Über den Umgang mit KI-Bildgeneratoren: Fotografien kennzeichnen – Scraping verhindern. Ein Leitfaden für Fotograf*innen

Viele von uns haben in den letzten Monaten über die möglichen Folgen des Einsatzes von Künstlicher Intelligenz in der Fotografie und Bilderstellung nachgedacht und diskutiert. Hier haben wir Handlungsoptionen für Fotograf*innen gesammelt, um die Authentizität von Fotografien nachzuweisen und das Scraping unserer Bilder zu unterbinden.

Text – Marco Urban / FREELENS

Der Wissenschaftsjournalist und Autor Ranga Yogeshwar wird in der **Augsburger Allgemeinen vom 17. Mai 2023** wie folgt zitiert:

> *»[…] wir erleben im Moment den größten Diebstahl in der Menschheitsgeschichte. Die reichsten Unternehmen der Welt wie Microsoft, Apple, Google, Meta oder Amazon bemächtigen sich der Summe des menschlichen Wissens. Also aller Texte, Kunstwerke, Fotografien und so weiter, die in digital verwertbarer Form existieren, um dieses Weltwissen dann in eigentumsrechtlich geschützten Produkten einzumauern. Es gibt dabei keine klare Offenlegung, mit welchen Lerndaten sie die KI trainieren. […] Das Urheberrecht wird missachtet – und zwar bewusst. Inzwischen kann per KI eine Massenproduktion von Plagiaten stattfinden, wobei ganze Berufsstände vor ihrem existenziellen Aus stehen.«*

Was können und müssen wir angesichts dieser rasanten Veränderung eigentlich tun? Was sind die Handlungsoptionen für Fotograf*innen, abgesehen von der eigenen Neugier, Bilder mit Künstlicher Intelligenz zu generieren? Wir haben die relevanten Punkte zusammengefasst, wobei sich sowohl die Fragestellungen als auch die Antworten täglich ändern können.

1. Kennzeichnen von Fotografien/Bildern mit den Kontaktdaten der Urheber*innen
Diese Maßnahme sollte in jedem Fall selbstverständlich sein. Sie ist absolut notwendig, um die Urheberschaft eines Werkes zu dokumentieren. Der IPTC-Standard ist Grundlage dafür in welcher Form diese Daten in den Meta- oder IPTC-Daten hinterlegt werden. Hier kann im Copyright-Status auch festgelegt werden, ob das Werk urheberrechtlich geschützt ist.

2. Kennzeichnen von Fotografien/Bildern entsprechend ihrer digitalen Herkunft
Werke sollten wahrheitsgemäß entsprechend des **IPTC-Standards »Digital Source Type«** oder »Quelltyp« gekennzeichnet werden. Er gibt Auskunft über die Art des digitalen Entstehungsprozesses der Bilddatei.

Aktuell gibt ~~es~~ sind folgende Typen vorgegeben:
– Digitale Originalaufnahme eines realen Motivs (Original digital capture sampled from real life)
– Digitalisiert von einem Filmnegativ (Digitised from a negative on film)
– Digitalisiert von einem Filmpositiv (Digitised from a positive on film)
– Digitalisiert von einem Druck oder nicht-transparenten Medium (Digitised from a print on non-transparent medium)
– Originalmedien mit geringfügigen menschlichen Bearbeitungen (Original media with minor human edits)
– Komposition aus aufgenommenen Elementen (Composite of captured elements)
– Algorithmisch angereicherte Medien (Algorithmically-enhanced media)
– Datengesteuerte Medien (Data-driven media)
– Digitale Kunst (Digital art)
– Virtuelle Aufnahme (Virtual recording)
– Kompositum mit synthetischen Elementen (Composite including synthetic elements)
– Geschulte algorithmische Medien (Trained algorithmic media)
– Komposition mit trainierten algorithmischen Medien (Composite with trained algorithmic media)
– Rein algorithmische Medien (Pure algorithmic media)

Bei Photoshop und Lightroom stehen aktuell nur die ersten vier Typen zur Verfügung.

3. Kennzeichnen von Fotografien/Bildern um die Verwendung als KI-Trainingsdaten zu verbieten.
IPTC arbeitet an einem Standard, in den IPTC-Daten ein Feld zu definierten, das kennzeichnet, ob und in welchem Umfang die Verwendung als KI-Trainingsmaterial erlaubt ist.
Dies wäre eine gute Maßnahme, um der gesetzlichen Forderung nach Widerspruch in maschinenlesbarer Form für jede einzelne Datei nachzukommen.

4. Kennzeichnen von Webseiten, um Scraping/Einlesen von Werken durch Trainings-Datenbanken zu verbieten.
Eigenen Webseiten sollten gem. UrhG mit einem maschinenlesbaren Hinweis versehen, um dem Scraping (engl. für kratzen oder schaben) zu widersprechen. Aktuell gehen wir davon aus, dass dies mit der robots.txt oder ähnlichen txt-Dateien geschehen kann. Robots.txt ist eine Textdatei, die in den Backend-Code einer Website eingefügt werden kann, um Web-Crawlern mitzuteilen, was sie durchsuchen dürfen und was nicht. Wer seine Website nicht von Google durchsuchen lassen möchte, schreibt den entsprechenden Befehl in die robots.txt Datei der Website. **Eine Anleitung findet ihr hier**.
Die New York Times, CNN und Australiens ABC **blockieren auf diese Weise** den Zugriff auf ihre Inhalte für OpenAIs GPTBot-Webcrawler.
Wie der Befehl lautet, der Scraping verhindert, ist allerdings bislang noch nicht festgelegt.
Die Voraussetzung nach UrhG ist allerdings lediglich, dass der Hinweis maschinenlesbar sein muss.
Technisch gesehen können die Befehle in der robots.txt allerdings ignoriert werden. Sie sind lediglich ein Hinweis an die Web-Crawler. Will man später gegen die Verwendung seiner Bilder zum KI-Training juristisch vorgehen, wird es aber wichtig sein, dass man diesen Hinweis korrekt platziert hat.
Eine weitere Möglichkeit hat das Projekt **https://spawning.ai** entwickelt: **Hier** kann man eine ai.txt Datei generieren lassen, um die Verwendung von Website-Inhalten für das Training von KI-Modellen zuzulassen oder zu verhindern.
Auch hier ist derzeit zweifelhaft, welche Crawler diese Datei tatsächlich auswerten. Einen Schutz kann die Datei nur für die Fotografien auf eben dieser Website haben, nicht aber, wenn die gleichen Fotos an anderer Stelle veröffentlicht werden.
Es kann aber nicht schaden, die txt-Dateien zu verwenden. Es ist davon auszugehen, dass man damit der gesetzlichen Forderung nach Widerspruch in maschinenlesbarer Form nachkommt.

5. Have I been Trained?
Die Website **https://haveibeentrained.com** wird ebenfalls von **https://spawning.ai** betrieben und bietet die Möglichkeit, nach Fotografien zu suchen, die von KI-Generatoren verwendet werden. Im Moment durchsucht das Programm die **Laion-5B** und **Laion-400M** Datenbanken. Diese werden von den meisten KI-Bildgeneratoren wie **Stable Diffusion** oder **Imagen** eingesetzt. Es gibt aber auch weitere, nicht öffentliche Trainingsdatensammlungen.
Es gibt hier auch die Möglichkeit, Bilder für das Training zu sperren.
Allerdings ist davon auszugehen, dass dies nicht nachträglich geschehen kann. Ist das KI-Modell erst einmal erstellt, können einzelne Trainingsbilder nicht mehr ausgenommen werden. Vermutlich halten sich auch nicht alle (oder sogar nur wenige) der mehreren tausend KI-Generatoren an diese Opt-Out-Anweisung. Bilder in möglicherweise größerer Zahl bei »**Have I been Traind**« einzeln zu sperren, dürfte also vergeblich sein.
Die Laion-Datenbanken sind schon einige Jahre alt und es werden keine neuen Links zu Bildern hinzugefügt.

6. Untersagen des Verwendens von Fotografien/Bilder als KI-Trainingsmaterial durch Kund*innen.
Vermutlich wird es für Kund*innen nicht möglich sein, sicherzustellen, dass Fotografien, die Fotograf*innen in ihrem Auftrag erstellt haben, nicht als KI-Trainingsmaterial verwendet werden. Selbst Fotografien von Druckerzeugnissen werden genutzt.
Fotograf*innen sollten jedoch durch AGB oder Verträge verbieten, dass die Kund*innen selbst deren Fotografien als Trainingsmaterial verwenden, zum Beispiel, um mit Hilfe von Porträtfotos KI-generierte Porträtbilder der gleichen Personen zu erstellen.
Solche Porträts generiert beispielsweise **Remini** oder **Generated Photos**.
Andererseits mag es aber fraglich sein, ob wir Kund*innen gänzlich verbieten sollten, Fotografien mit KI-Tools zu bearbeiten. Die App **https://piktid.com** anonymisiert beispielsweise Gesichter, indem sie diese durch KI-generierte ersetzt. Vermutlich lassen sich so Probleme mit Persönlichkeitsrechten umgehen, was bei Fotoaufträgen im öffentlichen Raum hilfreich sein mag.

7. Posten von Bildern auf Social Media Plattformen vermeiden
Posten Fotograf*innen eigene Fotografien/Bilder auf Social-Media-Plattformen, werden umfangreiche Nutzungsrechte an die Plattformen übertragen. Die Formulierungen über den genauen Umfang in den Online-Geschäftsbedingungen sind oft unklar.
In der Regel wird den Plattformen das Recht eingeräumt, sie auf jede erdenkliche Weise zu nutzen. Dies wird sicherlich auch die Integration in KI-Daten-Trainingseinheiten beinhalten.
X (Twitter) hat wohl gerade seine **Nutzerbedingungen dahingehend angepasst**, als dass Posts für KI-Training benutzt werden dürfen.

8. Bildagenturen erstellen KI-Modelle
Getty Images startete Ende September **seine eigene KI-Anwendung**, mit der Kund*innen selbst KI-Bilder generieren können. Der Trainings-Datensatz besteht aus Bildern aus dem Getty-Archiv. Versprochen wird eine Kompensation der Urheber*innen, sichere Verwendung bei unbegrenzter Haftungsfreistellung.
Wer also Fotos über Getty Images distribuiert, sollte klären, ob diese auch für das KI-Training verwendet werden und wie das honoriert wird.
Dass Adobe Stock die Fotos im Archivbestand für KI-Training nutzt, ist bekannt. Mit diesem Modell wird auch die Funktion »Generative Füllung« möglich gemacht.
Die Agentur Laif hat sich **klar gegen KI-generierte Bilder** ausgesprochen.
Von den großen, internationalen Bildagenturen ist jedoch ziemlich sicher zu erwarten, dass sie ihre Archive gegen entsprechendes Entgelt anderen für das KI-Training zur Verfügung stellen werden oder dies wie Getty gleich selbst machen.

9. Verwenden von Wasserzeichen
Getty Images verwendet für Preview-Ansichten grundsätzlich ein deutlich sichtbares Wasserzeichen mit dem Getty Images Logo. Offenbar wurden die Bilder von Getty Images in großer Menge eingelesen und als KI-Trainigsmaterial verwendet. Die KI hat den offenbar sehr hohen Anteil an Getty-Fotos derart interpretiert, dass Fussballbilder mit hoher Wahrscheinlichkeit ein Logo von Getty Images beinhalten. Getty Images konnte zum einen somit die Verwendung ihrer Fotografien leicht nachweisen, zum anderen waren die Bilder dadurch auch unbrauchbar. Das funktioniert natürlich nur bei einer sehr großen Anzahl Bilder.
Grundsätzlich ist es aber von Vorteil, wenn schon auf dem Foto erkennbar ist, wer der Urheber ist und dass es urheberrechtlich geschützt ist.

10. Kennzeichnung von KI-generierten Bildern durch KI-Generatoren
Es gibt **erste Zusagen** der großen Tech-Firmen, mit ihren Generatoren erstellte KI-Bilder zu kennzeichnen.
Wird diese Kennzeichnung von Medien und vor allen Dingen Sozialen Medien übernommen, wäre das ein großer Schritt, um Betrachter*innen zu ermöglichen, synthetische Bilder von authentischen Fotografien zu unterscheiden.

11. Kennzeichnung von Fotografien und Bildern in redaktionellen Medien
Fotograf*innen sollten sprachlich klar zwischen Fotos, die durch die Abbildung von realen Szenen durch Licht entstanden sind, und KI-generierten Bildern oder Illustrationen trennen. Unklare oder verschleiernde Bezeichnungen wie »KI-Fotos« sollten nicht verwendet werden.
Aufbauend auf die **Vereinbarung der Kennzeichnung von manipulierten Fotografien mit [M]** fordern wir von redaktionellen Medien auch die **Kennzeichnung** von authentischen Fotografien mit [A] und die Kennzeichnung von rein (oder in relevanten Teilen) KI-generierten Bildern mit [G]. Die Kennzeichnung soll mit dem Urhebernachweis am Bild erfolgen.
Darüber hinaus steht zur Debatte, im Zuge dessen auch die Kennzeichnung von Symbol- oder PR-Bildern zu fordern.

Die vergleichbare **Initiative »Writing with Light«** fordert, in der Credit-Line ein Symbol zu verwenden, das die Authentizität der Fotografie bestätigt. Weiter heißt es dort:
»Als Aufzeichnungen des Sichtbaren müssen journalistische Fotos das, was der Fotograf gesehen hat, wahrheitsgetreu und genau wiedergeben. Weder Veränderungen an einem Foto, welche die Öffentlichkeit in die Irre führen, noch die Inszenierung von Ereignissen, während sie als spontan dargestellt werden, sind im Journalismus akzeptabel. Ebenso wenig sollte man ein von einer künstlichen Intelligenz erstelltes fotorealistisches synthetisches Bild veröffentlichen und so tun, als sei es ein echtes Foto. Jede Abweichung von diesen Grundprinzipien muss in einer Bildunterschrift oder einem Bildnachweis erläutert und gegebenenfalls durch ein Symbol gekennzeichnet werden, das auf die Art der Manipulation hinweist und bei der Veröffentlichung auf oder unter dem Bild angebracht wird.«

12. CAI – Content Authenticity Initiative
Die **Content Authenticity Initiative** arbeitet daran, allen Arten von digitalen Inhalten, angefangen bei Fotos und Videos, durch Lösungen für die Herkunftsbestimmung und -zuordnung eine Ebene des überprüfbaren Vertrauens zu verleihen und damit Fehlinformationen, Fakes, zu bekämpfen. Sie besteht aus Medien- und Technologieunternehmen, Nichtregierungsorganisationen, Wissenschaftler*innen und anderen, die sich für die Einführung eines offenen Industriestandards für die Sicherung von Authentizität und Klärung der Herkunft von Inhalten einsetzen. FREELENS ist, wie große Verlage, Nachrichten- und Bildagenturen, Kamera- und Softwarehersteller Mitglied dieser Initiative.
CAI will ein sicheres End-to-End-System für die Herkunft digitaler Inhalte durch Open-Source-Entwicklung, branchenübergreifende Zusammenarbeit und Interoperabilität von Tools etablieren.

Dieses End-to-End-System soll den Weg von der Aufnahme einer Fotografie über die Bearbeitung bis zur Veröffentlichung oder Weitergabe in sozialen Medien dokumentieren und nachvollziehbar machen. Kurz erklärt passiert dabei Folgendes:

1. Aufnahme
Mit kryptografischen Asset-Hashing werden nachprüfbare, fälschungssichere Signaturen erstellt, die belegen, dass das Bild und die Metadaten nicht versehentlich oder heimlich verändert wurden. Bei der Erstellung wählen Fotograf*innen aus, welche Informationen an den zu erstellenden Inhalt angehängt werden sollen. Während des gesamten Prozesses kann der/die Ersteller*in von Inhalten wählen, ob er oder sie die Namensnennung beibehalten oder anonym bleiben möchte. Der Schutz der Privatsphäre und die Sicherheit von Fotograf*innen und anderen Urheber*innenn sollen an erster Stelle stehen.

2. Bearbeitung
Mit Werkzeugen wie Photoshop werden sichere Metadaten der Aufnahmen erhalten und mit Verlaufsdaten aller Änderungen am Inhalt ergänzt.

3. Veröffentlichung und Weitergabe
Durch Partnerschaften von CAI mit Nachrichtenorganisationen werden sichere Erfassungsinformationen und alle relevanten Inhaltsänderungen während des Veröffentlichungsprozesses durch die Integration in dem CMS der Verlage gesichert. Wenn Inhalte in sozialen Netzwerken geteilt werden, bleiben die CAI-Metadaten erhalten.

4. Überprüfung durch Nutzer*innen
Als Nutzer*n digitaler Inhalte kann jede*r über die **Verify-Website** historische Informationen über die Inhalte mit CAI-Metadaten einsehen. Während die CAI sich auf Systeme zur Bereitstellung von Kontext und Historie für digitale Medien konzentriert, übernimmt die **Coalition for Content Provenance and Authenticity (C2PA)** die Ausarbeitung der technischen Standards und Spezifikationen als Grundlage für eine diese universelle Inhaltsprovenienz. CAI bzw. deren Mitglieder werden also diese Standards in Kameras und Software integrieren, die Fotograf*innen nutzen.

UPDATE 26.10.2023:

13. Kudurru – Scraper identifizieren und abwehren
Neue Maßnahmen gegen Scraping (engl. für kratzen oder schaben) setzen nicht mehr auf das Entgegenkommen von Akteuren, die Daten für das KI-Training sammeln, wenn entsprechend gekennzeichnete Fotografien oder Webseiten vom Scraping durch Opt-Out ausgenommen werden sollen. Kudurru überwacht populäre KI-Datensätze auf Scraping-Verhalten und koordiniert sich im Netzwerk, um Scraper schnell zu identifizieren. Wenn ein Scraper identifiziert wird, wird seine Identität an alle geschützten Kudurru-Websites übermittelt. Alle Kudurru-Websites blockieren dann gemeinsam den Scraper für das Herunterladen von Inhalten von ihrem jeweiligen Host. Wenn der Scraper seine Arbeit beendet hat, informiert Kudurru das Netz, und der Datenverkehr kann wie gewohnt fortgesetzt werden.
Dabei werden die Scraper nicht nur zurückgewiesen, sondern es können auch alternative Bilder anstelle der von den Scrapern angeforderten Bilder ausgegeben werden. Diese Irreführung kann dazu führen, dass die Modelle falsche Assoziationen zu Ihrem Stil entwickeln und die Ausgabe beeinflussen, die sie produzieren.
Kudurru blockiert keine Suchmaschinen-Crawler oder Bots, wie z. B. Google Bot. Diese Bots sind von Google genau definiert (so dass man sie absichtlich nicht blockieren kann) und werden von Kudurru ignoriert. Kudurru hat keinen Einfluss auf das SEO-Ranking oder die Auffindbarkeit der Website.

Das Kudurru-Netzwerk verfügt schon jetzt über mehr als tausend aktive Websites, auf denen Millionen von Medien aus beliebten KI-Datensätzen gehostet werden.
Es gibt ein erstes Plug-In für WordPress-Websites und es sollen weiter für andere Plattformen entwickelt werden. Nach der Beta-Phase soll der Quellcode als Open Source Software zur Verfügung gestellt werden.

Die Macher von Kudurru sind Spawing, über deren Arbeit Ihr bereits in Punkt 4 und 5 etwas lesen konntet. Dabei handelte es sich aber um passive Abwehrmaßnahmen.

Weitere Informationen unter https://kudurru.ai

14. Nightshade – vergiftete Köder für die Scraper
Noch einen Schritt weiter geht das Tool Nightshade. Damit „können Künstler für Menschen unsichtbare Änderungen an den Pixeln ihrer Kunstwerke vornehmen, bevor sie diese online hochladen. Gelangen diese dann als Trainingsdaten in ein bildgenerierendes Sprachmodell, kann dies dazu führen, dass das

resultierende Modell auf chaotische und unvorhersehbare Weise gestört wird.“ (heise online - https://www.heise.de/news/Gift-fuer-Trainingsdaten-Neues-Tool-soll-Bilder-vor-KI-Bildgeneratoren-schuetzen-9343354.html)

Weiter heißt es „Nightshade (…) zielt darauf, die Arbeiten von Künstlern zu schützen, wenn KI-Firmen diese Werke ohne Erlaubnis der Urheber zum Trainieren ihrer Modelle verwenden. Die Verwendung des Tools, um diese Trainingsdaten zu "vergiften", könnte zukünftigen Iterationen von bildgenerierenden KI-Modellen wie DALL-E, Midjourney und Stable Diffusion schaden, indem es einige ihrer Ergebnisse unbrauchbar macht: aus Hunden werden Katzen, aus Autos werden Kühe und so weiter.
Nightshade wurde von Professor Ben Y. Zhao vom Department of Computer Science an der University of Chicago und seinem Team entwickelt und es künftig in Glace implementiert wird. Glace wiederum ist ein Tool, mit dem Kunstwerke vom Urheber „maskiert“ werden können, damit KI-Unternehmen dieses Bild nicht zu Trainingszwecken nutzen können. Die Verbindung zwischen Glace und Nightshade soll dann besonders effektiv sein, um KI-Trainingsmodelle zu „vergiften“.
Je mehr „vergiftete“ Bilder zu Trainingszwecken herangezogen werden, desto dramatischer die Auswirkungen. In einem ausführlichen Forschungs-Paper legen die Wissenschaftler dar, dass bereits wenige Bilder ausreichen, um eine KI falsch zu trainieren. So infiziert, erkennt die KI einen Hut als Kuchen, ein Auto als Kuh oder einen Cartoon als impressionistisches Werk. Getestet wurden die Auswirkungen von Glace und Nightshade an den neuesten Stable Diffusion Modellen. Hier haben bereits 300 manipulierte Bilder gereicht, um die KI-Trainings zu verwirren. (Connect Living - https://www.connect-living.de/news/nightshade-ki-tool-bilder-schutz-kuenstler-urheberrecht-3206362.html)

Mehr dazu in diesem Artikel der MIT Technology Review: https://www.technologyreview.com/2023/10/23/1082189/data-poisoning-artists-fight-generative-ai/

Diese Aufzählung stellt den Stand der Dinge im Oktober 2023 dar. Es gibt fast täglich neue Entwicklungen zum Thema. Wir werden euch weiterhin auf dem Laufenden halten.

»Photograph the world as it is. Nothing’s more interesting than reality.«
Mary Ellen Mark, Magnum-Fotografin

Bildunterschrift:
Mit den Prompts »Maschine bewirft kleines Mädchen mit tausenden Fotografien Fotografien fliegen umher Dystopie« generiertes Bild [G]. Erstellt mit Adobe durch Marco Urban.