

Kunst kommt von KI

Was Bildgeneratoren können und wie man sie einsetzt



Was KI-Bildgeneratoren können	Seite 64
Dall-E & Co. mit den richtigen Prompts bedienen	Seite 68
Sechs Bildgenerator-Dienste im Vergleich	Seite 72
Wie KI Bilder malt	Seite 80

Mehr als eine kurze Textvorgabe benötigen Bilder-KIs wie Dall-E, Midjourney oder Stable Diffusion nicht, um faszinierende Werke zu erschaffen. Die künstlichen Künstler werden so manchen grafischen Produktionsprozess umkrempeln. Die Technik wirft aber auch rechtliche und moralische Fragen auf.

Von Jo Bager und Pina Merkert

Unter der Adresse <https://beta.dreamstudio.ai> ist ein neues Künstlerkollektiv erreichbar. Wann immer Ihnen der Sinn nach einem Gemälde, einer Grafik oder einem Foto steht: Rufen Sie die Website auf und beschreiben Sie dort das Motiv sowie den Stil mit einem kurzen Text, zum Beispiel „a fox astronaut with the cosmos reflecting on the glass of his helmet dreaming of the stars“. Sofort setzt sich ein Künstler an seine Staffelei und malt das Bild für Sie. Innerhalb von ein paar Sekunden sehen Sie das Ergebnis.

Natürlich handelt es sich nicht um ein Kollektiv aus menschlichen Kreativen, sondern um einen virtuellen Künstler namens Stable Diffusion. Die KI hat gelernt, wie Rembrandt, Van Gogh, Dalí und Dutzende andere alte Meister zu malen und beherrscht noch Hunderte weiterer Stile. Sie entwirft Illustrationen, die von denen erfahrener Grafiker nicht zu unterscheiden sind, Comic-artige Szenen sowie „Fotos“ von Orten und Ereignissen, die es nicht gibt und auch nie gegeben hat.

Stable Diffusion ist nur ein Beispiel für eine ganze Klasse an KI-Modellen, die sich auf Text-zu-Bild-Konversion verstehen. Bei etlichen Onlinediensten kommen solche Modelle bereits zum Einsatz. Dieser und die folgenden Artikel stellen Ihnen das spannende und zugleich sehr unterhaltsame Thema der KI-Bildgeneratoren im Detail vor.

In diesem Artikel geht es darum, was die KI-Künstler können, welche rechtlichen Auswirkungen die KI-Kunst hat und wie sie die Arbeit von Kreativen beeinflusst. Ab Seite 72 vergleichen wir sechs

Online-Bildgeneratoren, die wir im Akkord Bilder malen lassen haben.

Bei Stable Diffusion und Konsorten handelt es sich streng genommen jeweils nicht um eine einzelne KI. Stattdessen haben die Entwickler es geschafft, mehrere spezialisierte neuronale Netze sinnvoll zusammenspielen zu lassen. Ab Seite 80 erklären wir den Prozess, in dem die Systeme Bilder erzeugen, im Detail.

Datenquelle: Internet

Wie alle nennenswerten KI-Errungenschaften der letzten Jahre funktionieren auch die Bildgeneratoren nur deswegen so gut, weil sie mit immens großen Datensätzen trainiert wurden. Firmen wie OpenAI und Google durchforsten das Internet auf der Suche nach Bildern und zugehörigen Textbeschreibungen. Das können Bildunterschriften von Online-Artikeln sein, Beschreibungen der Künstler auf Plattformen wie Instagram oder ArtStation oder auch der Inhalt des alt-Attributs beim img-Tag beliebiger HTML-Seiten.

Die Beschreibungen sind von sehr unterschiedlicher Qualität. Manchmal umschreiben ein oder zwei Sätze den Bildinhalt für blinde Webseitenbesucher. Fotografen geben mitunter Details zur verwendeten Kamera und zum Objektiv an. Maler nennen einen Bildtitel und dokumentieren, ob sie mit Ölfarben oder digital gearbeitet haben.

Manchmal hat das Trainingsmaterial aber auch das Potenzial, die KI auf einen Holzweg zu führen. Tweets zum Beispiel verstecken im Zusammenspiel zwischen Bild und Text gerne mal Witze. Ironie, Verneinungen, Rechtschreibfehler und Irrtümer: Es gibt viele Möglichkeiten, einer KI Schwierigkeiten zu bereiten.

Die knappen, mitunter widersprüchlichen und in die Irre führenden Textbe-

schreibungen sind einer der Gründe dafür, dass es eine Wissenschaft für sich ist, Formulierungen zu finden, mit denen man den KIs Bilder entlockt, die den eigenen Vorstellungen entsprechen. Der Fachbegriff für diesen Prozess lautet Prompt Engineering, weil die Entwickler der KI-Modelle den Eingabesatz „Prompt“ nennen.

Wir haben verschiedene Generator-KIs mit Dutzenden Prompts gefüttert und unsere Erfahrungen ab Seite 68 zusammengefasst. Unsere Tipps zeigen Ihnen, worauf Sie bei der Formulierung von Prompts achten sollten, um bei Ihren Versuchen mit Stable Diffusion, Midjourney & Co. schneller zu ansehnlichen Ergebnissen zu kommen.

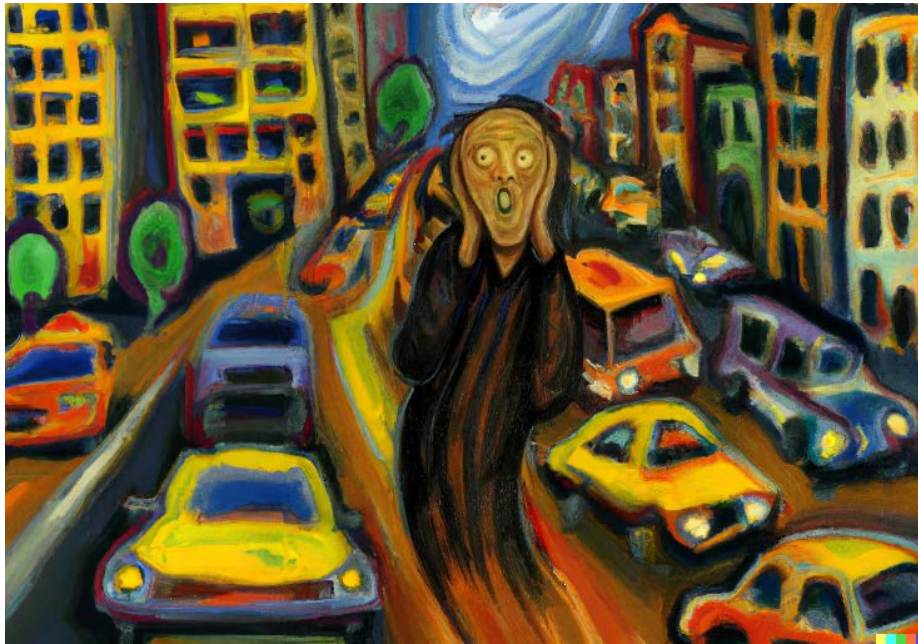
Nebenwirkungen

Mit den Trainingsdaten und ihrer Beschaffung hängen noch zahlreiche weitere Probleme der Bildgeneratoren zusammen. Sie reproduzieren zum Beispiel gesellschaftliche Klischees, die sie aus den Daten gelernt haben, wie unter anderem Experimente der Neuen Zürcher Zeitung zeigen: Auf die Eingabe des englischsprachigen „doctor“ hin generierten Dall-E, Midjourney und Stable Diffusion Bilder von Ärztinnen und Ärzten. Wurde „competent doctor“ vorgegeben, stiegen in allen Programmen der Männeranteil und das Alter. Zur Eingabe „prisoner“ erzeugten die Bildgeneratoren nur noch Männer – und anders als bei Ärzten tauchten nun auch Menschen mit dunkler Hautfarbe auf.

Nicht jeder mag damit einverstanden sein, dass sein Konterfei ungefragt für das Training einer KI herangezogen worden ist. Auf ähnliche Weise könnten Kreative



So stellt sich Stable Diffusion „a fox astronaut with the cosmos reflecting on the glass of his helmet dreaming of the stars“ vor.



Mit In- und Outpainting kann man in Dall-E 2 klassischen Motiven eine neue Bedeutung verpassen. Hier wurde aus dem Bild „Der Schrei“ von Edvard Munch der Hintergrund ausradiert und durch neue Frames mit dem Prompt „a crowded city with much traffic edvard munch“ ersetzt.

ein Problem damit haben, dass ihre Werke in das Training einfließen – Training für Systeme, die sich möglicherweise mal als Jobkonkurrenten herausstellen.

In vielen Fällen ist es aber gar nicht möglich nachzuvollziehen, ob die eigenen Bilder in ein KI-Modell eingeflossen sind. Das liegt daran, dass die Modelle aus Nutzersicht Black Boxes sind: Man kann nur einen Prompt eingeben und den Output ansehen – welche Bilder im Trainingsdatensatz steckten und welche Rolle diese wiederum für eine bestimmte Ausgabe spielten, lässt sich von außen nicht erkennen.

Die Betreiber des Projekts Spawning wollen ein Bewusstsein für das Problem schaffen. Mit ihrem Dienst „Have I Been Trained?“ erschließen sie die Bilddatenbank LAION-5B. Die ist mit 5,85 Milliarden Bildern und dazugehörigen Beschreibungstexten einer der größten öffentlich verfügbaren Datensätze für das Training von Bild-KIs. Besucher der Website können die Datenbank per Text oder per Bild durchsuchen.

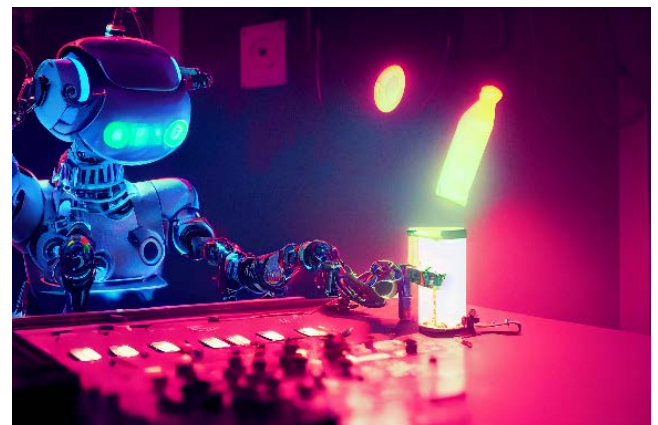
Urheberrecht: Kein Problem

Einig sind sich die meisten Juristen immerhin, dass per KI erstellte Bilder im Normalfall nicht unter das Urheberrecht fallen und daher frei genutzt werden dürfen. Dies ergibt sich hierzulande daraus, dass es sich für einen Schutz aus dem Urheber-

rechtsgesetz um ein Ergebnis eines menschlichen Schaffens handeln müsste. Daran fehlt es bei computergenerierten Grafiken offenkundig. Ähnlich hat auch das US-Copyright-Office entschieden, welches mehrfach die Einstufung von computergenerierten Bildern abgelehnt hat. Die Folge: Derartige Bilder dürfen frei genutzt werden.

Eine Ausnahme kann es allerdings dann geben, wenn die KI erkennbar große Teile eines bestehenden Werkes wiedergibt. Handelt es sich bei dem reproduzierten Bild um ein Foto eines Menschen, wäre das auch datenschutzrechtlich kritisch. Werden dagegen Fotos von Menschen nur für das Training einer KI genutzt, so gehen Juristen von einem sogenannten „berech-

Letztlich ist es Glückssache, was eine KI auswirft. Man produziert auch viel Ausschuss. Dieser deformierte Roboter entstand beim Versuch, eine Illustration für c't zu erzeugen.



tigten Interesse“ zur Nutzung dieser Bilder aus – und dürften ihr Okay geben. Allerdings ist das Thema der KI-Kunst noch neu und solche Fragen dürften die Experten noch einige Zeit beschäftigen.

Inhalte-Filter

Problematisch können auch erzeugte Bilder sein, die nicht jugendfrei sind oder anderweitigen Schaden anrichten könnten. Die Betreiber der Dienste versuchen, das durch restriktive Nutzungsbedingungen zu unterbinden. Der Dall-E-Entwickler OpenAI zum Beispiel verbietet es ausdrücklich, solche Bilder zu erstellen.

Zu den verbotenen Inhalten zählen Gewalt, „Belästigung“, Sexualität und Nacktheit. Tabu sind aber „große Verschwörungen oder Ereignisse im Zusammenhang mit wichtigen laufenden geopolitischen Ereignissen“, Inhalte, „die zur Beeinflussung des politischen Prozesses oder zum Wahlkampf verwendet werden können“ sowie „Bilder von Personen des öffentlichen Lebens“. Nutzer sind auch zur Transparenz angehalten. Sie dürfen nicht verschleiern, dass die erzeugten Bilder von einer KI stammen.

Offensichtlich nutzt OpenAI auch Wortfilter, um seine Regeln durchzusetzen. So weigerte sich Dall-E bei unseren Versuchen, ein Bild mit dem Prompt „elon musk flying on a blue bird digital art“ zu erzeugen.

Das Ende der Kunst?

Momentan entsteht ein großer Hype um die KI-Bildgeneratoren. Der Stockfoto-Anbieter Shutterstock etwa will künftig direkt auf seinem Portal mit Dall-E von OpenAI Bilder generieren – ebenso wie Microsoft mit seiner Suchmaschine Bing.

Derweil werden die KI-Dienste schnell weiterentwickelt und etwa um die Funktionen Inpainting und Outpainting

ergänzt. Inpainting erlaubt es dem Nutzer, Bildbereiche zu löschen, in die die KI mithilfe des Prompts neue Inhalte hinein-erfindet. Beim Outpainting wiederum kann eine KI ein Originalbild über den Rand hinaus erweitern. Dabei übernimmt sie wahlweise den Stil des Originals oder variiert ihn.

Die künstlichen Künstler haben ganz offensichtlich das Potenzial, so manchen grafischen Gestaltungsprozess zu verändern. Vielleicht ist Ihnen ja schon das eine oder andere von einer KI gestaltete Bildelement untergekommen – zum Beispiel in der c't: Das Aufmacherbild auf Seite 104 in Ausgabe 23 hat eine KI fabriziert.

Als im August ein von Midjourney generiertes Bild einen Kunstwettbewerb gewann, wurde auf Twitter und in einschlägigen Blogs schon diskutiert, ob die KI-Bildgeneratoren den „Tod der Kunst“ einläuten. Aber das scheint überzogen zu sein. Zwar finden viele Menschen zumindest einige KI-Bilder schön, der größte Teil ist aber auf eine unangenehme Art seltsam oder schlichtweg sinnloser Ausschuss. Vor allem aber ist es mit Prompt Engineering alleine nicht möglich, den Prozess so zu kontrollieren, dass das Ergebnis exakt den Erwartungen entspricht.

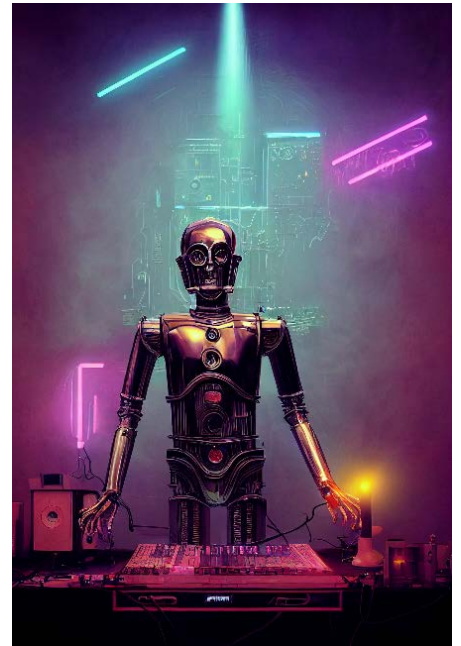
Genau aus diesem Grund haben wir gar nicht erst versucht, die Aufmacherbilder für die Artikel dieses Schwerpunkts von einer KI malen zu lassen, sondern einen Grafiker beauftragt. Die vier Bilder

sollten wiedererkennbare Elemente haben, die eine grafische Verbindung zwischen den Artikeln knüpfen. So etwas kann man bei KI-Generatoren nicht bestellen.

Dass die KI-Bilder schwer zu beeinflussen sind, wird sich auch mit den nächsten Generationen der Generatoren nicht ändern. Die Auflösungen werden sicherlich steigen, die Bezüge in den Bildern werden logischer werden, die Werke werden immer weniger wie Wimmelbilder aussehen. Aber es wird schwierig bleiben, bestimmte Ausgabe zu erzeugen, bei denen Komposition, Inhalte und Stil den eigenen Vorstellungen entsprechen.

Wer sich mehr Kontrolle erhofft, muss auf verbessertes Inpainting hoffen, das sicherlich in Zukunft eine größere Rolle in Bildbearbeitungsprogrammen spielen wird. Photoshop hat jetzt schon eine Funktion für KI-Inpainting. Die kann man aber bisher noch nicht mit einer Texteingabe steuern.

Für die Zukunft ist also zu erwarten, dass Bildgeneratoren Teil des Werkzeugkastens von Illustratoren und Künstlern werden, diese aber nicht ersetzen. Kreative werden die Grafik-KIs als Inspirationsquelle nutzen oder als Lieferanten von Rohmaterial – wie beim Aufmacher auf Seite 130 der c't 23. Dafür hat eine Kollegin mehrere Bilder mit Stable Diffusion generiert, zu einer Komposition zusammengestellt, in Photoshop einige Fehler retu-



Midjourney: „C3PO the robot god of music production wielding the magic“. Haben Sie's gesehen? Dieses Bild illustriert einen Artikel in Heft 23.

schert, das Mikro ausgebessert, die Augen verändert und die Farben mittels Filter aufgefrischt.

Spaß mit Bildern

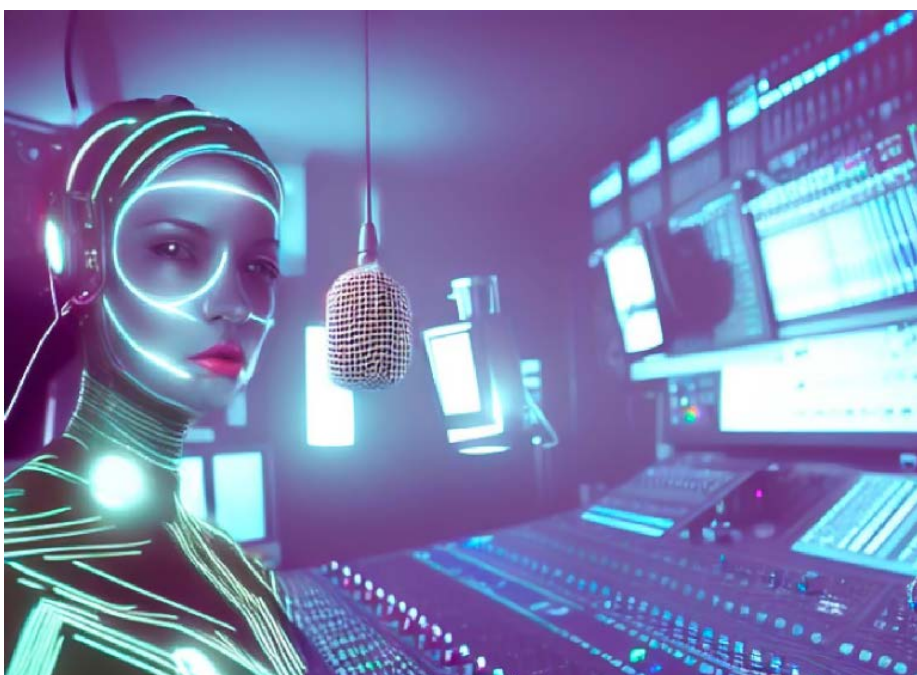
Man muss aber kein Profi sein, um die faszinierenden Bildgeneratoren zu nutzen. Die meisten Dienste sind zumindest in einer Basisversion kostenlos. Und es macht Spaß, sich an einem Werk zu versuchen – probieren Sie es einfach einmal aus. Einige c't-Kollegen haben bereits besonders gelungene Werke ausgedruckt und an die Wand gehängt.

Sie können auch ein Quiz daraus machen: Generieren Sie Bilder zu einem Songtitel, Bandnamen oder einer Textzeile und lassen Sie Ihre Freunde den Prompt raten. Oder sie veranstalten ein „Prompt Battle“ wie die Dozenten und Studenten der Hochschule für Technik und Wirtschaft in Dresden: Dabei geht es darum, durch die besseren Prompts Dall-E die eindrucksvolleren Bilder zu entlocken. (jo@ct.de) **ct**

Literatur

- [1] Dirk Hecker und Gerhard Paaß: Sprachversther, GPT-3 & Co. texten überzeugend, aber nicht faktentreu, c't 9/22, S. 64

Erwähnte Bildgeneratoren: ct.de/y299



Für dieses Aufmacherbild eines Artikels in c't 24 wurden mehrere Bilder mit StableDiffusion generiert, in Photoshop zusammengesetzt und nachbearbeitet.