.

Claude Was ist schiefgelaufen? (*Scheitern*)

Jennifer Also, wir kamen ab und zu in Debugging-Schleifen. Das war schon sehr nervig, aber ich habe dann besser gelernt, wie man die KI anleiten kann. Zum Beispiel, dass man öfter den Chat erneuern sollte, damit umso mehr Informationen im Kontext sind, desto schlechter die KI funktioniert. Es wurden auch falsche Ansätze gewählt, welche ich dann ablehnen musste bzw. zurückrudern musste.

Claude Hat sich eine eigene visuelle Sprache entwickelt? (*Ästhetik*)

Jennifer Ja, ein bisschen Kepler-DNA und der Matrix-Look. Ich mag dunkle Websites, und ich hatte mich ja vom Research Catalogue inspirieren lassen. Da waren die Artikel, die mich inspiriert haben, aber weiß, und ich wollte einen dunklen Look für dieses Projekt.

Synthese

Claude Du hast geschrieben, dass deine bestehende Web-Erfahrung entscheidend war, um KI-Vorschläge einzuschätzen. Glaubst du, jemand ohne diese Erfahrung hätte ein ähnliches Ergebnis erreichen können?

Jennifer Also ich glaube, es hängt sehr stark von der Person ab. Ich glaube, Leute, die gar keinen Hintergrund haben und auch nicht so viel Geduld, würden schnell an ihre Grenzen kommen. Also, es ist nicht so, dass man einfach alles an die KI geben kann. Garbage in, garbage out. Man muss schon ein bisschen mitlernen, obwohl es schon Ergebnisse gibt, die mit relativ viel wenig Arbeit gut sein können. Ich denke, wenn jemand einfach nur sagt: "Hey, kannst du mir einen Taschenrechner programmieren? Vielleicht in einer bestimmten Farbe, die mir gefällt?" – dann würde das schnell funktionieren, einfach nur weil es so viel vorhandener Code für das Programmieren von Taschenrechnern gibt.
Aber umso spezifischer die Anforderungen werden, umso länger der Chat und umso größer die codebase, desto schwieriger wird es das ganze für einen Laien zu managen.

Journal-Eintrag 8: KI wirklich verstehen - Teil 1 (11. Januar 2026)

Ich wollte Künstliche Intelligenz besser für mich selbst verstehen, auch außerhalb des Artistic Research Contexts, bevor ich dort weiter eintauchte. Ich wollte nicht nur die Tools benutzen, die aufkommen, sondern wirklich verstehen, was dort gerade vor sich geht.

Eigentlich wollte ich auf der Autofahrt einen Podcast hören. Aber die meisten, die ich gefunden habe, waren für sehr einfache Anwendungsfälle – wie können normale Menschen ChatGPT nutzen, um E-Mails zu schreiben. Ich wollte in die tieferen Themen rein. Ich hatte das Gefühl, dass manche Leute viel weiter sind als ich, und ich wollte genau die Lücken aufholen, wo ich durchfalle.

Dann dachte ich: Perfekt wäre eigentlich KI selbst, weil die ja immer verfügbar ist. Also Freisprechanlage an, Claude im Chat-Interface, das meine Stimme in Text transkribierte. Zugegebenermaßen etwas gefährlich am Anfang – das System stockte manchmal, ich musste Sätze wiederholen, war abgelenkt. Aber dann hatten wir uns eingespielt, und es wurde zu einem echten Gespräch über 30 Minuten Fahrtzeit.

Jennifer Ich will mehr über KI lernen. Ich will das wirklich tiefer verstehen.

Jennifer Vielleicht kannst du ein Gespräch mit mir führen und mein Level einschätzen, und mir dann ein Level höher erklären?

Claude Klar! Erzähl mir erstmal, was du schon weißt und womit du arbeitest.

Jennifer Also ich nutze Claude Code gerade, hab ein paar GitHub-Repos für Research, schreibe meine Masterarbeit, mache generelle Recherche zu PhD-Themen...

Jennifer Ich hab Web-Development-Erfahrung und bin generell im Computer Graphics Bereich. Also breites aber vielleicht nicht super tiefes Wissen.

Jennifer Ich will die Grundlagen von KI und KI-Forschung verstehen. Wie fängt der Weg von der Mathematik zur KI an?

Jennifer Gib mir ruhig lange Antworten. Ich kann gerade nicht lesen, also erkläre es ausführlich.

Claude Mathematik ist die Sprache, die KI spricht, weil Computer nur Zahlen verstehen. Wenn ein Mensch ein Foto sieht, verarbeitet das Gehirn Farben, Formen und Muster sofort – aber ein Computer sieht dasselbe Foto als Millionen von Zahlen, die Pixelwerte darstellen. Jede KI beginnt damit, reale Informationen in mathematische Repräsentationen umzuwandeln, die Computer manipulieren können. Der Weg von einfacher Mathematik zu KI beginnt mit Statistik und Wahrscheinlichkeit. Wenn Menschen Entscheidungen treffen, berechnen sie unbewusst Chancen basierend auf vergangenen Erfahrungen. Wer dunkle Wolken sieht, greift wahrscheinlich zum Regenschirm, weil er gelernt hat, dass dunkle Wolken oft Regen bedeuten.

Frühe KI-Forscher versuchten, genau das nachzuahmen: Systeme zu erschaffen, die aus Beispielen lernen, statt starren Regeln zu folgen. Statt einem Computer jeden möglichen Schachzug einzuprogrammieren, erkannten sie, dass man ihm Tausende von Schachpartien zeigen und ihn selbst Gewinnmuster entdecken lassen konnte.

Dafür brauchte es drei mathematische Durchbrüche: Erstens lineare Algebra, um massive Datenmengen gleichzeitig zu verarbeiten. Zweitens Analysis, um die Schätzungen des Systems zu optimieren und zu verbessern. Drittens Wahrscheinlichkeitstheorie, um mit Unsicherheit umzugehen und Vorhersagen auch bei unvollständigen Informationen zu treffen.

Jennifer Ja. Das wurde ja im letzten Jahrhundert ein Ding... Das basiert auf Mathe.

Jennifer Hilf mir, den Weg von 1920 bis 2020 zu verstehen. Wann hat KI eigentlich angefangen? Wo wurde geforscht? Wie kam es zu Durchbrüchen?

Claude Das formale Feld der künstlichen Intelligenz wurde auf der Dartmouth-Konferenz im Sommer 1956 geboren. Vier brillante Forscher überzeugten die Rockefeller Foundation, einen ambitionierten achtwöchigen Workshop zu finanzieren. Ihr Antrag war unglaublich kühn – sie glaubten, dass "jeder Aspekt des Lernens oder jede andere Eigenschaft der Intelligenz im Prinzip so präzise beschrieben werden kann, dass eine Maschine sie simulieren kann".

Jennifer Ja, die Prägung des Begriffs. Erzähl mir mehr.

Claude John McCarthy von Dartmouth erfand buchstäblich den Begriff "Artificial Intelligence" für diesen Workshop, weil sie einen einprägsamen Namen für ihren Förderantrag brauchten. Vorher nannte man das "thinking machines" oder "electronic brains" – zu sehr nach Science-Fiction. Die ursprüngliche Gruppe umfasste McCarthy, der später die Programmiersprache LISP erfand; Marvin Minsky von Harvard, der zum Vater der KI am MIT wurde; Claude Shannon von Bell Labs, der die Informationstheorie erschuf und damit digitale Kommunikation erfand; und Nathaniel Rochester von IBM, der einige der ersten elektronischen Computer entwarf. Sie luden zehn Teilnehmer ein, darunter Allen Newell und Herbert Simon von Carnegie Mellon, die ihren "Logic Theorist" mitbrachten – ein Programm, das tatsächlich mathematische Theoreme beweisen konnte. Das war das erste echte KI-Programm, das funktionierte.

Jennifer Oh, das ist super interessant. Das will ich in meiner Masterarbeit haben...

Jennifer Kannst du das noch mehr erklären? Das mit Claude Shannon?

Claude Claude Shannons Informationstheorie von 1948 ist fundamental für alles in der KI. Shannon fand heraus, wie man Information mathematisch in Bits messen kann. Er bewies, dass jede Information – ob Text, Bilder oder Ton – in Binärcode umgewandelt und mit perfekter Genauigkeit übertragen werden kann, wenn genug Fehlerkorrektur hinzugefügt wird.

Das ist gewaltig, weil es bedeutet, dass alles menschliche Wissen digitalisiert und von Maschinen verarbeitet werden kann. Shannons Arbeit bei Bell Labs ermöglicht direkt Computer, das Internet und jede digitale Technologie, die wir heute nutzen.

Information hat eine mathematische Struktur. Ein Münzwurf liefert genau ein Bit Information, weil er Unsicherheit zwischen zwei gleich wahrscheinlichen Ergebnissen auflöst. Eine Nachricht, die sagt "die Sonne wird morgen aufgehen", enthält fast keine Information, weil sie nahezu sicher ist.

Dieser mathematische Rahmen ermöglicht es Computern, Bedeutung zu verarbeiten – nicht nur Zahlen. Wenn jemand einen Satz tippt, reduziert jeder Buchstabe die Unsicherheit darüber, was als nächstes kommt, basierend auf Mustern in der Sprache. Moderne KI-Systeme nutzen dieses Prinzip, um das nächste Wort in einer Sequenz vorherzusagen.

Jennifer Okay. Super Erklärung. Kannst du da noch ein bisschen tiefer reingehen?

Jennifer Und wird das heute noch benutzt, die symbolischen Repräsentationen? Das sind ja Weights und Biases jetzt, oder?

Claude Symbolische KI und moderne neuronale Netze mit Gewichten und Bias sind fundamental unterschiedliche Ansätze zur Wissensrepräsentation.

Der symbolische Ansatz versucht, Wissen explizit in menschenlesbaren Regeln und Logik zu erfassen. Wenn der Logic Theorist weiß "Alle Menschen sind sterblich", existiert diese Regel buchstäblich im Code als Symbole, die gelesen und verstanden werden können.

Aber Gewichte und Bias in neuronalen Netzen repräsentieren Wissen auf eine völlig verteilte, implizite Weise. Ein neuronales Netz, das etwas über Sterblichkeit weiß, hat diese Regel nirgendwo geschrieben. Stattdessen entsteht dieses Wissen aus Millionen numerischer Gewichte, die statistische Muster über alle Trainingsdaten kodieren.

Der Schlüsselunterschied: Symbolische KI speichert Wissen so, wie Menschen über Wissen denken – mit klaren Konzepten und logischen Beziehungen. Neuronale Netze speichern Wissen so, wie Gehirne tatsächlich funktionieren könnten – als Verbindungsstärken zwischen einfachen Verarbeitungseinheiten.

Jennifer Okay. Wie realistisch ist es für mich, das zu verstehen?

Jennifer Ich glaub nicht, dass es so sehr vom Anwendungsfall abhängt... Ich meine, Large Language Models sind ja was anderes als, sagen wir, Blur-Filter in Photoshop, oder?

Jennifer Ich meine, wir hatten ja schon Noise Reduction und Color Grading. Wieso fühlt sich jetzt alles gleichzeitig an?

Claude Es gibt tatsächlich ein Spektrum von einfachen deterministischen Filtern bis zu moderner KI. Traditionelle Photoshop-Filter wenden feste mathematische Operationen an. Ein Blur-Filter konvolviert jedes Pixel mit seinen Nachbarn mit demselben Kernel. Deterministisch und vorhersagbar. Die erste Evolution waren adaptive Algorithmen. Noise Reduction in den frühen 2000ern begann, statistische Analyse zu nutzen, um Signal von Rauschen zu unterscheiden. Statt überall denselben Filter anzuwenden, analysierten diese Tools lokale Bildeigenschaften und passten ihr Verhalten an. Dann kam Machine Learning um 2010. Statt von Hand zu kodieren, wie man Kanten erkennt, trainierten Forscher Algorithmen auf Tausenden von gelabelten Beispielen. Das waren noch keine neuronalen Netze, sondern statistische Mustererkennung. Was alles veränderte, war Skalierung. GPUs machten es möglich, viel größere Netze auf viel mehr Daten zu trainieren. Dieselben mathematischen Prinzipien, die frühe Computer Vision antrieben, wurden plötzlich mächtig genug, um komplexe Muster in Bildern, Text und anderen Domänen zu erkennen.

Jennifer Und ja. Und dann, je nachdem... GPT-2 kam raus...

Claude GPT-2 war der Moment, in dem Forscher realisierten, dass Skalierung qualitativ neue Fähigkeiten erschafft – nicht nur bessere Leistung bei denselben Aufgaben. Vor GPT-2 dachte jeder, dass spezifische KI-Systeme für jede Aufgabe trainiert werden müssen. Will man Übersetzung? Trainiere ein Übersetzungsmodell. Will man Zusammenfassung? Trainiere ein Zusammenfassungsmodell. GPT-2 zeigte, dass ein genügend großes Sprachmodell, das einfach trainiert wurde, das nächste Wort vorherzusagen, plötzlich all diese Aufgaben erledigen konnte, ohne jemals explizit dafür trainiert worden zu sein. Das schockierte die Forscher. Sie erwarteten, dass größere Modelle besser darin werden, Text vorherzusagen, aber sie erwarteten nicht, dass sie spontan Reasoning-Fähigkeiten entwickeln würden, grundlegende Arithmetik oder die Fähigkeit, Code zu schreiben. Diese Fähigkeiten entstanden rein durch Skalierung.

Jennifer Ja. Jetzt ist es ja nicht mehr nur das Training der Modelle...

Jennifer Also ich nutze KI ja nicht nur für Text. Auch für Bilder, 3D, Video...

Jennifer MCP, und jetzt Skills versus Agents und so. Können wir noch eine Sekunde auf dieser Ebene bleiben?

Claude Die zentrale Einsicht ist, dass Transformer – die Architektur, die Sprachmodelle antreibt – jeden Datentyp verarbeiten können, sobald er in Tokens umgewandelt ist. Text wird in Wort-Teile tokenisiert, Bilder werden in Patches unterteilt, Audio wird zu Spektrogrammen. Einmal alles in Tokens, kann dieselbe mathematische Maschinerie alles verarbeiten.

Retrieval Augmented Generation (RAG), wie es Tools wie Claude Code nutzen, löst das Problem des Wissens-Cutoffs. Statt zu versuchen, alles während des Trainings zu memorieren, durchsuchen RAG-Systeme externe Datenbanken in Echtzeit und injizieren diese Information in den Kontext des Modells. MCP – Model Context Protocol – ist im Grunde ein standardisierter Weg für KI-Modelle, sich mit externen Tools und Datenquellen zu verbinden. Statt dass jede Firma eigene Integrationen baut, schafft MCP eine universelle Schnittstelle. Wie USB für KI.

Skills sind spezifische Fähigkeiten, die eine KI ausführen kann – wie "Spreadsheets analysieren" oder "Code generieren". Sie sind reaktive Tools. Agents sind autonome Systeme, die planen, mehrere Schritte ausführen und selbstständig auf Ziele hinarbeiten können.

Der vereinigende Faden sind Attention-Mechanismen – die mathematische Operation, die Transformern erlaubt, sich auf relevante Informationen zu fokussieren, unabhängig vom Datentyp. Deshalb explodieren KI-Fähigkeiten gleichzeitig über alle Domänen – dieselbe Kerntechnologie funktioniert überall, sobald das Tokenisierungsproblem gelöst ist.

Wie lerne ich das wirklich? *Gegen Ende der Fahrt wurde mir klar: Ich hatte jetzt all diese Informationen gehört – aber habe ich sie wirklich gelernt?*

Jennifer Okay, wenn ich dieses Thema wirklich tief verstehen will, und auch die Brücke schlagen will zu den Technologien, die ich benutze – Cursor, Claude Code für Research und Coding, ComfyUI, Houdini, Unreal Engine – was wäre der beste Weg dahin?

Jennifer Sollte ich den Text selbst schreiben? Auch wenn du ihn generiert hast? Es gibt ja diese Studien, dass das Verständnis nicht so tief ist, wenn Essays mit KI geschrieben werden. Aber ich brauche dieses Fundament, um weiterzumachen.

Jennifer Andererseits habe ich das Gefühl, dass ich Wissen schrittweise aufbaue. Manchmal wollte ich Computer Graphics wirklich tief verstehen. Hab reingeschaut und dann aufgehört. Aber ich glaube, ich habe danach ein höheres Verständnis. Oder bilde ich mir das nur ein?

Jennifer Zum Beispiel Houdini lernen – zum vierten oder fünften Mal. Ich mache langsam Fortschritte. Wie lernen menschliche Gehirne eigentlich?

Jennifer Ich hatte meinen ersten Computer mit 7. Einfache Spiele, Kekse dekorieren. Dann hab ich viel gezeichnet und war im Internet, dann hab ich eine Pokemon-Website gemacht, indem ich HTML und CSS von jemand anderem kopiert habe. In der Schule hatte ich Programmieren, ganz basic. Dann Innenarchitektur studiert mit 3ds Max und am Ende Unreal Engine. Das ist viel. Ein langer Weg, um all dieses spezialisierte Wissen zu sammeln.

Claude Du stellst die richtige Frage, und die ehrliche Antwort ist kompliziert.

Die Forschung, die du meinst, ist real – wenn du etwas selbst schreibst, wirst du gezwungen, dich mit Lücken in deinem Verständnis auseinanderzusetzen. Lesen oder Zuhören fühlt sich wie Lernen an, weil die Information glatt fließt, aber diese Flüssigkeit ist trügerisch. Du kannst bei einer Erklärung von Back-propagation nicken, ohne sie fünf Minuten später selbst erklären zu können. Das nennt man die Fluency Illusion.

Aber hier ist die Nuance: Es gibt verschiedene Arten von Wissen. Deklaratives Wissen ist "wissen dass" – Fakten, Konzepte, Geschichte. Prozedurales Wissen ist "wissen wie" – Dinge tatsächlich tun. Dein Weg von Kekse-Dekorieren über Pokemon-HTML zu 3ds Max zu Unreal ist prozedurales Wissen, das sich ansammelt. Das ist real und bleibt hängen auf eine Weise, die Lesen niemals erreicht.

Dein Muster, Houdini aufzugreifen, fallen zu lassen, Monate später zurückzukommen und etwas besser zu sein – so funktioniert tiefes Lernen tatsächlich. Dein Gehirn konsolidiert während der Pausen. Schlaf, Zeit weg, angrenzende Dinge tun – das sind keine Misserfolge, das ist Teil des Prozesses. Der vierte Versuch fängt nicht von vorne an, er baut auf neuronalen Bahnen auf, die sich während der Versuche eins bis drei gebildet haben.

Um KI-Verständnis mit deinen tatsächlichen Tools zu verbinden, würde ich vorschlagen: Schreib meinen Text nicht um. Stattdessen: Nimm einen Workflow, den du schon machst – vielleicht Depth-Map-Extraktion in ComfyUI – und verfolge die KI-Pipeline rückwärts. Welches Modell wird verwendet? Worauf wurde es trainiert? Was passiert mathematisch? Lernen, das an Tun geknüpft ist, bleibt hängen. Abstrakte Grundlagen ohne Anwendung schweben davon.

Jennifer Ja, ich fühle, dass es entscheidende Lücken gibt. Um es wirklich vollständig und tief zu verstehen.

Claude Wo fühlst du die Lücken?

Versuch mir zu erklären – der Akt des Artikulierens, wo dein Verständnis zusammenbricht, ist selbst diagnostisch. Kannst du zum Beispiel nachvollziehen, was passiert, wenn du einen Prompt in ComfyUI eingibst und ein Bild zurückbekommst? An welchem Punkt wird es zur Black Box?

Jennifer Nur ganz leicht. Ich verstehe Bild- oder Video-Generierungsmodelle nicht ganz, und ich bin mir nicht sicher, wie interessiert ich daran bin.

Jennifer Ich will eigentlich einen Lo-Fi Soviet Dystopian Winter Soundtrack mit KI erstellen. Vielleicht kann ich da anfangen.

Claude Das ist ein großartiger Einstiegspunkt – spezifisch, kreativ, und persönlich interessant. Du wirst mehr lernen, wenn du dem nachjagst, als wenn du dich zwingst, Image Diffusion zu verstehen, das dich nicht interessiert.

Was weißt du bisher über KI-Musikgenerierung? Hast du dir schon Tools angeschaut, oder ist das noch im Ideenstadium?

Reflexionsinterview Kontext

Claude Welche Tools hast du verwendet?

Jennifer Claude über die mobile App mit Spracheingabe, während einer 30-minütigen Autofahrt. Freisprechanlage an, quasi wie ein Telefonat.

Claude Was wolltest du erreichen?

Jennifer Ich wollte KI wirklich verstehen, nicht nur die Tools benutzen, sondern die Grundlagen. Was passiert mathematisch? Wie hängen Transformer, Diffusion Models und die Tools, die ich benutze, zusammen? Und wie lerne ich das am besten?

Die 8 Forschungsfragen

Claude Hat sich dein Arbeitsablauf verändert? (*Workflow*)

Jennifer Ein komplett neuer Lernmodus. Statt Podcast oder YouTube-Video hatte ich ein personalisiertes Gespräch. Die KI hat sich an mein Level angepasst, ich konnte jederzeit nachfragen. Dialog statt passives Konsumieren.

Claude Wie hast du deine Rolle erlebt? (*Autorschaft*)

Jennifer Die Fragende. Ich habe die Richtung bestimmt, welche Themen, wie tief, welche Beispiele. Claude hat erklärt, aber ohne meine Fragen wäre das Gespräch nirgendwohin gegangen.

Claude Wie hat die KI-Ausgabe deinen Prozess beeinflusst? (*Iteration*)

Jennifer Zum einen denke ich, dass ich mein bereits bekanntes Wissen noch einmal gestärkt habe und mich an ein paar Konzepte und historische Ereignisse erinnert habe, die ich schon vergessen hatte. Ich glaube aber, es könnte die Gefahr geben, dass ich der Illusion unterliege, dass ich es verstehe, obwohl ich es vielleicht nicht tue. Aber das Gleiche gilt ja auch, wenn ich mir ein Video anschauе.

Claude Was ist schiefgelaufen? (*Scheitern*)

Jennifer Es lief noch nicht ganz so flüssig. Ich glaube, die Implementation von ChatGPT ist ein bisschen besser.

Da wurde ich ein bisschen von der Autofahrt abgelenkt, ich würde das vielleicht nicht unbedingt noch einmal wiederholen. Wäre es jetzt wie ein Telefonat, dann ja, aber durch die Bugs lieber nicht. Aber sicherlich wird sich noch einiges tun in den nächsten Monaten.

Claude Hat sich eine eigene 'Lern-Sprache' oder Methode entwickelt? (*Ästhetik*)

Jennifer Ich denke ich würde jetzt öfters im Dialog mit der KI versuchen Dinge zu lernen.

Synthese

Claude Am Ende des Eintrags hast du geschrieben: 'Die Theorie wird erst real, wenn ich sie anwende. Nächster Schritt: Der Lo-Fi Soundtrack.' – Hast du das inzwischen gemacht?

Jennifer Ja, das Suno-Experiment war der direkte nächste Schritt. Claude hatte vorgeschlagen, KI-Verständnis an konkreten Projekten zu verankern statt abstrakt zu lernen, und genau das habe ich dann gemacht.

Journal-Eintrag 6: Suno-Experiment {12. Januar 2026}

Heute ein kleines Experiment außerhalb des Kepler-Kontexts. Ich arbeite an einem anderen Projekt – einer prozedural generierten Spielwelt mit sowjetischer Ästhetik – und wollte testen, wie gut KI-Musikgenerierung für spezifische Game-Audio-Anforderungen funktioniert.

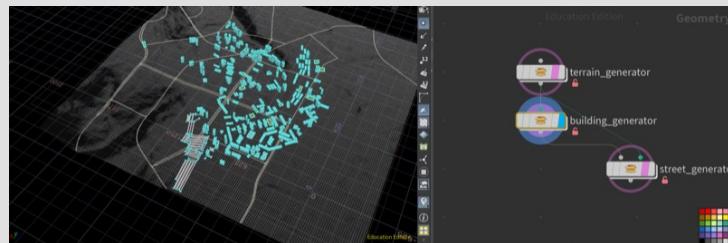


Abbildung 22: Houdini-Workflow mit den drei prozeduralen Generatoren: `terrain_generator`, `building_generator` und `street_generator` für die sowjetisch inspirierte Spielwelt.

Prozeduralismus und Pattern-basiertes Design Das Projekt „ПАНЕЛЬКИ“ (Panelki) entsteht im Rahmen des Masterkurses über prozedurale Weltgenerierung – Virtuelle Welten. Der Kern: Anstatt jedes Gebäude und jeden Straßenzug manuell zu platzieren, definiere ich *Regeln* und *Muster*, aus denen die Welt emergent entsteht. In Houdini habe ich drei HDAs (Houdini Digital Assets) entwickelt:

- **terrain_generator**: Echte Höhendaten aus Neryungi, Russland (ALOS PALSAR DEM)
- **building_generator**: Modularer Soviet-Architektur aus OpenStreetMap-Footprints
- **street_generator**: Straßennetzwerke, die die Gebäude verbinden

Das Ergebnis ist eine dystopische Winterstadt, bevölkert von anthropomorphen Tierfiguren im Stil sowjetischer Zeichentrickfilme.



Abbildung 23: Konzeptkunst: Wolf unter einer Straßenlaterne im ewigen Winter. Konzeptbild erstellt mit Google Gemini (Nano Banana Pro).

KI-Musik Was mich zu diesem Experiment gebracht hat: Ich sah ein Video eines Musikproduzenten in meinem YouTube Feed: viele kommerzielle Musik wird inzwischen KI-generiert – ohne dass es auffällt. Bei Bildern und Videos sehe ich noch deutliche Artefakte und Inkonsistenzen, KI-Werbespots werden regelmäßig „außenandergenommen“ und kritisiert. Bei Audio liegt es anscheinend für einen Laien nah am professionellen Standard. Suno ist wohl aktuell der Marktführer.

Das Experiment: Zwei Prompts Das Video hatte mich inspiriert: ich brauchte Hintergrundmusik für meine Welt, die einen Pixel-look haben sollte: loopbar, atmosphärisch, sowjetisch-dystopisch. Keine Vocals. Also habe ich mich angemeldet und zwei Ansätze auf Suno getestet.

Versuch 1: "Lo-fi" Ergebnis: Klingt zu kommerziell. "Lo-fi" war der falsche Begriff – ich wollte etwas Minimalistisches, Roheres. Die generierte Beschreibung klingt besser als der tatsächliche Output.

Versuch 2: "Minimal" Ergebnis: Nicht wirklich minimal. Der Fokus auf "pixel game soundtrack" hat dazu geführt, dass es teilweise wie ein Hero-Soundtrack aus einem Indie-Game klingt – nicht wie atmosphärische Hintergrundmusik für ein Spiel, in dem man explorativ durch eine verlassene Stadt läuft. Die Beschreibungen sind interessant ("aging arcade cabinet in an abandoned bunker"), ich finde der tatsächliche Sound geht in eine andere Richtung.



Abbildung 24: Pixel-Art-Ästhetik: Das Schaf auf sowjetischer Straße – genau dieser Stil bräuchte passende Musik. Konzeptbild erstellt mit Midjourney v7.

Die Beschreibungs-Output-Diskrepanz Was auffällt: Suno generiert wunderbar detaillierte Beschreibungen. "Bitcrushed chiptune pads over a sparse, detuned synth choir" klingt exakt nach dem, was ich will. Der tatsächliche Output? Klingt halt etwas zu "polished", wie Stock-Music.

Das ist dasselbe Muster wie bei meinen ComfyUI-Experimenten mit Kepler: Ich wollte einen voxelisierten, pixeligen Charakter ohne erkennbares Gesicht. Die KI lieferte typisch KI-generiert aussehende Ergebnisse mit Uncanny-Valley-Gesichtern, obwohl der Prompt explizit "voxel", "blocky", "no face" enthielt.

Die Erkenntnis: Generalistische KI-Tools sind wahrscheinlich einfach auf kommerzielle Mehrheitsästhetik trainiert. Spezialisierte Nischen – Pixel-Game-Soundtracks, voxelisierte Charaktere – erfor-

dern entweder sehr viele Iterationen, spezialisierte Fine-Tunes oder manuelle Nachbearbeitung. In diesem Fall hat das Tool die Anforderungen nicht erfüllt – der Output schien für mich “handwerklich solide”, aber stilistisch daneben, “too much”. Für kommerzielle Mainstream-Musik mag das funktionieren, für spezialisierte Nischen braucht es nach wie vor menschliche Expertise oder zumindest erhebliche Nachbearbeitung.

Technische Notizen Suno bietet zwei Modelle: - **v4.5**: Kostenlos, 50 Credits/Tag, “previously unimaginable production capabilities” - **v5**: \$30/Monat (Pro), noch höhere Qualität, v5-Samples werden parallel angeboten

Ich habe beide Modelle getestet: Die “Frozen”-Tracks (Frozen Terminals, Frozen Relay) sind mit v5 (beta) generiert – im Free Trial allerdings auf eine Minute begrenzt. Red Snow Protocol und Iron Frost Protocol stammen aus v4.5 und sind länger, gefallen mir aber weniger – **warum genau, kann ich nicht sagen.**

Reflexionsinterview Kontext

Claude Welche Tools hast du verwendet?

Jennifer Suno AI, im Browser. Ich habe sowohl v4.5 (kostenlos) als auch v5 (beta, im Free Trial) getestet.

Claude Was wolltest du erreichen?

Jennifer Einen loopbaren, atmosphärischen Soundtrack für meine sowjetisch-dystopische Spielwelt Panelki. Minimal, kalt, pixelig. Keine Vocals, eher Hintergrundmusik zum Explorieren.

Die 8 Forschungsfragen

Claude Hat sich dein Arbeitsablauf verändert? (*Workflow*)

Jennifer Es war ein schneller Prototyp. Ich konnte Ideen testen, ohne selbst Musik produzieren zu können. Prompt rein, Ergebnis in Sekunden. Aber genau diese Geschwindigkeit hat auch gezeigt, wie wenig Kontrolle ich über das Ergebnis hatte.

Claude Wie hast du deine Rolle erlebt? (*Autorschaft*)

Jennifer Enttäuschend. Ich war nur Kurator:in, ich konnte auswählen, aber nicht gestalten. Kein Regler, kein Parameter, den ich hätte drehen können. Nur der Prompt, und der wurde offensichtlich anders interpretiert als gemeint.

Claude Wie hat die schnelle KI-Ausgabe deinen Prozess beeinflusst? (*Iteration*)

Jennifer In dem Fall habe ich das Ganze ja verworfen, weil ich vermute, dass ich mit dem Modell nicht sonderlich weit komme.

Claude Was ist schiefgelaufen? (*Scheitern*)

Jennifer Der Output klang zu kommerziell, zu poliert, wie Stock-Music. Das Frustrierende: Die generierten Beschreibungen trafen genau meinen Geschmack ('Bitcrushed chiptune pads over a sparse, detuned synth choir'), aber die Musik selbst klang völlig anders.

Claude Hat sich eine eigene musikalische/ästhetische Sprache entwickelt? (*Ästhetik*)

Jennifer Nein, komplett verfehlt. Ich wollte minimal-dystopisch, bekommen habe ich generischen Indie-Game-Sound. Das zeigt mir, dass Suno auf kommerzielle Mehrheitsästhetik trainiert ist. Meine Nische liegt einfach außerhalb dessen, was das Modell liefern kann.

Synthese

Claude Im Eintrag schreibst du, dass Sunos generierte Beschreibungen ('Bitcrushed chiptune pads over a sparse, detuned synth choir') besser klingen als die Musik selbst. Ist das frustrierend – oder zeigt es, dass du die KI vielleicht als Text-Generator für Briefings an echte Musiker nutzen könntest?

Jennifer Also zeigt mir, dass für die Vision, die ich hatte, gibt es Begriffe, und es gibt mir Hoffnung, dass ich entweder bereits existierende Musik mit diesen Begriffen finden kann, ein anderes Tool verwenden könnte, um so etwas zu generieren, oder es einem echten Musiker geben könnte. Es ist sehr frustrierend.

Journal-Eintrag 7: Kepler-Recherche, zweiter Versuch (29. Januar 2026)

Im November hatte ich mit ComfyUI versucht, konsistente Bilder von Kepler zu generieren. Es war frustrierend. Die KI interpretierte hartnäckig Gesichter in seine Voxel-Maske hinein — egal, was ich promptete. Face Detailer, PuLID, InstantID: alle auf menschliche Gesichter trainiert, alle unfähig, einen Charakter ohne Gesicht zu respektieren.



Kepler in seiner Voxel-Form —
kein Gesicht, das ist Absicht.



Generiertes Bild aus dem
ComfyUI-Experiment: Die KI hat
ein Gesicht in die Maske
interpretiert.



Noch ein Versuch — wieder ein
Gesicht, wo keines sein sollte.

Seitdem habe ich das Thema nicht losgelassen. In den Wochen dazwischen habe ich andere Dinge gemacht — Suno ausprobiert, mit Claude über KI-Grundlagen geredet, die Thesis-Struktur weitergedacht. Aber im Hinterkopf blieb dieses Bild: Kepler mit einem fremden Gesicht. Es fühlte sich an, als hätte das System meinen Charakter nicht nur falsch dargestellt, sondern ihn *korrigiert*. Als wäre ein gesichtloses Wesen ein Fehler, den die KI beheben muss.

Ich wollte wissen: Liegt es an meinem Workflow? An den Tools? Oder ist das ein grundsätzliches Problem, das alle generativen KI-Modelle betrifft, wenn man nicht-normative Charaktere abbilden will?

Also habe ich mich einen Nachmittag hingesetzt und systematisch recherchiert, mit Claude als Sparringspartner.

Face-Tools sind das Kernproblem Mein erster Fehler im November war nicht die Wahl von ComfyUI — es war die Wahl der Face-Tools. PuLID, InstantID und Face Detailer sind alle auf menschliche Gesichter trainiert. Sie erwarten Augen, Nase, Mund. Kepler hat eine strukturierte Maske ohne erkennbare Gesichtszüge. Das System *will* ein Gesicht sehen, weil die Trainingsdaten sagen: Da, wo ein Kopf ist, muss ein Gesicht sein.

Das ist keine Limitierung meines Prompts. Das ist eine Limitierung der Repräsentation in den Trainingsdaten.

Nano Banana — nicht was ich dachte Bei der Recherche bin ich auf ComfyUI-NanoBanano gestoßen, das Google Gemini 2.5 Flash Image in ComfyUI integriert. Klingt vielversprechend, ist aber ein 2D-Bildgenerator — kein 3D-Tool. Für Kepler-Konzeptbilder potenziell nützlich, aber nicht für das eigentliche Problem: einen voxelisierten 3D-Charakter konsistent zu generieren.

Ich habe es trotzdem ausprobiert — nicht über das ComfyUI-Plugin, sondern über NanoBanano Pro direkt im Gemini Chat, mit dem Kepler-Referenzbild aus dem ersten ComfyUI-Experiment.

Chat 1: Pool-Szene

"Can you based on this reference generate an image where the voxel character is relaxing by a pool." → "Now him in the pool on a pool float." → "Just the pool float with him but from the top."



Gemini hält die 3D-Voxel-Ästhetik und fügt kein Gesicht hinzu.



Die Szene wurde erweitert, aber der zweite Kepler am Beckenrand blieb stehen.

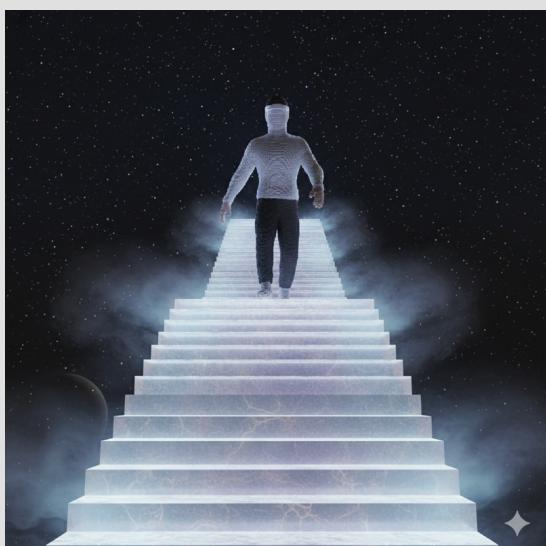


Gemini hat die Draufsicht umgesetzt, aber den liegenden Kepler am Rand nicht entfernt.

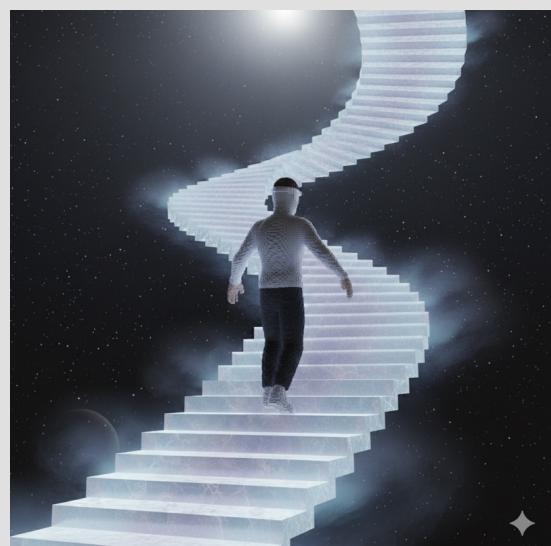
Beeindruckend: Gemini behält die geometrische Voxel-Form bei und fügt keine Gesichter hinzu. Aber jeder neue Prompt baut auf dem vorherigen Bild auf, statt es zu ersetzen — für einen wirklich neuen Stil müsste ich jeweils einen neuen Chat starten.

Chat 2: Treppe im Nachthimmel

"Show this character walking up a set of endless glowing stairs in the night sky. Make it a Portrait image." → "You did not end up making it a portrait, and I want him to be walking up, not down, and the staircase is a spiral staircase."



Beeindruckende Atmosphäre, aber kein Hochformat trotz explizitem Wunsch.



Die Spiraltreppe wurde umgesetzt, aber immer noch kein Portrait-Format. Und die Sterne liegen vor dem Planeten — physikalisch unmöglich.

Die Voxel-Ästhetik bleibt konsistent, kein Gesicht hinzugefügt. Aber Gemini ignoriert die Portrait-Anweisung zweimal, und die Sterne vor dem Planeten zeigen, dass das Modell kein räumliches Verständnis der Szene hat — es komponiert visuell, nicht physikalisch.

Chat 3: Brücke bei Nacht

“Das hier sind ein paar Beispielbilder von Kepler. Behalte seine Voxel-Ästhetik bei und generiere mir ein hochkant ästhetisches Cover. Wie er über eine Brücke in der Nacht fährt in einem Retro-Auto.” → “Sorge dafür, dass der Brückenzaun nicht leuchtet. Wähle eine andere Perspektive und entferne den Text.” → “Die Barrikaden der Brücke leuchten immer noch.”

Ich hatte nur “Cover” geschrieben, nicht “Album-Cover” — Gemini hat den Kontext selbst erschlossen und “NIGHT DRIVE / KEPLER” als Titel gesetzt.



Gemini generiert ungefragt ein Album-Cover mit Titel "NIGHT DRIVE / KEPLER" — es interpretiert "Cover" als Musikcover und fügt selbstständig Text hinzu.



Text entfernt, andere Perspektive — aber der Brückenzaun leuchtet weiterhin, und die Sterne-vor-Planeten-Problematik bleibt.



Dritter Versuch — die leuchtenden Barrikaden lassen sich nicht abschalten.

Chat 4: Typografie auf bestehendem Cover

“Kannst du mir auf dieses Bild einen Album title geben: ‘es tut mir leid’ in liquid metal cyber tribal aesthetic geben.” → “Jetzt soll er nur metallisch grau sein und etwas weiter oben eine andere Schriftart, die noch etwas mehr slanted ist. In den Farben des Covers selbst.” → “Ok, bitte ändert es noch mal komplett um und verschiebt den Text nach oben.”



Gemini setzt den Text in einer generischen Blackletter-Schrift mit leichtem Metallic-Effekt — weit entfernt von "liquid metal cyber tribal".



Kaum Veränderung — die Schrift ist minimal anders positioniert, aber weder slanted noch in den Cover-Farben.



Dritte Iteration — wieder fast identisch. "Komplett um" wurde ignoriert.

Besonders frustrierend: Drei Iterationen, kaum Unterschied. Keine der Anweisungen — liquid metal, cyber tribal, slanted, Cover-Farben, komplett anders — wurde wirklich umgesetzt. Typografie ist offensichtlich eine Schwachstelle: Gemini kann Text auf Bilder setzen, aber die ästhetische Kontrolle über Schriftart, Stil und Platzierung fehlt fast vollständig.

Zwischenfazit Gemini Die vier Chats zeigen ein klares Bild: Gemini ist beeindruckend darin, die Voxel-Ästhetik beizubehalten und Szenen zu generieren — als Inspirationsquelle funktioniert es. Aber für Art Direction ist das Chat-Fenster unbrauchbar. Manche Anweisungen werden präzise befolgt (Szenen-Änderungen, Text entfernen), andere hartnäckig ignoriert (Hochformat, Leuchteffekte, Typografie-Stil, Sterne-vor-Planeten-Physik). Für jeden neuen Anlauf muss man einen frischen Chat öffnen, und selbst dann ist die Kontrolle zu grob. Für feinere Steuerung bräuchte man die API oder spezialisierte Tools wie NanoBanano in ComfyUI. Das Potenzial ist da — aber Spaß macht es nicht.

2D-Workaround: ControlNet Depth + Inpainting Ein vielversprechender Ansatz: Statt Face-Tools zu nutzen, kann ich über Depth Maps arbeiten. [Flux.1-dev-Controlnet-Depth](#) extrahiert die räumliche Struktur eines Bildes und nutzt sie als Kontrollsiegel — ohne das Gesicht interpretieren zu müssen. Kombiniert mit Inpainting könnte ich Kepler-Referenzen mit Tiefenkarte einspeisen und die KI nur den Stil und die Umgebung generieren lassen, während die Silhouette erhalten bleibt.

Das umgeht das Face-Problem elegant: Die Depth Map sieht nur Geometrie, keine Gesichtszüge.

3D-native Modelle — der eigentliche Hoffnungsschimmer Was mich wirklich überrascht hat: Es gibt inzwischen Modelle, die nativ in 3D arbeiten und teilweise explizit mit Voxel-Repräsentationen.

TRELLIS 2 (Microsoft, MIT-Lizenz) arbeitet mit einer voxel-nativen Repräsentation — sogenannten O-Voxels (field-free sparse voxels). Das Modell generiert 3D-Assets aus Text oder Bildern und denkt dabei nativ in Voxeln. Keine 2D-Projektion, keine Face Detection, keine implizite Annahme über Gesichter.



Abbildung 25: TRELLIS 2 generiert 3D-Modelle nativ aus Voxel-Repräsentationen — genau die Datenstruktur, in der Kepler existiert.

XCube (NVIDIA) nutzt hierarchische Voxel-Strukturen für hochauflösende 3D-Generierung. Ein Forschungsprojekt, kein Produkt — aber der Ansatz, über Voxel-Hierarchien zu arbeiten, ist für Kepler natürlicher als jede 2D-basierte Methode.

Hunyuan3D-2 (Tencent, Open Source) ist ein vollständiges 3D-Generierungs-Ökosystem mit Text-to-3D und Image-to-3D. Weniger voxel-spezifisch, aber als offenes Modell experimentierfreundlich.



Abbildung 26: Hunyuan3D-2 (HY3D) — Tencents Open-Source 3D-Generierungsmodell.

VoxAI positioniert sich explizit als voxel-spezifisches KI-Tool. Noch sehr früh und wenig dokumentiert, aber allein die Existenz eines Tools, das sich auf Voxel-Generierung spezialisiert, zeigt: Es gibt Nachfrage für genau diesen Anwendungsfall.

TroublingGAN — ein verwandtes Artistic-Research-Projekt Bei der Recherche bin ich auf ein Projekt gestoßen, das mein eigenes Thema spiegelt: **TroublingGAN** von Lenka Hámošová und Pa-

vol Rusnák (JAR 31). Sie haben GANs bewusst an nicht-standardisierten Körpern scheitern lassen und das Scheitern selbst als ästhetische Strategie dokumentiert.

Ich habe fast aufgehört zu scrollen, als ich das gefunden habe. Jemand anderes hatte dasselbe Problem — und hat daraus Kunst gemacht, statt es als Bug abzutun. Das ist exakt mein Thema: Was passiert, wenn generative KI auf einen Charakter trifft, der die impliziten Normen der Trainingsdaten bricht? Bei mir ist es das fehlende Gesicht. Bei Hámošová & Rusnák waren es nicht-normative Körper. Das Ergebnis ist dasselbe: Das Modell *normalisiert*, weil es auf Normalität trainiert wurde.

Die Kern-Erkenntnis Die Repräsentationsform bestimmt, ob nicht-normative Ästhetik überlebt.

2D-Bildgeneratoren arbeiten mit Pixeln und impliziten Annahmen über Gesichter, Körper, Proportionen. Sie *wollen* normalisieren, weil ihre Trainingsdaten normativ sind. 3D-native Modelle arbeiten mit Geometrie — Voxel, Meshes, Point Clouds. Sie haben keine implizite Erwartung, dass an einer bestimmten Stelle ein Gesicht sein muss.

Das bedeutet: Nicht die KI als solche ist das Problem. Die Repräsentationsebene ist es. Und die kann ich wählen.

Tool-Vergleich

Tool	Typ	Voxel-nativ	Face-Bias	Lizenz	Status
ComfyUI + Face Tools	2D	Nein	Hoch	Open Source	Gescheitert
Control-Net Depth	2D	Nein	Niedrig	Open Source	Vielfersprechend
TRELLIS 2	3D	Ja	Keiner	MIT	Nächster Test
XCube Hunyu-an3D-2	3D	Ja	Keiner	Research	Beobachten
VoxAI	3D	Nein	Niedrig	Open Source	Ausprobieren
		Ja	Keiner	Kommerziell	Beobachten

Fal.ai — API-Zugang zu gesichtsfreien Pipelines Nachtrag, 15. Februar 2026

Nach der Recherche zu lokalen Tools (ComfyUI, TRELLIS 2) habe ich systematisch untersucht, welche Modelle über [fal.ai](#) als API verfügbar sind und für Kepler in Frage kommen. Fal.ai bietet gehostete Inferenz für dutzende Bildmodelle — kein lokales Setup, keine GPU nötig. Die Frage: Gibt es Modelle, die Keplers Voxel-Ästhetik respektieren, ohne ein Gesicht hineinzuinterpretieren?

Strategie 1: Depth Maps & Edge Detection (Face-Bias umgehen) Die vielversprechendste Kategorie. Diese Modelle arbeiten mit Struktursignalen statt Gesichtserkennung:

Modell	Input	Was es tut	Preis
FLUX.1 Control LoRA Depth	Depth Map + Text	Überträgt räumliche Struktur via Tiefenkarte auf generiertes Bild	\$0,04/MP
FLUX.1 Control LoRA Depth (i2i)	Bild + Depth Map + Text	Wie oben, aber mit zusätzlichem Referenzbild	\$0,04/MP
FLUX.1 Control LoRA Canny	Canny Edge Map + Text	Überträgt Kantenstruktur — perfekt für Keplers blockige Silhouette	\$0,04/MP
Z-Image Turbo ControlNet	Edge, Depth oder Pose + Text	Schnelles 6B-Modell mit ControlNet-Support	\$0,007/MP
FLUX.1 General	ControlNet + LoRA + IP-Adapter	Alles kombinierbar: Depth, Canny, LoRA, IP-Adapter	\$0,075/MP

Warum das für Kepler funktioniert: Eine Depth Map von Kepler im A-Pose enthält nur Geometrie — Voxel-Blöcke, Silhouette, räumliche Tiefe. Kein Modell wird versuchen, Augen in eine Tiefenkarte zu halluzinieren. Der Canny-Ansatz ist noch robuster: Kantenlinien einer Voxel-Figur sind geometrisch eindeutig und lassen keinen Raum für Gesichtsinterpretation.

Workflow-Idee: Kepler-Referenzbild → Depth Map extrahieren (lokal oder via API) → Depth Map + Prompt an FLUX Control LoRA → Kepler in neuer Szene, Silhouette erhalten.

Strategie 2: Referenzbild-basierte Konsistenz Diese Modelle nehmen ein oder mehrere Referenzbilder und generieren neue Bilder, die den Charakter beibehalten:

Modell	Input	Was es tut	Preis
MiniMax Subject Reference	Referenzbild + Text	Konsistente Charakterdarstellung aus Referenz	\$0,01/Bild
InstantCharacter	Referenzbild + Text	Charakter in neuen Posen und Stilen	\$0,10/MP
USO	Referenzbild + Text	Subject-driven Generation	\$0,10/MP
UNO	Referenzbild(er) + Text	Transformiert Referenz via Textprompt	\$0,05/MP
Vidu	Referenzbild + Text	Neue Bilder aus Referenz + Prompt	\$0,05/Bild
Reference-to-Image	Referenzbild + Text	Starke Referenzkontrolle für Edits	\$0,03/Bild
Kling O1 Image	Referenzbild + Text		

Risiko: Diese Modelle sind oft auf Gesichtskonsistenz trainiert (Porträts, Produktfotos). Ob sie einen gesichtslosen Voxel-Charakter als "Subjekt" akzeptieren, muss getestet werden. MiniMax Subject Reference ist mit \$0,01 pro Bild billig genug für schnelle Tests. InstantCharacter klingt vielversprechend, aber der Name ("Character") impliziert möglicherweise menschliche Figuren.

Strategie 3: Image Editing (Kepler-Foto als Ausgangspunkt) Statt von Null zu generieren: Ein existierendes Kepler-Bild als Input nehmen und per Textanweisung modifizieren.

Modell	Input	Was es tut	Preis
FLUX 2 Edit	Bild + Text	Natürlichsprachliche Bildbearbeitung	\$0,012/MP
FLUX 2 Pro	Bild + Text	Höhere Qualität, Stiltransfer	\$0,03/MP
FLUX Kontext Pro	Referenz + Text	Lokale Edits + komplexe Transformationen	\$0,04/Bild
FLUX Kontext Pro	Mehrere Referenzen + Text	Mehrere Referenzbilder kombinierbar	\$0,04/Bild
Multi			
Bria Reimagine	Strukturbild + Text	Struktur erhalten, neuer Stil	\$0,04/Bild

Bester Kandidat: FLUX Kontext Pro Multi — ich kann Kepler im A-Pose *und* in der Casual-Pose als Referenz einspeisen und per Text eine neue Szene beschreiben. Das Modell sieht die Voxel-Ästhetik aus mehreren Blickwinkeln und muss sie nicht “erraten”.

Strategie 4: 3D & Multiview

Modell	Input	Was es tut	Preis
Era 3D	Einzelbild	Generiert Multiview + Normalen aus einem Bild	\$0,001/Sek
Hunyuan World	Einzelbild	Einzelbild → Panorama / 3D-Welt	—

Era 3D ist besonders interessant: Aus einem einzigen Kepler-Foto könnte ich automatisiert Multiviews generieren — verschiedene Blickwinkel *mit* Normal Maps. Die Normalen könnten dann wiederum als ControlNet-Input für weitere Generierungen dienen. Quasi ein 2D\$→3D→\$2D-Pipeline, die Keplers Geometrie über mehrere Stufen konsistent hält. Und das für weniger als einen Cent pro Durchlauf.

Strategie 5: Video (Kepler animieren)

Modell	Input	Was es tut	Preis
Wan 2.1	Bild + Text	Kepler-Bild → animiertes Video	\$0,40/Video
Image-to-Video			
DreamActor v2	Bild + Motion-Video	Motion Transfer — explizit für nicht-menschliche Charaktere	\$0,05/Sek

Modell	Input	Was es tut	Preis
Kling O3 Reference-to-Video	Referenzbild + Text	Charakter-konsistentes Video	0,28/Sek [Wan2.6Reference-to – Video](https://fal.ai/models/wan/v2.6/reference-video/flash) Referenzbild+Text Referenz→\$Video, extrem günstig

DreamActor v2 ist der Gamechanger: ByteDance bewirbt es explizit mit “Great performance for non-human and multiple characters”. Das ist exakt Keplers Anwendungsfall — ein nicht-menschlicher Charakter, dessen Bewegungen von einem Motion-Referenzvideo übertragen werden. Ich könnte mich selbst filmen und die Bewegung auf Kepler übertragen, ohne dass das Modell nach einem Gesicht sucht.

Zwischenfazit fal.ai Die API-Landschaft hat sich seit meinem ComfyUI-Experiment im November fundamental verändert. Statt lokal mit Face-Tools zu kämpfen, kann ich über fal.ai auf Modelle zugreifen, die das Face-Problem gar nicht erst haben:

1. **Depth/Canny ControlNets** umgehen Face-Bias, weil sie nur Geometrie sehen
2. **Era 3D** ermöglicht einen 2D→Multiview\$→\$2D-Loop für geometrische Konsistenz
3. **DreamActor v2** ist das erste Modell, das *explizit* nicht-menschliche Charaktere unterstützt
4. **FLUX Kontext Multi** kann mehrere Kepler-Referenzen gleichzeitig verarbeiten

Die Preise sind niedrig genug für systematisches Experimentieren: Ein Batch von 100 Depth-Control-Bildern kostet ~\$4. Das macht eine *agentic pipeline* realistisch — Claude Code könnte systematisch Prompts variieren, Referenzbilder kombinieren und die Ergebnisse evaluieren. Genau das Experiment 8 aus dem Roadmap, nur mit API statt lokalem ComfyUI.

Ausblick Mein nächstes Experiment wird ein Hybrid-Ansatz: TRELLIS 2 für 3D-Generierung von Kepler-ähnlichen Voxel-Formen, kombiniert mit dem ControlNet-Depth-Inpainting-Workflow für 2D-Konzeptbilder. Statt das Face-Problem zu lösen, umgehe ich es — indem ich Tools wähle, die gar nicht erst nach Gesichtern suchen.

Zusätzlich werde ich die fal.ai-Modelle systematisch testen — beginnend mit den günstigsten (Z-Image ControlNet, MiniMax Subject Reference, Era 3D) und aufsteigend zu den leistungsstärkeren (FLUX Kontext Multi, DreamActor v2). Jedes Modell wird mit demselben Kepler-Referenzset getestet: Casual-Pose, A-Pose und — falls Era 3D funktioniert — automatisch generierte Multiviews.

Ob das funktioniert, weiß ich nicht. Vielleicht scheitert TRELLIS 2 auf eine völlig andere Weise. Aber zum ersten Mal habe ich nicht nur eine Hypothese, sondern ein konkretes Arsenal an Tools, die ich testen kann — und eine API, die systematisches Experimentieren erschwinglich macht.

Eidesstattliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderweitigen fremden Äußerungen entnommen wurden, sind als solche einzeln kenntlich gemacht.

Der Einsatz generativer KI-Systeme (Claude, Cursor) bei der Erstellung dieser Arbeit ist im Abschnitt "Einsatz von KI in dieser Arbeit" vollständig offengelegt.

Detmold, Februar 2026

Jennifer Meier