



TECHNISCHE HOCHSCHULE
OSTWESTFALEN-LIPPE
UNIVERSITY OF
APPLIED SCIENCES
AND ARTS



EVERYTHING MACHINE

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

Jennifer Meier

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

Artistic Research

Prof. Dr. Aristotelis Hadjakos

Februar 2026

Abstract

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

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

Inhaltsverzeichnis

Einleitung	5
Hintergrund und Motivation	5
Forschungsfrage und Überblick	5
Stand der Forschung	6
Artistic Research als Erkenntnisform	6
Technische Grundlagen generativer KI	6
Kreativität und Mensch-KI-Kollaboration	7
Forschungslücke	8
Methodologie	8
Methodischer Rahmen: Practice-Led Research	8
Konkrete Methoden	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
Experiment 3: Musik für ein Alter Ego (Suno AI)	12
Experiment 4: KI-Dialog als Reflexionswerkzeug	12
Experiment 5: Automatisierte Bildpipeline (fal.ai API + Claude Code)	12
Experiment 6: Video-Generierung (fal.ai API + Claude Code)	27
Zentrale Beobachtungen aus dem Prozess	32
1. Unvorhersehbarer Flow	32
2. Speed vs. Control	33
3. Verschiebung der Autorschaft	33
4. KI-Burnout und KI-Enthusiasmus	34
Rückbezug auf die Forschungsfrage	34

Diskussion und Interpretation	35
Zusammenfassung der Ergebnisse	35
Theoretische Implikationen	36
Praktische Implikationen	36
Stärken und Limitierungen	37
Stärken	37
Limitierungen	37
Practice-Based Intention, Practice-Led Realität	37
Ausblick	38
Literaturverzeichnis	39

Einleitung

Hintergrund und Motivation

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. 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?

Dass diese Werkzeuge nahezu gleichzeitig verfügbar wurden, ist kein Zufall, sondern Folge konvergierender Durchbrüche: Generative Adversarial Networks legten 2014 die Grundlage für Bildgenerierung, die Transformer-Architektur ermöglichte 2017 die Skalierung auf Milliarden von Parametern, und Latent Diffusion Models machten 2022 hochauflösende Bildgenerierung erstmals auf Consumer-Hardware möglich – das direkte Fundament der Tools, mit denen diese Arbeit operiert.

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

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

Forschungsfrage und Überblick

Die zentrale Forschungsfrage lautet:

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

Im Laufe der Praxis veränderte sich nicht nur der kreative Prozess selbst, sondern auch die Art, ihn zu dokumentieren: Die Website nahm einen größeren Raum ein als geplant, das KI-gestützte Reflexionsinterview etablierte sich als Methode, wurde dann als zu knapp kritisiert. Diese Verschie-

bungen bestätigten, dass die Prozessfrage die richtige war: Nicht das Ergebnis, sondern die Veränderung des kreativen Prozesses durch KI erwies sich als zentrale Erkenntnis.

Zur Beantwortung dieser Frage verbindet die Arbeit:

1. **Praktische Experimente:** Iterative Entwicklung von Kepler durch verschiedene KI-Tools, beobachtet auf Prozessveränderungen
2. **Theoretische Reflexion:** Einordnung in Diskurse zu Kreativität, Autorschaft und Human-AI Interaction. Die strukturierten Reflexionsinterviews orientieren sich dabei an fünf Subfragen: Wie verändert sich der Workflow? Wie verschiebt sich die Autorschaft? Welche Rolle spielt Iteration? Wie wird mit Scheitern umgegangen? Und wie entwickelt sich die Ästhetik im Prozess?
3. **Methodische Transparenz:** Öffentliche Dokumentation des gesamten Prozesses

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

Stand der Forschung

Artistic Research als Erkenntnisform

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

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

Turkle (1995) ergänzt diese methodologische Perspektive um die Dimension digitaler Identität: Digitale Identitäten sind nicht bloße Abbilder physischer Personen, sondern eigenständige Konstruktionen, die in der Interaktion mit digitalen Umgebungen entstehen und sich verändern. Diese Perspektive ist für die vorliegende Arbeit zentral, da Kepler als digitale Persona genau in einem solchen Konstruktionsprozess steht.

Technische Grundlagen generativer KI

Die aktuelle Welle generativer KI-Werkzeuge basiert auf drei konvergierenden Durchbrüchen. Goodfellow et al. (2014) führten 2014 mit Generative Adversarial Networks (GANs) das Prinzip ein, dass zwei konkurrierende Netzwerke lernen, realistische Daten zu erzeugen — die Grundlage für alle nachfolgenden Bildgenerierungsmodelle. Die Transformer-Architektur von Vaswani et al. (2017) ersetzte 2017 rekurrente Netzwerke durch einen Attention-Mechanismus, der parallele Verarbeitung

ermöglichte und die Skalierung auf Milliarden von Parametern erlaubte — die technische Voraussetzung für GPT, DALL-E und die multimodalen Systeme, mit denen diese Arbeit operiert. Rombach et al. (2022) verschoben 2022 den Diffusionsprozess in einen komprimierten latenten Raum und machten 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 ist für die vorliegende Arbeit nützlich als Analysewerkzeug: Sie erlaubt zu beschreiben, welche Art von Kreativität die KI zeigt und welche bei mir als Künstlerin verbleibt.

Galanter (2003) definiert generative Kunst als jede Praxis, bei der ein autonomes System zum Produktionsprozess beiträgt — vom Computerprogramm über physikalische Prozesse bis zu Regelwerken. KI-gestützte Kunstproduktion steht damit in einer langen Tradition. Gleichzeitig ist der aktuelle Moment qualitativ anders: Multimodale KI-Systeme operieren in Text, Bild und Audio gleichzeitig, mit 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. In dieser Arbeit geht es nicht darum, ob die KI kreativ „ist“, sondern wie sich der kreative Prozess in der Kollaboration verändert.

Ein zentrales Phänomen tritt beim praktischen Arbeiten mit generativer KI wiederholt auf: Das *Uncanny Valley* (Mori et al., 2012) beschreibt das Unbehagen, das entsteht, wenn künstliche Darstellungen menschenähnlich, aber nicht perfekt realistisch sind. In der Arbeit mit Kepler zeigt sich dieses Phänomen als konkretes Hindernis — Bildmodelle halluzinieren Gesichtszüge in seine Maske, weil sie die Abwesenheit eines Gesichts nicht akzeptieren. ## Artistic Research mit generativer KI

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

Torres Núñez del Prado (2025) untersucht in „Encoding Culture“, wie kulturelle Kodierungen in KI-Trainingsdaten eingebettet sind und wie Künstler*innen diese sichtbar machen und unterlaufen können — eine Perspektive, die für Keplers Konfrontation mit normativen Gesichtsmodellen direkt relevant ist.

Diese Projekte bleiben jeweils in einer Modalität. Mein Ansatz geht über Bild, Audio, 3D und Code

- gerade weil Kepler in all diesen Medien existiert.

Forschungslücke

Autoethnografische Studien zu KI im kreativen Prozess gibt es – in der Musikkomposition, in der visuellen Kunst. Was fehlt: Arbeiten, die den iterativen Prozess über mehrere Medien hinweg dokumentieren, aus der Perspektive der Künstlerin selbst.

Methodologie

Methodischer Rahmen: Practice-Led Research

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

Konkrete Methoden

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

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

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

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

Ethische Überlegungen und KI-Offenlegung

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

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

Meine Forschung: Der Weg zu Kepler

Ausgangspunkt und erste Schritte

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

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

Entscheidungen und Abzweigungen

Die Wahl des visuellen Stils

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

Genau diese Designentscheidung erwies sich als fundamentale Herausforderung für die Arbeit mit generativer KI. Bildgenerierungsmodelle wie Flux oder Stable Diffusion sind auf Trainingsdaten 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 nicht oder kaum repräsentiert ist. Was bei einer realistischen, Anime- oder Cartoon-Figur funktioniert hätte, scheiterte an Keplers bewusster Andersartigkeit.

Die Website als Forschungsartefakt

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

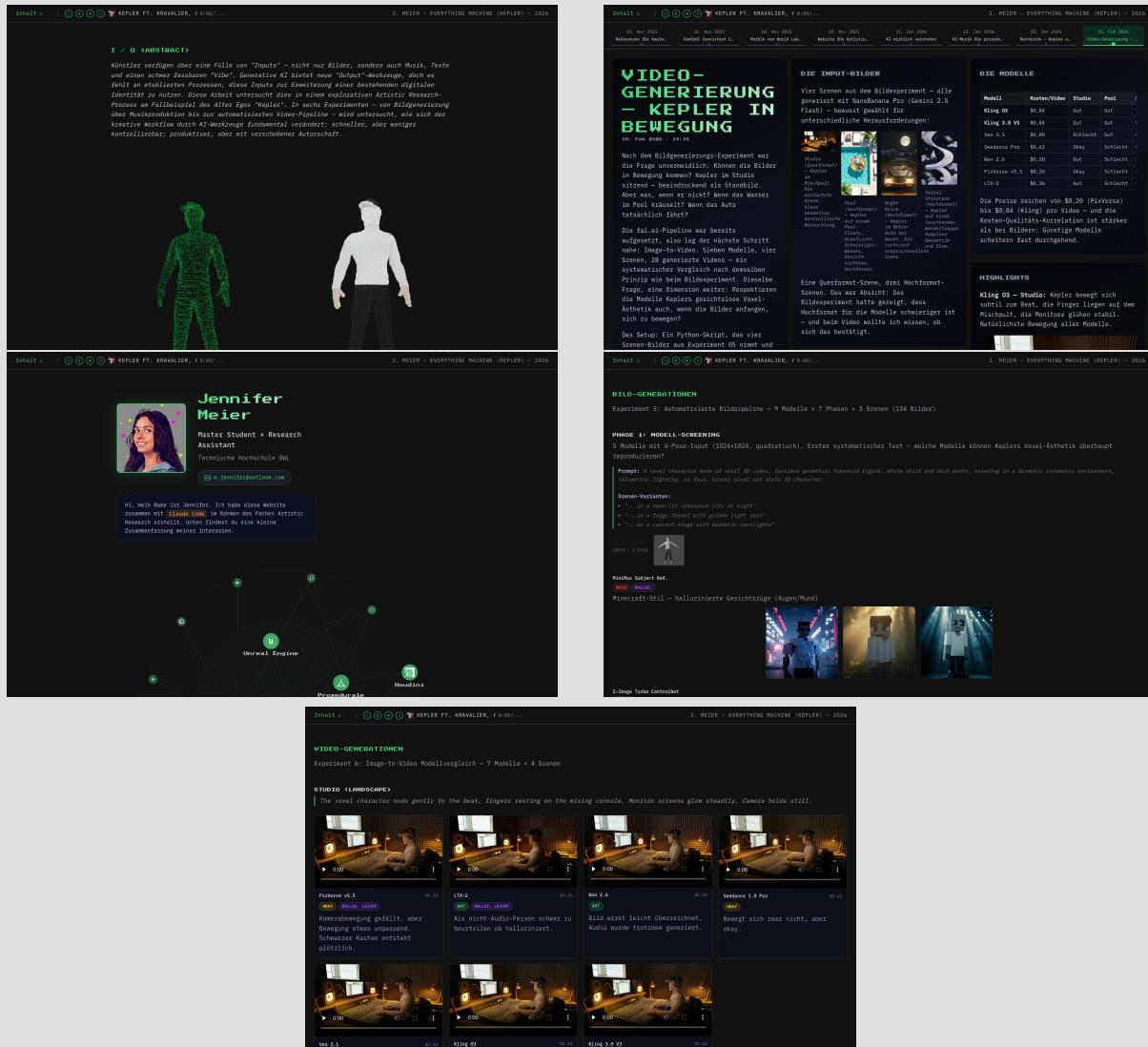


Abbildung 1: Die fünf Seiten der Website (<https://jenniferied.github.io/everything-machine/>): Projektübersicht mit 3D-Viewern, Prozess-Logbuch, Über mich mit Skill-Netzwerk, Bild-Generationen und Video-Generationen.

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

Von der Intention zum Experiment

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

Experimente und Ergebnisse

Experiment 1: Visuelle Konsistenz (ComPyUI + Flux)

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

Die Ergebnisse waren gemischt. Wenn ich ein Foto als Referenz nutzte, auf dem Kepler einen Basketball hielt, wurde der Basketball im generierten Bild beibehalten, obwohl er dort nicht hingehörte. Das Modell hat kein semantisches Verständnis davon, welche Elemente zusammengehören. Gravierender war das Gesichtsproblem: Face Detailer, eigentlich zur Verbesserung von Gesichtsdetails entwickelt, interpretierte Gesichter in Keplers Maske hinein – eine Art Face Hallucination, die selbst durch negative Prompts nicht vollständig unterdrückt werden konnte. Es handelt sich um das, was Mori (2012) als Uncanny Valley beschreibt, allerdings in einer unerwarteten Richtung: nicht ein fast-menschliches Gesicht, das Unbehagen auslöst, sondern ein Nicht-Gesicht, das vom Modell zwanghaft vervollständigt wird. Die Halluzinationen driften in den Bereich der Unordnung ab — 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 nicht, den Prozess aufzugeben, sondern die Parameter zu variieren, den Workflow anzupassen, neue Referenzbilder zu testen. Die Modelle sind auf Mainstream-Ästhetiken trainiert. Keplers Nische fällt durch das Raster.

Experiment 2: Generierte Welten (Marble / World Labs)

Nachdem Kepler als Charakter in der Bildgenerierung an Grenzen gestoßen war, verlagerte ich den Fokus auf Hintergrundwelten. 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 intendierte Workflow.

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, eine Perspektive, die sich durch alle folgenden Experimente ziehen sollte.

Experiment 3: Musik Für ein Alter Ego (Suno AI)

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

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

Experiment 4: KI-Dialog als Reflexionswerkzeug

Ein unkonventionelleres Experiment: Auf einer 30-minütigen Autofahrt führte ich ein Voice-Gespräch mit Claude über KI-Grundlagen – von Shannons Informationstheorie über die Dartmouth Conference bis zur Transformer-Architektur. Das war kein Experiment im engeren Sinne, sondern eine Erkundung der KI als personalisiertem Lernpartner.

Die wichtigste Erkenntnis betraf die Fluency Illusion: Das Gespräch fühlte sich an wie tiefes Verstehen, aber ob ich die Konzepte tatsächlich internalisiert hatte, war unklar. Claude selbst wies mich darauf hin, dass es einen Unterschied gibt zwischen deklarativem Wissen ("wissen, dass") und prozedurellem Wissen ("wissen, wie"). Die Theorie wird erst real, wenn ich sie anwende. Erkenntnis entsteht nicht im Hören oder Lesen, sondern im Tun.

Experiment 5: Automatisierte Bildpipeline (Fal.ai API + Claude Code)

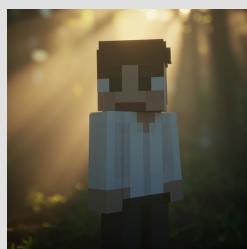
Nach vier Einzelexperimenten stellte sich eine methodische Frage: Was passiert, wenn ich den Prozess systematisiere? Gemeinsam mit Claude Code baute ich eine automatisierte Pipeline, die über die fal.ai API neun verschiedene Bildgenerierungsmodelle mit denselben Prompts und Referenzbildern ansteuerte.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

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

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



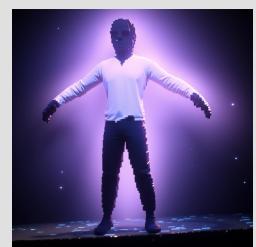
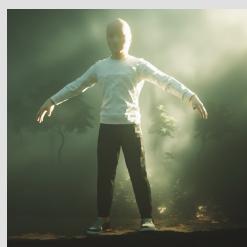
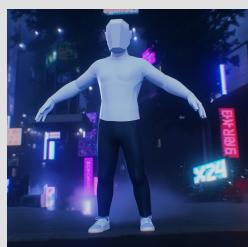
Abbildung 2: Input: Kepler A-Pose (1024×1024)



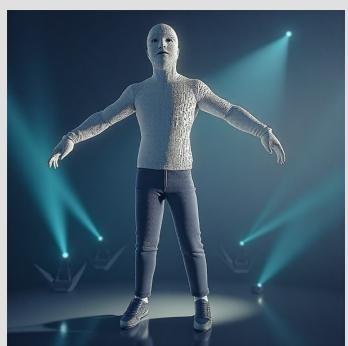
MiniMax Subject Reference



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.

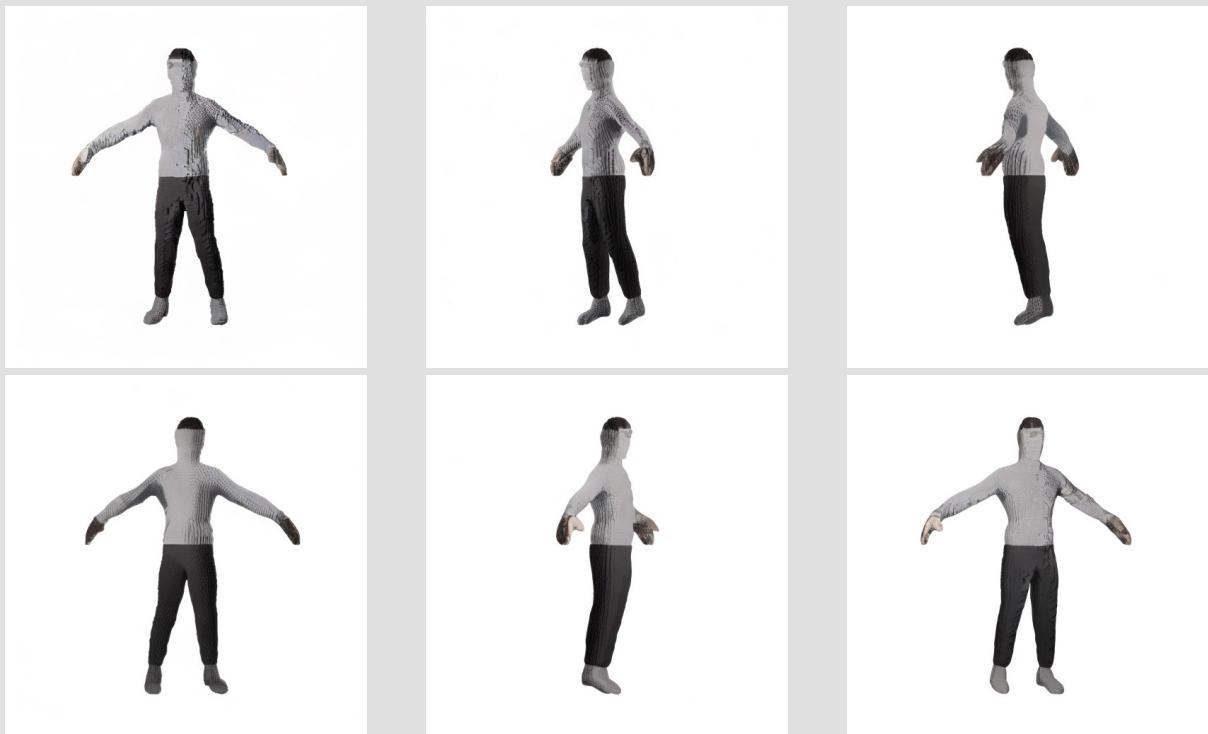


Abbildung 4: Phase 1: Era3D Multiview – nutzlos im Vergleich zu eigenen Renders. Bild-Galerie

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

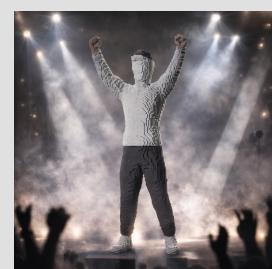
Verworfene Modelle. Keines der fünf Modelle lieferte überzeugende Ergebnisse. MiniMax Subject Reference interpretierte Keplers Voxel-Ästhetik als Minecraft-Welt und halluzinierte menschliche Gesichtszüge, obwohl der Prompt “faceless” explizit vorgab. FLUX Depth LoRA bewahrte zwar die Körperhaltung über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

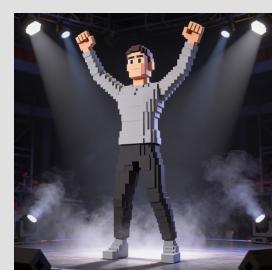
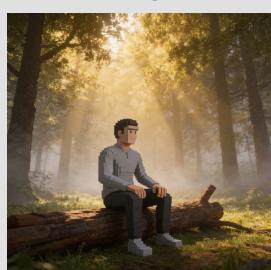
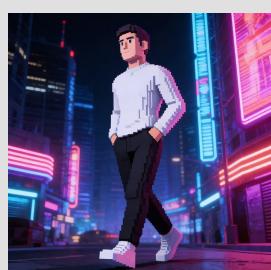
Phase 2: Top-Modelle in Szenen (A-Pose) Die vier vielversprechendsten Modelle wurden in Szenen-Kontexte eingeführt, jeweils drei Varianten mit demselben A-Pose-Input.



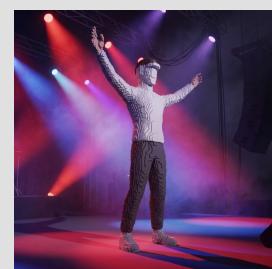
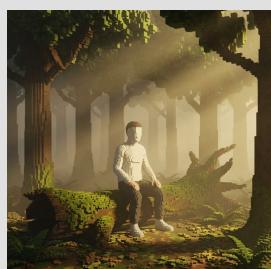
FLUX Kontext Pro



GPT-Image-1.5



SeedDream v4.5



NanoBanana 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
NanoBanana Pro	gut	teilweise	Beste Phase-2-Ergebnisse

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

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

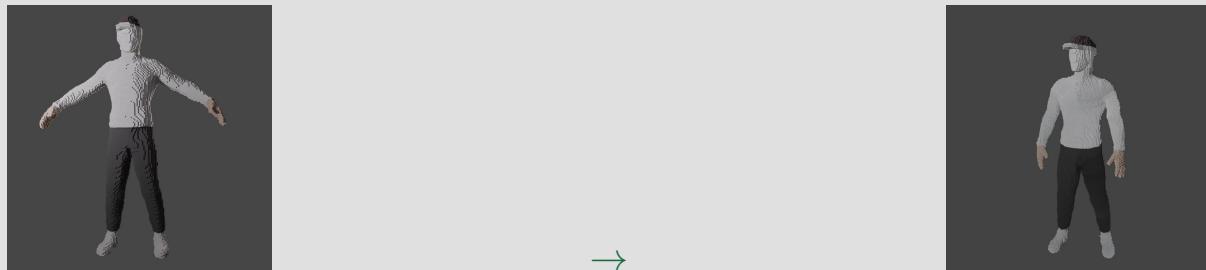


Abbildung 6: Input-Wechsel: A-Pose → Posed – natürliche Haltung statt T-Pose

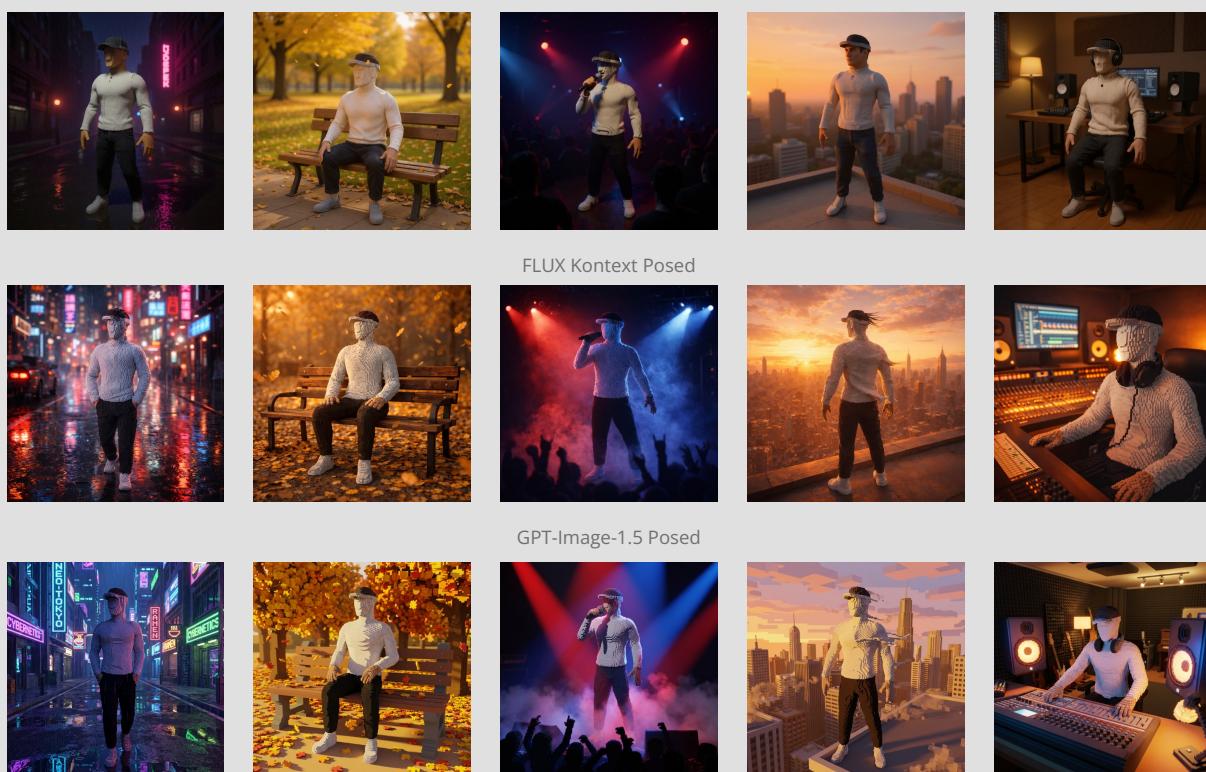


Abbildung 7: Phase 2b: Drei Finalisten mit Posed Input (je 5 Varianten). Bild-Galerie

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
NanoBanana Posed	ja	nein	Stärkste Voxel-Ästhetik

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

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



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



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



NanoBanana Portrait (14) – korrektes Format



NanoBanana Landscape (14) – korrektes Format

Abbildung 8: Phase 3a: GPT-1.5 mit Ratio-Bug (oben, quadratisch), NanoBanana korrekt (unten). [Bild-Galerie](#)

Tabelle 4: Phase 3a: Signature Scenes v1

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



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

Abbildung 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	NanoBanana	portrait	gemischt	nein	Stretching, keine Halluz.
15/16	NanoBanana	landscape	gemischt	teilweise	Weniger Stretching, gut
17/18	GPT-1.5	portrait	gemischt	teilweise	1:1 + leichte Halluz.
17/18	GPT-1.5	landscape	gemischt	teilweise	Dasselbe
17/18	NanoBanana	portrait	gemischt	-	-
17/18	NanoBanana	landscape	stark	-	-

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

In der zweiten Iteration (Ordner 17/18) wurden die Prompts verfeinert: Voxel-Treppen statt Marmor, präzisere Szenenbeschreibungen. Das Stretching blieb jedoch bestehen, da der Fehler in der Input-Vorbereitung lag, nicht im Prompting. GPT-Image-1.5 produzierte erneut quadratische Outputs (1024×1024) trotz Portrait-Anfrage – derselbe API-Bug wie in Phase 3a. NanoBanana Pro hingegen lieferte korrekte Formate und die stärksten Landscape-Ergebnisse der gesamten Pipeline. Die Phasen 3a-3b zeigen 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.



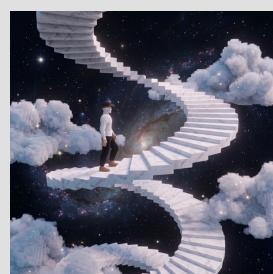
GPT-1.5 Portrait (15)



NanoBanana Portrait (16)



GPT-1.5 Portrait (17)



NanoBanana 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	-	-
NanoBanana	portrait	gemischt	-	-
NanoBanana	landscape	stark	-	-

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

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



GPT-1.5 Portrait (19)



NanoBanana Portrait (20)



GPT-1.5 Landscape (19)



NanoBanana Landscape (20)

Abbildung 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.

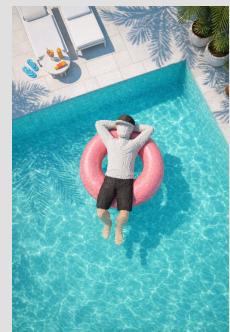


Abbildung 13: Input-Vergleich Phase 3d: Posed (links) vs. A-Pose (rechts)

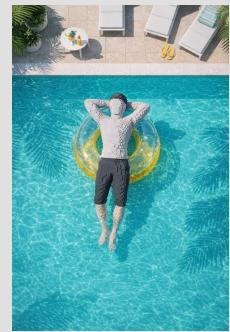
Das Ergebnis war eindeutig: Posed Input war A-Pose in fast allen Szenen überlegen. GPT-Image-1.5 produzierte die atmosphärischsten Ergebnisse, halluzinierte aber häufiger Details. NanoBanana Pro war konsistenter, aber weniger ausdrucksstark.

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

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



GPT-1.5 Posed Portrait (21)



GPT-1.5 A-Pose Portrait (23)



NanoBanana Posed Portrait (22)



NanoBanana A-Pose Portrait (24)

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



NanoBanana Posed Landscape (22)

NanoBanana A-Pose Landscape (24)

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

Gesamtbewertung Die folgende Tabelle fasst alle 52 Einzelbewertungen aus der Pipeline zusammen. Die verwendeten Prompts sind im Prompt-Verzeichnis nach der Tabelle dokumentiert.

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

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

Prompt-Verzeichnis Die folgende Übersicht dokumentiert alle in der Pipeline verwendeten Prompts, gruppiert nach Phase. Ab Phase 3b wurden für GPT-Image-1.5 und NanoBanana Pro unterschiedliche Prompt-Varianten verwendet – GPT erhielt strukturierte Prompts (Background/Center/Details/Constraints), NanoBanana natürlichsprachliche. Hier ist jeweils die NanoBanana-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 + NanoBanana, 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. [...]"

Zusammenfassung und theoretische Einordnung Dieses Experiment markiert eine Verschiebung: Meine Rolle war nicht mehr die der Experimentierenden, die direkt mit einem Tool interagiert, sondern die einer Orchestratorin – ich definierte das „Was“ und „Warum“, während Claude Code das „Wie“ implementierte. 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 explorativen 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. Und eine neue Frage drängte sich auf: Wenn ich nicht mehr selbst prompte, sondern einen KI-Agenten beauftrage, der für mich promptet – wo endet dann meine Autorschaft?

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

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

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

- **Studio** (Landscape, 1376×768): Kepler am Mischpult – kontrollierte Innenszene, wenig Bewegung
- **Pool** (Portrait, 768×1376): Kepler auf dem Pool-Ring, Vogelperspektive – Wasser, ungewöhnlicher Blickwinkel
- **Night Drive** (Portrait, 768×1376): Kepler im Retro-Auto auf Wüstenhighway – Nachtszene, 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.

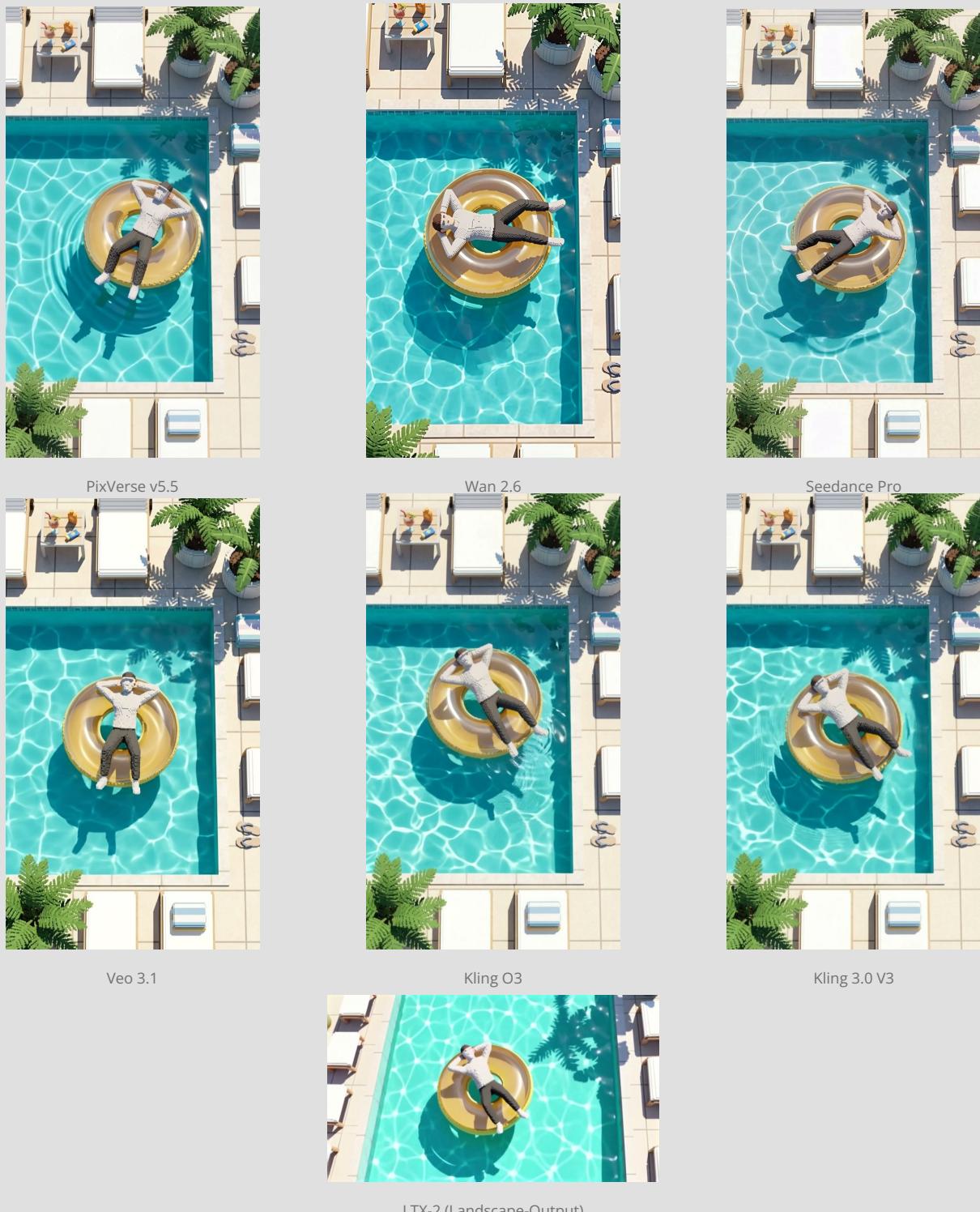


Abbildung 17: Mittlere Frames, Pool-Szene (Portrait, 7 Modelle). Verlinkt auf die [Video-Galerie](#).

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

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

Das auffälligste Phänomen war die **Gesichts-Halluzination bei Portrait-Szenen**. Keplers bewusst gesichtsloser Voxel-Kopf wurde von mehreren Modellen (Wan, Seedance, PixVerse) mit realistischen Gesichtszügen versehen – ein Problem, das bei der Landscape-Studio-Szene nicht auftrat. Die Modelle scheinen auf menschliche Gesichter als Default trainiert zu sein und können die bewusste Abwesenheit eines Gesichts nicht beibehalten. Dieses Phänomen spiegelt die Erfahrung aus Experiment 1 wider: Auch ComfyUI halluzinierte Gesichter in Keplers Maske. Hámová 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 erweiterte dieses Experiment die Orchestrator-Rolle aus Experiment 5: Dieselbe Pipeline, dasselbe Interview-basierte Evaluationsformat, nun auf Video angewandt. Die Bewertungskriterien – „Gefällt mir“, „Halluzination“, „Kommentar“ – erwiesen sich als robust über Medientypen hinweg. Die zentrale Erkenntnis für die Praxis: Video-Generierung mit Kepler ist möglich, aber nur mit den teuersten Modellen zuverlässig. Der kreative Prozess bleibt ein iteratives Auswählen aus einem unvorhersehbaren Möglichkeitsraum.

Zentrale Beobachtungen aus dem Prozess

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

1. Unvorhersehbarer Flow

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

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

2. Speed vs. Control

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

3. Verschiebung der Autorschaft

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

Die wechselnden Rollen bewegen sich auf einem Spektrum — mal näher am „Mensch als Schöpfer“ (Referenzen sammeln), mal in der Mixed-Initiative-Zone (Website-Build), mal näher am „Computer als Schöpfer“ (Suno). Diese Rollenvielfalt ist nicht Unschärfe, sondern Qualität. Die wechselnden Rollen – Kuratorin, Dirigentin, Kollaboratorin, Lernende – sind Ausdruck einer Autorschaft, die sich fundamental verschoben hat: nicht mehr alleinige Urheberschaft, nicht vollständige Delegation. Domänenwissen erweist sich erneut als Voraussetzung für produktive Kollaboration. Wer nicht

weiß, was gut ist, kann nicht kuratieren. In Experiment 5 kam eine weitere Rolle hinzu: die der Orchestratorin, die nicht mehr direkt promptet, sondern einen KI-Agenten beauftragt, der seinerseits Bildmodelle steuert – eine dreistufige Kette, die die Frage nach der Autorschaft weiter verkompliziert.

4. KI-Burnout und KI-Enthusiasmus

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

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

Rückbezug auf die Forschungsfrage

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

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

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

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

Welche Rolle spielt Iteration? Iteration ist nicht Optimierung, sondern Erkenntnismethode. Jeder gescheiterte Prompt, jede unerwartete Halluzination wurde zum Ausgangspunkt für reflektier-

te Anpassung — Schöns (1983) „Situation, die zurückspricht“ in digitaler Form. Die automatisierte Pipeline in Experiment 5 zeigt, dass systematische Iteration den explorativen Raum erweitert, ohne die zentrale Spannung zwischen Kontrolle und Generierung aufzulösen.

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

Wie entwickelt sich die Ästhetik im Prozess? Die Ästhetik entsteht nicht trotz, sondern durch die Einschränkungen der KI. Keplers Nischen-Ästhetik ließ sich nicht direkt generieren, aber die Auseinandersetzung mit den Grenzen der Modelle — Halluzinationen, Mainstream-Drift, fehlende Feinsteuerung — schärfe das ästhetische Urteil und führte zur bewussten Integration von „Fehlern“ als Gestaltungsmittel.

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

Diskussion und Interpretation

Zusammenfassung der Ergebnisse

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

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

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

Theoretische Implikationen

Die Ergebnisse legen nahe, dass bestehende Kreativitätstheorien im Kontext der Mensch-KI-Kollaboration neu betrachtet werden sollten. Bodens (2004) Unterscheidung zwischen explorativer und transformativer Kreativität erweist sich hier als besonders aufschlussreich. 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 Kreativität ging in keinem der sechs Experimente von den Modellen aus, sondern von meinen Entscheidungen als Künstlerin. Die KI liefert Möglichkeiten; der Mensch entscheidet, welche davon neu ist. Der unvorhersehbare Flow, den ich beschreibe, folgt aus dieser Asymmetrie: Die KI exploriert schnell, aber ob das Ergebnis wirklich etwas verändert, erfordert menschliches Urteil.

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

Borgdorff (2012) argumentiert, dass künstlerische Forschung Wissen freilegt, das sich nur durch die Praxis selbst zeigt. Das bestätigt sich hier konkret: Mein Vorwissen über Keplers Ästhetik – die dystopische Klangwelt, die voxelisierte Maskierung, die Spannung zwischen Lo-Fi und Hyperdigitalem – ließ sich als Prompt-Anweisung an KI-Systeme nur begrenzt übertragen. Ohne Domänenwissen produziert die Kollaboration generische Ergebnisse — das bestätigt den in Abschnitt 2 eingeführten Zusammenhang zwischen Expertise und Outputqualität. Dieses Wissen zeigte sich erst *im* iterativen Prozess, nicht vor ihm.

Besonders aufschlussreich ist die Verbindung zu Schön (1983): Reflection-in-Action, das spontane Reagieren auf unerwartete Situationen in der Praxis, beschreibt präzise den Modus, in dem ich auf KI-Halluzinationen, unerwartete Outputs und produktive Fehler reagierte. Das Scheitern eines Prompts war kein Abbruchkriterium, sondern Auslöser für reflektierte Anpassung. 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 resoniert mit meiner vierten Beobachtung: Der KI-Burnout-Zyklus ist möglicherweise ein Symptom dafür, dass aktuelle KI-Tools die Mixed-Initiative-Balance noch nicht gefunden haben — sie sind entweder zu autonom (Suno: keine Eingriffsmöglichkeit) oder zu abhängig von Expertenwissen (ComfyUI: hohe technische Hürde). Die produktivsten Momente meiner Forschung entstanden dort, wo die Balance gelang: beim Website-Build, wo die KI repetitive Aufgaben übernahm und ich die ästhetische Vision steuerte.

Praktische Implikationen

Für Künstler*innen und Medienproduzent*innen, die generative KI in ihre Praxis integrieren wollen, ergeben sich aus dieser Arbeit konkrete Erkenntnisse. Die wichtigste: Domänenwissen ist keine

optionale Zugabe, sondern Voraussetzung für ästhetisch relevante Ergebnisse. Wer ohne klare ästhetische Vision promptet, erhält den kleinsten gemeinsamen Nenner, die Mainstream-Ästhetik, auf die Generalist-Modelle trainiert sind. Technisch lässt sich das als eine Form von Overfitting beschreiben: Die Modelle haben so viele Mainstream-Beispiele gesehen, dass sie dazu tendieren, genau diesen Durchschnitt zu reproduzieren. Nischen-Ästhetiken, für die es weniger Trainingsdaten gibt, werden entweder in Richtung Mainstream gezogen oder scheitern ganz — ein Underfitting auf die spezifische ästhetische Domäne. Keplers Nischen-Ästhetik ließ sich nur dort realisieren, wo ich spezifisches Vokabular, präzise Referenzen und ein geschultes Urteilsvermögen einbrachte.

Ein durchgängiger Speed-versus-Control-Trade-off zeigte sich ebenfalls: KI-Tools beschleunigen die Produktion erheblich, reduzieren aber gleichzeitig die Kontrolle über ästhetische Details. Für die Medienproduktion bedeutet das, dass KI-Werkzeuge sich besonders für Ideation, Prototyping und Exploration eignen, während die finale ästhetische Entscheidung bei der Künstlerin verbleiben muss. Die Rolle verschiebt sich von der Ausführenden zur Kuratorin.

Stärken und Limitierungen

Stärken

Der practice-led Ansatz mit autoethnografischem Journaling hat mir 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 zudem die Komplexität zeitgenössischer Medienproduktion ab, anstatt sie auf ein einzelnes Medium zu reduzieren.

Limitierungen

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

Der erste betrifft 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.

Der zweite betrifft 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.

Der dritte betrifft 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. Paradoxerweise war die Website das Artefakt mit dem meisten Flow – obwohl als Dokumentationstool konzipiert, wurde sie zum vielleicht künstlerischsten Element des Projekts. Aber auch hier liegt der Erkenntnisbeitrag nicht in der Website selbst, sondern 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.

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

Ausblick

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

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

Literaturverzeichnis

- Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd Aufl., S. 344). Routledge.
- Borgdorff, H. (2012). *The Conflict of the Faculties: Perspectives on Artistic Research and Academia*. Leiden University Press. <https://doi.org/10.5117/9789400600997>
- Candy, L. (2006). *Practice Based Research: A Guide* (CCS Report 1.0; S. 1-19). Creativity & Cognition Studios, University of Technology Sydney. <https://www.creativityandcognition.com/resources/PBR%20Guide-1.1-2006.pdf>
- 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>
- Frayling, C. (1993). Research in Art and Design. *Royal College of Art Research Papers*, 1(1), 1-5.
- Galanter, P. (2003). What is Generative Art? Complexity Theory as a Context for Art Theory. *Generative Art 2003: Proceedings of the International Conference*.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, 27, 2672-2680.
- Hámová, 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>
- 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>
- 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/>
- Turkle, S. (1995). *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster.
- 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.