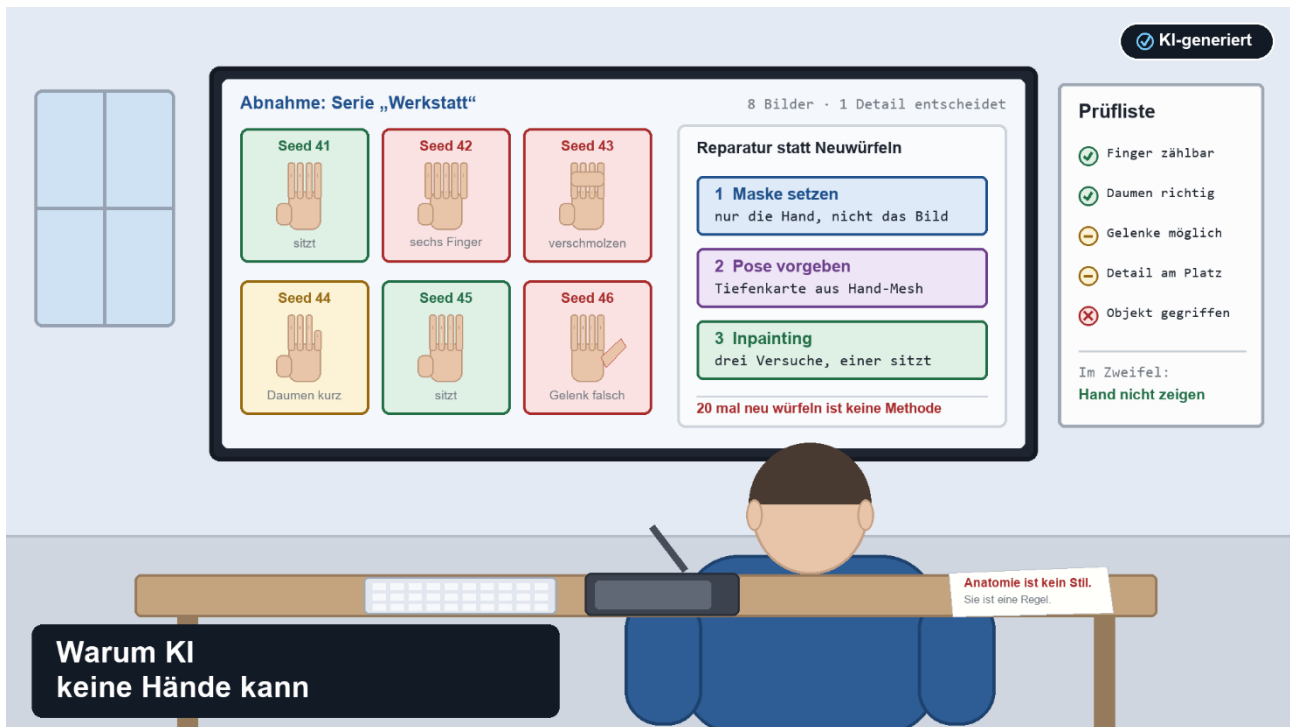


Warum KI keine Hände kann



Sechs Versuche, zwei brauchbare Hände: So sieht eine ehrliche Abnahmerunde aus. Der Ausweg ist nicht der siebte Versuch, sondern die Maske auf der kaputten Stelle.

Wissen	Beratung	Schulung
Alle Grundlagen zur KI-Bildgenerierung an einer Stelle – vom Diffusionsprinzip bis zur Governance. KI-Bildgenerierung	Wir bauen dir einen Bildprozess, in dem Anatomiefehler auffallen, bevor sie beim Kunden ankommen. KI-Bildgenerierung Beratung	Dein Team lernt Maskieren, Posensteuerung und Inpainting statt Dauerwürfeln. KI-Bildgenerierung Schulung

Es gibt in jedem Workshop diesen einen Moment. Wir haben zwei Stunden über Diffusion, Modelle und Prompts geredet, alle sind einigermaßen beeindruckt, und dann meldet sich jemand aus der zweiten Reihe: „Aber KI kann keine Hände, oder?“ Danach nicken alle, weil das der eine Satz über KI-Bilder ist, den wirklich jede und jeder kennt. Und dann muss ich die unbequeme Doppelantwort geben: Der Satz stimmt nicht mehr so pauschal wie vor zwei Jahren – und er beschreibt trotzdem ein Problem, das keiner der großen Anbieter bisher sauber gelöst hat.

Der Klassiker sind sechs Finger, verschmolzene Knöchel, ein Daumen, der auf der falschen Seite ansetzt, ein Gelenk, das sich in eine Richtung biegt, die kein Mensch überleben würde. Der zweite, weniger berühmte Fall ist die Platzierung: Das Tattoo, das an den rechten Unterarm sollte, sitzt am linken. Das Firmenlogo, das auf die Brusttasche gehört, klebt auf dem Ärmel. Beide Fehler haben dieselbe Wurzel, und wenn du die verstanden hast, weißt du auch, warum kein noch so eleganter Prompt sie zuverlässig wegzaubert.

Die Kurzfassung vorweg: Das Modell malt nach, was es oft gesehen hat. Es weiß nicht, dass eine Hand fünf Finger hat, es weiß nur, dass auf Bildern von Händen meistens ungefähr fünf fingerförmige Dinge zu sehen sind. „Meistens ungefähr“ ist bei einer Wolke kein Problem. Bei einer Hand schon. Dieser Beitrag zeigt dir die strukturellen Ursachen, nimmt die Platzierung als zweites

Beispiel dazu, ordnet ein, wo die Modelle im Herbst 2026 wirklich stehen, und liefert dir den Werkzeugkasten, mit dem du das Problem in der Produktion erledigst statt es zu beklagen.

Kurz gefasst

Hände sind kein Software-Bug, sondern die Folge von vier zusammenfallenden Eigenschaften: enormer Zustandsraum, winzige Bildfläche, häufige Verdeckung im Trainingsmaterial und ein Modell, das Muster reproduziert statt Regeln anzuwenden.

Exakte Platzierung – „rechter Unterarm“, „linke Brusttasche“ – scheitert an derselben Ursache: Das Modell kennt die Begriffe, aber kein Koordinatensystem des Körpers.

Die Modellgeneration seit FLUX.2 und Nano Banana Pro trifft einfache Handposen verlässlich. Bei verschränkten Fingern, zwei greifenden Händen und kleinen Objekten liegt die Fehlerquote weiterhin deutlich über null.

Der schnellste Fix bleibt unspektakulär: die Hand gar nicht zeigen. Der zuverlässigste ist Inpainting mit Maske – nicht das Bild neu würfeln, sondern die kaputte Stelle neu erzeugen.

Warum ausgerechnet Hände

Wenn ein Bildmodell eine Kaffeetasse malt, darf es sich erstaunlich viel erlauben. Henkel etwas zu dick, Ellipse minimal schief, Schatten physikalisch fragwürdig – niemand merkt es, weil es keine allgemein bekannte Sollform für eine Tasse gibt. Bei einer Hand gibt es die. Jeder Mensch hat seit dem ersten Lebensjahr Millionen Hände gesehen und trägt zwei davon mit sich herum. Das Toleranzfenster für Abweichungen ist bei Händen praktisch geschlossen, und genau deshalb fällt jeder Fehler sofort auf, der bei anderen Motiven durchgehen würde.

Vier Ursachen, die zusammenfallen

Jede für sich wäre beherrschbar. Alle vier zusammen ergeben das bekannteste Fehlerbild der Bildgenerierung.



Fazit der Skizze: Kein Prompt behebt eine dieser vier Ursachen. Werkzeuge, die Struktur vorgeben oder nachträglich reparieren, schon.

Vier Ursachen, die sich gegenseitig verstärken. Keine davon lässt sich durch bessere Formulierung im Prompt beheben – sie liegen in den Trainingsdaten und in der Funktionsweise des Modells.

Ursache 1: Der Zustandsraum ist gigantisch

Eine menschliche Hand besteht aus 27 Knochen und rund zwanzig beweglichen Gelenken. Jedes davon kann unabhängig gebeugt werden, dazu kommen Drehung des Handgelenks, Blickwinkel und perspektivische Verkürzung. Dieselbe Hand sieht ausgestreckt, zur Faust geballt, einen Stift

haltend oder halb gedreht aus wie vier völlig verschiedene Objekte. Ein Gesicht hat dagegen eine erstaunlich stabile Topologie: Zwei Augen über einer Nase über einem Mund, und das in praktisch jedem Bild der Welt. Das Modell kann bei Gesichtern auf einer zuverlässigen Struktur aufbauen. Bei Händen gibt es diese Struktur nicht, sondern eine fast unbegrenzte Menge gleichberechtigter Erscheinungsformen.

Ursache 2: Zu wenig Fläche, zu wenig Strafe

Auf einem typischen Porträt belegt das Gesicht leicht ein Fünftel der Pixel. Eine Hand liegt oft unter einem Prozent. Diese Fläche entscheidet doppelt. Beim Training bekommt das Modell aus jedem Bild proportional zur Fläche Signal – wenige Pixel bedeuten wenig Information über die feine Struktur. Und während der Bilderzeugung ist der Beitrag einer schlecht getroffenen Hand zur Gesamtbewertung minimal: Ein Bild mit falscher Fingerzahl ist rechnerisch kaum schlechter als eines mit korrekter. Deshalb hilft es messbar, wenn du das Bildformat so wählst, dass die Hand mehr Fläche bekommt – wie sich Seitenverhältnis und Kantenlänge auswirken, steht in [Auflösung & Seitenverhältnis bei KI-Bildern](#). Eine Nahaufnahme der Hände auf 1024×1024 gelingt deutlich häufiger als dieselbe Hand als Randdetail in einer Totalen.

Ursache 3: Im Trainingsmaterial sind Hände selten vollständig

Sieh dir die nächsten zwanzig Fotos an, die dir in einem sozialen Netzwerk begegnen. Wie viele davon zeigen eine komplett sichtbare, unverdeckte, scharfe Hand? In der Realität stecken Hände in Taschen, umschließen Gegenstände, liegen hinter dem Rücken, sind durch Bewegungsunschärfe verwischt oder vom Bildrand abgeschnitten. Das Modell hat also überproportional viele Beispiele gelernt, in denen Hände unvollständig, verdeckt oder unscharf sind. Es hat gewissermaßen gelernt, dass eine Hand auch ein diffuser fleischfarbener Klumpen sein darf – weil das im Trainingsmaterial oft genug so aussah.

Ursache 4: Nachahmen statt verstehen

Das ist der Kern, und er erklärt Hände und Platzierung in einem Aufwasch. Ein Diffusionsmodell hat kein Konzept von Anatomie. Es hat keine Regel gespeichert, die besagt: Eine Hand hat genau fünf Finger, davon einen opponierbaren Daumen, und Gelenke beugen sich nur in eine Richtung. Es hat gelernt, welche Pixelnachbarschaften in welchen Zusammenhängen wahrscheinlich sind. Fünf Finger sind für das Modell keine Notwendigkeit, sondern eine Häufigkeit – und Häufigkeiten darf man unterschreiten und überschreiten. Wie dieses Grundprinzip funktioniert, ist in [Basiswissen KI-Bildgenerierung: Wie Diffusion funktioniert](#) ausführlich beschrieben.

Diese vier Ursachen greifen ineinander. Ein großer Zustandsraum wäre lernbar, wenn genug Fläche und genug saubere Beispiele da wären. Wenig Fläche wäre verschmerzbar, wenn eine Regel die Struktur absichern würde. Erst die Kombination erzeugt das Fehlerbild, das inzwischen als Running Gag durchs Internet läuft. Und sie erklärt auch, warum die typischen Problemzonen alle derselben Beschreibung folgen: Zähne, Brillenbügel, Fahrradspeichen, Gitarrensaiten, Schmuckketten, Uhrzeiger. Klein, feingliedrig, mit einer Sollstruktur, die jeder kennt.

Der Zusammenhang, der alles erklärt

Überall dort, wo der Mensch eine exakte Sollstruktur im Kopf hat und das Motiv gleichzeitig klein und variabel ist, patzt das Modell sichtbar. Hände sind nur der prominenteste Fall dieser Regel, nicht der einzige. Wer das verstanden hat, weiß bei jedem neuen Motiv vorher, wo er hinschauen muss.

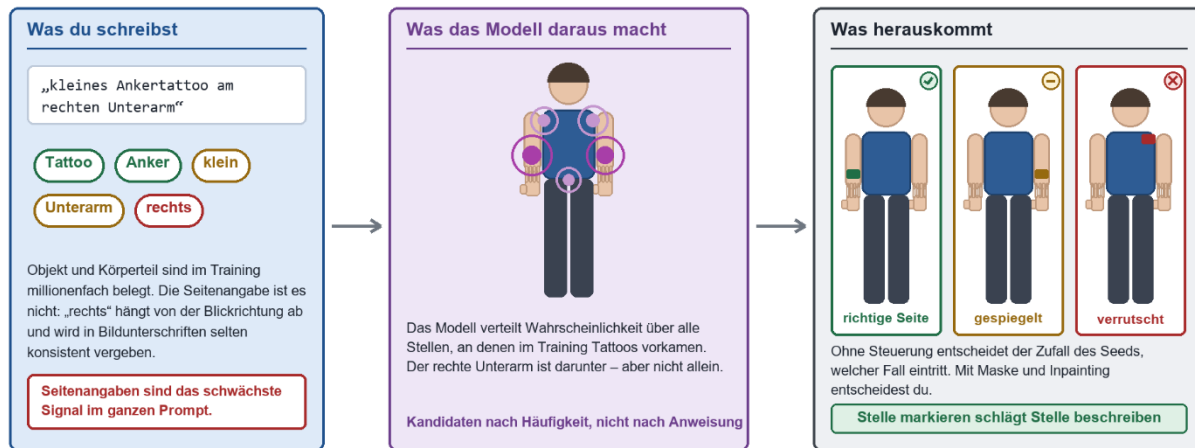
Das Tattoo am falschen Arm

Der zweite Fall wird seltener belacht, kostet in der Praxis aber mehr Zeit. Du schreibst „kleines Ankertattoo am rechten Unterarm“ und bekommst der Reihe nach: das Tattoo am linken

Unterarm, das Tattoo auf der Schulter, zwei Tattoos, kein Tattoo und irgendwann auch einmal das Richtige. Der Ärger darüber ist verständlich, die Ursache aber dieselbe wie bei den Händen: Es fehlt ein inneres Modell des Körpers.

Begriffe kennt das Modell, Koordinaten nicht

Dieselbe Ursache wie bei den Händen: Es gibt kein inneres Modell des Körpers, auf dem eine Stelle angesteuert werden könnte.



Dasselbe gilt für „Logo auf der linken Brusttasche“, „Ring am Ringfinger“ oder „Narbe über dem rechten Auge“: Sobald eine exakte Stelle gefragt ist, rät das Modell plausibel.

Der Ausweg ist immer derselbe – nicht besser formulieren, sondern die Fläche vorgeben.

Der Prompt liefert Begriffe, das Modell verteilt Wahrscheinlichkeit über alle plausiblen Stellen. Die Seitenangabe ist dabei das schwächste Signal im ganzen Satz.

Warum „rechts“ das schwächste Wort im Prompt ist

„Tattoo“ und „Unterarm“ sind im Training millionenfach belegt und eindeutig mit Bildinhalten verknüpft. „Rechts“ ist beides nicht. Es hängt von der Blickrichtung ab: Der rechte Arm der abgebildeten Person erscheint im Bild links. Bildunterschriften und Alternativtexte, aus denen die Modelle ihre Sprachverknüpfung lernen, handhaben das völlig uneinheitlich – mal aus Sicht der Person, mal aus Sicht der Kamera, meistens gar nicht. Das Modell hat also kaum verlässliches Signal dafür, was „rechts“ überhaupt bedeuten soll. Dieselbe Schwäche trifft „links neben“, „hinter“, „über“ und jede Form von Zählangabe. Warum Modelle Anweisungen generell nicht wie ein Befehl abarbeiten, ist in [Warum die KI den Prompt ignoriert](#) ausführlich beschrieben.

Was du stattdessen formulieren kannst

Ein bisschen lässt sich holen, bevor man zum Werkzeug greift. Beschreibe nicht die Koordinate, sondern die Situation: Eine Hand, die eine Tasse hält, ist eine Handlung mit klarer Form; eine Hand „mit gespreizten Fingern in die Kamera gehalten“ ist eine seltene Pose in ungünstiger Verkürzung. Verankere die Stelle an einem sichtbaren Bezugspunkt statt an einer Seitenangabe – „Tattoo direkt unterhalb der hochgekrempelten Ärmelkante“ funktioniert besser als „am rechten Unterarm“, weil die Ärmelkante im Bild existiert und „rechts“ nicht.

So schreiben es die meisten	Was daraus wird	Was besser funktioniert
„Tattoo am rechten Unterarm“	Seite zufällig, Größe zufällig	Stelle maskieren und gezielt hineinmalen
„Hand mit gespreizten Fingern“	seltene Pose, hohe Fehlerquote	„Hand umfasst den Griff der Tasse“
„hält ein Smartphone“	Finger verschmelzen mit dem Gerät	Objekt größer und Griff eindeutig beschreiben
„zwei Personen geben sich die Hand“	der klassische Fingerbrei	Bildausschnitt so wählen, dass der Griff nicht im Fokus liegt
„Logo auf der linken Brusttasche“	Logo verrutscht oder erscheint	Logo nachträglich als Ebene

	doppelt	einsetzen
„Ring am Ringfinger“	Ring wandert, Finger verschmelzen	Detail per Inpainting ergänzen

Der ehrliche Teil dieser Tabelle steht in der rechten Spalte: In der Hälfte der Fälle lautet die beste Antwort nicht „anders formulieren“, sondern „anders arbeiten“. Wie viel Einfluss Formulierung und Negativ-Prompt tatsächlich haben und wo die Grenze verläuft, steht in [Prompt und Negativ-Prompt in der KI-Bildgenerierung](#).

Wo die Modelle im Herbst 2026 stehen

Jetzt der Teil, bei dem der alte Running Gag zu korrigieren ist. Die pauschale Aussage „KI kann keine Hände“ war 2023 zutreffend, 2024 meistens, und sie ist heute zu grob. Die aktuelle Modellgeneration hat bei Anatomie deutlich zugelegt – aus drei Gründen: größere und besser kuratierte Trainingsdaten mit gezielt ergänztem Handmaterial, deutlich größere Modelle mit mehr Kapazität für feine Struktur, und Architekturen, die Text und Bild enger koppeln.

Die Lage bei den großen Modellen

Black Forest Labs hat FLUX.2 im November 2025 veröffentlicht: ein Rectified-Flow-Transformer, dessen offenes Basismodell 32 Milliarden Parameter mitbringt, dazu die API-Stufen pro, flex und max. Die kleineren, destillierten Klein-Modelle mit 4 und 9 Milliarden Parametern kamen am 15. Januar 2026 als offene Gewichte hinzu.

Lizenz ist kein Nebenthema: Nur die kleinste Klein-Variante steht unter Apache 2.0. Das 32B-Modell und die 9B-Klein-Variante laufen unter der nicht-kommerziellen FLUX-Lizenz – die erzeugten Bilder darfst du kommerziell nutzen, das Modell selbst in einem Produkt betreiben nur mit bezahlter Lizenz.

Google liefert Bilder über die Nano-Banana-Reihe; Nano Banana Pro erzeugt bis 4K und hält Referenzpersonen ohne Feinabstimmung stabil. Seit Februar 2026 ist Nano Banana 2 das Standardmodell in der Gemini-App und in der Google-Suche.

Alibabas Qwen-Image steht unter Apache 2.0 und ist damit die freizügigste Basis für eigene Feinabstimmung; Qwen-Image-Edit ergänzt seit August 2025 das instruktionsbasierte Bearbeiten, das in der Praxis häufig zuverlässiger auf freie Anweisungen reagiert als die Alternativen.

Midjourney rendert in Version 8.1 standardmäßig 2048 × 2048 Pixel, OpenAI hat mit ChatGPT Images 2.5 vor allem Latenz und Stiltreue über mehrere Bearbeitungsschritte verbessert.

Besser heißt nicht gelöst

Wenn ich heute mit einem aktuellen Modell eine Person zeige, die entspannt steht, sitzt, etwas Größeres hält oder die Hände sichtbar ruhen lässt, bekomme ich in der überwiegenden Zahl der Versuche brauchbare Hände. Das ist ein echter Unterschied zu früher. Sobald die Szene anspruchsvoller wird, kippt das Bild aber schnell: zwei Hände, die sich gegenseitig greifen. Finger um ein kleines Objekt, eine Nadel, ein Kabel, einen Kugelschreiber. Musikinstrumente, wo Fingerhaltung und Saiten gleichzeitig stimmen müssen. Handschuhe, Ringe, Verbände. Und alles, was gleichzeitig in Bewegung ist.

Ein zweiter Effekt wird gern übersehen: Die kleinen, schnellen Varianten verhalten sich anders als ihre großen Geschwister. Die destillierten Klein-Modelle liefern in wenigen Rechenschritten beeindruckende Bilder, produzieren aber spürbar häufiger Anatomiefehler. Wer dort von vier auf acht bis zwölf Schritte geht, sieht den größten Qualitätssprung genau bei Händen und Gesichtern – darüber hinaus verpufft der Effekt. Was Schrittzahl bewirkt und wo sie aufhört zu helfen, steht in [Sampling-Steps erklärt](#); welche Rolle Sampler und Scheduler dabei spielen, in [Sampler & Scheduler erklärt](#).

Szene	Aktuelle Großmodelle	Kleine, schnelle Varianten	Empfehlung
Person steht, Hände hängen locker	meist sauber	oft sauber	einfach prüfen, selten reparieren
Hand hält eine Tasse oder ein Buch	meist sauber	gemischt	kleiner Stapel, bestes Bild nehmen
Finger zeigen auf etwas	gemischt	häufig fehlerhaft	Pose vorgeben oder Ausschnitt ändern
Zwei Hände greifen ineinander	häufig fehlerhaft	praktisch immer fehlerhaft	Inpainting einplanen
Finger um kleines Objekt	häufig fehlerhaft	praktisch immer fehlerhaft	Objekt vergrößern oder verdecken
Instrument, Werkzeug, Tastatur	unzuverlässig	unbrauchbar	Ausschnitt so wählen, dass es nicht zählt
Detail an exakter Körperstelle	unzuverlässig	unzuverlässig	maskieren statt beschreiben

Iteratives Bearbeiten frisst Hände auf

Jeder weitere Bearbeitungsdurchgang – hochskalieren, Stil anpassen, Hintergrund tauschen, „mach es noch etwas heller“ – kostet Detailtreue genau dort, wo feine Strukturen liegen. Hände, Zähne, Schmuck und Saiten degradieren schneller als der Rest des Bildes. Wenn nach der dritten Korrekturrunde plötzlich ein Finger fehlt, liegt das nicht am Zufall, sondern am Verfahren. Konsequenz für die Praxis: Anatomie zuletzt prüfen, nach dem letzten Bearbeitungsschritt – nicht vorher.

Was bedeutet das für die Modellauswahl

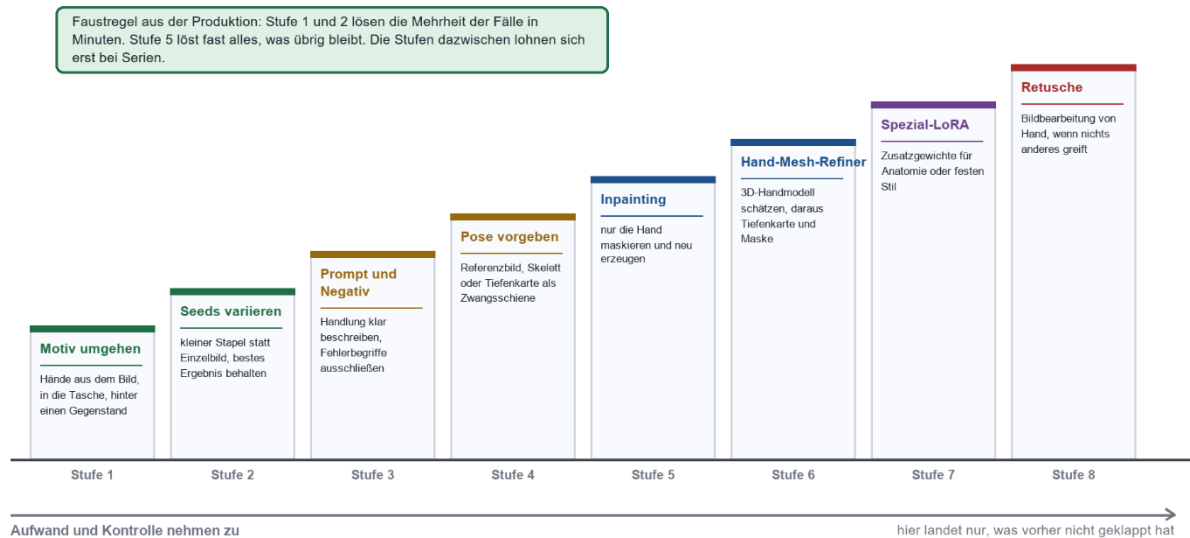
Ein Modell allein wegen der Hände auszuwählen, wäre unsinnig. Trotzdem gehört der Punkt in die Entscheidungsmatrix, wenn du regelmäßig Menschen zeigst. Große Modelle mit offenen Gewichten geben dir die Reparaturwerkzeuge in die Hand; geschlossene Dienste liefern oft schönere Bilder, lassen dich aber bei einem schiefen Detail im Regen stehen, weil du keine Maske setzen kannst. Der Vergleich der Werkzeuge steht in [Übersicht der wichtigsten KI-Bildgeneratoren](#) und [Die 15 wichtigsten Modelle zur lokalen Bildgenerierung](#); die grundsätzliche Abwägung zwischen Dienst und eigener Maschine in [Cloud vs. lokale KI-Bildgenerierung](#).

Der Werkzeugkasten: was wirklich hilft

Genug Theorie. In der Produktion zählt, wie du in möglichst wenigen Minuten zu einem Bild kommst, das durch die Abnahme geht. Ich arbeite mich dabei immer von unten nach oben durch die Leiter – nicht weil das elegant wäre, sondern weil die unteren Stufen den Großteil der Fälle erledigen und die oberen Zeit kosten.

Die Eskalationsleiter: erst ausweichen, dann steuern, dann reparieren

Von links nach rechts steigen Aufwand und Kontrolle. Wer oben anfängt, verbrennt Zeit; wer unten stehen bleibt, liefert kaputte Hände aus.



Acht Stufen, sortiert nach Aufwand. Die Stufen 1 und 2 lösen die Mehrheit der Fälle in Minuten, Stufe 5 erledigt fast alles, was übrig bleibt. Alles dahinter lohnt sich erst bei Serien.

Stufe 1 und 2: ausweichen und würfeln

Der meistunterschätzte Trick ist, das Problem gar nicht erst ins Bild zu lassen. Hände in die Jackentasche, hinter den Rücken, um einen großen Gegenstand, aus dem Ausschnitt heraus, hinter einem Tisch verdeckt. Kein gezeigtes Detail, kein Fehler. Bei Porträts und den meisten Unternehmensmotiven ist das keine Notlösung, sondern eine ganz normale fotografische Entscheidung. Die zweite Stufe ist fast genauso billig: Erzeuge nicht ein Bild, sondern einen kleinen Stapel mit verschiedenen Startwerten und nimm das beste. Die Zusammenhänge zwischen Startwert, Stapel und Wiederholbarkeit stehen in [Der Seed bei der KI-Bildgenerierung](#) und [batch_size und Reproduzierbarkeit](#). Wichtig ist nur die Disziplin: vier bis sechs Versuche, dann wird die Methode gewechselt.

Stufe 3: Prompt und Negativ-Prompt

Begriffe wie „verformte Hände“, „zusätzliche Finger“ oder „fehlerhafte Anatomie“ im Negativ-Prompt senken die Fehlerquote messbar – aber nur bei Modellen, die überhaupt einen echten Negativ-Prompt kennen. Klassische Diffusionsmodelle mit freier Guidance tun das. Viele destillierte und API-basierte Modelle nicht: Dort ist das Feld entweder gar nicht vorhanden oder es wird ignoriert, weil die Guidance fest eingebaut ist. Was genau dahintersteckt, erklärt [CFG-Scale: Guidance in KI-Bildgeneratoren](#). Und eine Warnung an dieser Stelle: Ein überladener Negativ-Prompt mit vierzig Begriffen macht das Bild flau, ohne die Hände zu retten.

Stufe 4: Pose vorgeben statt beschreiben

Hier wird es wirksam. Statt die Haltung zu beschreiben, gibst du sie als Bild vor: eine Strichfigur mit Gelenkpunkten, eine Tiefenkarte, eine Kantenzeichnung oder ein Referenzfoto. Das Modell bekommt damit eine Zwangsschiene, an der es sich entlanghangeln muss, und genau diese Zwangsschiene fehlt ihm bei Händen sonst. In lokalen Werkzeugketten sind diese Steuermodule Standard; in den großen Diensten übernimmt ein Referenzbild eine ähnliche Rolle, wenn auch mit weniger Kontrolle. Für Serien mit wiederkehrenden Motiven ist das die Stufe, ab der du planbar wirst statt hoffend.

Stufe 5: Inpainting, der eigentliche Arbeitsferd-Fix

Wenn ich in Projekten genau eine Technik mitgeben darf, dann diese. Du behältst das fertige Bild, markierst nur den kaputten Bereich – die Hand, das Tattoo, den Ring – und lässt allein diesen Bereich neu erzeugen. Der Rest bleibt unberührt. Das ist schnell, es ist wiederholbar, und es beantwortet zugleich das Platzierungsproblem: Eine markierte Fläche ist eine Koordinate, und Koordinaten versteht das Verfahren sehr wohl. Für das Tattoo heißt das: Du malst es nicht per Beschreibung an die richtige Stelle, du markierst die Stelle und lässt es dort entstehen.

Inpainting-Rezept, das bei mir fast immer funktioniert

Maske großzügig setzen: Die Hand plus ein bis zwei Zentimeter Handgelenk. Zu knappe Masken erzeugen sichtbare Nähte.

Für den maskierten Bereich einen eigenen, kurzen Prompt schreiben – „entspannte Hand, fünf Finger, greift den Griff“ – statt den kompletten Bildprompt zu wiederholen.

Vier Varianten mit verschiedenen Startwerten erzeugen und nebeneinander vergleichen, statt einzeln zu bewerten. Der Unterschied springt im Vergleich sofort ins Auge.

Am Ende auf 100 Prozent Zoom die Übergangskante prüfen. Wenn dort ein Farbsprung sitzt, war die Maske zu eng oder die Stärke zu hoch.

Stufe 6: Hand-Mesh-Refiner

Die spezialisierte Variante davon ist seit einigen Jahren im lokalen Ökosystem etabliert und immer noch die beste Antwort auf hartnäckige Fälle: Ein Handerkennungsmodell schätzt aus dem misslungenen Bild ein dreidimensionales Handmodell, leitet daraus eine saubere Tiefenkarte und eine passgenaue Maske ab, und genau diese Tiefenkarte steuert anschließend das Inpainting. Das Verfahren wurde als HandRefiner bekannt und steckt heute in den gängigen Knotenbibliotheken für lokale Werkzeugketten. Es repariert Hände, an denen freies Inpainting scheitert, weil es dem Modell die Handstruktur nicht beschreibt, sondern vorgibt.

Stufe 7 und 8: LoRA und Retusche

Es gibt kleine Zusatzgewichte, die gezielt auf Handanatomie trainiert sind und sich an ein Basismodell anhängen lassen. Sie helfen, sie sind aber kein Allheilmittel und sie ziehen häufig den Bildstil mit – wie diese Technik arbeitet und was sie kostet, steht in [Was ist ein LoRA?](#). Für einen Stil, den du ohnehin trainierst, lohnt es sich, Handmotive bewusst in die Trainingsdaten aufzunehmen. Und ganz oben auf der Leiter steht die ehrlichste Option: Bildbearbeitung von Hand. Einen Finger wegzustempeln dauert zwei Minuten und ist manchmal schlicht der kürzeste Weg. Es gibt keinen Preis dafür, ein Problem ausschließlich mit KI gelöst zu haben.

Die Seed-Lotterie ist kein Arbeitsprozess

Das häufigste Muster, das ich in Teams sehe: Jemand würfelt zwanzig, dreißig, fünfzig Mal denselben Prompt in der Hoffnung auf die perfekte Hand. Das ist teuer, es ist nicht reproduzierbar, und es endet oft damit, dass aus Erschöpfung ein mittelmäßiges Bild freigegeben wird. Setz dir ein hartes Budget – bei mir sind es vier Versuche, dann wird maskiert oder das Motiv geändert.

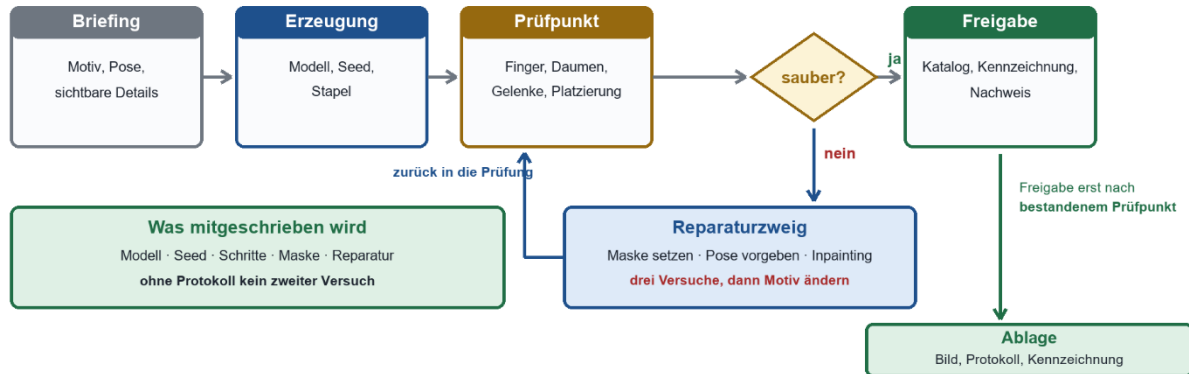
Aus dem Ärgernis einen Prozess machen

Der Unterschied zwischen einem Team, das sich über kaputte Hände ärgert, und einem, das sie beherrscht, liegt nicht im Modell. Er liegt darin, ob die Prüfung ein fester Punkt im Ablauf ist oder ein Zufallsfund in der Abnahme. Bei einem Kunden aus dem Maschinenbau ging eine Anzeige mit sechsfingriger Hand durch drei Freigaben und wurde erst vom Druckdienstleister bemerkt. Die

Nachbesserung hat einen Tag gekostet, die internen Diskussionen danach eine Woche. Seitdem gibt es dort eine Prüfliste mit vier Punkten, und das Thema ist erledigt.

Der Prüfpunkt gehört in den Ablauf, nicht in den Geschmack

Wer Hände erst in der Abnahme bemerkt, zahlt den Umbau doppelt: einmal in Zeit, einmal in Glaubwürdigkeit.



Drei Reparaturversuche sind das Budget. Wird es überschritten, ist nicht die Hand das Problem, sondern das Motiv.

Der Prüfpunkt sitzt nach dem letzten Bearbeitungsschritt und vor der Freigabe. Der Reparaturzweig hat ein festes Budget – wird es überschritten, ändert sich das Motiv und nicht die Geduld.

Die Prüfliste, die in jede Abnahme gehört

- Finger zählbar und einzeln abgrenzbar, an jeder sichtbaren Hand.
- Daumen auf der richtigen Seite, Handrücken und Handfläche konsistent zur Körperhaltung.
- Gelenke in physikalisch möglichen Winkeln, keine Beugung gegen die Gelenkrichtung.
- Geforderte Details an der geforderten Stelle: Logo, Tattoo, Ring, Abzeichen, Aufnäher.
- Griff plausibel – die Hand berührt das Objekt tatsächlich, statt daneben zu schweben.
- Symmetriefalle geprüft: Wenn beide Hände sichtbar sind, gehören sie zur selben Person und zeigen dieselbe Hauttextur.

Diese Liste kostet in der Anwendung dreißig Sekunden pro Bild. Sie ist damit billiger als jede nachträgliche Korrekturschleife, und sie hat einen Nebeneffekt: Sie macht die Bewertung überprüfbar. „Gefällt mir nicht“ ist keine Rückmeldung, mit der die Bildproduktion arbeiten kann. „Punkt 3 der Prüfliste, linke Hand“ schon.

Protokollieren, sonst war es Glück

Wenn ein Bild gut geworden ist, willst du wissen, warum. Modell, Startwert, Schrittzahl, Maskenbereich, verwendete Zusatzgewichte – ohne diese Angaben ist jedes gelungene Bild ein Einzelfall statt eines Rezepts. Wie eine Bildablage aussieht, die das mitschreibt und wiederauffindbar hält, steht in [Digital Asset Management für KI-Bilder mit SharePoint Online und Power Platform](#); wie der Erzeugungsteil dazu aufgebaut wird, in [Aufbau einer skalierbaren Bild-Maschine mit ComfyUI und Flux](#). In meinen eigenen Projekten übernimmt UB.Dossier den Katalog- und Nachweisteil: Was für ein Bild gilt, welche Werte gesetzt waren und wer es freigegeben hat, steht dort an einer Stelle und muss nicht aus Chatverläufen rekonstruiert werden. UB.Forge sitzt einen Schritt davor im Workflow und kippt eine Serie nicht kommentarlos in die Freigabe, sondern hängt den Plausibilitäts-Check zwischen Produktion und Abnahme.

Wenn das Bild nach außen geht

Zwei Dinge, die in der Aufregung um Finger gern untergehen. Erstens: Ein veröffentlichtes Bild mit offensichtlichem Anatomiefehler wird sofort als KI-Bild erkannt, und zwar unabhängig davon, wie gut der Rest ist – kaputte Hände sind der bekannteste Verräter überhaupt, wie [KI-generierte Bilder erkennen](#) zeigt. Zweitens: Die Kennzeichnungspflichten aus der KI-Verordnung sind seit dem 2. August 2026 anwendbar, für die maschinenlesbare Markierung läuft eine Frist bis zum 2. Dezember 2026. Was davon für dich gilt, hängt an deiner Rolle – die Einordnung steht in [EU AI Act: Was seit dem 2. August 2026 gilt](#). Ein sauber repariertes Bild entbindet dich nicht von der Kennzeichnung, und eine korrekte Kennzeichnung rettet kein Bild mit sechs Fingern.

Wer Bilder in Serie und markenkonform produziert, sollte den Prüfpunkt ohnehin nicht isoliert betrachten – wie ich das in meiner eigenen Produktion handhabe, steht in [Markenkonsistente KI-generierte Bilder](#) und [Dogfooding: Ich setze ein, was ich berate](#). Und wenn du den ganzen Ablauf einmal mit jemandem durchgehen willst, der die Stolpersteine kennt: dafür gibt es die [KI-Bildgenerierung Beratung](#) und die [KI-Bildgenerierung Schulung](#).

Fazit

„KI kann keine Hände“ ist als Satz zu einfach geworden, aber das dahinterliegende Problem ist nicht verschwunden. Die Ursachen sind strukturell: ein gewaltiger Zustandsraum, zu wenig Bildfläche, überwiegend verdeckte Beispiele im Training und ein Verfahren, das Muster reproduziert statt Regeln zu befolgen. Dieselben Ursachen erklären, warum das Tattoo am falschen Arm landet und das Logo auf dem Ärmel. Die aktuellen Modelle haben das deutlich abgemildert, und bei einfachen Posen musst du dir kaum noch Gedanken machen. Gelöst ist es nicht.

Die praktische Konsequenz ist unspektakulär und funktioniert seit Jahren: ausweichen, wo es geht, reparieren, wo es nötig ist, und aufhören zu würfeln. Ein Bildausschnitt ohne Hände kostet nichts. Eine Maske über der kaputten Hand kostet zwei Minuten. Dreißig Versuche auf der Suche nach dem perfekten Zufall kosten einen Nachmittag und liefern kein besseres Ergebnis. Wenn du aus diesem Beitrag genau eine Sache mitnimmst, dann diese: Die Stelle markieren schlägt die Stelle beschreiben – bei Händen genauso wie bei jedem anderen Detail, das genau dort sitzen soll, wo du es hinhaben willst.

Häufige Fragen

Warum kann die KI keine Hände?

Weil vier Dinge zusammenkommen. Eine Hand hat 27 Knochen und rund zwanzig bewegliche Gelenke, sieht also in jeder Stellung anders aus. Sie belegt auf typischen Bildern weniger als ein Prozent der Fläche und liefert dem Modell damit wenig Lernsignal. Im Trainingsmaterial ist sie meist verdeckt, verkürzt oder unscharf. Und das Modell arbeitet mit Wahrscheinlichkeiten statt mit Regeln: Fünf Finger sind für es kein Gesetz, sondern eine Häufigkeit.

Warum landet mein Tattoo an der falschen Stelle?

Weil das Modell kein Koordinatensystem des Körpers hat. „Tattoo“ und „Unterarm“ sind eindeutig belegt, „rechts“ ist es nicht – die Seitenangabe hängt von der Blickrichtung ab und wird in Bildbeschreibungen uneinheitlich verwendet. Das Modell verteilt die Wahrscheinlichkeit über alle plausiblen Stellen und trifft mal richtig, mal daneben. Der verlässliche Weg ist Inpainting: Stelle markieren, Detail dort erzeugen. Warum Anweisungen generell nicht wie Befehle ausgeführt werden, steht in [Warum die KI den Prompt ignoriert](#).

Werden die Modelle nicht immer besser bei Händen?

Ja, und zwar erheblich. Die Generation seit FLUX.2 und Nano Banana Pro trifft einfache Posen verlässlich – stehende Personen, ruhende Hände, große Gegenstände. Bei verschränkten Fingern, zwei ineinandergreifenden Händen, kleinen Objekten und Instrumenten liegt die Fehlerquote weiterhin deutlich über null. Das Grundproblem ist abgemildert, nicht behoben. Verlass dich also nicht darauf, dass es „inzwischen einfach geht“.

Was ist der schnellste Weg zu sauberen Händen?

Die Hände nicht zeigen. In der Tasche, hinter dem Rücken, um einen großen Gegenstand, hinter einem Tisch oder außerhalb des Ausschnitts. Das ist keine Kapitulation, sondern eine gewöhnliche gestalterische Entscheidung. Müssen die Hände sichtbar sein, ist Inpainting mit Maske der zuverlässigste Fix – vier Varianten erzeugen, die beste nehmen, fertig.

Hilft ein Negativ-Prompt gegen kaputte Hände?

Er senkt die Fehlerquote, garantiert aber nichts. Begriffe wie „zusätzliche Finger“ oder „verformte Hände“ wirken nur bei Modellen mit echtem Negativ-Prompt; viele destillierte und API-basierte Modelle haben dieses Feld gar nicht oder ignorieren es. Und überladene Negativlisten verschlechtern das Gesamtbild, ohne die Hand zu retten. Details dazu in [Prompt und Negativ-Prompt in der KI-Bildgenerierung](#) und [CFG-Scale: Guidance in KI-Bildgeneratoren](#).

Gibt es spezielle Modelle oder Zusätze für bessere Hände?

Ja. Es gibt Zusatzgewichte, die gezielt auf Handanatomie trainiert sind und sich an ein Basismodell anhängen lassen; sie verbessern die Trefferquote, ziehen aber oft den Stil mit. Wie diese Technik funktioniert, erklärt [Was ist ein LoRA?](#). Spezialisierter und meist wirksamer ist ein Hand-Mesh-Refiner: Er schätzt ein dreidimensionales Handmodell aus dem misslungenen Bild und steuert damit das Inpainting.

Warum sind die Hände nach dem Hochskalieren wieder kaputt?

Weil jeder weitere Durchgang feine Strukturen neu interpretiert. Hochskalieren, Stiländerungen und Nachbearbeitungsschritte kosten Detailtreue zuerst dort, wo die Strukturen klein sind: Hände, Zähne, Schmuck, Schrift. Prüfe Anatomie deshalb immer nach dem letzten Bearbeitungsschritt, nicht vorher – und repariere in derselben Reihenfolge.

Hilft eine höhere Auflösung?

Indirekt ja, aber nicht so, wie die meisten denken. Entscheidend ist nicht die Gesamtgröße, sondern wie viel Fläche die Hand im Bild bekommt. Eine Nahaufnahme gelingt deutlich häufiger als dieselbe Hand als kleines Detail in einer Totalen. Wer einfach nur die Kantenlänge hochdreht, bekommt eher Bildfehler als bessere Finger – warum das so ist, steht in [Auflösung & Seitenverhältnis bei KI-Bildern](#).

Warum klappt es bei einem Bild und beim nächsten nicht?

Weil der Startwert das gesamte Rauschmuster festlegt, aus dem das Bild entsteht. Ein günstiger Startwert legt zufällig eine Struktur an, die zu einer sauberen Hand führt, ein anderer nicht. Genau deshalb lohnt ein kleiner Stapel statt eines Einzelbildes – und genau deshalb musst du den Startwert notieren, wenn ein Bild gelungen ist. Mehr dazu in [Der Seed bei der KI-Bildgenerierung](#) und [batch size und Reproduzierbarkeit](#).

Was ist mit Füßen, Zähnen und Brillen?

Dasselbe Muster, nur weniger berühmt. Überall dort, wo ein Motiv klein und variabel ist und der Mensch trotzdem eine exakte Sollstruktur im Kopf hat, wird es schwierig: Zehen, Zahnreihen,

Brillenbügel, Fahrradspeichen, Gitarrensaiten, Uhrzeiger, Treppengeländer. Die Gegenmittel sind identisch – verdecken, maskieren, Struktur vorgeben.

Lohnt sich für gute Hände ein Modellwechsel?

Nur selten als alleiniger Grund. Wichtiger als das Modell ist, ob du Reparaturwerkzeuge hast: Ein Dienst mit schöneren Bildern, bei dem du keine Maske setzen kannst, ist im Zweifel die schlechtere Wahl als ein etwas nüchterneres Modell mit vollem Zugriff. Die Gegenüberstellung steht in [Cloud vs. lokale KI-Bildgenerierung](#) und [Übersicht der wichtigsten KI-Bildgeneratoren](#).

Wie halte ich das im Team nachvollziehbar?

Mit einer kurzen Prüfliste in der Abnahme und einem Protokoll zu jedem freigegebenen Bild. Die Prüfliste braucht fünf bis sechs Punkte und dreißig Sekunden pro Bild: Finger zählbar, Daumen richtig, Gelenke möglich, Detail am Platz, Griff plausibel. Das Protokoll hält Modell, Startwert, Schrittzahl und Reparaturschritte fest. Beides zusammen verwandelt Glück in ein Rezept.

Passend zum Thema

- › Warum Modelle Anweisungen generell nicht wie Befehle ausführen: [Warum die KI den Prompt ignoriert](#)
- › Formulierung, Negativ-Prompt und ihre Grenzen: [Prompt und Negativ-Prompt in der KI-Bildgenerierung](#) · [CFG-Scale: Guidance in KI-Bildgeneratoren](#)
- › Wiederholbarkeit und Stapel statt Dauerwürfeln: [Der Seed bei der KI-Bildgenerierung](#) · [batch_size und Reproduzierbarkeit](#)
- › Rechenaufwand und Qualität feiner Strukturen: [Sampling-Steps erklärt](#) · [Sampler & Scheduler erklärt](#)
- › Bildfläche, Format und warum Details untergehen: [Auflösung & Seitenverhältnis bei KI-Bildern](#)
- › Zusatzgewichte gezielt einsetzen: [Was ist ein LoRA?](#) · Basismodelle und Dekoder: [Checkpoints, Modelle und VAE](#)
- › Wie diese Bilder technisch entstehen: [Basiswissen KI-Bildgenerierung: Wie Diffusion funktioniert](#)
- › Welche Werkzeuge aktuell was liefern: [Übersicht der wichtigsten KI-Bildgeneratoren](#) · [Die 15 wichtigsten Modelle zur lokalen Bildgenerierung](#) · [Cloud vs. lokale KI-Bildgenerierung](#)
- › Produktion und Ablage im Unternehmen: [Aufbau einer skalierbaren Bild-Maschine mit ComfyUI und Flux](#) · [Digital Asset Management für KI-Bilder mit SharePoint Online und Power Platform](#) · [KI-Bildgenerierung im Unternehmen – Die strategische Alternative zur Cloud](#)
- › Kaputte Hände als Erkennungsmerkmal: [KI-generierte Bilder erkennen](#)
- › Kennzeichnung und Markenführung: [EU AI Act: Was seit dem 2. August 2026 gilt](#) · [Markenkonsistente KI-generierte Bilder](#) · [Dogfooding: Ich setze ein, was ich berate](#)
- › Bildarbeit direkt in Office: [Black Forest FLUX.2 in PowerPoint und Copilot](#)
- › Beratung zum Bildprozess: [KI-Bildgenerierung Beratung](#) · Schulung für das Team: [KI-Bildgenerierung Schulung](#)