

Emotionswahrnehmung in KI-generierten Gesichtsausdrücken

Emil Ojo¹ und Mathias Walther^{1,*} 

¹Technische Hochschule Wildau, Germany

*Correspondence: Mathias Walther, mathias.walther@th-wildau.de

Abstract. Diese Studie analysiert, ob Gesichtsausdrücke in generierten Bildern genauso wahrgenommen werden wie in realen Bildern. Dabei werden die Wahrnehmung der Emotionsintensität und Sympathie, die Fähigkeit zur Emotionserkennung sowie die Unterscheidung zwischen echten und generierten Bildern untersucht. Zur Beantwortung dieser Fragen wurde eine Online-Umfrage mit mehr als 200 Teilnehmern durchgeführt. Die Ergebnisse zeigen geringfügige bis moderate Unterschiede in der Wahrnehmung zwischen generierten und realen Bildern, allerdings wird auch deutlich, dass die Unterscheidung zwischen beiden zunehmend schwieriger wird.

Keywords: Künstliche Intelligenz, Generative KI, Emotionswahrnehmung

1. Grundlagen

1.1 Motivation

Generative KI und KI-Bild-Generatoren wie „DALL-E“, „Midjourney“ und „Stable Diffusion“ stehen aufgrund ihrer zunehmenden Verbreitung seit einiger Zeit im Zentrum zahlreicher Diskussionen. Bildgeneratoren ermöglichen das synthetische Erstellen von Bildern jeglicher Art: von abstrakten Kunstwerken bis hin zu fotorealistischen Porträts. Das kann für Unternehmen, besonders im Bereich des Marketing und Online-Brandings, viele Vorteile bieten. Beispielsweise erreichen die KI-Influencer „Lilmiquela“ und „Imma.gram“ ein Publikum von 2,6 Millionen und 391.000 Followern, was sie für viele Firmen als Werbepartner attraktiv macht [1], [2]. Infolgedessen hat bereits der Autohersteller BMW mit „Lilmiquela“ kooperiert und ihre Persona in Foto- und Videowerbespots miteinbezogen [3], [4]. Als weitere Beispiele für die Zusammenarbeit zwischen Unternehmen und KI-Influencern dienen die Modemarken Prada und Calvin Klein [5], [6]. Daraus lässt sich mutmaßen, dass in Zukunft der Einsatz von generativer KI im Marketing vermehrt zunehmen könnte. Darüber hinaus beweist generative KI auch im Bereich der Mensch-Maschinen-Interaktion Potenzial: Beispielsweise könnte generative KI die Produktivität von Angestellten steigern, wie aus der Studie von Brynjolfsson, Li und Raymond hervorgeht [7]. Ein weiteres Anwendungsfeld für generative KI stellen medizinische und therapeutische Anwendungen dar, die Avatare nutzen, um mit dem Patienten zu kommunizieren. Zur Zeit müssen künstliche Avatare per Hand entwickelt werden, was sehr zeitintensiv ist [8]. Hier würde die Nutzung von realistischen

synthetisch erstellten Avataren die Entwicklungszeit solcher Applikationen drastisch verkürzen. Was für all diese Einsatzmöglichkeiten jedoch von entscheidender Bedeutung ist, ist wie überzeugend menschlich die generierte Persona wirkt, was wiederum maßgeblich von ihrer Fähigkeit zur menschlichen Kommunikation, insbesondere des nonverbalen Kommunizierens und des Kommunizierens von Emotionen, abhängt. Um letzteres Thema zu untersuchen, wurde die Wahrnehmung von Grundemotionen in generierten Gesichtsausdrücken von E. Ojo [9] untersucht. Dieser Artikel fasst die Haupterkenntnisse der Arbeit zusammen.

1.2 Nonverbale Kommunikation von Emotionen

Die menschliche Kommunikation umfasst sowohl auditive als auch visuelle Elemente, wie etwa Bewegungen beim Sprechen, Gesichtsausdrücke oder die Körperhaltung des Sprechers [10]. Ein wesentlicher Teil der Kommunikation betrifft daher nicht nur die Sprache selbst sondern auch Gestik und Mimik. Diese nonverbale Kommunikation kann als „Verhalten des Gesichts, Körpers, oder der Stimme ohne den sprachlichen Inhalt, [...]“ [11] angesehen werden. Es ist noch immer unklar, welchen genauen Anteil nonverbale Kommunikation in der Kommunikation einnimmt. Es gibt eine oft zitierte und falsch interpretierte Aussage von Mehrabian, die nahelegt, dass 55% der Kommunikation aus Gesichtsausdrücken, 38% aus der Tonalität der Stimme und nur 7% aus dem gesprochenen Wort bestehen [12]. Allerdings gab es erhebliche Kritik im Bezug auf Methodik, Stichprobengröße, Versuchsaufbau, der statistischen Auswertung und allgemeinen Anwendbarkeit von Mehrabians Studie, die zu dieser Aussage führte [13]. Auch wenn Mehrabians Erkenntnisse nicht generalisiert werden können, könnten seine Experimente trotzdem andeuten, welche Bedeutung Gesichtsausdrücke in der Kommunikation einnehmen und dass sie möglicherweise sogar eine größere Rolle spielen als das gesprochene Wort. Dies gilt umso mehr für Bilder, denen naturgemäß jegliche verbalen Aspekte der Kommunikation fehlen, was den Gesichtsausdruck in beispielsweise einem Porträt sicherlich zum bedeutendsten Kommunikationsmittel macht [14].

Die nonverbalen Kommunikationskanäle übertragen Informationen in Form von sogenannten sozialen Signalen, die Ausdruck und Kommunikationsmittel von sozialer Intelligenz und grundlegendem Sozialverhalten sind, das sowohl Zustimmung und Ablehnung als auch Empathie, Höflichkeit umfasst [15]. Zu diesem kommunizierten Signalen gehören insbesondere auch Emotionen [16], welche im Folgenden untersucht werden sollen. Auf Grund der langen Forschungsgeschichte existieren mittlerweile über 100 verschiedene Definitionen von Emotionen in verschiedenen wissenschaftlichen Disziplinen, wie zum Beispiel der Physiologie, der Lernforschung oder der Anthropologie [17]. Die ältesten stammen aus der Mitte des 19. Jahrhunderts und berufen sich auf Darwin [16], [18]. In der Forschung werden Emotionen oft durch einen kategorischen Ansatz beschrieben. Dieser basiert auf sogenannten Basisemotionen und beinhaltet die folgende Annahme: Einige Emotionen, die als Primär- oder Basisemotionen bezeichnet werden, haben sich im Verlauf der Evolution als Lösung für spezifische evolutionäre Herausforderungen entwickelt. Dabei beinhalten die meisten Definitionen folgende Grundemotionen: Freude, Ärger, Wut, Angst, Sorge und Trauer, wobei sich Trauer noch weiter in stille und weinende Trauer unterteilen lässt [19]. Ein Ansatz, welcher den Kern vieler dieser Definitionen repräsentiert, ist die Definition der sieben Grundemotionen nach Ekman. Diese Definition umfasst die Emotionen Ärger, Verachtung, Ekel, Angst, Freude, Trauer und Überraschung [20] und dient als wissenschaftliche Grundlage der in diesem Artikel beschriebenen Studie. Abbildung 1 zeigt typische Beispiele

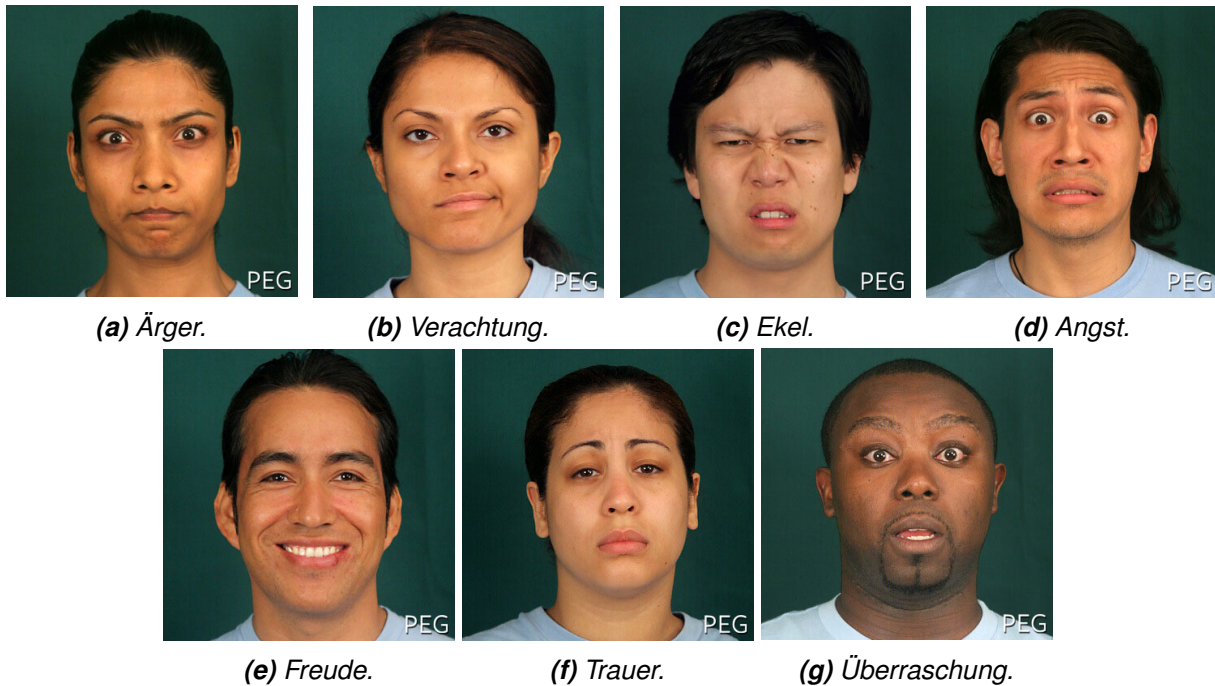


Abbildung 1. Die sieben Grundemotionen nach Ekman [20].

zur Veranschaulichung der entsprechenden Gesichtsausdrücke. Eine ausführliche Beschreibung kann in [20], [21] eingesehen werden.

2. Methode

2.1 Emotionsdatenbanken

Für die Studie wurde ein Satz von echten Porträts mit den sieben Grundemotionen von Ekman benötigt. Basierend auf einem Katalog von Porträt-Datenbanken [22] wurden daher verschiedene Quellen hinsichtlich der folgenden vier Aspekten evaluiert: Erstens wurde geprüft, ob die Datenbank spezifisch Gesichtsausdrücke enthält, da der Katalog auch zahlreiche Datenbanken mit neutralen Gesichtsbildern umfasst. Zweitens wurde bewertet, ob die Datenbank frei zugänglich ist. Drittens wurde betrachtet, wie viele von Ekmans Grundemotionen durch die Datenbank abgebildet werden. Zuletzt die Bildqualität geprüft. Insgesamt erfüllten sechs Datenbanken die Grobanforderungen:

1. Amsterdam Dynamic Facial Expression Set (ADFES) [23],
2. Cohn-Kanade AU-Coded Facial Expression Database (CKFED) [24]
3. Complex Emotion Expression Database (CEED) [25],
4. FACES database (FACES) [26],
5. FG-NET Database with Facial Expressions and Emotions (FG-NET) [27],
6. Karolinska Directed Emotional Faces (KDEF) [28],

Die CKFED und CEED Datenbanken wurden ausgeschlossen, da sie nicht zugänglich waren. FG-Net und KDEF wurden ebenso ausgeschlossen, da die Bilder entweder nicht die erforderliche Bildqualität hatten beziehungsweise nicht genügend von Ekmans Grundemotionen enthielten. Die endgültige Wahl fand zwischen der ADFES und FACES Datenbank statt, wobei die Bildqualität der entscheidende Faktor war. Schlussendlich wurde die FACES-Datenbank ausgewählt, auch wenn diese zwei Grundemotionen, Überraschung und Verachtung, nicht beinhaltete.

2.2 Bilderzeugung mit generativer KI

Im Frühling 2024 wurden einige KI-Bildgeneratoren getestet, darunter DALL-E 3, Midjourney und Stable Diffusion (v1.5), wobei Stable Diffusion die vielversprechendsten Ergebnisse lieferte. Dieses Ergebnis spiegelte auch die Studie von Borji wider, welche einen quantitativen Vergleich zwischen den drei genannten Bildgeneratoren zieht und ebenfalls zu dem Ergebnis kam, dass Stable Diffusion zu dem Zeitpunkt die besten Resultate liefert [29]. Daher wurde Stable Diffusion zum Mittel der Wahl für diese Studie.

Die Bilder für das Experiment wurden wie folgt generiert: Zuerst wurden Standardprompts geladen, um einige generellen Eigenschaften des Bildes festzulegen. Diese wurden dann um Negativprompts ergänzt, um Abnormalitäten in den Bildern zu vermeiden. Die genutzten Negativprompts entsprachen dabei größtenteils den für das Modell empfohlenen Negativprompts [30].

Standardprompts: portrait, symmetrical, centered, head and shoulders, neutral grey background, neutral shot, neutral lighting, average looking, imperfect, expressive, grey t-shirt

Standard-Negativprompts: (deformed iris, deformed pupils, semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, anime), text, cropped, out of frame, worst quality, low quality, jpeg artifacts, ugly, duplicate, morbid, mutilated, poorly drawn face, mutation, deformed, blurry, dehydrated, bad anatomy, bad proportions, extra limbs, cloned face, disfigured, gross proportions, long neck, black and white

Beispiel eines emotionsspezifischen Prompts (Ekel - Frau): middle-aged woman, disgusted expression, straight simple hair, teeth

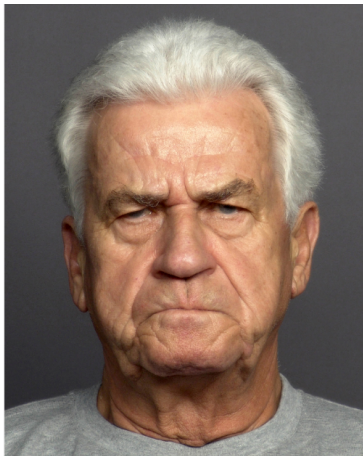
Abbildung 2 zeigt die wichtigsten Schritte des Generierungsprozesses. Ziel war es, Bilder zu generieren, die den Bildern der FACES-Datenbank so ähnlich wie möglich waren (Abb. 2a). Daher beinhaltete der Prompt auch Stichwörter, um die T-Shirts und den Hintergrund der generierten Bilder den T-Shirts und dem Hintergrund der FACES-Bilder anzupassen. Zusätzlich wurde meist noch das Werkzeug „ControlNet“ [31] verwendet, um die Position des Gesichts in der Mitte des Bildes zu zentrieren mit geradem Blick nach vorne. Mit diesen Einstellungen wurden Sätze zu je vier Bilder generiert, bis ein Bild passend schien. In einigen Fällen, in denen die Prompts alleine keine befriedigenden Ergebnisse erzielen konnten kamen sogenannte LoRA-Modelle zum Einsatz, welche Stable-Diffusion-Modelle nochmal nachjustieren [32]. Erfreulicherweise existieren LoRA-Modelle, welche spezifisch für das Generieren von Gesichtsausdrücken erstellt wurden fünf von Ekmans sieben Grundemotionen abdecken: Überraschung, Trauer, Ärger, Angst, Ekel [30]. Weiterhin ließ sich die Freude auch leicht ohne Nutzung eines LoRA-Modells generieren. Daher ließen sich alle von Ekmans Grundemotionen außer Verachtung generieren. Abbildung 2b zeigt ein Beispiel. Zum Schluss wurden die generierten Bilder noch leicht nachbearbeitet, um dem Stil der realen Gegenstände noch stärker zu entsprechen (Abb. 2c). Das beinhaltete meistens die Anpassung von Kontrast und Farbe des Hintergrunds.

2.3 Umfrage

Für die Studie wurden folgende zwei Hypothesen aufgestellt:

Hypothese 1: Emotionen werden in generierten Bildern gleich wahrgenommen wie in echten Bildern.

Gleich wahrgenommen zu werden bedeutet diesbezüglich, dass keine bedeutsamen Unterschiede zwischen echten und generierten Bildern festgestellt werden



(a) Reales Bild der
FACES-Datenbank [26].



(b) Mit Stable Diffusion
erzeugtes Bild.



(c) Generiertes und
nachbearbeitetes Bild.

Abbildung 2. Generierung und Nachjustierung von den Bildern am Beispiel von Ärger.

können im Bezug auf die Intensität der Emotionen, die Fähigkeit die dargestellten Emotionen klar zu erkennen und im Bezug auf die empfundene Sympathie hinsichtlich den Emotionen.

Hypothese 2: Generierte und echte Bilder können nicht unterschieden werden.

Um diese Hypothesen zu testen, wurde eine Online-Umfrage durchgeführt, in welcher Probanden Fragen zu jeweils sechs generierten und sechs echten Bildern gestellt wurde. Auf den Bildern waren fünf von Ekman's sieben Grundemotionen, sowie ein neutraler Gesichtsausdruck, beides jeweils einmal echt und einmal generiert, abgebildet. Die Emotionen die von der Umfrage nicht abgedeckt wurden waren „Verachtung“ und „Überraschung“. Für beide fehlte in der FACES Datenbank ein Vergleichsbild und weiterhin lies sich die Verachtung mit dem genutzten Modell nicht generieren, weswegen beide Emotionen in der Umfrage ausgelassen wurden. Insgesamt sollten die Teilnehmer daher Bilder zu folgenden sechs Emotionen evaluieren: Ärger, Ekel, Freude, Angst, Trauer und einen neutralen Ausdruck. Da das Umfrage-Tool keine Zufallsreihenfolge der Bilder unterstützte, wurden allen Teilnehmern die Bilder in gleicher Reihenfolge gezeigt. Die Reihenfolge wurde bei Erstellung der Umfrage zufällig festgelegt. Jeder Teilnehmer musste die selben vier Fragen zu jedem Bild beantworten:

1. „Welche Emotion wird auf dem aktuellen Bild dargestellt?“
2. „Wie intensiv wird diese Emotion dargestellt?“
3. „Wie sympathisch wirkt die Person auf dem Bild?“
4. „Ist das Bild echt oder generiert?“

Bei Frage 1 gab es zusätzlich zu den 7 oben beschriebenen Zuständen noch die Antwortoption „Überraschung“.

Die Umfrage wurde im Mai 2024 unter Studierenden und Mitarbeitern der Technischen Hochschule Wildau durchgeführt. Insgesamt schlossen 219 Teilnehmer die Umfrage ab. Die Geschlechterverteilung belief sich auf 50% (109) Männer, 49% (108) Frauen und 1% (2) Divers, was ein erwartbares Ergebnis war. Dabei lag die Altersspanne zwischen 18 und 62 Jahren, wobei der Median bei 25 Jahren lag und das häufigste Alter ebenfalls 25 Jahre war. Auch die Altersverteilung war erwartbar für eine Hochschule.

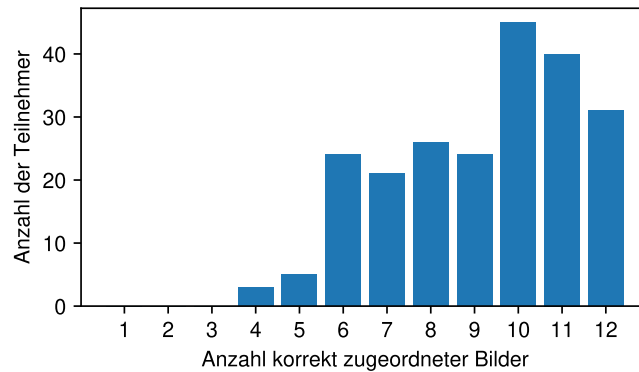


Abbildung 3. Unterscheidung von echten und generierten Bildern [9].

Tabelle 1. Übergreifender Vergleich zwischen echten und generierten Bildern [9].

	Intensität	Sympathie	Unterscheidungs- güte (echt/gen.)	Emotionserken- nung
Echt	3,94	2,88	70 %	73 %
Generiert	3,62	2,78	83 %	72 %

3. Ergebnisse

3.1 Gesamtübersicht

Abbildung 3 zeigt die Ergebnisse der Teilnehmer bei der Unterscheidung zwischen echten und generierten Bildern. Es wurde gezählt, wie viele der 12 Bilder korrekt zugeordnet werden konnten. Wie in Abbildung 3 zu sehen, erzielte der Großteil der Teilnehmenden ein gutes Ergebnis von 10 bis 12 korrekten Zuordnungen. Das spiegelt sich auch in Median und Modalwert wieder, die beide bei 10 liegen. Weiterhin gab es niemanden, der weniger als vier korrekte Zuordnungen schaffte. Außerdem erzielten nur weniger als ein Viertel (32) der Teilnehmenden sechs Treffer oder weniger.

Tabelle 1 zeigt, wie unterschiedlich echte und generierte Bilder wahrgenommen wurden. Für echte und generierte Bilder sind jeweils die Mittelwerte aller Antworten für Intensität und empfundene Sympathie eingetragen. Die Skala liegt jeweils zwischen 1 und 5, von „überhaupt nicht intensiv“ (1) bis „sehr intensiv“ (5) bzw. „überhaupt nicht sympathisch“ (1) bis „sehr sympathisch“ (5). Die Unterscheidungsgüte stellt dar, in wie viel Prozent der Fälle die echten Bilder als echt und die generierten Bilder als generiert klassifiziert wurden. Die Spalte Emotionserkennung gibt an, in wie viel Prozent der Fälle die dargestellten Emotionen der richtigen Grundemotion zugeordnet wurden.

Auf den ersten Blick impliziert Tabelle 1 keinen Unterschied in der Wahrnehmung von Emotionen zwischen echten und generierten Bildern. Insbesondere die Fähigkeit zur Erkennung der dargestellten Emotionen scheint sehr ähnlich für echte und generierte Bilder. Wo allerdings ein merkbarer Unterschied vorliegt, ist in der Unterscheidungsgüte. Hier wurden echte Bilder in 70 % der Fälle korrekt klassifiziert, während generierte Bilder zu 83 % korrekt identifiziert wurden. Das bedeutet, dass die Teilnehmer Bilder tendenziell als generiert angesehen haben, was möglicherweise einen generellen Skepsis nahelegt. Um die statistische Signifikanz aller Ergebnisse zu überprüfen, wurde ein Mann-Whitney-U-Test mit einem Konfidenzniveau von 95 % durchgeführt [33]. Für die Intensität, empfundene Sympathie, und Unterscheidungsgüte konnte die

Tabelle 2. Erkennung der einzelnen Emotionen [9].

Emotion	Ärger	Ekel	Freude	Angst	Trauer	Neutral
Echt	97 %	91 %	98 %	28 %	71 %	56 %
Generiert	74 %	95 %	87 %	47 %	63 %	67 %

Tabelle 3. Durchschnittliche Bewertungen in Intensität und Sympathie für die individuellen Emotionen [9].

	Emotion	Ärger	Ekel	Freude	Angst	Trauer	Neutral
Intensität	Echt	4,53	4,02	3,86	4,26	4,05	2,89
	Generiert	3,82	4,47	3,43	3,93	3,16	2,88
Sympathie	Echt	2,13	2,80	3,67	2,78	3,21	2,69
	Generiert	2,08	2,50	4,02	2,75	2,80	2,50

Nullhypothese widerlegt werden, was einen signifikanten Unterschied zwischen den Gruppen der echten und generierten Bildern bedeutet. Auf der anderen Seite kann für die Emotionserkennung die Nullhypothese nicht widerlegt werden, was bedeutet, dass es in diesem Punkt tatsächlich keinen Unterschied in der Wahrnehmung von echten und generierten Bildern gibt. Zusammengefasst lässt sich daher sagen, dass echte und generierte Bilder unterschiedlich in den Punkten Intensität, empfundene Sympathie und Unterscheidungsgüte wahrgenommen werden, nicht jedoch bei der Genauigkeit der Emotionserkennung.

3.2 Einzelne Emotionen im Vergleich

Tabelle 2 offenbart, dass die Emotionen gut erkannt werden konnten, mit Trefferquoten häufig über 70 %, vier sogar über 90 % und drei davon über 95 %. Allerdings gibt es zum Teil gravierende Unterschiede in der Trefferrate zwischen echten und generierten Bildern: bei Freude und dem neutralen Ausdruck beträgt der Unterschied 11 %, bei Angst 19 % und bei Ärger 23 %. Was zudem noch auffällt ist, dass die Emotion Angst nur sehr schlecht erkannt wurde mit Werten von 28 % und 47 %. Eine mögliche Erklärung dafür ist, dass sich die Emotionen Angst und Überraschung, die als Bild zwar nicht enthalten, aber dennoch eine Antwortmöglichkeit war, viele Charakteristika teilen.

Tabelle 3 zeigt die durchschnittliche Intensität der einzelnen Emotionen gruppiert nach echten und generierten Bildern. Auch hier reicht die Skala von „überhaupt nicht intensiv“ (1) bis „sehr intensiv“ (5). Die neutrale Emotion wurde als am wenigsten intensiv empfunden, was erwartbar war, da sie keine bestimmte Emotion darstellt. Auf der anderen Seite wurden der Ärger (echt) und der Ekel (generiert) am intensivsten wahrgenommen. Außerdem wurde die gleiche Emotion oft als unterschiedlich intensiv in den echten und generierten Bildern wahrgenommen.

Die Werte für die empfundene Sympathie sind in ebenfalls in Tabelle 3 dargestellt. Wie auch zuvor sind hier die Werte Mittelwerte der von den Teilnehmer vergebenen Punkte auf einer Skala von 1 bis 5, von „überhaupt nicht sympathisch“ bis „sehr sympathisch“. Wie zu erwarten schneidet die Freude, charakterisiert durch ein lächelndes Gesicht, am besten ab. Und das auch in beiden Fällen: echt und generiert. Auf der anderen Seite wurde der Ärger als am wenigsten sympathisch empfunden. Ein weitere nennenswerte Beobachtung ist, dass außer die Freude alle Emotionen einen Sympathie-Wert von unter drei haben, wobei drei in der Mitte der Skala liegt. Das impliziert, dass die gezeigten Personen im Allgemeinen als leicht unsympathisch

Tabelle 4. Cluster der Unterscheidungsmerkmale.

Cluster	Nennungen	Cluster	Nennungen
Fehler & Anomalien	47	Hautbild	39
Beleuchtung	32	Haare	32
Kleidung & Accessoires	31	„Zu perfektes“ Aussehen	29
„Zu weiches“ Aussehen	26	Gesichtsausdruck	17
Bildqualität	17	Hautfalten	17
Unnatürliches Aussehen	16	„Details“	16

Tabelle 5. Cluster für potenzielle Vorteile generierter Bilder.

Cluster	Nennungen
Marketing	37
Geld- und Zeitersparnis	27
Lizenzen, Rechte und Datenschutz	21
Bildmaterial für Präsentationen o.ä.	17
Inspiration und Ideengenerierung	16

empfundene wurden. Diese Hypothese wird auch dadurch gestützt, dass der neutrale Gesichtsausdruck einen durchschnittlichen Wert von 2,6 hatte, was möglicherweise bedeutet, dass selbst ein neutraler Ausdruck leicht unsympathisch ist.

3.3 Offene Fragen

Im letzten Abschnitt der Umfrage wurden die Teilnehmenden dazu befragt, welche Faktoren ihnen am meisten dabei geholfen haben, zwischen echten und generierten Bildern zu unterscheiden. Außerdem wurden sie gefragt, welchen potenziellen Nutzen und welche Gefahren sie in dieser Technologie sehen. Während die erste Frage darauf abzielte, mögliche Erklärungen für die Ergebnisse zu finden, dienten die letzten beiden Fragen dazu weitere Aspekte für eine Diskussion zu identifizieren. Die im Freitext-Format gegebenen Antworten wurden zu thematischen Clustern gruppiert. Zur Frage, was den Teilnehmenden geholfen hat, echt von generiert unterscheiden zu können, gab es 166 Antworten, die in Tabelle 4 zusammengefasst sind. An erster Stelle verriet offensichtliche Fehler und Anomalien, dass ein Bild generiert war. Ein Beispiel hierfür ist das generierte Bild für die Emotion Angst, in dem eine alte Dame abgebildet war, deren Halskette in der Mitte des Halses endete. Ein weiteres Unterscheidungsmerkmal war das Hautbild, welches beispielsweise „zu perfekt für ein bestimmtes Alter“ war oder eine unnatürliche Hautfarbe sowie eine unnatürliche Menge an Hautfalten aufwies. An dritter Stelle stehen sowohl die Beleuchtung, die von manchen als zu perfekt, unnatürlich oder unstimmig beschrieben wurde, als auch die Haare, die zwar meist nicht näher beschrieben, aber dennoch häufig erwähnt wurden.

Tabelle 5 fasst die Cluster von 150 Antworten nach den möglichen Verwendungen für generierte Emotionsdarstellungen zusammen und deutet an, dass der Hauptvorteil generierter Bilder in der Verwendung als Marketingmaterial liegen könnte. Darüber hinaus könnte die Verwendung generierter Bilder Geld und Zeit sparen, da sie möglicherweise teure Fotoshootings mit Models ersetzen oder Künstler und Designer mit schnellen Erstentwürfen und Designvarianten unterstützen könnte. An dritter Stelle waren viele Teilnehmenden der Ansicht, dass generierte Bilder im Allgemeinen lizenzfrei

Tabelle 6. Cluster für potenzielle Nachteile generierter Bilder.

Cluster	Nennungen
Fake News und Propaganda	75
Deepfakes und Verleumdung	39
Täuschung und Manipulation	29
Straftaten und Betrug	25
Identitätsdiebstahl und falsche Identitäten	25
Unterscheidungsprobleme zwischen Realität und Fälschung	20

wären, und somit von Unternehmen und Einzelpersonen ohne mögliche Lizenz- und Datenschutzverletzungen verwendet werden könnten. Diese Aussage ist allerdings umstritten, da die zugrundeliegenden KI-Modelle erst einmal auf Bilder trainiert werden müssen, die möglicherweise selbst nicht lizenz- und urheberrechtsfrei sind. Weiterhin könnten generierte Bilder verwendet werden, um Bildmaterial für Präsentation oder ähnliches zu erstellen. Als letztes könnten generierte Bilder als Inspiration oder erster Entwurf für neue Ideen dienen.

Tabelle 6 zeigt auf Basis von 169 Antworten, dass die Teilnehmer dieser Studie in generierten Bildern, die Verwendung dieser Bilder zur Verbreitung von Fake News und Propaganda als Hauptgefahr ansahen. Dieses Bedenken, etwas real erscheinen zu lassen, das nicht real ist, scheint sich über alle der in der Tabelle erwähnten Cluster zu ziehen. Am zweithäufigsten genannt wurde die mögliche Verwendung von generierten Inhalten für Deepfakes und Verleumdung. Das dritte Cluster befasst sich allgemein mit dem Thema Täuschungs- und Manipulationspotenzial. Obwohl „Täuschung“ ein spezifisches und häufig verwendetes Schlüsselwort war, weshalb dies ein zusätzlicher Cluster ist, könnte dieser Cluster auch als allgemeines Thema für die möglichen Gefahren generierter Bilder betrachtet werden. Dasselbe gilt für das vierte Cluster, da man sich vorstellen kann, dass der Missbrauch von generierten Bildern im Allgemeinen kriminelle Absichten hat. Als letztes erwähnenswert ist der Punkt, „Unterscheidungsprobleme zwischen Realität und Fälschung“, da er die Frage aufwirft, wie sehr die Menschen in Zukunft überhaupt in der Lage sein werden, irgendeiner Art von Medien zu vertrauen, da die Medienlandschaft mit Falschinformationen überflutet werden könnte, die möglicherweise sogar die Journalisten vertrauenswürdiger Medieninstitutionen täuschen könnten. Daher dürfen die möglichen Gefahren generierter Bilder, insbesondere die Erstellung und Verbreitung von Falschinformationen jeglicher Art, nicht unterschätzt werden.

3.4 Zusammenfassung und Diskussion

Im Kontext dieser Studie wurde Hypothese 1 — Emotionen in generierten Bildern werden gleich wahrgenommen wie Emotionen in echten Bildern — widerlegt, da signifikante Unterschiede zwischen den Bildgruppen gezeigt werden konnten. Erstens wurden die Emotionen in realen Bildern als intensiver wahrgenommen als die in generierten Bildern zu einer vermutlich moderaten Effektstärke. Zweitens, wurden die realen Personen als sympathischer wahrgenommen als die generierten, wobei hier die Effektstärke vermutlich nur gering ist. Und zuletzt waren die generierten Bilder leichter als solche zu erkennen als die echten Bilder, in diesem Fall auch mit einer vermuteten leichten Effektstärke. Abgesehen vom letzten Punkt, der bei der Diskussion der zweiten Hypothese behandelt wird, stehen diese Erkenntnisse nicht im offensichtlichen Widerspruch zu den Erkenntnissen in der Literatur. Während es

in den berichteten Intensitäten und Emotionserkennungsgenauigkeiten Unterschiede in dieser Studie im Vergleich zu den Ergebnissen von Ekman et al. [34] gab, scheint keiner dieser Unterschiede unerklärlich zu sein, da selbst bei den Ergebnissen in der Literatur zum Teil hohe Schwankungen zu finden sind. Darüber hinaus könnten einige dieser Unterschiede möglicherweise korreliert sein, da in einigen Fällen, in denen die Intensität der Emotionen als deutlich höher empfunden wurde als in der Literatur, die selbe Emotion auch besser erkannt wurde, während wenn die Emotion als weniger intensiv wahrgenommen wurde, die Emotion auch weniger gut erkannt wurde. Dies ist jedoch nicht bei allen Emotionen der Fall und ob es eine tatsächliche Korrelation gibt, bleibt offen für zukünftige Untersuchungen. Für einen detaillierteren Vergleich mit anderen Studien sei auf [9] verwiesen.

Hypothese 2 — generierte Bilder können von echten Bildern nicht unterschieden werden — konnte in dieser Studie nicht widerlegt werden. Auch wenn die Ergebnisse in dieser Studie suggerieren, dass diese Hypothese falsch ist, haben verschiedene andere Studien bereits bewiesen, dass diese Hypothese stimmt. In diesen Studien konnten die Teilnehmer echte und generierte Bilder sogar nur mit einer Genauigkeit von weniger als 50 % korrekt unterscheiden [35], [36], [37].

Daher ist die finale Schlussfolgerung, dass Emotionen in künstlich generierten Bildern zwar immer noch leicht bis moderat unterschiedlich wahrgenommen werden, es aber nicht mehr möglich ist echte und generierte Bilder zu unterscheiden. Dies kann im Marketing von Vorteil sein, wo generierte Bilder in Werbung und Branding eingesetzt werden könnten. Weiterhin könnten Vorteile dieser Technologie Geld- und Zeiteinsparung sein, sowie das schnelle generieren von Material für Präsentationen oder ähnliches. Auf der anderen Seite birgt diese Technologie aber auch große Gefahren, die nicht außer Acht gelassen werden dürfen. Generierte Bilder können beispielsweise auch für die Erstellung und Verbreitung von Falschinformationen missbraucht werden.

Diese Studie hat außerdem aufgezeigt, dass das Erstellen von realistischen Porträts mit generativer KI relativ einfach ist. Was sich aber als schwierig herausstellte, war es, durch den KI-Bildgenerator ein exaktes und spezifisches Wunschergebnis zu erstellen. Dabei haben manche Prompts wie „A man displaying the expression of disgust“ also „Ein Mann, der die Emotion Ekel ausdrückt“ schlichtweg nicht funktioniert und auch das gezielte Ansprechen spezifischer Gesichtsregionen in Prompts, beispielsweise „inner corners of the eyebrows raised and drawn together“ also „gehobene und zusammengezogenen innere Ecken der Augenbrauen“ funktionierte nicht. Allerdings werden sich vermutlich auch diese Probleme durch die rasante Entwicklung der künstlichen Intelligenz bald lösen.

Beiträge der Autoren

Emil Ojo: Methodology, Data curation, Investigation, Writing – original draft

Mathias Walther: Conceptualization, Supervision, Writing – review & editing

Interessenkonflikte

Die Autoren erklären, dass keine Interessenkonflikte bestehen.

References

- [1] lilmiquela [@lilmiquela]. "Instagram profile". eprint: [Instagram](#), Accessed: Jan. 16, 2025. [Online]. Available: <https://www.instagram.com/lilmiquela/>.
- [2] imma [@imma.gram]. "Instagram profile". eprint: [Instagram](#), Accessed: Jan. 16, 2025. [Online]. Available: <https://www.instagram.com/imma.gram/>.
- [3] lil Miquela]. "The first-ever bmw ix2 × lil miquela. youtube video". eprint: [youtube](#), Accessed: Jan. 16, 2025. [Online]. Available: <https://www.youtube.com/watch?v=MXSaZQ2aAAc%7D>.
- [4] bmw and lilmiquela. "Instagram post". eprint: [Instagram](#), Accessed: Jan. 16, 2025. [Online]. Available: <https://www.instagram.com/p/CyP2zNuoOEV/>.
- [5] Prada. "Prada candy - rethink reality - sensations. youtube video". eprint: [Youtube](#), Accessed: Jan. 16, 2025. [Online]. Available: <https://www.youtube.com/watch?v=E0skE>.
- [6] lilmiquela. "I am here. that is my truth. #mytruth #mycalvins". eprint: [Instagram](#), Accessed: Jan. 16, 2025. [Online]. Available: <https://www.instagram.com/p/Bxh4AdrnTel/>.
- [7] E. Brynjolfsson, D. Li, and L. Raymond, *Generative AI at Work*. 2023. DOI: [10.3386/w31161](#).
- [8] E. Rykova and M. Walther, "Aphadigital – digital speech therapy solution for aphasia patients with automatic feedback provided by a virtual assistant", in *57th Hawaii International Conference on System Sciences, HICSS 2024, Honolulu, Hawaii, USA, 2024*, pp. 3385–3394.
- [9] E. Ojo, "Perception of basic emotions in facial expressions created by generative ai: Tools and empirical tests", Bachelor's thesis, Technical University of Applied Sciences Wildau, 2024.
- [10] A. Vinciarelli, M. Pantic, H. Bourlard, and A. Pentland, "Social signal processing: State-of-the-art and future perspectives of an emerging domain", in *Proceedings of the 16th ACM international conference on Multimedia*, ACM, 2008, pp. 1061–1070.
- [11] J. A. Hall, T. G. Horgan, and N. A. Murphy, "Nonverbal communication", *Annual Review of Psychology*, vol. 70, 1 Jan. 2019. DOI: [10.1146/annurev-psych-010418-103145](#). [Online]. Available: <http://dx.doi.org/10.1146/annurev-psych-010418-103145>.
- [12] A. Mehrabian and S. R. Ferris, "Inference of attitudes from nonverbal communication in two channels.", *Journal of Consulting Psychology*, vol. 31, no. 3, pp. 248–252, 1967, ISSN: 0095-8891. DOI: [10.1037/h0024648](#).
- [13] D. Lapakko, "Three cheers for language: A closer examination of a widely cited study of nonverbal communication", *Communication Education*, vol. 46, 1 Jan. 1997. DOI: [10.1080/03634529709379073](#).
- [14] P. Ekman, *Emotions Revealed, Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York: Times Books, 2003.
- [15] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain", *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, 2009.
- [16] U. Hess and P. Thibault, "Darwin and emotion expression.", *American Psychologist*, vol. 64, 2 Feb. 2009. DOI: [10.1037/a0013386](#). [Online]. Available: <http://dx.doi.org/10.1037/a0013386>.
- [17] P. R. Kleinginna and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition", *Motivation and emotion*, vol. 5, no. 4, pp. 345–379, 1981.
- [18] C. Darwin, *On the Expression of the Emotions in Man and Animals*. London: John Murray, 1872. [Online]. Available: <http://books.google.de/books?id=NWAJAAAAQAAJ>.
- [19] A. Paeschke, *Prosodische Analyse emotionaler Sprechweise* (Mündliche Kommunikation). Berlin: Logos, 2003, ISBN: 3-8325-0146-0, 978-3-8325-0146-4.
- [20] Paul Ekman Group, *Universal emotions | what are emotions? | paul ekman group*, Online. <https://www.paulekman.com/universal-emotions/>, Accessed: Feb. 13, 2025, no date.

- [21] E. Ojo and M. Walther, "Empirical evaluation of basic emotion perception in ai-generated facial expressions", in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, im Druck, Götheburg, Schweden.
- [22] M. Testerman, *Face image databases*, Online. <https://libguides.princeton.edu/facedatabases>, Accessed: Jan. 30, 2024, 2022.
- [23] J. van der Schalk, S. T. Hawk, A. H. Fischer, and B. Doosje, "Moving faces, looking places: Validation of the amsterdam dynamic facial expression set (adfs).", *Emotion*, vol. 11, no. 4, pp. 907–920, 2011, ISSN: 1528-3542. DOI: [10.1037/a0023853](https://doi.org/10.1037/a0023853).
- [24] T. Kanade, Y. Li, and J. Cohn, "Comprehensive database for facial expression analysis", in *Proceedings of International Conference on Face and Gesture Recognition*, 2000, pp. 46–53.
- [25] K. S. Scherf and M. S. Benda, "The complex emotion expression database: A validated stimulus set of trained actors", 2019. DOI: [10.17910/B7.874](https://doi.org/10.17910/B7.874). [Online]. Available: <http://dataverse.org/volume/874>.
- [26] N. C. Ebner, M. Riediger, and U. Lindenberger, "Faces – a database of facial expressions in young, middle-aged, and older women and men: Development and validation", *Behavior Research Methods*, vol. 42, 1 Feb. 2010. DOI: [10.3758/brm.42.1.351](https://doi.org/10.3758/brm.42.1.351).
- [27] F. Wallhoff, B. Schuller, M. Hawellek, and G. Rigoll, "Efficient recognition of authentic dynamic facial expressions on the feedtun database", *2006 IEEE International Conference on Multimedia and Expo*, Jul. 2006. DOI: [10.1109/icme.2006.262433](https://doi.org/10.1109/icme.2006.262433).
- [28] D. Lundqvist, A. Flykt, and A. Öhman, "Karolinska directed emotional faces", 1998. [Online]. Available: <https://api.semanticscholar.org/CorpusID:150212888>.
- [29] A. Borji, "Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2", Oct. 2022. DOI: [10.48550/arXiv.2210.00586](https://doi.org/10.48550/arXiv.2210.00586). eprint: [arXiv:10.48550/arXiv.2210.00586](https://arxiv.org/abs/2210.00586). [Online]. Available: <http://arxiv.org/abs/2210.00586v2>.
- [30] civitai.com, *Faces by thegipper | lora-collection*, Online. <https://civitai.com/tag/faces>, Accessed: Jul. 25, 2024, 2024.
- [31] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models", Feb. 2023. DOI: [10.48550/arXiv.2302.05543](https://doi.org/10.48550/arXiv.2302.05543). eprint: [arXiv:10.48550/arXiv.2302.05543](https://arxiv.org/abs/2302.05543). [Online]. Available: <http://arxiv.org/abs/2302.05543v3>.
- [32] K. Mehreen, *Using lora in stable diffusion*, Online. <https://machinelearningmastery.com/using-lora-in-stable-diffusion/>, Accessed: Jul. 25, 2024, 2024.
- [33] N. Nachar, "The mann-whitney u: A test for assessing whether two independent samples come from the same distribution", *Tutorials in Quantitative Methods for Psychology*, vol. 4, no. 1, pp. 13–20, 2008. DOI: [10.20982/tqmp.04.1.p013](https://doi.org/10.20982/tqmp.04.1.p013). [Online]. Available: <http://www.tqmp.org/RegularArticles/vol04-1/p013/p013.pdf>.
- [34] P. Ekman et al., "Universals and cultural differences in the judgments of facial expressions of emotion.", *Journal of Personality and Social Psychology*, vol. 53, 4 1987. DOI: [10.1037/0022-3514.53.4.712](https://doi.org/10.1037/0022-3514.53.4.712). [Online]. Available: <http://dx.doi.org/10.1037/0022-3514.53.4.712>.
- [35] J. Frank et al., "A representative study on human detection of artificially generated media across countries", Dec. 2023. DOI: [10.48550/arXiv.2312.05976](https://doi.org/10.48550/arXiv.2312.05976). eprint: [arXiv:10.48550/arXiv.2312.05976](https://arxiv.org/abs/2312.05976). [Online]. Available: <http://arxiv.org/abs/2312.05976v1>.
- [36] S. J. Nightingale and H. Farid, "Ai-synthesized faces are indistinguishable from real faces and more trustworthy", *Proceedings of the National Academy of Sciences*, vol. 119, 8 Feb. 2022. DOI: [10.1073/pnas.2120481119](https://doi.org/10.1073/pnas.2120481119).
- [37] F. Lago, C. Pasquini, R. Bohme, H. Dumont, V. Goffaux, and G. Boato, "More real than real: A study on human visual perception of synthetic faces [applications corner]", *IEEE Signal Processing Magazine*, vol. 39, 1 Jan. 2022. DOI: [10.1109/msp.2021.3120982](https://doi.org/10.1109/msp.2021.3120982).