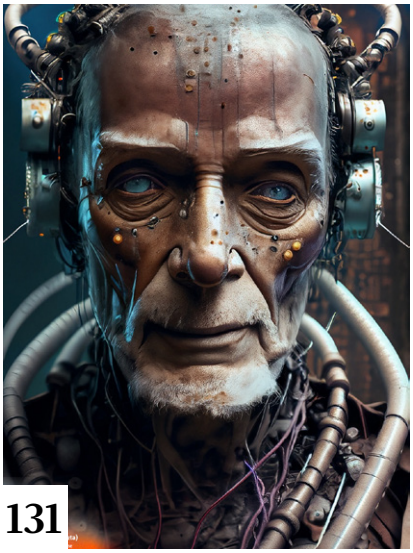
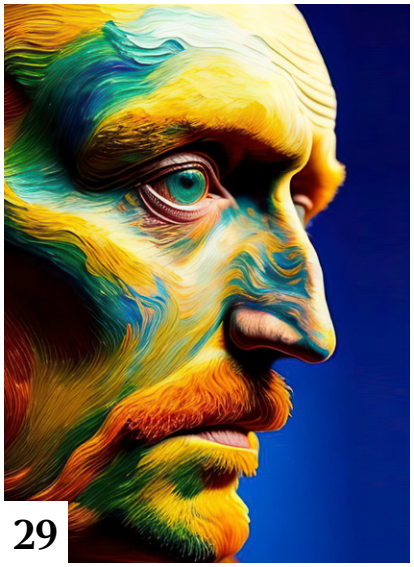
Inhalt

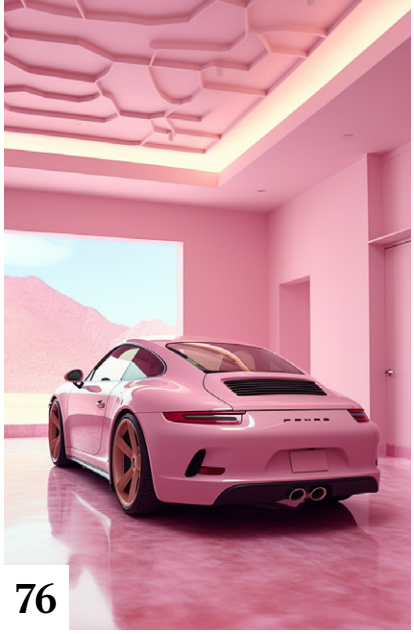

107


131


64


70

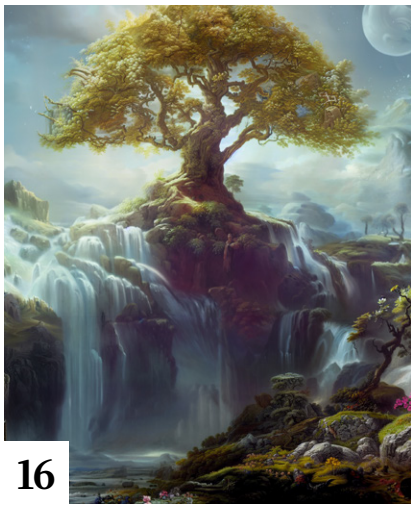

29


76


52


168


87


16


62


95


25

Teil III

Bilder auf Zuruf

Wer ein Bild mit bestimmten Eigenschaften braucht, konnte es bisher selbst fotografieren oder bei einem Stockfoto-Anbieter auf die Suche nach einem geeigneten Motiv gehen. Künftig könnten neuronale Netze solche Bilder auf Zuruf generieren – man muss ihnen nur sagen oder skizzieren, was man benötigt, und die Künstliche Intelligenz erfüllt fast alle Wünsche – auch völlig absurde.
Michael J. Hußmann

Nicht nur uns DOCMA-Redakteuren geht es gelegentlich so: Wir brauchen ein Foto, sei es als Aufmacher, als Illustration oder als Versatzstück in einer Montage – aber woher nehmen, wenn man nichts Passendes im eigenen Archiv findet? Für ein Strandmotiv kurz auf die Seychellen zu jetten, wäre nicht wirklich praktikabel, und so bleiben Stockfotos, unter denen man erst einmal etwas finden muss, das den eigenen Vorstellungen nahe kommt und sich idealerweise für kleines Geld lizenzieren lässt.

Dabei stand uns das Bild, das wir suchten, doch schon lebhaft vor Augen – wie könnte man daraus auf einem direkteren Weg ein Digitalbild machen? Verfahren der Künstlichen Intelligenz zeigen neuerdings einen Lösungsweg auf. Man braucht die gewünschte Szene nur grob zu skizzieren, und ein neuronales Netz berechnet daraus eine fotorealistische Version. Der Grafikkarten-Hersteller Nvidia verfolgt schon seit längerer Zeit sein GauGAN-Projekt *(www.docma.info/22382)* – der Name soll an den Maler Paul Gauguin erinnern–, bezieht sich aber auch auf die hier eingesetzte Architektur der „Generational Adversarial Networks, kurz „GAN“. In einem GAN arbeiten während der Lernphase zwei Netze gegeneinander: Eines wird trainiert, die eigentliche Leistung zu erbringen, während ein anderes lernt, den Output des ersten Netzes mit realen Bildern zu vergleichen und auf dieser Basis immer treffender zu kritisieren. Nvidias Interesse an KI-Projekten rührt daher, dass sie es können: Grafikprozessoren mit Tausenden von parallel arbeitenden Prozessorkernen sind auch ideal dazu geeignet, neuronale Netze zu simulieren, und sind daher die bevorzugte Hardware für das Training solcher Netze.

GauGAN2 kann nach dem Vorbild eines Fotos ähnliche Bilder generieren, aber es genügen auch schon grobe Skizzen, in denen man die Stellen markiert, an denen es beispielsweise Strand, Himmel, Wolken, Bäume oder Berge geben soll **[1]**. GauGAN2s Königsdisziplin ist allerdings, ein

Bild: GauGAN2 / Nvidia

Nach einer groben Skizze hat GauGAN2 diese Strandansicht erzeugt. Die Herkunft des dunklen Flecks unten rechts bleibt rätselhaft.

Bild aus einer englischsprachigen Beschreibung wie „Blue sky over the beach“ zu generieren **[2]**. Wer einen Windows-PC mit einer Nvidia-RTX-Grafikkarte besitzt, kann dies mit der „Studio Canvas“-App *(www.docma.info/22383)* ausprobieren; eine browser-basierte Version steht auf *www.docma.info/22384* auch allen anderen zur Verfügung.

An solchen Projekten wird derzeit von vielen Forschungseinrichtungen gearbeitet. OpenAI beispielsweise hat DALL-E entwickelt *(www.docma.info/22385)*, benannt nach Salvador Dalí und dem Roboter WALL·E aus dem gleichnamigen Pixar-Film. Dieses System generiert aus – teilweise recht komplexen – Beschreibungen dazu passende Bilder. Was können solche KI-Lösungen heute bereits leisten, woran scheitern sie noch und worin liegen ihre Beschränkungen?

Grenzen der Fantasie

GauGAN2 wurde mit Millionen von Landschaftsfotos trainiert, deren Elemente sich als Versatzstücke zu immer neuen Bildern kombinieren lassen, aber es kann nichts produzieren, das es nie gesehen hat, und große Teile der Welt sind ihm anscheinend unbekannt. Wenn man eine Wüstenlandschaft (englisch „desert“) verlangt, wird das generierte Bild unweigerlich an Arizona erinnern, nicht an die Sahara oder die Wüste Gobi, und darin platzierte Bäume werden keine Palmen. GauGAN2 kann zwar verschiedene Motive in einem Bild kombinieren, beispielsweise „A desert with mountains in the background“, aber wenn man eine Flusslandschaft erzeugen möchte, tut sich das System schwer – der Fluss ähnelt dann einer Straße oder ist gar nicht zu erkennen **[3]**. Generell geraten Kombinationen, die in dieser Form real vorkommen und in den Trainingsdaten enthalten waren, realistischer, als wenn man mit GauGAN2 fantastische Welten zu erschaffen versucht.

Bild: GauGAN2 / Nvidia

Der Fluss, den das neuronale Netz in die Berglandschaft eingefügt hat, wirkt wenig realistisch.

Das System ist darauf trainiert, Bilder zu erzeugen, keine dreidimensionalen Szenen. Auch die Skizzen, in denen man Kategorien wie „Berg“, „Baum“ oder „Wolke“ auswählt und mit einem Pinsel- oder Füllwerkzeug die Flächen in das Bild malt, in denen solche Elemente erscheinen sollen, haben nur zwei Dimensionen. Die Tiefenstaffelung lässt sich nicht andeuten, und man muss sich darauf verlassen, dass sich aufgrund des Trainings mit realen Fotos ein stimmiger Bildaufbau ergibt. Das ist nicht immer gewährleistet: Manchmal ragt ein ebener Teil einer Landschaft steil in den Himmel, weil die KI eine Ansicht aus der Vogelperspektive und ein Foto aus Augenhöhe miteinander verquickt **[2]**. ▶

Als Interpretation von „Blue sky over a beach“ hat GauGAN2 unvereinbare Perspektiven in einem Bild kombiniert.

Bild: GauGAN2 / Nvidia

Den Versuch, einen Dolmen zu skizzieren, versteht GauGAN2 falsch, da die dritte Dimension unberücksichtigt bleibt.

Bild: GauGAN2 / Nvidia

5

6

Bilder: GauGAN2 / Nvidia

*Wenn GauGAN2 keinen Anhaltspunkt für die Bildkomposition findet, da es in dem im Training präsentierten Bildmaterial kein passendes Vorbild gab, werden vorgegebene Formen – hier die einer Kuh – mit teils kuh-artigen **[5]**, teils völlig surrealen Versatzstücken **[6]** gefüllt.*

Es fehlt auch an Wissen darüber, welche typische Gestalt verschiedene Objekte haben. Das zeigt sich, wenn man ein weiteres Feature des Systems nutzt und ein reales Foto lädt, um es von GauGAN2 analysieren zu lassen und daraus wiederum ein neues Bild zu generieren. Die erkannten Konturen von beispielsweise einer Kuh werden mit mehr oder minder kuhförmigen Details gefüllt, aber das Ergebnis sieht aus wie eine Kreatur Frankensteins **[5, 6]**.

Jedes Bildergebnis kann GauGAN2 mit verschiedenen Lichtstimmungen rendern, wobei ein Schwerpunkt der vorgegebenen Stile bei Sonnenauf- und -untergängen, nächtlichen Szenen und Nebel liegt. Einige dieser Varianten lassen vorhandene Mängel weniger offensichtlich hervortreten, da sie der Nebel oder die Dämmerung kaschiert.

GauGAN2 ist auch kaum in der Lage, Beziehungen zwischen Wörtern zu erkennen, so dass zwar alle Substantive berücksichtigt, Verben und Präpositionen aber offenbar ignoriert werden. Je präziser man weiß, was man will, und je ungewöhnlicher die eigenen Vorstellungen sind, desto eher wird das System an der Textinterpretation scheitern und ein unbefriedigendes Resultat abliefern. Selbst wenn die Vorgabe hinreichend gut verstanden wird, zeigt GauGAN2 Schwächen, wenn es die verschiedenen Elemente in einen plausiblen Zusammenhang bringen soll.

Unmögliches möglich machen

Das DALL·E-System von OpenAI ist auf die Umwandlung von Text in Bilder beschränkt, leistet in dieser Disziplin jedoch mehr als GauGAN2. Es wurde mit 250 Millionen Fotos und den dazugehörigen Bildunterschriften trainiert, wobei der Input des neuronalen Netzes nicht allein die Texte waren, zu denen es ein Äquivalent des zugehörigen Bildes generieren sollte. Vielmehr bestand der Input in der Trainingsphase aus Bild *und* Text; die Aufgabe des neuronalen Netzes lag darin, durch Abstraktion eine interne Repräsentation dieser Inputs zu konstruieren, aus der sich neue Bilder erzeugen lassen.

Auf Basis solcher Abstraktionen hat das neuronale Netz Fähigkeiten erworben, die gar nicht Teil des Lehrplans waren. So kann man DALL·E Bilder eines Tapirs generieren lassen, der die Form eines Akkordeons hat, und das System wird verschiedene Varianten finden, beides in einem Objekt zu vereinigen **[7]**. Einfachere Aufgaben, wie einer Katze einen Hut, eine Fliege, eine Sonnenbrille oder Kopfhörer aufzusetzen, bereiten DALL·E erst recht keine Probleme. Auch besondere Perspektiven wie eine Vogelperspektive oder eine Fisheye-Ansicht lassen sich auswählen und werden überzeugend umgesetzt **[9]**.

Das neuronale Netz beherrscht auch, anders als GauGAN2, Ansätze der sogenannten Variablenbindung: Wenn es eine Zeichnung eines Baby-Igels in einem Pullover mit weihnachtlichem Muster erzeugen soll, der einen kleinen Hund Gassi führt, dann versteht die KI prinzipiell, dass der Igel und nicht der Hund den Pullover tragen soll. Jedenfalls funktioniert dies in vielen Fällen; manchmal allerdings bekommt auch der Hund einen Pullover, oder am Ende der Leine befindet sich statt des Hundes ein weiterer, kleinerer Igel **[8]**. Solche noch weiter auszubauende Leistungen erfordern

7

Bilder: DALL·E / OpenAI

Das neuronale Netz in DALL·E hat keine Schwierigkeiten, einen Tapir in Gestalt eines Akkordeons zu imaginieren.

Bilder: DALL·E / OpenAI

DALL·E s Illustrationen eines Baby-Igels in einem weihnachtlich gemusterten Pullover, der einen Hund Gassi führt.

nicht nur eine aufwendigere Textanalyse, sondern vor allem die Fähigkeit der KI, komplexe Zusammenhänge zwischen den Elementen einer Szene zu repräsentieren.

Auch DALL·Es Fähigkeiten können Sie im Browser ausprobieren (*www.docma.info/22385*). Allerdings ist hier keine freie Texteingabe möglich; es lassen sich nur Varianten verschiedener Mustersätze wählen. Das System erzeugt dann jeweils eine große Zahl von Varianten. Die Auflösung der generierten Bilder ist leider eher gering, so dass sie noch nicht für praktische Anwendungen in Betracht kommen.

Wunsch und Wirklichkeit

Der Vorstellung, man müsse nur klar formulieren, was für ein Bild man sich wünscht, damit ein KI-System dieses produziert, wird der aktuelle Stand der Forschung noch nicht gerecht. Selbst wenn die generierten Bilder in sich stimmig sind und realistisch wirken, sind sie für die meisten Anforderungen noch zu gering aufgelöst.

Mit StyleGAN2, einem weiteren KI-System Nvidias, lassen sich sehr realistische Porträts produzieren, wovon man sich unter *https://thispersondoesnotexist.com* überzeugen kann – jeder Aufruf dieser URL ergibt ein neues, lebensecht wirkendes Porträt eines Menschen, der nicht existiert **[10]**. Dieser Erfolg ist vor allem dadurch zu erklären, dass alle menschlichen Gesichter die prinzipiell gleiche Form haben – die relativen Positionen von Augen, Nase und Mund lassen nur wenig Spielraum. Schon die vom gleichen System erzeugten Bilder von Tieren haben eine schlechtere Qualität, weil es hier weit mehr Variabilität gibt. Da StyleGAN2 ebenso wenig wie GauGAN2 über Wissen zu den dargestellten Motiven verfügt, zeigen beispielsweise die generierten Pferdebilder zwar pferdeartige Details, aber nicht immer den Körperbau eines Pferdes. Auch Landschaftsbilder sind zu vielgestaltig, als dass es einfache Lösungen gäbe.

Grundsätzliche Fortschritte solcher KI-Systeme werden wir vermutlich erst dann sehen, wenn sie sich auf ein

Bilder: DALL·E / OpenAI

DALL·E gelingt es oft, wenn auch nicht in allen vorgeschlagenen Varianten, wie gefordert einen Fuchs in einer Fisheye-Perspektive abzubilden.

enzyklopädisches Wissen über Menschen, Tiere, Pflanzen wie auch unbelebte Motive wie Berge, Häuser oder Maschinen stützen können, und dazu auf das Wissen über die Gesetze der Perspektive. Bevor wir uns von einer KI exakt maßgeschneiderte und lizenzfreie konkrete Landschaftsbilder auf Zuruf generieren lassen können, wird noch etwas Zeit vergehen. ■

Bilder: StyleGAN2 / Nvidia

StyleGAN2 von Nvidia erzeugt ziemlich realistische Porträts nicht existierender Personen.

PROMPT-ENGINEERING

Wie man mit einer KI spricht

Eine künstliche Intelligenz zur Bilderzeugung versteht nicht, was Sie ihr vorgeben, sondern stellt nur nach, was sie aus Milliarden von Bildern und zugehörigen Beschreibungen und Kommentaren gelernt hat. Vor diesem Hintergrund müssen Sie versuchen, der KI ein Angebot zu machen, das sie nicht ablehnen kann. Olaf Giermann zeigt, welche Faktoren dabei eine Rolle spielen.

Die Mafia-Filmreihe „Der Pate“ von Francis Ford Coppola wird oft zu den besten Filmen aller Zeiten gezählt. Ein wiederkehrendes Thema ist, dass Freunde der „Familie“ den Paten um einen Gefallen bitten (wie im Aufmacherbild). Er erfüllt diesen dann häufig mit einem „Angebot, das man nicht ablehnen kann“, setzt Widersacher also auch mit Gewalt unter Druck.

Nun brauchen Sie im Umgang mit einer Text-zu-Bild-KI nicht kriminell zu werden, aber Sie müssen verstehen, dass sie schwer von Begriff ist und genaue Anweisungen benötigt. Eine solche Anweisung bezeichnet man als Prompt.

Künstliche Text-zu-Bild-Intelligenzen sind noch recht begriffsstutzig und wenig intuitiv. Diese werden jedoch wahrscheinlich schnell sehr viel besser werden. Es kann aber nie schaden, Prompts möglichst klar und strukturiert zu formulieren.

Prompt-Engineering

Prompt-Engineering steht für das systematische Anlegen eines Prompts, um das gewünschte Bildergebnis zu erzeugen. Es gibt auch den oft synonym gebrauchten Begriff *Prompt-Crafting*, der eher den kreativen Weg zu einem funktionierenden Prompt beschreibt. Ein Prompt besteht dabei immer aus der Textbeschreibung dessen, was Sie im Bild gern sehen wollen, vielleicht auch einem Referenzbild und Parametern, die spezifisch für jede KI-Software sind. Das Wichtigste ist die Beschreibung, die ganz anders ausfallen kann, als wenn Sie einem Menschen ein Bild beschreiben würden.

Prompt: „ein foto eines haus im wald“. Während eine korrekte Grammatik kaum ein Rolle spielt, ist vor allem die Reihenfolge der Worte wichtig. Deutsch kann wie hier funktionieren, mit Englisch haben Sie mehr Erfolg.

Die Textbeschreibung

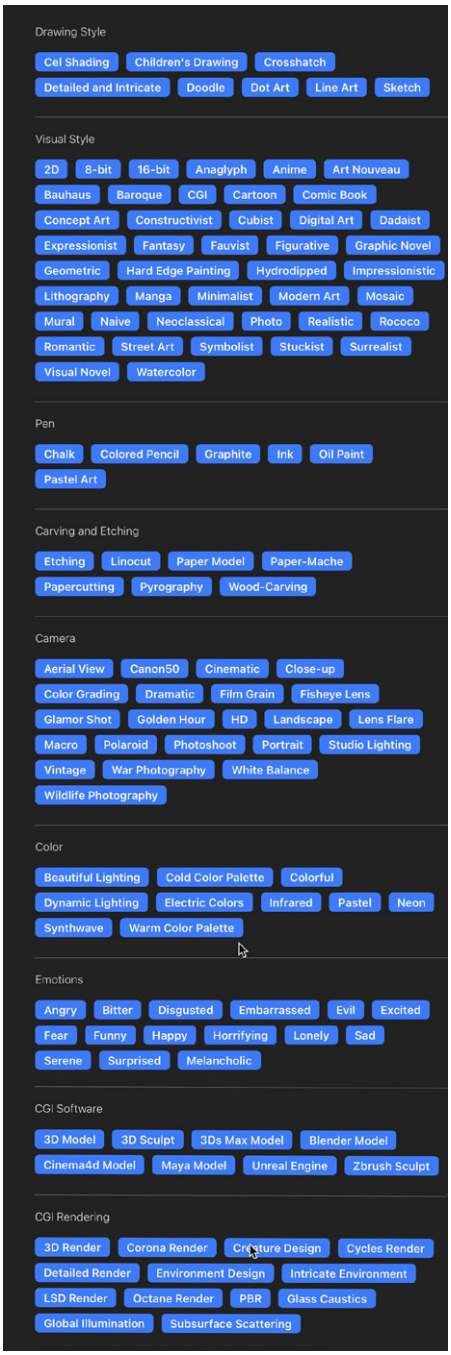
Die Stable-Diffusion-KI wurde anhand von fünf Milliarden Bild-Text-Paaren aus dem Internet trainiert. Damit hat sie mehr Bilder „gesehen“, als jeder Mensch es in einem Leben könnte. Da die meisten Bilder im Web eher mit englischen als deutschen Beschreibungen versehen sind, sollten Sie ebenfalls englische Begriffe verwenden, obwohl Deutsch auch funktionieren kann **[1]**. Im Zweifel übersetzen Sie Ihre deutschen Prompts einfach mit *Deepl.com* ins Englische.Wenn Sie einem Menschen nun ein Bild beschreiben müssten, würden Sie wahrscheinlich das Hauptmotiv benennen, etwas über den Hintergrund, die Farbe und die wichtigen Details sagen. Der Hörer hat dann eine ungefähre Vorstellung. Aber keine genaue. Nehmen wir an, ich bitte Sie, eine Giraffe für mich zu zeichnen. Die Fähigkeit zu zeichnen setzen wir einfach einmal voraus. Aber sie hängt sehr eng damit zusammen, was Sie wissen und was Sie sich vorstellen können. Eine Giraffe hat einen langen Hals, das weiß jedes Kind. Und dann? Welche Form haben ihre Ohren? Der Schwanz? Die Augen? Wie genau war noch einmal das Fell der Giraffe gemustert? Solche Details und Versatzstücke sind für eine KI gar kein Problem. Die hat sie zu Abertausenden beim maschinellen Lernen kennengelernt. Irgendeine ▶

erkennbare Giraffe spuckt Ihnen jede KI deshalb im Handumdrehen aus **[2]**. Und dann müssen Sie präziser beschreiben, was genau Sie möchten.

Künstler kopieren?

Aktuell ist es in Midjourney und Stable Diffusion 1.5 noch möglich, einfach den Namen der Künstler oder Fotografen im Prompt anzugeben, deren Stil Sie nachahmen möchten. Rechtlich und moralisch ist das noch eine Grauzone, aber der wohl schnellste Weg zu konkreten Ergebnissen. Die Verwendung tausender geschützter Fotos und Artworks für das Machine Learning hat insbesondere Midjourney zu dem beeindruckenden Werkzeug gemacht, das es jetzt ist. In Stable Diffusion 2.x wurden die Werke vieler lebender Künstler bereits aus dem Datensatz entfernt. Es bleibt abzuwarten, wie sich die Situation weiterentwickelt und ob Künstler für ihre Vorlagen nicht auch entsprechend entlohnt werden können. Kunstrichtungen und Namen früherer Künstler können Sie selbstverständlich auch weiterhin in Prompts verwenden.

Praktisch: Die auf Stable Diffusion basierte Mac-Software „DiffusionBee“ hält eine Menge nützlicher Stichwörter zur stilistischen Bildbeschreibung bereit, die per Mausklick eingetragen werden.



Wie eine Maschine denken

Falls Sie ein besonders hochwertig wirkendes Bild ohne Künstler-Prompts erzeugen möchten, könnten Sie zwar Begriffe wie „hochwertig, elegant, Bokeh“ eingeben, aber die KI versteht das vielleicht nicht richtig. Jedoch kennt die KI auch die Metadaten der Fotos, die sie beim Machine-Learning vorgesetzt bekommen hat. Und das können Sie sich zunutze machen, indem Sie einfach technische Parameter eingeben. Das kann sogar ganz detailliert sein, wie etwa „Canon EOS 5D Mk IV DSLR, f/1.4 Aperture, 1/125 second shutter speed, ISO 100, professional, Adobe Photoshop“. Also alles Merkmale von hochwertigem Equipment und professionellen Workflows.

Hinweise auf 3D-Programme können für hochwertige Ergebnisse mit einem „CGI-Look“ sorgen. Einige derartige Stichworte finden Sie in der Stichwortliste links ganz unten.

Der Weg zum Prompt

Letztlich gibt es (noch) kein Patentrezept zum richtigen Prompt. Fangen Sie am besten mit einer einfachen Beschreibung an und ergänzen Sie vom Ergebnis ausgehend immer weitere Hinweise. Hier bietet es sich an, auch Details einzufügen, die Sie mit dem gesuchten Bild assoziieren. Im Falle der Aufmacher-Illustration ergänzte ich beispielsweise den Filmtitel („The Godfather“), die Bezeichnung des Schauspielers (Marlon Brando) und seiner Rolle im Film (Don Vito Corleone).

Negative Prompts

Mit negativen Prompts teilen Sie der KI mit, was Sie *nicht* möchten. In Midjourney funktioniert das per Tag „--no“ gefolgt von einer Beschreibung der unerwünschten Inhalte. In Stable Diffusion gibt es je nach gewählter Bedienoberfläche entweder ein gesondertes Eingabefeld für negative Prompts (bei Automatic1111, siehe *www.docma.info/22716*) oder Sie schließen Begriffe mit eckigen Klammern aus (bei Invoke AI, siehe *www.docma.info/22717*)

Bild-zu-Bild

Um ein erstes Ergebnis weiter zu verfeinern, gibt es in Stable Diffusion die Möglichkeit, es an den „Image2Image“-Modus weiterzureichen und dann Variationen über weitere Prompts zu erzeugen. Am einfachsten ist das in Midjourney gelöst: Sie suchen sich das beste aus vier Ergebnissen aus und lassen von diesem Bild Variationen erzeugen.

KÜNSTLICHE INTELLIGENZEN IM VERGLEICH

Drei große Text-zu-Bild-Generatoren bestimmen die KI-Szene: DALL-E 2 und Midjourney sind proprietäre, kostenpflichtige Online-Programme, während Stable Diffusion eine Open-Source-Software ist, die Sie auch kostenlos lokal auf Ihrem eigenen Rechner ausführen können.

Auf einen Blick

- DALL-E 2: Man zahlt pro Prompt. Gut für Gelegenheitsanwender.
  www.openai.com/dall-e-2
- Midjourney: Abo-Modell mit künstlerisch ansprechenden Ergebnissen. Interaktion per Discord-Messenger.
  www.midjourney.com
- Stable Diffusion: Open-Source für Tüftler. Alles ist möglich, aber vergleichsweise kompliziert.
  www.stablediffusionweb.com

Massen-Berechnung
Der große Vorteil bei einer lokalen Installation von Stable Diffusion ist, dass Sie ohne Kosten Hunderte Varianten berechnen lassen können, sich dann einfach das beste Ergebnis aussuchen und dieses hochskalieren. Empfehlenswert ist hier das erwähnte *Automatic1111* für PCs mit einer Nvidia-Grafikkarte oder die *DiffusionBee*-App für Macs. So erzeugte Variationen sind meist subtil und kaum über einen gezielten Prompt zu erreichen. In Midjourney lädt das Abomodell zum Experimentieren ein – im Relaxed Modus werden die Bilder berechnet, sobald Serverkapazität frei wird. DALL-E 2 wird dagegen mit seiner Abrechnung jeder einzelnen Berechnung schnell teuer.

Bildbearbeitung
Als DOCMA-Leser haben Sie den großen Vorzug, sich (hoffentlich) mit Bildbearbeitung auszukennen. Statt also zu versuchen, jedes noch so kleine Detail von der künstlichen Intelligenz optimieren, hinzufügen oder entfernen zu lassen, nutzen Sie einfach die Möglichkeiten von Photoshop oder Affinity Photo. Retuschieren Sie störende Elemente oder kombinieren Sie das Beste aus mehreren KI-Ergebnissen über Ebenen und Masken.

Prompt-Inspiration
Um ein Gefühl für funktionierende Prompts zu bekommen, lassen Sie sich von anderen KI-Nutzern inspirieren. Dafür empfehlen sich beispielsweise die Prompt-Presets bei *www.openart.ai/presets*, für Midjourney-Anwender vor allem die Showcase-Seite *www.midjourney.com/showcase/recent* oder auch generell *www.lexica.art* und *www.arthub.ai*. Dort sehen Sie nicht nur die für die jeweiligen Bilder verwendeten Prompts, sondern können diese meist auch direkt per Schaltfläche in die Zwischenablage einfügen. Was Sie allerdings nicht sehen, ist der Aufwand an Variationsberechnung und etwaige Image-to-image-Prozesse. Wundern Sie sich also nicht, wenn Sie nicht sofort ähnlich eindrucksvolle Ergebnisse erhalten. ■

EIN TYPISCHER WORKFLOW MIT MIDJOURNEY

2

Prompt: giraffe

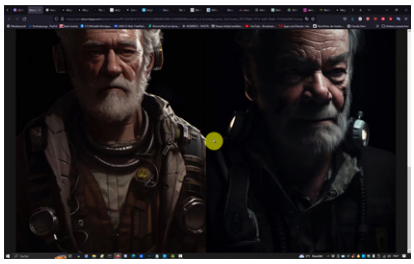
Prompt: a tiny giraffe in a coffee cup with its neck sticking out

Detaillierterer Prompt: a tiny giraffe within a coffee cup, with its neck sticking out, white porcellan cup, photorealistic, detailed, macro, Canon EOS 5D Mk IV DSLR

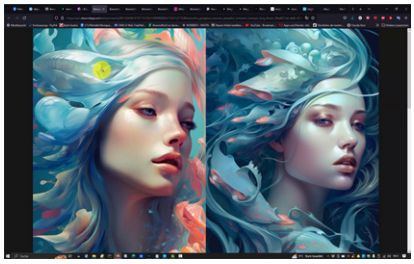
Hochskalierte Variante.

Der Prompt des Aufmacherbildes in Midjourney: The offer, scene from godfather, man wispering into the ear of the godfather, photorealistic, intricate details, 50 mm, f 1.8, --ar 2:3.

unten: zwei Ergebnisvarianten in Stable Diffusion mit demselben Prompt, ergänzt um „Don Vito Corleone“

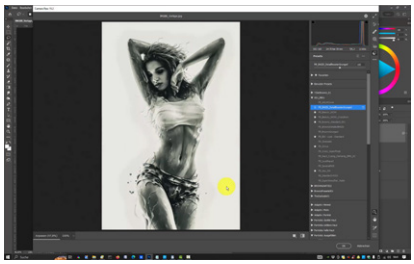


VIDEO 1 (21 min) *Erste Experimente mit der neuen Midjourney-Version 5*

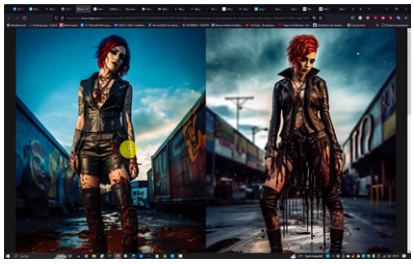


VIDEO 2 (20 min) *Eingabebilder mit Prompts variieren*

VIDEO 3 (17 min) *Studiofotos und Midjourney kombinieren. Vorsicht bei Aktfotos!*



VIDEO 4 (18 min) *Bonusvideo: (Teil-)Aktbilder mit der KI von www.mage.space*



VIDEO 5 (28 min) *Midjourney 5: der neue Befehl »Describe«*

DIE VIDEOS UND WEITERE BILDER FINDEN SIE UNTER
www.docma.info/22778

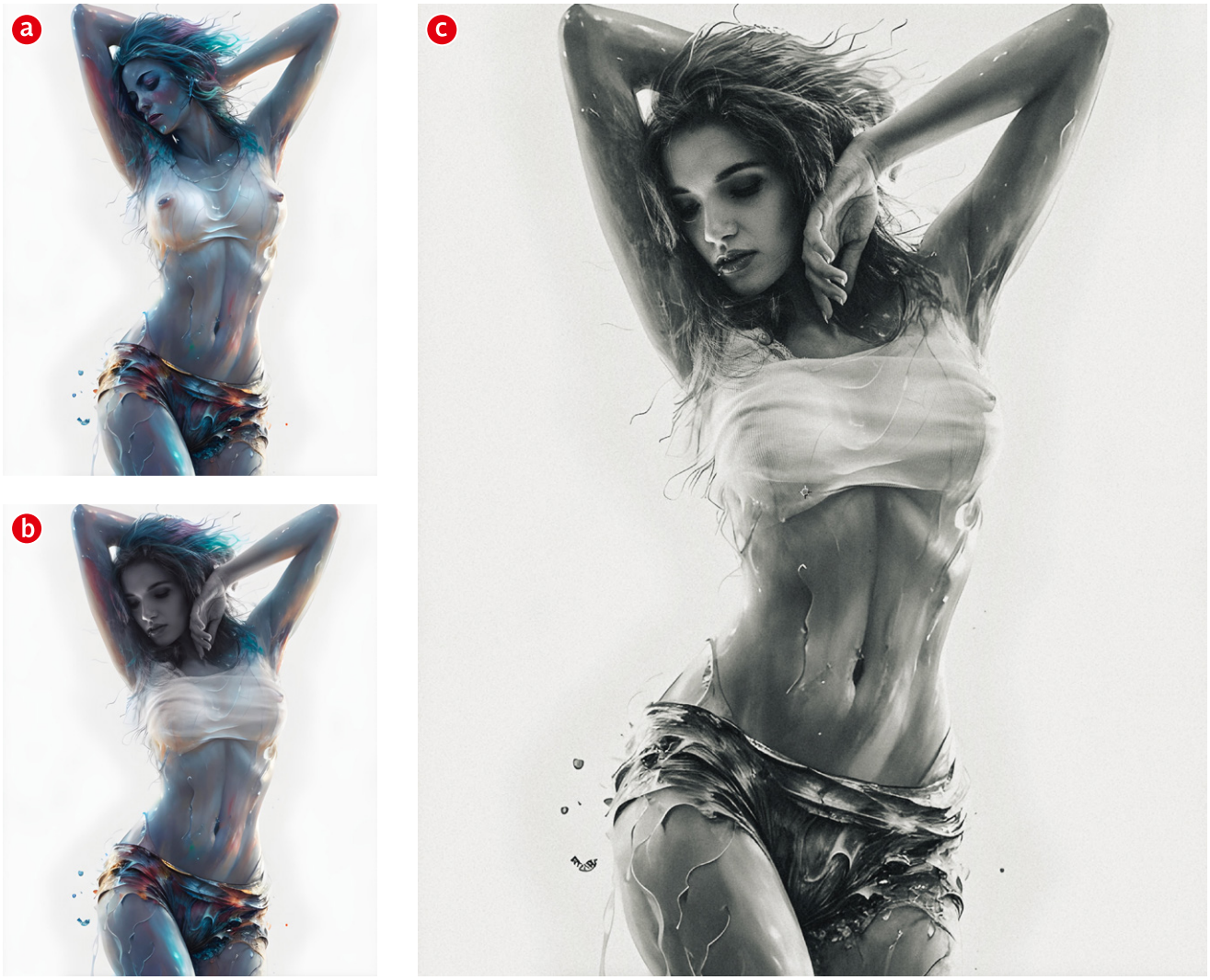
+ Video

MIDJOURNEY

Fotografisches mit Version 5

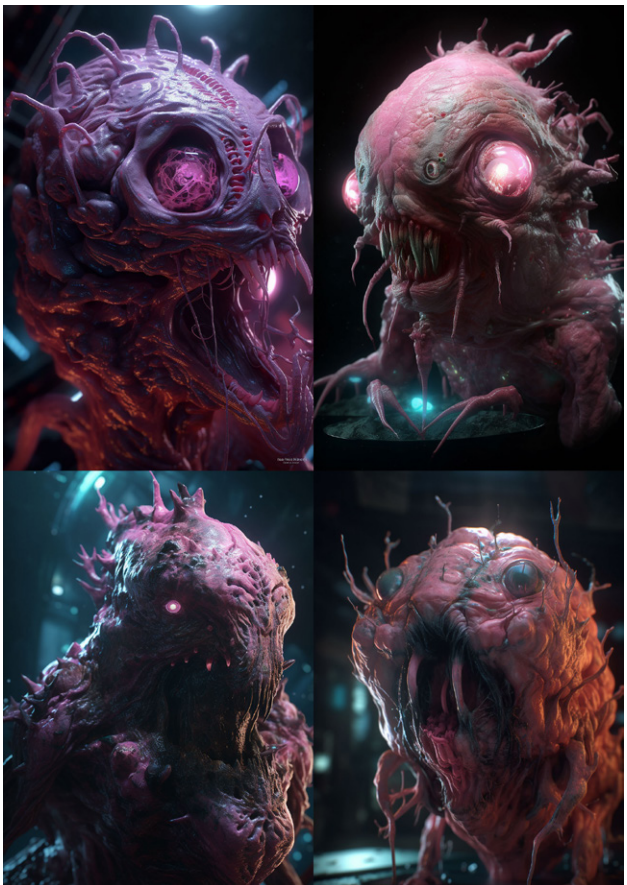
Peter Braunschmid zeigt in fünf Videos, was es in der beeindruckenden Midjourney-Version 5 Neues gibt und wie Sie Bilder und Skizzen in synthetische Artworks verwandeln.

Die Entwicklung von Midjourney verläuft weiterhin rasant. Die neue Version 5 ist ein Meilenstein der Text-zu-Bild-KIs. Midjourney 5 liefert mitunter bereits so fotorealistische Ergebnisse, dass man rätseln muss, ob es sich nicht doch „nur“ um ein Foto handelt – produziert aber auch häufig noch Midjourney-typische KI-Fehler im Detail und im Stil. In seinen Videos bringt Sie Peter Braunschmid auf den aktuellen Stand seiner Arbeitsweise mit Midjourney – zu der auch der Einsatz der Describe-Funktion zählt (siehe Video 5). Sie ermöglicht das Ableiten eines Prompts aus Bildern. ▶



Für dieses Bild hat Peter Braunschmid die meisten Likes bekommen, seit er bei Instagram ist. Von einem Studiofoto ausgehend erzeugte er die KI-Variante **(a)** *– aufgrund der Nacktheit mit www.mage.space, das auf Stable Diffusion basiert – und kombinierte in Photoshop den Kopf aus dem Foto sowie den Körper aus dem KI-Ergebnis* **(b)** *für das finale Bild mit einem Color Grading* **(c)**. *Siehe Video 4.*

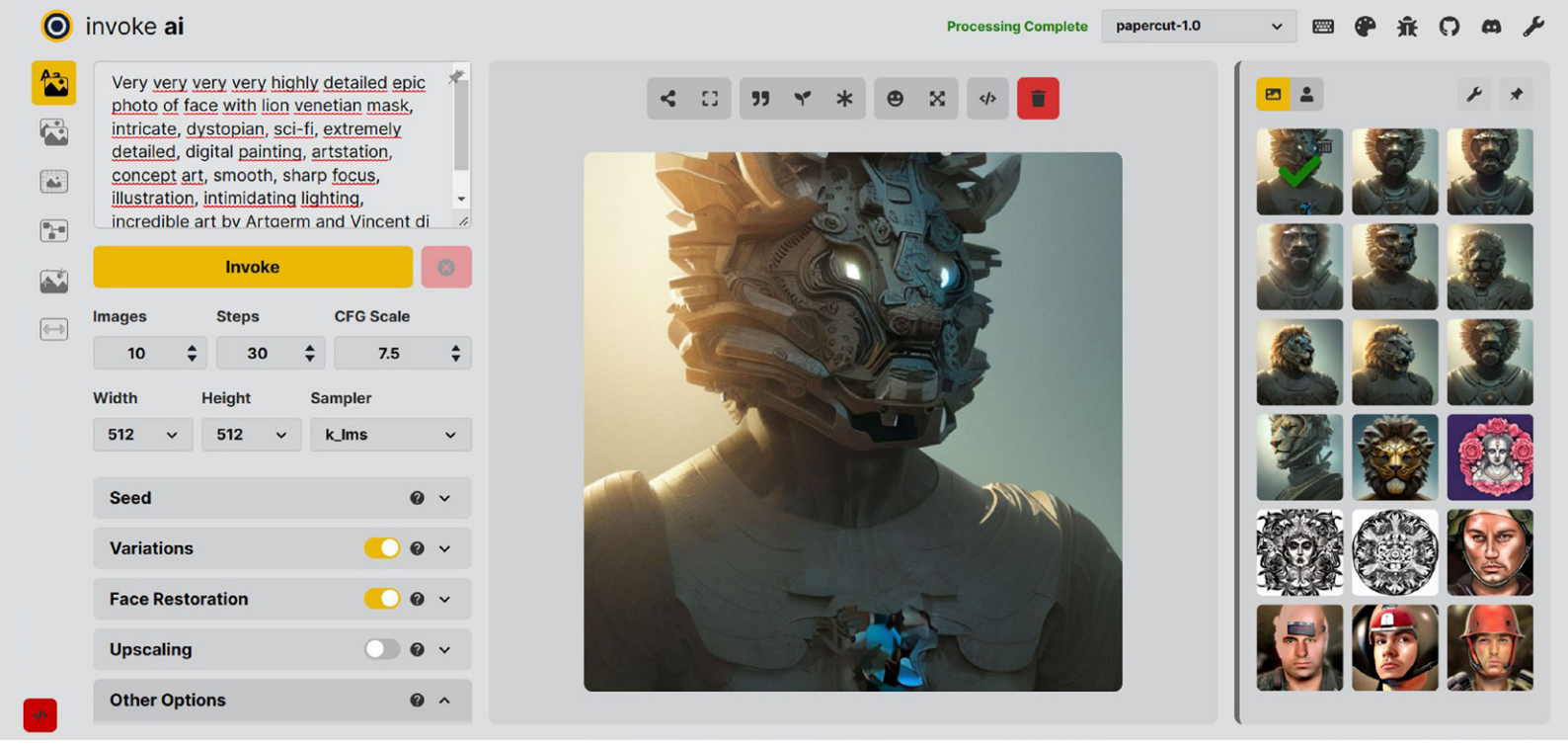
In Video 1 zeigt Peter Braunschmid, wie er seine Prompts aus verschiedenen Quellen zusammenstellt und für Midjourney 5 kombiniert.



*Das Ausgangsfoto **(d)** ergibt in Kombination mit dem unten stehenden Prompt Bildvariationen wie rechts gezeigt **(e)**. Siehe Video 3.*
***Prompt:** »fairytale nymph queen :: gorgeous woman, peaceful, innocent, tranquil, long flowing hair, liquid rococo fractal, James Jean, Rutkowski, fluid acrylic, elegant gradients, intricate, octane render, depth, Kaluta, detailed eyes, incredibly detailed, hyperrealistic, pastel colors, Artgerm, WLOP, fractal, 8k octane, style of Anna Dittmann, digital art --ar 9:16 --s 255 --v5« (Letztere sind Parameter für Bildformat, Stilisierung und die Midjourney-Version).*

*Aus dem Foto unten **(f)** wurden mit der »Describe«-Funktion von Midjourney der folgende Prompt und damit die nebenstehenden Bilder **(g)** und das Aufmacherbild generiert.*
Prompt: *»Woman with golden eye mask, in the style of mixes realistic and fantastical elements, exquisite clothing detail, dark gold and light crimson, intricate embellishments, romantic emotion, strong facial expression, intricately sculpted« (og)*

Foto: Nejron Photo – Adobe Stock



BILDGENERIERUNG MIT STABLE DIFFUSION

Bild-KI kostenlos auf dem eigenen Rechner nutzen

Für die Generierung von Bildern mit Midjourney und DALL-E müssen Sie regelmäßig Geld ausgeben. Stable Diffusion benötigt dagegen nur einen Computer, der die Hardwarevoraussetzungen erfüllt. Olaf Giermann zeigt, wie Sie die kostenlose Open-Source installieren und nutzen können.

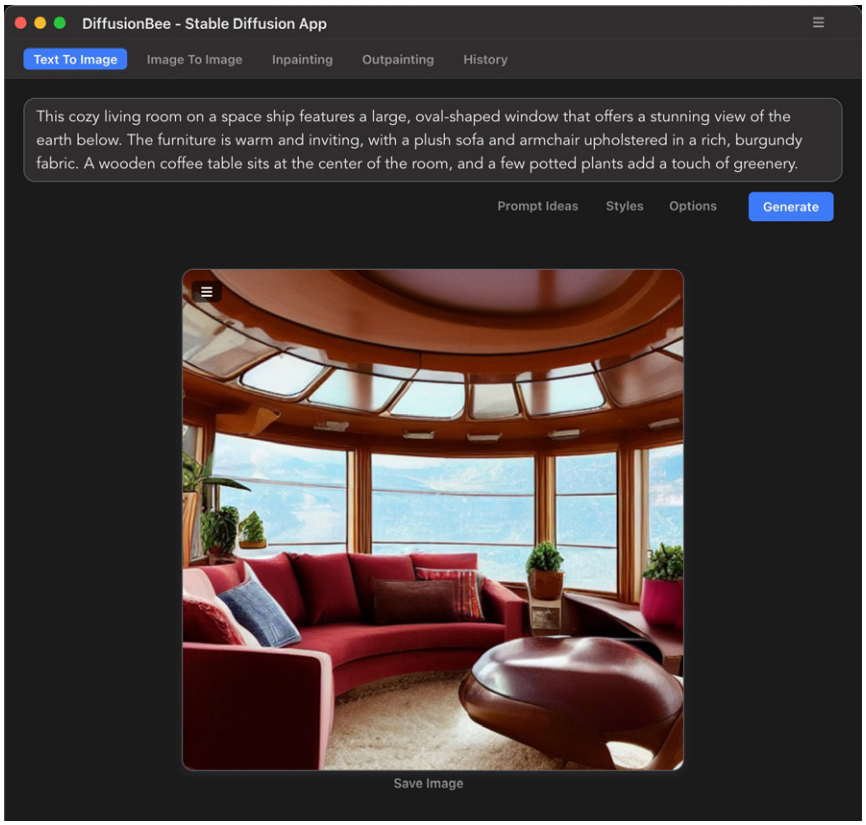
Stable Diffusion ist ein Open-Source-KI-Modell, das Bilder auf der Grundlage von Textbeschreibungen erzeugt. Eine Reihe von Diensten basiert auf diesem Modell, etwa Websites wie Nightcafe Studio, Apps wie das für Social-Media-Avatare gehypte Lensa oder Photoshop-Plug-ins wie Flying Dog oder Alpaca. Ihnen gemeinsam ist, dass die Berechnung online auf Servern erfolgt. Das hat den Vorteil, dass Sie als Nutzer keinerlei Einrichtungsaufwand haben, jedoch den Nachteil, dass Sie für diesen Komfort zur Kasse gebeten werden, sobald die üblichen Frei-Credits verbraucht sind. Verständlich, denn Strom und Rechenleistung sind teuer. Sie können Stable Diffusion aber auch herunterladen und auf dem eigenen Rechner betreiben, wenn dieser die Systemvoraussetzungen erfüllt. Sie benötigen neben mindestens 12 GB RAM und 12 GB freiem Festplattenspeicher mindestens einen:

- Windows-PC mit Nvidia-Grafikkarte mit 4, besser 8 GB VRAM oder
- Linux-Rechner mit entweder Nvidia- oder AMD-Grafikkarte oder
- Mac, ideal mit M1- oder M2-Chip.

Aufgrund der deutlich höheren Berechnungsgeschwindigkeit ist der Einsatz einer Nvidia-GPU empfehlenswert. Da kann auch ein ansonsten extrem schneller Mac Studio Ultra nicht mithalten.

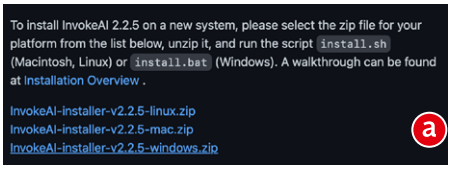
Für die Nutzung von Stable Diffusion benötigen Sie eine Bedienoberfläche, für die es verschiedene Alternativen gibt. Am einfachsten haben es hier die Anwender von Macs, denn die App **DiffusionBee** (*www.diffusionbee.com*) laden Sie einfach herunter, starten sie, und alle benötigten Zusatzmodule werden automatisch geladen und installiert. Für PC-Nutzer empfiehlt sich die Web-UI **Automatic1111**. Diese wird häufig aktualisiert und unterstützt sämtliche Funktionen von Stable Diffusion (*www.docma.info/22723*).

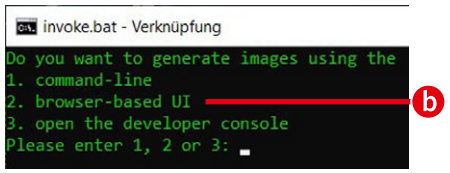
Mehr Komfort bei der Installation und Nutzung bietet **InvokeAI**, das für alle Betriebssysteme verfügbar ist und alle notwendigen Dateien automatisch herunterlädt.



Auf Macs empfiehlt sich die App DiffusionBee aufgrund der Einklick-Installation und der einfachen Oberfläche. InvokeAI bietet dagegen mehr Möglichkeiten.

Bei der Installation von InvokeAI oder Automatic1111 kann man sich beinahe wie ein Hacker der Matrix fühlen.





Nach dem Start der invoke.bat/sh drücken Sie „2“ und Enter, um InvokeAI zu starten.

Installation von InvokeAI

Einen Überblick der Installation finden Sie auf *www.docma.info/22718*. Englisch-Kenntnisse sind nötig, um den einzelnen Anweisungen auf dem Bildschirm folgen zu können. Die wichtigsten Schritte:

1. Installieren Sie Python

Sie benötigen Version 3.9.1 oder höher. Python 3.11 wird von InvokeAI nicht empfohlen. Eine Anleitung für Ihr Betriebssystem finden Sie unter *www.docma.info/22719*

2. Laden Sie den InvokeAI-Installer

Laden Sie auf *www.docma.info/22724* die aktuelle ZIP-Datei-Version für Ihr Betriebssystem **(a)** und entpacken Sie sie in einen Ordner auf der Festplatte.

3. Starten Sie die Installation und … warten Sie geduldig

Unter Windows doppelklicken Sie zuallererst auf „WinLongPathsEnabled.reg“ und dann auf „install.bat“, auf Apple- und Linux-Computern führen Sie einen Doppelklick auf „install.sh“ aus. Folgen Sie den Anweisungen auf dem Bildschirm, um den Installationspfad festzulegen und die gewünschten CKPT-Modelle auszuwählen (im Zweifel nehmen Sie „recommended“). Anschließend werden mehrere Gigabyte an Dateien aus dem Netz geladen.

4. Starten Sie InvokeAi

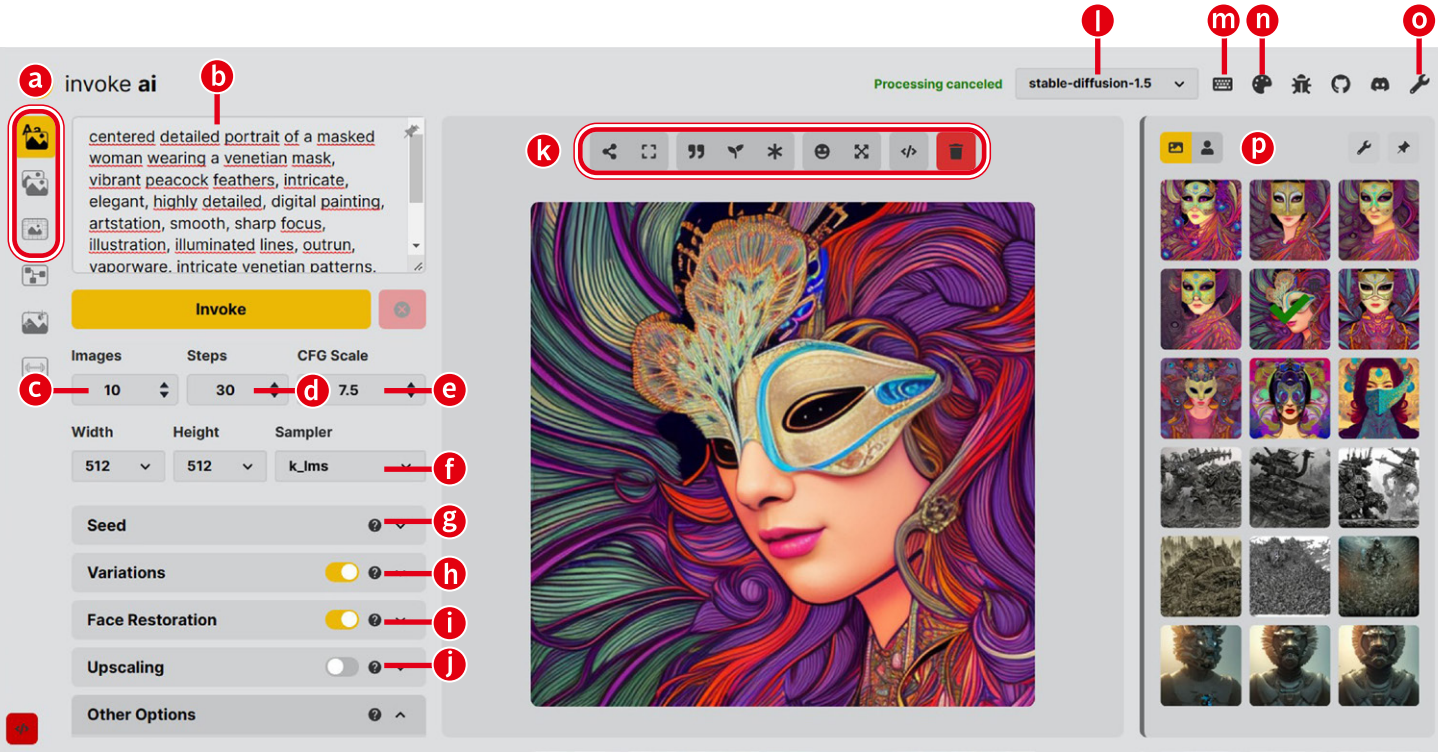
Doppelklicken Sie dazu im Installationsverzeichnis auf „invoke.bat“ (Win) beziehungsweise „invoke.sh“ (macOS/Linux). Wählen Sie im sich öffnenden Kommandozeilen-Fenster Option 2 **(b)** und drücken Sie Enter. Das Fenster darf während der gesamten Sitzung nicht geschlossen werden. Geben Sie die Web-Adresse „*http://localhost:9090*“ in einem Internetbrowser ein, um die InvokeAI-UI zu öffnen. Einen Rundgang durch die Oberfläche finden Sie auf der nächsten Seite. ▶

KI-BEGRIFFE

Über einige Begriffe werden Sie bei der Beschäftigung mit generativen KI immer wieder stolpern. Das bedeuten sie:

- Python ist eine verbreitete, plattformübergreifend einsetzbare Programmiersprache, von der viele KI-Oberflächen Gebrauch machen.
- Bei GitHub handelt es sich um einen Online-Dienst für den Austausch und die Versionsverwaltung von Software.
- CKPT steht kurz für „Checkpoint“. Sie enthalten die durch Machine Learning oder Kombination mit anderen CKPT-Dateien entstandenen KI-Modelle.
- Hugging Face ist eine Plattform, die es Usern erlaubt, Machine-Learning-Modelle und Datensätze auszutauschen.
- Inpainting ist mit der inhaltsbasierten Retusche von Photoshop vergleichbar, aber textbasiert.
- Outpainting erlaubt das textbasierte Erweitern eines Bildes über seine Grenzen hinaus.

InvokeAI im Überblick



In der **Seitenleiste (a)** wechseln Sie zwischen »*Text to image*« (der Standardmodus), »*Image to image*« (Bild-zu-Bild-Variationen) und »*Unified Canvas*«. Letzter bietet wahrscheinlich aktuell den besten Arbeitsbereich für das In- und Outpainting aller KI-Tools.

Im **Eingabefeld (b)** tragen Sie Ihren Text-Prompt ein. Negative Prompts zum Ausschluss von Begriffen bei der Bildgenerierung fassen Sie in eckige Klammern. Klicken Sie darunter auf »*Invoke*«, um den Prozess zu starten, und gegebenenfalls auf die „x"-Schaltfläche daneben, um ihn vorzeitig abzubrechen.

Über den Wert für **Images (c)** legen Sie fest, wie viele Bilder in einem Durchgang berechnet werden sollen. Sie müssen also nicht zehn Mal auf »*Invoke*« klicken, um 10 Bildvarianten für einen einzelnen Prompt zu erzeugen.

Stable Diffusion erzeugt ein Bild, indem es mit einer Leinwand voller Rauschen beginnt und dieses schrittweise „zu einem Bild" entrauscht. Der Parameter **Steps (d)** steuert die Anzahl dieser Entrauschungsschritte. Normalerweise ist eine höhere Zahl besser, es dauert aber länger und kann bei hohen Werten das Ergebnis auch verschlechtern. Bleiben Sie am besten zunächst bei den Standardwerten. Der „Classifier Free Guidance", kurz **CFG-Scale (e)**, legt fest, wieviel Freiheit Sie Stable Diffusion beim Generieren des Bildes lassen möchten. Bei einem Wert von null wird der Prompt völlig ignoriert und irgendein Bild erzeugt. Der voreingestellte Wert ist eine gute Balance zwischen Prompt-Befolgung und Variation.

Die **Sampling-Methoden (f)** unterscheiden sich darin, wie beim schrittweisen „Entrauschen" jeweils der nächste Schritt in der Bilderzeugung berechnet wird. Dementsprechend erhalten Sie mit jeder Methode unterschiedliche Bildergebnisse bei gleichem Prompt. Empfehlenswert für den Einsteig ist der Sampler DDIM, da er schnell funktioniert und mit nur 10 Schritten ansehnliche Ergebnisse erzeugt. So macht das Experimentieren und Verfeinern mehr Spaß.

Der **Seed (g)** steuert das Anfangsrauschen und wird standardmäßig bei jeder Generierung geändert, so dass Sie jedes Mal ein anderes Bild erhalten. Hat der Seed einen konkreten Wert, erhalten Sie bei jedem Prompt bei auch sonst identischen Einstellungen dasselbe Bildergebnis. Der **Variations**-Wert **(h)** ist eine Möglichkeit in InvokeAI, die Streuung des Seeds feiner zu steuern.

Die **Face Restoration (i)** versucht, Gesichter in den Bildern zu erkennen und typische Artefakte wie verbogene Augen und Nasen zu korrigieren. Höhere Werte sorgen für eine kräftigere Korrektur und sollten bei jedem Porträt zum Einsatz kommen. Das geht auch nachträglich, indem Sie in der Symbolleiste **(k)** über dem generierten Bild auf den Smiley und dann dort auf »*Restore Faces*« klicken.

Über die Einstellungen im **Upscaling-Panel (j)** vergrößern Sie die standardmäßig mit einer Größe von 512 × 512 Pixeln generierten Bilder auf das Zwei- oder Vierfache.

In der **Symbolleiste (k)** finden Sie verschiedene Optionen, um das im Imagebrowser **(p)** ausgewählte Bild an die anderen Module der Seitenleiste **(a)** zu übergeben. Sie können die Invoke-Einstellungen verbergen, den Prompt, den Seed oder beides aus einem angezeigten Bild in die Einstellungen laden, nachträglich die Gesichter verbessern oder das Bild vergrößern, Informationen einblenden oder das Bild löschen.

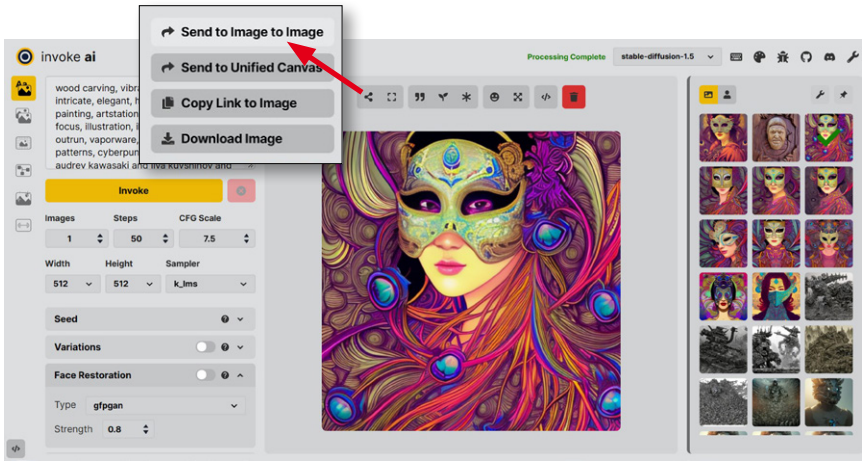
Installierte Checkpoint-**Modelle** laden Sie über ein Dropdown-Menü **(l)**. Mit den Schaltflächen daneben blenden Sie die praktischen **Tastaturkürzel (m)** von InvokeAI ein oder schalten zwischen verschiedenen Helligkeiten der **Bedienoberfläche (n)** um. Weitergehende Einstellungen finden Sie im **Settingsmenü (o)**. Eine Übersicht aller generierten Bilder sehen Sie im **Image-Browser (p)**.

Image to image und Outpainting

01 Bild weiterreichen

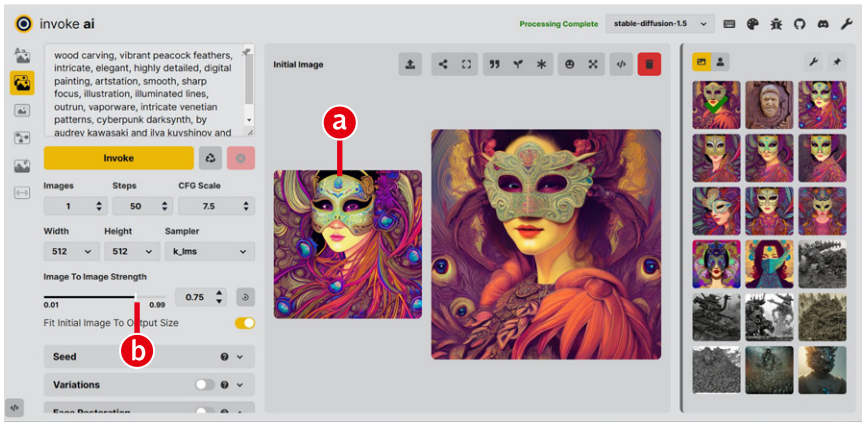
Haben Sie ein Bild erzeugt, das grundsätzlich in die Ihnen vorschwebende Richtung geht, aber stilistisch noch nicht passt? Dann geben Sie es in den *»Image to image«*-Bereich von InvokeAi. Klicken Sie dafür auf die erste Schaltfläche in der Symbolleiste und auf *»Send to image to image«*.

***TIPP**: Im »Image to image«-Bereich können Sie auch eigene Fotos hochladen und dann per Prompt variieren.*



02 Prompt verfeinern

Variieren Sie den Prompt, bis Ihnen das Ergebnis gefällt. InvokeAI ändert dementsprechend das Ausgangsbild **(a)**. Wieviel Freiheiten die KI dabei hat, legen Sie über den Parameter *»Image to image strength«* fest **(b)**. Höhere Werte erlauben dabei eine größere Abweichung vom Originalbild. Der Standard von 0,75 ist ein guter Kompromiss für die meisten Anwendungsfälle.



03 Outpainting

Über dasselbe Menü wie im ersten Schritt reichen Sie das Ergebnis aus Schritt 2 für das In- oder Outpainting mit *»Send to unified canvas«* weiter. Hier können Sie mit der Pipette **(b)** Farben aus dem Bild aufnehmen und mit dem Pinsel **(a)** Linien und Formen ins Bild malen. Mit dem Verschieben-Werkzeug **(c)** bewegen Sie den um das Bild liegenden Rechteckrahmen nach außen, so dass noch etwas Überlappung besteht. Auf diese Weise markierte Bereiche können Sie dann mit einem Klick auf *»Invoke«* mit dem Inhalt des Prompts füllen lassen und so das Bild nahtlos erweitern **(d)**. Dieser Vorgang lässt sich beliebig oft – mit demselben oder einem angepassten Prompt – wiederholen. Die aufgefüllten Bereiche zwischen den vorgenommenen Pinselstrichen **(e)** passen hier nicht so richtig ins Bild. Glücklicherweise haben Sie nach jedem *»Invoke«* die Möglichkeit zu entscheiden, ob Sie ein Ergebnis übernehmen oder verwerfen möchten **(f)**. Mit dem Radierwerkzeug löschen Sie gemalte Strukturen einfach vor dem nächsten Versuch. ■