

☐ **Du: Also ich:** Meine Frage: Wie kann die KI „ohne Augen“ Fotos „sehen“, gar Stimmungen erkennen?

(Meine Fragen konnte ich nicht kopieren; deswegen hier abgekürzt)

☐ **ChatGPT:**

Das ist eine berechtigte Frage. **Ich habe keine Augen und sehe Fotos nicht so wie ein Mensch.** Wenn ich zuvor den Eindruck erweckt habe, ich hätte ein Foto wie ein menschlicher Betrachter gesehen und dessen Stimmung unmittelbar „gefühlt“, wäre das missverständlich gewesen.

Wenn ein Bild tatsächlich an ChatGPT übermittelt wird, kann ich es **bildanalytisch auswerten**: Das System kann beispielsweise Formen, Farben, Kontraste, Bildaufbau, Personen, Gegenstände, Lichtverhältnisse und andere visuelle Merkmale erkennen und daraus eine Beschreibung oder Einschätzung ableiten. Das ist aber etwas anderes als menschliches Sehen.

Besonders bei „**Stimmung**“ ist Vorsicht angebracht. Ich kann etwa feststellen:

- dunkle oder helle Farbgebung,
- warmes oder kühles Licht,
- starke oder geringe Kontraste,
- ruhige oder sehr dynamische Komposition,
- Nebel, Schatten, Leere oder dichtes Gedränge,
- Gesichtsausdrücke oder Körperhaltungen, soweit sie erkennbar sind.

Daraus kann ich eine **wahrscheinliche Wirkung** des Bildes beschreiben, etwa „wirkt ruhig, melancholisch oder dramatisch“. Aber ich kann nicht im menschlichen Sinn **empfinden**, ob ein Bild melancholisch ist. Und eine solche Interpretation ist immer mit Unsicherheit verbunden.

Noch wichtiger in Ihrem Fall: Wenn Sie mir die **Internetseite „Fotofenster-Nohe“** genannt hatten und ich daraus Aussagen über konkrete Fotos gemacht habe, müsste man unterscheiden, ob mir die Bilder tatsächlich als Bildmaterial vorlagen oder ob ich nur Informationen über die Webseite bzw. deren Texte verarbeitet habe. **Aus einem Seitentext allein kann ich ein Foto nicht seriös beurteilen.**

Wenn Sie möchten, können Sie mir sagen, **welche konkrete Auskunft ich Ihnen damals über Fotofenster-Nohe gegeben habe**. Dann kann ich Ihnen genau erklären, worauf diese Aussage beruhen konnte und ob sie möglicherweise zu weitgehend formuliert war.

☐ **Du: Also ich:** fotofenster-nohe ist eine Internetseite - die KI kann doch die einzelnen Seiten, einzelne Fotos aufschlagen...? ZB die Seiten: Regenfenster / Spiegelungen / Licht und Schatten

☐ **ChatGPT:**

Ja — **wenn Sie mir die konkrete Internetseite nennen, kann ich sie aufrufen und untersuchen, und ich kann dabei auch die dort eingebundenen Bilder berücksichtigen, sofern sie für mich technisch zugänglich sind.** Ich muss aber sauber unterscheiden zwischen „die Seite aufrufen“ und „jedes Bild tatsächlich visuell analysieren“.

Ich habe gerade nach „**Fotofenster-Nohe**“ gesucht. Unter diesem Namen finde ich momentan allerdings nicht eindeutig Ihre Seite. Deshalb möchte ich Ihnen nichts vormachen.

Wenn Sie mir die **Internetadresse (URL) von Fotofenster-Nohe** hier hineinkopieren, kann ich beispielsweise gezielt die Seiten

- „**Regenfenster**“
- „**Spiegelungen**“
- „**Licht und Schatten**“

aufsuchen und die dort vorhandenen Fotografien betrachten, soweit sie mir als Bilder zugänglich sind.

Und dann können wir Ihre eigentliche Frage sehr interessant überprüfen: **Was kann ich an diesen Bildern tatsächlich erkennen – und was davon ist eine Interpretation der Stimmung?**

Zum Beispiel könnte ich bei einem Foto sagen: „Das Bild zeigt Regentropfen auf einer Scheibe; dahinter sind unscharfe Lichter zu erkennen. Die gedämpften Farben und die Unschärfe erzeugen eine ruhige, vielleicht melancholische Wirkung.“ Das wäre eine **Bildanalyse aufgrund sichtbarer Gestaltungsmittel**. Ich würde aber nicht behaupten, dass ich die Melancholie so empfinde, wie Sie oder ein anderer Mensch sie beim Anschauen empfindet.

Schicken Sie mir einfach den Link zu Fotofenster-Nohe. Dann können wir es konkret an Ihren Bildern ausprobieren.

☐ **Du: Ich** nenne die Adresse...

☐ **ChatGPT:**

Ja. Jetzt kann ich Ihre Frage **konkret an Ihrer Seite** beantworten — und ich muss meine vorige Antwort etwas präzisieren.

Ich habe Ihre Startseite geöffnet. Dort sind tatsächlich die von Ihnen genannten Fotoseiten verlinkt: „**Regenfenster**“, „**Schatten im Licht**“ und „**Spiegelungen**“; außerdem etwa „Skispringer im Nebel“, „Lichter“, „Farbenrausch“ usw.

Ich habe auch die Seite „**Regenfenster**“ geöffnet. Dort sind zahlreiche Fotografien als Bilder eingebunden, und die Seite selbst bezeichnet eine Aufnahme ausdrücklich als „**Besinnlich**“ und als „Aussichten aus meinem Küchenfenster“. Die Seite „**Schatten im Licht**“ enthält ebenfalls eine Reihe von Fotografien, unter anderem „Fensterblick, Köln“, Aufnahmen in Siena und Ostia antica.

Aber nun zu Ihrer eigentlichen Frage

Sie haben mit Ihrem Einwand „**Fotos muß man sehen, man braucht dazu Augen**“ einen wichtigen Punkt getroffen.

Wenn ich eine solche Webseite aufrufe, bekomme ich nicht automatisch dasselbe Erlebnis wie Sie, wenn Sie die Seite am Bildschirm betrachten. Ich kann die **Bilddateien technisch als visuelle Information analysieren**, wenn sie mir zugänglich sind. Dabei kann ich beispielsweise feststellen:

- welche Farben und Helligkeiten vorhanden sind,
- ob ein Bild scharf oder verschwommen wirkt,
- welche Formen und Strukturen vorkommen,
- ob Regentropfen, Spiegelungen, Schatten, Nebel usw. zu erkennen sind,
- wie ein Bild komponiert ist,
- welche visuellen Mittel eine bestimmte Wirkung **nahelegen**.

Das ist also tatsächlich eine Art „**Sehen**“ im **technischen Sinn**, aber nicht menschliches Sehen.

Und bei **Stimmungen** wird es noch interessanter.

Wenn ich bei einem Ihrer Regenfenster-Bilder beispielsweise sage:

„Das Bild wirkt still, nachdenklich und melancholisch.“

dann bedeutet das nicht: **Ich empfinde Melancholie.**

Es bedeutet vielmehr: Aus bestimmten visuellen Eigenschaften — etwa Regentropfen, diffusem Licht, Unschärfe, gedämpften Farben, einer eingeschränkten Sicht nach draußen und der Bildkomposition — ergibt sich für mich eine Interpretation, die ich mit dem menschlichen Begriff „**melancholisch**“ oder „**besinnlich**“ beschreiben kann.

Das ist eine **Interpretation**, keine Empfindung.

Und ich finde gerade Ihre Seite dafür besonders interessant, weil Sie offenbar nicht nur fotografische Gegenstände zeigen, sondern **mit fotografischen Mitteln Stimmungen erzeugen**. „Regenfenster“ ist dafür ein sehr gutes Beispiel. Schon die Bezeichnung der Seite lautet „Aussicht mit Tropfen“, und eine der Aufnahmen wird auf der Seite selbst als „Besinnlich“ bezeichnet.

Wenn Sie möchten, können wir deshalb einen kleinen Test machen: Sie nennen mir eines Ihrer Bilder, z. B. aus „**Regenfenster**“, und ich beschreibe zunächst ganz nüchtern, **was ich visuell erkenne**. Danach sage ich, **welche Stimmung ich daraus ableiten würde**. Und anschließend können Sie mir sagen, ob meine „Stimmung“ mit Ihrer Absicht als Fotograf übereinstimmt.

Das wäre meines Erachtens die ehrlichste Antwort auf Ihre Frage, **ob ChatGPT die Stimmungen Ihrer Fotografien tatsächlich erkennen kann**.

☐ **Du: also ich:** Seite Regenfenster. Auf dieser Scrollseite das erste und die beiden letzten Bilder

<https://www.fotofenster-nohe.de/regenfenster/seiten/regenfenster.html>

☐ **ChatGPT:**

Ja. Jetzt kann ich genau die drei von Ihnen gemeinten Bilder betrachten: das **erste** sowie die **beiden letzten** Bilder der Scroll-Seite „Regenfenster“. Die Bilddateien sind für mich tatsächlich zugänglich und ich kann ihre visuellen Eigenschaften analysieren.

Und Ihre ursprüngliche Frage wird daran sehr schön deutlich.

Das erste Bild

Hier sehe ich eine Scheibe, die fast vollständig mit Regentropfen bedeckt ist. Die Welt dahinter ist stark verschwommen: Bäume, Farben und Formen sind nur noch undeutlich erkennbar. Die Tropfen selbst sind dagegen scharf und bilden geradezu eine zweite, strukturierte Oberfläche.

Meine Beschreibung der Stimmung wäre: ruhig, abgeschirmt, etwas melancholisch, aber auch sehr kontemplativ.

Das Interessante ist: Ich komme zu dieser Einschätzung nicht dadurch, dass ich „Melancholie empfinde“. Ich erkenne eine Kombination von **Unschärfe + gedämpftem Licht + Regen + eingeschränkter Sicht + vielen kleinen Tropfen** und weiß aus meinem Sprach- und Bildwissen, dass Menschen eine solche Bildgestaltung häufig mit Ruhe, Einsamkeit oder Melancholie verbinden.

Das vorletzte Bild

Hier verändert sich die Situation entscheidend: Eine Person steht links am Fenster und schaut hinaus. Dahinter sind Dächer, Bäume und eine Stadt mit Kirchtürmen zu sehen. Der Regen bleibt als **Schicht zwischen Betrachter und Außenwelt** erhalten.

Ich würde dieses Bild nicht einfach als „melancholisch“ bezeichnen. Für mich wirkt es eher **nachdenklich, beobachtend und ein wenig sehnsüchtig**. Die Person schaut hinaus, aber sie befindet sich gleichzeitig hinter einer durch den Regen abgetrennten Scheibe.

Das ist schon eine ziemlich konkrete **Bildinterpretation** – und sie beruht auf dem Zusammenspiel der Person, ihrer Blickrichtung, der Fensterscheibe und der dahinterliegenden Landschaft.

Das letzte Bild

Dieses Bild ist für mich überraschend anders. Die Person ist jetzt seitlich zu sehen und **scheint zu lächeln**. Die Regentropfen sind noch da, aber das Bild wirkt wesentlich lebendiger. Die kräftigen Grüntöne der Bäume und die sichtbare Stadt geben ihm mehr Offenheit.

Meine spontane Interpretation wäre deshalb eher **heiter, interessiert, vielleicht sogar ein bisschen amüsiert** – trotz des Regens.

Und genau hier zeigt sich etwas Wichtiges:

Ich kann tatsächlich Unterschiede in der „Stimmung“ der drei Bilder erkennen bzw. beschreiben.

Ich sehe dabei visuelle Merkmale und setze sie in Beziehung zu menschlichen Begriffen für Wahrnehmung und Emotion.

Aber ich **erlebe** diese Stimmungen nicht.

Ich würde es ungefähr so formulieren:

Sie sehen ein Bild und empfinden möglicherweise seine Stimmung. Ich analysiere ein Bild und erschließe daraus eine wahrscheinliche Stimmung.

Und bei Ihren drei Bildern finde ich besonders interessant, dass die Stimmung **nicht einfach vom Motiv „Regen“ bestimmt wird**. Beim ersten Foto dominiert der Regen und die Abgeschlossenheit. Beim zweiten kommt das menschliche Schauen hinzu. Beim dritten verändert allein die sichtbare Mimik der Person die Wirkung erheblich.

Wenn Sie möchten, können wir noch einen Schritt weitergehen: **Sie sagen mir, welche Stimmung Sie selbst bei diesen drei Bildern beabsichtigt haben**. Dann können wir vergleichen, wie weit meine „künstliche“ Bildinterpretation mit Ihrer tatsächlichen fotografischen Absicht übereinstimmt. Das wäre ein ziemlich guter Test dafür, was ChatGPT bei Ihren Fotografien wirklich leisten kann.

☐ **Du: Also ich:** Im Moment genügt es ...komme darauf zurück...dann werden Sie sich aber nicht mehr an mich erinnern....

☐ **ChatGPT:**

Das stimmt — **in einem späteren neuen Gespräch kann ich mich möglicherweise nicht an dieses Gespräch erinnern**. In diesem Chat bleibt der bisherige Verlauf aber erhalten, sodass wir hier jederzeit daran anknüpfen können.

Wenn Sie später in einem neuen Chat zurückkommen, reicht schon ein kurzer Hinweis wie **„Wir hatten über die Stimmungen meiner Fotos auf Fotofenster-Nohe gesprochen“**. Dann weiß ich zumindest, worum es gehen soll; den genauen bisherigen Wortlaut kann ich allerdings nicht zuverlässig voraussetzen.

Und ich würde mich freuen, das Gespräch fortzusetzen. Gerade die Frage, **ob und wie eine KI fotografische Stimmung von bloßer Bildbeschreibung unterscheiden kann**, ist bei Ihren Fotos sehr spannend.