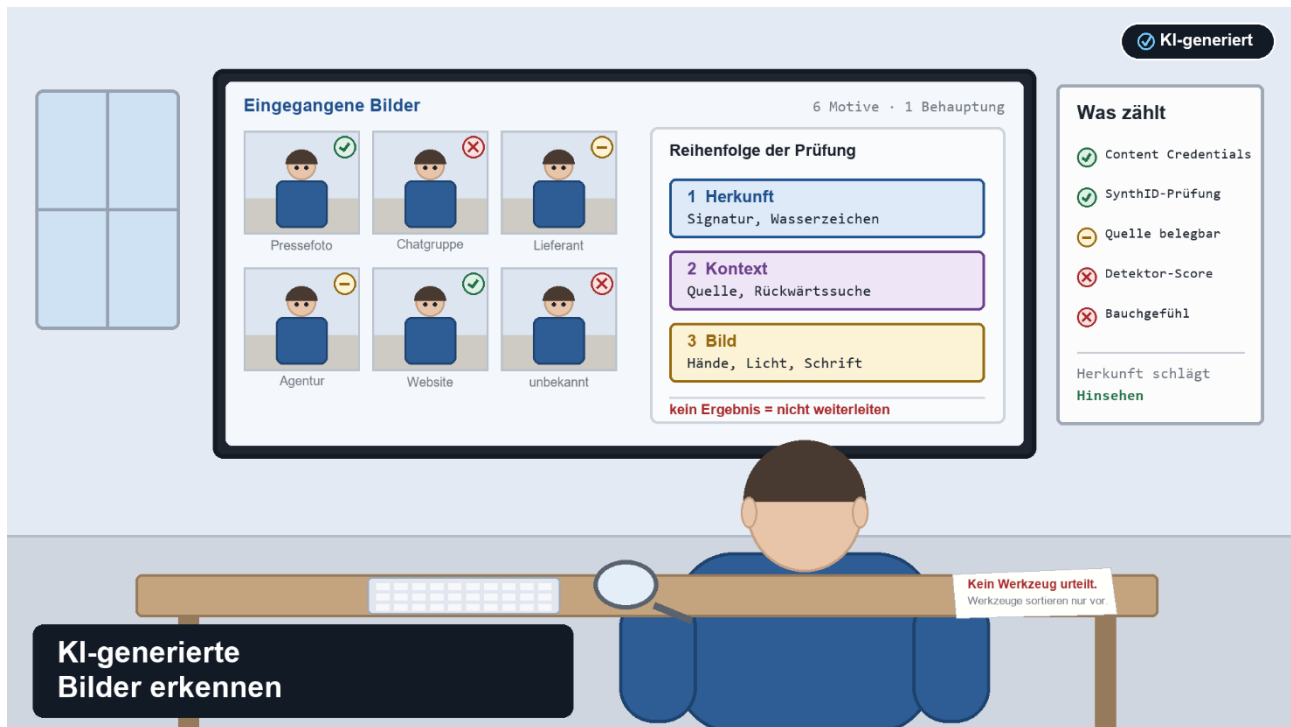


KI-generierte Bilder erkennen



Sechs eingegangene Bilder, eine Behauptung: Die Prüfung beginnt nicht beim Pixel, sondern bei der Herkunft. Wer mit den Fingern anfängt, verliert Zeit und am Ende meistens auch die Diskussion.

Wissen	Beratung	Schulung
Alle Grundlagen zur KI-Bildgenerierung an einer Stelle – vom Diffusionsprinzip bis zur Governance.	Wir bauen dir eine Prüf- und Kennzeichnungsroutine, die im Alltag hält und im Zweifel belegbar ist.	Dein Team lernt, Bilder methodisch zu prüfen statt auf Fingerzählen zu hoffen.
KI-Bildgenerierung	KI-Bildgenerierung Beratung	KI-Bildgenerierung Schulung

„Das sieht man doch sofort.“ Diesen Satz höre ich in Workshops ungefähr alle zwanzig Minuten, und er hat eine Halbwertszeit von genau einer Übung. Ich lege dem Team zwölf Bilder vor, sechs echte Fotos, sechs erzeugte, gemischt und ohne Dateinamen. Dann darf jede und jeder tippen. Das Ergebnis liegt seit über einem Jahr zuverlässig dort, wo auch reines Raten landen würde – mit dem kleinen Unterschied, dass die Selbstsicherheit vorher deutlich höher war als die Trefferquote hinterher.

Besonders schön war ein Termin bei einem Maschinenbauer im Sauerland. Die Kommunikationsabteilung war überzeugt, KI-Bilder immer zu erkennen, und lag bei sieben von zwölf richtig. Peinlich wurde es erst bei der Auflösung: Zwei der Bilder, die man einhellig als „ganz klar KI“ eingestuft hatte, waren Aufnahmen aus der eigenen Werkhalle, geschossen von einer Fotografin, die das Unternehmen seit fünfzehn Jahren begleitet. Die Bilder waren zu sauber ausgeleuchtet. So weit sind wir also: Professionelle Fotografie gerät unter Verdacht, weil sie zu gut aussieht.

Trotzdem lohnt sich ein geschultes Auge. Nur eben anders, als die meisten erwarten. Dieser Beitrag zeigt dir die klassischen Verräter und sagt ehrlich dazu, wie viel sie heute noch wert sind. Danach geht es um das, was wirklich trägt: Herkunftsnachweise, Kontextprüfung und eine Routine, die du

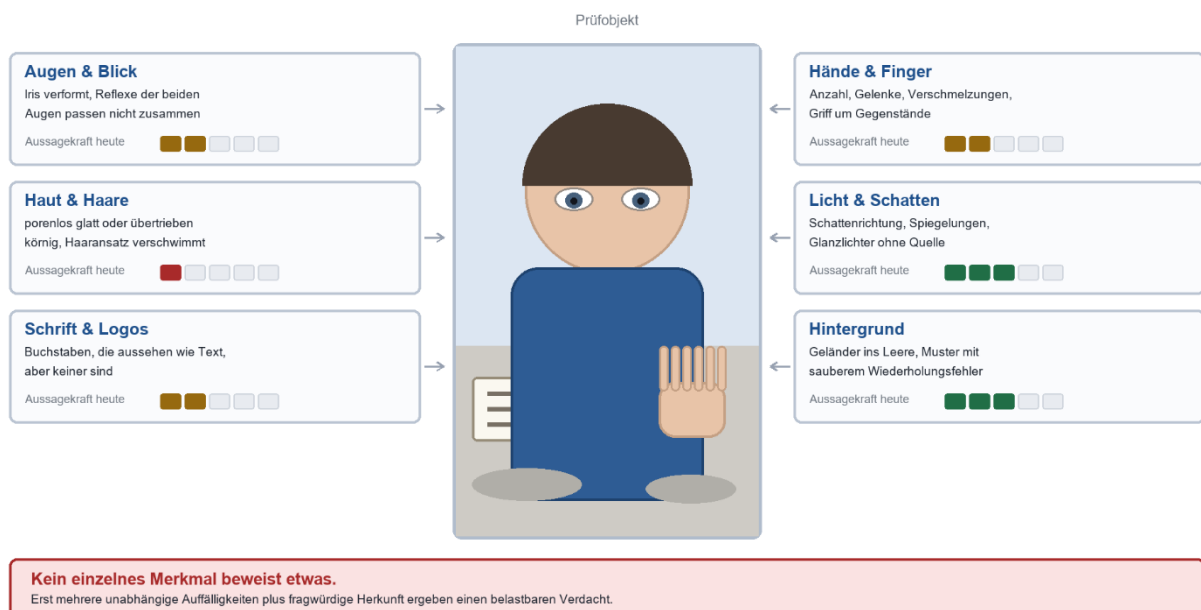
auch dann erklären kannst, wenn jemand widerspricht. Die unbequeme Kurzfassung gleich vorweg – eine sichere Methode, ein gutes KI-Bild allein am Bild zu erkennen, gibt es nicht mehr.

Die klassischen Verräter und ihr Verfallsdatum

Fangen wir mit dem an, was du vermutlich suchst: die Merkmale, an denen Bildmodelle traditionell scheitern. Sie sind nicht wertlos. Sie sind nur kein Beweis, sondern ein Anfangsverdacht – und sie halten unterschiedlich lange. Wichtig ist die Denkweise dahinter: Ein Diffusionsmodell versteht die Welt nicht, es hat gelernt, wie Bilder der Welt üblicherweise aussehen. Es reproduziert statistische Regelmäßigkeiten. Überall dort, wo Plausibilität und Physik auseinanderlaufen, entstehen die Fehler.

Die klassischen Verräter

Wo Modelle historisch patzen – und wie viel Aussagekraft davon 2026 übrig ist



Sechs Prüfstellen und die Aussagekraft, die 2026 davon übrig ist. Der Balken zeigt, wie verlässlich das Merkmal heute noch trennt – nicht, wie oft es früher funktioniert hat.

Hände, Finger und Gelenke

Der Dauerbrenner, und immer noch der erste Blick. Zu viele Finger, zu wenige, verschmolzene Knöchel, Daumen an der falschen Seite, Gelenke, die sich in eine Richtung biegen, in die kein menschliches Gelenk geht. Die Ursache ist handwerklich erklärbar: Hände erscheinen im Training in unzähligen Verkürzungen und Verdeckungen, und kleine Fehler in einem kleinen Bildbereich kosten das Modell beim Lernen kaum etwas. Wer die Mechanik dahinter nachlesen will: [Warum KI keine Hände kann](#) erklärt, warum ausgerechnet dieses Detail so hartnäckig ist.

Der Haken: Genau dieses Merkmal war so berühmt, dass die Modellhäuser gezielt nachgebessert haben. Bei den aktuellen Spitzenmodellen sind saubere Hände der Normalfall, nicht die Ausnahme. Was bleibt, sind Randfälle – zwei Hände, die einander greifen, Hände an Werkzeugen, Hände hinter Glas. Wenn du auf einem Bild aus dem Jahr 2026 fünf korrekte Finger siehst, weißt du genau gar nichts.

Augen, Zähne, Ohren und Schmuck

Augen sind ergiebiger als Hände, weil sie zwei Prüfungen auf einmal erlauben. Erstens die Form: Iris leicht oval statt rund, Pupille nicht mittig, Lidkanten unterschiedlich scharf. Zweitens – und

das ist der bessere Test – die Reflexe. In einem echten Foto zeigen beide Augen dieselbe Lichtquelle in leicht versetzter Position. Erzeugte Bilder bauen Glanzlichter oft dekorativ statt konsistent, mit unterschiedlichen Formen in linkem und rechtem Auge. Dasselbe Prinzip gilt für Ohrringe, Brillenbügel und Zähne: Alles, was paarweise oder in Serie auftritt, muss zueinander passen, und genau das ist statistisch schwer.

Haut, Haare und Stoff

Die alte Regel „KI-Haut ist zu glatt“ stimmt heute eher umgekehrt. Die aktuelle Modellgeneration liefert Hautporen, Härchen und Unreinheiten in einer Detailtiefe, die frühere Versionen nicht hatten; FLUX.2 wird in Vergleichen genau dafür gelobt. Was übrig bleibt, ist ein feineres Signal: Die Textur ist oft über die gesamte Fläche gleichmäßig verteilt. Echte Haut hat Zonen – die Nase anders als die Wange, die Stirn anders als der Hals. Ähnlich bei Haaren: Einzelne Strähnen sehen perfekt aus, aber der Übergang zur Kopfhaut verschwimmt, oder eine Strähne endet einfach im Nichts. Bei Stoff lohnt der Blick auf Nähte und Knopfleisten, die gern die Richtung wechseln.

Licht, Schatten und Spiegelungen

Hier wird es für Modelle unangenehm, weil Licht keine Textur ist, sondern eine Konsequenz. Prüfe die Schattenrichtung mehrerer Objekte gegeneinander: Alle Schatten in einer Szene müssen zu derselben Lichtquelle passen, und ihre Länge muss zusammenstimmen. Spiegelungen in Fenstern, Brillengläsern, Wasser oder polierten Flächen sind der härteste Test überhaupt, weil ein Spiegelbild eine geometrisch korrekte Zweitansicht derselben Szene wäre. Modelle malen dort meistens etwas Spiegelähnliches hin. Von allen Bildmerkmalen ist dieses das langlebteste – es setzt ein Raumverständnis voraus, das über Mustererkennung hinausgeht.

Schrift, Logos und Zahlen

Das klassische Kauderwelsch auf Schildern ist weitgehend Geschichte. Aktuelle Modelle setzen kurze Texte sauber, FLUX.2 und Nano Banana Pro werden für ihr Textrendering ausdrücklich gelobt. Aufgeben musst du das Merkmal trotzdem nicht, du musst nur genauer hinschauen: Ein korrekt gesetzter Markenname ist nichts wert, ein Firmenlogo mit leicht falschen Proportionen schon. Ergiebig sind lange Texte, Kleingedrucktes, Nummernschilder, Preisschilder, Fahrpläne. Je mehr Zeichen, desto höher die Wahrscheinlichkeit, dass irgendwo ein Buchstabe kippt oder eine Zeile in Pseudoschrift ausfranst.

Hintergrund, Muster und Physik

Der Hintergrund ist die faulste Region eines jeden Bildes – beim Modell wie beim Betrachter. Geländer laufen ins Leere, Treppenstufen ändern die Höhe, Fliesenraster verziehen sich, ein Fenster hat innen mehr Sprossen als außen. Besonders verräterisch sind Wiederholungen: Menschenmengen mit demselben Gesicht in drei Variationen, Bücherregale mit identischen Rücken, Parkplätze mit demselben Auto. Und dann die Physik: Wasser, das nicht fällt, Rauch ohne Quelle, Kleidung, die die Schwerkraft ignoriert, Schnittkanten zwischen Vorder- und Hintergrund, die zu sauber sind.

Merkmal	Worauf du achtest	Aussagekraft heute	Warum es verfällt
Hände und Finger	Anzahl, Gelenkrichtung, Griff um Objekte	gering bis mittel	gezielt nachtrainiert, weil es das bekannteste Merkmal war
Augen	Irisform, Symmetrie, identische Reflexe	mittel	Form wird besser, Reflexlogik bleibt schwierig
Haut und Haare	gleichmäßige Textur ohne Zonen, Haaransatz	gering	aktuelle Modelle liefern Poren und Unreinheiten von selbst
Licht und Schatten	Schattenrichtung, Länge, Spiegelbilder	hoch	verlangt Raumverständnis, nicht nur Musterwissen

Schrift und Logos	lange Texte, Kleingedrucktes, Proportionen	mittel	kurze Texte sitzen in zwischen, lange nicht
Hintergrund	Linienverlauf, Wiederholungen, Rasterfehler	mittel bis hoch	Nebensache im Training, deshalb weiter fehleranfällig
Gesamteindruck	zu perfekt, zu symmetrisch, zu ausgeleuchtet	gering	lässt sich per Prompt gezielt zerstören

Ein Merkmal ist ein Verdacht, keine Feststellung

Die häufigste Fehlanwendung dieser Liste ist nicht das Übersehen eines KI-Bildes, sondern der falsche Vorwurf gegen ein echtes. Professionell ausgeleuchtete Fotos, retuschierte Porträts und Produktaufnahmen tragen genau die Merkmale, die hier als verdächtig gelten: makellose Haut, perfekte Symmetrie, kontrolliertes Licht.

Wer auf dieser Grundlage öffentlich behauptet, jemand habe ein KI-Bild verbreitet, landet schneller in einer Richtigstellung, als ihm lieb ist. Ein Verdacht rechtfertigt eine Rückfrage beim Urheber. Er rechtfertigt keine Behauptung.

Wenn es kein Fehler ist, sondern nur ein Gefühl

Manchmal findest du keinen einzigen konkreten Fehler und trotzdem sträubt sich etwas. Das Bild wirkt wie das Idealbild einer Szene und nicht wie ein eingefangener Moment: alles im Fokus, nichts im Weg, keine abgeschnittene Person am Bildrand, kein Kabel, kein Fleck, keine Unordnung. Dieses Unbehagen hat einen Namen – Uncanny Valley, das unheimliche Tal – und es ist tatsächlich schneller als der bewusste Verstand. Als Vorsortierung taugt es.

Als Beweis taugt es nicht mehr, und zwar aus einem sehr banalen Grund: Der Effekt lässt sich wegprompten. Wer „Handyfoto, Abendlicht, leicht verwackelt, Rauschen“ schreibt, bekommt genau das. Die Modelle können Unvollkommenheit inszenieren, und wer Bilder zur Täuschung erzeugt, hat einen sehr guten Grund, genau das zu tun. Die Schnappschuss-Ästhetik ist heute ein Stilfilter wie jeder andere. Damit hat sich das Bauchgefühl in ein Signal verwandelt, das gegen Amateure funktioniert und gegen jeden versagt, der es ernst meint.

Der praktikabelste Einzeltest: hineinzoomen, wo niemand hinschaut

Nicht das Hauptmotiv prüfen, sondern die Peripherie. Bildränder, dritter Plan, Reflexionen in Nebensachen, das Muster auf einem Teppich, die Beschriftung auf einer Verpackung im Hintergrund. Bildmodelle investieren ihre Qualität dort, wo die Aufmerksamkeit liegt – und der Rest wird plausibel, nicht korrekt.

Zweiter Handgriff: Bild stark vergrößern und den Übergang zwischen Objekt und Hintergrund verfolgen. Erzeugte Bilder haben dort oft eine verdächtig saubere Kante oder einen weichen Halo, echte Aufnahmen haben Tiefenunschärfe mit Verlauf.

Warum das reine Hinsehen die Schlacht verloren hat

Alle Bildmerkmale teilen eine gemeinsame Eigenschaft: Sie sind Fehler. Und Fehler werden behoben. Jedes Merkmal, das öffentlich als Erkennungszeichen gehandelt wird, landet auf der Verbesserungsliste der Modellhäuser – die Bekanntheit eines Verräters ist die Ursache seines Verschwindens. Bei den Händen hat man das in Zeitlupe beobachten können, bei der Schrift ebenfalls.

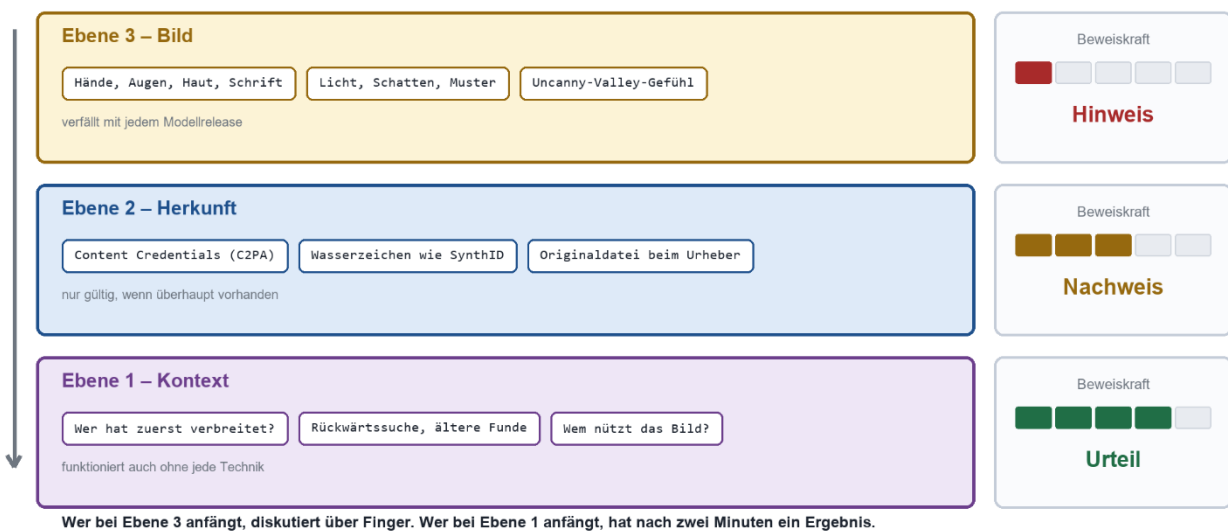
Der Stand von heute: Die Spitzenmodelle verfolgen unterschiedliche Stärken. FLUX.2 liefert den rohen Fotorealismus mit Hauttextur und dichter Typografie, Nano Banana Pro punktet bei

Anweisungstreue, Bildbearbeitung und Charakterkonsistenz über mehrere Motive hinweg. Für die Erkennung ist die Arbeitsteilung egal, das Ergebnis nicht: Für das bloße Auge ist ein sorgfältig erzeugtes Bild vom Foto nicht mehr zu trennen. Welche Werkzeuge gerade welche Stärken haben, ordnet die [Übersicht der wichtigsten KI-Bildgeneratoren](#) ein; wer lokal arbeitet, findet die Landschaft in [Die 15 wichtigsten Modelle zur lokalen Bildgenerierung](#) sortiert.

Dazu kommt ein Punkt, der in der Erkennungsdebatte fast immer untergeht: Die interessanten Fälle sind gar keine reinen KI-Bilder. Das Bild ist ein echtes Foto, bei dem ein Schild ausgetauscht, eine Person entfernt oder ein Himmel ersetzt wurde. Solche Teilmanipulationen tragen alle Echtheitssignale des Originals – Kameradaten, Rauschmuster, natürliche Unordnung – und sind mit Merkmalslisten praktisch nicht zu fassen. Wie diese Bilder überhaupt entstehen und warum Modelle Muster nachahmen statt zu verstehen, steht ausführlich in [Basiswissen KI-Bildgenerierung: Wie Diffusion funktioniert](#) sowie in [Checkpoints, Modelle und VAE](#).

Drei Ebenen, drei Beweiskräfte

Je weiter du dich vom Pixel entfernst, desto belastbarer wird die Aussage



Wer bei Ebene 3 anfängt, diskutiert über Finger. Wer bei Ebene 1 anfängt, hat nach zwei Minuten ein Ergebnis.

Drei Prüfebene mit sehr unterschiedlicher Beweiskraft. Die Ebene, die am meisten Spaß macht, liefert am wenigsten – und die langweiligste liefert das Urteil.

Damit verschiebt sich die Frage. Sie lautet nicht mehr „Sieht das Bild echt aus?“, sondern „Woher kommt es, wer steht dafür gerade, und wem nützt es?“. Das ist keine Kapitulation, sondern eine Rückkehr zu einer sehr alten journalistischen Disziplin. Fotos waren nie ein Beweis für sich selbst, sie waren ein Beweis, weil eine nachvollziehbare Kette dahinterstand. Diese Kette ist jetzt wieder das Entscheidende.

Herkunft statt Aussehen: C2PA, SynthID und die Lücken

Die Branche hat auf das Problem mit zwei technischen Antworten reagiert, die sich ergänzen und die du auseinanderhalten solltest. Die eine hängt Daten an die Datei, die andere versteckt ein Signal in den Pixeln. Beide sind seriös, beide haben eine klar benennbare Grenze.

Content Credentials: der signierte Beipackzettel

Content Credentials nach dem C2PA-Standard sind eine kryptografisch signierte Herkunftsaakte, die in der Datei mitreist: Womit wurde das Bild erzeugt oder aufgenommen, was wurde seither daran verändert, wer hat unterschrieben. Die Spezifikation ist inzwischen bei Version 2.4 angekommen, veröffentlicht im April 2026; die Vorversion 2.3 vom Januar 2026 hatte die

Herkunftsnachweise auf Livestreams ausgedehnt. Wichtiger als die Versionsnummer ist das Konformitätsprogramm dahinter, das prüft, ob ein Produkt die Daten auch sicher erzeugt und korrekt validiert. Das Pixel 10 war das erste Gerät, das die höchste Konformitätsstufe für gesicherte Aufnahme und Signatur erreicht hat.

Gleichzeitig zeigt sich dort die erste unangenehme Wahrheit: Die großen Kamerahersteller bauen ihre Implementierungen bis heute auf der älteren Architektur der Version 1.4 auf und nehmen am Konformitätsprogramm nicht teil. Ihre Signaturen sind innerhalb des eigenen Ökosystems gültig, werden von konformen Plattformen aber nicht zwingend als vertrauenswürdig behandelt. Ein Siegel, das nur der eigene Hersteller liest, ist ein halbes Siegel.

Unsichtbare Wasserzeichen: das Signal in den Pixeln

Der zweite Ansatz arbeitet unterhalb der Wahrnehmungsschwelle. SynthID von Google DeepMind verändert bei der Erzeugung minimal die Pixel – und bei Text die Wortwahrscheinlichkeiten – so, dass ein Detektor das Muster später wiederfindet, auch nach Zuschchnitt, Skalierung, Filtern und verlustbehafteter Kompression. Bis Mai 2026 waren nach Angaben des Anbieters über 100 Milliarden Inhalte auf diesem Weg markiert. Geprüft wird entweder über ein eigenes Portal, das zunächst Redaktionen offensteht, oder ganz pragmatisch durch Nachfragen in der Gemini-App.

Bemerkenswert ist, was im Mai 2026 passiert ist: OpenAI ist dem C2PA-Lenkungsausschuss beigetreten und bettet seither SynthID-Wasserzeichen zusätzlich zum C2PA-Manifest in die Bilder aus ChatGPT und der eigenen Programmierschnittstelle ein, samt öffentlichem Prüfwerkzeug, das beides gleichzeitig abfragt. Zwei Konkurrenten, die sich auf ein gemeinsames Doppelverfahren einigen – das ist in dieser Branche selten genug, um es zu erwähnen. Für dich heißt das praktisch: Bei Bildern aus den großen kommerziellen Diensten hast du inzwischen eine realistische Chance auf einen positiven Nachweis.

Die Lücke, die sich beide Verfahren teilen

Jetzt der Teil, den die Marketingfolien gern auslassen. Soziale Netzwerke kodieren praktisch jedes hochgeladene Bild neu, um Speicherplatz und Bandbreite zu sparen. Eingebettete Metadaten überleben das nicht – Instagram, Facebook, X und YouTube entfernen C2PA-Manifeste beim Upload genauso zuverlässig, wie sie seit einem Jahrzehnt EXIF-Daten entfernen. Ein Bild mit vollständiger, korrekter Herkunftsakte kommt beim Publikum ohne jede Herkunftsakte an. Das Wasserzeichen im Pixel übersteht diese Verarbeitung besser, aber es existiert nur, wenn der erzeugende Dienst es gesetzt hat.

Und genau da klafft die zweite Lücke: Wer lokal mit offenen Modellen arbeitet, erzeugt Bilder ohne jede Markierung, wenn er sie nicht selbst einbaut. Das ist kein Vorwurf an die lokale Bildgenerierung – es ist eine Gestaltungsaufgabe. Wer eine eigene Produktionsstrecke betreibt, wie sie in [Aufbau einer skalierbaren Bild-Maschine mit ComfyUI und Flux](#) beschrieben ist, muss Kennzeichnung und Nachweis zum Bestandteil des Workflows machen. Die Abwägung zwischen beiden Welten samt ihrer Compliance-Folgen steht in [Cloud vs. lokale KI-Bildgenerierung](#).

Verfahren	Was ein Treffer beweist	Was ein Fehlen bedeutet	Übersteht Upload und Weitergabe
Content Credentials (C2PA)	Signierte Herkunft und Bearbeitungshistorie, prüfbar bis zum Aussteller	Nichts. Die große Mehrheit aller Bilder trägt keine.	Nein – Plattformen entfernen die Manifeste beim Neukodieren
Unsichtbares Wasserzeichen (z. B. SynthID)	Das Bild stammt aus einem markierenden Dienst	Nichts. Nur markierende Anbieter setzen es überhaupt.	Meist ja – es überlebt Zuschchnitt, Skalierung und Kompression
Sichtbares Badge im Bild	Der Urheber kennzeichnet freiwillig und erkennbar	Nichts – lässt sich wegschneiden	Ja, solange niemand den Bildausschnitt ändert

Erkennungsdienst (Detektor)	Eine Wahrscheinlichkeit, sonst nichts	Eine Wahrscheinlichkeit, sonst nichts	Unabhängig davon – aber die Trefferquote sinkt mit jedem Modellrelease
Eigener Produktionsnachweis	Modell, Prompt, Seed und Freigabe im eigenen Katalog belegbar	Gilt nur für eigene Bilder	Ja – der Nachweis liegt im System, nicht in der Datei

Der teuerste Denkfehler: fehlende Kennzeichnung als Echtheitsbeweis

Aus „kein Wasserzeichen gefunden“ folgt nicht „echtes Foto“. Es folgt ausschließlich: kein Nachweis. Die meisten Bilder im Umlauf tragen keinerlei Markierung – weil sie aus nicht markierenden Werkzeugen stammen, weil sie über eine Plattform gelaufen sind, oder weil jemand die Metadaten bewusst entfernt hat, wofür ein einziger Befehl reicht.

Umgekehrt gilt die gute Nachricht: Ein Treffer ist belastbar. Herkunftsprüfung liefert starke positive Aussagen und praktisch keine negativen. Wer das nicht trennt, baut seine Kommunikation auf einer Beweisumkehr auf, die es nicht gibt.

Erkennungsdienste: Wahrscheinlichkeit, kein Urteil

Es gibt eine ganze Reihe von Diensten, die per Upload ausrechnen, wie wahrscheinlich ein Bild KI-generiert ist. Sie werden gern als Schiedsrichter benutzt, und genau dafür sind sie nicht geeignet. Sie sind selbst Klassifikatoren, trainiert auf den Ausgaben bekannter Modelle, und ihre Trefferquote fällt in dem Moment, in dem ein neues Modell erscheint, das sie nicht kennen.

Was ein Test von fünf verbreiteten Erkennungsdiensten ergab

NewsGuard hat im Mai 2026 fünf gängige Bilddetektoren mit 45 Bildern geprüft: 15 authentische Aufnahmen von Nachrichtenagenturen und großen Redaktionen, dazu jeweils leicht und stark bearbeitete Fassungen derselben Motive.

Im Schnitt erklärten die Dienste echte Bilder in 13,33 Prozent der Fälle für KI-generiert. Der schwächste Dienst lag bei 40 Prozent Falschalarm auf echten Fotos, ein weiterer bei 20 Prozent. Zwei Dienste kamen bei den authentischen Bildern ohne Fehlalarm durch.

Der eigentlich aufschlussreiche Wert steht am Ende: Bei 35 der 45 Bilder kam mindestens ein Dienst zu einem anderen Urteil als die übrigen. Wer sich einen Detektor aussucht, sucht sich damit auch ein Ergebnis aus.

Was folgt daraus für die Praxis? Ein Detektor ist ein Trichter, kein Richter. Er darf entscheiden, welche von zweihundert eingereichten Bildern du dir zuerst genauer ansiehst. Er darf niemals allein darüber entscheiden, ob jemand getäuscht hat. Der Unterschied klingt akademisch, bis der erste Fall im Haus ist: Wenn eine Bewerbungsunterlage, ein Schadensfoto einer Versicherung oder ein Lieferantennachweis auf Grundlage eines Prozentwerts abgelehnt wird, steht am Ende eine Behauptung im Raum, die du vor niemandem begründen kannst.

Detektorwerte gehören nicht in Entscheidungen über Menschen

Personalauswahl, Leistungsbewertung, Leistungsablehnung, Verdachtsäußerungen gegenüber Beschäftigten oder Geschäftspartnern: Hier hat ein Wahrscheinlichkeitswert aus einem Erkennungsdienst nichts zu suchen. Falsch-Positive treffen immer denjenigen, der sich am schlechtesten wehren kann.

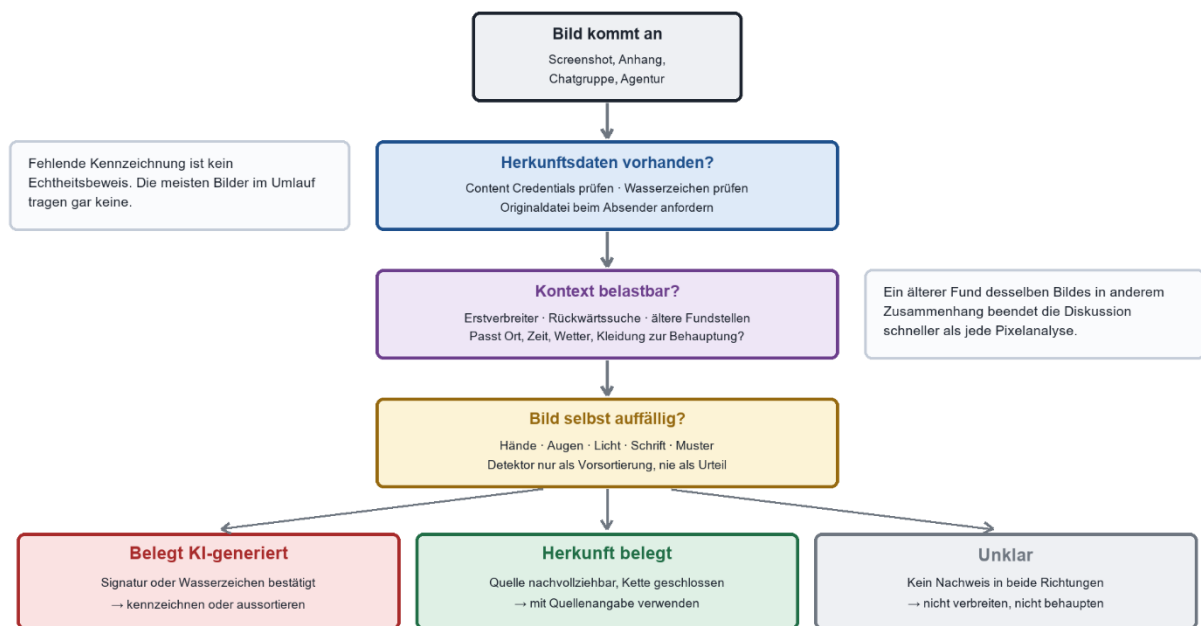
Tragfähig ist der umgekehrte Weg: nicht Verdacht automatisieren, sondern Nachweis ermöglichen. Wer ein Originalformat mit intakter Herkunftsakte anfordert, verlagert die Frage von „beweise deine Unschuld“ zu „zeig mir die Datei, mit der du gearbeitet hast“.

Die Prüfroutine, die im Alltag hält

Genug Theorie. Was du brauchst, ist ein Ablauf, den auch jemand befolgen kann, der sich für Diffusionsmodelle nicht interessiert, und der eine Aussage produziert, die vor der Geschäftsführung Bestand hat. Die Reihenfolge ist dabei nicht beliebig – sie geht vom Belastbaren zum Weichen und nicht umgekehrt.

Der Prüfablauf

Vom eingehenden Bild zu einer Aussage, die du auch begründen kannst



Der Ablauf von oben nach unten, mit drei möglichen Ausgängen. „Unklar“ ist ein zulässiges Ergebnis und in vielen Fällen das ehrlichste.

Schritt eins: nach der Originaldatei fragen

Der wirkungsvollste Handgriff kostet keine Technik. Frag den Absender nach der Originaldatei, nicht nach dem Screenshot aus der Chatgruppe. Screenshots sind Herkunftsvernichter: kein Manifest, keine Kameradaten, oft nicht einmal die ursprüngliche Auflösung. Kommt die Originaldatei, prüfst du die Content Credentials und lässt ein Wasserzeichen abfragen. Kommt sie nicht, hast du schon ein Ergebnis, nur eben nicht das technische.

Schritt zwei: Kontext und Rückwärtssuche

Wer hat das Bild zuerst verbreitet, und wann? Existiert dasselbe Motiv anderswo, älter, in einem anderen Zusammenhang? Eine Rückwärtssuche entlarvt nach wie vor mehr Fälschungen als jede Pixelanalyse – auch deshalb, weil ein erheblicher Teil der Desinformation gar nicht erzeugt, sondern umetikettiert ist: echtes Foto, falsches Ereignis. Dazu die Plausibilitätsfragen, für die man keine Technik braucht: Passen Jahreszeit, Wetter, Vegetation, Kleidung, Beschilderung und Fahrzeugmodelle zu dem, was behauptet wird?

Schritt drei: erst jetzt das Bild selbst

Und wenn du bei den Bildmerkmalen ankommst, arbeitest du sie ab wie eine Checkliste und nicht wie ein Ratespiel: Licht und Schatten zuerst, danach Spiegelungen, dann lange Texte und Beschriftungen, dann Hintergrundgeometrie und Wiederholungen, ganz zum Schluss die Körperdetails. Diese Reihenfolge steht bewusst kopf gegenüber dem, was die meisten instinktiv tun – sie folgt der Haltbarkeit der Merkmale, nicht ihrer Bekanntheit.

Anlass	Angemessener Aufwand	Wer entscheidet	Was als Ergebnis reicht
Bild in der privaten Chatgruppe	zwei Minuten: Quelle und Rückwärtssuche	du selbst	„weiß ich nicht“ – und dann nicht weiterleiten
Bild für eine Unternehmenspublikation	Originaldatei, Herkunftsprüfung, Lizenznachweis	Kommunikation, nach festem Ablauf	belegbare Quelle oder Verzicht auf das Motiv
Bild als Nachweis in einem Geschäftsvorgang	Originaldatei verpflichtend, Herkunftsprüfung, Vier-Augen-Blick	Fachbereich plus Compliance	dokumentierte Prüfung, nicht nur ein Bauchgefühl
Bild in einem öffentlichen Konflikt	vollständiger Ablauf über alle drei Ebenen	Leitung, mit juristischer Rücksprache	belastbarer Nachweis oder erkennbare Zurückhaltung

Baue das Nichtwissen in den Prozess ein

Die meisten Prüfabläufe scheitern nicht an fehlenden Werkzeugen, sondern daran, dass es keinen definierten Ausgang für „unklar“ gibt. Wenn ein Prozess nur „echt“ und „KI“ kennt, wird bei jedem Zweifel geraten – und Raten fällt zugunsten dessen aus, der lauter argumentiert.

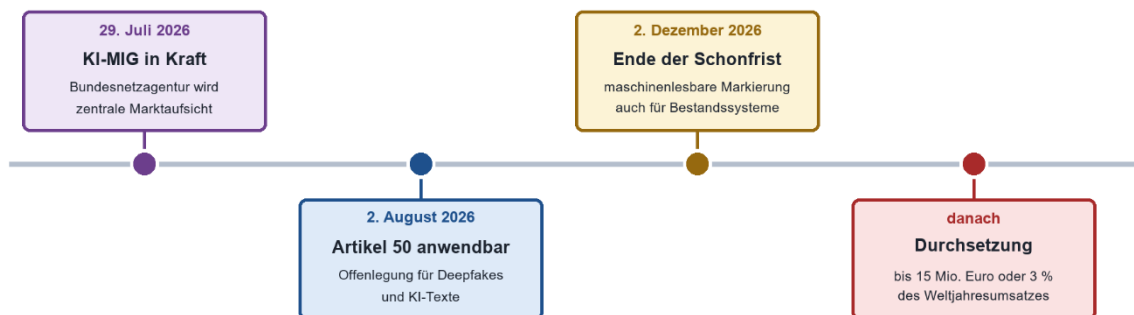
Definiere die dritte Möglichkeit ausdrücklich: kein Nachweis in beide Richtungen, also keine Verwendung und keine Behauptung. Das ist keine Schwäche, das ist die einzige Aussage, die in dieser Lage überhaupt stimmt.

Die andere Seite des Problems: selbst kennzeichnen

Bis hierher ging es um Bilder, die bei dir ankommen. Der Teil, den du vollständig in der Hand hast, ist der umgekehrte: die Bilder, die dein Haus verlässt. Und dort ist die Lage seit dem Sommer 2026 keine Stilfrage mehr, sondern geltendes Recht.

Kennzeichnen ist längst Pflicht

Was für Anbieter und Betreiber in Deutschland und der EU gilt



Wer Bilder selbst erzeugt, sitzt auf der anderen Seite des Problems: Kennzeichnung ist der Teil, den du vollständig in der Hand hast.

Die Fristen für Kennzeichnung und Offenlegung. Die Schonfrist betrifft ausschließlich die maschinenlesbare Markierung bei Bestandssystemen, nicht die Offenlegungspflicht.

Was der AI Act konkret verlangt

Die Transparenzpflichten aus Artikel 50 der KI-Verordnung sind seit dem 2. August 2026 anwendbar. Für Anbieter generativer Systeme heißt das: Die Ausgabe muss maschinenlesbar als KI-erzeugt markiert und als solche erkennbar sein. Für Systeme, die vor dem Stichtag auf dem Markt waren, läuft eine Übergangsfrist bis zum 2. Dezember 2026. Betreiber, die Deepfakes oder KI-generierte Texte zu Fragen von öffentlichem Interesse veröffentlichen, müssen die künstliche

Herkunft offenlegen; eine Ausnahme greift nur bei echter redaktioneller Verantwortung mit klarer Zurechenbarkeit. Der Rahmen sieht Sanktionen von bis zu 15 Millionen Euro oder drei Prozent des weltweiten Jahresumsatzes vor. Was davon für dein Haus tatsächlich relevant ist, ordnet das Briefing [EU AI Act: Was seit dem 2. August 2026 gilt](#) ein.

In Deutschland ist die Zuständigkeit inzwischen geklärt: Das nationale Durchführungsgesetz ist seit dem 29. Juli 2026 in Kraft und macht die Bundesnetzagentur zur zentralen Marktüberwachungs- und Koordinierungsstelle, flankiert von einem Kompetenzzentrum mit Service-Desk für Unternehmen. Der praktische Effekt: Es gibt einen Adressaten. Das ist mehr, als man über die Umsetzung mancher anderer EU-Verordnung sagen kann.

Was das im Betrieb bedeutet

Die gute Nachricht: Für die meisten Unternehmen ist das kein Projekt, sondern eine Handvoll Festlegungen. Die schlechte: Wer sie nicht trifft, merkt es erst, wenn jemand fragt. Ich kennzeichne die Bilder auf boddenberg.de sichtbar direkt im Bild – nicht im Kleingedruckten unter dem Beitrag, weil Bilder wandern und Bildunterschriften nicht mitwandern. Wie dieses Badge gemeint ist und was hinter dem Anspruch der Markenkonsistenz steckt, steht in [Markenkonsistente KI-generierte Bilder](#); warum ich das überhaupt selbst betreibe, in [Dogfooding: Ich setze ein, was ich berate](#).

Rolle	Worum es geht	Was konkret zu tun ist
Marketing und Kommunikation	Bilder, die nach außen gehen	Sichtbare Kennzeichnung festlegen, Regel schriftlich fixieren, Ausnahmen benennen
Redaktion und Fachbereiche	Bilder, die hereinkommen	Prüfablauf mit drei Ebenen einführen, Ausgang „unklar“ ausdrücklich zulassen
Produktion und Bildarchiv	Nachweisbarkeit der eigenen Bilder	Modell, Version, Prompt, Seed und Freigabe je Motiv festhalten
Recht und Compliance	Pflichten und Fristen	Anbieter- und Betreiberrolle je System klären, Fristen dokumentieren
IT	Werkzeuge und Ablage	Kennzeichnung in den Workflow einbauen, Metadaten beim Export nicht verwerfen

Nachweis führen, wenn die Datei nichts mehr verrät

Weil eingebettete Metadaten den Weg durch Plattformen nicht überleben, hilft nur ein Nachweis, der außerhalb der Datei liegt. Genau dafür betreibe ich UB.Dossier: Dort steht je Motiv, mit welchem Modell und welcher Version es entstanden ist, welcher Prompt und welcher Seed verwendet wurden, wer freigegeben hat und wo das Bild eingesetzt wird. Der Seed ist dabei mehr als Buchhaltung – er macht ein Ergebnis reproduzierbar, was in [Der Seed bei der KI-Bildgenerierung](#) und [batch_size und Reproduzierbarkeit](#) ausführlich beschrieben ist. Wer lieber auf vorhandener Microsoft-365-Infrastruktur aufsetzt, findet den Architekturweg in [Digital Asset Management für KI-Bilder mit SharePoint Online und Power Platform](#).

Die zweite Hälfte ist der Produktionsweg. In UB Forge laufen die Workflows, die Bilder erzeugen, und dort sitzt auch der Plausibilitäts-Check vor der Freigabe: Finger gezählt, Schrift gelesen, Schattenrichtung geprüft, Kennzeichnung gesetzt. Dass diese Prüfung automatisiert vor der Veröffentlichung stattfindet und nicht danach im Kommentarbereich, ist der eigentliche Gewinn. Sprachmodelle sind dabei eine brauchbare zweite Instanz – nicht als Orakel, sondern als geduldiger Prüfer, der zwanzig Kriterien abarbeitet, ohne müde zu werden; mehr dazu unter [Claude im Unternehmen](#).

Kennzeichnen ist billiger als Erklären

Der häufigste Einwand lautet, eine Kennzeichnung entwerte das Bild. Meine Erfahrung aus mehreren

Projekten sagt das Gegenteil: Wer von sich aus kennzeichnet, nimmt die Diskussion vorweg. Unangenehm wird es ausschließlich für diejenigen, denen man es nachweist.

Der Aufwand ist überschaubar – eine Badge-Vorlage, eine Regel, eine Zeile im Freigabeprozess. Verglichen mit einer Richtigstellung nach einem geplatzten Kampagnenmotiv ist das ein Schnäppchen.

Fazit

Die klassischen Verräter sind nicht verschwunden, aber sie sind zu Wahrscheinlichkeitshinweisen geschrumpft, und sie schrumpfen weiter. Hände und Schrift, die beiden bekanntesten Merkmale, tragen am wenigsten – gerade weil sie so bekannt waren. Am längsten halten die Merkmale, die Verständnis statt Musterwissen verlangen: Licht, Schatten, Spiegelungen und die Geometrie im Hintergrund. Aber auch die sind ein Verdacht und kein Beweis.

Belastbar ist heute nur der Weg über die Herkunft. Ein Treffer bei Content Credentials oder einem Wasserzeichen sagt dir etwas Verlässliches; ihr Fehlen sagt dir gar nichts, weil Plattformen Metadaten entfernen und die meisten Bilder ohnehin nie markiert waren. Erkennungsdienste bleiben eine Vorsortierung mit messbar unterschiedlichen Urteilen – als Trichter brauchbar, als Schiedsrichter untauglich, und in Entscheidungen über Menschen fehl am Platz.

Der nachhaltigere Gewinn ist deshalb keine Checkliste, sondern eine Haltung plus eine Routine: Ein Bild im Netz beweist von sich aus nichts mehr. Wer das akzeptiert, prüft in der richtigen Reihenfolge, lässt „unklar“ als Ergebnis zu und kennzeichnet die eigenen Bilder freiwillig, bevor jemand danach fragt. Den technischen Unterbau dazu – vom Verfahren über die Werkzeugwahl bis zur Governance – findest du gebündelt unter [KI-Bildgenerierung](#).

Häufige Fragen

Worauf schaue ich zuerst, um ein KI-Bild zu erkennen?

Nicht auf das Bild, sondern auf die Herkunft: Originaldatei anfordern, Content Credentials und Wasserzeichen prüfen, danach Quelle und Rückwärtssuche. Erst wenn das nichts ergibt, gehst du ins Bild – und dort in dieser Reihenfolge: Licht und Schatten, Spiegelungen, lange Texte, Hintergrundgeometrie, zuletzt Hände und Augen. Je mehr unabhängige Auffälligkeiten zusammenkommen, desto belastbarer der Verdacht.

Sind die klassischen Erkennungsmerkmale noch zuverlässig?

Immer weniger, und zwar mit erkennbarem Muster: Je bekannter ein Merkmal war, desto gezielter wurde es wegtrainiert. Sechs Finger und Kauderwelsch-Schrift sind bei aktuellen Modellen die Ausnahme. Was sich hartnäckiger hält, sind Lichtkonsistenz, Spiegelungen und Hintergrundgeometrie. Behandle alle Merkmale als Hinweise mit Verfallsdatum, nie als Beweis.

Kann ich mit einem Werkzeug sicher feststellen, ob ein Bild KI ist?

Nein. Erkennungsdienste liefern Wahrscheinlichkeiten und sind sich untereinander häufig uneins – in einem Test von fünf verbreiteten Diensten mit 45 Bildern im Mai 2026 wichen die Urteile bei 35 Bildern voneinander ab, und echte Fotos wurden im Schnitt in gut 13 Prozent der Fälle für KI-generiert erklärt. Als Vorsortierung taugen solche Dienste, als Entscheidungsgrundlage nicht.

Bedeutet ein fehlendes Wasserzeichen, dass ein Bild echt ist?

Nein, und das ist der teuerste Denkfehler in diesem Thema. Die meisten Bilder tragen überhaupt keine Markierung: weil das erzeugende Werkzeug keine setzt, weil Plattformen Metadaten beim

Hochladen entfernen, oder weil jemand sie bewusst gelöscht hat. Ein Fund ist ein Nachweis, ein Nichtfund ist keiner.

Was sind Content Credentials und bringen sie mir etwas?

Content Credentials sind eine signierte Herkunftsakte in der Datei: womit ein Bild entstanden ist, was daran verändert wurde, wer dafür unterschreibt. Der zugehörige C2PA-Standard liegt seit April 2026 in Version 2.4 vor, es gibt ein Konformitätsprogramm, und moderne Smartphones signieren Aufnahmen bereits auf der höchsten Stufe. Der Nutzen ist real – allerdings nur, solange die Datei nicht durch eine Plattform gelaufen ist, die Metadaten entfernt.

Überlebt eine Kennzeichnung den Weg durch soziale Netzwerke?

Eingebettete Metadaten in aller Regel nicht: Instagram, Facebook, X und YouTube kodieren Bilder beim Hochladen neu und werfen dabei die Manifeste. Unsichtbare Wasserzeichen im Pixel halten Kompression, Zuschnitt und Skalierung deutlich besser aus. Ein sichtbares Badge im Bild ist die robusteste Variante – es verschwindet nur, wenn jemand den Ausschnitt bewusst ändert.

Warum hat KI so oft Probleme mit Händen und Schrift?

Weil das Modell Muster nachbildet, statt zu verstehen. Hände treten in unzähligen Stellungen und Verkürzungen auf, und Schrift verlangt exakte Zeichenfolgen statt plausibler Formen – beides ist über Wahrscheinlichkeiten schwer exakt zu treffen. Die ausführliche Erklärung steht in [Warum KI keine Hände kann](#); warum Modelle auch sonst nicht einfach ausführen, was dasteht, in [Warum die KI den Prompt ignoriert](#).

Was mache ich, wenn ich es nicht sicher erkennen kann?

Den Fokus vom Bild auf den Kontext verlagern: Wer hat zuerst verbreitet, taucht das Motiv älter und anderswo auf, passt es zu Ort, Zeit und Behauptung? Und wenn auch das nichts ergibt, ist „unklar“ ein zulässiges Ergebnis. Nicht weiterverbreiten, nichts behaupten – besonders bei allem, was politisch oder persönlich brisant ist.

Muss ich KI-Bilder im Unternehmen kennzeichnen?

In vielen Fällen ja. Die Transparenzpflichten aus Artikel 50 der KI-Verordnung sind seit dem 2. August 2026 anwendbar; für die maschinenlesbare Markierung bestehender Systeme läuft eine Frist bis zum 2. Dezember 2026. Betreiber müssen Deepfakes und KI-Texte zu Fragen von öffentlichem Interesse offenlegen. In Deutschland ist die Bundesnetzagentur seit dem 29. Juli 2026 die zuständige Marktaufsicht. Welche Pflichten dich genau treffen, hängt an deiner Rolle – die Einordnung dazu steht in [EU AI Act: Was seit dem 2. August 2026 gilt](#).

Was ist mit Bildern, die nur teilweise KI sind?

Das ist der praktisch wichtigste und am schlechtesten erkennbare Fall: ein echtes Foto, bei dem eine Person entfernt, ein Schild ausgetauscht oder der Himmel ersetzt wurde. Solche Bilder tragen alle Echtheitssignale des Originals. Merkmalslisten helfen hier kaum; nur eine Bearbeitungshistorie in den Content Credentials oder der Vergleich mit einer älteren Fassung des Originals bringt Klarheit.

Hilft es, sich die EXIF-Daten anzusehen?

Als Zusatzinformation ja, als Beweis nein. Kameradaten, Objektiv und Aufnahmezeit können eine Geschichte stützen, lassen sich aber in Minuten fälschen und verschwinden beim Hochladen auf den meisten Plattformen ohnehin. Fehlende EXIF-Daten sind kein Verdachtsmoment, sondern der Normalfall für alles, was über ein soziales Netzwerk zu dir gekommen ist.

Darf ich öffentlich behaupten, jemand habe ein KI-Bild verbreitet?

Nur mit einem Nachweis, nicht mit einem Verdacht. Ein Detektorwert oder ein auffälliges Detail reichen dafür nicht – professionell ausgeleuchtete und retuschierte Fotos zeigen genau dieselben Auffälligkeiten. Der saubere Weg ist die Rückfrage beim Urheber mit der Bitte um die Originaldatei. Was danach kommt oder eben nicht kommt, ist deutlich aussagekräftiger als jede Bildanalyse.

Passend zum Thema

- › Warum ausgerechnet Hände so hartnäckig schwierig sind: [Warum KI keine Hände kann](#)
- › Warum Modelle nicht einfach ausführen, was im Prompt steht: [Warum die KI den Prompt ignoriert](#)
- › Wie diese Bilder technisch überhaupt entstehen: [Basiswissen KI-Bildgenerierung: Wie Diffusion funktioniert](#) · [Checkpoints, Modelle und VAE](#)
- › Welche Werkzeuge gerade welchen Realismus liefern: [Übersicht der wichtigsten KI-Bildgeneratoren](#) · [Die 15 wichtigsten Modelle zur lokalen Bildgenerierung](#)
- › Kennzeichnung und Nachweis in der eigenen Produktion: [Aufbau einer skalierbaren Bild-Maschine mit ComfyUI und Flux](#) · [Digital Asset Management für KI-Bilder mit SharePoint Online und Power Platform](#)
- › Wo Bilder erzeugt werden und was das für Compliance bedeutet: [Cloud vs. lokale KI-Bildgenerierung](#) · [KI-Bildgenerierung im Unternehmen – Die strategische Alternative zur Cloud](#)
- › Wie ich selbst kennzeichne und warum: [Markenkonsistente KI-generierte Bilder](#) · [Dogfooding: Ich setze ein, was ich berate](#)
- › Was der AI Act seit August 2026 konkret verlangt: [EU AI Act: Was seit dem 2. August 2026 gilt](#)
- › Reproduzierbarkeit als Nachweisgrundlage: [Der Seed bei der KI-Bildgenerierung](#) · [batch_size und Reproduzierbarkeit](#)
- › Prompt-Arbeit direkt in Office: [Black Forest FLUX.2 in PowerPoint und Copilot](#)
- › Sprachmodelle als geduldige Prüfinstanz: [Claude im Unternehmen](#)
- › Beratung zur Prüf- und Kennzeichnungsroutine: [KI-Bildgenerierung Beratung](#) · Schulung für das Team: [KI-Bildgenerierung Schulung](#)