



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 Musiker-Persona, werden vier Experimente und zwei Exkurse mit KI-Tools aus den Bereichen Bild- und Videogenerierung, Musikproduktion und Texterstellung dokumentiert. Die Analyse identifiziert vier zentrale Prozessveränderungen: Flow entsteht anfangs dort, wo er nicht geplant war; der agentische Workflow kann ihn später gezielt herstellen. Der Website-Build gelang produktiv, die Bild- und Videogenerierung scheiterte an Keplers Nischen-Ästhetik. KI beschleunigt Exploration und Prototyping, reduziert aber die Kontrolle über ästhetische Details. Die Autorschaft verschiebt sich von der Ausführenden zur Kuratorin und Orchestratorin. Zugleich erweist sich der emotionale Zyklus zwischen KI-Enthusiasmus und KI-Burnout als eigenständige Belastung im kreativen Prozess.

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

Inhaltsverzeichnis

Einleitung	5
Stand der Forschung	6
Artistic Research als Erkenntnisform	6
Technische Grundlagen generativer KI	6
Kreativität und Mensch-KI-Kollaboration	6
Artistic Research mit generativer KI	7
Forschungslücke	8
Methodologie	8
Methodischer Rahmen	8
Ethische Überlegungen und KI-Offenlegung	8
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	11
Experimente und Ergebnisse	11
Experiment 1: Visuelle Konsistenz (ComfyUI + Flux)	11
Experiment 2: Generierte Welten (Marble / World Labs)	11
Exkurs: Musik-Generierung (Suno AI)	12
Exkurs: KI als Lernpartner	12
Experiment 3: Automatisierte Bildpipeline (fal.ai API + Claude Code)	12
Experiment 4: Video-Generierung (fal.ai API + Claude Code)	26
Zentrale Beobachtungen aus dem Prozess	32
Unvorhersehbarer Flow	32
Speed vs. Control	32
Verschiebung der Autorschaft	33
KI-Burnout und KI-Enthusiasmus	34
Diskussion und Interpretation	34
Theoretische Implikationen	34
Praktische Implikationen	35

Stärken und Limitierungen	36
Practice-Based Intention, Practice-Led Realität	36
Abschließende autoethnografische Synthese: Die Maschine als Werkbank	37
Ausblick	37
Literaturverzeichnis	38
Abbildungsverzeichnis	39
Anhang A – Autoethnographische Interviews	40
Anhang B – Promptverzeichnis (Experiment 3: Automatisierte Bildpipeline)	52
Anhang C – Gesamtbewertung Bildpipeline (Experiment 3)	53
Anhang D – Gesamtbewertung Video-Generierung (Experiment 4)	54
Eidesstattliche Erklärung	55

Einleitung

Die vorliegende Arbeit versteht Artistic Research 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. Dass diese Werkzeuge nahezu gleichzeitig verfügbar wurden, ist kein Zufall, sondern Folge konvergierender Durchbrüche in der KI-Forschung zwischen 2017 und 2022. 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-Pixelform über eine 3D-Voxelfigur hin zu einer real gefilmten 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 mächtig und begrenzt zugleich – und genau diese Spannung macht sie zum Gegenstand künstlerischer Forschung. 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, sondern auch die Art, ihn zu dokumentieren: Die Website nahm mehr Raum ein als geplant, das KI-gestützte Reflexionsinterview etablierte sich als Methode, wurde dann aber als zu knapp kritisiert. Diese Verschiebungen bestätigten, dass die Prozessfrage die richtige war: Die Veränderung des kreativen Prozesses durch KI erwies sich als die eigentliche Erkenntnis. Zur Beantwortung dieser Frage verbindet die Arbeit:

- **Praktische Experimente:** Iterative Entwicklung von Kepler durch verschiedene KI-Tools, beobachtet auf Prozessveränderungen
- **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?
- **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.

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.

Technische Grundlagen generativer KI

Die aktuelle Welle generativer KI-Werkzeuge basiert auf drei konvergierenden Durchbrüchen. Die Transformer-Architektur von Vaswani et al. (2017) ersetzte 2017 rekurrente Netzwerke — also Modelle, die Daten Schritt für Schritt nacheinander verarbeiten — durch einen Attention-Mechanismus, der parallele Verarbeitung ermöglichte und die Skalierung auf Milliarden von Parametern erlaubte. Vereinfacht gesagt lernte das Modell damit, alle Teile eines Textes gleichzeitig zu betrachten statt Wort für Wort. Das war die technische Voraussetzung für GPT, DALL-E und die multimodalen Systeme, mit denen diese Arbeit operiert. Radford et al. (2021) trainierten 2021 mit CLIP ein Modell auf Hunderten Millionen von Bild-Text-Paaren aus dem Internet, das Texte und Bilder in einem gemeinsamen Bedeutungsraum verortet. Dadurch kann ein Modell verstehen, welches Bild zu welcher Beschreibung passt — die Brücke, die natürlichsprachliche Prompts für Bildgenerierungsmodelle interpretierbar macht. Rombach et al. (2022) verschoben 2022 den Diffusionsprozess in einen komprimierten latenten Raum: Statt ein Bild in voller Auflösung zu berechnen, arbeitet das Modell mit einer kleineren Zwischenrepräsentation und spart so enorm Rechenleistung. Das machte hochauflösende Bildgenerierung erstmals auf Consumer-Hardware möglich — Stable Diffusion, das direkte Fundament von ComfyUI, dem zentralen Tool in Experiment 1.

Kreativität und Mensch-KI-Kollaboration

Boden (2004) unterscheidet drei Formen von Kreativität. *Kombinatorische* Kreativität verbindet Bekanntes auf neue Weise. *Explorative* Kreativität lotet die Möglichkeiten innerhalb eines bestehenden Stils oder Regelsystems aus — sie probiert Variationen durch, ohne die Regeln selbst zu ändern. *Transformative* Kreativität bricht die Regeln: Sie verändert den Rahmen selbst und schafft etwas, das vorher undenkbar war. Diese Unterscheidung nutze ich im Folgenden als Analysewerkzeug: Sie erlaubt zu beschreiben, welche Art von Kreativität die KI zeigt und welche bei mir verbleibt.

Was ist generative Kunst? Galanter (2003) definiert sie als jede Praxis, bei der ein autonomes System zum Produktionsprozess beiträgt. KI-gestützte Kunstproduktion steht damit in einer langen Tradition. Der aktuelle Moment ist aber qualitativ anders: Multimodale KI-Systeme operieren in Text, Bild und Audio gleichzeitig, mit einer Geschwindigkeit und Zugänglichkeit, die es so noch nicht gab. 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 im Folgenden 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. Mir geht es nicht darum, ob die KI kreativ „ist“, sondern wie sich der kreative Prozess in der Kollaboration verändert.

Beim praktischen Arbeiten mit generativer KI tritt ein weiteres Phänomen auf: Das *Uncanny Valley* (Mori et al., 2012) beschreibt das Unbehagen, das entsteht, wenn künstliche Darstellungen menschenähnlich, aber nicht perfekt realistisch sind. Dieses Konzept wird in den Experimenten als konkretes Hindernis wiederkehren.

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) nutzen den ausgestorbenen King-Island-Zwergemu — ein Tier, von dem kaum verlässliche Bilder existieren — als Fallstudie für KI-gestützte Bildgenerierung. Konkret geben sie denselben Prompt an DALL-E 2 und Midjourney und vergleichen die Ergebnisse mit händischen Collagen aus historischen Quellen. Dabei zeigen sie, dass die Formulierung von Prompts selbst zum Denkakt wird: „Prompting as Thinking-With“, eine relationale Erkenntnismethode im Sinne Donna Haraways. Weil kein „richtiges“ Bild existiert, wird jede KI-Ausgabe zum Forschungsdatum — die substituierten Trainingsdaten-Muster offenbaren, was das Modell für „Emu“ hält.

Torres Núñez del Prado (2025) untersucht in „Encoding Culture“ die strukturelle Parallelie zwischen Datensatz-Kuration und Kanonbildung. Ihr Team hat Stable Diffusion 3.5 gezielt auf einem kuratierten Datensatz schwedischer Kunstgeschichte nachtrainiert — Fine-Tuning bedeutet, dass ein vortrainiertes Modell mit zusätzlichen, spezifischen Daten weiter spezialisiert wird. Die Arbeit zeigt, dass die Auswahl der Trainingsdaten dieselben Entscheidungen über Wertigkeit und Repräsentativität impliziert wie die Zusammenstellung eines kulturellen Kanons: Welche Werke fließen ein, welche nicht? Der Fokus liegt darauf, welche Ästhetiken als kanonisch eingebettet werden — und welche fehlen.

Hámošová und Rusnák (2023) trainieren in „TroublingGAN“ ein StyleGAN2-Modell auf Krisenfotografie (Reuters „Photos of the Week“ aus 2020). Die Heterogenität des Datensatzes verhindert strukturell Fotorealismus — das Ergebnis sind visuell ambige Bilder mit Schlieren, Blobs und Hohlräumen, wo Gesichter wären. Ihr zentrales Argument: GAN-Scheitern ist nicht Defekt, sondern erkenntnistheoretisch bedeutsam. Die Unfähigkeit des Modells zu konvergieren offenbart die Komplexität des Trainingsgegenstands.

Forschungslücke

Alle drei Projekte untersuchen jeweils ein Medium und verfolgen einen konzeptuell-künstlerischen Ansatz. Keines dokumentiert, ob sich Prozessmuster über Mediengrenzen hinweg wiederholen. Keines reflektiert systematisch, wie sich die Rolle der Künstlerin im Verlauf der Kollaboration verschiebt. Und keines fragt pragmatisch, wie sich KI-Tools in einen produktionsorientierten Workflow integrieren lassen. Hier setzt meine Arbeit an: Sie untersucht die Mensch-KI-Kollaboration autoethnografisch und multimodal aus der Perspektive einer Medienproduzentin. Kepler existiert in Bild, Audio, 3D und Code, und der Wechsel zwischen den Medien macht sichtbar, dass Muster wie der Speed-vs-Control-Tradeoff und die Rollenverschiebung zur Kuratorin medienübergreifend auftreten.

Methodologie

Methodischer Rahmen

Die Forschung verbindet drei Stränge: iterative Experimentierung, eine öffentliche Website als Forschungsartefakt und KI-gestütztes autoethnografisches Journaling. Der Rhythmus war pragmatisch: etwas ausprobieren, darüber schreiben, im Interview reflektieren, anpassen. 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.

Generative KI kam in dieser Arbeit auf vier Ebenen zum Einsatz: als Experimentgrundlage, als Reflexionspartner, als Schreibassistenz und als Entwicklungstool. Experiment 1 nutzte [ComfyUI](#), ein komplexes, node-basiertes Werkzeug, in dem Bildgenerierungs-Workflows visuell als Knotennetzwerke zusammengesteckt werden. Experiment 2 testete [Marble](#) von World Labs für 3D-Weltengenerierung, ein Exkurs erprobte [Suno AI](#) in der kostenlosen Stufe für Musikgenerierung. Für die automatisierte Bild- und Videogenerierung in Experiment 3 und 4 stieg ich auf die API (Application Programming Interface, eine Programmierschnittstelle) von [fal.ai](#) um, einem Anbieter, bei dem man Guthaben einzahlt und darüber verschiedene Generierungsmodelle ansprechen kann, ohne jedes einzeln installieren oder konfigurieren zu müssen. Nach jedem Experiment wurde eine strukturierte Reflexion im Dialog mit Claude durchgeführt (vollständige Transkripte auf der Projektwebsite sowie im Anhang). Claude unterstützte außerdem bei Strukturierung und Formulierung dieser Dokumentation. Die Interviews und ein Großteil der Instruktionen an Claude Code wurden per Spracheingabe über [Wispr Flow](#) diktiert, ein KI-gestütztes Diktionstool, dessen Spracherkennung durch ein eigenes Sprachmodell deutlich präziser arbeitet als herkömmliche Diktiersoftware und dadurch die

Reibung zwischen Gedanken und geschriebenem Text verringert.

Die Projektwebsite wurde zunächst mit [Cursor](#) gebaut, einer KI-gestützten Entwicklungsumgebung. Cursor ist ein Fork, eine eigenständige Abspaltung, von [Visual Studio Code](#) (VS Code), dem verbreiteten Open-Source-Editor von Microsoft. Im Laufe der Arbeit stieg ich auf [Claude Code](#) um, ein terminalbasiertes KI-Werkzeug von Anthropic, das direkt in der Kommandozeile arbeitet statt in einer grafischen Oberfläche und sich mit VS Code kombinieren lässt. Das Besondere an Cursor und Claude Code ist der Zugriff auf lokale Dateien und das Betriebssystem: Damit lassen sich Dokumente, Scripts und Code erstellen und ausführen, weit über die Fähigkeiten eines Chatbots hinaus. Die Arbeit erforderte auch eine finanzielle Investition: Claude (Max-Abonnement), Cursor (Lizenz) und fal.ai (Guthaben) waren kostenpflichtig.

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 mehr Bilder, mehr Videos, eine größere visuelle Welt. Kepler existierte bereits als musikalische Identität mit einem gewachsenen visuellen Fundament: 2D-Pixel-Art, verschiedene 3D-Modelle, zwei Alben (Blue Edition und Yellow Edition), Fotos mit echtem Kostüm und Maske. Ein Passion-Projekt ohne Budget, zusammengehalten von einer ästhetischen Intuition, die ich teilte mit Gavin, meinem musikalischen Kollaborateur. Was fehlte, war nicht die Identität, sondern die visuelle Konsistenz und Skalierbarkeit. 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.

Entscheidungen und Abzweigungen

Die Wahl des visuellen Stils

Kepler trägt eine Maske. Er hat kein sichtbares Gesicht. In seiner 3D-Form ist sein Stil voxelisiert, blockhaft, bewusst abstrahiert. Kepler ist ein alter Künstler, und die Maske verbirgt seine echte Identität — das ist keine ästhetische Spielerei, sondern konstitutiv: Die Maske ist das, was Kepler von einem gewöhnlichen Avatar unterscheidet. Diese Anonymität soll auch digital gewahrt bleiben.

Genau diese Designentscheidung wurde zur zentralen Herausforderung für die Arbeit mit generativer KI. Bildgenerierungsmodelle wie Flux oder Stable Diffusion sind auf Trainingsdaten aufgebaut, 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 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 von Grund auf mit Cursor zu bauen – einem KI-gestützten Code-Editor, statt auf ein fertiges CMS zurückzugreifen. 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.

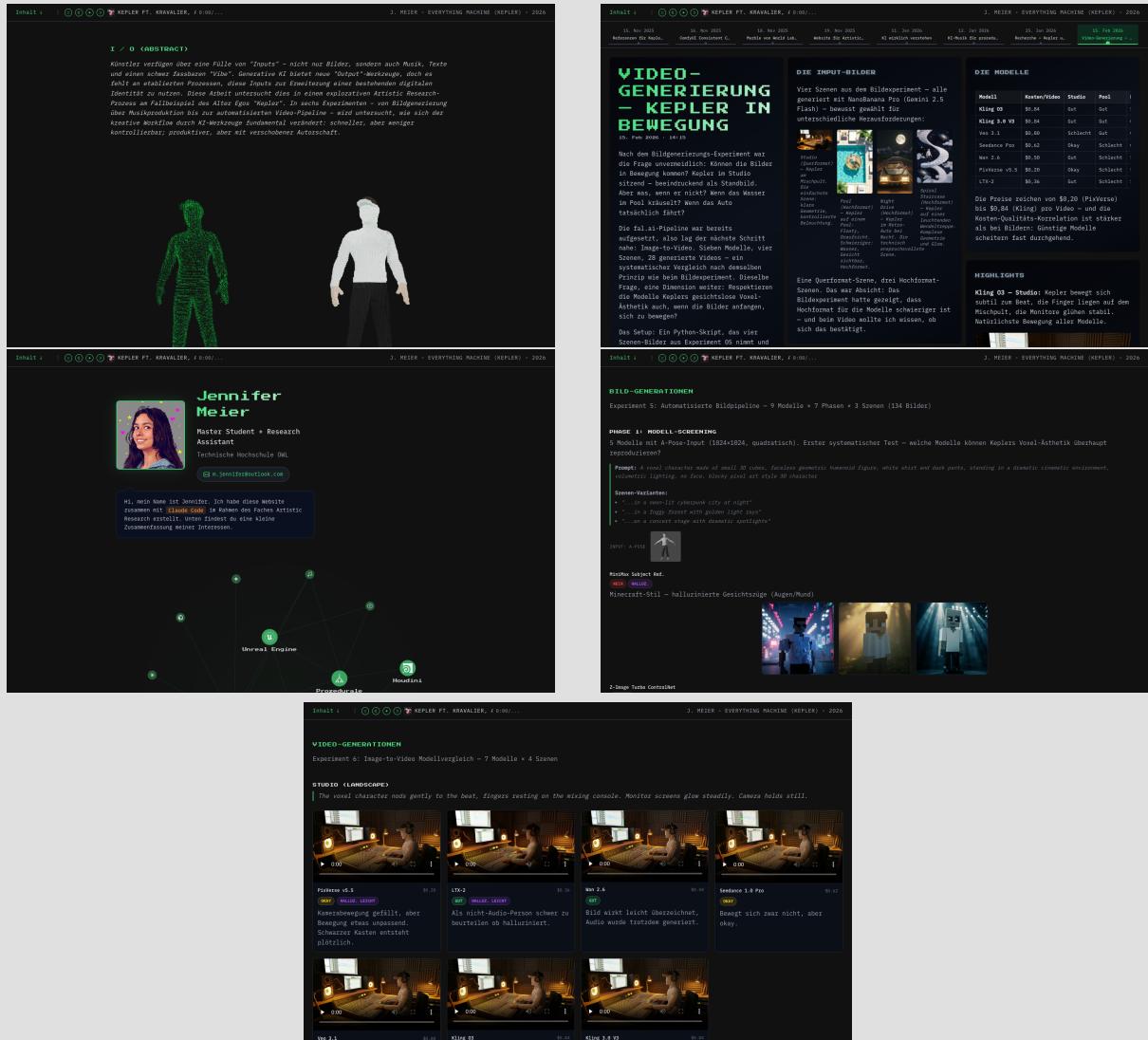


Abbildung 1: Die fünf Seiten der Website (<https://jenniferied.github.io/everything-machine/>): Projektübersicht mit 3D-Viewern, Prozess-Logbuch, Über mich mit Interessen-Netzwerk, Bild-Generationen und Video-Generationen.

Der Bauprozess mit Cursor war eine eigene Erfahrung. Ich fühlte mich, wie ich in meinem Journal notierte, „wie eine Managerin mit Gesamtvision, die einen Junior Developer dirigiert“ (Eintrag 4). „Garbage in, garbage out“, notierte ich — je spezifischer die Anforderungen, desto besser die Ergebnisse.

Von der Intention zum Experiment

Die Forschungsfrage funktionierte als Kompass für die Experimentwahl. Ich wollte verschiedene Modalitäten testen: Bild, 3D, Code, Audio. Weil Kepler multimodal existiert und weil ich verstehen wollte, ob die Muster der Kollaboration über Domänen hinweg stabil bleiben.

Experimente und Ergebnisse

Experiment 1: Visuelle Konsistenz (ComfyUI + Flux)

Mein erster ernsthafter Versuch, Kepler visuell konsistent mit KI zu generieren, nutzte ComfyUI, ein node-basiertes Werkzeug für Bildgenerierung, mit verschiedenen Erweiterungen für Konsistenz und Detailverbesserung. Die Installation auf meinem M1 MacBook Pro dauerte Stunden und verschlang viele Gigabyte, schon das ein Hinweis darauf, dass generative KI kein "Zaubertool" ist, als das sie 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. Die Halluzinationen kippen ins Chaotische — weg von der ästhetischen Balance, die Keplers Identität 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, die Parameter zu variieren, den Workflow anzupassen, neue Referenzbilder zu testen. Die Modelle sind auf Mainstream-Ästhetiken trainiert. Keplers Nische fällt durch das Raster. Torres Núñez del Prado (2025) beschreibt dieses Phänomen als strukturellen Bias: Die Auswahl der Trainingsdaten definiert, welche Ästhetiken das Modell reproduzieren kann — und Nischen, für die es weniger Beispiele gibt, werden in Richtung Mainstream gezogen.

Experiment 2: Generierte Welten (Marble / World Labs)

Nachdem Kepler als Charakter in der Bildgenerierung an Grenzen gestoßen war, verlagerte ich den Fokus auf Hintergrundwelten. Ich testete Marble (World Labs), ein browserbasiertes Tool, das 3D-Welten aus Gaussian Splats generiert. 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 geplante Workflow. Monin und Sadokierski (2025) beschreiben genau diese Dynamik als „Prompting as Thinking-With“: Die Formulierung

des Prompts ist nicht bloß Instruktion, sondern ein eigenständiger Denkprozess, bei dem die Aus-einandersetzung mit der KI das Verständnis des Gegenstands vertieft.

Bei Personen und Spiegelungen zeigte sich erneut das Uncanny Valley: verzerrte Proportionen, gruselige Doppelgänger. In der Reflexion entstand die Idee, genau diese "Fehler" als ästhetisches Mittel einzusetzen: „Man könnte tatsächlich wirklich damit visuell spielen. [...] Ich könnte auch 3D gerenderte Bilder in einen Gaussian Splat Generator eingeben“ (Eintrag 3). Diese Perspektive – Scheitern nicht als Defekt, sondern als Ausgangspunkt – zog sich durch alle folgenden Experimente.

Exkurs: Musik-Generierung (Suno AI)

Parallel zu den visuellen Experimenten testete ich Suno AI für die Generierung von Musik (minimal, loopbar, sowjetisch-dystopisch). Dieser Exkurs gehörte nicht direkt zur Kepler-Bildentwicklung, gab aber Aufschluss über ein Muster, das sich auch in den visuellen Experimenten zeigte.

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 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. 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.

Exkurs: KI als Lernpartner

Auf einer 30-minütigen Autofahrt führte ich ein Voice-Gespräch mit Claude über KI-Grundlagen: von Shannons Informationstheorie über die Dartmouth Conference 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. „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 anschau“ (Eintrag 5). Claude selbst wies mich darauf hin, dass es einen Unterschied gibt zwischen deklarativem Wissen ("wissen, dass") und prozeduralem Wissen ("wissen, wie"). Die Theorie wird erst real, wenn ich sie anwende. Dieses Phänomen ist relevant für jeden, der KI als Lernwerkzeug einsetzt.

Experiment 3: Automatisierte Bildpipeline (Fol.ai API + Claude Code)

An dieser Stelle nimmt die Autoethnografie einen sehr technischen Charakter an. Der Bau einer Pipeline aus APIs und Skripten war kein kreativer Befreiungsschlag im klassischen Sinne, sondern entsprang der schieren Frustration über die manuelle Bedienung von Chatbots und node-basierter Software. Für pipeline-basierte Kontrolle über Nischen-Ästhetiken muss man vorübergehend zum Engineer werden. Doch genau in dieser technischen Orchestrierung, dem gemeinsamen Debuggen und Iterieren mit Claude Code, entstand ein unerwarteter Flow. Ob das Errichten einer solchen Architektur noch klassische "künstlerische Praxis" ist oder reines Engineering, trat im Prozess in den Hintergrund: Es war der pragmatische Weg, um in einem extrem schnelllebigen Technologie-

Umfeld handlungsfähig zu bleiben und die Skalierung der Experimente überhaupt erst zu ermöglichen.

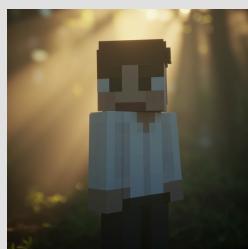
Nach den Einzelexperimenten und Exkursen stellte sich eine methodische Frage: Was passiert, wenn ich den Prozess systematisiere? Gemeinsam mit Claude Code baute ich eine automatisierte Pipeline, die über die fal.ai API neun verschiedene Bildgenerierungsmodelle mit denselben Prompts und Referenzbildern 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. Die folgenden Seiten dokumentieren die einzelnen Phasen.

Phase 1: Modell-Screening (5 Modelle, A-Pose) Als Referenzbild diente Keplers A-Pose-Render (ein quadratischer 1024×1024 Render mit seitlich ausgestreckten Armen). Fünf Modelle wurden getestet:



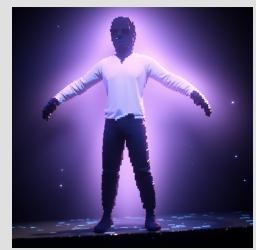
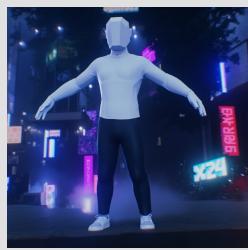
Abbildung 2: Input: Kepler A-Pose (1024×1024)



MiniMax Subject Reference



Z-Image Turbo ControlNet



FLUX Depth LoRA



FLUX Canny LoRA

Abbildung 3: Phase 1: Modell-Screening – MiniMax, Z-Image, FLUX Depth, FLUX Canny. Bilder in der Bild-Galerie vergrößerbar.

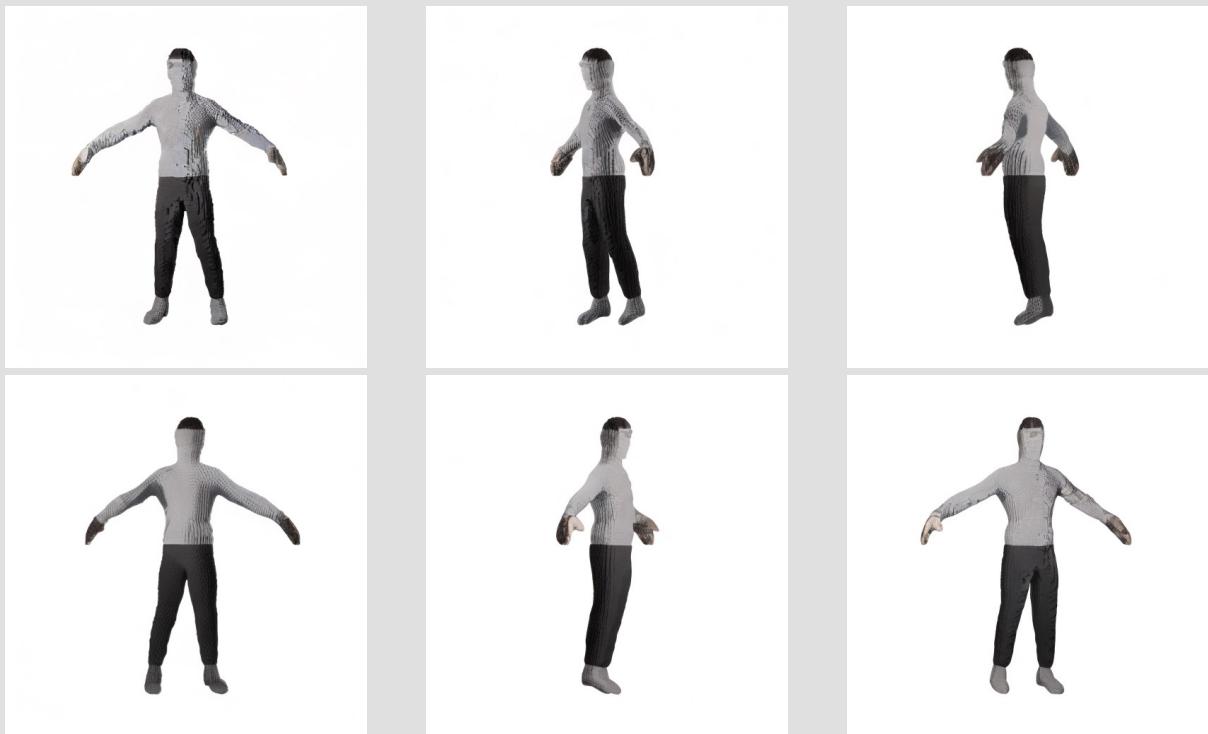


Abbildung 4: Phase 1: Era3D Multiview – nutzlos im Vergleich zu eigenen Renders. Bild-Galerie

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

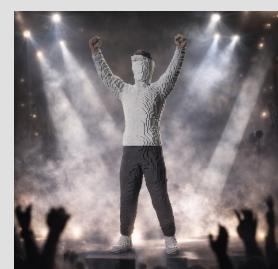
Verworfene Modelle. Keines der fünf Modelle lieferte überzeugende Ergebnisse. MiniMax Subject Reference interpretierte Keplers Voxel-Ästhetik als Minecraft-Welt und halluzinierte menschliche Gesichtszüge, obwohl der Prompt “faceless” explizit vorgab. FLUX Depth LoRA bewahrte zwar die Körperhaltung über 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 (FLUX Kontext Pro, GPT-Image-1.5, SeedDream und Nano Banana Pro), die sich in Benchmarks als State-of-the-Art für Charakter-Konsistenz erwiesen hatten.

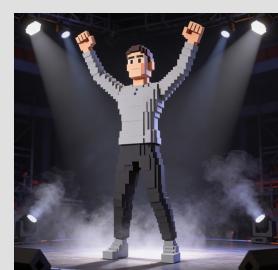
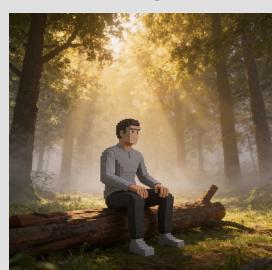
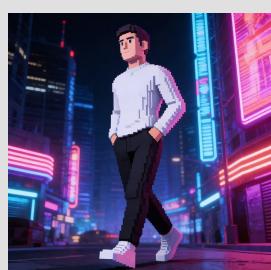
Phase 2: Top-Modelle in Szenen (A-Pose) Die vier vielversprechendsten Modelle wurden in Szenen-Kontexte eingeführt, jeweils drei Varianten mit demselben A-Pose-Input.



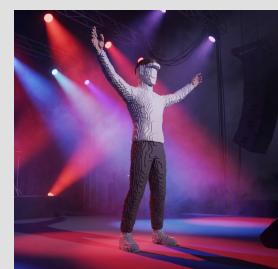
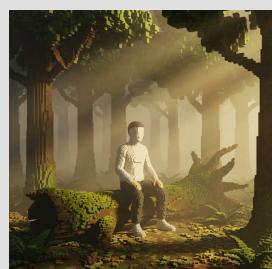
FLUX Kontext Pro



GPT-Image-1.5



SeedDream v4.5



Nano Banana Pro

Abbildung 5: 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
Nano Banana Pro	gut	teilweise	Beste Phase-2-Ergebnisse

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

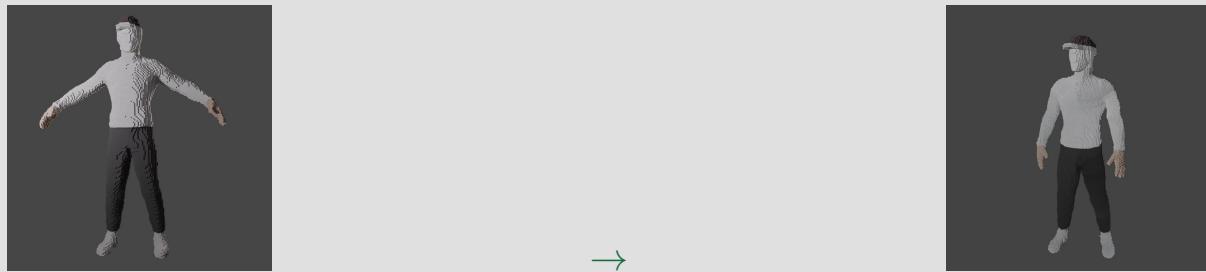


Abbildung 6: Input-Wechsel: A-Pose → Posed – natürliche Haltung statt T-Pose

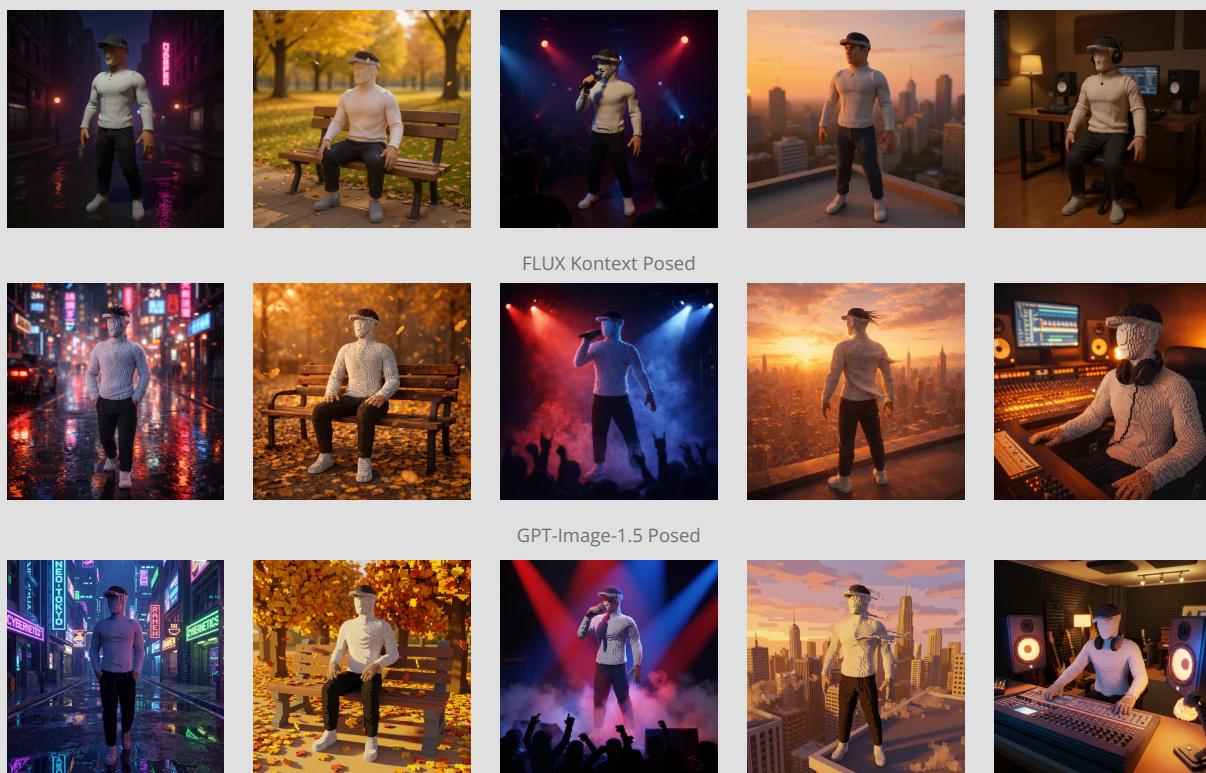


Abbildung 7: Phase 2b: Drei Finalisten mit Posed Input (je 5 Varianten). Bild-Galerie

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
Nano Banana Posed	ja	nein	Stärkste Voxel-Ästhetik

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

Phase 3a: Signature Scenes v1 – Ratio-Bug Ab Phase 3a wurden GPT-Image-1.5 und Nano Banana Pro parallel in "Signature Scenes" getestet: Studio, Retro Car Bridge, Spiral Staircase, Pool Floaty. Ein Ratio-Bug in der Pipeline betraf GPT: Alle Outputs kamen quadratisch statt im gewünschten Format. Nano Banana Pro lieferte korrekte Portrait- und Landscape-Formate.



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



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



Nano Banana Portrait (14) – korrektes Format



Nano Banana Landscape (14) – korrektes Format

Abbildung 8: Phase 3a: GPT-1.5 mit Ratio-Bug (oben, quadratisch), Nano Banana 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
Nano Banana	portrait	gemischt	teilweise	Halb cool, halb Halluz.
Nano Banana	landscape	gemischt	teilweise	Ohne Halluz. stark



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

Abbildung 9: 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	Nano Banana	portrait	gemischt	nein	Stretching, keine Halluz.
15/16	Nano Banana	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	Nano Banana	portrait	gemischt	-	-
17/18	Nano Banana	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 Nano Banana 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: leuchtende 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. Nano Banana Pro hingegen lieferte korrekte Formate und die stärksten Landscape-Ergebnisse der gesamten Pipeline. Die Phasen 3a-3b zeigen eine wiederkehrende Dynamik: 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.



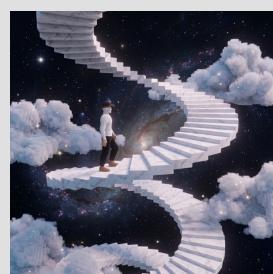
GPT-1.5 Portrait (15)



Nano Banana Portrait (16)



GPT-1.5 Portrait (17)



Nano Banana Landscape (18)

Abbildung 10: Phase 3b: Outputs mit gestrechten Referenzbildern. [Bild-Galerie](#)



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

Abbildung 11: 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	-	-
Nano Banana	portrait	gemischt	-	-
Nano Banana	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 (siehe Anhang B).

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. Nano Banana 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)



Nano Banana Portrait (20)



GPT-1.5 Landscape (19)



Nano Banana Landscape (20)

Abbildung 12: 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 13: 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. Nano Banana Pro war konsistenter, aber weniger ausdrucksstark.

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

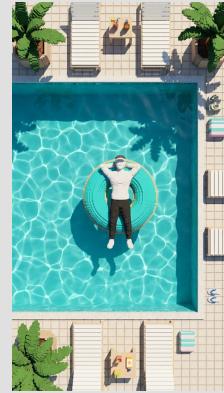
Ordner	Modell	Input	Format	Szene	Bew.	Halluz.
21	GPT-1.5	posed	portrait	night_drive	ja	ja
21	GPT-1.5	posed	portrait	staircase	ja	nein
21	GPT-1.5	posed	portrait	pool	ja	nein
21	GPT-1.5	posed	landscape	night_drive	ja	ja
21	GPT-1.5	posed	landscape	staircase	ja	nein
21	GPT-1.5	posed	landscape	pool	neutral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



GPT-1.5 Posed Portrait (21)



GPT-1.5 A-Pose Portrait (23)



Nano Banana Posed Portrait (22)



Nano Banana A-Pose Portrait (24)

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



Nano Banana Posed Landscape (22)

Nano Banana A-Pose Landscape (23)

GPT-1.5 Posed Landscape (21)

GPT-1.5 A-Pose Landscape (23)

GPT-1.5 Posed Landscape (21)

GPT-1.5 A-Pose Landscape (23)

Abbildung 15: Phase 3d: Landscape-Vergleich – Posed (21/22) vs. A-Pose (23/24). [Bild-Galerie](#)

Analyse und Reflexion Dieses Experiment markiert einen Rollenwechsel: Ich definierte das "Was" und "Warum", während Claude Code das "Wie" implementierte. Sieben Phasen steckten den Möglichkeitsraum zwischen neun Modellen, zwei Input-Typen und drei Szenen systematisch ab. Wo ich in Experiment 1 stundenlang an ComfyUI-Parametern drehte, durchlief die Pipeline denselben Prozess in Minuten.

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. Eine neue Frage drängte sich auf: Wenn ein KI-Agent für mich promptet – wo endet dann meine Autorschaft?

Im Reflexionsinterview wurde mir der zentrale Shift bewusst: Der Workflow wechselte vom Überredungs-Problem zum Automatisierungs-Problem. In Experiment 1 versuchte ich stundenlang, ein einzelnes Modell dazu zu bringen, Keplers Maske zu respektieren. In Experiment 3 verglich ich neun Modelle systematisch miteinander. „Ich war technische Problemlöserin. Instruktionsbefolgerin. Claude Code hat mir detaillierte Instruktionen gegeben, wie ich das Ganze aufsetzen soll“ (Eintrag 7). Dabei fiel mir etwas auf, das ich vorher nicht so klar gesehen hatte: Die Autorschaft war in diesem Projekt bereits vor dem KI-Einsatz verteilt. Keplers 3D-Modell stammt von Gavin und anderen Artists, nicht von mir. Die systematische Bildpipeline funktionierte; die Frage war nun, ob sich der Ansatz auf andere Medien übertragen ließ. Die Gesamtbewertung aller 52 Einzelergebnisse ist in Anhang C dokumentiert, die verwendeten Prompts in Anhang B.

Experiment 4: Video-Generierung (Fol.ai API + Claude Code)

Die automatisierte Bildpipeline aus Experiment 3 warf eine naheliegende Frage auf: Wenn statische Bilder von Kepler überzeugend gelingen, lässt sich derselbe Ansatz dann auf Bewegtbild übertragen? Image-to-Video-Modelle versprechen, aus Einzelbildern kurze Clips zu generieren. Wo Bildgenerierung schon mit Halluzinationen kämpft, potenziert Video das Problem: 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** (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, Fahrbewegung
- **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." (siehe Ergebnisse). 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 16: Mittlere Frames, Studio-Szene (Landscape, 7 Modelle). Verlinkt 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.



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, Pool-Szene (Portrait, 7 Modelle). Verlinkt 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.
Veo 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 18: Mittlere Frames, Night-Drive-Szene (Portrait, 7 Modelle). Verlinkt 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 19: Mittlere Frames, Spiral-Staircase-Szene (Portrait, 7 Modelle). Verlinkt 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
Veo 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

Die vollständigen Einzelbewertungen aller 28 Videos sind in Anhang D dokumentiert.

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 sind offenbar auf menschliche Gesichter als Default trainiert 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. Hámóšová und Rusnák (2023) zeigen in ihrem Projekt *TroublingGAN*, dass das Scheitern generativer Modelle ästhetisch produktiv sein kann – visuelle Ambiguität als bewusstes Gestaltungsmittel statt als Fehler. Keplers Gesichts-Halluzinationen sind ein konkreter Fall dieser Spannung: Was das Modell als Defizit behandelt (die Abwesenheit eines Gesichts), ist für die Künstleridentität konstitutiv.

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. Veo 3.1 scheiterte beim Studio, glänzte aber beim Pool. Kein Modell war durchgehend überlegen; die beste Wahl hing von der Szene ab.

Methodisch war dieses Experiment eine Erweiterung von Experiment 3: 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.

Meine Rolle beschrieb ich in der Reflexion als „neugierige Kuratorin“ (Eintrag 8). Das systematische Bewerten über Medientypen hinweg machte Spaß, und der agentische Workflow gab mir die Zuversicht, dass brauchbare Ergebnisse wiederholbar erreichbar sind. Aber diese technisch brauchbaren Ergebnisse sind der Startpunkt, nicht das Ziel. Die eigentliche kreative Iteration mit dem Künstler steht noch bevor: Gavins ästhetische Vision einbringen, mit Pinterest-Referenzen den Stil schärfen, das Prompting für Stil-Experimente nutzen.

Zentrale Beobachtungen aus dem Prozess

Am Anfang stand eine praktische Frage: Kann generative KI die visuelle Konsistenz liefern, die Kepler braucht? Die Antwort: nicht zuverlässig. Bildgenerierungsmodelle scheitern an Keplers Nischen-Ästhetik, halluzinieren Gesichter in seine Maske, driften in Mainstream-Ästhetiken ab. Aber das Scheitern brachte Erkenntnisse. Aus der vergleichenden Analyse der Journal-Einträge und der 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 beantworten zugleich die fünf Subfragen, die die strukturierten Reflexionsinterviews geleitet haben.

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 nicht nur repetitive, sondern auch technisch anspruchsvolle Aufgaben, die meine vorhandenen Web-Kenntnisse (HTML, CSS, JavaScript) überstiegen, 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. 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 wechselseitig 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.

Experiment 3 und 4 zeigten allerdings eine Gegenbewegung: Der agentische Workflow mit Claude Code und fal.ai stellte Flow gezielt her. „Ich sage der KI einfach: Recherchiere die neuesten Modelle und generiere mir fünf Beispielbilder, und dann habe ich ein Resultat“ (Eintrag 8). Der reproduzierbare Workflow war die direkte Antwort auf die Frustration der Einzelversuche.

KI-Tools verschieben den Workflow von linearer Ausführung zu iterativer Navigation: explorativer, weniger planbar, stärker abhängig von der Fähigkeit zu kuratieren statt zu produzieren.

Speed vs. Control

KI beschleunigt und begrenzt den kreativen Prozess zugleich. Exploration und Prototyping werden 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 beschreibt, was hier passiert: Das Modell bewegt sich innerhalb eines Stils und probiert Variationen durch. Im Suno-Experiment — einem Exkurs in die KI-Musikgenerierung, der nicht direkt zur Kepler-Entwicklung gehörte — konnte das Modell den gewünschten Stil sprachlich präzise beschreiben, ihn aber klanglich nicht umsetzen. Es exploringte den Raum der Möglichkeiten.

ten auf der Beschreibungsebene, scheiterte aber an der Übersetzung in tatsächlichen Sound. Die Diskrepanz zwischen Beschreibung und Output zeigt den Speed-vs-Control-Tradeoff konkret: Generierung ist schnell, aber Feinsteuierung ist nicht gegeben. Die automatisierte Pipeline aus Experiment 3 illustriert beide Seiten: 134 Bilder in 52 Einzelbewertungen entstanden in Minuten, aber Details wie Keplers Cap oder die Cockpit-Perspektive blieben unkontrollierbar.

Das Tempo selbst wurde dabei zum Problem. „Das Tempo verleitet zur Oberflächlichkeit“, reflektierte ich nach dem Website-Build (Eintrag 4). Die KI-Ausgabe war sofort da, und manchmal akzeptierte ich zu schnell, ohne den Code richtig zu verstehen. In Experiment 3 führte eine ungenaue Instruktion dazu, dass die Referenzbilder gestreckt statt erweitert wurden. Der Fehler war vermeidbar, kostete aber Geld und Zeit, weil ich zu schnell die Instruktion geben und zu schnell das Ergebnis haben wollte (Eintrag 7).

Iteration wird in diesem Kontext zur Erkenntnismethode: Jeder gescheiterte Prompt, jede unerwartete Halluzination wurde zum Ausgangspunkt für reflektierte Anpassung. Die Ästhetik entsteht dabei durch die Einschränkungen der KI, nicht trotz ihnen: Die Auseinandersetzung mit Halluzinationen und Mainstream-Drift schärfe das ästhetische Urteil.

Verschiebung der Autorschaft

Über die Experimente hinweg verschob sich meine Rolle ständig. Beim Referenzen-Sammeln war ich Kuratorin — eine Tätigkeit, die ganz ohne KI auskam und auf ästhetischem Urteilsvermögen beruhte. Bei ComfyUI war ich Lernende und Experimentierende, orientiert am Workflow einer anderen Person. Bei Marble war ich „einfach nur jemand, der experimentiert und ein Produkt ausprobiert“ (Eintrag 3) – die flachste Autorschaftsrolle im ganzen Projekt. 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 bewegen sich auf einem Spektrum — mal ganz ohne KI (Referenzen sammeln), mal in der Mixed-Initiative-Zone (Website-Build), mal näher am „Computer als Schöpfer“ (Suno). Diese Rollenvielfalt ist kein Zufall, sondern eine Qualität des Prozesses. Die wechselnden Rollen sind Ausdruck einer Autorschaft, die sich verschoben hat: nicht mehr alleinige Urheberschaft, nicht vollständige Delegation. In Experiment 3 kam eine weitere Rolle hinzu: die der Orchestratorin, die einen KI-Agenten beauftragt, der seinerseits Bildmodelle steuert. Eine dreistufige Kette, die die Frage nach der Autorschaft weiter verkompliziert.

Im Video-Experiment vertiefte sich die Kuratorinnen-Rolle. Das systematische Bewerten machte Spaß, die Spannung zwischen quantitativen und qualitativen Ergebnissen trieb den Prozess (Eintrag 8). Aber das Reflexionsinterview nach Experiment 3 stellte eine grundlegendere Frage: War die Autorschaft in diesem Projekt jemals ungeteilt? Keplers Grunddesign stammt von Gavin und anderen 3D-Artists. Meine Rolle war von Anfang an die der Modifiziererin, nicht der Schöpferin. „Ich fühle mich nicht wie eine Künstlerin. Macht mir aber nichts aus“ (Eintrag 7). Die KI hat die verteilte Autorschaft nicht erzeugt, sondern sichtbar gemacht. Der Kreativprozess begann schon beim Kuratieren der Pinterest-Referenzen mit Gavin: „Die visuelle Sprache entsteht hier durch Kuration, nicht durch Generierung“ (Eintrag 1). Die Rollenverschiebung über die Experimente hinweg verläuft nicht linear. Von der Kuratorin über die Lernende und Dirigentin zur Orchestratorin und schließlich zur „Instruktionsbefolgerin“ (Eintrag 7) — gemeint ist die Erfahrung, in Experiment 3 detaillierte technische Anweisungen von Claude Code zu erhalten und sie auszuführen, statt selbst

die Richtung vorzugeben. Welche Rolle ich einnehme, hängt vom Kontext ab, nicht vom Fortschritt.

KI-Burnout und KI-Enthusiasmus

Die emotionale Dimension der KI-Arbeit war eine Prozessveränderung, die ich nicht erwartet hatte. Es gibt einen Zyklus: Begeisterung, wenn etwas unerwartet gut funktioniert – der Website-Build, 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 eine 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 aufgerissen 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 3 wirkte dem entgegen: Sieben strukturierte Phasen mit klaren Evaluationskriterien reduzierten die Frustration gegenüber den Einzelversuchen erheblich. Die Wirkung hielt bis ins Video-Experiment an: „Es hat mir Confidence gegeben, dass KI auch coole, schnelle Ergebnisse liefern kann“ (Eintrag 8). Es war „erleichternd und cool, mal interessante Ergebnisse zu kriegen“ (Eintrag 7). Der Pipeline-Ansatz war nicht nur methodisch effizient, sondern auch emotional stabilisierend. Der Umgang mit Scheitern ist dabei auch eine psychologische Herausforderung. ComfyUI scheiterte an Keplers Gesicht, Suno an der gewünschten Klangästhetik – beide Erfahrungen schärften das Verständnis dafür, wo die Grenzen aktueller Modelle liegen.

Diskussion und Interpretation

Die produktivsten Momente entstanden dort, wo Intention und die Logik der Modelle in Spannung zueinander traten, nicht wo sie sich deckten.

Theoretische Implikationen

Bestehende Kreativitätstheorien lassen sich im Kontext der Mensch-KI-Kollaboration produktiv anwenden und differenzieren. Bodens (2004) Unterscheidung zwischen explorativer und transformativer Kreativität erweist sich hier als nützlich. In meiner Praxis übernahm die KI durchgehend die explorative Rolle: Sie generierte Variationen, probierte Stile durch, lieferte schnell viele Optionen. Aber die Entscheidung, welche Option weiterverfolgt wird, welche Richtung stimmt und wann ein Ergebnis die bisherigen Grenzen sprengt, das lag bei mir. Transformative Impulse gingen in keinem der Experimente von den Modellen aus. Sie lagen bei meinen Entscheidungen. Die KI liefert Möglichkeiten; der Mensch entscheidet, welche davon neu ist. Der unvorhersehbare Flow folgt aus dieser Asymmetrie: Die KI exploriert schnell, aber ob das Ergebnis wirklich etwas verändert, erfordert menschliches Urteil.

Galanters (2003) Perspektive ergänzt diesen Befund: Generative Kunst ist als Methode zu verstehen, relevant ist das *Wie* der Produktion. Der Speed-vs-Control-Tradeoff lässt sich als Navigation auf seiner Ordnung-Chaos-Skala beschreiben: KI-Tools beschleunigen die Exploration des konzeptuellen Raums, aber die Kontrolle über die Zone effektiver Komplexität bleibt beim Menschen. Was Borgdorff (2012) über künstlerische Forschung argumentiert, bestätigt sich hier: Mein Vorwissen

über Keplers Ästhetik (die dystopische Klangwelt, die voxelisierte Maskierung, die Spannung zwischen Lo-Fi und Hyperdigitalem) ließ sich als Prompt-Anweisung an KI-Systeme nur begrenzt übertragen. Ohne Domänenwissen produziert die Kollaboration generische Ergebnisse. Dieses Wissen zeigte sich erst *im* iterativen Prozess, nicht vor ihm.

Dieser iterative Prozess folgt dem, was Schön (1983) Reflection-in-Action nennt: das spontane Reagieren auf unerwartete Situationen in der Praxis. Das Scheitern eines Prompts war Auslöser für reflektierte Anpassung. Bei Schön spricht physisches Material zurück: Ton, der reißt, Farbe, die verläuft. In der KI-Kollaboration antwortet ein probabilistisches System mit eigener Logik. 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 trifft meine vierte 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 bietet keine Eingriffsmöglichkeit) oder zu abhängig von Expertenwissen (ComfyUI setzt hohe technische Hürden). Die produktivsten Momente meiner Forschung entstanden dort, wo die Balance gelang: beim Website-Build, wo die KI nicht nur repetitive, sondern auch technisch anspruchsvolle Aufgaben übernahm und ich die ästhetische Vision steuerte.

Die Erfahrungen vergleichbarer Projekte bestätigen und differenzieren diese Befunde. Monin und Sadokierski (2025) zeigen, dass Prompting als „Thinking-With“ funktioniert, und in meiner Praxis bestätigt sich das: Die produktivsten Prompts waren nicht präzise Instruktionen, sondern exploratives Denken mit der Maschine (vgl. Experiment 2). Dass Keplers Nischen-Ästhetik systematisch in Richtung Mainstream gezogen wird, erklärt Torres Núñez del Prado (2025) strukturell: Trainingsdaten-Kuration gleicht einer Kanonbildung, und die Modelle reproduzieren den ästhetischen Kanon, auf dem sie trainiert wurden. Hámošová und Rusnák (2023) zeigen, dass KI-Scheitern ästhetisch produktiv sein kann. In meinem Fall stößt diese Perspektive allerdings an ihre Grenzen. Hámošová kann als Künstlerin die visuellen Artefakte des GAN-Scheiterns bewusst als ästhetisches Element einsetzen. In meiner Praxis war das Gesichts-Halluzinieren schlicht ein Hindernis: Kepler braucht eine intakte Maske, und ein halluziniertes Gesicht zerstört seine Identität. Dieses Scheitern als Feature umzudeuten war keine Option, es musste funktionieren.

Praktische Implikationen

Für Künstler*innen und Medienproduzent*innen, die generative KI in ihre Praxis integrieren wollen, ergeben sich aus dieser Arbeit konkrete Erkenntnisse. Für ästhetisch spezifische Ergebnisse jenseits des Mainstream ist Domänenwissen Voraussetzung. Wer ohne klare ästhetische Vision promptet, erhält den kleinsten gemeinsamen Nenner: die Mainstream-Ästhetik, auf die Generalist-Modelle trainiert sind. Die Modelle zeigen eine Regression zur ästhetischen Mitte der Trainingsdaten. Nischen-Ästhetiken, für die es weniger Trainingsdaten gibt, werden entweder in Richtung Mainstream gezogen oder scheitern ganz. Keplers Nischen-Ästhetik ließ sich nur dort realisieren, wo ich spezifisches Vokabular, präzise Referenzen und ein geschultes Urteilsvermögen einbrachte.

KI-Werkzeuge eignen sich für Ideation, Prototyping und Exploration, aber die finale ästhetische Entscheidung verbleibt beim Menschen. Ein pragmatischer Ansatz zur Bewältigung dieses Speed-vs-Control-Tradeoffs erwies sich als die Kombination aus Claude Code als Orchestrator und der fal.ai API als Modell-Backend: ein reproduzierbarer Workflow für systematisches Modell-Screening, der

den explorativen Raum effizient absteckt und zugleich die Frustration der Einzelversuche reduziert.

Stärken und Limitierungen

Der Practice-Led Ansatz mit autoethnografischem Journaling hat erlaubt, den Prozess in einer Tiefe festzuhalten, die von außen nicht möglich wäre. Dass alles auf der Website öffentlich liegt (auch die Fehlversuche), macht die Ergebnisse nachprüfbar. Die multimodale Perspektive (Bild, Audio, Code, Text) bildet die Komplexität zeitgenössischer Medienproduktion ab, anstatt sie auf ein einzelnes Medium zu reduzieren.

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. 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: 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. Candy (2006) unterscheidet die beiden Ansätze entlang einer zentralen Frage: Liegt der neue Wissensbeitrag *in* den Artefakten oder *über* die Praxis? In einer practice-based Arbeit verkörpert das Kunstwerk selbst neues Wissen, das sich nicht vollständig in Textform übersetzen lässt. In einer practice-led Arbeit generiert die Praxis Erkenntnisse, die sich als propositionales Wissen formulieren lassen. Im Verlauf hat sich meine Arbeit als practice-led erwiesen – aus drei Gründen.

Erstens die Wissensform: Die vier Beobachtungen über Prozessveränderungen (unvorhersehbarer Flow, Speed-vs-Control-Tradeoff, Rollenverschiebung zur Kuratorin, Burnout-Zyklus) lassen sich vollständig in Textform vermitteln. Die generierten Bilder und Clips dienen als Belege für diese Erkenntnisse, nicht als eigenständige Wissensträger. In einer practice-based Arbeit müssten die Artefakte selbst etwas zeigen, das der Text nicht kann. Hier ist es umgekehrt: Der Text erklärt, was die Artefakte allein nicht sichtbar machen – nämlich die Prozessveränderungen, die zu ihnen geführt haben.

Hinzu kommt die Fragestellung. „Wie verändert der Einsatz multimodaler KI-Systeme den kreativen Prozess?“ fragt nach dem *Wie* der Veränderung, nicht nach dem *Was* des Ergebnisses. Eine practice-based Arbeit hätte eine andere Frage gebraucht – etwa: „Welche neuen ästhetischen Formen entstehen durch die Kollaboration mit KI?“ Meine Frage zielt auf Prozesswissen, und Prozesswissen ist propositional: Es lässt sich benennen, strukturieren, diskutieren. Die Methode folgt der Frage, nicht umgekehrt.

Schließlich die Artefakte selbst: Die KI-generierten Bilder fühlen sich wie Zwischenergebnisse an, ästhetisch teils interessant, aber nicht „meine Kunst“. Wenn ich die Artefakte selbst nicht als eigenständige Werke betrachte, in denen sich unausgesprochenes Wissen verkörpert, fehlt die Grundlage für eine practice-based Einordnung. Paradoxe Weise war die Website das Artefakt mit dem meisten Flow. Obwohl als Dokumentationstool konzipiert, wurde sie zum vielleicht künstlerischs-

ten Element des Projekts. Aber auch hier liegt der Erkenntnisbeitrag in der Beobachtung, *dass* ausgerechnet das Dokumentationstool zum kreativsten Moment wurde. Die Erkenntnis ist über den Prozess, nicht im Artefakt.

Diese Verschiebung von practice-based zu practice-led ist kein Scheitern der ursprünglichen Intention, sondern selbst ein Ergebnis der Forschung. Sie spiegelt das Spannungsfeld, in dem sich viele Kreative befinden, die zum ersten Mal mit generativer KI arbeiten: Die Unsicherheit, ob das hier Kunst oder Technikerprobung ist, lässt sich nicht vorab klären. Sie zeigt sich erst im Prozess.

Abschließende autoethnografische Synthese: Die Maschine als Werkbank

Die Automatisierung des Prozesses hat mich der Figur Kepler weder nähergebracht noch weiter von ihr entfremdet. Da das grundlegende 3D-Design nicht meine alleinige Kreation ist, betrachte ich die massenhaft generierten Bilder nicht als emotional aufgeladenes Endprodukt, sondern als erarbeitetes Fundament. Der Prozess war letztlich der Prolog: Erst jetzt, da die Pipeline steht und die Grenzen der Modelle verstanden sind, kann die eigentliche Arbeit beginnen. Die nächsten Schritte führen zurück zu mehr Kontrolle – sei es das Extrahieren spezifischer Stile aus unseren Pinterest-Boards, die gezielte Steuerung durch ControlNet oder das Erstellen von groben Block-outs in einer Game Engine, meiner eigentlichen technologischen Heimat.

Der im Prozess identifizierte emotionale Zyklus aus KI-Burnout und Erschöpfung hat sich durch den automatisierten Workflow vorerst wieder in Richtung KI-Enthusiasmus und Neugier verschoben. Diese ständige Oszillation scheint der Normalzustand in der Arbeit mit einer Technologie zu sein, die nie stillsteht. Als jemand mit starkem technischen Hintergrund fühle ich mich in diesem experimentellen Raum wohl. Doch diese Autoethnografie zeigt auch deutlich: Für Medienschaffende, die nicht bereit sind, sich tief in Code, API-Dokumentationen und technische Frustrationstoleranz einzuarbeiten, bleibt die vermeintliche ‘Everything Machine’ für Nischen-Ästhetiken vorerst ein starres Werkzeug, das den kreativen Prozess eher diktiert als befreit.

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, wird diesen Zustand der Vorläufigkeit als methodische Grundbedingung berücksichtigen müssen.

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>
- 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>
- Galanter, P. (2003). What is Generative Art? Complexity Theory as a Context for Art Theory. *Generative Art 2003: Proceedings of the International Conference*.
- Hámošová, L., & Rusnák, P. (2023). TroublingGAN: Generated Visual Ambiguity as a Speculative Alternative to Photojournalism. *Journal for Artistic Research*, (31). <https://jar-online.net/en/exposition/abstract/troublinggan-generated-visual-ambiguity-speculative-alternative-photojournalism>
- 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>
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models from Natural Language Supervision. *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 139, 8748–8763.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684–10695. <https://doi.org/10.1109/CVPR52688.2022.01042>
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action* (S. 374). Basic Books.
- Torres Núñez del Prado, P. (2025). Encoding Culture. *PARSE Journal*. <https://parsejournal.com/article/encoding-culture/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

Abbildungsverzeichnis

1	Die fünf Seiten der Website (https://jenniferied.github.io/everything-machine/): Projektübersicht mit 3D-Viewern, Prozess-Logbuch, Über mich mit Interessen-Netzwerk, Bild-Generationen und Video-Generationen.	10
2	Input: Kepler A-Pose (1024×1024)	14
3	Phase 1: Modell-Screening – MiniMax, Z-Image, FLUX Depth, FLUX Canny. Bilder in der Bild-Galerie vergrößerbar.	14
4	Phase 1: Era3D Multiview – nutzlos im Vergleich zu eigenen Renders. Bild-Galerie . . .	15
5	Phase 2: Top-Modelle in Szenen (A-Pose). Bild-Galerie	16
6	Input-Wechsel: A-Pose → Posed – natürliche Haltung statt T-Pose	17
7	Phase 2b: Drei Finalisten mit Posed Input (je 5 Varianten). Bild-Galerie	17
8	Phase 3a: GPT-1.5 mit Ratio-Bug (oben, quadratisch), Nano Banana korrekt (unten). Bild-Galerie	18
9	Phase 3b: Fehlerhafte Referenzbilder (gestrechlt statt gepaddet)	19
10	Phase 3b: Outputs mit gestrechten Referenzbildern. Bild-Galerie	20
11	Phase 3c: Korrekt gepaddete Referenzbilder	21
12	Phase 3c: Outputs mit korrekt gepaddeten Inputs. Bild-Galerie	22
13	Input-Vergleich Phase 3d: Posed (links) vs. A-Pose (rechts)	23
14	Phase 3d: Portrait-Vergleich – Posed (21/22) vs. A-Pose (23/24). Bild-Galerie	24
15	Phase 3d: Landscape-Vergleich – Posed (21/22) vs. A-Pose (23/24). Bild-Galerie	25
16	Mittlere Frames, Studio-Szene (Landscape, 7 Modelle). Verlinkt auf die Video-Galerie. .	27
17	Mittlere Frames, Pool-Szene (Portrait, 7 Modelle). Verlinkt auf die Video-Galerie. . . .	28
18	Mittlere Frames, Night-Drive-Szene (Portrait, 7 Modelle). Verlinkt auf die Video-Galerie.	29
19	Mittlere Frames, Spiral-Staircase-Szene (Portrait, 7 Modelle). Verlinkt auf die Video-Galerie.	30

Anhang A - Autoethnographische Interviews

Die folgenden Interviews wurden als autoethnographische Methode begleitend zu jedem Experiment durchgeführt. In einem strukturierten Dialog stellte Claude fünf standardisierte Fragen zu den Forschungsdimensionen Workflow, Autorschaft, Iteration, Scheitern und Ästhetik, gefolgt von einer kontextspezifischen Synthesefrage. Die Interviews fanden unmittelbar nach jeder Session statt und wurden per Spracheingabe aufgezeichnet. Die vollständigen Journal-Einträge sind im [Prozess-Logbuch](#) auf der Website einsehbar.

1 --- Referenzen sammeln · 18. November 2022

 **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).

 **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.

 **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.

2 --- ComfyUI Consistent Character · 16. November 2023

 **Claude** Welche Tools hast du in dieser Session verwendet?

 **Jennifer** ComfyUI mit einem Network von Mickmumpitz.

 **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.

 **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.

 **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.

3 --- Marble von World Labs · 18. November 2022

 **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 ComfyUI-Workflow, hatte mich interessiert, ob wir vielleicht Hintergrundwelten generieren könnten.

 **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.

 **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.

4 --- Website mit Cursor aufsetzen · 19. November 2022

 **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.

 **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 verleitet 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.

 **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.

8 --- KI wirklich verstehen · 11. Januar 2026

 **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?

 **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.

 **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.

6 --- Suno-Experiment · 18. Januar 2026

 **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.

 **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.

 **Claude** Im Eintrag schreibst du, dass Sunos generierte Beschreibungen ('Bitcrushed chip-tune 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.

7 --- Kepler und die KI, ein zweiter Versuch · 28. Januar 2026

 **Claude** Welche Tools hast du in dieser Session verwendet?

 **Jennifer** Ich habe erstmal meine Referenz, die ich in Blender gerendert habe, in Gemini gepackt. Die hat sich dann NanoBanano Pro bedient. Das war mir dann aber zu tricky, deswegen habe ich fal.ai benutzt. Ich benutze mittlerweile gar nicht mehr Cursor, sondern VS Code, weil das kostenlos ist, und ich bin komplett auf Claude Code umgestiegen.

 **Claude** Was wolltest du erreichen?

 **Jennifer** Ich wollte endlich erfolgreiche Generationen machen und visuell cool aussehende Bilder erstellen, die nahe an finaler Qualität sind. Und ich wollte die neuen Modelle ausprobieren, ohne extremen Aufwand zu betreiben — es sind ständig neue da und es ist schwer zu wissen, was das Beste ist. Ich wollte auch mit Automatisierung und agentischen Workflows in Claude Code spielen.

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

 **Jennifer** Der gesamte Workflow hat sich fundamental verändert. Anstatt dieses schwierigen Arbeitens mit ComfyUI hatte ich plötzlich dieses total automatisierte Setup mit Claude, wo ich zu den einzelnen Bildern interviewt wurde. Ich konnte mir durch die verschiedenen Ergebnisse ein besseres Bild machen. Durch die objektive Bewertung per Interview und die Erstellung der Tabelle konnte ich am Ende wirklich sagen: das funktioniert oft gut und das nicht. Dadurch habe ich mit Hilfe der KI recherchiert, was Best Practices sind. Die Ergebnisse wurden dadurch teilweise besser — zum Beispiel das quadratische Aspect Ratio, welches ich dann ins Portrait- oder Landschaftsformat bringen konnte.

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

 **Jennifer** Ich war technische Problemlöserin. Instruktionsbefolgerin. Claude Code hat mir wirklich detaillierte Instruktionen gegeben, wie ich das Ganze aufsetzen soll. Aber ich bin selber auf die Idee gekommen, habe mich auf der Seite registriert, den API Key geholt, das Ganze bezahlt. Es war eigentlich genau Kuratorin: Ich habe etwas Technisches ausprobiert und dann bewertet. Davor habe ich die technischen Komponenten nach Anleitung von Claude Code zusammengestellt. Ich fühle mich nicht wirklich wie eine Künstlerin gerade. Macht mir aber auch nicht unbedingt was aus. Ich fand es einfach erleichternd und cool, mal interessante Ergebnisse zu kriegen.

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

 **Jennifer** Ich habe viel mehr Modelle ausprobiert, als es ohne KI jemals möglich gewesen wäre. Es ist ein reproduzierbarer Workflow — das werde ich in der Zukunft wieder benutzen. Da es immer wieder neue Modelle geben wird, und ich jetzt weiß, dass ich in dieser Domäne zu Ergebnissen kommen kann, werde ich es weiterhin so automatisiert mit Claude Code über einen API Key machen.

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

 **Jennifer** Da ich das Bild nicht selber breiter oder höher gemacht habe, sondern die KI gesagt hat, sie soll es machen, hat sie mich missverstanden und hat das Bild gestrechelt. Dadurch habe ich Geld und Zeit verloren mit verzerrten Ergebnissen. Genauso schnell wurde es behoben, aber manchmal ist es frustrierend, die KI zu korrigieren. Da hilft es natürlich, sich präzise auszudrücken. Ich tendiere dazu, schnell die Instruktion zu geben und schnell das Ergebnis haben zu wollen — dann gebe ich mir nicht so viel Mühe, genau zu beschreiben, was ich will.

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

 **Jennifer** Nicht unbedingt eine eigene Sprache, denn die Ästhetik wurde durch Kepler selbst — das 3D-Modell, welches wir erstellt haben, und das, was die KI dazu gedacht hat — erstellt. Ich habe da noch nicht viel mit dem Prompting rumgespielt, ob man etwas Stil reinbringen kann.

 **Claude** Du beschreibst den Shift vom "Überredungs-Problem" zum "Automatisierungs-Problem". Gleichzeitig sagst du, du fühlst dich nicht wie eine Künstlerin, sondern wie eine Kuratorin. Hat dieser Wechsel etwas an deiner Beziehung zu den Ergebnissen verändert?

 **Jennifer** Die Bilder fühlen sich von Anfang an nicht nach etwas an, das ich selber gemacht habe. Klar, ich habe das Modell gerendert und ein bisschen daran gearbeitet, aber das Grund-3D-Modell wurde von jemand anderem erzeugt — von Gavin. Er hat mit anderen 3D Artists den Voxel-Kepler entworfen, und ich habe ihn nur modifiziert. Also sehe ich die 3D-Figur nicht als meine Kreation. Hätte ich das Charakterdesign, den Charakter selber und die komplette 3D-Modellierung von Anfang an erstellt, dann hätte ich wahrscheinlich ein viel höheres Gefühl der Autorschaft. So hält es sich in Grenzen. Aber allgemein ist es in diesem Projekt so, dass ich nicht die einzige Autorin bin. Das fällt mir übrigens gerade erst ein, wie du mir diese Frage stellst.

8 --- Video-Generierung · 18. Februar 2026

 **Claude** Welche Tools hast du in dieser Session verwendet?

 **Jennifer** Ich habe wieder fal.ai API und Claude Code verwendet. Claude Code hat für mich ein Python-Skript geschrieben, in dem ich sieben verschiedene Video-Modelle ansprechen konnte.

 **Claude** Was wolltest du erreichen?

 **Jennifer** Ich wollte diesen Workflow noch mal an Videos testen. Ich war endlich zufrieden mit den Bildergebnissen und dachte, ich könnte diese als Ausgangsbilder für Image-to-Video-Modelle benutzen. Das bringt mich näher an mein Endziel: Cover-Grafiken und Spotify-Reels für Kepler. Ich wollte schauen, wie nah ich an ein brauchbares Ergebnis kommen kann.

 **Claude** Hat sich dein Arbeitsablauf verändert im Vergleich zum Bild-Experiment? (*Workflow*)

 **Jennifer** Es war einfacher durch die bestehende Pipeline als zuvor. Ansonsten war es sehr ähnlich und auch sehr interessant. Die Halluzinationen waren besonders interessant — es sind einfach viel mehr Elemente, die schiefgehen können oder interessant sein könnten.

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

 **Jennifer** Wieder Kuratorin — Anweisungen geben und dann kuratieren. Vielleicht war ich hier auch ein bisschen mehr Erforscherin. Ich habe mich neugierig gefühlt, was die Ergebnisse angeht. Ich hatte mich in dieses systematische Bewerten eingegroovt und war gespannt auf die quantitativen und qualitativen Ergebnisse.

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

 **Jennifer** Ich konnte mehr Experimente machen, genauso wie bei den Bildern. Ich habe jetzt einen reproduzierbaren Workflow und kann jedes neue Modell einfach ausprobieren, ohne Burnout. Ich sage der KI einfach: "Recherchiere die neuesten Modelle und generiere mir fünf Beispielbilder" — und dann habe ich ein Resultat. Es hat mir Confidence gegeben, dass KI auch coole, schnelle Ergebnisse liefern kann, bei denen man schnell iterieren kann. Es ist viel aufwändiger, alles in einen Chatbot zu schreiben und rein- und rauszukopieren — durch den agentischen Workflow mit Claude Code wird das automatisiert.

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

 **Jennifer** [Keine spezifischen Probleme über die im Journal dokumentierten hinaus.]

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

 **Jennifer** Es ist gar nicht meine eigene visuelle Sprache! Ich habe in den Prompts gar nicht mit dem Stil experimentiert — das ist eine Frage für weitere Experimente. An sich finde ich den Style cool, aber er ist nicht meiner. Oder der von Kepler. Auch wenn Kepler dort teilweise gut drin aussieht.

 **Claude** Die Gesichtshalluzination passiert im Video progressiv — Frame für Frame wächst ein Gesicht in die Maske. Siehst du darin etwas über Keplers Identität, oder ist das eher ein technisches Problem?

 **Jennifer** Ich bin gar nicht die Schöpferin von Kepler — ich setze seine Vision um. Seine ästhetischen Qualitäten sind in unserem Pinterest Board zu finden. Ich bin noch gar nicht an dem Punkt, wo ich unsere Inputs verwende oder ihn mit reinbringe. Jetzt, wo die Modelle funktionieren — und das ist relativ neu, diese Prompt Adherence, wahrscheinlich durch die Kombination von Large Language Models mit bildgenerierenden Modellen — ist der Moment, wo ich zum Künstler gehe und mit ihm zusammen iteriere. Oder erstmal in unsere Referenzen gehe und versuche, genau diesen Stil nachzubilden.

Anhang B - Promptverzeichnis (Experiment 3: Automatisierte Bildpipeline)

Die folgende Übersicht dokumentiert alle in der Pipeline verwendeten Prompts, gruppiert nach Phase. Ab Phase 3b wurden für GPT-Image-1.5 und Nano Banana Pro unterschiedliche Prompt-Varianten verwendet: GPT erhielt strukturierte Prompts (Background/Center/Details/Constraints), Nano Banana natürlichsprachliche. Hier ist jeweils die Nano Banana-Variante dokumentiert; die GPT-Variante beschrieb dieselbe Szene in strukturierter Form.

Phase 1 – Modell-Screening (alle Modelle, identischer Prompt)

Base: "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"
+ *Szenen:* "...in a neon-lit cyberpunk city at night" / "...in a foggy forest with golden light rays" / "...on a concert stage with dramatic spotlights"

Phase 2 – Top-Modelle in Szenen (alle Modelle, identischer Prompt)

Cyberpunk: "Place this voxel character in a neon-lit cyberpunk city at night, walking down the street with hands in pockets, cinematic lighting"
Forest: "...in a foggy forest clearing, sitting on a fallen tree, golden light rays filtering through the canopy"
Concert: "...on a concert stage, performing with arms raised, dramatic spotlights and fog machine effects"

Phase 2b – Posed-Input-Test (5 Szenen, identischer Prompt)

Cyberpunk: "Place this voxel character in a neon-lit cyberpunk city at night, walking down a rain-soaked street with hands in pockets, cinematic lighting, reflections on wet ground"
Park: "...sitting on a park bench in autumn, golden leaves falling around him, soft warm afternoon light"
Concert: "...on a dark concert stage, performing with a microphone, dramatic red and blue spotlights, fog machine, crowd silhouettes"
Rooftop: "...standing on a rooftop at sunset, looking over a vast city skyline, wind blowing, golden hour lighting"
Studio: "...in a recording studio, sitting at a mixing desk with headphones around neck, warm ambient lighting, monitors glowing"

Phase 3a – Signature Scenes v1 (GPT + Nano Banana, identischer Prompt)

Studio: "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 [...]"
Retro Car: "...driving a vintage 1970s convertible car across a long suspension bridge at night, city lights reflecting on the water below [...]"
Spiral Staircase: "...walking up an endless spiral staircase that extends infinitely into the sky, passing through layers of clouds and stars, surreal dreamlike atmosphere, the staircase is white marble [...]"
Pool Floaty: "...lying on a pink flamingo pool float in a turquoise swimming pool, seen from directly above as a top-down bird's eye view [...]"

Phase 3b – Fixed Scenes (separate GPT/NB-Prompts, hier NB)

Studio: "Show this voxel character in a cozy recording studio. He's sitting at a big mixing desk with his hands on the faders, wearing headphones. Warm amber lighting [...] IMPORTANT: The character must remain completely faceless [...]"
Tokyo Rain: "This voxel character is walking through a rainy Tokyo alley at night. Neon signs everywhere in Japanese, steam rising from street vents [...] IMPORTANT: The character must remain completely faceless [...]"
Spiral Staircase: "...walks up an endless spiral staircase made of white marble, floating through deep space. Stars and galaxies surround the staircase [...] IMPORTANT: The character must remain completely faceless [...]"
Pool Floaty: "Top-down bird's eye view of a turquoise swimming pool. This voxel character is lying on an inflatable pool ring (donut shape with a hole in the middle) [...] IMPORTANT: The character must remain completely faceless [...]"

Phase 3c – Final Scenes (hier NB-Variante)

Studio: wie Phase 3b
Night Drive (neu, ersetzt Tokyo): "This voxel character is driving a vintage 1970s convertible down a long empty road at night. A big full moon and stars fill the dark sky [...] IMPORTANT: The character must remain completely faceless [...]"
Spiral Staircase (update): "...walks up an endless spiral staircase [...] The staircase is built from the same voxel cubes as the character — blocky geometric steps that glow with soft blue light. [...] NOT marble, NOT stone."
Pool Floaty: wie Phase 3b, erweitert um Poolside-Details (Liegen, Cocktails, Pflanzen)

Phase 3d – Refined + A-Pose-Vergleich (hier NB-Variante)

Night Drive (refined): "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. [...] Retro synthwave vibes with teal and magenta tones. IMPORTANT: The character must remain completely faceless [...]"
Spiral Staircase (refined): "...walks up an endless spiral staircase floating through a dreamy cloudy sky. The sky is soft pastel blue and lavender with layered clouds [...] NOT intense galaxies, just a calm dreamy twilight atmosphere. [...] blocky geometric steps that glow with soft blue neon light. [...]"
Pool Floaty (enriched): "Top-down bird's eye view of a turquoise swimming pool at a luxurious summer resort. [...] Around the pool edge: white sun loungers, a small table with a cocktail and sunscreen, tropical potted plants, neatly folded towels, flip-flops on the tiles. [...]"

Anhang C - Gesamtbewertung Bildpipeline (Experiment 3)

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	Nano Banana	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	Nano Banana	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	Nano Banana	posed	portrait (sq.)	gemischt	ja	Halb cool, halb Halluz.
14	Nano Banana	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	Nano Banana	posed	portrait	gemischt	nein	Stretching, keine Halluz.
16	Nano Banana	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	Nano Banana	posed	portrait	gemischt	nein	-
18	Nano Banana	posed	landsc.	stark	nein	-
19	GPT-1.5	posed	portrait	gemischt	teilweise	-
19	GPT-1.5	posed	landsc.	gemischt	teilweise	-
20	Nano Banana	posed	portrait	gemischt	nein	-
20	Nano Banana	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	Nano Banana	posed	portrait	ja	nein	ND
22	Nano Banana	posed	portrait	nein	nein	SS: Stufen nicht schön
22	Nano Banana	posed	portrait	ja	nein	PF: Sehr cool
22	Nano Banana	posed	landsc.	ja	ja	ND: Cockpit falsch
22	Nano Banana	posed	landsc.	nein	ja	SS: Doppelte Treppe
22	Nano Banana	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	Nano Banana	a-pose	portrait	ja	nein	ND
24	Nano Banana	a-pose	portrait	nein	ja	SS: Treppe nicht schön
24	Nano Banana	a-pose	portrait	ja	nein	PF: Super
24	Nano Banana	a-pose	landsc.	neutral	ja	ND: Halluz. stören
24	Nano Banana	a-pose	landsc.	nein	ja	SS: Treppen nicht schön
24	Nano Banana	a-pose	landsc.	neutral	ja	PF: Pool Floaty falsch

Anhang D - Gesamtbewertung Video-Generierung (Experiment 4)

Modell	Szene	\$	Sek.	Bew.	Halluz.	Kommentar
<i>Studio (Landscape, 1376x768)</i>						
PixVerse v5.5	Studio	0.20	37	okay	leicht	Kamerabewegung gefällt, Bewegung unpassend
LTX-2	Studio	0.36	79	gut	leicht	Schwer zu beurteilen ob halluziniert
Wan 2.6	Studio	0.50	68	gut	nein	Bild leicht überzeichnet, Audio trotzdem generiert
Seedance Pro	Studio	0.62	53	okay	nein	Bewegt sich nicht, aber okay
Veo 3.1	Studio	0.80	49	schlecht	leicht	Slider bewegen sich, Bewegung gefällt nicht
Kling O3	Studio	0.84	165	gut	leicht	Audio Meter unscharf, Bewegung gefällt sehr
Kling 3.0 V3	Studio	0.84	105	gut	leicht	Audio Meter schwammig, Bewegung cool
<i>Pool (Portrait, 768x1376)</i>						
PixVerse v5.5	Pool	0.20	33	schlecht	stark	Voxel-Gesicht verändert sich, Bewegung zu schnell
LTX-2	Pool	0.36	78	schlecht	stark	Aspect Ratio geändert, Kamera bewegt sich zu sehr
Wan 2.6	Pool	0.50	59	schlecht	stark	Gesicht entsteht mit der Zeit, Floaty-Bewegung gelungen
Seedance Pro	Pool	0.62	57	schlecht	stark	Gesicht entsteht, Zehen in Schuhen
Veo 3.1	Pool	0.80	64	gut	leicht	Gesicht verändert sich leicht, Floaty-Bewegung cool
Kling O3	Pool	0.84	129	gut	leicht	Ripples falsch, Wassersimulation top
Kling 3.0 V3	Pool	0.84	108	gut	nein	Das beste!
<i>Night Drive (Portrait, 768x1376)</i>						
PixVerse v5.5	Night Drive	0.20	30	schlecht	stark	Gesicht wird Lego-Figur, Autofahrt unnatürlich
LTX-2	Night Drive	0.36	89	schlecht	stark	Auto fährt rückwärts, Aspect Ratio geändert
Wan 2.6	Night Drive	0.50	60	schlecht	leicht	Gesicht wird generiert, Animation sonst gut
Seedance Pro	Night Drive	0.62	55	gut	nein	Überraschend gut
Veo 3.1	Night Drive	0.80	65	okay	leicht	Gut, fährt aber rückwärts
Kling O3	Night Drive	0.84	113	schlecht	stark	Halluziniert am Rand, zoomt heraus
Kling 3.0 V3	Night Drive	0.84	138	okay	leicht	Etwas besser als Kling O3
<i>Spiral Staircase (Portrait, 768x1376)</i>						
PixVerse v5.5	Spiral St.	0.20	37	okay	leicht	Stufen unnatürlich, interessanter Glow
LTX-2	Spiral St.	0.36	78	schlecht	stark	Läuft fast von Stufe, Aspect Ratio ändert sich
Wan 2.6	Spiral St.	0.50	58	okay	leicht	Bewegung leicht unnatürlich
Seedance Pro	Spiral St.	0.62	52	okay	leicht	Kopf verändert sich, Glow okay
Veo 3.1	Spiral St.	0.80	70	okay	leicht	Überspringt Stufen, Animation relativ okay
Kling O3	Spiral St.	0.84	107	gut	leicht	Eine Stufe funktioniert nicht, Leuchten gefällt
Kling 3.0 V3	Spiral St.	0.84	123	schlecht	leicht	Stufen funktionieren nicht, Kamera unvorhersehbar

Eidesstattliche Erklärung

Ich versichere, die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im Schriftenverzeichnis angegebenen Hilfsmittel nicht bedient zu haben. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Im Rahmen dieser Arbeit wurden KI-gestützte Werkzeuge (u. a. Large Language Models und Bildgenerierungsmodelle) als Recherche-, Formatierungs- und Produktionshilfen eingesetzt. Sämtliche Inhalte wurden von der Autorin geprüft, kuratiert und verantwortet. Die eingesetzten Werkzeuge und Methoden sind in Abschnitt „Ethische Überlegungen und KI-Offenlegung“ dokumentiert.

Detmold, den 18. Februar 2026



Jennifer Meier