Wahrscheinlichkeit und Blickordnung

Generative KI als Spiegel gesellschaftlicher Ambivalenzen

Analytische Befunde zu generativen Bild- und Videosystemen

Günter Schaden

Unabhängiger Befundbericht

Version 1.0 · 2026

Inhaltsverzeichnis

Executive Summary

Generative Bild- und Videosysteme erzeugen Inhalte auf Grundlage statistischer Wahrscheinlichkeiten. Diese Wahrscheinlichkeiten spiegeln die Trainingsverteilungen wider, auf denen die Modelle beruhen. Bestimmte Darstellungen von Personen, ihre Konstellationen, Körperhaltungen und Bildordnungen sind in diesen Daten dichter repräsentiert als andere.

Das vorliegende Dossier untersucht, wie sich diese Verteilungen in der Darstellung von Personen im Output bemerkbar machen. Ausgangspunkt ist eine bewusst reduzierte Versuchsanordnung: eine offene, skizzenhafte Szene eines jungen Paares, fortgeschrieben unter strengen formalen Einschränkungen. Trotz minimaler Vorgaben zeigte sich eine Stabilisierung der Darstellung. Linien wurden kohärenter, Bewegungen lesbarer, Kontraste deutlicher.

Diese Verschiebung wird hier nicht als ästhetische Verbesserung verstanden, sondern als Ausdruck probabilistischer Struktur. Was im Trainingsmaterial häufig vorkommt, besitzt eine höhere Chance, erneut erzeugt zu werden. Diese strukturelle Bevorzugung wird im Dossier als Normalisierung bezeichnet: statistisch dominante Muster erscheinen im generierten Bild als vertraut und plausibel – als selbstverständlicher als weniger dicht repräsentierte Alternativen.

Die Analyse zeigt, dass diese Logik nicht nur Motive betrifft, sondern auch Wahrnehmungsordnungen. Fixierte Kamerapositionen, rhythmisch stabilisierte Bewegungen und partielle Hervorhebungen erzeugen implizite Betrachterpositionen. Aufmerksamkeit wird nicht explizit angeordnet, aber ihre Wahrscheinlichkeit wird strukturell erhöht.

Dabei entsteht eine Ambivalenz: Die dargestellte Szene bleibt motivisch harmlos, während die Wahrnehmung gebunden wird. Diese Gleichzeitigkeit verweist auf eine tiefere Ebene. Generative Systeme reproduzieren nicht nur einzelne Inhalte, sondern verdichten statistisch dominante Bildordnungen. In ihnen spiegeln sich kulturelle Verteilungen von Sichtbarkeit – einschließlich der Spannungen, die mit ihnen verbunden sind.

Das Dossier versteht generative KI daher nicht als moralisches Subjekt, sondern als probabilistischen Verdichtungsmechanismus. Es fragt, wie Trainingsdichte, Wahrscheinlichkeit und Wahrnehmung zusammenwirken – und wie sich kulturell vertraute Bildmuster im scheinbar neutralen Output wiedererkennen lassen.

Kapitel 1

Ausgangspunkt und Versuchsanordnung

Der Analyse ging ein bewusst einfach gehaltener Prompt voraus. Die Ausgangsszene sollte keine Dramaturgie entfalten, sondern einen Moment im Vorübergehen festhalten. Ziel war es, die Ernsthaftigkeit eines verliebten Paares – ihre leichte Selbstbezogenheit, ihre Konzentration aufeinander – mit einem humorvollen Unterton darzustellen.

Der Prompt lautete in verkürzter Form:

A rough pencil storyboard sketch of a young couple walking through a fairground.
Drawn quickly while observing a moving scene.

Es sollte sich um eine skizzenhafte, suchende Darstellung handeln:

The drawing is unfinished and searching.
Lines overlap and follow the act of looking, not the object.
Multiple loose construction lines are visible.

Explizit ausgeschlossen wurden stilistische Verdichtungen:

No clean outlines.
No shading gradients.
No polished details.
White paper background.
Graphite pencil only.
No cinematic composition.

Auch inhaltlich war die Szene bewusst reduziert. Die Frau blickt zum Mann, der Mann wendet den Kopf leicht zu einem entfernten Luftballonverkäufer. Keine dramatische Interaktion, keine zugespitzte Emotion, kein Konflikt:

The couple appears in motion, slightly blurred through repeated sketch lines.
The woman looks at the man.
The man's head is slightly turned toward a balloon vendor in the distance.

Wichtig war dabei die Formulierung:

Drawn before decisions are made.

Die Szene sollte den Charakter eines vorläufigen Beobachtungsmoments behalten – noch nicht festgelegt, noch nicht entschieden.

Die anschließende Iteration war ebenso restriktiv formuliert. Sie enthielt keine neue inhaltliche Setzung, sondern lediglich eine zeitliche Verschiebung:

Keep the same characters, same clothing, same proportions, same camera angle, same background, and same rough pencil storyboard style.

Und weiter:

Imagine this scene approximately 5 minutes later while they continue walking naturally through the fairground.

Dabei wurde ausdrücklich gefordert:

Do not change composition.
Do not introduce new objects.
Do not redesign the characters.
Preserve sketch quality and line structure.

Die Entwicklung sollte sich ausschließlich als natürliche Zeitprogression innerhalb derselben formalen und räumlichen Konstellation vollziehen.

Gerade diese Einschränkungen sind für die folgende Analyse entscheidend. Die später beschriebenen Verschiebungen entstanden nicht aus einer dramaturgischen Vorgabe, sondern innerhalb eines bewusst begrenzten Rahmens. Das Experiment war kein Versuch, eine erotische Szene zu erzeugen oder Aufmerksamkeit gezielt zu lenken, sondern eine Untersuchung dessen, was geschieht, wenn minimale Angaben von einem probabilistischen System fortgeschrieben werden.

Aus dieser Versuchsanordnung ergibt sich die zentrale Fragestellung: Welche Strukturen treten hervor, wenn Wahrscheinlichkeit eine offene Szene weiterrechnet?

1. Ausgangsbild, Fortschreibung und Trainingsverteilung

Ausgangspunkt der Untersuchung war eine bewusst reduzierte Versuchsanordnung. Ein kurzer Prompt beschrieb eine skizzenhafte Szene: ein junges Paar, das über einen Jahrmarkt geht. Die Darstellung sollte unvollständig bleiben, mit überlagernden Linien, ohne Ausarbeitung, ohne klare ästhetische Entscheidung. Die anschließende Iteration sollte lediglich eine zeitliche Fortschreibung zeigen – bei unveränderter Komposition, unverändertem Stil und ohne Einführung neuer Elemente.

2. Das Ausgangsbild

Im generierten Ausgangsbild zeigte sich das Paar in einer klar lesbaren visuellen Konstellation, obwohl viele Merkmale nicht spezifiziert worden waren.

Die weibliche Figur war schlank gezeichnet, mit langen Haaren und einem schulterfreien, körpernahen Top. Ihr Oberkörper war zum Mann hin orientiert, der Kopf nach oben geneigt. Der Blick war auf ihn gerichtet. Die Linienführung im Gesicht ließ einen weichen, zugewandten Ausdruck erkennen.

Die männliche Figur trug einen Hoodie und hatte kurzes Haar. Sein Körper war leicht von der Partnerin weg orientiert; sein Blick ging nach vorne, seine Aufmerksamkeit war hauptsächlich auf den Ballonverkäufer gerichtet.

Beide Figuren wurden als weiß dargestellt.

Diese Merkmale – Körperbau, Kleidung, Hautfarbe – waren im Prompt nicht festgelegt. Sie entstanden unter minimaler Spezifikation.

Zwischen den Figuren bestand eine subtile Asymmetrie: Nähe war angedeutet, aber nicht vollständig gespiegelt. Die überlagernden Linien der Skizze ließen mehrere mögliche Fortsetzungen offen. Das Bild enthielt Spannung, aber keine Festlegung.

Abbildung 1 - Ausgangsbild

3. Die Videoiteration

In der anschließenden Videoiteration blieb die Komposition formal unverändert. Kamera, Bildausschnitt und Figurenpositionen wurden beibehalten. Die Veränderung lag in der zeitlichen Organisation.

Die zuvor offenen Linien wurden in Bewegung überführt. Der Gang der weiblichen Figur stabilisierte sich: Die Schritte erschienen klein und gleichmäßig gesetzt. Über mehrere Frames hinweg verlagerte sich das Körpergewicht in nahezu identischer Amplitude von einer Seite zur anderen. Die seitliche Bewegung des Beckens wiederholte sich rhythmisch.

Zugleich wurde die Jeans der weiblichen Figur unterhalb der Hüfte deutlich blau eingefärbt, während der obere Bereich skizzenhaft blieb. Dieser Kontrast war im Ausgangsbild nicht vorhanden.

Die Bewegung reduzierte Mehrdeutigkeit. Was im statischen Bild als mögliche Lesarten nebeneinander bestehen konnte, wurde im zeitlichen Verlauf sequenziell festgelegt.

Abbildung 2 - Standbild aus der Videoiteration

Videoiteration der beschriebenen Szene:

https://www.guenter-schaden.at/ki-dossier-video

4. Trainingsverteilung und Wahrscheinlichkeit

Um diese Verschiebung zu verstehen, ist die Funktionslogik generativer Systeme entscheidend. Bild- und Videomodelle erzeugen Inhalte auf Grundlage statistischer Wahrscheinlichkeiten. Diese Wahrscheinlichkeiten spiegeln die Trainingsverteilungen wider, auf denen die Modelle beruhen.

Bestimmte Konstellationen – etwa typische Paarhaltungen, vertraute Kleidungskombinationen oder gängige Bewegungsmuster – können im Trainingsmaterial dichter vertreten sein als andere. Statistische Dichte erhöht die Wahrscheinlichkeit ihrer Reproduktion.

Die im Ausgangsbild entstandene visuelle Konstellation sowie die im Video beobachtete Stabilisierung lassen sich vor diesem Hintergrund als Ausdruck solcher Gewichtungen verstehen. Unter minimaler Spezifikation werden jene Varianten wahrscheinlicher, die im Modell stärker repräsentiert sind.

Wahrscheinlichkeit betrifft daher nicht nur die Auswahl eines Motivs, sondern auch dessen Fortführung. Trainingsverteilungen formen Sichtbarkeit – und beeinflussen, welche Konstellationen stabiler erscheinen als andere.

Kapitel 2

Der Default-Raum als statistische Verdichtung

Wenn Wahrscheinlichkeit Ausdruck einer Trainingsverteilung ist, dann ist der generierte Bildraum kein neutraler Möglichkeitsraum. Er ist gewichtet. Manche Konstellationen sind im Trainingsmaterial häufiger vertreten als andere. Diese Häufigkeit beeinflusst, welche Varianten unter minimaler Spezifikation erscheinen.

Ein einfaches Beispiel verdeutlicht diesen Mechanismus. Der Begriff „couple“ ist semantisch offen. Er legt weder Geschlecht noch Alter noch Beziehungsform eindeutig fest. Dennoch zeigte sich in den Iterationen eine Tendenz: Ohne weitere Spezifikation generierte das System wiederholt eine heterosexuell gelesene Konstellation eines jungen Mannes und einer jungen Frau.

In technischen Systemen bezeichnet „default“ eine Voreinstellung, auf die zurückgegriffen wird, wenn keine abweichenden Parameter gesetzt werden. Übertragen auf generative Bildsysteme lässt sich hier von einem Default-Raum sprechen: einem statistisch verdichteten Möglichkeitsbereich, in den unspezifizierte Eingaben mit erhöhter Wahrscheinlichkeit fallen.

Diese Wiederkehr ist kein Beweis für normative Setzung, sondern legt nahe, dass entsprechende Konstellationen im Trainingsmaterial dichter vertreten sind als andere.

Der Default-Raum entsteht dort, wo minimale Eingaben auf häufig repräsentierte Bildkonstellationen treffen. Unspezifizierte Begriffe fallen nicht in einen neutralen Durchschnitt, sondern in jene Varianten, die im Trainingsmaterial besonders oft kombiniert und wiederholt vorkommen. Wahrscheinlichkeit wirkt hier als Rückfallmechanismus in statistisch dominante Muster.

In den durchgeführten Varianten mit veränderten Geschlechterkonstellationen waren alternative Paarformen grundsätzlich erzeugbar.

Sie zeigten jedoch in der Fortschreibung eine größere Varianz. Die Darstellung schwankte stärker zwischen nebeneinander her Gehen und betonter Nähe; Abstand und Berührung veränderten sich deutlicher, und die Blickrichtung der Figuren war weniger konstant aufeinander bezogen als in der Ausgangskonstellation.

Abbildung 3 - Variante 1

Abbildung 4 - Variante 2

Auffällig ist dabei nicht nur die Wahl der Paarform, sondern auch die Modalität körperlicher Nähe. In den vorliegenden Iterationen erscheinen Frauenpaare durch explizites Handhalten verbunden, während Nähe zwischen Männern über seitliche Armberührungen innerhalb einer parallelen Bewegungsachse entsteht. Die Kontaktzonen unterscheiden sich damit sichtbar: einmal als klar markierte Beziehungsgeste, einmal als beiläufige körperliche Annäherung im gemeinsamen Gehen.

Auch der formal offene Begriff „jung“ wurde in den Iterationen in geschlechtlich klar markierten Erscheinungsformen konkretisiert. In den vorliegenden Iterationen erschienen weibliche Figuren konsistent mit mindestens schulterlangem Haar und Pony, während männliche Figuren durchgehend kurzes Haar trugen. Der Prompt spezifizierte weder Frisur noch Stilrichtung. Die Differenz lässt sich als Ausdruck eines statistisch verdichteten Bildraums lesen, in dem Jugend geschlechtlich codiert erscheint.

Diese Beobachtungen legen nahe, dass Trainingsdichte nicht nur Häufigkeit, sondern auch strukturelle Stabilität beeinflusst. Je häufiger eine Konstellation im Trainingsmaterial vorkommt, desto stärker sind ihre Regelmäßigkeiten im Modell repräsentiert. In der Generierung kann sich dies als geringere Varianz bei der Fortführung zeigen.

Der Default-Raum ist daher nicht bloß eine Ansammlung häufiger Motive. Er beschreibt jene Konstellationen, die unter Variation konsistenter fortgeschrieben werden als andere. Diese Konsistenz wirkt im Output wie Selbstverständlichkeit.

Entscheidend ist dabei: Der Default-Raum ist keine bewusste Setzung des Modells. Er ist das Resultat einer gewichteten Trainingsverteilung. Was im Trainingsmaterial häufig vorkam, wird probabilistisch wahrscheinlicher reproduziert.

Damit verschiebt sich die Analyse weiter. Wenn statistisch dominante Konstellationen stabiler fortgeschrieben werden, betrifft dies nicht nur Figuren und Abstände, sondern auch Blickachsen und Bewegungsrichtungen. Der Default-Raum bereitet so jene Wahrnehmungsordnungen vor, die später als Default-Blick beschrieben werden.

Kapitel 3

Von der Verteilung zur Bewegung

Die im vorangegangenen Kapitel beschriebene Trainingsdichte betrifft zunächst statische Konstellationen: Figuren, Abstände, Blickrichtungen. Mit der Einführung von Bewegung verschiebt sich die Fragestellung. Nun geht es nicht mehr nur darum, welche Konstellation erscheint, sondern wie sie im zeitlichen Verlauf fortgeschrieben wird.

Generative Videomodelle erzeugen keine isolierten Einzelbilder. Sie berechnen Übergänge. Zwischen zwei Zeitpunkten dürfen keine abrupten Inkonsistenzen entstehen. Bewegungsabläufe müssen über mehrere Frames hinweg anschlussfähig bleiben. Diese Sequenzanforderung macht Regelmäßigkeiten besonders sichtbar.

Im beschriebenen Beispiel zeigte sich dies im Gang der Figur. Die ursprünglich mehrfach überlagerten Linien der Skizze ließen offen, wie das Körpergewicht im nächsten Moment verteilt sein würde. In der generierten Bewegung wurde diese Offenheit reduziert. Die Schritte erschienen klein und gleichmäßig, die Gewichtsverlagerung erfolgte in einer rhythmisch wiederholbaren Seitwärtsbewegung. Die Verschiebung des Beckens blieb über mehrere Frames konsistent.

Diese Stabilisierung lässt sich als Ausdruck der zugrunde liegenden Wahrscheinlichkeitslogik verstehen. Bewegungssequenzen, die im Modell gut repräsentiert sind, können unter minimaler Spezifikation konsistenter fortgeführt werden als ungewöhnliche oder abrupt wechselnde Abläufe. Die Beobachtung legt nahe, dass vertraute Gangmuster im Generierungsprozess eine geringere Varianz aufweisen.

Die Stabilisierung betrifft dabei nicht nur grobe Körperpositionen, sondern auch feine Bewegungsdetails wie Schrittgröße, Taktung und Amplitude der Gewichtsverlagerung. Bewegung wirkt dadurch nicht dramatischer, sondern berechenbarer.

Im statischen Bild können mehrere Lesarten nebeneinander bestehen. In der Bewegung hingegen wird jede Entscheidung sequenziell fortgesetzt. Wahrscheinlichkeitsverteilungen treten hier deutlicher hervor als im Einzelbild, weil jede Variation Anschluss finden muss. Was im Modell stärker verankert ist, erscheint im zeitlichen Verlauf konsistenter.

Bewegung macht damit sichtbar, wie Verteilungen wirken. Sie übersetzt statistische Gewichtung in rhythmische Struktur. Erst in dieser sequenziellen Organisation wird deutlich, dass Wahrscheinlichkeit nicht nur auswählt, sondern fortschreibt.

An dieser Stelle berührt die Analyse erstmals die Wahrnehmungsebene. Regelmäßige Bewegung erzeugt Erwartbarkeit. Erwartbarkeit strukturiert Aufmerksamkeit. Doch bevor dieser Zusammenhang explizit entfaltet wird, ist die räumliche Organisation der Szene genauer zu betrachten.

Kapitel 4

Aufmerksamkeit und Betrachterposition

Im statischen Ausgangsbild war die Position des Betrachters nicht festgelegt. Die Skizze zeigte das Paar im Gehen, doch sie definierte keinen eindeutig fixierten Standpunkt. Der Blick konnte sich im Bild frei orientieren. Da keine Bewegung stattfand, gab es keinen zeitlichen Zwang, einem bestimmten Element zu folgen.

In der Videoiteration veränderte sich diese Situation grundlegend. Die Kamera blieb statisch. Der Betrachterstandpunkt war festgelegt. Die Figuren bewegten sich aus diesem festen Blickfeld heraus nach vorne, vom Betrachter weg.

Diese Konstellation hat eine spezifische Wirkung: Der Betrachter bleibt räumlich unbeweglich, während sich die Figuren entfernen. Der Blick kann nicht mitgehen, er bleibt an derselben Stelle im Raum verankert. Dadurch durchqueren bestimmte Körperregionen wiederholt das Zentrum des Sichtfelds.

Im konkreten Beispiel wurde die Gewichtsverlagerung des Körpers – insbesondere im Bereich des Beckens – durch die rhythmische Stabilisierung der Bewegung klarer lesbar. Da die Figuren sich vom fixierten Betrachter wegbewegten, wurde genau diese Bewegung in der zentralen Perspektive sichtbar.

Hier zeigt sich eine Verschiebung gegenüber dem statischen Bild. Nicht nur das Motiv wird generiert, sondern auch eine räumliche Beziehung zwischen Szene und Betrachter. Die Fixierung der Kamera erzeugt eine implizite Zuschauerposition: einen ruhenden Beobachtungspunkt inmitten einer sich bewegenden Szene.

In der Wahrnehmungspsychologie wird beschrieben, dass Bewegung und rhythmische Wiederholung besonders aufmerksamkeitsstark sind. Diese Eigenschaft – Salienz – bedeutet, dass bestimmte Reize automatisch hervortreten. In Verbindung mit einer fixierten Perspektive kann daraus eine stabile Blickbindung entstehen.

Die Irritation im beschriebenen Beispiel entstand nicht aus einer expliziten inhaltlichen Setzung, sondern aus dieser Konstellation: statischer Betrachter, sich entfernende Figuren, rhythmisch stabilisierte Bewegung im zentralen Sichtfeld. Die Aufmerksamkeit wurde nicht befohlen, aber strukturell wahrscheinlicher gelenkt.

Damit erweitert sich der Begriff der Normalisierung um eine weitere Dimension. Wahrscheinlichkeit beeinflusst nicht nur, welche Bilder erscheinen, sondern auch, welche Wahrnehmungsposition sie nahelegen. Der Default-Raum wird zum Default-Blick.

Kapitel 5

Wahrnehmung und Ambivalenz

Im Ausgangsbild blieb die Szene offen. Die Frau blickte zum Mann auf, leicht lächelnd, ihr Körper ihm zugewandt. Der Mann war nach vorne orientiert. Diese Asymmetrie erzeugte Spannung, ohne sie festzulegen. Die überlagerten Skizzenlinien ließen mehrere mögliche Fortsetzungen zu. Nähe war angedeutet, aber nicht eindeutig erwidert.

In der Videoiteration veränderte sich nicht die Grundkonstellation, sondern ihre Bestimmtheit. Der Gang stabilisierte sich. Die unterhalb der Hüfte blau eingefärbten Jeans erzeugten einen Kontrast innerhalb der Figur.

Auffällig ist, dass die Einfärbung ausschließlich die Jeans der weiblichen Figur betrifft, obwohl beide Personen vergleichbare Kleidung tragen. Der Farbkontrast etabliert damit eine visuelle Hierarchie innerhalb der Szene, indem eine Figur farblich akzentuiert wird, während die andere monochrom bleibt. Eine rein funktionale Erklärung – etwa im Sinne einer Bewegungsstabilisierung – würde diese Asymmetrie nicht zwingend erwarten lassen. Die Selektivität der Hervorhebung wirft daher die Frage auf, ob neben technischen Parametern auch statistisch verdichtete Bildmuster wirksam werden.

Zugleich bewegten sich weitere Figuren quer durch den Raum. Diese Überlagerung erzeugte eine zusätzliche visuelle Ebene, während die Kamera selbst unbeweglich blieb. Der Bildrahmen veränderte sich nicht; der Betrachterstandpunkt blieb fixiert. Die rhythmische Seitwärtsbewegung durchquerte wiederholt denselben zentralen Bildbereich.

Aus der Verschränkung von rhythmischer Wiederholung, farblichem Kontrast und räumlicher Fixierung entsteht eine spezifische Wahrnehmungssituation. Die eingefärbte Jeans im Beckenbereich sowie die rhythmisch stabilisierte Bewegung lenken Aufmerksamkeit wiederholt auf denselben Körperbereich.

Die Szene bleibt motivisch harmlos – zwei Personen gehen nebeneinander her. Gleichzeitig wird der Blick strukturell gebunden. Bestimmte Zonen werden wahrscheinlicher wahrgenommen als andere.

Ambivalenz bedeutet hier die Gleichzeitigkeit zweier Ebenen: einer beiläufigen Alltagsszene und einer formalen Organisation, die einen kulturell sensiblen Körperbereich hervorhebt.

Diese Ambivalenz betrifft sowohl die Position des Betrachtens als auch die des Betrachtetwerdens. Die wiederholte Bewegung innerhalb eines fixierten Rahmens kann als visuelle Hervorhebung erfahren werden; die eigene Fixierung auf einen bestimmten Bildbereich kann als unfreiwillige Blickbindung erlebt werden. Beides entsteht aus der Organisation von Raum, Rhythmus und Kontrast, nicht aus einer expliziten Setzung.

Vor dem Hintergrund der zuvor beschriebenen Trainingsverteilungen wird diese Ambivalenz verständlich. Wenn bestimmte Konstellationen im Trainingsmaterial dichter repräsentiert sind, erhöht sich ihre Wahrscheinlichkeit der Reproduktion und ihrer konsistenten Fortführung. Die formale Stabilität, die im Video sichtbar wird, kann als Ausdruck solcher Gewichtungen gelesen werden.

Das System erzeugt keine neue Spannung. Es macht sichtbar, was in der Wahrscheinlichkeitsstruktur bereits angelegt ist.

Schlusskapitel

Generative Systeme als Spiegel gesellschaftlicher Verteilungen

Generative Bild- und Videosysteme erscheinen häufig als eigenständige Produzenten von Wirklichkeit. Die Analyse dieses Dossiers legt jedoch eine andere Perspektive nahe. Die Modelle erzeugen keine kulturellen Muster aus sich heraus. Sie berechnen Wahrscheinlichkeiten auf Grundlage von Trainingsdaten – also auf Grundlage von Bildern, die zuvor von Menschen produziert, verbreitet und betrachtet wurden.

Im konkreten Beispiel zeigte sich unter minimaler Spezifikation eine schlanke, als weiß lesbare Konstellation. Während Alter („jung“) und Paarform durch den Prompt definiert waren, waren Körperbau, Kleidung und Hautfarbe nicht spezifiziert. Sie entstanden im Ausgangsbild als statistisch plausible Konstellation.

In der Videoiteration stabilisierte sich diese Darstellung weiter: kleine, gleichmäßige Schritte, rhythmische Gewichtsverlagerung, partielle Farbbetonung. Was zunächst offen und skizzenhaft war, wurde im zeitlichen Verlauf konsistenter fortgeführt.

Diese Entwicklung kann als Ausdruck einer zugrunde liegenden Verteilung verstanden werden. Bestimmte Körperbilder, Paarhaltungen und Bewegungsrhythmen sind in großen Bildarchiven typischerweise häufiger repräsentiert als andere. Unter minimaler Spezifikation erhöht sich daher die Wahrscheinlichkeit, dass solche Konstellationen erscheinen und konsistent fortgeschrieben werden.

Der Spiegel zeigt sich genau hier: Nicht das Modell setzt eine Norm, sondern es reproduziert statistische Gewichtungen. Was im Trainingsmaterial dichter vertreten ist, erscheint im Output stabiler und selbstverständlicher. Die generierte Szene wirkt beiläufig und zugleich strukturiert, weil sie auf wiederholten Mustern basiert.

Die im Video erfahrene Ambivalenz – die Gleichzeitigkeit von motivischer Harmlosigkeit und struktureller Blickbindung – verweist auf diesen Zusammenhang. Die Irritation entsteht nicht aus einer expliziten Botschaft, sondern aus der Wahrnehmung von Gewichtung. Wiederholung erzeugt Vertrautheit; statistisch verdichtete Bildordnungen werden seltener als Auswahl erkannt und eher als naheliegende Form gelesen.

In diesem Sinn fungieren generative Systeme als Spiegel gesellschaftlicher Sichtbarkeiten. Sie zeigen keine autonome Setzung, sondern die statistische Struktur dessen, was zuvor vielfach produziert wurde. Ihre Bilder machen sichtbar, welche Konstellationen häufiger repräsentiert sind und dadurch wahrscheinlicher erscheinen.

Die Irritation entsteht nicht, weil etwas Neues eingeführt wird, sondern weil Verteilung sichtbar wird.

Die weiterführende Frage lautet daher nicht, was das Modell über sich selbst aussagt, sondern was diese Verteilungen über die Bildordnungen verraten, die wir als selbstverständlich akzeptieren. Wenn generative Systeme statistische Gewichtungen berechnen, spiegeln sie nicht nur Daten, sondern auch wiederkehrende Muster kollektiver Sichtbarkeit.

Die interessante Frage ist somit: Was sagt die Häufigkeit bestimmter Konstellationen über die visuellen Präferenzen, Gewohnheiten und Ausschlüsse einer Gesellschaft aus, deren Bilder in Trainingsdaten eingehen?

Weiterführende Perspektiven

Die folgenden Arbeiten berühren angrenzende Fragestellungen zu Trainingsdaten, Modelllogiken und Normalisierung.

Kate Crawford & Trevor Paglen – Excavating AI: The Politics of Images in Machine Learning Training Sets (2021)

https://doi.org/10.1007/s00146-021-01162-8

Timnit Gebru et al. – Datasheets for Datasets (2018)

https://arxiv.org/abs/1803.09010

Pranav Seshadri et al. – The Bias Amplification Paradox in Text-to-Image Generation (2024)

https://aclanthology.org/2024.naacl-long.353

Emily M. Bender et al. – On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? (2021)

https://doi.org/10.1145/3442188.3445922