

Achim Mohné

Versuchsreihe zur Auslotung ästhetischer Möglichkeiten K.I.-generierender Mittel in bildgebenden Verfahren anhand der Arbeit

Rotkäppchen

In den 1980er Jahren prophezeite Vilém Flusser die Verdrängung der Texte durch technische Bilder – als einen kulturellen Übergang von der Schrift- zur Bildkultur, durch den Texte ihre bisher dominante Stellung verlieren¹.

Diese Kulturrevolution ist inzwischen Wirklichkeit. Oft stellt sie sich heute als von Rationalität losgelöste Affekthandlung in Politik und Gesellschaft dar. Technische Bilder regieren die Welt nicht zuletzt, weil sie nicht mehr als herkömmliche Bilder auftreten, sondern als in Bilder verkleidete Datenpakete. So verbreiten sie, trojanischen Pferden gleich, meist unbemerkt ihre Metadaten unter den User*innen. Das wird besonders in den sozialen Medien deutlich, die sich für die Generation der *digital natives* zum beliebtesten, wenn nicht zum einzigen Kommunikationsmedium entwickelt haben.

Dalle, Midjourney, Stable Diffusion und diverse andere KI-Konzerne veröffentlichen seit 2020 nahezu täglich neue KI-Verfahren zur Generierung synthetischer Bilder, die durch die Eingabe von Texten erzeugt werden, wodurch sich das Flusser'sche *Universum der technischen Bilder* extrem erweitert und eine neue Ära des Verhältnisses von Text und Bild entsteht. Was bisher im Verborgenen unter oftmals ausbeuterischen Bedingungen durch Programmierer*innen geschah, nämlich Bilder und Text zuzuordnen, übernehmen dieser Tage die User*innen selbst, oft ohne sich dessen bewusst zu sein. Jedes generierte Katzen-Bildchen, sei es noch so spaßig gemeint, trainiert die KI und lässt sie lernen.

Semantisch haben diese Texte kaum Bezug zu unserem Umgang mit Sprache, da ihre *Interpretation* in mathematischen Prozessen erfolgt. Dabei sind die Vorgänge der Transformation von Sprache zu Bildern durch Binärcodes und Algorithmen derart komplex und datenintensiv, dass die sogenannten „latenten Räume“ eines Diffusionsmodells oder Autoencoders bis zu 1024 Dimensionen nutzen, um ein Bild zu beschreiben.

Das Schreiben von Texten als instrumentelle Prompts wird zu einer neuen Form sprachlichen Ausdrucks – wenn auch nicht um der Sprache selbst willen, sondern mit dem Ziel, die gewünschten

¹ „Wenn Texte von Bildern verdrängt werden, dann erleben, erkennen und bewerten wir die Welt und uns selbst darin nicht mehr linear, prozessual, historisch, sondern als Fläche, als Kontext, als Szene.“ Vilém Flusser, *Ins Universum der technischen Bilder*. Göttingen 1985.

Bilder und Bildeffekte und ihre Vermarktbarkeit zu optimieren. Expert*innen, die verlässlich prompten können werden, dringend gesucht, bestens bezahlt und beginnen, Fotograf*innen zu ersetzen. Sprache wird zum Werkzeug, um möglichst wahrscheinliche, anwendbare Bilder zu kreieren.

Mir diese latenten Räume vorzustellen, gelingt mir, sooft ich es auch versuche, genauso wenig wie viele der internen Vorgänge innerhalb dieser Super-Mega-Blackbox. Mir helfen dabei künstlerische Experimente, die ich als wichtig ansehe, um die rasante Entwicklung der KI besser zu begreifen.

Was dabei herauskommen kann, ist im besten Fall „Kunst über KI“, mit Sicherheit aber keine „Kunst mit KI“ oder gar „KI-Kunst“. Denn der für die Kunst-Definition entscheidende Unterschied liegt in der Haltung gegenüber dem Werkzeug.

Eine „Kunst über KI“ impliziert die kritische Auseinandersetzung mit Bedingungen, Logiken und Machtstrukturen der KI-Technologien. Eine solche Praxis reflektiert die Funktionsweisen der Systeme, ihre impliziten Ideologien, aber auch deren ökonomische Infrastrukturen. Im Sinne von Flusser lässt sich dies als ein Arbeiten „mit dem Apparat gegen den Apparat“ verstehen. Der künstlerische Prozess erschöpft sich nicht im Bedienen eines Programms, sondern versucht dessen vorgegebene Programme zu unterlaufen und offenzulegen. Statt die von Algorithmen vorgeschlagenen Bildwelten zu reproduzieren, verschiebt, irritiert oder dekonstruiert eine solche Praxis deren ästhetische und semantische Automatismen.

Ein mögliches Experiment in diesem Kontext stellt meine Arbeit *Rotkäppchen* dar, die ich als eine ungefragte Kollaboration mit Jacob Grimm, Wilhelm Grimm und der *Luma AI Dream Machine* sehe. Das Experiment wird zweiteilig präsentiert – als ein Kurzfilm und als eine Reihe von Standbildern, die mit dem Text der jeweiligen Prompts untertitelt sind.

1. Video

Zur Generierung der kurzen Clips wurde die KI nicht nur mit den üblichen Bildbeschreibungen oder handlungsanweisenden Prompts gefüttert, sondern auch mit dem originalen Prosatext der Gebrüder Grimm, aus dem Jahr 1812 der ebenfalls auf der Ebene der Prompts eingesetzt wurde. Der originale Text wurde im genauen Wortlaut abschnittsweise in je ein bis drei Sätzen separiert und als Prompt in die *Luma AI Dream Machine* eingesetzt. Aus jedem Textabschnitt generierte die KI eine Videosequenz von fünf Sekunden Länge. Für den finalen Film wurden die generierten Videotracks dem Text folgend montiert.

Beim Editieren wurde keine Auswahl der einzelnen Sequenzen getroffen, keine der Sequenzen wurde ein weiteres Mal generiert. Alle Clips wurden mit einem kostenlosen User-Account, also ohne Abonnement und auf KI-Laien-Niveau erstellt.

Beide Methoden des Promptens, also die übliche Bildbeschreibung als Handlungsanweisung bzw. der unveränderte Prosatext werden im Video montiert und gegenübergestellt. Im finalen sind die chronologisch montierten Sequenzen mit einem Voiceover vertont. Der Schauspieler Frank Smilies hat dafür den Originaltext des Grimm'schen Märchens *Rotkäppchen* eingesprochen. Die Verwendung der menschlichen Stimme emotionalisiert die Abfolge der synthetisierten Bewegtbilder.

2. Reihe von Stillbildern / Prompts

Dem Video gegenüber gestellt werden Standbilder aus den Clips, die ausschließlich mit dem originalen Märchentext geprompted wurden. Sie sind unternitelt mit dem jeweilig eingegebenen Prompt. Die Montage zeigt auf, wie Text und Bild interagieren.

Innerhalb dieses Konzepts wurden der KI keine weiteren optisch-gestalterischen Parameter vorgegeben, wie es in Verfahren mit „Text zu Bild“, durch Anweisungen wie „Gestalte im Stil von...“ möglich und üblich ist. Es wurden keine weiteren Hinweise über Prompts gegeben, wie sie den Originaltext zu interpretieren hat oder mit welcher emotionalen Wirkung auf den Betrachter sie die Sequenzen generieren soll, szenische Anweisungen wie die Kameraperspektive, Kamerafahrt, Objektive, Beleuchtung wurden nicht definiert. Eine Ausnahme bilden die immer vorangestellten Attribute „hyperrealistic“, „photorealistic“ und „4K“, um eine kindlich-naiv-zeichnerische Darstellung zu vermeiden. Die KI durfte also den Text „selbst interpretieren“.

Das Experiment stellt beide Verfahren des „Promptens“ einander als Narrativ und als expositorische Anweisung gegenüber. Das ist ein syntaktischer Shift, der die Semantik der Bilder bestimmt. Textliche Handlungsanweisungen werden mit Prosa-Texten kombiniert bzw. ausgetauscht. In beiden Fällen wird Text/Sprache zum maschinenproduzierten (Bewegt-)Bild. Ziel des Experiments ist es, die neue Möglichkeit des Promptens als Handlungsanweisung durch Texte (textbasierte KI-Bildgenerierung) im Hinblick auf Sprache künstlerisch zu beleuchten.