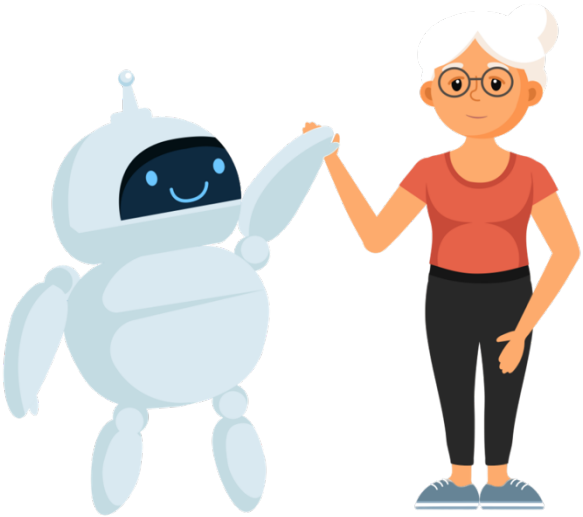
Universität Bremen

Fakultät 3: Mathematik und Informatik

Ein Erklärvideo zu dem Thema „Bildgenerierende Künstliche Intelligenz“ für Studierende des ersten Semesters

Bachelor-Thesis zur Erlangung des akademischen Grades Bachelor of Science (B.Sc.) im Studienfach Digitale Medien



Caroline Dietz
Matrikelnr.: 606743
Abgabedatum: 27.05.2025

Erstprüfer: Prof. Dr. Ing. Udo Frese
Zweitprüfer: Prof. Dr. Thomas Dieter Barkowsky
Betreuer: Prof. Dr. Ing. Udo Frese

Universität Bremen

Nachname Dietz Matrikelnr. 6067432

Vorname Caroline

A) Eigenständigkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Teile meiner Arbeit, die wortwörtlich oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht. Gleiches gilt auch für Zeichnungen, Skizzen, bildliche Darstellungen sowie für Quellen aus dem Internet, dazu zählen auch KI-basierte Anwendungen oder Werkzeuge. Die Arbeit wurde in gleicher oder ähnlicher Form noch nicht als Prüfungsleistung eingereicht. Die elektronische Fassung der Arbeit stimmt mit der gedruckten Version überein. Mir ist bewusst, dass wahrheitswidrige Angaben als Täuschung behandelt werden.

☐ Ich habe KI-basierte Anwendungen und/oder Werkzeuge genutzt und diese im Anhang "Nutzung KI basierte Anwendungen" dokumentiert.

B) Erklärung zur Veröffentlichung von Bachelor- oder Masterarbeiten

Die Abschlussarbeit wird zwei Jahre nach Studienabschluss dem Archiv der Universität Bremen zur dauerhaften Archivierung angeboten. Archiviert werden:

1) Masterarbeiten mit lokalem oder regionalem Bezug sowie pro Studienfach und Studienjahr 10 % aller Abschlussarbeiten
2) Bachelorarbeiten des jeweils ersten und letzten Bachelorabschlusses pro Studienfach und Jahr.

☐ Ich bin damit einverstanden, dass meine Abschlussarbeit im Universitätsarchiv für wissenschaftliche Zwecke von Dritten eingesehen werden darf.

☐ Ich bin damit einverstanden, dass meine Abschlussarbeit nach 30 Jahren (gem. §7 Abs. 2 BremArchivG) im Universitätsarchiv für wissenschaftliche Zwecke von Dritten eingesehen werden darf.

☐ Ich bin nicht damit einverstanden, dass meine Abschlussarbeit im Universitätsarchiv für wissenschaftliche Zwecke von Dritten eingesehen werden darf.

C) Einverständniserklärung zur Überprüfung der elektronischen Fassung der Bachelorarbeit / Masterarbeit durch Plagiatssoftware

Eingereichte Arbeiten können nach § 18 des Allgemeinen Teil der Bachelor- bzw. der Master-prüfungsordnungen der Universität Bremen mit qualifizierter Software auf Plagiatsvorwürfe untersucht werden.
Zum Zweck der Überprüfung auf Plagiate erfolgt das Hochladen auf den Server der von der Universität Bremen aktuell genutzten Plagiatssoftware.

☐ Ich bin damit einverstanden, dass die von mir vorgelegte und verfasste Arbeit zum oben genannten Zweck dauerhaft auf dem externen Server der aktuell von der Universität Bremen genutzten Plagiatssoftware, in einer institutionseigenen Bibliothek (Zugriff nur durch die Universität Bremen), gespeichert wird.

☐ Ich bin nicht damit einverstanden, dass die von mir vorgelegte und verfasste Arbeit zum o.g. Zweck dauerhaft auf dem externen Server der aktuell von der Universität Bremen genutzten Plagiatssoftware, in einer institutionseigenen Bibliothek (Zugriff nur durch die Universität Bremen), gespeichert wird.

Datum

Datum Unterschrift

Zusammenfassung

Die vorliegende Bachelorthesis widmet sich der Frage, wie bildgenerierende Künstliche Intelligenz, insbesondere auf Basis von Diffusionsmodellen, verständlich und zielgruppengerecht erklärt werden kann. Ziel war die Konzeption, Gestaltung und Umsetzung eines animierten Erklärvideos, das den komplexen technischen Prozess der KI-Bildgenerierung didaktisch reduziert und für Studienanfänger des Studiengangs „Digitale Medien“ an der Universität Bremen aufbereitet. Im theoretischen Teil wurden grundlegende Konzepte der Künstlichen Intelligenz sowie der Bildgenerierung mithilfe des Diffusionsprozesses dargestellt und relevante Prinzipien für Erklärvideos erläutert. Aufbauend darauf erfolgte die Entwicklung eines Video-Konzepts, das sich dramaturgisch am Modell der „Heldenreise“ orientiert. Die Wirkung des Videos wurde abschließend in einer Evaluation untersucht, um Rückschlüsse auf Verständlichkeit, gestalterische Wirkung und didaktische Wirksamkeit zu ziehen.

Abstract

This bachelor's thesis is dedicated to the question of how image-generating artificial intelligence, in particular on the basis of diffusion models, can be explained in an understandable and target group-oriented way. The goal was the conception, design and implementation of an animated explanatory video that didactically reduces the complex technical process of AI image generation and prepares it for first-year students of the “Digital Media” course at the University of Bremen. In the theoretical part, basic concepts of artificial intelligence and image generation using the diffusion process were presented and relevant principles for explanatory videos were explained. Building on this, a video concept was developed that is dramaturgically based on the “hero's journey” model. Finally, the impact of the video was examined in an evaluation in order to draw conclusions about comprehensibility, creative impact and didactic effectiveness.

Inhaltsverzeichnis

I. Abbildungsverzeichnis

II. Abkürzungsverzeichnis

KL	**Künstliche Intelligenz**
ML	**Maschinelles Lernen**
NLP	**Natural Language Processing**
NIP	**Natural Image Processing**
GANs	**Generative-Adversarial-Networks**
dt.	**deutsch**
engl.	**englisch**

1. Einleitung

Etwa 35 Tage malte der spanische Künstler Pablo Picasso an seinem Werk **“**Guernica” (siehe Abb.1). Darauf zu sehen sind drei Tiere (ein Stier, ein verletztes Pferd, ein Vogel) und fünf menschliche Figuren, darunter eine trauernde Mutter, eine Frau mit einer Lampe und eine sterbende Person mit erhobenen Armen. Im Auftrag der spanischen Republik wurde das Bild zum Anlass des Luftangriffs auf die baskische Stadt Guernica angefertigt. Es ist ein Ereignis, das Pablo Picasso ausschließlich durch Pressefotos kannte.[1]

Abb. 1: "Guernica" v. Pablo Picasso (Quelle: Museu National Centro de Arte Reina Sofia)

Etwa 84 Sekunden benötigte die bildgenerierende Künstliche Intelligenz *DALL-E 3*, um das nachfolgenden Bild mit dem Text-Prompt: “Male ein Gemälde der 3 Bremer Stadtmusikanten im Stil von Pablo Picassos “Guernica” in schwarz weiss” zu generieren (siehe Abb. 2).

Kunstwerke, in denen Stunden, Tage, Monate, gar jahrelange Arbeit steckt, können heutzutage in wenigen Sekunden von Künstlichen Intelligenzen mit Hilfe einer simplen Texteingabe generiert werden.

Diese Entwicklung verändert die gesamte Kreativbranche – doch wie funktioniert die Generierung von Bildern mit Hilfe von Künstlicher Intelligenz? Was steckt hinter den Begriffen *Diffusion, neuronale Netze* und *multidimensionaler Raum*?

Abb. 2: Bremer Stadtmusikanten (Quelle: generiert mit DALL-E 3)

All diese Begriffe sind sehr abstrakt und um den Einstieg in das Thema zu erleichtern, wurde im Rahmen dieser Arbeit ein animiertes Erklärvideo produziert, das die Funktionsweise der bildgenerierenden KI einfach, verständlich und anschaulich erklärt.

Dabei richtet sich das Video speziell an Studierende des ersten Semesters des Studiengangs “Digitale Medien”, da das Thema der bildgenerierenden KI ein Teil des Lehrplans im Modul “Medieninformatik I” ist.

Ziel dieser Arbeit ist es, die Konzeption, Umsetzung und didaktische Gestaltung dieses Videos zu dokumentieren und zu reflektieren.

2. Grundlagen

Um im Verlauf dieser Arbeit auf die bildgenerierenden Künstlichen Intelligenzen eingehen zu können, Bedarf es zuvor dem Verständnis für die Künstliche Intelligenz (KI) im Allgemeinen. Im folgenden Teil wird der Begriff definiert, die Historie kurz erläutert und daraufhin auch die Funktionsweise von KI erklärt.

2.1 Definitionsversuch

Abercrombie definiert den Begriff der Künstlichen Intelligenz wie folgt:

> *"KI sind menschenähnliche Intelligenzsysteme."*[2]

Da stellt sich zunächst die Frage, was unter Intelligenz bzw. unter *natürlicher* Intelligenz zu verstehen ist. Schon hier wird deutlich, dass es keine einheitliche Definition gibt: Intelligenz kann sich auf verschiedene Bereiche beziehen, etwa auf logisch-mathematische Fähigkeiten, sprachliche Kompetenz oder emotionale Intelligenz und vieles mehr.

Versucht man, diese unterschiedlichen Dimensionen zusammenzufassen, lässt sich Intelligenz allgemein als die Fähigkeit beschreiben, Wissen zu erwerben, anzuwenden und daraus neues Wissen zu generieren, um Probleme zu lösen.

> *"Intelligenz ist ein vielschichtiges Konzept, das die Fähigkeit eines Individuums beschreibt, komplexe Informationen zu verstehen, zu verarbeiten und daraus angemessene Schlussfolgerungen zu ziehen. Diese Fähigkeit umfasst kognitive Prozesse wie Wahrnehmung, Lernen, Erinnern, Problemlösung, kritisches Denken und Entscheidungsfindung."* [3]

Künstliche Intelligenz (KI) lässt sich als eine Nachbildung dieser natürlichen Intelligenz verstehen: Maschinen sind in der Lage, Wissen aufzubauen, es praktisch einzusetzen, eigenständig Entscheidungen zu treffen und komplexe Aufgaben zu bewältigen.[4]

[2] Vgl. C. Abercrombie, 2022, S. 27
[3] Vgl. NeuroNation, im Internet
[4] Vgl. U. Engelke, B. Engelke, 2024, S.11-12

2.2: Die historische Entwicklung

Die Geschichte der Künstlichen Intelligenz ist geprägt von Visionären und technologischen Durchbrüchen. Ein historischer Überblick zeigt, wie sich Konzepte und Systeme über Jahrzehnte hinweg zu den heutigen Anwendungen weiterentwickelt haben.

Den Grundstein für die Künstliche Intelligenz legte 1936 der britische Mathematiker Alan Turing als er seine Theorie zu der “Turing Maschine” veröffentlichte. Darin bewies er, dass Rechenmaschinen in der Lage sind kognitive Prozesse auszuführen – unter der Voraussetzung, dass sich die Prozesse in einzelne Schritte unterteilen und sich in einem Algorithmus darstellen lassen.[5]

Den ersten Startschuss für Künstliche Intelligenz setzte Alan Turing 1950 mit der Veröffentlichung seines Papers “Computing Machinery and Intelligence". Hier warf der britische Mathematiker erstmals die Frage auf, ob Computer *denken* können.
Um dies zu herauszufinden, entwickelte er einen Test, der später als der *Turing Test* bekannt wurde. Bei diesem Test führt eine Person über ein Computerprogramm ein Gespräch mit zwei Gesprächspartnern. Bei einem der Beiden handelt es sich um einen Computer, bei dem anderen um einen Menschen. Die Frage ist, ob während eines Gespräches erkannt werden kann, wer von den Beiden der Computer und wer der Mensch ist (siehe Abb. 3). [6]

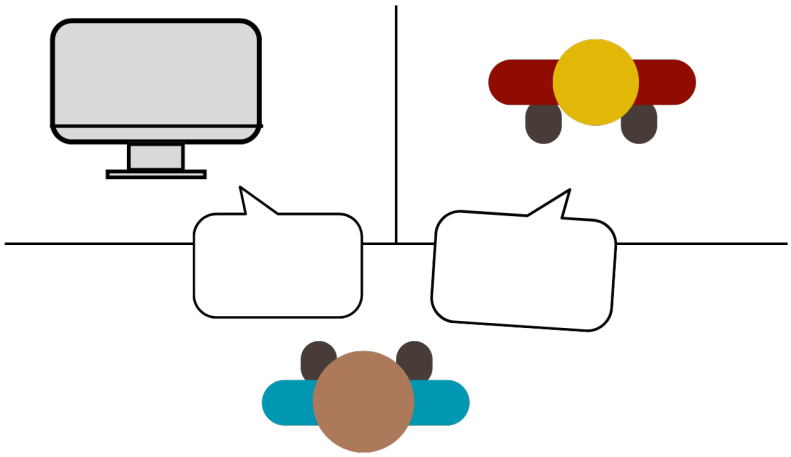


Abb. 3: Der Turing-Test (Quelle: eigene Darstellung)

[5] Vgl. Schmidt, Robominds, im Internet
[6] Vgl. Mebis, im Internet

Der *Begriff* Künstliche Intelligenz, wie wir ihn heute kennen, wurde 1956 von dem Informatiker John McCarthy auf einer Konferenz in New Hampshire geprägt. Dort führte er zusammen mit weiteren Kollegen das “Summer research project on artificial intelligence“ durch.[7]

1958 entwickelte der amerikanische Psychologe Frank Rosenblatt das erste neuronale Netz, das sogenannte “Perzeptron” und legte damit den Grundstein für das maschinelle Lernen.

Mit dem Programm “ELIZA” wurde 1966 der erste Chatbot geboren. Das Wort *Chatbot* ist eine Fusion der Worte *Chat* und *Robot.* Der MIT-Professor Joseph Weizenbaum entwickelte somit ein Programm, das ein Gespräch simulieren kann.[8] Dass das Programm funktioniert, belegte offenbar auch der Turing Test. Benutzer gaben an, dass sie nicht bemerkten, dass sie mit einem Maschine sprachen. ELIZA wurde das Vorbild für viele Chatbots, die wir heute kennen. [9]

In den 1970er Jahren brach die Forschung rund um das Thema KI ein: die Technik stieß an ihre Grenzen, das Interesse sank, die Finanzierungen wurden zurückgefahren und somit führte es zum ersten *KI-Winter. [10]*

Terrence J. Sejnowski und Charles Rosenberg bringen 1986 erstmals einen Computer zum Sprechen. Das von ihnen entwickelte Programm „NETtalk“ ist in der Lage, Wörter zu lesen, auszusprechen und auch anzuwenden.[11]

Gegen Ende der Achtziger Jahre erlitt die Branche wiederum den zweiten KI-Winter: hohe Kosten, begrenzte Leistung und enttäuschte Erwartungen führten erneut zu massiven Investitionsrückgängen.

In den 1990er Jahren lag ein Schwerpunkt der KI-Forschung auf der Sprachverarbeitung und dem maschinellen Lernen[12] und so besiegte die KI-Schachmaschine „Deep Blue”, die von IBM entwickelt wurde, den amtierenden Schachweltmeister Garri Kasparow im Jahr 1997.[13]

[7] Vgl. Santner, 2024, S. 76
[8] Vgl. Mebis, im Internet
[9] Vgl. Santner, 2024, S. 76
[10]Vgl. U. Engelke, B. Engelke, 2024, S. 13
[11] Vgl. Mebis, im Internet
[12] Vgl. U. Engelke, B. Engelke, 2024, S. 14
[13] Vgl. Mebis, im Internet

Ab den 2000er Jahren beschleunigte sich die Entwicklung deutlich, da leistungsfähigere Technologien verfügbar wurden und die Menge an digital gespeicherten Daten stetig zunahm – nicht zuletzt dank der Erfindung des Internets. Diese Entwicklungen leiteten schließlich die Ära von *Big Data* ein.[14]

Ein herausragendes Beispiel für diese neue Ära ist *ImagNet:* eine der weltweit größten Bilddatenbanken, initiiert 2009 von Fei-Fei Li.
Fei-Fei Li machte sich zum Ziel, Maschinen das "Sehen" beizubringen und erlangte mit ImageNET einen Durchbruch im Bereich der *Computer Vision* (Bilderkennung).
Im Jahr 2020 umfasste ImageNet über 21 Millionen Bilder, die in rund 21.000 Objektklassen kategorisiert waren.[15]

Auf ImageNet folgte kurze Zeit später *AlexNet*, das "erste tief lernende neuronale Netz" (vgl. Konecny, S.89).
2012 nahm AlexNet an einer von ImageNet veranstalteten Challenge teil, der "ImageNet Large Scale Visual Recognition Challenge" (kurz: ILSVRC) und belegte den ersten Platz.
Ziel dieses Wettbewerbs ist es, ein System zu entwickeln, das möglichst viele Bilder korrekt klassifiziert – bei möglichst geringer Fehlerquote.
AlexNet erreichte eine Fehlerquote von *nur* 15,3 Prozent, während der bis dahin beste Wert bei rund 25 Prozent lag. Dieser Durchbruch gilt als Wendepunkt in der Geschichte der Computer Vision und wird oft als der Moment bezeichnet, in dem Maschinen erstmals ein „Sehvermögen" zugesprochen wurde.[16]

2014 fand die Einführung von Generative Adversarial Networks (GANs) durch Ian Goodfellow statt, die es Maschinen ermöglichen, Bilder zu generieren, die fast nicht mehr von echten Bildern zu unterscheiden sind.[17]

Nur ein Jahr später wurde das Unternehmen gegründet, das später eines der einflussreichsten KI-Systeme entwickeln sollte: 2015 entstand in den USA die Forschungsorganisation *OpenAI*. Von Beginn an formulierten sie ein Ziel: Künstliche Intelligenz sollte nicht wenigen, sondern der gesamten Menschheit zugutekommen.[18]

[14] Vgl. U. Engelke, B. Engelke, 2024, S. 14
[15] Vgl. Konecny, 2020, S. 87-88
[16] Vgl. Konecny, 2020, S. 89-90
[17] Vgl. Goodfellow et al. 2014
[18] Vgl. U. Engelke, B. Engelke, 2024, S. 15

2017 wurde die Transformer-Architektur mit dem Paper „Attention is All You Need“ vorgestellt. Entwickelt wurde es von Ashish Vaswani, einem Team bei Google Brain, und einer Gruppe von der University of Toronto. Die Transformermodelle können gezielt relevante Teile von Eingaben analysieren und sind deshalb besonders leistungsfähig bei Textverarbeitung und -generierung.[19]

Im selben Jahr veröffentlichte OpenAI dann die erste Version ihres Chabots *ChatGPT*. Dabei steht *GPT* für “Generative Pretrained Transformer*”* – das bedeutet, dass die KI zuvor mit Texten trainiert wurde, selbst Texte verfassen kann und auf einem Transformer-Modell basiert.
In den Jahren 2019 und 2021 veröffentlichte OpenAI neue und verbesserte Versionen von ChatGPT, die bis zum November 2022 nur für ein Fachpublikum zugänglich waren. Nach der Veröffentlichung von ChatGPT-3 für die breite Masse, erreichte es in den ersten fünf Tagen bereits über eine Millionen Nutzer.[20]
Heutzutage (Stand Februar 2025) hat ChatGPT über 400 Millionen Nutzer weltweit und ist somit die meistgenutzte Künstliche Intelligenz weltweit (Siehe Abbildung 4).[21]

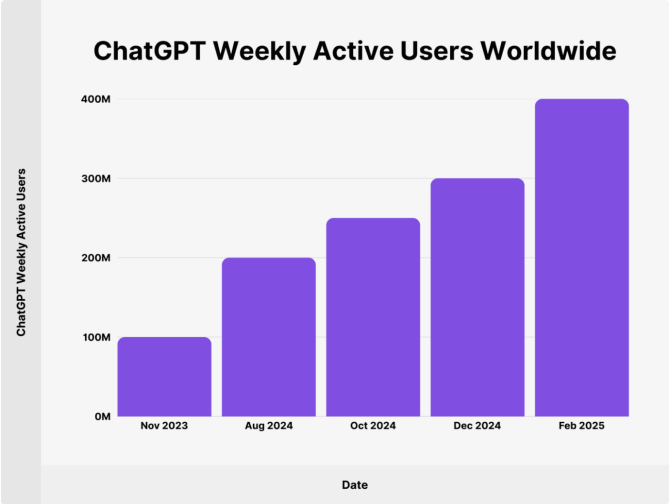


Abb. 4: Statistik der ChatGPT Nutzer 2023-2025 (Quelle: Backlinko)

Im Jahr 2022 wurde Stable Diffusion von der CompVis-Gruppe an der LMU München zusammen mit Stability AI und Runway entwickelt. Es ist ein Text-zu-Bild-Diffusionsmodell, das realistische Bilder aus einfachen Anweisungen erzeugt.

[19] Vgl. IBM, im Internet
[20] Vgl. Engelke und Engelke, 2024, S. 32-33
[21] Vgl. Backlinko, Im Internet

Im Gegensatz zu vielen anderen Modellen ist Stable Diffusion Open Source – das heißt, der Quellcode ist öffentlich zugänglich, kann angepasst und direkt auf den Rechnern der Nutzer ausgeführt werden. Dies löste eine Welle an Innovationen aus: Weltweit begannen Entwickler, eigene Bildgeneratoren zu entwerfen, die auf der Architektur von Stable Diffusion basieren.[22]

Was einst als theoretisches Konzept begann, hat sich über Jahrzehnte hinweg zu einer Technologie entwickelt, die heute Bilder malt, Texte schreibt und ganze Konversationen führt.

Um zu verstehen, wie solche Systeme funktionieren, braucht es jedoch mehr als nur einen historischen Überblick: Im nächsten Kapitel werden daher die zentralen Grundlagen erläutert, auf denen heutige KI-Systeme basieren.

2.3 Arten von KI

Künstliche Intelligenz lässt sich je nach Leistungsumfang und Zielsetzung in verschiedene Typen gliedern. Die folgende Übersicht zeigt die wichtigsten Kategorien, die sich in ihrem Grad an Selbstständigkeit und Komplexität unterscheiden.

2.3.1 Starke KI (Artificial General Intelligence)

Eine starke KI verfügt über ein Bewusstsein, ein Empfindungsvermögen und ein Selbstbewusstsein. Somit ist die KI aufgrund des Selbstbewusstseins in der Lage ihre Umwelt wahrzunehmen und diese Umwelt aufgrund ihres Bewusstseins *subjektiv* wahrzunehmen und besitzt letztlich auch Fähigkeit Dinge subjektiv zu empfinden, mithilfe ihres Empfindungsvermögen.[23]

Starke Künstliche Intelligenz ist in der Lage, eigenständig Aufgaben zu erkennen, Probleme zu analysieren und selbstständig neues Wissen aus dem jeweiligen Anwendungsbereich zu erarbeiten. Im Gegensatz zur schwachen KI kann sie flexibel und kreativ auf neue Situationen reagieren und eigenständige Lösungen entwickeln – auch ohne vorgegebene Methodik.[24]

[22] Vgl. Stability.ai, 2022, im Internet
[23] Vgl. Uria-Recio, 2024, S. 641
[24] Vgl. Technische Hochschule Würzburg-Schweinfurt, im Internet

2.3.2 Schwache KI (Artificial Narrow Intelligence/weak AI)

Im Gegensatz zu der starken KI verfügt die *schwache KI* weder über ein Bewusstsein oder Selbstbewusstsein, noch über ein Empfindungsvermögen. [25]

Ihre Stärken liegen darin, konkrete Probleme zu lösen, Muster zu erkennen und mit großen Datenmengen umzugehen, z.B. durch maschinelles Lernen. Typische Einsatzbereiche sind automatisierte Prozesse, Spracherkennung, Text- und Bilderkennung sowie Navigations- oder Übersetzungssysteme.[26]

2.3.3 Die Superintelligenz

Als *Superintelligenz* bezeichnet man eine Form von Intelligenz, die der menschlichen in nahezu allen Bereichen deutlich überlegen ist, etwa in wissenschaftlicher Kreativität, allgemeiner Problemlösungskompetenz und sozialen Fähigkeiten. Wie eine solche Intelligenz technisch umgesetzt wird – ob durch digitale Systeme, vernetzte Computer oder biologische Strukturen – bleibt offen. [27]

2.4 Einsatzbereiche

Künstliche Intelligenz findet heute in zahlreichen Anwendungsfeldern Einsatz und unterstützt bei unterschiedlichsten Aufgaben. Im Folgenden werden zentrale Bereiche vorgestellt, in denen KI-Technologien besonders prägend wirken: die Verarbeitung natürlicher Sprache (Natural Language Processing), die Analyse visueller Daten (Natural Image Processing), Expertensysteme sowie Robotik.

2.4.1 Natural Language Processing

Damit natürliche menschliche Sprache verarbeitet werden kann, sind Algorithmen erforderlich, die sowohl das Verstehen als auch das Erzeugen von Sprache ermöglichen. Dieser Teilbereich der Künstlichen Intelligenz nennt sich *Natural Language Processing (NLP).* [28]

[25] Vgl. Uria-Recio, S. 641
[26] Vgl. Technische Hochschule Würzburg-Schweinfurt, im Internet
[27] Vgl. Bostrom, 1997, im Internet
[28] Vgl. Dahm, Zehnder 2023, S.23

Die nachfolgende Tabelle zeigt die verschiedenen Arten von NLP:

Input/Output	Beschreibung	Beispiel
Speech-to-Text	KI überträgt gesprochenes Wort in digitalen Text	Diktierfunktion im Smartphone
Speech-to-Speech	KI verarbeitet gesprochenes Wort und gibt gesprochenes Wort durch Natural Language Generation aus	Sprachassistenten wie Siri und Alexa
Text-to-Speech	KI verarbeitet Text und gibt diesen in Form von Sprache wieder	Vorlese-Funktion von Web-Browsern oder Smartphones (z. B. für Sehbehinderte)
Text-to-Text	KI verarbeitet digitalen Text in eine andere Form von digitalem Text	Übersetzungssoftware wie DeepL oder Chatbots wie ChatGPT

Abb.5: Arten von NLP (Quelle: Dahm und Zehnder, 2023, S.23)

2.4.2 Natural Image Processing

Ähnlich wie im Natural Language Processing (NLP) werden auch beim *Natural Image Processing* (NIP) eingehende Bilddaten zunächst in eine maschinenlesbare Form übersetzt, ihre Inhalte interpretiert und darauf basierende Handlungsentscheidungen getroffen.

NIP kommt in einer Vielzahl von Anwendungsfeldern zum Einsatz: von der Überwachung von Objekten und Personen über die biomedizinische Bildanalyse bis hin zur Kunst- und Kreativbranche.[29]

2.4.3 Expertensysteme

Ein weiteres Teilgebiet der Künstlichen Intelligenz sind die *Expertensysteme*.

Diese Systeme sind darauf ausgelegt, menschliches Expertenwissen in einem bestimmten Fachgebiet nachzuahmen. Dafür werden Regeln aufgestellt, die das Fachwissen des Experten abgeleitet werden. Des Weiteren verfügt das System über eine spezifische Entscheidungslogik, um Probleme ähnlich wie der Experte lösen zu können. [30]

[29] Vgl. Dahm, Zehnder 2023, S.24
[30] Vgl. Dahm, Zehnder 2023, S.26

2.4.4 Robotik

Künstliche Intelligenz und *Robotik* sind eng miteinander verbunden: Während die Robotik sich mit der Konstruktion und Steuerung von Maschinen befasst, sorgt KI dafür, dass diese Maschinen intelligentes Verhalten zeigen können. Durch KI-Systeme können Roboter ihre Umgebung wahrnehmen, interagieren und sich flexibel an neue Situationen anpassen. [31]

2.5 Die Funktionsweise der Künstlichen Intelligenz

Künstliche Intelligenz basiert auf einer Vielzahl von Algorithmen – also klar definierten Rechenvorschriften zur Lösung bestimmter Aufgaben. Die Funktionsweise eines Algorithmus lässt sich gut mit einem Kochrezept vergleichen: Es gibt eine Abfolge von Anweisungen, die ausgeführt werden müssen, um aus bestimmten Zutaten (den Eingabedaten) ein fertiges Gericht (die Ausgabedaten) zu erzeugen.

Ein zentrales Merkmal moderner KI-Systeme ist ihre Fähigkeit zu lernen: Auf Basis großer Datenmengen kann ein KI-Modell seine Rezepte, also seine Problemlösungsstrategien, kontinuierlich verbessern und an neue Situationen anpassen. Dieses Lernen erfolgt durch statistische Auswertung und Anpassung interner Parameter.[32]

Wie genau aber lernen diese Systeme? Um das zu verstehen, müssen wir das Prinzip der neuronalen Netze näher betrachten.

2.5.1: Neuronale Netze

Die neuronalen Netze (auch “künstliche neuronale Netze” - kurz: KNN - genannt) basieren auf Algorithmen und sollen genauso wie die Neuronen des menschlichen Gehirns miteinander kommunizieren.

Dabei besteht ein neuronales Netz aus mehreren Schichten, durch die Informationen Schritt für Schritt verarbeitet werden. Am Anfang steht die Eingabeschicht, dann folgen eine oder mehrere Zwischenschichten und ganz am Schluss folgt die Ausgabeschicht.

[31] Vgl. Dahm, Zehnder 2023, S.26
[32] Vgl. Bundeszentrale für politische Bildung, 2024, im Internet

In jeder Schicht sitzen viele kleine Neuronen, die miteinander verbunden sind: Jede Schicht bekommt ihre Infos von der vorherigen.
Dabei hat jede Verbindung ein Gewicht, das beim Lernen angepasst wird. Dieses Gewicht bestimmt, wie stark ein Signal weitergegeben wird (siehe Abb. 6).

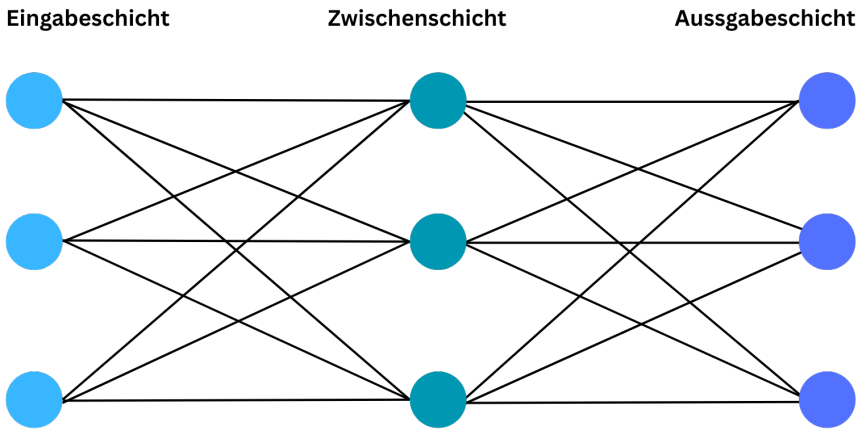


Abb. 6: Aufbau eines neuronalen Netzes (Quelle: eigene Darstellung)

Sie werden primär zur Analyse und Klassifikation von Objekten oder Signalen in Bereichen wie Sprachverarbeitung, visueller Erkennung und Regelungstechnik verwendet.[33]

2.5.2: Maschinelles Lernen

Maschinelles Lernen (ML) ist ein zentraler Bestandteil der Künstlichen Intelligenz. Dabei geht es darum, dass ein System eigenständig Muster in Daten erkennt und auf dieser Grundlage Wahrscheinlichkeiten berechnet sowie Entscheidungen trifft.[34]

Dabei wird unter drei verschiedenen Arten des Lernens unterschieden:

- Überwachtes Lernen
- Unüberwachtes Lernen
- Bestärkendes Lernen

[33] Vgl. Mathworks, im Internet
[34] Vgl. Bundeszentrale für politische Bildung, 2024, im Internet

2.5.2.1 Überwachtes Lernen (supervised learning)

Beim Überwachten Lernen wird dem neuronalen Netz während des Trainingsprozesses zu jeder Aufgabe die entsprechende Lösung vorgegeben.

Jaromir Konecny vergleicht es mit dem Besuch in einem Tierpark: eine Mutter erklärt ihrem Sohn die verschiedenen Tiere: “Das ist ein Panda, und das ist ein Löwe.” (vgl. Konecny, S. 59). Im Gegensatz zu dem Computer reicht es dem Kind aus, wenn man ihm das Tier zweimal erklärt. Ein neuronales Netz braucht Zehntausende Daten, bis es verstanden hat, was einen Löwen von einem Panda unterscheidet. [35]

2.5.2.2 Unüberwachtes Lernen (unsupervised learning)

Beim Unüberwachten Lernen lernt das neuronale Netz – wie der Name bereits verrät – ohne die Hilfe eines Menschen. Hier arbeitet das Netz mit dem sog. *Clustering*. Bekommt es z.B. einen Datensatz mit Bildern von verschiedenen Obstsorten und die Aufgabe, diese voneinander zu unterscheiden, würde das neuronale Netz damit beginnen, die Sorten in *Cluster* einzuteilen. Da die Bilder nicht vorher von einem Menschen gekennzeichnet wurden, versucht das neuronale Netz die Cluster selbst zu erstellen anhand von Mustererkennung.[36]

2.5.2.3 Bestärkendes Lernen (reinforcement learning)

Beim bestärkenden Lernen entwickelt das neuronale Netz sein Verhalten, indem es mit der Umgebung interagiert und Rückmeldungen in Form von Belohnung oder Bestrafung erhält – abhängig davon, wie erfolgreich seine Aktionen sind.[37]

2.5.3: Deep Learning

Das Trainingslager für das maschinelle Lernen findet beim sogenannten Deep Learning (dt. „tiefes Lernen“) statt. Es ist ein Teilbereich des maschinellen Lernens, bei dem künstliche neuronale Netze mit vielen Schichten – sogenannten „tiefen“ Netzwerken – eingesetzt werden, um komplexe Muster in großen Datenmengen zu erkennen. Der Begriff „deep“ verweist auf die Vielzahl der Verarbeitungsschichten, durch die die Daten schrittweise abstrahiert und analysiert werden.

[35] Vgl. Konnecny, 2020, S. 59-60
[36] Vgl. Konnecny, 2020, S. 61
[37] Vgl. Konnecny, 2020, S. 62

Die Funktionsweise von Deep Learning ist der Struktur des menschlichen Gehirns nachempfunden: Informationen durchlaufen mehrere Ebenen, in denen sie jeweils weiterverarbeitet und gewichtet werden. Aus den gelernten Mustern kann das System schließlich Vorhersagen treffen oder neue Inhalte erzeugen. Deep Learning bildet somit das Fundament für viele moderne KI-Anwendungen, darunter auch bildgenerierende Systeme.[38]

[38] Vgl. Datasolut, 2024, im Internet

3. Bildgenerierung mit KI

Mithilfe künstlicher Intelligenz lassen sich auf Grundlage kurzer Texteingaben (Prompts) Bilder generieren, die sowohl spezifische Inhalte als auch Stile berücksichtigen. Die Bildgenerierung basiert auf dem Zusammenspiel verschiedener Deep-Learning-Architekturen, die in einem Text-Bild-Modell miteinander verbunden sind und Informationen effizient verarbeiten und umsetzen.[39]

3.1 Prompting

Um ein Bild von einer Künstlichen Intelligenz generieren zu lassen, bedarf es neben der KI selbst noch einer weiteren Komponente: dem *Prompt*.
Bei einem *Prompt* handelt es sich um eine Anweisung, um eine passende Antwort von der KI zu erhalten: dabei kann es sich um einen Text, Bild oder Audio handeln.
Der Prompt ist also das Schlüsselelement der Mensch-KI-Interaktion. Entscheidend ist hierbei: je klarer und präziser der Prompt formuliert, desto besser können die Ergebnisse sein, die von der Künstlichen Intelligenz erzeugt werden. Das gilt natürlich auch für die Generierung von Bildern.[40]

3.2 Text-Bild-Modell

Um aus natürlichsprachlichen Texteingaben Bilder zu erzeugen, muss der *Prompt* zunächst von einem Sprachmodell verarbeitet und in eine für den Bildgenerator nutzbare Repräsentation überführt werden.

Ein solches Text-zu-Bild-System basiert im Wesentlichen auf zwei zentralen Komponenten: einem Sprach-Bild-Modell, das darauf trainiert ist, Texte mit passenden Bildinhalten zu verknüpfen, und einem generativen Modell, das in der Lage ist, aus dieser Repräsentation neue Bilder zu erzeugen.

Das Ziel dieser Kombination ist es, das generative Modell so zu trainieren, dass es Bilder erzeugt, die vom Sprach-Bild-Modell als inhaltlich passend zum jeweiligen Prompt bewertet werden.

[39] Vgl. Nöcker-Prior, 2023, S. 1
[40] Vgl. Internationale Hochschule Akademie, im Internet

Dieses Zusammenspiel bildet die Grundlage moderner Text-Bild-Modelle, wie sie beispielsweise in DALL-E 3, Midjourney oder Stable Diffusion eingesetzt werden.[41]

3.3. Bild

Um zu verstehen, wie die Bildgenerierung durch Künstliche Intelligenz funktioniert, muss zuvor definiert werden, was ein *Bild* ist.
Zieht man dabei ChatGPT zu Rate, lassen sich unter Anderem folgende Aussagen treffen:

Ein Bild ist eine visuelle Darstellung. Diese visuelle Darstellung dient der Vermittlung von Informationen, Vorstellungen oder Eindrücken. Während im Englischen zwischen immateriellen (engl. “image”) und materiellen (engl. “picture”) Bildern unterschieden wird, gibt es im Deutschen nur den einen Begriff *Bild*, der beides miteinander vereint.

Aus technischer Sicht besteht ein Bild aus Linien, Farben, Texturen und Formen. Diese Komposition erzeugt einen Eindruck, eine Aussage, gar eine emotionale Wirkung. Dabei ist diese Wirkung nicht festgelegt, sondern liegt im Auge des Betrachters. Ein und dasselbe Bild kann also unterschiedliche Aussagen transportieren – je nachdem, wer es betrachtet und in welchem Kontext es steht.[42]

Eines der aktuell leistungsfähigsten und meistgenutzten Modelle zur Generierung von Bildern mit KI ist das *Diffusionsmodell*. Im Folgenden wird dieses näher erläutert.

3.4 Diffusionsmodell

Zum besseren Einstieg in das Thema sowie als Grundlage für das weitere Verständnis empfiehlt es sich, zunächst das im Rahmen dieser Arbeit produzierte Erklärvideo anzusehen. Dieses kann entweder über den beiliegenden USB-Stick oder über den folgenden QR-Code aufgerufen werden:

[41] Vgl. Nöcker-Prior, 2023, S. 1
[42] Vgl. Chat mit ChatGPT-4o vom 18.04.2025, 13:28 Uhr

Diffusionsmodelle gehören zu den leistungsfähigsten Verfahren der KI-gestützten Bildgenerierung. Das zentrales Prinzip beruht darauf, Bilder in mehreren Schritten zu verrauschen und anschließend zu rekonstruieren (entrauchen). Dieser Prozess wird in dem Erklärvideo vereinfacht dargestellt und hier erneut in schriftlicher, detaillierter Form erklärt.

Grundsätzlich lässt sich die Bildgenerierung mit einem Diffusionsmodell in drei Abschnitte unterteilen:

- Vorwärtsprozess (Forward Process)
- Rückwärtsprozess (Reverse Process)
- Bildgenerierung

Bevor der Vorwärtsprozess gestartet werden kann, wird die KI mit einem großen Datensatz an Bildern ausgestattet. Mit diesem Datensatz beginnt das *Training* der KI.

Der Vorwärts- und Rückwärtsprozess sind Teil der Trainingsphase und finden im *latenten* Raum statt. Bei dem latenten Raum (engl. “latent space”) handelt es sich um einen *multidimensionalen* Raum, in dem ein Bild in eine kompakte Darstellung aus Merkmalen übersetzt wird: das nennt man die *latente Repräsentation*. Siehe auch in Kapitel 3.4.4 Latente Diffusion (S. 26).

3.4.1 Vorwärtsprozess (Forward Process)

Der Vorwärtsdiffusionsprozess in einem Diffusionsmodell dient dazu, die Bilder aus dem Datensatz schrittweise in reines Rauschen zu überführen. Dies geschieht durch das wiederholte Hinzufügen von *Gaußschem Rauschen* in vielen kleinen Schritten. Dieser Prozess orientiert sich an dem Modell der *Markow-Kette:* das bedeutet, dass die Wahrscheinlichkeit von dem zukünftigen Zustand (wie das Bild im nächsten Schritt aussehen wird) immer nur vom aktuellen Zustand abhängt.
“*Einfach ausgedrückt: x_t, der Zustand der Markow-Kette x zum Zeitpunkt t, wird nur direkt von x_{t-1} beeinflusst.*” (vgl. Bergmann, Stryker).
So wird das ursprüngliche Bild Schritt für Schritt verrauscht, bis es komplett zufälligem Rauschen entspricht (siehe Abbildung 7).

3.4.2 Rückwärtsprozess (Reverse Process)

Bei dem Rückdiffusionsprozess wird ein neuronales Netz darauf trainiert, den Vorwärtsprozess umzukehren. Ziel ist es, aus reinem Gaußschen Rauschen wieder ein realistisches Bild zu rekonstruieren.

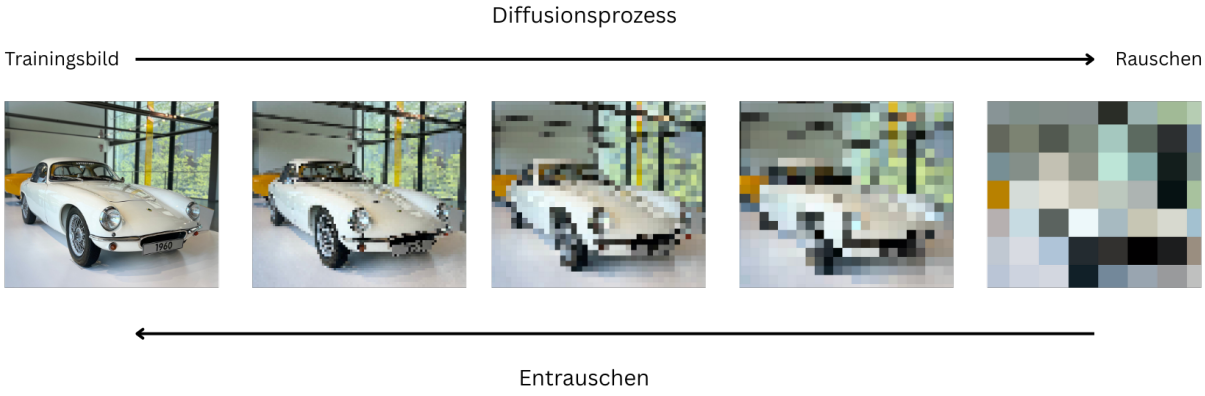


Abb. 7: Diffusionsprozess (Quelle: eigene Darstellung)

3.4.3 Geführte Diffusion (Guided Diffusion)

Im Training erfolgt der Vorwärts- und Rückwärtsprozess ohne weitere Eingaben: die Bilder aus dem Datensatz werden nur verrauscht und rekonstruiert. Für die tatsächliche Bildgenerierung hingegen muss das Modell eine Eingabe erhalten, z. B. in Form eines Text-Prompts.

Ein sogenanntes *Large Language Model* (LLM) wie *CLIP* von OpenAI interpretiert diesen Prompt und übersetzt ihn in eine numerische Repräsentation (z.B. ein Vektor), mit der der Diffusionsprozess gezielt gesteuert werden kann. Das System wird dadurch zu einem *geführten* Diffusionsmodell.[43]

3.4.3.1 CLIP

CLIP steht für *Contrastive Language–Image Pre-training* und dabei handelt es sich um ein neuronales Netzwerk, das Texte und Bilder gemeinsam verarbeitet. CLIP besteht aus zwei Teilen: einem Text-Encoder, der einen Text-Prompt in einen Vektor übersetzt, und aus

[43] Vgl. IBM, im Internet

einem Bild-Encoder, der dasselbe für ein Bild tut. Beide Vektoren liegen im selben Raum, sodass das Modell lernt, welche Texte zu welchen Bildern gehören.[44]

3.4.4 Latente Diffusion

Der Diffusionsprozess findet im *latenten* Raum statt. Der latente Raum orientiert sich an dem Konzept der *Variational Autoencoder* (VAEs) (siehe Abbildung 8). Ein VAE ist in der Lage, Daten – in diesem Fall *Bilder* – zu komprimieren (kodieren) und nur die wichtigen Strukturen und Merkmale des Bildes zu behalten und diese im latent space abzulegen. Diese komprimierten Daten können aus dem latenten Raum wieder entnommen werden und die ursprünglichen Daten (Bilder) werden rekonstruiert (dekodiert).

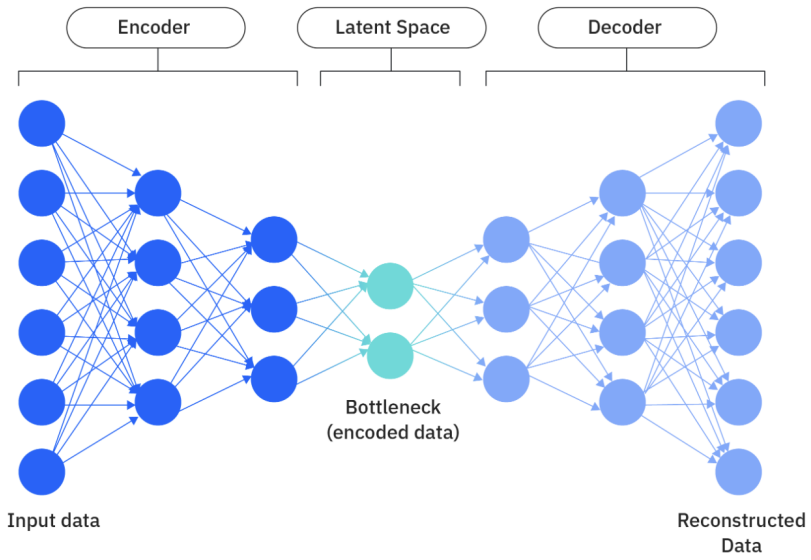


Abb. 8: Aufbau eines Variational Autoencoder (Quelle: IBM)

Der große Vorteil dieser Methode liegt in der *Effizienz*: Da nicht mit hochauflösenden Pixelbildern, sondern mit kompakten latenten Repräsentationen gearbeitet wird, sinkt der Rechenaufwand erheblich, trotz gleichzeitig hoher Bildqualität.

[44] Vgl. OpenAi, im Internet

3.4.3 Bildgenerierung

Nach der Trainingsphase kann ein Diffusionsmodell genutzt werden, um neue Bilder zu erzeugen. Dazu wird ein zufälliges verrauschtes Bild ausgewählt und Schritt für Schritt entrauscht. Durch den Einsatz eines gewissen Zufallanteils im Prozess entstehen dabei nicht einfach Kopien der Trainingsbilder, sondern neue, ähnliche Bilder. [45]

3.5 Die aktuellen Systeme

Im folgenden Abschnitt werden ein paar der gängigsten bildgenerierenden Künstlichen Intelligenzen vorgestellt, die auf einem Diffusionsmodell basieren.

3.5.1 Stable Diffusion

Stable Diffusion ist ein latentes Text-zu-Bild-Diffusionsmodell, das auf Millionen von Bildern trainiert wurde. Es nutzt einen CLIP-Textencoder, um Texteingaben als Prompt zu interpretieren. [46]
Eine Besonderheit: Bei Stable Diffusion handelt es sich um ein Opensource Modell. Es ist somit einsehbar, bearbeitbar und direkt auf den Computern der Nutzer ausführbar. Somit bildet Stable Diffusion die Basis für viele weitere Bildgeneratoren, die heutzutage einen Großteil der Bilder generieren – dies verdeutlicht auch die Grafik aus Abbildung 9.

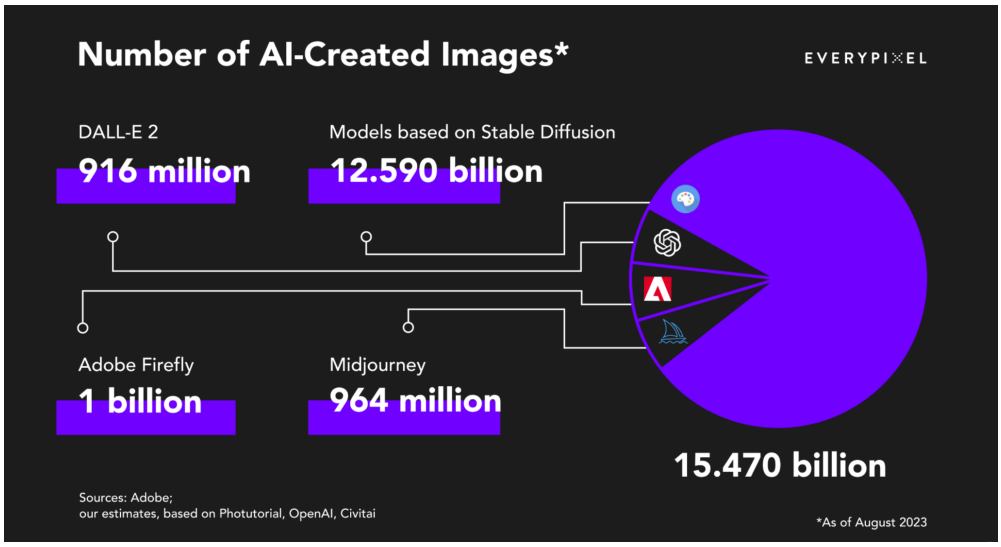


Abb. 9: Statistik der populärsten Bildgenerierenden KIs (Quelle: Journal Everypixel)

[45] Vgl. IBM, im Internet
[46] Vgl. Github, im Internet

Der Autor Uria-Recio beschreibt es wie folgt: “Stable Diffusion ist für die Bilderzeugung das, was ChatGPT für die Sprachverarbeitung ist, mit dem wichtigen Unterschied, dass Stable Diffusion Opensource ist.” (vgl. Uria-Recio, 2024, S. 204).[47]

3.5.2 DALL-E 3

DALL-3 wurde von OpenAI entwickelt, erschien im September 2023, und wurde auf dem Large Language Modell von ChatGPT gebaut. Nutzer können somit in der Version ChatGPT-4o direkt Bilder von DALL-E 3 generieren lassen. Dafür nutzen sie entweder einen reinen Text-Prompt oder einen Text-Prompt in Kombination mit einem Bild.

Besonders viral ging ein Trend, bei dem Nutzer Bilder im Stil des Animationsstudios *Studio Ghibli* teilten. Dieser Trend löste eine regelrechte Bilderflut aus und markierte den Moment, in dem Millionen von Menschen begannen, über ChatGPT eigene Bilder zu generieren (siehe Abbildung 10 und 11).[48]

Abb. 10: Meme “Disaster Girl” (Quelle: Wikipedia)

Abb. 11: "Disaster Girl" im Studio Ghibli Stil (Quelle: euronews)

DALL·E 1 wurde im Jahr 2021 veröffentlicht und nutzte ein Transformer-Modell, das Bilder direkt aus Texteingaben erzeugte. Bei DALL·E 2 kam erstmals ein Diffusionsmodell zum Einsatz, das gemeinsam mit dem CLIP-Modell arbeitete. DALL·E 3 baut auf diesen Ansätzen auf und kombiniert ein verbessertes Diffusionsmodell mit einem Sprach-Bild-System, das durch die Integration von GPT-4 deutlich besser darin ist, auch komplexe Texteingaben zu verstehen und passende Bilder zu erzeugen.[49]

[47] Vgl. Uria-Recio, 2024 S. 204-206
[48] Vgl. Euronews, im Internet
[49] Vgl. Pooja M M, 2025, S.2

3.5.3 Midjourney

Midjourney ist ein KI-System zur Bildgenerierung, das im Jahr 2022 erstmals veröffentlicht wurde. Entwickelt wurde es vom gleichnamigen Forschungslabor Midjourney, Inc. mit Sitz in San Francisco.

Der Zugang zu Midjourney erfolgt primär über einen Discord-Bot sowie über eine zugehörige Webanwendung. Innerhalb von Discord kann die Nutzung über den offiziellen Server, über Direktnachrichten an den Bot oder durch die Integration in einen externen Server erfolgen (siehe Abbildung 12).

Bei Discord handelt es sich um eine Kommunikationsplattform, die Sprach-, Video- und Textübertragungen unterstützt und vor allem im Bereich digitaler Communities verbreitet ist.[50]

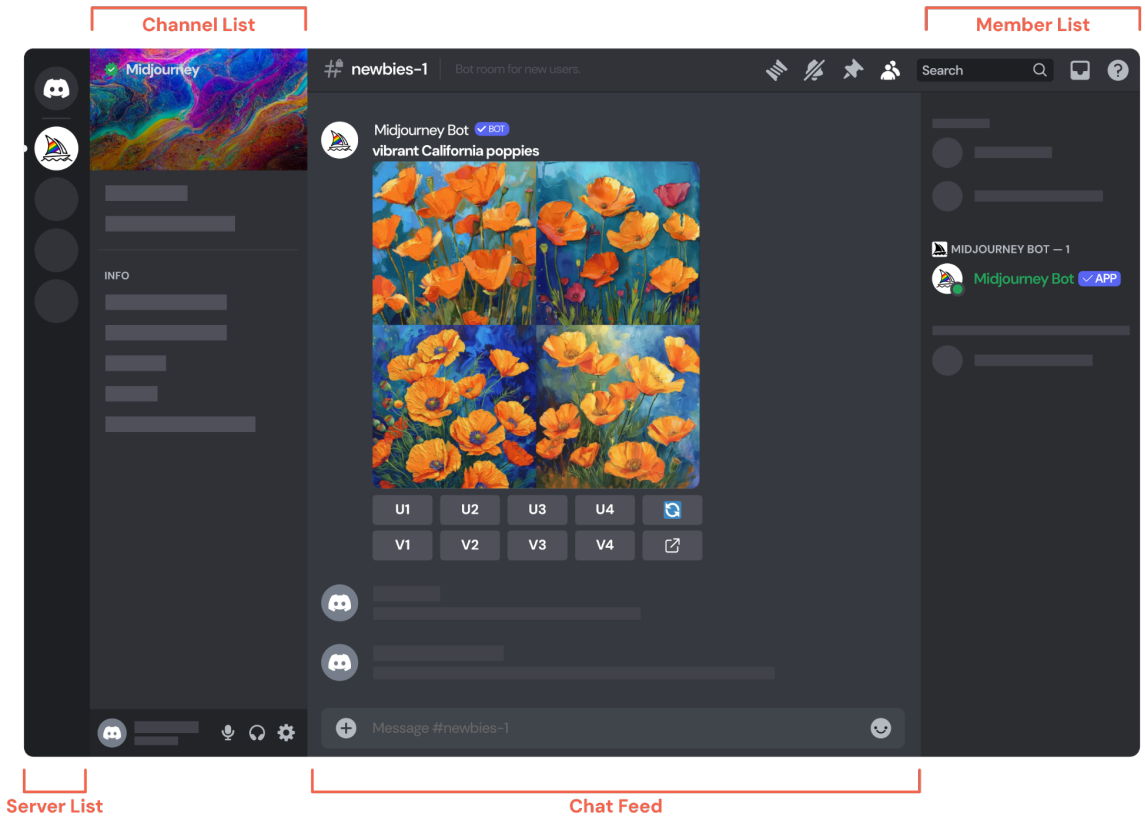


Abb. 12: Benutzeroberfläche Midjourneys Discord Kanal (Quelle: Midjourney)

[50] Vgl. Epidemic Games, im Internet

4. Bedarfsermittlung

4.1 Einleitung

Vor der Konzeption des Erklärvideos wurde eine Umfrage unter Studierenden des 1. Semester im Studiengangs *Digitale Medien* durchgeführt, um den Kenntnisstand sowie die Einstellung der Zielgruppe gegenüber Künstlicher Intelligenz und insbesondere KI-gestützten Bildgeneratoren zu erfassen. Ziel der Erhebung war es, den Bedarf für ein erklärendes Video zu prüfen und konkrete Anhaltspunkte für die inhaltliche und didaktische Gestaltung des Videos zu gewinnen.

Insgesamt nahmen 121 Personen an der Umfrage teil.

Die Mehrheit der Befragten war unter 22 Jahre alt (86 %), 13 % befanden sich im Altersbereich zwischen 23 und 28 Jahren und 4 % waren zwischen 29 und 35 Jahre alt. Hinsichtlich des Geschlechts identifizierten sich 54 % der Teilnehmenden als weiblich, 43 % als männlich und 3 % als divers.

Die Studierenden wurden unter anderem gefragt, ob sie bereits Tools mit Künstlicher Intelligenz (z. B. ChatGPT) genutzt oder selbst Erfahrungen mit KI-gestützter Bildgenerierung gesammelt haben. Darüber hinaus sollten sie ihr eigenes Wissen über Künstliche Intelligenz und KI-Bildgeneratoren einschätzen sowie Vorteile und mögliche Probleme solcher Systeme benennen.

4.2 Ergebnisse

85,1% der Befragten gaben an, bereits mit einer Künstlichen Intelligenz gearbeitet zu haben – wiederum nur 50,4% von ihnen haben mit einer KI ein Bild generiert.

Um den Wissensstand der Studierenden sowohl im Bereich der allgemeinen Künstlichen Intelligenz als auch speziell der bildgenerierenden KI zu erfassen, wurden sie gebeten, ihr eigenes Wissen auf zwei separaten Skalen von 1 (sehr gering) bis 5 (sehr gut) einzuschätzen (siehe Abbildung 13 und 14).

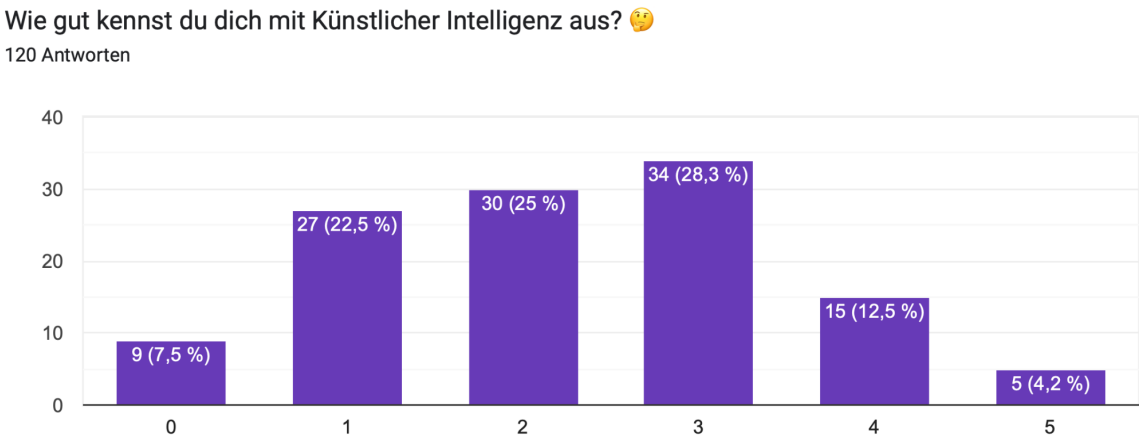


Abb. 13: Vorkenntnisse zu Künstlicher Intelligenz (Quelle: eigene Darstellung auf Basis von Google Forms)

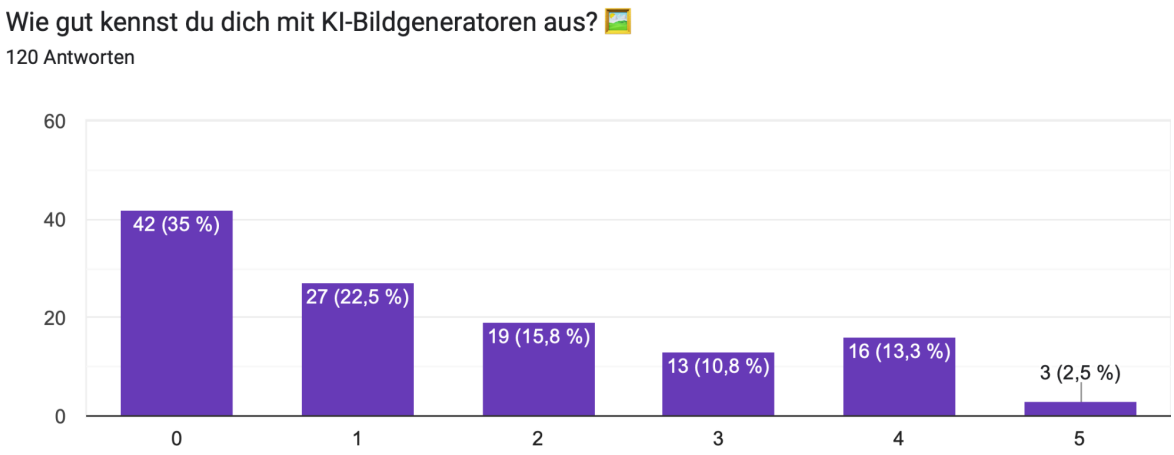


Abb. 14: Vorkenntnisse zu bildgenerierender Künstlichen Intelligenz (Quelle: eigene Darstellung auf Basis von Google Forms)

Bei der offenen Frage „Was wissen Sie über bildgenerierende Künstliche Intelligenz?“ reichen die Antworten von sehr allgemeinen Aussagen wie „Es wird eine große Datenbank an Bildern bezogen, um neue Bilder zu generieren“ bis hin zu differenzierteren Einschätzungen wie „Funktioniert mit neuronalen Netzwerken. Größte Vertreter z.B. Midjourney oder DALL-E. Trainiert mit Millionen von Bildern aus dem Internet. Möglichkeit mit wenigen Worten ein noch nie dagewesenes Bild in Sekunden zu erschaffen“.

Einzelne Teilnehmende äußerten sich auch kritisch, etwa: „Das die meisten Künstler nicht darüber erfreut sind“ oder „Es nimmt schon existierende Bilder und trainiert dadurch und ist deswegen Urheberrechtlich kritisch“.

Diese Bandbreite an Antworten verdeutlicht, dass zwar erste Vorstellungen vorhanden sind, jedoch häufig Unsicherheiten über die genaue Funktionsweise bestehen.

Auf die offene Frage „Was meinen Sie, welche Vorteile ein KI-Bildgenerator bietet?“ nannten die Teilnehmenden vielfältige positive Aspekte.

Die Teilnehmenden nannten vor allem die schnelle und kostengünstige Erstellung von Bildern, die einfache Zugänglichkeit auch ohne künstlerische Vorkenntnisse sowie die Unterstützung bei kreativen Prozessen als zentrale Vorteile von KI-Bildgeneratoren.

Zudem wurden Möglichkeiten der Personalisierung und die Vereinfachung gestalterischer Aufgaben hervorgehoben. Kritische Stimmen blieben vereinzelt und bezogen sich meist auf die noch unvollkommene Bildqualität. Insgesamt sehen die Studierenden großes Potenzial in Effizienzsteigerung, kreativer Unterstützung und dem erleichterten Zugang zu individuellen Bildern.

4.3 Auswertung

Die Auswertung der Umfrage zeigt, dass grundsätzlich ein breites Interesse an Künstlicher Intelligenz besteht. Während sich viele Befragte im Bereich allgemeiner KI-Anwendungen als relativ sicher einschätzen, gaben deutlich weniger Teilnehmende an, sich mit bildgenerierender KI gut auszukennen. Insbesondere bei der Selbsteinschätzung zum Thema KI-Bildgeneratoren zeigte sich eine Wissenslücke, die den Bedarf nach einer niedrigschwelligen, verständlichen Einführung deutlich macht.

Auch die freien Antworten auf die Fragen nach Vorteilen und bisherigen Kenntnissen verdeutlichen, dass die Studierenden zwar grundlegende Potenziale von KI-Bildgeneratoren – wie Zeitersparnis, kreative Unterstützung und Zugänglichkeit – erkennen, jedoch oft keine tiefergehenden Kenntnisse über Funktionsweise, technische Hintergründe oder Grenzen dieser Systeme besitzen. Unsicherheiten in Bezug auf Qualität, Originalität und Urheberrecht wurden ebenfalls vereinzelt geäußert.

Insgesamt lässt sich festhalten, dass sowohl ein grundlegendes Interesse als auch ein klarer Informationsbedarf in der Zielgruppe besteht. Aus diesen Erkenntnissen ergibt sich die Notwendigkeit, zentrale Begriffe, technische Abläufe und Zusammenhänge rund um bildgenerierende KI verständlich und anschaulich aufzubereiten. Die Erstellung eines Erklärvideos bietet hierfür eine geeignete Möglichkeit, da es komplexe Sachverhalte visuell unterstützt erklären und damit die Einstiegshürde für die Studierenden senken kann.

5. Das Erklärvideo

5.1 Didaktischer Hintergrund

Erklärvideos sind kurze audiovisuelle Formate, in denen komplexe Inhalte oder Sachverhalte verständlich aufbereitet und erläutert werden. Sie dienen dazu, Funktionen, Abläufe oder abstrakte Konzepte anschaulich zu vermitteln und werden häufig eigens für didaktische oder kommunikative Zwecke produziert.[51]

Insbesondere bei Zielgruppen ohne fachliche Vorkenntnisse, wie Studierenden im ersten Semester, helfen Erklärvideos dabei, einen niedrigschwelligen Zugang zu einem Thema zu ermöglichen.

Vor diesem Hintergrund wurde auch für diese Bachelorarbeit ein Erklärvideo entwickelt, das die grundlegende Funktionsweise bildgenerierender Künstlicher Intelligenz vermittelt. Die didaktische Gestaltung orientiert sich dabei an den grundlegenden Prinzipien der Lernpsychologie, insbesondere an den Konzepten des *Multimedia Learning* nach Richard E. Mayer. Dazu wurde eine Kombination aus gesprochener Sprache, animierten Bildern und Typografie eingesetzt, um verschiedene Sinne gleichzeitig anzusprechen und das Lernen zu unterstützen.[52]

Des Weiteren ist es so, dass heutzutage viele Menschen an kurze, schnelle und stark visuelle Inhalte gewöhnt sind, wie man sie zum Beispiel auf Plattformen wie TikTok oder Instagram findet. Dort passiert ständig etwas Neues im Bild und Ton, um die Aufmerksamkeit zu halten. In diesem Zusammenhang wird umgangssprachlich manchmal vom sogenannten „TikTok Brain“ gesprochen. Damit ist gemeint, dass viele Zuschauer inzwischen eine kürzere Aufmerksamkeitsspanne haben und schnell das Interesse verlieren, wenn Inhalte nicht abwechslungsreich gestaltet sind. [53]

[51] Vgl. Becker und Brehmer, 20217, S. 1
[52] Vgl. Ludwig-Maximiloams-Universität München, im Internet
[53] Vgl. Spektrum, im Internet

5.2 Zielgruppe und Lernziel

Die Zielgruppe des Erklärvideos sind Studierende im ersten Semester des Studiengangs *Digitale Medien*. Im Rahmen des Moduls *Medieninformatik I* wird dort unter anderem das Thema Künstliche Intelligenz, insbesondere *bildgenerierende* Künstliche Intelligenz behandelt.

Das übergeordnete Lernziel besteht darin, den Studierenden ein grundlegendes Verständnis für die Funktionsweise bildgenerierender KI zu vermitteln. Dabei soll das Video zentrale Begriffe und Abläufe anschaulich erklären und den Zusammenhang zwischen Texteingabe und Bildausgabe nachvollziehbar darstellen. Ziel ist es, die Inhalte so aufzubereiten, dass sie sowohl informativ als auch anregend für eine weiterführende Beschäftigung mit dem Thema wirken.

5.3 Die Konzeption

Ziel des animierten Erklärvideos ist es, den Prozess der Bildgenerierung sowohl verständlich als auch visuell ansprechend und anregend zu gestalten. Das Video soll dabei nicht ausschließlich informativ sein, sondern auch eine gewisse Unterhaltung bieten, ohne dabei auf wesentliche Fachbegriffe zu verzichten. Es wird eine Balance zwischen Wissenstransfer und Unterhaltung angestrebt.

5.3.1 Die Storyentwicklung

Ursprünglich bestand die Idee darin, dass eine Figur den Prozess der Bildgenerierung erklärt. Diese erste Konzeption war informativ, jedoch ohne erzählerischen Rahmen. Schnell wurde deutlich, dass dem Konzept eine narrative Struktur fehlte, die den Zuschauer stärker einbindet. Daraus entwickelte sich die Idee, das Video an das Modell der Hero’s Journey (dt. *Heldenreise*) anzulehnen (vgl. Kapitel 5.3.3 Die Heldenreise**)**. Damit war klar: Die Geschichte benötigt eine Hauptfigur, die eine Entwicklung durchläuft.

Die Figur, die den Prozess erklärt, fungiert eher in der Rolle der *Mentorin*, daher war schnell ersichtlich, dass eine zweite Figur die eigentliche Heldenrolle übernehmen musste. In Ansprachen wendet sich die Mentorin immer direkt an eine Person: den Zuschauer. Daraus entstand die Überlegung, den Zuschauer symbolisch in die Rolle des Helden zu versetzen. Doch wer wäre innerhalb der Geschichte ein sinnvoller Adressat für die Erklärungen zur Bildgenerierung? Die naheliegende Antwort: ein Computer.

So entstand das Konzept, dass die Mentorin einem Computer beibringt, wie man Bilder generiert – oder, einfacher gesagt: wie man „malen“ lernt. Daraus ergibt sich auch der Titel des Projekts: „Ein Computer lernt malen“. Der Zuschauer übernimmt damit die Perspektive des Computers, der schrittweise lernt Bilder zu erzeugen.

Die zentrale Herausforderung bestand darin, diese Rollenverteilung für den Zuschauer verständlich zu machen. Dies wurde über die Sprache gelöst: Bereits zu Beginn des Videos stellt die Mentorin klar, an wen sie sich wendet, etwa mit den Worten:

> *“So you would like to learn how to draw?* ***But are you, as a computer,*** *even capable of drawing?”*

Auch im weiteren Verlauf wird immer wieder deutlich, dass sich die Ansprache an einen Computer richtet – etwa durch Aussagen wie:

> *“But there is a problem here:* ***as a computer****,* ***you are*** *not able to read a text or interpret a picture – you can only work with numbers.”*

5.3.2 Die Story

Die Geschichte beginnt damit, dass eine ältere Dame an die „Glasscheibe“ des Bildschirms klopft und den Zuschauer fragt, ob er das Malen lernen möchte. Doch schon im nächsten Moment stellt sie selbst fest, dass das gar nicht so einfach werden könnte, denn schließlich ist das Gegenüber ein Computer, der weder lesen noch einen Pinsel halten kann. Trotz dieser Hürden streckt sie ihm die Hand entgegen und bietet an, ihm das Malen beizubringen.

Damit beginnt der eigentliche Lernprozess. Die alte Dame, mit dem Namen Ruby, erklärt, dass der Computer zunächst lernen muss, Bilder ihren passenden Bildbeschreibungen zuzuordnen – ähnlich wie bei einem Memory-Spiel. Da ein Computer jedoch keine Sprache versteht und auch keine Bilder interpretieren kann, scheint diese Aufgabe auf den ersten Blick unlösbar.
Ruby bietet jedoch eine Lösung: Der Computer soll zu jedem Bild und jeder Bildbeschreibung einen eigenen Vektor erstellen – eine Art Übersetzung in seine „Sprache“. Auf diese Weise kann er die Paare als Zahlenkombinationen erfassen und speichern. Die so gebildeten Vektoren werden in einem multidimensionalen Raum abgelegt, das ähnlich wie eine Bibliothek funktioniert, in der die Inhalte nach Kategorien geordnet werden.

Nachdem der Computer diese Phase des Trainings erfolgreich abgeschlossen hat, steht er vor seiner finalen Herausforderung: dem Diffusionsprozess.
Dafür wird eines der bekannten Bilder Schritt für Schritt mit Rauschen überlagert. Der Computer erhält nun die Aufgabe, dieses verrauschte Bild wiederherzustellen, allein auf Basis der dazugehörigen Bildbeschreibung und des gespeicherten Vektors. Gelingt ihm das, wartet die letzte und schwerste Prüfung auf ihn:
Er soll nun erstmals eigenständig ein Bild generieren, das nicht Teil der Trainingsdaten war.

Dazu bekommt der Computer einen Text-Prompt, den passenden Vektor und ein vollständig verrauschtes Bild, das keine erkennbare Struktur enthält. Doch auch diese Aufgabe meistert er schließlich: Aus dem Rauschen entsteht ein neues, klares Bild.

Damit hat der Computer seine finale Prüfung bestanden und die Fähigkeit erlangt, selbständig Bilder zu malen.

5.3.3 Die Heldenreise

Die „Heldenreise“ (engl. *Hero’s Journey*) ist ein universelles Erzählmuster, das auf die mythologischen Studien von Joseph Campbell zurückgeht. In seinem Werk *The Hero with a Thousand Faces* (1949) beschreibt Campbell die strukturellen Gemeinsamkeiten zahlreicher Mythen, Sagen und Geschichten unterschiedlicher Kulturen. Daraus entwickelt er ein zyklisches Modell, das aus mehreren Phasen besteht – darunter der Ruf zum Abenteuer, die Bewältigung von Prüfungen, das Treffen mit einem Mentor, die Konfrontation mit dem Unbekannten sowie die Rückkehr mit einer transformierenden Erkenntnis (siehe Abbildung 15).[54]

Abb. 15: Die Heldenreise (Quelle: eigene Darstellung)

[54] Vgl. Masterclass, im Internet

Im Rahmen der vorliegenden Arbeit wurde die Heldenreise als erzählerisches Fundament für das Erklärvideo gewählt. Obwohl das Genre des Erklärvideos primär informativen Charakter hat, ermöglicht der Einsatz der Heldenreise eine emotionale Rahmung der Wissensvermittlung. In der narrativen Struktur übernimmt die Figur „Ruby“ die Rolle der Mentorin, während der Zuschauer in der Rolle des unerfahrenen „Computers“ schrittweise in eine neue Welt – die Welt der künstlichen Intelligenz und der Bildgenerierung – eingeführt wird. Er ist also der *Held* der Geschichte. Ruby initiiert den Lernprozess (Ruf zum Abenteuer), führt durch verschiedene Stationen der Erkenntnisgewinnung (Prüfungen und Transformation), bis der Zuschauer schließlich erkennt, dass er selbst ein Bild erschaffen kann. Er bewältigt somit die *große Prüfung* und kehrt in die vertaute Welt zurück, mit der neuen Fähigkeit des Bildgenerierens (beide Welten vereint).

5.3.4 Einsatz von Fachbegriffen

Wie im Verlauf dieser Arbeit deutlich wurde, ist das Themenfeld der bildgenerierenden Künstlichen Intelligenz mit vielen, komplexen Fachbegriffen verbunden, zum Beispiel *neuronale Netze*, *Diffusion* oder der *latente Raum*. Da das Erklärvideo eine Balance zwischen fachlicher Tiefe und unterhaltsamer, vereinfachter Darstellung anstrebt, war es wichtig, die inhaltliche Komplexität gezielt zu steuern. Ziel war es, die Zuschauer nicht zu überfordern, gleichzeitig jedoch zentrale Fachbegriffe und Konzepte nicht vollständig auszublenden. Daher musste sorgfältig abgewogen werden, welche Begriffe im Video aufgegriffen und welche weggelassen oder vereinfacht dargestellt werden können.

Der Begriff „latenter Raum“ wird in dem Video nicht explizit verwendet. Stattdessen wird er als „multidimensionaler Raum“ bezeichnet. Diese Entscheidung basiert auf zwei Überlegungen: Zum einen ist der Ausdruck „latenter Raum“ im deutschen Sprachgebrauch wenig geläufig, zum anderen erklärt sich der Begriff „latent“ in diesem Zusammenhang nicht von selbst. Die Bezeichnung „multidimensionaler Raum“ vermittelt hingegen anschaulicher, worum es geht: um einen Raum mit vielen Dimensionen, in dem komplexe Zusammenhänge mathematisch dargestellt werden.

Die Begriffe „Rauschen“ (engl. *Noise*) und “Entrauschen” (engl. Denoising) werden verwendet, um den zentralen Mechanismus des Diffusionsprozesses zu veranschaulichen. Er beschreibt genau das, was im Video gezeigt wird: Ein Bild wird schrittweise mit Rauschen überlagert und anschließend wieder davon befreit. Aus diesem Grund erschien es sinnvoll, diesen Begriff in dem Video einzusetzen.

Auf die Verwendung komplexerer Begriffe wie „neuronales Netz“ oder „Markow-Kette“ wurde bewusst verzichtet, da sie nicht selbsterklärend sind und eine ausführlichere Erklärung erfordern würden. Dies würde von der zentralen Aussage des Videos – der anschaulichen Darstellung des Diffusionsprozesses – ablenken.

5.3.5 Sprache

Das Erklärvideo wurde bewusst auf Englisch erstellt, da Englisch in der Informatik und insbesondere im Bereich Künstliche Intelligenz die zentrale Sprache ist. Viele Begriffe wie „diffusion model“, „prompt“ oder „neural network“ stammen aus dem Englischen und werden in der Fachliteratur sowie in KI-Tools (z. B. DALL·E, Midjourney) meist nicht übersetzt.

Des Weiteren macht die Wahl der englische Sprache das Video das Video für mehr Menschen zugänglich und einsatzfähig. Somit kann es innerhalb der Universität von allen Studierenden genutzt werden.

5.3.6 Storyboard

Bevor mit der gestalterischen Ausarbeitung und der anschließenden Videoproduktion begonnen werden konnte, musste die zuvor entwickelte Erzählung in ein Storyboard überführt werden.

Zu diesem Zweck wurde eine tabellarische Struktur gewählt, die sich in die Kategorien Szene, Kameraeinstellung, Handlung, Voice-Over-Text, Bildreferenz sowie Dauer unterteilt.

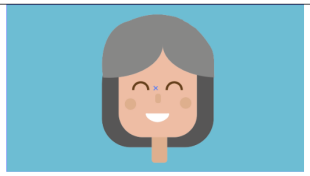
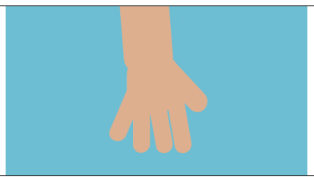
Scene	Camera	plot	Voice Over	Reference	Duration in seconds
1.1	Closeup to medium long shot	Ruby appears on the screen really close, knocks against the screen and talks to the audience. Then she backs up a bit. She is quirky and excited as she talks to the audience.	*-Excited-* *„Hi!* *So you would like to learn how to draw?”* *-(sceptical-* *“hmm, but are computers even capable of drawing?* *You can’t even hold a pencil ..* *and can not even read?”* **thinking** *(hopeful)* *„You know what? I think we can make it work!”*		8sec.
1.2	Top view close up	She puts out her hand towards the sceen,	*“Do you trust me?”*		3sec.

Abb. 16: Ausschnitt des ersten Version des Storyboards (Quelle: eigene Darstellung)

Die Geschichte wurde auf dieser Grundlage in einzelne Szenen und Kameraeinstellungen unterteilt und systematisch nach dem gewählten Storyboard-Schema aufgebaut (siehe Abbildung 16).

Die Skizzen, die im Storyboard zu sehen sind, sind teilweise selbst in Adobe Illustrator angefertigt worden und teilweise aus dem Internet von der Internetseite Canva.com.

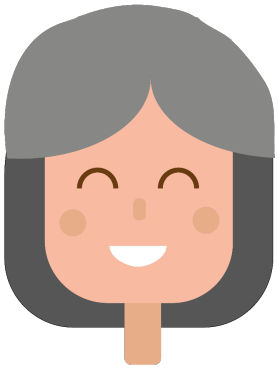
Das vollständige Storyboard ist im Anhang zu finden.

5.4 Umsetzung

5.4.1 Gestaltung

Die visuelle Gestaltung des Erklärvideos trägt wesentlich dazu bei, die Inhalte anschaulich und verständlich zu vermitteln. Ziel war es, ein einheitliches und ansprechendes Erscheinungsbild zu schaffen, das die Erklärungen sinnvoll unterstützt und den Zuschauern hilft, sich im Video zurechtzufinden.
Der gewählte Illustrationsstil ist bewusst reduziert und klar gehalten. Dadurch wirken die gezeigten Elemente nicht überladen und die Aufmerksamkeit bleibt auf den wichtigen Inhalten. Die Hauptfigur Ruby ist vereinfacht und freundlich gezeichnet. Diese bewusste gestalterische Entscheidung sorgt dafür, dass sie leicht erkennbar ist und gleichzeitig sympathisch wirkt. So fällt es den Zuschauern leichter, ihr zu folgen und sich auf ihre Erklärungen einzulassen.

5.4.1.1 Die Hautfigur



Die Gestaltung der Hauptfigur *Ruby* begann mit einem einfachen Sketch ihres Kopfes in dem Programm *Adobe Illustrator (Siehe Abbildung 17)*. Ruby soll ein vereinfacht und freundlich gezeichneter Charakter sein.
Diese bewusste gestalterische Entscheidung sorgt dafür, dass sie leicht erkennbar ist und gleichzeitig sympathisch wirkt.

Abb. 17: Ruby Sketch (Quelle: eigene Darstellung)

Nach ein paar Iterationen war Ruby fertig gestellt: eine schlanke, ältere Dame, mit einer dunklen Hose, einem T-Shirt in der Farbe Orange und einer Brille auf der Nase. Ruby hat weißes, hochgestecktes Haar und ein Lächeln auf den Lippen (siehe Abbildung 18). Sie trägt warme Farben, die sich sie deutlich vom Hintergrund abheben und machen sie zur zentralen Figur.

Abb. 18: Ruby finalisierte Illustration (Quelle: eigene Darstellung)

Des Weiteren fiel die Entscheidung auf bunte Kleidung und Turnschuhe, da sie den Charakter quirlig und jung wirken lassen.

Diese Hauptfigur ist angelehnt an *Oma Grete,* aus der bekannten Hörspiel- und Zeichentrickserie *Bibi Blocksberg*, die als visuelle und charakterliche Inspirationsquelle diente (siehe Abbildung 19).

Oma Grete ist eine Nebenfigur der Serie und die Großmutter der Hauptfigur Bibi Blocksberg. Charakterlich zeichnet sich Oma Grete durch ihre moderne und unkonventionelle Art aus. [55]

In ihrer Rolle als Bibis Großmutter steht sie ihr immer zur Seite und fungiert auch als eine Art Mentorin.

Abb. 19: Oma Grete (Quelle: Bibi Blocksberg)

5.4.1.2 Die Farbauswahl

Die Farben im Video wurden gezielt eingesetzt. Ruby trägt warme Farben, die Vertrauen und Kompetenz ausstrahlen. Diese Farbwahl hebt sie deutlich vom Hintergrund ab und macht sie zur zentralen Figur. Der Hintergrund selbst bleibt eher schlicht. In einigen Szenen werden Farbübergänge genutzt, um neue Abschnitte einzuleiten oder Inhalte voneinander abzugrenzen. Dadurch wird das Video übersichtlicher und leichter verständlich (siehe Abbildung 20).

Abb. 20: Ruby auf blauem Grund (Quelle: Screenshot Erklärvideo)

[55] Vgl. Bibi Blocksberg, im Internet

5.4.2 Die Videoproduktion in Adobe After Effects

Für die Umsetzung des Erklärvideos wurde Adobe After Effects verwendet – eine Software, die speziell für Animation und Bewegtbildgestaltung entwickelt wurde. Die Grafiken, die zuvor in Adobe Illustrator erstellt wurden, konnten in After Effects als einzelne Ebenen (Layers) importiert und dort unabhängig voneinander animiert werden. So war es möglich, zum Beispiel Arme, Augen oder Objekte zu bewegen.

5.4.2.1 Lippensynchronisation mit SF-Caddy

Um Rubys Lippenbewegung auf das Gesprochene anzupassen, bedarf es einer Lippensynchronisation. Hierfür wurde ein Plugin namens *SF-Caddy* in Adobe After Effects eingesetzt.
SF-Caddy ermöglicht es, verschiedene Mundformen (sogenannte „Mouth Shapes“ oder „Phoneme“) als Buttons in einem kleinen Panel in After Effects anzuzeigen. Diese Mundformen sind in einer Komposition. Mit einem Klick auf den entsprechenden Button kann man dann direkt zur passenden Mundform im Zeitverlauf springen oder sie an der aktuellen Stelle einfügen.

5.4.2.1 After Effects trifft auf Stable Diffusion

Alle Illustrationen im Animationsvideo wurden eigenständig in Adobe Illustrator erstellt – mit zwei Ausnahmen: die Hundebilder und das Bild des grünen Autos stammen aus dem KI-System *Stable Diffusion*. Der Grund dafür ist, dass im Video der Diffusionsprozess als zentraler Bestandteil der Bildgenerierung visuell erklärt wird.
Dazu wurde mithilfe eines Prompts ein Bild in Stable Diffusion generiert. Mit Unterstützung eines Kommilitonen konnten die einzelnen Zwischenschritte der Bildentstehung aus dem Programm exportiert und in die Animation eingebunden werden, um den Ablauf der Diffusion nachvollziehbar, authentisch und realistisch darzustellen.

5.5 Iterationen und gestalterische Anpassungen

Alle Aspekte dieser Videoproduktion gingen durch mehrere Iterationen: Story, Gestaltung, Animation.

5.5.1 Darstellung des Diffusionsprozesses

Ein wesentlicher Aspekt, der sich im Verlauf des Projekts mehrfach verändert hat, war die Darstellung des Diffusionsprozesses.
In der ersten Version wurde dieser über eine metaphorische Bildsprache erklärt: Der Computer blickt auf ein Bild, das nach und nach hinter einer vernebelten Glasscheibe verschwindet. Die Idee war, dass der Computer das Bild durch eine einfache Wischbewegung mit der Hand wieder sichtbar macht: so, als würde er den Nebel oder das Kondenswasser entfernen (siehe Abbildung 21).

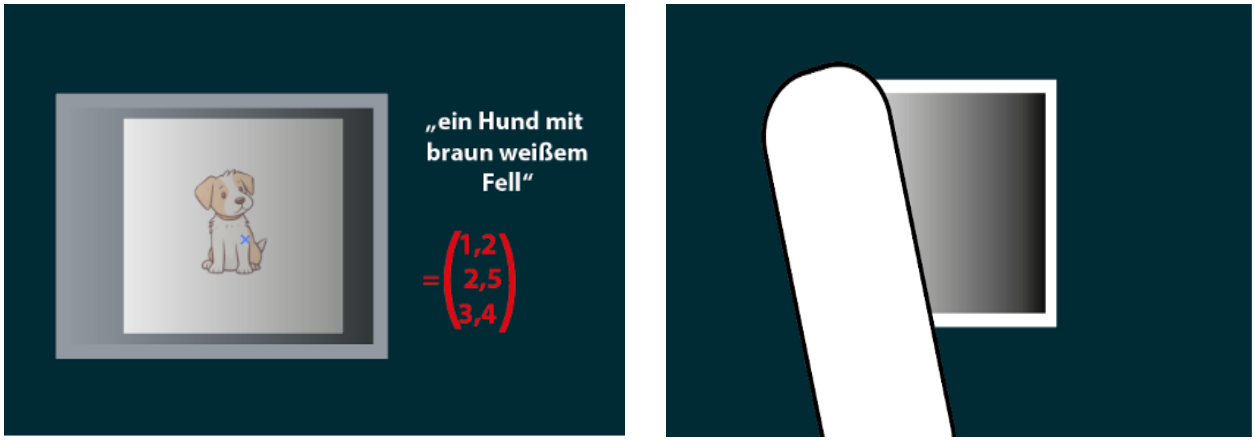


Abb. 21: erster Versuch der Darstellung von Diffusion

Eine zweite Variante ergänzte diese Szene um eine Staffelei mit einer Leinwand: Das Bild auf der Glasscheibe verschwand, und der Computer rekonstruierte es durch eine neue Zeichnung auf der Leinwand (siehe Abbildung 22).

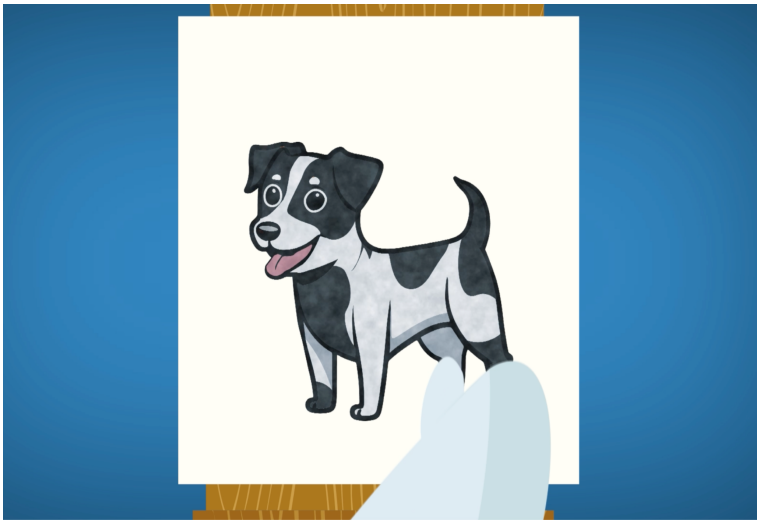


Abb. 22: zweiter Versuch der Darstellung von Diffusion

Beide Ansätze erwiesen sich jedoch als zu weit von der eigentlichen Funktionsweise des Diffusionsmodells entfernt. Erste Rückmeldungen von außenstehenden Personen machten deutlich, dass die Metapher zwar visuell verständlich war, doch zu abstrakt und zu weit weg von dem, was eigentlich bei dem Prozess passiert. Somit war das neue Ziel, die Diffusion doch realistischer darzustellen.

Daraufhin wurde eine zweite Iteration umgesetzt: Das Bild wurde – dem realen Vorgang nachempfunden – in After Effects verrauscht, indem es schrittweise verpixelt wurde. Diese Herangehensweise kam dem tatsächlichen Ablauf bereits näher, war jedoch visuell noch nicht überzeugend.

In einer dritten und finalen Iteration wurde schließlich eine authentische Darstellung des Diffusionsprozesses integriert: Mit Unterstützung eines Kommilitonen konnten die originalen Zwischenschritte aus Stable Diffusion exportiert und direkt in After Effects eingebunden werden (siehe Abbildung 23). Dadurch zeigt das Video nun den Prozess der Bildgenerierung auf Basis eines Diffusionsmodells realitätsnah und nachvollziehbar – und bildet damit die technische Grundlage deutlich präziser ab als die früheren Entwürfe.

Abb. 23: finale Version der Darstellung von Diffusion

5.5.2 Trainingsphase verdeutlichen

In der ersten Version des Videos wurde der Begriff *Training* oder *Trainingsphase* nicht erwähnt – also der Teil, in dem die KI mit Daten (in diesem Falle mit Bildern) trainiert wird. Der Prozess des Trainings wird zwar Teil des Videos und wird auch vorgestellt, aber nicht mit dem Wort *Training* deklariert. Somit wurde auch nicht klar, ab wann das Training beginnt, wann es aufhört und wann die Bildgenerierung startet.

Um dieses Problem zu lösen wurden visuell als auch auditiv zwei Dinge angepasst: Die Mentorin Ruby erwähnt zu Beginn des Videos, dass sie nun mit der Trainingsphase beginnen und passend dazu ändert sich die Hintergrundfarbe von Blau zu Rosa und ein Textoverlay mit dem Wort “Trainingpase” wird eingeblendet.

5.6 Audio

5.6.1 Voice over

Für das Voiceover des Erklärvideos wurde die Text-zu-Audio-KI *ElevenLabs* verwendet. Damit konnte eine Stimme erzeugt werden, die klar, natürlich und passend für den Charakter klingt. Es wurde eine Frauenstimme mit guter englischer Aussprache gewählt, um die Inhalte freundlich und verständlich zu vermitteln. Durch die Nutzung der KI ist es möglich, die Stimme genau an den Text und das Video anzupassen und auch im Nachhinein noch Änderungen vorzunehmen. So können Textpassagen flexibel überarbeitet und die entsprechenden Audiospuren jederzeit aktualisiert werden.

5.6.2 Soundeffekte

Wie bereits in Kapitel 6.1 erwähnt, soll das Erklärvideo sowohl auf visueller als auch auf auditiver Ebene ansprechend gestaltet sein. Daher wurden gezielt Hintergrundmusik sowie Soundeffekte eingesetzt, um die audiovisuelle Wirkung zu verstärken.

Die verwendeten Audioelemente stammen von der Plattform *Epidemic Sound*, die nach Registrierung Zugang zu einer Sammlung von lizenzfreier Musik und Soundeffekte bietet, die für Veröffentlichungen genutzt werden dürfen.

Um das Video auch auditiv dynamisch zu gestalten, kamen unter anderem sogenannte „Swoosh“-Geräusche sowie andere Soundeffekte zum Einsatz, die bestimmte visuelle Aktionen, wie das Umdrehen von Karten oder den Rauschprozess auf Bildern, unterstreichen.

6. Evaluation

6.1 Empirische Studie

Das animierte Erklärvideo wurde am 23. Februar 2025 im Rahmen einer Vorlesung in Medieninformatik I den Studierenden an der Universität Bremen vorgestellt. An diesem Tag stand das Thema *Bildgenerierung durch Künstliche Intelligenz* im Mittelpunkt der Lehrveranstaltung.

Im Anschluss an die Vorführung hatten die Studierenden die Möglichkeit, über einen QR-Code an einer Umfrage teilzunehmen. Diese war in zwei Hauptbereiche gegliedert: Zum einen wurden Einschätzungen zur Verständlichkeit und Gestaltung des Videos erfasst, zum anderen wurde das vermittelte Wissen überprüft, also ob die Zuschauer tatsächlich etwas gelernt haben.

Am Ende der Umfrage konnten die Teilnehmenden das Video auf einer Skala von 0 bis 5 Sternen bewerten und zusätzliche Anregungen oder Verbesserungsvorschläge einreichen.

6.2 Ergebnisse

An der abschließenden Umfrage nahmen insgesamt 35 Personen teil.

Zu Beginn wurde erhoben, wie verständlich das Video insgesamt wahrgenommen wurde. 65,7 % der Teilnehmenden gaben an, dass sie das Video als *sehr verständlich* empfanden. Weitere 26,6 % stuften es als *verständlich* ein, während lediglich 5,7 % angaben, das Video sei *teils/teils* verständlich gewesen. Eine negative Bewertung erhielt das Video in diesem Punkt nicht.

Ein zentrales Ergebnis zeigt sich in der Frage nach der grundsätzlichen Wirksamkeit animierter Erklärvideos: 94,3 % der Befragten stimmten der Aussage zu, dass ihnen ein Erklärvideo hilft, komplexe Konzepte besser zu verstehen (siehe Abbildung 24)..

Abb.24: Nutzen von Erklärvideos für das Verständnis (Quelle: eigene Darstellung auf Basis von Google Forms)

Auch die gestalterischen Mittel wurden positiv bewertet: 80 % der Teilnehmenden gaben an, dass die Visualisierungen im Video ihnen konkret beim Verständnis geholfen haben.

Etwas durchmischter fiel das Ergebnis auf die Frage: “Auf einer Skala von 1-5, wie sehr haben dir die Metaphern im Video geholfen, die Konzepte besser zu verstehen?”.

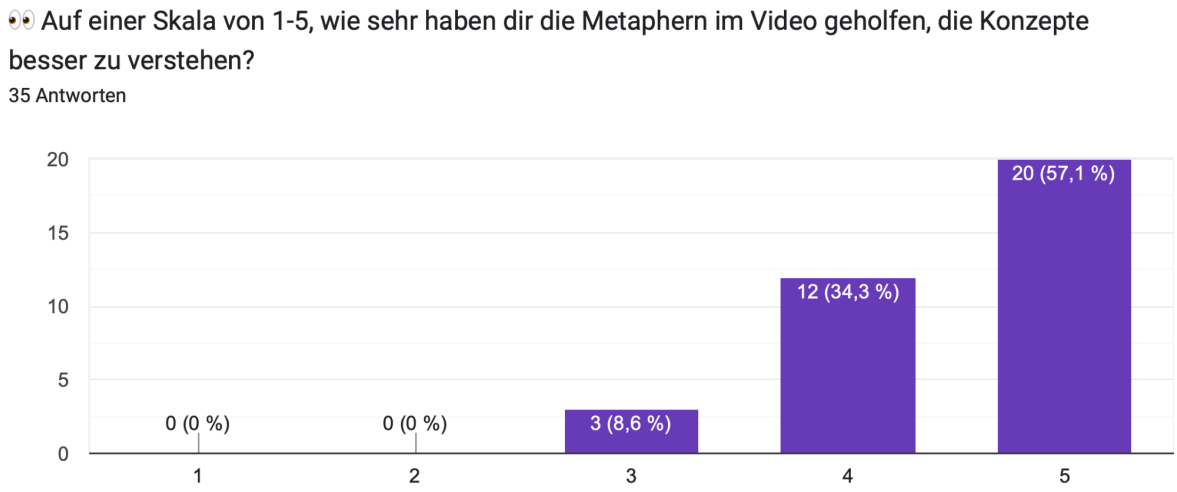


Abb. 25: Nutzen von Metaphern in Erklärvideos für das Verständnis (Quelle: eigene Darstellung auf Basis von Google Forms)

Trotz des Einsatzes bildhafter Vergleiche wurden die Inhalte insgesamt als klar vermittelt wahrgenommen: 57,1 % bezeichneten die Begriffe und Konzepte im Video als *sehr klar*, 40 % als *klar* und lediglich 2,9 % bewerteten sie als *neutral*.

Ein besonderes Ergebnis zeigte sich bei der Frage nach der allgemeinen Nachvollziehbarkeit: 97,3 % der Teilnehmenden gaben an, dem Video durchgehend folgen zu können.

Um diese subjektive Einschätzung zur Verständlichkeit weiter zu überprüfen, folgte im zweiten Teil der Umfrage eine Wissensabfrage mit konkreten Verständnisfragen.

Die erste Frage lautete: *„In welchem Raum werden all die Daten, mit denen die KI trainiert, gespeichert?“* – eine offene Frage, die das im Video eingeführte Konzept des „multidimensional space“ prüfen sollte. Der Großteil der Teilnehmenden konnte diesen Begriff wiedergeben. Nur einige wenige wählten ungenaue Begriffe wie „Raum“, „Speicherraum“ oder „Computer“.

Die zweite Verständnisfrage bezog sich auf den technischen Prozess der Texterfassung: *„Wie kommt die KI dazu, den Prompt zu verstehen?“* Hier gaben 82,9 % die korrekte Antwort *„Die KI übersetzt den Text in einen Vektor“* an (siehe Abb. 25).

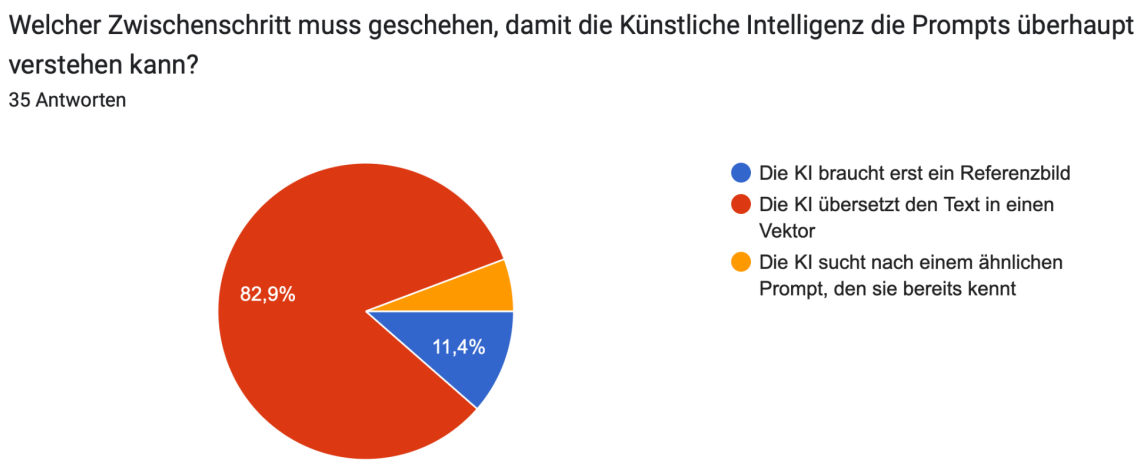


Abb. 26: Verständnisfrage Diffusionsprozess (Quelle: eigene Darstellung auf Basis von Google Forms)

Bei der abschließenden Wissensfrage handelt es sich um eine offene Frage: die Befragten wurden gebeten den Prozess der Bildgenerierung mit KI in bis zu drei Sätzen zu beschreiben – so als würden sie es einer Person erzählen, die noch nie etwas davon gehört hat. Die Ergebnisse waren zu einem Großteil sehr positiv. Hier folgen eine paar Auszüge der Antworten der Befragten:

“Bilder werden Beschreibung zugeordnet dann im multidimesional space abelegt (wie so eine Bibliothek geordnet) und dort sortiert. Dann werden die Bilder alle verrauscht und entrauscht und somit lernt der computer Bilder zu generieren”

“First it needs to be trained with enough data to undertsand a prompt, than it turns the description that we give for the pic into a number and finds the closest matching pic and at the end it generates a new pic from pixels”

“Bilder-Text-Pärchen werden gefunden und abgespeichert. Dann kriegen die Bilder eine Noise oben drauf, bis man sie nicht mehr erkennt. Im Training lernt die KI diese Bilder dann wieder herzustellen.”

“Ich würde sagen, dass die Bild-KI mit vielen Daten (Bilder + Beschreibungen) gefüttert wird, diese werden im multidimsionalen Raum zugeordnet/abgelegt und auf Basis dieser Daten kann die KI, wenn sie ein verrauschtes Bild als Eingabe hat, das Bild entrauschen und das gewünschte Bild (Prompt) erzeugen.”

Am Schluss wurden alle Befragten gebeten dem Video eine Bewertung zwischen 0 (=schlecht) und 5 (=sehr gut) zu geben.

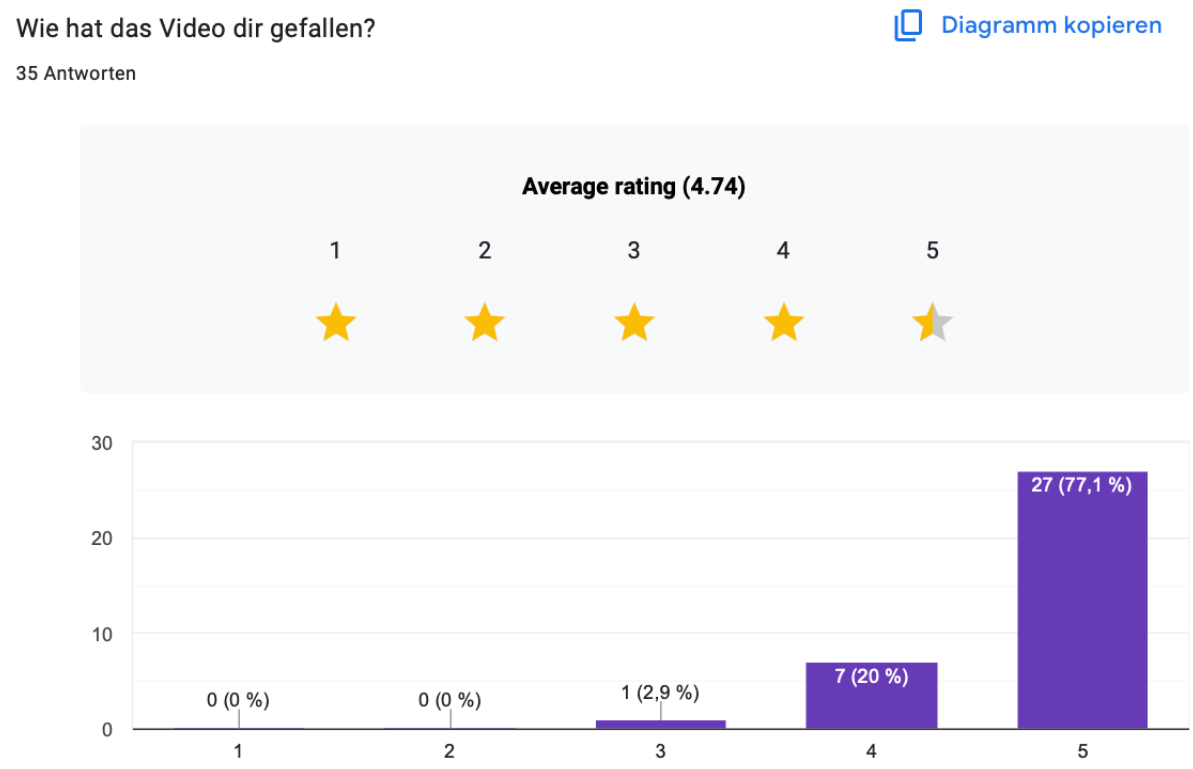


Abb. 27: Bewertung des Erklärvideos (Quelle: eigene Darstellung aus Basis von Google Forms)

6.3 Auswertung

Die quantitativen Ergebnisse legen nahe, dass das Animationsvideo sowohl inhaltlich als auch gestalterisch erfolgreich konzipiert wurde. Besonders die visuellen Darstellungen und Metaphern wurden positiv bewertet: sie halfen den Befragten dabei, die Inhalte klarer zu erfassen. Besonders auffällig ist der nahezu durchgängige Konsens hinsichtlich der Nachvollziehbarkeit des Videos, was für eine gelungene didaktische Struktur spricht.

Insgesamt lässt sich sagen, dass das Video gut gestaltet ist und komplexe Inhalte auf einfache und anschauliche Weise erklärt. Die positiven Rückmeldungen sprechen dafür, dass animierte Erklärvideos ein wirksames Mittel zur Wissensvermittlung sein können.

7. Reflexion

Die Produktion des Erklärvideos “ein Computer lernt malen” dauerte von der Konzeption über die Produktion bis hin zu dessen Evaluation etwa 62 Tage. In dieser Zeit hat OpenAI bereits die nächste Künstliche Intelligenz veröffentlicht: *Sora.*

Sora ist in der Lage, nicht nur Bilder, sondern auch kurze Videos auf Basis von Texteingaben zu generieren. Aktuell können diese Videos maximal 20 Sekunden lang sein, aber wenn uns die Geschichte eines gezeigt hat, dass die Entwicklung von Künstlicher Intelligenz rasant voranschreitet. Vielleicht wird das nächste Erklärvideo, das seinen Platz in der Lehre der Universität Bremen findet, teilweise von Sora oder seinen Nachfolgern produziert oder unterstützt. Die Ergebnisse sehen jedenfalls vielversprechend aus.

Aus der eigenen Arbeit habe ich viele Erkenntnisse gewonnen – am eindrücklichsten war die Erfahrung, wie wichtig das Feedback der Zielgruppe für die Wirksamkeit eines didaktischen Mediums ist. Dinge, die für mich als Produzentin ganz offensichtlich oder verständlich sind, werden von den Zuschauern ganz anders wahrgenommen. Auf der anderen Seite hilft das Feedback ungemein, um die Konzepte aus einem anderen Blickwinkel zu betrachten und das Projekt viel weiter nach vorne zu bringen.

Rückblickend würde ich in einer zukünftigen Produktion noch mehr Feedback aus der Zielgruppe einholen: Was für einen Charakter findet die Zielgruppe besonders ansprechend? Muss das Video eher wie ein kurzes Video, wie auf Social Media sein, damit es reizvoll ist?

Abschließend bin ich froh, dass ich im Vorfeld der Produktion eine Bedarfsanalyse durchgeführt habe, um ein grundsätzliches Bild davon zu bekommen, wie der Wissensstand bei der Zielgruppe ist und ob sie sich überhaupt für das Thema interessieren.

8. Fazit

Die vorliegende Arbeit widmete sich der Frage, wie bildgenerierende Künstliche Intelligenz – insbesondere auf Basis von Diffusionsmodellen – für Studienanfänger verständlich und didaktisch sinnvoll vermittelt werden kann. Durch die Verbindung theoretischer Grundlagen, empirischer Bedarfsanalyse und gestalterischer Umsetzung wurde ein ganzheitlicher Ansatz verfolgt, der sowohl den technischen Hintergrund als auch die pädagogische Aufbereitung berücksichtigt.

Zentrale Erkenntnis der Bedarfsanalyse war, dass bei der Zielgruppe grundlegendes Interesse an KI-basierten Bildgeneratoren besteht, jedoch vielfach Unsicherheiten in Bezug auf deren Funktionsweise, Begriffe und technische Zusammenhänge vorhanden sind. Diese Erkenntnisse bildeten die Grundlage für die Konzeption des Videos

Das entwickelte Erklärvideo wurde so gestaltet, dass es komplexe Inhalte leicht verständlich vermittelt. Es erklärt die technischen Abläufe der KI-Bildgenerierung mithilfe einer klaren, schrittweisen Struktur und nutzt dabei einfache Sprache, visuelle Elemente und gezielte Animationen. Die Erzählweise orientiert sich an der sogenannten Heldenreise: Eine Figur begleitet den Zuschauer durch die einzelnen Lernschritte und veranschaulicht dabei zentrale Begriffe wie Prompt, Vektorraum und Diffusionsprozess.

Besonders hervorzuheben ist die Kombination aus Fachinhalt und erzählerischer Gestaltung, die das abstrakte Thema greifbarer macht. Auch die bewusst reduzierte visuelle Gestaltung sorgt dafür, dass die Aufmerksamkeit auf den Inhalt gelenkt wird. Der reale Ablauf des Diffusionsprozesses wurde möglichst authentisch dargestellt, um ein realistisches Verständnis zu fördern.

Die anschließende Evaluation zeigte, dass das Video von den meisten Studierenden gut verstanden wurde. Viele konnten zentrale Begriffe und Abläufe korrekt wiedergeben. Das zeigt, dass ein animiertes Erklärvideo gut geeignet ist, um komplexe Themen wie die Bildgenerierung mit KI verständlich zu machen.

Insgesamt bestätigt diese Arbeit, dass visuelle und erzählerische Mittel dabei helfen können, schwierige Inhalte in der Hochschullehre anschaulich zu vermitteln. Das Beispiel zeigt auch, wie digitale Medien in der Lehre sinnvoll eingesetzt werden können, um das Lernen zu unterstützen.

KI-gestützte Bildgenerierung ist nicht nur ein technologisch faszinierendes Feld ist, sondern bringt auch neue didaktische Herausforderungen mit sich. Die Verbindung von Theorie und kreativer Wissensvermittlung bietet großes Potenzial, um komplexe Inhalte einem breiteren Publikum zugänglich zu machen.

9. Quellenverzeichnis

9.1 Literaturquellen

Abercrombie, Cortnie: *KI: Wenn wir wüssten...: Was künstliche Intelligenz alles über uns weiß und was wir über sie wissen sollten.* Plassen Verlag. 2022

Becker*, Sebastian & Brehmer, Jana: *Erklärvideos
...als eine andere und/oder unter- stützende Form der Lehre. 2017
Online verfügbar: https://www.uni-goettingen.de/de/document/download/5d0fa49e220547bded74a21f21d44fc0.pdf/03_Erklärvideos.pdf (Stand 04.05.2025)

Bostrom, **Nick**: *HOW LONG BEFORE SUPERINTELLIGENCE?* 1997
Online verfügbar:
https://www.cs.ucf.edu/~lboloni/Teaching/CAP5636_Fall2023/homeworks/Reading%202%20-%20Nick%20Bostrom-How%20long%20before%20superintelligence.pdf (Stand 10.04.2025)

Dahm, Markus H. & Zehnder, Valentin: *Moderne Personalführung mit Künstlicher Intelligenz. Chancen und Risiken*. Wiesbaden: Springer Gabler. 2023

Engelke, Barbara & Engelke Urlich*: ChatGPT – Mit KI in ein neues Zeitalter.* mitp Verlag. 2024

Goodfellow, Ian J. et al.: Generative Adversarial Nets. 2014.
Online verfügbar: https://arxiv.org/pdf/1406.2661 (Stand 10.04.205)

Hecker, Dirk & Paaß, Gerhard: *Künstliche Intelligenz: Was steckt hinter der Technologie der Zukunft?* Springer Vieweg. 2020

Konecny, Jaromir: *Ist das intelligent oder kann das weg?* LMV. 2020

Pooja M M: *DALL-E 3: Advanced AI image generation model*. 2025
Online verfügbar:
https://d197for5662m48.cloudfront.net/documents/publicationstatus/253506/preprint_pdf/4cb2f05bc756940dd11fd314e70de37d.pdf (Stand 09.04.2025)

Santner, Christoph: „*Alles KI? – Die neue Welt der Künstlichen Intelligenz verstehen und nutzen,* Goldmann, 2024

Uria-Recio, Pedro: „*Wie KI unsere Zukunft gestalten wird: Künstliche Intelligenz verstehen, um einen Schritt voraus zu sein.* Verlegt von Pedro Uria-Recio. 2024

9.2 Internetquellen

Backlinko – ChatGPT / OpenAI Statistics: How Many People Use ChatGPT?
https://backlinko.com/chatgpt-stats (Stand 15.04.2025)

Bundeszentrale für politische Bildung – Was ist KI und welche Formen von KI gibt es? 2024
https://www.bpb.de/lernen/bewegtbild-und-politische-bildung/555997/was-ist-ki-und-welche-formen-von-ki-gibt-es/ (Stand 11.04.2025)

Bibi Blocksberg – Oma Grete
https://www.bibiblocksberg.de/bibis-welt/charaktere/oma-grete?utm_source=chatgpt.com
(Stand 06.05.2025)

Datasolut – Deep Learning: Definition, Beispiele & Frameworks – Wuttke, Laurent – 2023
https://datasolut.com/was-ist-deep-learning/ (Stand 16.04.2025)

Euronews - ChatGPT's viral Studio Ghibli-style images: 'An insult to life itself' - 28/03/2025 - David Mouriquand
https://www.euronews.com/culture/2025/03/28/chatgpts-viral-studio-ghibli-style-images-an-insult-to-life-itself (Stand 28.04.2025)

Epidemic Games - Was ist und wozu benutzt man Discord? – 2022
https://store.epicgames.com/de/news/what-is-discord-and-what-is-it-used-for
(Stand 22.04.2025)

Github – Stability-AI/**stablediffusion**
https://github.com/Stability-AI/stablediffusion (Stand 01.05.2025)

Ludwig-Maximiloams-Universität München - Theorie zum multimedialen Lernen nach Mayer
https://www.didaktik.physik.uni-muenchen.de/multimedia/lernen_mit_multimedia/psycho_theo/multimedia_mayer/index.html
(Stand 06.05.2025)

Masterclass Writing 101: What Is the Hero's Journey? 2 Hero's Journey Examples in Film
https://www.masterclass.com/articles/writing-101-what-is-the-heros-journey (Stand 07.05.2025)

MathWorks – Die Bedeutung von neuronalen Netzen
https://de.mathworks.com/discovery/neural-network.html (Stand 15.04.2025)

Mebis Magazin – Die Geschichte der künstlichen Intelligenz

https://mebis.bycs.de/beitrag/ki-geschichte-der-ki#sec2 (Stand 03.04.2025)

NeuroNation – Intelligenz: Was ist das genau?

https://www.neuronation.com/science/de/definition-der-intelligenz-was-ist-das-eigentlich/

(Stand 15.05.2025)

Nöcker-Prior, Philipp: Bildgenerierung mit Künstlicher Intelligenz, 2023
Online verfügbar: https://ai.hdm-stuttgart.de/downloads/student-white-paper/Winter-2223/Bildgenerierung_mit_KI.pdf (Stand 23.04.2025)

OpenAI – CLIP: Connecting text and images – 2021
https://openai.com/index/clip/ (Stand 25.04.2025)

Robominds – Die Geschichte der KI: von der Turingmaschine bis Deep Learning – 2024
https://www.robominds.de/blog/die-geschichte-der-ki-von-der-turingmaschine-bis-deep-learning (Stand 03.04.2025)

Stability.ai – Stable Diffusion Launch Announcement – 2022
https://stability.ai/news/stable-diffusion-announcement (Stand 15.04.2025)

Spektrum - Was machen TikTok & Co mit unserem Gehirn? – 2023
https://scilogs.spektrum.de/hirn-und-weg/was-machen-tiktok-co-mit-unserem-gehirn/ (Stand 06.05.2025)

Technische Hochschule Würzburg-Schweinfurt – Schwache vs. Starke KI
https://ki.thws.de/thematik/starke-vs-schwache-ki-eine-definition/ (Stand 11.04.2025)

IBM – Was ist ein Transformator-Modell?
https://www.ibm.com/de-de/think/topics/transformer-model (Stand 10.04.2025)

BM – Bergmann, Dave & Stryker, Cole – Was sind Diffusionsmodelle? – 21.08.2024
https://www.ibm.com/de-de/think/topics/diffusion-models (Stand 01.05.2025)

Internationale Hochschule Akademie – Was ist Prompting? – 2023
https://www.iu-akademie.de/blog/was-ist-prompting/ (Stand 29.04.2025)

Epidemic Sound - https://www.epidemicsound.com/music/featured/?override_referrer=

10. Nutzung KI basierte Anwendungen

1. **Audiogenerierung mit Elevenlabs**

Die Text -Audio-KI Elevenlabs wurde für dieses Projekt genutzt, um das Voice Over für das Video zu erstellen.

Link: https://elevenlabs.io/de

2. **Bildgenerierung mit DALL- E 3**

In der Einleitung der Arbeit wurde ein Bild eingesetzt, das innerhalb von ChatGPT durch DALL-E 3 generiert wurde.

Link: https://chatgpt.com/

3. **Definieren mit ChatGPT 4o**

In dem Kapitel 3.3 wurde die Künstliche Intelligenz ChatGPT genutzt, um den Begriff “Bild” zu definieren.

Link: https://chatgpt.com/

4. **Grammatik mit ChatGPT-4o**

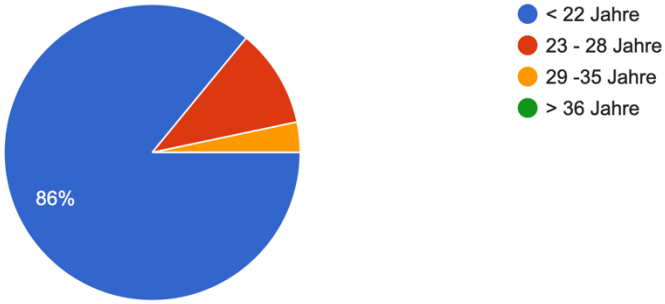
ChatGPT (OpenAI, GPT-4o, Stand: Mai 2025) wurde punktuell zur grammatikalischen Überarbeitung und als Formulierungshilfe genutzt.

Link: https://chatgpt.com/

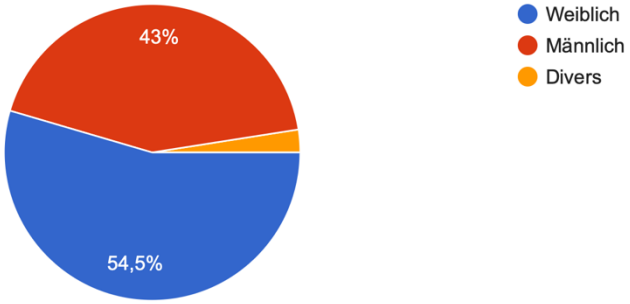
11. Anhang

11.1 Ergebnisse der Bedarfsanalyse

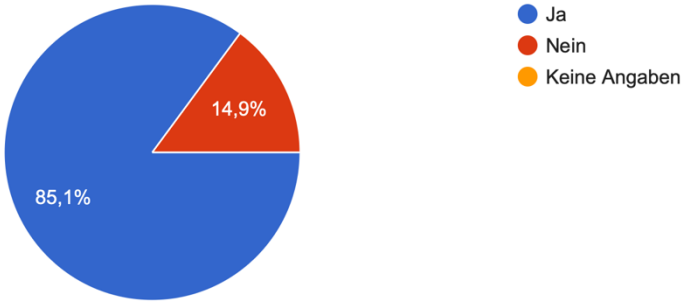
Wie alt bist du?
121 Antworten



Dein Geschlecht:
121 Antworten

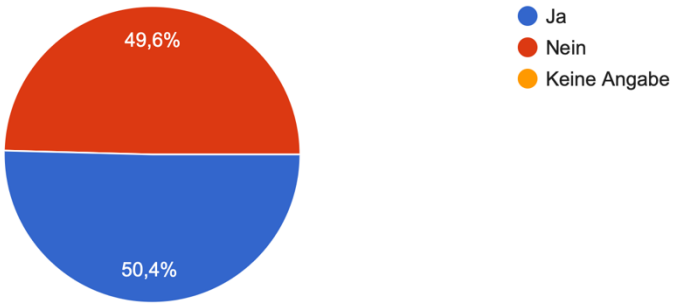


Hast du schonmal ein Tool mit Künstliche Intelligenz genutzt (z.B. ChatGPT etc.)? 💻
121 Antworten



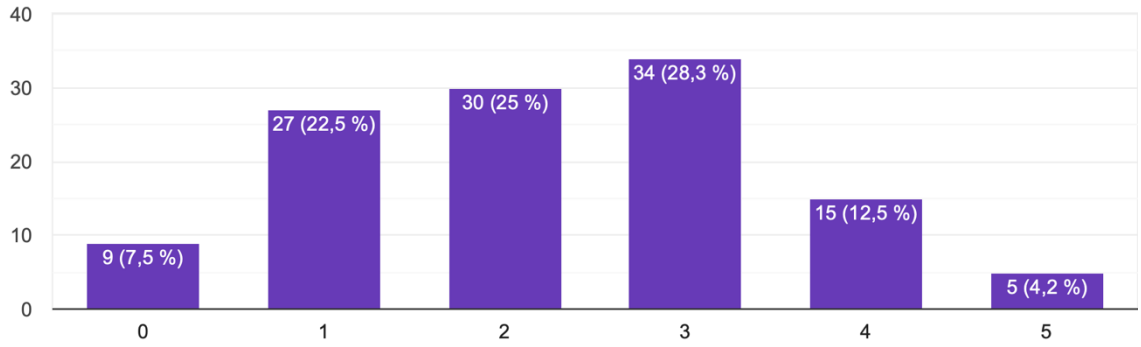
Hast du schonmal mit Hilfe einer Künstlichen Intelligenz ein Bild generiert? 📸

121 Antworten



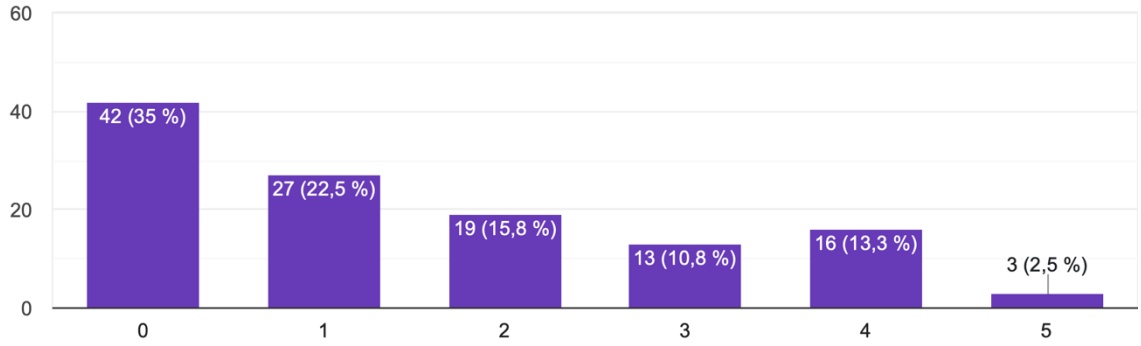
Wie gut kennst du dich mit Künstlicher Intelligenz aus? 🤔

120 Antworten



Wie gut kennst du dich mit KI-Bildgeneratoren aus? 🖼️

120 Antworten



Was weisst du über bildgenerierende Künstliche Intelligenz?

Antworten:

1.Nichts

2.Sie können keine Hände darstellen

3. Die Bilder die generiert werden basieren auf tausende aus dem Internet gefunden Bilder zu gewissen Key-Wörtern. Noch ist sie nicht 100% genau und man kann gut erkennen was mit KI generiert wurde, doch die Grenzen verschwimmen immer mehr.

4. Sie nutzen eine große Datenbank anderer Bilder und generieren zu meinem Input daraus ein am wahrscheinlichsten passendes Bild

5. es ist mit Computer Vision verbunden (und alles wie convolution, filter...)

6. Dall-E, Stable Diffusion, Midjourney, Local Inference, Technologie hinter Stable Diffusion, Modelle, Training

7. Wurde mit Millionen von Bildern trainiert

8. noch nicht so viel

9. Sie Nutzten bereits besthende Bilder, werten diese aus und versuchen auf deren Grudlage ein neues Bild zu erschaffen.

10. Man gibt vorgaben und die Ki generiert einen mehrere Beispiele

11. Das sie mittlerweile sehr reale Bilder erstellen kann

12. Anwendung, Funktionsweise, Einschränkungen

13. Zu wenig, Hintergründe oder kleiner Details sind aber manchmal fehlerhaft.

14. Das die meisten Künstler nicht darüber erfreut sind

15. Sie generieren Bilder je nach positive und negative prompt die jeweiligen Modelle können Bilder nach verschiedenen Kriterien generieren.

16. Es werden für die gängigen Modelle wie Midjourney und Co riesige Mengen an Bildmaterial für das Training benötigt. Diese Bilder werden meist ohne Zustimmung der Urheber genutzt.

17. Ich habe mich vor allem damit beschäftigt, mit welchen Informationen KIs trainiert wurde. Ob es ein eigen Werk ist, Kreativität gennant werden kann, urheberrechtliche Probleme

18. Das sich die KI Bilder aus dem gesamten Internet raussucht, die passende Schlagwörter enthält und daraus dann ein Bild mit den vorgegebenen Schlagwörtern generiert

19. Es lernt durch verschiedene im Internet verfügbare Informationen.

20. Ich weiß, wie man die richtigen „Befehle“ formuliert, um der KI den richtigen Input zu geben und ein Ergebnis zu erhalten, mit dem man zufrieden ist.

21. In einer Datenbank sind Wörter mit Bilder verbunden. Durch trainieren der KI lernt sie verschiede Wortzusammenstellungen auch als Bild zusammenzustellen

22. Sie arbeitet mit einem vorgegebenen Datensatz und versucht dann anhand von prompts das zu generieren was verlangt wird.

23. Es wird immer verbreiteter.

24. ich bin nur nutzer

25. Es gibt beliebte Programme dafür

26. Benutzen Bilder aus dem Internet und verarbeiten diese zu neuen?

27. Die KI muss vorher mit Informationen (z.B. vielen Bildern, Daten) gefüttert werden.

28. generieren Bilder durch das angeben von stichwörtern durch Personen. dies ist möglich weil die KI zuvor durch die große Datenbank der Bilder im Netz die Fähigkeit erlernt hat wörter mit bildern zu verbinden

29. Etwas wie sie funktionieren, sie werden zb trainiert mit schon vorhanden bildern(zb Hundebilder), die KI guckt sich bzw über 1000 bilder an die im internet vorhanden sind und wird dann darauf trainiert die merkmale zuerkennen und wieder zu geben. Aber so im Detail kenne ich mich nicht aus.

30. Das die KI aus bereits existierenden Bilder neue macht, wie eine professionelle Collage

31. Es gibt verschiedene Engines oder Models mit denen man Bilder generieren kann (z.B. Stable Diffusion oder Redshift). Die Engins haben Unterschiede wie z.B. Stiel. Dazu kann man auch Sachen wie LoRA, Control, Stärke oder den Seed einstellen. Bei verschiedenen Seeds kommt mit den gleichen Angaben ein unterschiedlicher Output. Es gibt aber auch Ki-Systeme, mit denen man z.B. die Gesichter von Personen in Videos oder live vor Kamera ändern kann.

32. Es wird eine große Datenbank an Bildern bezogen, um neue Bilder zu generieren

33. Es nimmt schon existierende Bilder und trainiert dadurch und ist deswegen Urheberrechtlich kritisch.

34. Nichts wirklich

35. Das Bild welches generiert werden soll kann mithilfe von Beschreibungen (Wörtern) generiert werden. So genauer die Beschreibung ist so genauer ist auch das generierte Bild.
36. KI bilden aus einem Vorrat an verschiedenen Bildmaterial, entweder aus dem Internet oder speziell ausgesuchtes, und werden damit trainiert indem z.b. viele Ölgemälde vorgegeben werden und als Input der Begriff Ölgemälde damit verknüpft wird, damit eine KI mithilfe eines Prompts etwas generieren kann, was eine Mischung aus allen in der Art ist

37. Logos erstellen, Presentationen uzw.

38. Benutzen entweder das Internet an sich oder eine große datenbank and referenzen um Bilder zu generieren

39. nicht viel

40. Funktioniert mit neuronalen Netzwerken. Größte Vertreter z.B. Midjourney oder DALL-E. Trainiert mit Millionen von Bildern aus dem Internet. Möglichkeit mit wenigen Worten ein noch nie dagwesenes Bild in Sekunden zu erschaffen

41. So gut wie nichts

42. Beginnt mit Noise und generiert daraus nach und nach das Bild, basierend auf der Eingabe und den Trainingsdaten

43. dass man mit wenig Aufwand ein Foto bearbeiten und daraus eine völlig neue Kreation machen kann. Wenn Sie früher stundenlang in Photoshop damit arbeiten mussten, übernimmt KI das jetzt

44. nicht viel höchstens NFTS

45.Bei einer bildgenerierenden KI ist man in der Lage durch Beschreibungen des gewünschten Bildes, ein solches Bild zu erstellen. Ebenso kann man ein Bild damit „erweitern“

46.Dass sie auf bereits bestehenden Bildern basiert und diese beim generieren umwandelt/anpasst

47.Es gibt einige KI, die beschriebene Wörter in Bilder umwandeln.

48.Bildgenerierende KIs sind Software, die u.a. über die text-to-image Funktion, wie es der Name bereits sagt, Bilder generieren. Mit Text prompts wird der KI eine Beschreibung gegeben, nach der sie ein Bild konstruiert. Die KI nutzt hierfür einen Datensatz, mit dem sie „trainiert“ wurde. Im Falle von Bildgenerierender KI wären dies Bildquellen, denen bestimmte Charakteristiken zugeordnet wurden.

49.Eine bildgebende KI wird mit Tausenden von Bildern bestimmter Objekte oder "Ideen" gefüttert. Der Schöpfer der KI gibt ihr z. B. eine Menge verschiedener Autobilder und sagt ihr, dass diese "Auto" heißen, damit sie die "Idee" eines Autos versteht. Nachdem sie mehrere

Ideen gelernt hat, kann sie diese sogar miteinander kombinieren. Wenn sie also gelernt hat, was die Farbe "Rot" und was "Auto" ist, kann man "Rotes Auto" eintippen, und sie gibt einem die Kombination. Und so weiter.

50.Sie entwickelt anhand von Vergleichs Bilder im Internet ein gefragt es Idealbild.

51.Kann neue Bilder zu einem bestimmten Thema erstellen, z.B. year book trend

52.Sie basieren auf anderen Bildern

53.KI Bildgeneratoren eine schnelle und kostenlose Methode, Grafiken zu produzieren. Diese sind jedoch nicht zwingend gut und haben häufig Ähnlichkeiten mit bestehenden Grafiken. Bei komplexeren Darstellungen ist es meist sehr aufwendig, genau das Bild zu bekommen, was man haben will/sich vorstellt.

54.nicht wirklich

55.Dass das Bild einmalig ist (also eigentlich)

56Bereits existierenden Bildern werden als samples verwendet für die neu generierten Bilder. Es sind Modelle, die an großen Datenmengen, also Bildern, trainiert wurden hinsichtlich ihrer Erkennung von Strukturen. Ansonsten kenne ich das buzzword "Deep Learning" ohne tiefere Kenntnisse.

57.Mit DALL-E von Open AI habe ich mehrere Versuche unternommen, die gewünschte Referenzbildtafel zu erstellen, wobei ich verschiedene algorithmische Wörter ausprobiert habe, um herauszufinden, wie ich das Aussehen und die Wirkung des gewünschten Bildes erreichen kann. Ich habe auch Midjourney verwendet, und in letzter Zeit habe ich mich mit KI zur Videogenerierung beschäftigt.

58.in der Lage lustige Bilder zu erstellen, erzeugte Bilder gerade realitätsnahe Bilder sehen häufig noch unecht aus

59.man sollte sie am besten kennzeichnen

60.Die KI wird mithilfe vorhandener Bilder trainiert und sucht über die Beschreibung des zu generierenden Bildes nach Ähnlichkeiten mit anderen (ihr bekannten) Bildern. Daraus generiert sie ein neues Bild.

61.Die Bilder werden von einer KI erstellt, die mit sehr großen Datasets trainiert wurden. Die KI versucht dann, Bilder mithilfe der Trainingsdaten so zu erstellen, so dass der Stil der Bilder imitiert wird.

62. nichts

63.Gar nichts

64.Ich habe aufmerksam die Entwicklungen sowohl im Erstellen von Bilder mit Ki, sowie Erkennung von KI generierten Bildern verfolgt.

65.Das man Anhand von spezifischen Eingaben ein komplettes Bild erstellen kann

66. Ich weiß dass die KI auf eine vordefinierte Datenbank von menschlich generierten Bildern zugreift und diese als Referenz nutzt. Auch weiß ich dass man bei der Angabe von Definitionen und Beschreibungstexten sehr generisch und akribisch vorgehen. uss da die Ki lange oder komplizierte Beschreibungen nicht immer genau versteht. Stichpunkte sind der schlüssel.

67. Sie nehmen Referenzbilder und mischen sie zusammen bzw. kreieren dadurch neue

68.In vielen Fällen werden Bilder zum trainieren der KI gestohlen. Als Abwehrmethode gibt es das sogenannte glaze, dass verhindert, dass die AI Bilder replizieren kann. Es gibt verschiedene Generatoren u.a. open source stable diffusion xl. Um ein gewünschtes Bild zu erhalten, kann man verschiedene promts verwenden, dabei ist in persönlicher Erfahrung weniger manchmal mehr.

68. Erstellung von visuellen Darstellungen

69.Internet

70.Kaum etwas

71.Sie kopiert keine Bilder. Wenn man z.B ein Hund bild generiert haben will, macht er das alleine Lernen, indem sie Daten aus vorhandenen Bildern sammeln

72.Nicht viel

73. bildgenerierende Künstliche Intelligenzen klauen Kunst von Künstler*innen.

74. erstellt ein Bild, dass durch eine Aussage vorher beschrieben wurde. Wurde mit Milliarden von Texten und Bildern trainiert, mit denen sie die gegebene Aussage vergleicht, um dann ein vergleichsmäßig passendes Bild zu generieren.

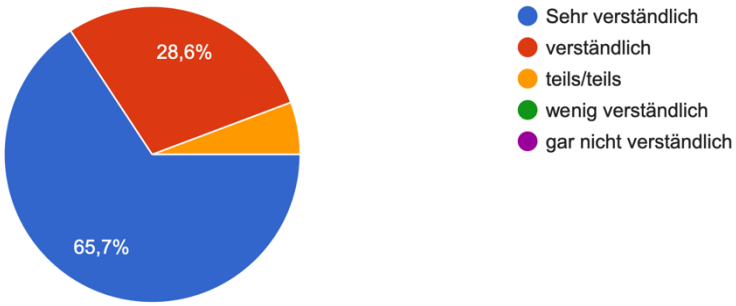
75. Werden von Milliarden an Bildern trainiert und sind seitdem sie öffentlich verfügbar sind (ca. 1,5) Jahre um ein vielfaches besser geworden. Es gibt open-source Versionen die man auch lokal auf seinem PC laufen lassen kann. Für die Generierung wird die Leistung der Grafikkarte beansprucht. Je höher die VRAM desto besser.

76. Bei manchen Programmen kann man wohl Anweisungen eingeben, wie das Bild generiert werden soll

11.2 Ergebnisse der Umfrage zu der Verständlichkeit des Erklärvideos

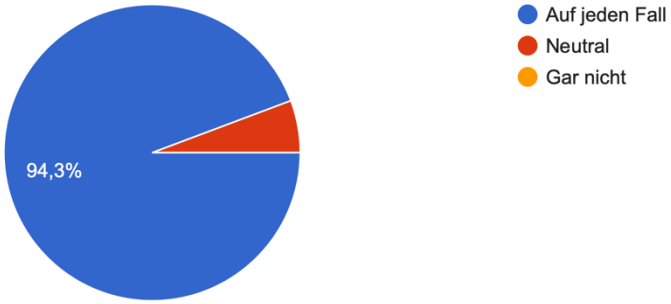
👉🏼 Wie verständlich findest du das Video?

35 Antworten



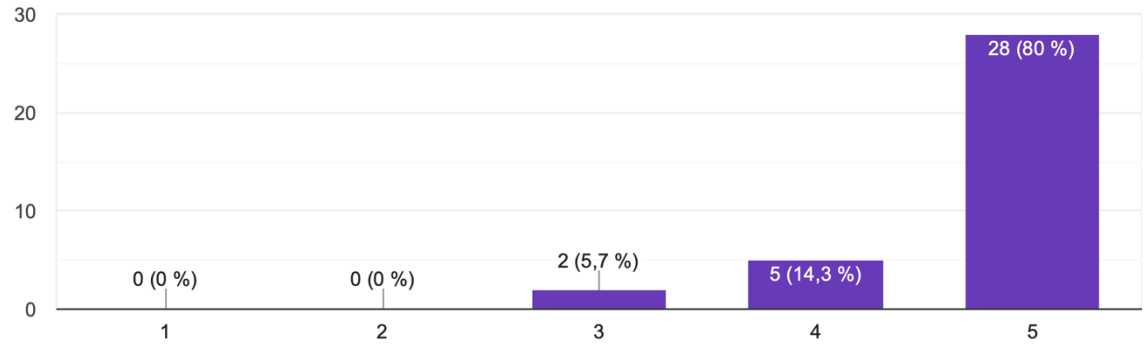
🤔 Hilft es dir, das Konzept zu verstehen, wenn du es in Form eines animierten Erklärvideos schaust?

35 Antworten



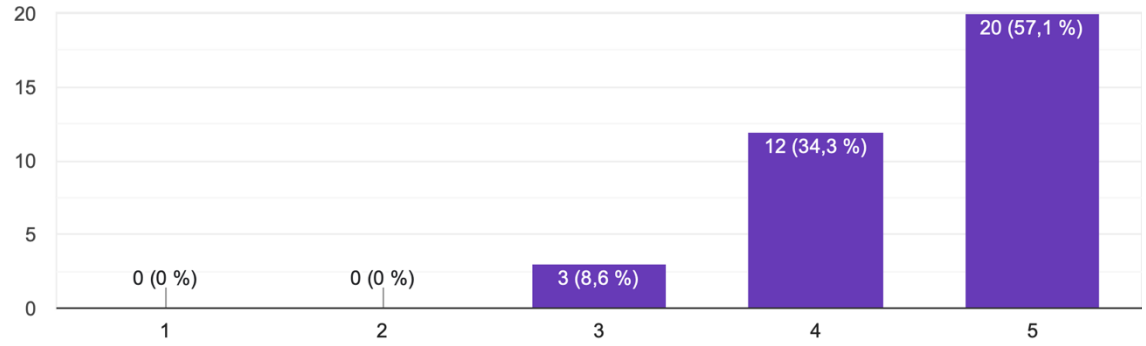
🖼️ Auf einer Skala von 1-5, wie sehr haben die die visuellen Darstellungen im Video geholfen, die Konzepte besser zu verstehen?

35 Antworten



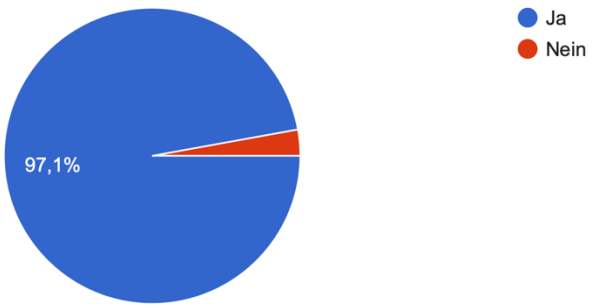
👀 Auf einer Skala von 1-5, wie sehr haben dir die Metaphern im Video geholfen, die Konzepte besser zu verstehen?

35 Antworten



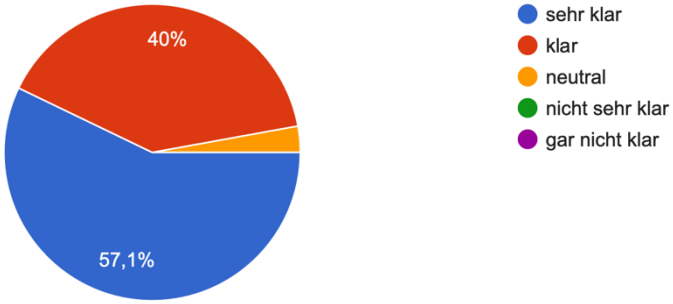
🤸 Konntest du dem Video die gesamte Zeit über folgen?

35 Antworten



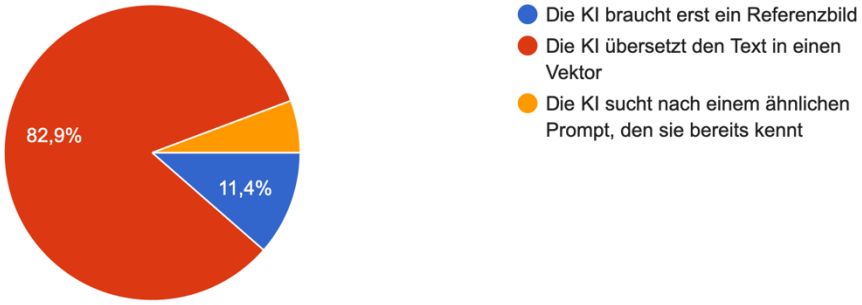
🔎 Wie klar waren die im Video verwendeten Begriffe und Konzepte erklärt?

35 Antworten



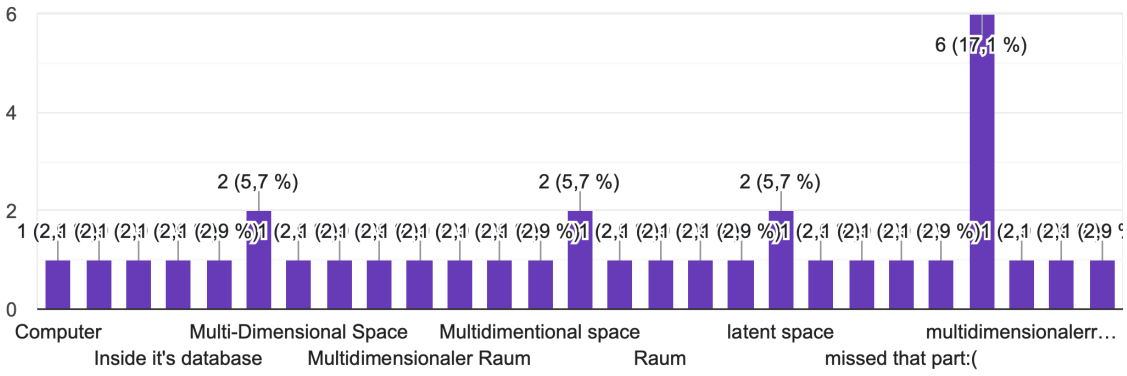
Welcher Zwischenschritt muss geschehen, damit die Künstliche Intelligenz die Prompts überhaupt verstehen kann?

35 Antworten



In welchem Raum werden all die Daten, mit denen die KI trainiert, gespeichert?

35 Antworten



Wenn du einer 3. Person, das Konzept von KI-Bildgenerierung in 3 Sätzen erklären müsstest - was würdest du sagen?

Antworten:

1. 1. Daten sammeln 2. Paar bilden 3. Paare abspeichern 4. Rauschen von Bildern entfernen lernen 5. Selber Bilder generieren, aus reinem Rauschen

2. AI gets better by looking at lots of pictures, once the AI understand the pattern, it can turn text into images.

3. beschreiben das Bild, das KI bilden sollte, verstauschen Trainingbild künstlich und machen das Trainingbild rauschfrei

4. It's like teaching a computer to draw using words, the AI will try to generate examples based on what it learned, it draws new images based one the context giver.

5. Bilder und Texte werden miteinander verbunden und dann im latent space abgespeichert. Danach werden die Bilder verpixelt und wieder entpixelt: das ganze tausende Male. Danach kann die KI selbst Bilder erstellen aus reinen Pixeln.

6. Es muss trainiert werden mit Sätzen und einem verpixelten Bild, damit die KI Ansatzpunkte für Kanten hat. Um die KI selbstständig ein Bild erstellen zu lassen aufgrund von Text, kann man ihm dadurch helfen.

7. Bilder werden Beschreibung zugeordnet dann im multidimesional space abelegt (wie so eine Bibliothek geordnet) und dort sortiert. Dann werden die Bilder alle verrauscht und entrauscht und somit lernt der computer Bilder zu generieren

8. Die KI muss erst durch den Menschen trainiert werden. Dann kann sie anhand von Ähnlichkeiten Bilder generieren, da das Bild mit dem Text einen ähnlichen Vektor hat, Vektoren sind Zahlen mit denen der Computer kommuniziert

9. Ein KI-Bildgenerator mit Diffusionsmodell startet mit einem zufälligen Bildrauschen und verbessert es Schritt für Schritt, bis ein realistisches Bild entsteht

10. First it needs to be trained with enough data to undertsand a prompt, than it turns the description that we give for the pic into a number and finds the closest matching pic and at the end it generates a new pic from pixels

11. KI kriegt Daten zum üben, damit besitzt sie Wissen. Dieses Wissen benutzt sie um Bilder zu enttäuschen und richtige Bilder zu erstellen, nach Texteingabe.

12. Bilder enttäuschen mit Trainingsdaten vom Anfang

13. It sees a lot of photos and learns patterns, use these patterns to start with random noise and changes it step by step, keeps adjusting until the result keeps improving.

14. AI trains to learn what things should look like, then give it a prompt, generate a image based on what the AI guessed it should be.

15. dass die KI mit Bilder trainiert wird und dort verrauscht und dann entrauscht wird. Und am Ende kann die KI aus reinem Rauschen neue Bilder generieren.

16. Bilder-Text-Pärchen werden gefunden und abgespeichert. Dann kriegen die Bilder eine Noise oben drauf, bis man sie nicht mehr erkennt. Im Training lernt die KI diese Bilder dann wieder herzustellen.

17. hat was mit Bilder und Beschreibung zu tun, die man zusammenführen muss und dann mit bekommen die Bilder so eine "Noise", die wieder entfernt wird

18. the ai will be trained with a bunch of date: pictures and descriptions. This data will be stored in the latent space. Afterwards it practices to add noise to the data and then remove this noise - over and over again and in the end, the ai is capable of creating a picture from random noise when getting a prompt. :)

19. Ich würde sagen, dass die Bild-KI mit vielen Daten (Bilder + Beschreibungen) gefüttert wird, diese werden im multidimsionalen Raum zugeordnet/abgelegt und auf Basis dieser Daten kann die KI, wenn sie ein verrauschtes Bild als Eingabe hat, das Bild entrauschen und das gewünschte Bild (Prompt) erzeugen.

20. Die KI trainiert mit einem Datensatz an Bilder und Texten. All diese Bilder werden dann nach Kategorien sortiert, nachdem sie in computer-mäßige Sprache übersetzt werden = Zahlen bzw. Vektoren. Dann können die Bilder einmal verrauscht werden und wieder entrauscht werden. Über diesen Prozess lernt die KI Bilder selbst aus reinem Rauschen zu generieren.

21. Die KI hat ein Set von Trainingsdaten mit Prompts und Bildern, die zusammengehören, was die KI durch die Gleichheit von Vektoren, die Prompt und Trainingsbild darstellen, erkennt. Die werden dann in einem Multi-Dimensional Space gespeichert. Bei der KI-Bildgenerierung wird dann ein rauschendes Bild genommen und die KI fügt Details, die sie aus den Trainingsdaten kennt, hinzu, sodass ein Bild entsteht.

22. Bilderpaare werden gebildet, dann werden die Paare nach Kategorien sortiert und am Schluss werden sie genoised und entnoised und mit diesem Prozess lernt das System, selbst Bilder zu generieren aus reiner Noise.

23. It sees many images, learns the patterns, try to generate images following the patterns.

24. das könnte ich nicht denke ich

25. The AIs study the images received, it takes a sentence and creates its version of what it should look like, it keeps improving repeting this process multiple times.

26. The AI learns from a other pictures and words. When you give it a sentence, it will try to figure out what it should look like, the AI will use it's current data base to generate a new image.

27. KI lernt mit Trainingsdaten Bilder zu deuten und muss dann im Generierungsprozess das Rauschen in ein Bild verwandeln.

28. The AI trained with images and descriptions, it tries to guess how it should look like, repeat the process until the result is good.

29. Bilder werden gepaired und dann abgespeichert und am Schluss kriegen sie rauschen und dann wird das entfernt. Am Schluss lernt die KI über diesen Prozess und kann letztlich eigene Bilder aus Rauschen erstellen

30. After the AI being exposed and trained on many visual and texts, it interprets the content to guess the result. It adds noises when generate new images.

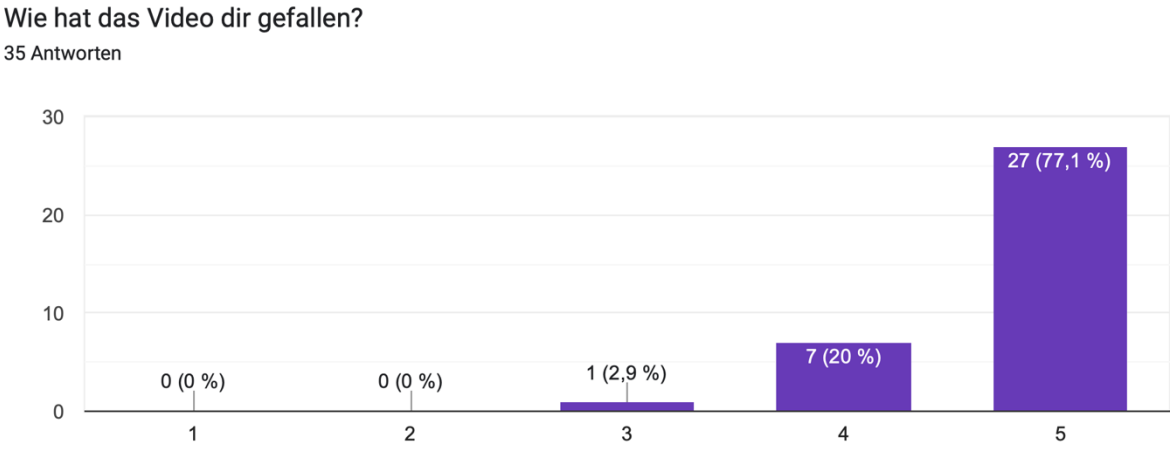
31. Erst Bilder und Text zusammenhänge lernen. Dann lernen Rauschen aufzulösen anhand von bereits gelerntem. Zum Schluss dann Rauschen von einem Bild entfernen und ein passendes Bild zum Prompt erstellen, ohne dass dieses Bild eigentlich im Rausch n enthalten war.

32. It learns from a huge picture database, then given a sentence, it starts drawing pictures using what it learned, it keeps adjusting until it looks right.

33. Die KI trainiert mit Bilder + Texten und lernt diese in Zahlen zu übersetzen und nach Kategorien zu sortieren, wie in einer Bibliothek. Danach bekommt jedes Bild Noise und die KI lernt auch, diese wieder zu entfernen. Letztlich kann die KI selbst aus Rauschen Bilder zaubern.

34. First it trains with images and text, second it get some images and try to guess the text and third, it gets the text and try to generate an image based on past experiences that it learned.

35. It connects words and images through practice, trying to create patterns links, it generate images based on previous patterns.



Wie könnte ich das Video verbessern?

Antworten:

1.Gar nicht

2. maybe the last part in which it is explained how the ai creates a new pic from pixels could be more detailed but it is in general very understandable video

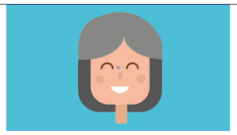
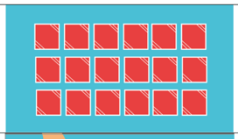
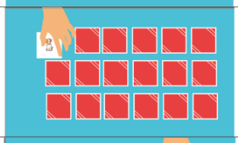
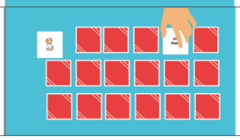
3. Mir fällt ehrlich gesagt nichts ein, was es zu verbessern gäbe

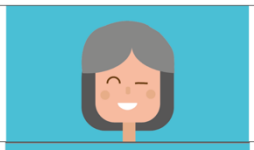
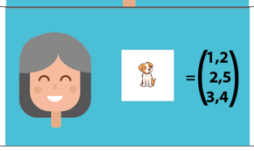
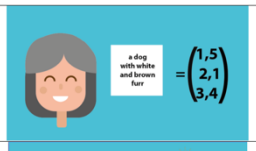
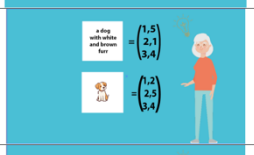
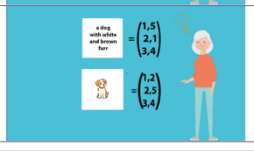
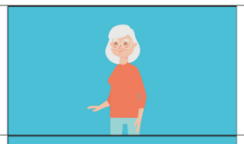
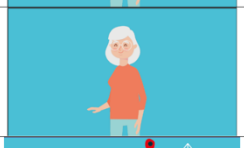
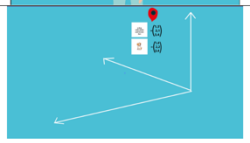
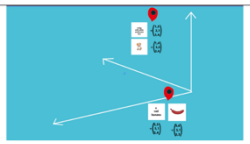
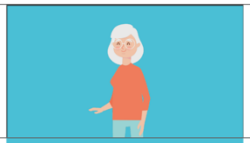
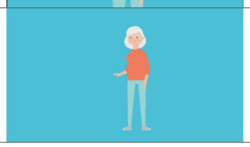
4. ist super!

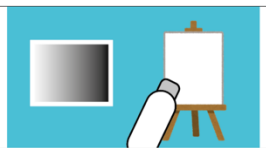
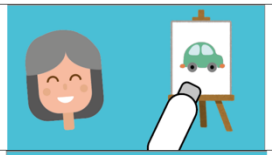
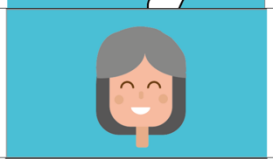
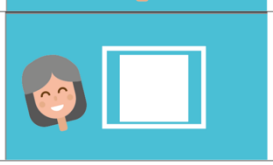
5. wichtige Punkte highlight mit Untertitel

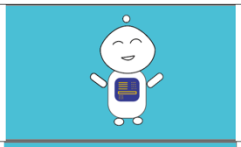
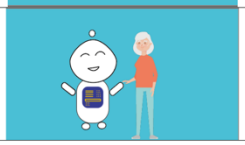
6. Alles verstanden!

Storyboard

Scene	Camera	plot	Voice Over	Reference	Duration in seconds
1.1	Closeup to medium long shot	Ruby appears on the screen really close, knocks against the screen and talks to the audience. Then she backs up a bit. She is quirky and excited as she talks to the audience.	*-Excited-* *„Hi!* *So you would like to learn how to draw?"* *-(sceptical-* *"hmm, but are computers even capable of drawing?* *You can't even hold a pencil .. and can not read?"* **thinking** *(hopeful)* *„You know what? I think we can make it work!"*		8sec.
1.2	Top view close up	She puts out her hand towards the sceen,	*"Do you trust me?"*		3sec.
2.1	Top view, Long shot	A field of cards appears one after another on the screen.	*"For step one, I will give you a big set of data – in this case: a big set of cards."*		5sec.
2.2	Top view, Long shot	Rubys Hand reaches from the top of the screen towards one card. She turns it around and a picture of a dog appears.	*„I will turn one card around.* *(..)* *As you can see: it's a picture of a dog."*		4sec.
2.3	Top view, Long shot	Ruby turns around another card, "a dog" is written on it.	*„Now a second card."* *(..)* *„a dog with white and brown furr"* *"seems like a describtion of the first card."*		4sec.

2.4	Frontally, Half-Long shot	Ruby appears on the screen She talks and has a big smile on her face. She holds up the 2 cards with her 2 hands (not displayed here but we see half of her body and her hands are holding the cards)	*„These to cards are a match, like in the memory card game.* *Cool, right?"*		3sec.
2.5	Zoom in, closeup	Ruby talks to the audience and in the end she winks at the audience.	*„Of course I understand: as a computer you do not know what do with that – you are more of the math-type.* *But don't worry – we can work with that!"*		6sec.
3.1	Zoom out, Half long shot	The card with the picture of the dog appears. Next to it a vector. Ruby is also in a half long shot on the side.	*„For each picture you memorize a numerical code as a mnemonic, in this case: a vector."*		4sec.
3.2	Half long shot	The card with the text of the dog appears. Next to it a vector. Ruby is also in a half long shot on the side.	*"Same thing for the text describtion."*		2sec.
3.3	Long shot	The 2 cards appear and the 2 vectors. Ruby is standing text to it and explaining.	*"What do the text and the picture have in common?* *That's right, they mean the same thing,* *only in different forms of representation."*		6sec.
3.4	Long shot	A Graph with 3 lines appears. At one point a "Maps dot" appears and the 2 pictures with their vector beneath.	*"That means you can put them in one place in a multidimensional space."*		3sec.
3.5	Zomm in, Half-long shot	Ruby is talking to the audience. She is excited to share all the knowledge.	*"What is a multidimensional Space, you ask?"*		2sec
3.6	Half-long-shot	Ruby is talking to the audience. She is excited to share all the knowledge.	*"It refers to a lower-dimensional, abstract representation of data where similar data points are closer together, and dissimilar ones are farther apart."*		6sec.
3.7	Long shot	A three dimensional space builds up one after another on the screen.	*"Look at this example: Each dimension represents certain paramenters. This one could represent furry animals. Therefor our picture and the discribtion of the dog are placed or everything that is brown and white."*		8sec.
3.8	Long shot	Same screen as before. Now on the lower part appears another red dot. Then the picture of a red banana and the discribtion and the 2 very similar vectors.	*"In another dimension, you see this red banana. Maybe all red and roundy things are placed here.* *Humans only know up to 3 dimensions x,y,z."*		5sec.
3.9	Closeup and slow zoom in	Ruby talks to the viewer. She gets very enthusiastic, almost emotional about it.	*"But you, you have gift. You can work with hundreds of dimensions.* *In the multidimensional space."*		3sec.
4.1	Zoom out to long shot	Ruby points at the viewer.	*"Let's use your great talent and move on!"*		2sec.

4.2	Long shot	The picture, describtion and the vector are on the screen.	*"We have the picture, discribtion and vector of the dog."*		3sec.
4.3	Long shot	A frosted glas (about 40% opacity) appears over the picture. Then the opacity gets higher and higher (up to 100%).	*"What happens if I put a frosted glas in front of it? (..) It becomes harder and harder by the second to see the dog."*		5sec.
4.4	Long shot	Next to the frosted glas appears an easel and Ruby explaning and making gestures towards the easel.	*"Now it's your turn: I want you to recreate the picture."*		3sec.
4.5	longshot	The arm of the viewer (robot, computer) reaches towards the easel and starts "drawing"	-		3sec.
4.6	Long shot	As the robotarm is moving, a picture of a dog appears.	*"oh oh, that's not the dog we need. I needs to have brown and white furr, remember? Try again!"*		5sec.
4.7	Long shot	The robot arm whipes over the easel, everything becomes pixely (pixel all over the easel)	-		2sec.
4.8	Long shot	The "correct" dog appears on the easel.	*"Amazing work!"*		2sec.
5.1	Zomm in, medium long shot	Ruby in a half long shot (we see her head + shoulders). She holds up another frosted glas in her hand. Next to it appears another describtion ("a green car") and another vector.	*" And now I will make it a bit harder! (..) Recreate a picture of a green car. You have the vector and the discribtion and the frosted glas. (..) Let's go!"*		6sec.
5.2	Zoom out Long shot	The robot arm starts "drawing" on the easel. Pixel appear on the easel, they move around.	-		2sec.
5.3	Half long shot	Ruby appears and is super enthusiastic. On the easel is now the green car.	*"Oh wow, perfect."*		3sec.
5.4	Close up	Ruby talks to the viewer. She is a bit shy in the beginning and then very confident.	*"but I have to admit something .. I tricked you .. but for your own good!"*		3sec.
5.5	Zoom out, half long shot	Ruby moves to the side and holds up the frosted glas, with a white picture behind it.	*"There was no picture of a green car behind the frosted glas! You created it yourself!"*		5sec

5.6	Half	She turns the "camera around" and now we see the "Viewer" which is the computer, that now learned how to draw.	*"Can you believe it?*		3sec.
5.7	Long shot	Ruby gives the computer a high five.	*"Anything is possible, if you just believe in yourself."*		4sec.