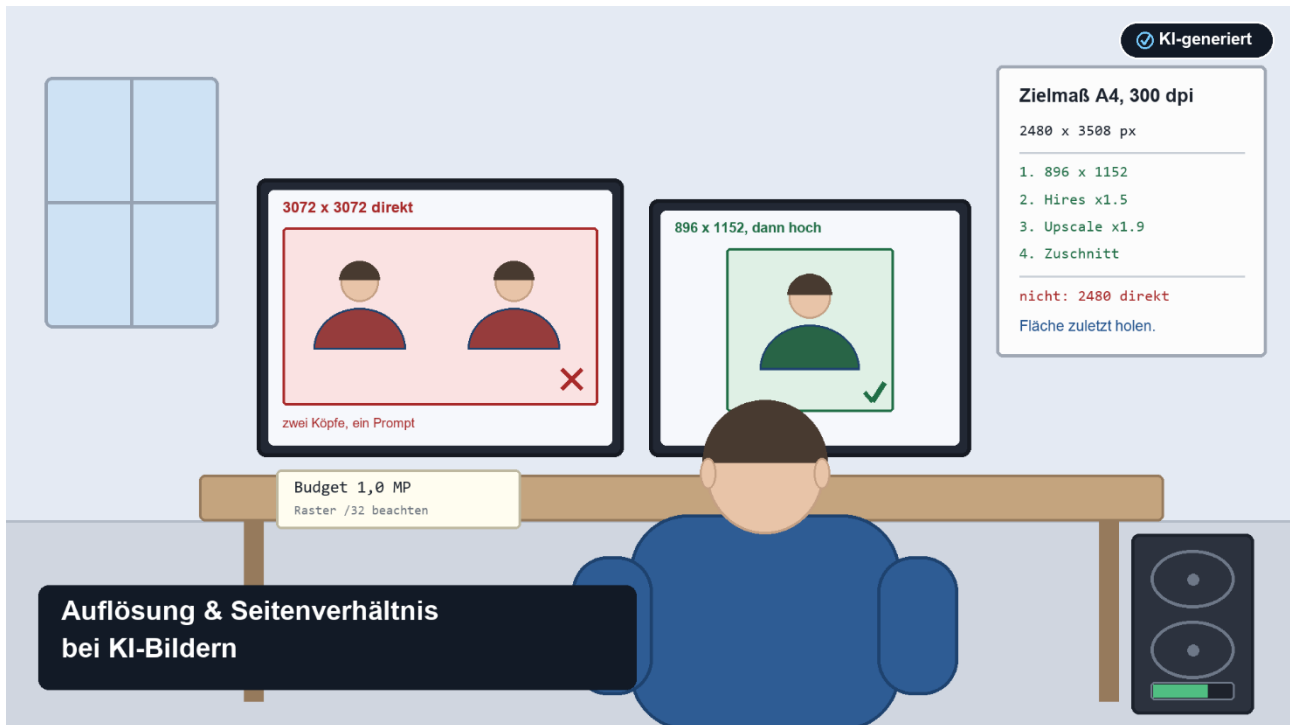


Auflösung & Seitenverhältnis bei KI-Bildern



Links das, was passiert, wenn man die Zahl einfach hochdreht. Rechts derselbe Prompt, in zwei Stufen gerechnet. Bild selbst erzeugt und als KI-generiert gekennzeichnet.

Wissen	Beratung	Schulung
Alle Grundlagen zur KI-Bildgenerierung an einer Stelle – vom Diffusionsprinzip bis zur Governance. KI-Bildgenerierung	Vom Einzelbild zur Formatmatrix, die für jeden Kanal reproduzierbar das richtige Maß liefert. KI-Bildgenerierung Beratung	Dein Team lernt, Fläche und Form auseinanderzuhalten – und den Upscaler statt der Brechstange zu nehmen. KI-Bildgenerierung Schulung

Es gibt in der KI-Bildgenerierung einen Fehler, den praktisch jeder genau einmal macht, und zwar in den ersten zwanzig Minuten. Du brauchst ein großes Bild. Im Werkzeug stehen zwei Felder für Breite und Höhe. Also tippst du eine große Zahl hinein, drückst auf Start, wartest deutlich länger als sonst – und bekommst ein Motiv mit zwei Köpfen. Oder drei Händen. Oder einem Horizont, der in der Bildmitte noch einmal von vorne anfängt.

Das ist kein Bug, kein Zufall und auch nicht dein Prompt. Das ist die logische Folge davon, wie diese Modelle gebaut und trainiert wurden. Wer das einmal verstanden hat, hört auf, gegen das Modell zu arbeiten, und bekommt saubere Bilder in exakt der Größe, die er braucht – nur eben auf einem anderen Weg als dem naheliegenden.

Auflösung und Seitenverhältnis sind zwei Regler, bei denen das Bauchgefühl besonders zuverlässig danebenliegt. Sie sehen aus wie Qualitätsregler, sind aber keine. Dieser Beitrag gehört zum Themenbereich [KI-Bildgenerierung](#) und klärt drei Dinge: was die native Auflösung im KI-Bildgenerator wirklich ist, warum zu groß generierte Bilder das Motiv verdoppeln – und wie du trotzdem an große, scharfe Ausgaben kommst, ohne dich mit dem Modell zu prügeln.

Fläche und Form: zwei Regler, die ständig verwechselt werden

Bevor wir über Doppelköpfe reden, muss die Begriffslage sauber sein. Denn die beiden Einstellungen tun grundverschiedene Dinge, und der Ärger entsteht fast immer daran, dass man den einen Regler benutzt, obwohl man den anderen gemeint hat.

Auflösung ist eine Fläche, keine Qualitätsstufe

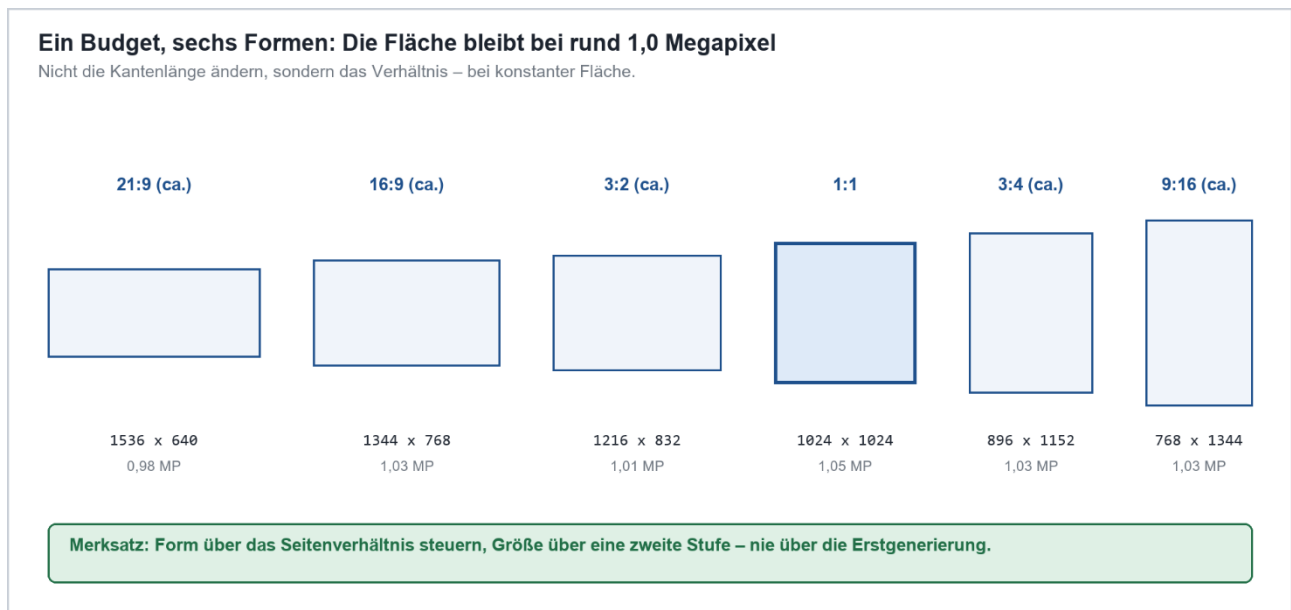
Die Auflösung ist schlicht die Größe des Bildes in Bildpunkten, etwa 1024×1024 Pixel. Entscheidend für die Generierung ist dabei nicht die einzelne Kantenlänge, sondern das Produkt aus beiden: die Gesamtfläche. 1024×1024 sind rund 1,05 Millionen Pixel, also etwa ein Megapixel. 1344×768 sind ebenfalls rund ein Megapixel, obwohl das Bild deutlich breiter ist. Für das Modell sind diese beiden Bilder gleich groß. Für dein Auge überhaupt nicht.

Das ist die wichtigste Denkhaltung im ganzen Thema: Denk in Fläche, nicht in Kantenlänge. Ein Modell, das bei einem Megapixel sauber arbeitet, arbeitet bei 1536×640 ebenfalls sauber, obwohl die lange Kante die 1024 deutlich übersteigt. Umgekehrt geht 1600×1600 schief, obwohl die Kante „nur“ anderthalbmal so lang ist – die Fläche ist nämlich zweieinhalbmal so groß.

Seitenverhältnis ist Bildsprache

Das Seitenverhältnis ist die Form, also das Verhältnis von Breite zu Höhe. 1:1 ist quadratisch, 3:4 und 2:3 sind Hochformate, 16:9 und 21:9 sind Querformate. Und hier passiert etwas, das viele unterschätzen: Das Seitenverhältnis ändert nicht nur den Ausschnitt, sondern die Komposition. Dasselbe Motiv, derselbe Prompt, derselbe Seed – im Hochformat rückt das Modell die Person nach vorn und zeigt weniger Umgebung, im Querformat macht es daraus eine Szene mit Raum drumherum.

Das ist gelerntes Verhalten. Porträts im Trainingsmaterial sind überwiegend hochformatig, Landschaften überwiegend quer. Das Modell hat gelernt, welche Art von Bild in welche Form gehört, und komponiert entsprechend. Wenn du also „Porträt“ promptest und ein 21:9-Format wählst, kämpfen zwei Signale gegeneinander – ein Fall, den auch [Warum die KI den Prompt ignoriert](#) behandelt. Das Format ist ein stiller Teil deines Prompts.



Skizze 1: Sechs Formen, eine Fläche. Alle sechs Rahmen liegen bei rund einem Megapixel – für das Modell sind sie gleich groß, für die Bildwirkung sind sie es überhaupt nicht.

Tipp: So rechnest du dir jedes Format selbst aus

Du brauchst keine Tabelle im Kopf. Das Budget in Pixeln nennen wir B (bei einem Megapixel also 1.048.576), das gewünschte Verhältnis $r = \text{Breite} / \text{Höhe}$. Dann ist die Breite die Wurzel aus B mal r, und die Höhe die Wurzel aus B geteilt durch r.

Beispiel 16:9 bei einem Megapixel: Wurzel aus $(1.048.576 \times 1,778)$ ergibt 1365, Wurzel aus $(1.048.576 \div 1,778)$ ergibt 768. Beide Werte auf das Raster des Modells runden – fertig: 1344×768 .

Danach immer die Gegenprobe: Breite mal Höhe teilen durch eine Million. Liegt das Ergebnis im Budget des Modells, passt es.

Ausgabegröße ist nicht Auflösung

Der dritte Begriff, der hier regelmäßig hineinrutscht: die Ausgabegröße in Zentimetern. Ein Bild hat keine physische Größe, solange es auf einem Bildschirm liegt – es hat nur Pixel. Erst beim Druck kommt die Punktdichte ins Spiel. Bei den üblichen 300 dpi braucht eine A4-Seite 2480×3508 Pixel, also fast neun Megapixel. Das ist ein Vielfaches dessen, was irgendein heutiges Modell in einem einzigen Durchgang sauber hinkriegt. Genau deshalb ist das zweistufige Vorgehen weiter unten kein Trick für Feinschmecker, sondern der Normalfall.

Zielformat	Benötigte Pixel	Fläche	In einem Lauf machbar?
Social-Beitrag quadratisch	1080×1080	1,2 MP	Ja, direkt im Budget
LinkedIn-Beitragsbild	1200×627	0,8 MP	Ja, direkt im Budget
Web-Hero, 16:9	1920×1080	2,1 MP	Nur bei FLUX.2 und Qwen-Image
PowerPoint-Vollbild, scharf	2560×1440	3,7 MP	Nur FLUX.2, sonst Upscaling
Flyer A5, 300 dpi	1748×2480	4,3 MP	Grenzfall, besser Upscaling
Seite A4, 300 dpi	2480×3508	8,7 MP	Nein – Upscaling zwingend
Plakat A2, 150 dpi	3508×4961	17,4 MP	Nein – Upscaling zwingend
Roll-up 85×200 cm, 100 dpi	3346×7874	26,3 MP	Nein – Upscaling zwingend

Wer regelmäßig für PowerPoint produziert, kennt die Frage aus dem Alltag: Das Bild sieht auf dem Notebook gut aus und fällt auf dem Beamer auseinander. Die Formatfrage entscheidet sich also lange vor dem Prompt. Wie sich das in der Office-Welt konkret anfühlt, zeigt das Briefing zu [Black Forest FLUX.2 in PowerPoint und Copilot](#).

Die native Auflösung: ein Pixelbudget, kein Lieblingswert

Jetzt zum Kern. Jedes Modell hinter einem KI-Bildgenerator wurde auf Bildern einer bestimmten Größe trainiert. Dieser Bereich heißt native Auflösung, und er ist der Grund, warum dasselbe Modell in manchen Größen brillant und in anderen unbrauchbar ist. Nur ist „native Auflösung“ ein etwas irreführender Begriff: Es handelt sich nicht um genau einen Wert, sondern um ein Flächenbudget mit einer Bandbreite an zulässigen Formen.

Wie das Budget zustande kommt

Diffusionsmodelle arbeiten nicht direkt auf Pixeln, sondern in einem komprimierten Latentraum – das erklärt [Basiswissen KI-Bildgenerierung: Wie Diffusion funktioniert](#) im Detail. Ein Autoencoder, der VAE, verkleinert das Bild vor dem Training um den Faktor 8 pro Kante. Aus 1024×1024 Pixeln werden also 128×128 Latent-Felder. Das eigentliche Modell sieht nie ein Pixel, sondern immer nur dieses Raster. Wie VAE und Modellgewichte zusammenhängen, steht in [Checkpoints, Modelle und VAE](#).

Auf diesem Raster hat das Modell gelernt, wie ein Gesicht aussieht, wie groß ein Kopf im Verhältnis zum Bild ist, wo ein Horizont liegt. Diese Größenverhältnisse sind fest in die Gewichte eingebrannt. Ein Kopf ist für ein SDXL-Modell rund 40 Latent-Felder breit – nicht „ein Drittel des Bildes“,

sondern eine absolute Größe im Raster. Gibst du ihm ein doppelt so großes Raster, macht es den Kopf nicht doppelt so groß. Es malt einen zweiten Kopf daneben.

Dass überhaupt mehrere Formen funktionieren, liegt an einem Trainingskniff namens Aspect-Ratio-Bucketing. Statt alles auf Quadrate zu beschneiden, sortiert man die Trainingsbilder in Behälter mit ähnlichem Seitenverhältnis, aber vergleichbarer Fläche. SDXL wurde auf einer solchen Liste feinjustiert, bei der die Gesamtzahl der Pixel jeweils bei oder unter 1.048.576 liegt – von 512×2048 bis 2048×512 . Deshalb gibt es nicht eine native Größe, sondern einen Satz nativer Größen mit gemeinsamem Flächenbudget.

Faktenkasten: Was die Zahlen wirklich bedeuten

Native Auflösung ist ein Flächenbudget, kein Kantenmaß. Maßgeblich ist Breite mal Höhe, nicht die längste Seite.

Der VAE komprimiert um den Faktor 8 pro Kante. Ein 1-Megapixel-Bild ist für das Modell ein Raster von etwa 128×128 Feldern.

SDXL wurde auf Seitenverhältnis-Behältern mit maximal 1.048.576 Pixeln trainiert. Die bekannten Werte 1152×896 , 1216×832 , 1344×768 und 1536×640 stammen genau aus dieser Liste – und sind bewusst nur Näherungen an 4:3, 3:2, 16:9 und 21:9.

Stability empfiehlt für Stable Diffusion 3.5 Large eine Gesamtfläche von rund einem Megapixel und Kantenlängen, die durch 64 teilbar sind. Die Medium-Variante deckt einen Bereich von 0,25 bis 2 Megapixeln ab.

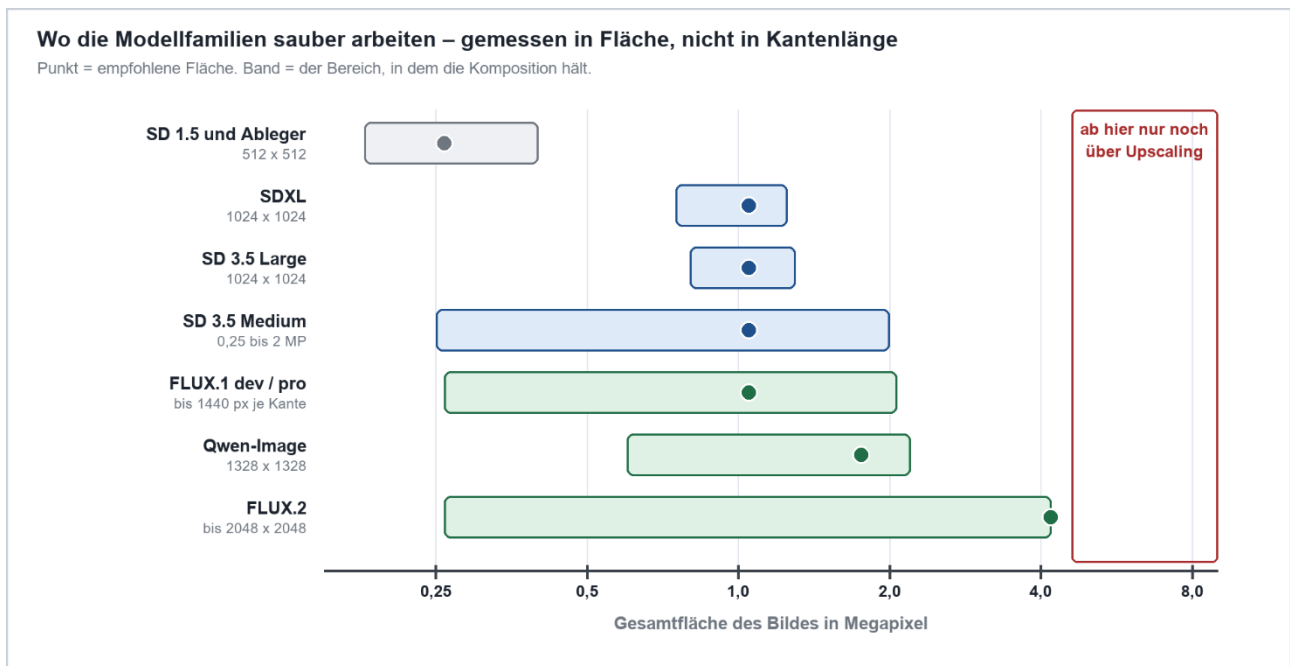
FLUX.1 dev und die Pro-Varianten akzeptieren Kantenlängen von 256 bis 1440 Pixeln. FLUX.2 erzeugt und bearbeitet nativ bis zu 4 Megapixel, also etwa 2048×2048 .

Qwen-Image ist auf 1328×1328 trainiert, also rund 1,76 Megapixel, mit passenden Varianten wie 1664×928 für 16:9.

Was heute je Modellfamilie gilt

Die pauschale Merkregel „ältere Modelle 512, neuere 1024“ trägt nicht mehr. Sie war richtig, solange alle Modelle derselben Architekturfamilie entstammten. Inzwischen liegen zwischen der kleinsten und der größten sinnvollen Fläche mehr als anderthalb Größenordnungen. Wer heute mit derselben Voreinstellung über alle Modelle fährt, verschenkt entweder Qualität oder produziert Ausschuss.

Modellfamilie	Empfohlene Fläche	Typische Maße	Raster	Obergrenze in einem Lauf
Stable Diffusion 1.5 und Ableger	ca. 0,26 MP	512×512 , 512×768	64	etwa 0,4 MP, danach Verdopplung
SDXL und darauf basierende Checkpoints	ca. 1,05 MP	1024×1024 , 1216×832	64	etwa 1,3 MP
Stable Diffusion 3.5 Large	ca. 1,0 MP	1024×1024 , 1344×768	64	etwa 1,3 MP
Stable Diffusion 3.5 Medium	0,25 bis 2,0 MP	1024×1024 bis 1536×1280	64	2,0 MP laut Hersteller
FLUX.1 dev, FLUX.1.1 pro	ca. 1,0 bis 1,5 MP	1024×1024 , 1344×768	32	1440 px je Kante, also 2,07 MP
Qwen-Image	ca. 1,76 MP	1328×1328 , 1664×928	32	etwa 2,2 MP
FLUX.2 dev, flex, pro	1,0 bis 4,2 MP	1024×1024 bis 2048×2048	32	4 MP nativ, auch beim Bearbeiten



Skizze 2: Dasselbe Diagramm, das jede Diskussion über „welche Größe nehme ich denn jetzt“ in zehn Sekunden beendet. Der Punkt ist die empfohlene Fläche, das Band der Bereich, in dem die Komposition noch hält.

Bei den geschlossenen Diensten sieht die Sache anders aus, weil sie die native Auflösung gar nicht erst offenlegen. Sie geben stattdessen feste Stufen vor. gpt-image-1 kennt genau drei Maße: 1024×1024 , 1536×1024 und 1024×1536 . Nano Banana Pro, das Bildmodell auf Gemini-3-Basis, liefert gestaffelt in 1K, 2K und 4K bei einer festen Liste von Verhältnissen von 1:1 bis 21:9. Midjourney, seit dem 24. Juli 2026 standardmäßig in Version 8.2, trennt ebenfalls zwischen Standardausgabe und einem HD-Modus mit deutlich höherer Auflösung. In allen drei Fällen macht der Anbieter die Zweistufigkeit intern für dich – du bekommst sie nur nicht zu sehen. Der Überblick über die Anbieterlandschaft steht in [Übersicht der wichtigsten KI-Bildgeneratoren](#), die offenen Modelle behandelt [Die 15 wichtigsten Modelle zur lokalen Bildgenerierung](#).

Wichtig: Der Checkpoint entscheidet, nicht die Oberfläche

Ein feinjustierter Checkpoint erbt das Budget seiner Basis. Ein SDXL-Fine-Tune bleibt ein Ein-Megapixel-Modell, auch wenn er 2026 veröffentlicht wurde und der Name nach etwas Modernem klingt. Dasselbe gilt für Zusatzgewichte: Ein [Was ist ein LoRA?](#), das auf 512er-Material trainiert wurde, zieht das gesamte Ergebnis in Richtung dieser Fläche – auch auf einem 1024er-Modell.

Merke: Nicht das Werkzeug legt die native Auflösung fest, sondern die Gewichte, die du geladen hast. Die Voreinstellung der Oberfläche ist eine Vermutung, keine Angabe.

Das Raster: durch 8 teilbar reicht längst nicht mehr

Eine Regel aus der Frühzeit hält sich hartnäckig: Kantenlängen müssen durch 8 teilbar sein. Das war für die UNet-Modelle der ersten Generation richtig, weil der VAE um Faktor 8 verkleinert. Für die heutigen Transformer-Modelle ist es zu wenig. Die zerlegen das Latentbild zusätzlich in Kacheln, üblicherweise 2×2 Felder. Damit liegt der tatsächliche Teiler bei 16 – und die Schnittstelle von Black Forest Labs verlangt für FLUX schlicht Vielfache von 32, um die Sache eindeutig zu halten. Stability empfiehlt für die 3.5er-Reihe 64.

Warnung: Krumme Maße scheitern leise

Wenn eine Kantenlänge nicht auf dem Raster liegt, bricht selten etwas ab. Die Werkzeuge runden still auf den nächsten gültigen Wert – und du bekommst ein Bild, das ein paar Pixel anders ist als

angefordert. Bei einem Einzelbild egal. In einer Serie, die pixelgenau in ein Layout muss, ärgerlich. Der praktisch relevantere Fall: Manche Zielmaße sind auf einem 32er-Raster gar nicht erreichbar. 4K-UHD mit 3840×2160 ist so ein Fall – 2160 ist nicht durch 32 teilbar. Du generierst dann 3840×2176 und schneidest 16 Zeilen ab. Das gehört in den Workflow, nicht in die Nacharbeit von Hand. Faustregel für den Alltag: Runde immer auf Vielfache von 64. Damit bist du auf jedem Modell auf der sicheren Seite, ohne dir merken zu müssen, welche Architektur gerade geladen ist.

Das Doppelkopf-Phänomen und seine Verwandten

Jetzt können wir erklären, was am Anfang passiert ist. Und zwar so, dass du das Symptom beim nächsten Mal sofort richtig zuordnest, statt am Prompt herumzuschrauben.

Warum das Modell das Motiv zweimal anfängt

Das Modell hat keine Vorstellung von „dem Bild als Ganzem“. Es hat gelernt, lokale Nachbarschaften plausibel zu füllen, und es hat dabei eine feste Vorstellung davon entwickelt, wie groß die Dinge im Raster sind. Gibst du ihm mehr Fläche, als es kennt, ist jede einzelne Bildregion für sich betrachtet weiterhin plausibel – nur ergibt das Ganze keine einzelne Szene mehr, sondern zwei halbe, die sich in der Mitte treffen.

Deshalb sind die Symptome so typisch: zwei Köpfe auf einem Körper, ein zweiter Oberkörper am Bildrand, eine dritte Hand, ein Horizont, der sich wiederholt, ein Muster, das plötzlich gekachelt aussieht. Bei Händen kommt erschwerend hinzu, dass die ohnehin die Schwachstelle sind – das ist das Thema von [Warum KI keine Hände kann](#). Eine zu große Leinwand macht aus einem schwierigen Problem ein aussichtsloses.

Das trainierte Fenster passt nur einmal – oder eben zweimal
Gestrichelt: die Bildgröße, in der das Modell komponieren gelernt hat.

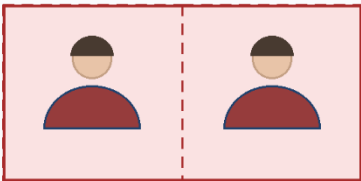
A – im Budget
1024 x 1024



✓

Leinwand = trainiertes Fenster.
Das Modell setzt das Motiv
genau einmal an.

B – weit über Budget
2048 x 1024 bei 1-MP-Modell



✗

Die Leinwand fasst das Fenster zweimal.
Das Modell füllt die Restfläche, indem es
das Motiv noch einmal anfängt.

C – unter Budget
512 x 512



✗

Zu wenig Fläche für das
gelernte Detail: matschige
Gesichter, verlorene Struktur.

Beide Fehler sind derselbe Fehler: Die Leinwand passt nicht zu der Fläche, auf der das Modell komponieren gelernt hat.

Skizze 3: Das trainierte Fenster ist gestrichelt eingezeichnet. Links passt es genau, in der Mitte passt es zweimal – und genau das produziert das Modell dann auch.

Nach unten geht es genauso schief

Der umgekehrte Fehler bekommt weniger Aufmerksamkeit, weil er nicht so spektakulär aussieht. Generierst du deutlich unter dem Budget, hat das Modell schlicht zu wenig Raster für das Detail, das es gelernt hat. Das Ergebnis ist nicht verdoppelt, sondern matschig: Gesichter verlieren

Struktur, Text zerfällt, feine Muster verschmieren. Ein SDXL-Modell auf 512×512 liefert schlechtere Bilder als ein SD-1.5-Modell auf derselben Größe – weil es für diese Fläche nie gebaut wurde.

Beide Fehler sind derselbe Fehler mit umgekehrtem Vorzeichen. Und beide lassen sich durch dieselbe Maßnahme abstellen: das Bild in der Fläche rendern, für die das Modell gebaut ist.

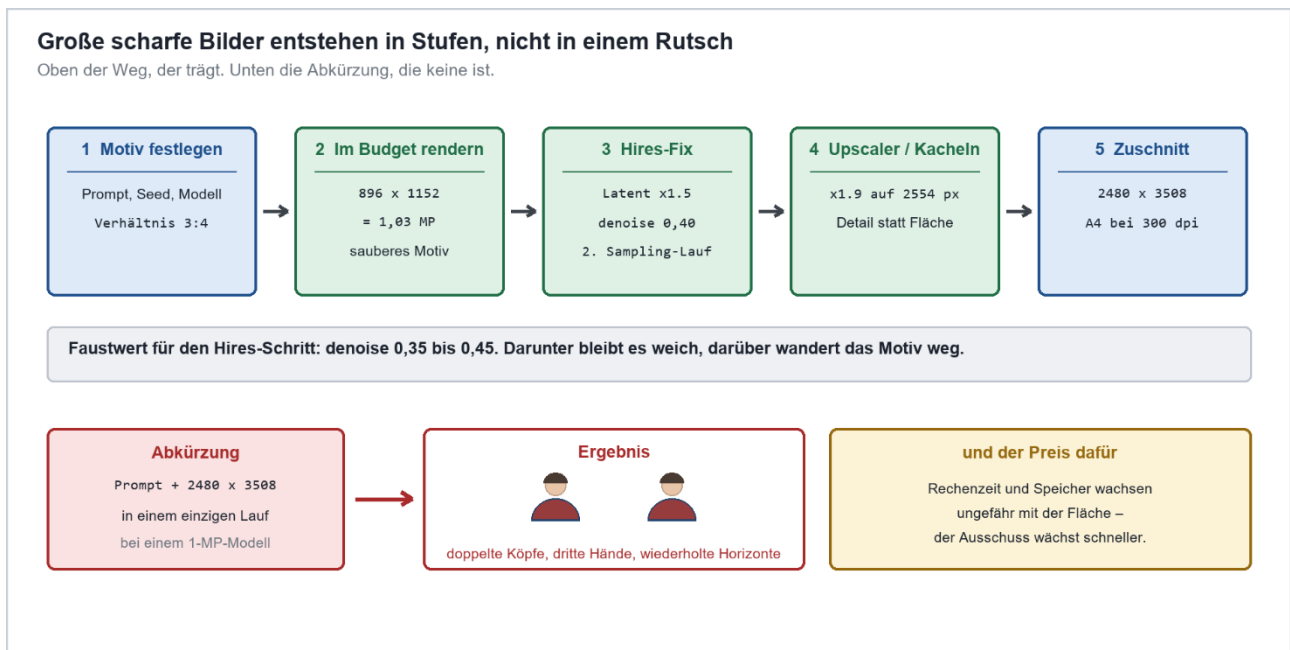
Diagnose: Symptom, Ursache, Abhilfe

Was du siehst	Woran es fast immer liegt	Was hilft
Zwei Köpfe, doppelter Oberkörper	Leinwand deutlich über dem Flächenbudget	Zurück auf das Budget, Größe danach per Upscaling holen
Horizont oder Muster wiederholt sich	Extremes Panoramaformat bei zu großer Fläche	Fläche reduzieren, Verhältnis beibehalten
Zusätzliche Gliedmaßen am Bildrand	Dasselbe Problem, nur weniger offensichtlich	Fläche prüfen, bevor du am Negativ-Prompt schraubst
Matschige Gesichter, verlorenes Detail	Unter dem Budget generiert	Hoch auf die empfohlene Fläche
Bild wirkt scharf, aber leer	Fläche stimmt, Motiv trägt das Format nicht	Seitenverhältnis ändern, nicht die Fläche
Streifen oder unsaubere Kante unten/rechts	Kantenlänge liegt nicht auf dem Raster	Auf Vielfache von 64 runden
Nach dem Hires-Schritt ein anderes Gesicht	Denoise im zweiten Durchgang zu hoch	Auf 0,35 bis 0,45 zurück
Sichtbare Nähte bei großem Upscale	Kachelüberlappung zu klein oder Denoise zu hoch	Überlappung erhöhen, Denoise senken

Bemerkenswert an dieser Tabelle ist, wie wenige Zeilen etwas mit dem Prompt zu tun haben. Wer bei einem Doppelkopf anfängt, „two heads“ in den Negativ-Prompt zu schreiben, behandelt ein Symptom mit einem Werkzeug, das dafür nicht gebaut ist – mehr dazu in [Prompt und Negativ-Prompt in der KI-Bildgenerierung](#). Die Fläche zu korrigieren dauert fünf Sekunden und wirkt zuverlässig.

Der Weg zu großen, scharfen Bildern

„Im Budget bleiben“ heißt nicht, dass du auf große Bilder verzichtest. Es heißt nur, dass du die Fläche in einem zweiten Arbeitsgang holst und nicht im ersten. Das Prinzip ist immer dasselbe: Das Motiv entsteht klein und sauber, die Größe kommt danach.



Skizze 4: Oben der Weg, der trägt – vier Schritte von der Bildidee bis zum Druckmaß. Unten die Abkürzung samt ihrem üblichen Ergebnis.

Stufe 1: im Budget rendern

Wähle die Fläche nach dem Modell und die Form nach dem Verwendungszweck. Hier wird auch der Seed festgehalten, denn ab jetzt willst du genau dieses Motiv weiterverarbeiten und kein anderes – warum das ohne notierten Seed nicht funktioniert, steht in [Der Seed bei der KI-Bildgenerierung](#). Wer in dieser Phase mehrere Kandidaten braucht, findet in [batch size und Reproduzierbarkeit](#) die Regeln für den Stapelbetrieb.

Stufe 2: der Hires-Fix

Der Hires-Fix, in manchen Oberflächen auch Highres-Fix oder Two-Pass genannt, ist der elegantere Teil. Er vergrößert das fertige Latentbild um einen moderaten Faktor, typischerweise 1,5 bis 2, und lässt das Modell einen zweiten, kürzeren Sampling-Durchgang darüberlaufen. Der Clou liegt im Denoise-Wert dieses zweiten Durchgangs: Er bestimmt, wie viel Freiheit das Modell bekommt.

Bei 0,35 bis 0,45 bleibt die Komposition erhalten, und das Modell füllt nur die neu hinzugekommene Fläche mit Detail. Unter 0,3 passiert zu wenig, das Bild bleibt weich. Über 0,6 fängt das Modell an, das Motiv umzuinterpretieren – Gesichter verändern sich, Kleidung wechselt die Farbe. Welcher Sampler diesen zweiten Durchgang macht, ist keine Nebensache; das behandelt [Sampler & Scheduler erklärt](#). Auch die Schrittzahl darf im zweiten Lauf niedriger liegen als im ersten, dazu mehr in [Sampling-Steps erklärt](#).

Warnung: Upscaling ist kein neutraler Vorgang

Jeder diffusionsbasierte Vergrößerungsschritt erfindet Detail hinzu, das im Ausgangsbild nicht steckte. Meistens ist das genau das, was du willst. Bei Gesichtern ist es ein Risiko: Ein Markenbotschafter, ein wiederkehrender Charakter oder ein Produkt können sich im zweiten Durchgang messbar verändern. Wie man Bildwelten trotzdem konsistent hält, steht unter [Markenkonsistente KI-generierte Bilder](#).

Wenn Konsistenz zählt: Denoise niedrig halten, den Ausschnitt nach dem Upscaling gegen das Original prüfen und den Faktor lieber in zwei kleinen Schritten holen als in einem großen.

Stufe 3: Upscaler und Kachelverfahren

Für alles jenseits von Faktor 2 brauchst du eigene Werkzeuge. Da gibt es zwei grundverschiedene Ansätze, die gern in einen Topf geworfen werden – zu Unrecht, denn sie taugen für Unterschiedliches.

Verfahren	Was dabei passiert	Sinnvoller Faktor	Stärke	Schwäche
Hires-Fix im Generator	Latent vergrößern, zweiter Sampling-Lauf	1,5 bis 2	Ein Arbeitsgang, sehr gute Detailqualität	Braucht viel Speicher, Motiv kann driften
Klassischer Pixel-Upscaler	Gelerntes Hochrechnen ohne Diffusionsschritt	2 bis 4	Schnell, sparsam, verändert das Motiv nicht	Erfindet kein echtes neues Detail
Kachelweise Diffusion	Bild in überlappende Kacheln zerlegen, jede neu rechnen	4 und mehr	Beliebig große Ausgaben, viel neues Detail	Langsam, Nahtstellen, Motiv-Drift je Kachel
Anbieterseitige Hi-Res-Stufe	Der Dienst macht die Zweistufigkeit intern	fest vorgegeben	Kein Aufwand, gut abgestimmt	Keine Kontrolle, kein Zwischenstand

In der Praxis kombiniert man: Hires-Fix für den ersten Faktor, ein klassischer Upscaler für die reine Fläche, kachelweise Diffusion nur da, wo wirklich neues Detail gebraucht wird. Wie so eine Kette als belastbarer Produktionsablauf aussieht, beschreibt [Aufbau einer skalierbaren Bild-Maschine mit ComfyUI und Flux](#). Und weil jede Stufe Speicher kostet, ist die Frage nach Cloud oder eigener Hardware hier besonders greifbar – siehe [Cloud vs. lokale KI-Bildgenerierung](#).

Was Auflösung nicht kann

Eine höhere Auflösung macht kein Bild besser. Sie gibt ihm mehr Pixel. Die Komposition wird davon nicht spannender, ein schiefer Bildaufbau nicht gerade und ein austauschbares Motiv nicht interessant. Wer ein langweiliges Bild hochrechnet, hat danach ein großes langweiliges Bild. Der Inhalt kommt aus Prompt und Modell, die Bildqualität aus dem sauberen Entstehen – und dafür sind Schrittzahl, Sampler und der Guidance-Wert ungleich wichtiger als die Pixelzahl. Zu Letzterem lohnt [CFG-Scale: Guidance in KI-Bildgeneratoren](#).

Auch der Umkehrschluss stimmt nicht: Mehr Pixel heißt nicht automatisch mehr Schärfe. Ein auf 4000 Pixel hochgerechnetes Bild, das aus einer unsauberen Erstgenerierung stammt, ist ein großes unscharfes Bild mit hinzuerfundenen Details. Die Schärfe entsteht in Stufe 1, nicht in Stufe 3.

Formate rückwärts planen statt vorwärts raten

Bis hierhin war alles Technik. Der Teil, an dem Projekte in der Praxis scheitern, ist ein anderer: Niemand hat vorher festgelegt, welche Maße überhaupt gebraucht werden. Dann generiert eine Abteilung quadratisch, die nächste im Hochformat, und drei Wochen später fragt jemand nach der Druckvorlage für die Messe.

Die Formatmatrix kommt vor dem ersten Prompt

Der Ablauf, der funktioniert, ist immer derselbe: Erst den Kanal festlegen, dann das Zielmaß in Pixeln, dann das Seitenverhältnis, dann die Generierungsfläche, dann den Upscaling-Faktor. Fünf Zeilen pro Kanal, einmal sauber aufgeschrieben. Das ist kein Bürokratieprojekt, das ist eine Tabelle mit acht Zeilen, die man an einem Vormittag erstellt und danach nie wieder diskutiert.

Ein Beispiel aus einem Projekt, das ich nicht näher benennen werde: Ein Mittelständler hatte rund 900 KI-Bilder für einen Produktkatalog erzeugt, alle in 1024×1024 . Der Katalog wurde im Hochformat gedruckt. Beschnitten sah jedes Bild aus wie ein Ausschnitt aus einem anderen Bild –

weil das Modell quadratisch komponiert hatte. Die Neuproduktion im richtigen Verhältnis dauerte zwei Tage. Die Diskussion darüber, wer schuld war, dauerte drei Wochen.

Tipp: Zwei Werkzeuge, die genau hier greifen

Den Formatkatalog – welcher Kanal welches Zielmaß, welche Generierungsfläche, welchen Upscaling-Weg bekommt – führst du als Katalog in UB.Dossier. Dort lassen sich die Werte ändern, ohne dass jemand eine Vorlage anfassen muss, und pro Asset bleibt nachweisbar, nach welcher Formatregel es entstanden ist.

Den eigentlichen Ablauf – Fläche, Raster, Hires-Faktor, Zuschnitt – legst du in UB.Forge ab. Der Plausibilitäts-Check dort fängt die beiden klassischen Fehler vor dem Lauf ab: eine Kantenlänge, die nicht auf dem Raster liegt, und eine Fläche, die über dem Budget des geladenen Modells liegt. Beides kostet in der Produktion sonst eine komplette Serie.

Was das für Nachweis und Governance bedeutet

Sobald Bilder in Serie entstehen, gehört das Format zu den Metadaten, die du mitschreiben solltest: Generierungsfläche, Seitenverhältnis, Upscaling-Kette, Zielmaß. Ohne diese Angaben lässt sich ein Asset später nicht reproduzieren und auch nicht sauber nachschärfen. Wie sich so etwas in einer bestehenden Microsoft-Umgebung ablegen lässt, zeigt [Digital Asset Management für KI-Bilder mit SharePoint Online und Power Platform](#). Der strategische Rahmen dahinter steht in [KI-Bildgenerierung im Unternehmen – Die strategische Alternative zur Cloud](#).

Und noch ein Punkt, der mit Auflösung direkt zu tun hat, auch wenn er zunächst nicht danach klingt: Kennzeichnungspflichten. Ein Upscaling-Schritt durchläuft in vielen Ketten eine Neucodierung des Bildes – dabei gehen eingebettete Herkunftsangaben gern verloren. Wer sich auf technische Marker verlässt, sollte nach jeder Vergrößerungsstufe prüfen, ob sie noch da sind. Was seit dem 2. August 2026 gilt, steht in [EU AI Act: Was seit dem 2. August 2026 gilt](#); wie belastbar solche Marker überhaupt sind, behandelt [KI-generierte Bilder erkennen](#).

Wenn du die Formatmatrix und die Upscaling-Kette nicht selbst aufbauen willst: Genau das ist der Gegenstand meiner [KI-Bildgenerierung Beratung](#). Soll dein Team es selbst können, ist die [KI-Bildgenerierung Schulung](#) der kürzere Weg. Dass ich die beschriebenen Abläufe selbst einsetze und nicht nur beschreibe, steht unter [Dogfooding: Ich setze ein, was ich berate](#). Für die Automatisierung rund um Formatberechnung und Prüfregele arbeite ich überwiegend mit Sprachmodellen von Anthropic – dazu [Claude im Unternehmen](#).

Häufige Fragen zu Auflösung und Seitenverhältnis

Warum hat mein Bild zwei Köpfe oder doppelte Motive?

Weil die Leinwand deutlich größer ist als die Fläche, auf der das Modell trainiert wurde. Die Größenverhältnisse im Bild sind fest gelernt; bekommt das Modell mehr Fläche, füllt es sie, indem es das Motiv noch einmal anfängt. Abhilfe: zurück auf das Flächenbudget des Modells, sauber generieren, danach vergrößern. Am Prompt zu schrauben hilft hier nicht.

Was genau ist die native Auflösung im KI-Bildgenerator?

Die native Auflösung ist das Flächenbudget, auf dem ein Modell trainiert wurde und in dem es am zuverlässigsten komponiert. Es ist kein einzelner Wert, sondern ein Bereich zulässiger Formen bei ungefähr gleicher Pixelzahl. SDXL und Stable Diffusion 3.5 Large liegen bei rund einem Megapixel, Qwen-Image bei rund 1,76 Megapixeln, FLUX.2 bei bis zu vier Megapixeln. Die alten SD-1.5-Modelle liegen bei rund 0,26 Megapixeln.

Wie bekomme ich große, scharfe Bilder?

In Stufen. Erst im Flächenbudget generieren, wo das Motiv sauber entsteht. Dann den Hires-Fix mit Faktor 1,5 bis 2 und einem Denoise-Wert von 0,35 bis 0,45. Für alles darüber ein eigenes Upscaling-Werkzeug oder kachelweise Diffusion. Zum Schluss auf das exakte Zielmaß beschneiden. Direkt riesig zu generieren liefert keine großen Bilder, sondern große Fehler.

Welches Seitenverhältnis soll ich nehmen?

Das entscheidet der Verwendungszweck, nicht der Geschmack. 1:1 für Social-Kacheln, 3:4 und 2:3 für Personen und Porträts, 16:9 für Präsentationen und Web-Hero-Bilder, 21:9 für Banner. Wichtig ist nur: Die Form änderst du über das Verhältnis bei ungefähr konstanter Fläche – nicht, indem du eine Kante verlängerst.

Wird ein größeres Bild automatisch besser?

Nein. Mehr Pixel sind mehr Pixel, sonst nichts. Komposition, Detailtiefe und Bildwirkung kommen aus Prompt, Modell und einem sauberen Entstehungsprozess. Jenseits des Budgets generiert wird das Bild sogar messbar schlechter, weil die Komposition zerfällt.

Müssen die Maße wirklich durch 8 teilbar sein?

Durch 8 ist das Minimum und stammt aus der Zeit der ersten UNet-Modelle, wo der Autoencoder um Faktor 8 verkleinert. Heutige Transformer-Modelle zerlegen das Latentbild zusätzlich in 2×2 -Kacheln, womit der Teiler bei 16 liegt; die FLUX-Schnittstelle verlangt Vielfache von 32, Stability empfiehlt für die 3.5er-Reihe 64. Wer generell auf 64 rundet, hat mit keinem Modell Ärger.

Kann ich mit einem 1-Megapixel-Modell überhaupt drucken?

Ja, aber nicht in einem Durchgang. Eine A4-Seite bei 300 dpi braucht 2480×3508 Pixel, also fast neun Megapixel. Der Weg dorthin führt über eine saubere Erstgenerierung bei rund einem Megapixel im richtigen Verhältnis und anschließend über zwei Vergrößerungsstufen. Das Ergebnis ist regelmäßig besser als ein direkt groß generiertes Bild – und braucht weniger Rechenzeit.

Warum sieht mein Bild im Hochformat ganz anders aus als im Querformat?

Weil das Seitenverhältnis mitkomponiert. Das Modell hat gelernt, welche Motive in welchen Formaten vorkommen, und passt Bildausschnitt und Anordnung entsprechend an. Das ist kein Fehler, sondern ein Gestaltungsmittel: Wenn du mehr Umgebung sehen willst, nimm Querformat; wenn die Person tragen soll, nimm Hochformat.

Verändert Upscaling das Gesicht meines Motivs?

Es kann, und zwar genau dann, wenn der Denoise-Wert im zweiten Durchgang zu hoch steht. Ab etwa 0,6 interpretiert das Modell das Motiv neu, statt es nur nachzuschärfen. Für wiedererkennbare Personen oder Produkte gehört der Wert auf 0,35 bis 0,45, und der Faktor sollte lieber in zwei kleinen Schritten geholt werden als in einem großen. Rein rechnerische Upscaler ohne Diffusionsschritt verändern das Motiv gar nicht – dafür fügen sie auch kein Detail hinzu.

Fazit

Auflösung ist kein Qualitätsregler, sondern eine Flächenangabe. Jedes Modell hat ein Flächenbudget, in dem es sauber komponiert, und außerhalb dieses Budgets produziert es zuverlässig Unsinn – nach oben durch Verdopplung, nach unten durch Matsch. Das Seitenverhältnis wiederum ist ein Gestaltungsmittel, das die Komposition beeinflusst, und kein Nebeneffekt der Größenwahl.

Die drei Sätze, die in der Praxis den Unterschied machen: Denk in Fläche statt in Kantenlänge. Ändere die Form über das Verhältnis, nicht über die Größe. Und hol dir die Größe in einer zweiten Stufe, nicht in der ersten. Wer sich daran hält, hat das Doppelkopf-Problem nicht gelöst, sondern gar nicht erst.

Und weil die Budgets inzwischen weit auseinanderliegen – von 0,26 Megapixeln bei den alten Modellen bis zu vier Megapixeln bei FLUX.2 – lohnt sich der Blick in die Modellangaben vor dem ersten Lauf. Es ist eine Zahl. Sie nachzuschlagen dauert eine Minute. Sie nicht nachzuschlagen kostet einen Nachmittag.

Weiterlesen

- › Welche native Auflösung gilt, entscheidet das geladene Modell: [Checkpoints, Modelle und VAE](#)
- › Was überhaupt im Bild landet, regeln Prompt und Negativ-Prompt: [Prompt und Negativ-Prompt in der KI-Bildgenerierung](#)
- › Für die Bildqualität zählen Schrittzahl und Sampler mehr als die Pixelzahl: [Sampling-Steps erklärt](#) und [Sampler & Scheduler erklärt](#)
- › Ohne notierten Seed lässt sich kein Bild nachschärfen: [Der Seed bei der KI-Bildgenerierung](#)
- › Die Upscaling-Kette als belastbarer Produktionsablauf: [Aufbau einer skalierbaren Bild-Maschine mit ComfyUI und Flux](#)
- › Alle Grundlagen im Überblick: [KI-Bildgenerierung](#)