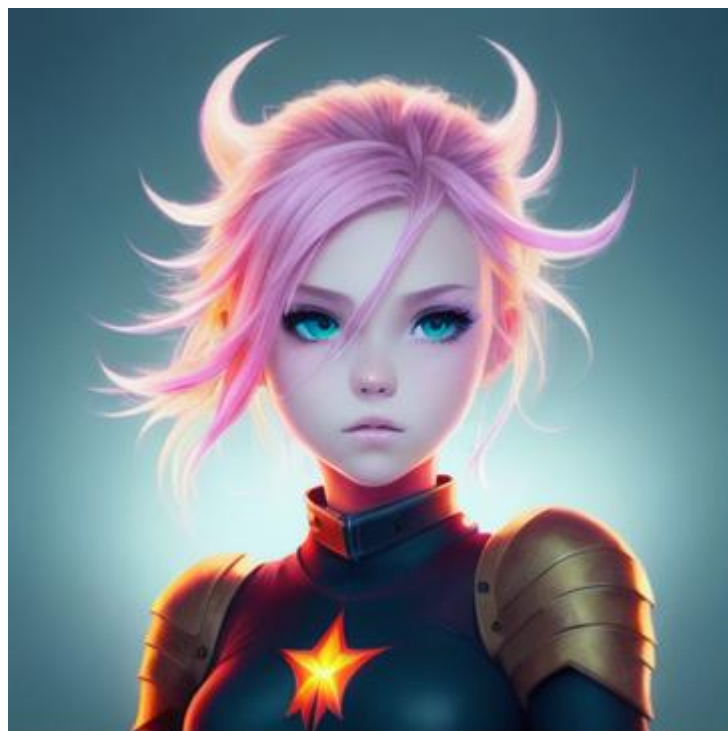


Multimediagenerierung mittels Künstliche Intelligenz

Anleitung zum Schülerworkshop



Oleg Bissing
HTW Dresden
2023

Inhalt

1. Motivation	3
2. Zielstellung.....	3
3. Glossar.....	3
4. Infrastruktur zur Workshopdurchführung	4
Schritt 1: Zugriff auf Google Colab	4
Schritt 2: Hochladen Ihres Notebooks	4
Schritt 3: Nutzung von Colab Notebooks	4
Schritt 4: Ausführen von Code.....	5
Schritt 5: Ressourcen und Beschleunigung nutzen	5
5. Workshopablauf	7
5.1. Der Workshop besteht aus fünf Modulen:	7
5.2. Modularer Aufbau	7
5.3. Konzepte und Möglichkeiten	7
5.4. Ergebnisse	8
5.5. Schlussfolgerung	8
6. Bildgenerierung.....	9
6.1. Theoretischer Teil	9
6.2. Praktischer Teil	12
7. Sprachgenerierung.....	14
7.1. Theoretischer Teil	14
7.2. Praktischer Teil	15
8. Musikgenerierung	16
8.1. Theoretischer Teil	16
8.2. Praktischer Teil	18
9. Videogenerierung.....	20
9.1. Theoretischer Teil	20
9.2. Praktischer Teil	21
10. Textgenerierung	22
11. Bekannte Fehler(meldungen)	23
11.1. „Out of memory“.....	23
11.2. „Too many sessions“	23
12. Fazit.....	24
12.1. Beurteilung der Modelle.....	24
12.2. Einige konkrete Erkenntnisse, die die Teilnehmerinnen und Teilnehmer aus dem Workshop mitnehmen können	24
13. Ausblick	25

1. Motivation

1. Keine AI-Literacy: Viele Menschen haben nur ein grundlegendes Verständnis von KI und sind sich der potenziellen Risiken und Vorteile dieser Technologie nicht bewusst.
2. Kein Interesse an MINT-Fächern: Viele junge Menschen interessieren sich nicht für MINT-Fächer, wie Informatik, Mathematik und Naturwissenschaften. Dies hat zur Folge, dass es einen Fachkräftemangel in diesen Bereichen gibt.
3. Diskrepanz zwischen Nutzung von KI in der Arbeitswelt und Schule/Bildung: In der Arbeitswelt wird KI bereits in vielen Bereichen eingesetzt. In der Schule und in der Bildung wird KI hingegen noch wenig eingesetzt. Dies führt dazu, dass junge Menschen nicht die Fähigkeiten und Kenntnisse erwerben, die sie benötigen, um in einer von KI geprägten Arbeitswelt erfolgreich zu sein.

2. Zielstellung

1. Dieser Workshop trägt zur Lösung 1. Motivation Problem bei, indem er Schülerinnen und Schüler mit generativer KI vertraut macht und ihnen die Möglichkeit gibt, diese Technologie selbst zu erkunden.
2. Die Anwendung von KI-Modellen trägt zur AI-Literacy bei, indem Schülerinnen und Schüler lernen, wie KI funktioniert und wie sie eingesetzt werden kann.
3. Die Gefahren von KI werden aufgezeigt, indem Schülerinnen und Schüler lernen, wie KI missbraucht werden kann und wie sie sich dagegen schützen können.

3. Glossar

Erklärungen zum folgenden Begriffen sind als hilfreiche Mittel für Workshopteilnehmern gedacht.

Prompts sind kurze Befehle oder Fragen, die man bei einem Neuronetz (wie z.B. ChatGPT) eingibt, damit das Tool eine Antwort (Medien usw.) generiert. Dabei gilt: Je besser und präziser die Anfrage (der Prompt), desto besser das Ergebnis.

Ein **Seed** ist eine Startnummer, die die Zufälligkeit in einem Algorithmus kontrolliert. Durch das Festlegen eines Seeds können Ergebnisse reproduzierbar gemacht werden, da der gleiche Seed jedes Mal den gleichen Zufallsprozess startet.

Inference Steps (Anzahl_Inferenzschritte) bezieht sich auf die spezifischen Prozesse, die während der Inferenzphase in der Multimediagenerierung durchgeführt werden, um neue Medien basierend auf dem gelernten Modell zu generieren. Dabei werden Eingabedaten in das Modell eingegeben, Vorhersagen generiert und gegebenenfalls die Ergebnisse verfeinert. Diese Schritte sind entscheidend, um hochwertige multimediale Ergebnisse zu erzielen.

Mit der "**Orientierungsskala**" sagt man, wie sehr sich das Modell an die Prompts halten soll. Ist der Faktor sehr hoch, folgt das Modell den Anweisungen sehr genau, ist er sehr niedrig, wird es (zu) kreativ.

4. Infrastruktur zur Workshopdurchführung

Für diesen Workshop ist ein Google Account erforderlich. Zur Codeausführung wird Google Colab, auch bekannt als Google Colaboratory benutzt. Es ist eine Cloud-basierte Plattform von Google, die es Benutzern ermöglicht, Python-Notebooks auszuführen und zu teilen. Es basiert auf der beliebten Open-Source-Jupyter-Plattform und bietet viele Vorteile für KI-Projekte.

Hier sind einige Gründe, warum Google Colab für unsere KI-Projekt nützlich ist:

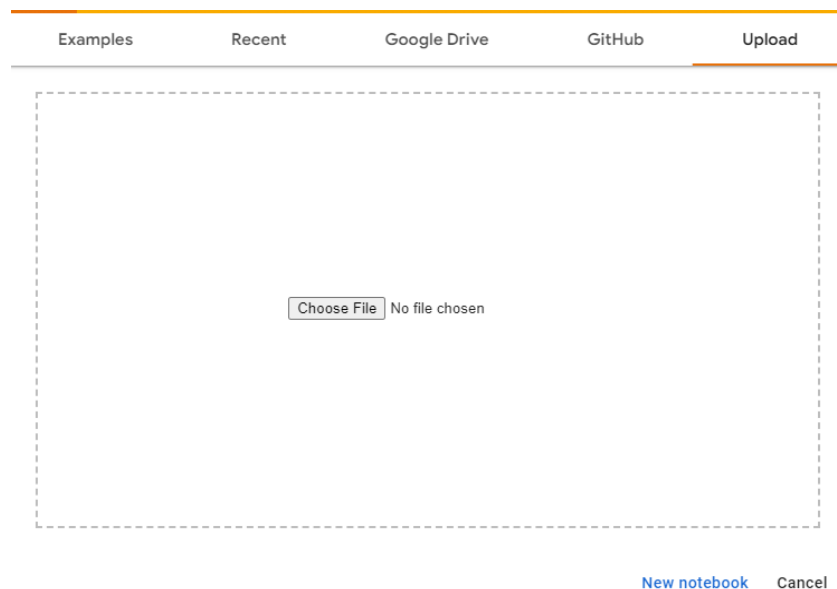
1. Kostenlose Cloud-Plattform zum Ausführen von Python-Notebooks.
2. Unterstützung von GPUs und TPUs für schnelles Deep Learning.
3. Vorinstallierte KI-Bibliotheken für einfache Anwendungsentwicklung.
4. Einfache Zusammenarbeit und Teilen von Projekten.
5. Cloud-basierte Umgebung, keine eigene Hardware nötig.

Schritt 1: Zugriff auf Google Colab

Um Google Colab zu nutzen, benötigen Sie lediglich einen Webbrowser und ein Google-Konto. Öffnen Sie Ihren bevorzugten Webbrowser und besuchen Sie die Webseite von Google Colab¹.

Schritt 2: Hochladen Ihres Notebooks

Nachdem Sie die Colab-Webseite geöffnet haben, sehen Sie die Option "Notebook hochladen" (Upload). Klicken Sie darauf und wählen Sie das Notebook von Ihrem PC aus, das Sie auf Google Colab nutzen möchten. Ihr Notebook wird dann auf die Plattform hochgeladen.



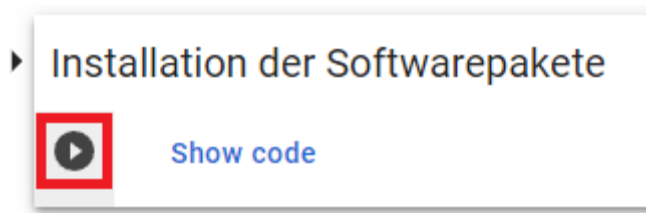
Schritt 3: Nutzung von Colab Notebooks

Sobald Ihr Notebook auf Colab hochgeladen ist, können Sie es bearbeiten und ausführen, genau wie bei einem lokalen Jupyter-Notebook. Sie können Codezellen schreiben und ausführen, Textzellen für Dokumentation und Erklärungen hinzufügen und multimediale Inhalte wie Bilder und Videos einbetten.

¹ <https://colab.research.google.com/>

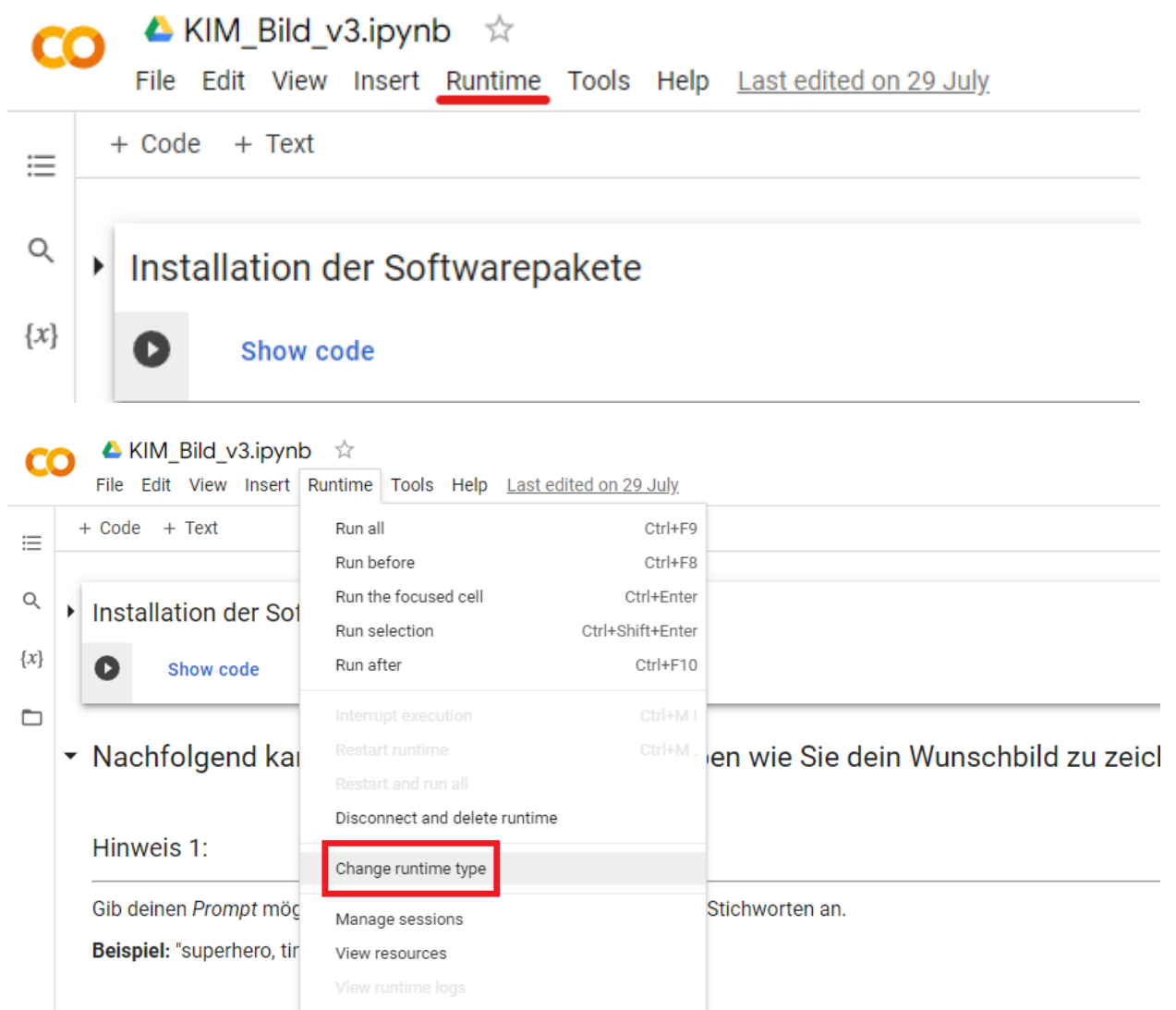
Schritt 4: Ausführen von Code

Mit Google Colab können Sie Python-Codezellen interaktiv ausführen. Klicken Sie einfach auf eine Codezelle und drücken Sie "Shift + Enter", um den Code in dieser Zelle auszuführen. Ergebnisse werden direkt unter der Codezelle angezeigt.



Schritt 5: Ressourcen und Beschleunigung nutzen

Wenn Sie in Ihrem Notebook rechenintensive KI-Modelle erstellen oder große Datenmengen verarbeiten möchten, können Sie von den leistungsstarken Ressourcen von Google Colab profitieren. Gehen Sie dazu auf "Laufzeit" in der Menüleiste und wählen Sie "Laufzeittyp ändern". Dort können Sie zwischen CPUs, GPUs oder TPUs als Hardware-Beschleuniger wählen.



Change runtime type

Runtime type

Python 3 ▼

Hardware accelerator

- ☐ CPU ☒ T4 GPU ☐ A100 GPU ☐ V100 GPU
- ☐ TPU

Want access to premium GPUs? [Purchase additional compute units](#)

Cancel [Save](#)

5. Workshopablauf

5.1. Der Workshop besteht aus fünf Modulen:

- Modul 1: Bildgenerierung. In diesem Modul lernen die Teilnehmer, wie man mit KI Bilder generiert. Sie erfahren, wie KI-Modelle funktionieren, und wie man sie zur Erstellung von realistischen oder abstrakten Bildern einsetzt. *Modell: Deliberate (Stable Diffusion)*².
- Modul 2: Sprachgenerierung. In diesem Modul lernen die Teilnehmer, wie man mit KI Stimme klonen kann. Sie erfahren, wie KI-Modelle funktionieren, und wie man sie zur Erstellung von realistischen Stimme einsetzt. *Modell: Tortoise*³.
- Modul 3: Musikgenerierung. In diesem Modul lernen die Teilnehmer, wie man mit KI Musik generiert. Sie erfahren, wie KI-Modelle funktionieren, und wie man sie zur Erstellung von verschiedenen Musikstilen einsetzt. *Modell: Audiocraft*⁴.
- Modul 4: Videogenerierung. In diesem Modul lernen die Teilnehmer, wie man mit KI Videos generiert. Sie erfahren, wie KI-Modelle funktionieren, und wie man sie zur Erstellung von realistischen oder animierten Videos einsetzt. *Modell: Stable Diffusion*⁵.
- Modul 5: Textgenerierung. In diesem Modul lernen die Teilnehmer, wie man mit KI Texte generiert. Sie erfahren, wie KI-Modelle funktionieren, und wie man sie zur Erstellung von hochqualitativen Texts einsetzt. *Modell: unterschiedliche Varianten sind möglich; z.B. LLaMA*^{6,7}, *ChatGPT*⁸ oder *Bard*⁹.

5.2. Modularer Aufbau

Jedes Modul ist in mehrere Abschnitte unterteilt. Die Abschnitte sind so konzipiert, dass die Teilnehmer die Konzepte der KI-Modelle Schritt für Schritt kennenlernen können.

5.3. Konzepte und Möglichkeiten

In jedem Modul werden verschiedene Konzepte und Möglichkeiten der KI-Modelle vermittelt. Dazu gehören:

- Generelles Prompting: Die Teilnehmer lernen, wie man KI-Modelle mit einfachen Textprompts oder fortgeschrittenen Techniken steuert.
- Einsatzmöglichkeiten: Die Teilnehmer lernen, wie man KI-Modelle in verschiedenen Bereichen einsetzt, wie zum Beispiel in der Kunst, der Musik, der Videoproduktion und der Bildung.
- Gefahren: Die Teilnehmer lernen auch über die potenziellen Risiken der KI-Modelle, wie zum Beispiel das Klonen der Stimme oder die Verdrängung von Berufsbildern.

² <https://civitai.com/models/4823/deliberate>

³ <https://github.com/neonbjb/tortoise-tts>

⁴ <https://github.com/facebookresearch/audiocraft>

⁵ <https://github.com/nateraw/stable-diffusion-videos>

⁶ https://huggingface.co/docs/transformers/main/model_doc/llama

⁷ <https://chat.lmsys.org/>

⁸ <https://chat.openai.com/>

⁹ <https://bard.google.com/>

5.4. Ergebnisse

Am Ende des Workshops werden die Teilnehmer in der Lage sein, die folgenden Ergebnisse zu erzielen:

- Verständnis der Funktionsweise von KI-Modellen
- Fähigkeit, KI-Modelle zur Generierung von Bildern, Texten, Musik und Videos einzusetzen
- Bewusstsein der potenziellen Risiken der KI-Modelle

5.5. Schlussfolgerung

Dieser Workshop bietet den Teilnehmern eine Einführung in die KI-basierte Mediengenerierung. Die Teilnehmer lernen die Funktionsweise von KI-Modellen kennen und wie man sie zur Erstellung von verschiedenen Medienformaten einsetzt. Der Workshop vermittelt auch ein Bewusstsein für die potenziellen Risiken der KI-Modelle.

6. Bildgenerierung

6.1. Theoretischer Teil

Stable Diffusion ist ein generatives KI-Modell, das Bilder aus Textbeschreibungen generieren kann. Es wurde von einem Team aus Forschern und Ingenieuren der Non-Profit-Organisation LAION, der Forschungsabteilung des KI-Unternehmens Stability AI sowie CompVis, einer Gruppe an der Ludwig-Maximilians-Universität München (LMU), entwickelt.

Stable Diffusion funktioniert, indem es ein Bild in einen latenten Raum zerlegt, in dem die Informationen über das Bild in einer abstrakteren Form gespeichert sind. Dieser latente Raum wird dann mit einem Textprompt gefüttert, der beschreibt, was das Bild darstellen soll. Das Modell verwendet dann den Textprompt, um das Bild wieder aus dem latenten Raum zu rekonstruieren.

In diesem Workshop wird Deliberate benutzt. Es ist eine neue Version von Stable Diffusion, die zu genaueren, realistischeren und ansprechenderen Bildern führt. Deliberate verwendet einen größeren latenten Raum, einen neuen Algorithmus und eine neue Verlustfunktion, um diese Verbesserungen zu erzielen.

Stable Diffusion (und auch Deliberate) ist eine sehr leistungsstarke Bildgenerierungstechnik. Es kann Bilder von einer Vielzahl von Objekten und Szenen generieren, einschließlich Menschen, Tiere, Landschaften und Gebäude. Die Bilder sind oft sehr realistisch und detailliert.

Dabei ist es sehr wichtig die richtige Prompts¹⁰ einzugeben. Diese kann man sowohl auf Englisch als auch auf Deutsch eingeben. Wenn Workshopteilnehmern Tipps für Prompt-Engineering brauchen, dann kann man folgende verwenden:

- **Sei so detailliert wie möglich.** Je genauer du den gewünschten Bildinhalt beschreibst, desto wahrscheinlicher ist es, dass das Modell ein Bild generiert, das deinen Erwartungen entspricht.
- **Verwende spezifische Wörter und Ausdrücke.** Vermeide allgemeine Begriffe wie "schön" oder "interessant". Stattdessen solltest du Wörter und Ausdrücke verwenden, die den gewünschten Bildinhalt genau beschreiben.
- **Verwende konkrete Beispiele.** Wenn es dir schwerfällt, den gewünschten Bildinhalt mit Worten zu beschreiben, kannst du auch konkrete Beispiele verwenden. Zum Beispiel könntest du statt "ein schönes Bild" schreiben "ein Bild einer goldenen Sonnenuntergangsszene am Meer".
- **Sei kreativ.** Denke nicht nur an die offensichtlichen Dinge. Versuche, auch kreative und ungewöhnliche Bildideen zu generieren.
- **Experimentiere mit verschiedenen Prompts.** Probiere verschiedene Formulierungen und Ansätze aus, um herauszufinden, was am besten funktioniert.
- **Sei geduldig.** Es kann einige Versuche erfordern, um ein Bild zu generieren, das deinen Erwartungen entspricht.
- **Habe Spaß!** Bildgenerierung ist eine großartige Möglichkeit, deiner Kreativität freien Lauf zu lassen.

¹⁰ Siehe Glossar auf s. 3

Hier sind einige Beispiele gute Prompts:

1. [superpower*], teenager girl, super power, drawing, artwork, fantasy hero character portrait *(z.B. Firestarter, Plant Whisperer, Wind Rider usw.)



2. steampunk city, christmas, sunset, photorealism, from above



3. japan street, night, raining, moonlight, high detail, anime style



4. mystical forest, vibrant colors, towering trees




5. dramatic battle between elemental forces, with fire, water, earth, and air colliding in a mesmerizing display of power and energy

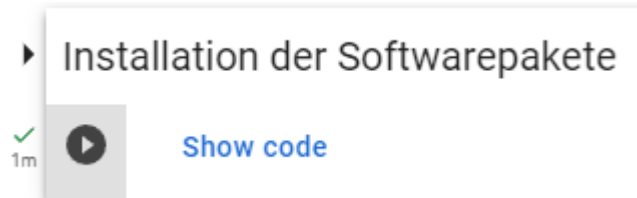


Während der Bildgenerierung ist es möglich, dass an der Stelle einige Bilder wird ein schwarzes Rechteck generiert. Das bedeutet, dass generiertes Bild vermutlich nicht jugendfrei ist und wird deswegen mittels einer NSFW-Filter versteckt.

6.2. Praktischer Teil


Gegebene Vorlage für die Bildgenerierung lässt innerhalb wenigen Minuten alle nötige Vorbereitungen wie laden von Bibliotheken vornehmen und mit der Bildgenerierung anfangen.

Laden Sie bitte KIM_Bild_v3.ipynb in Google Colab hoch. Der erste Schritt ist „Installation der Softwarepakete“. Dafür kann man auf „Play-Button“  klicken. Dann muss man ca. 1 Minute warten, bis alle nötigen Bibliotheken hochgeladen sind. Inzwischen kann man danach folgende Hinweise lesen. Wenn alles geladen ist, steht ein grünes Häkchen daneben.





Ab sofort kann man mit Generierung anfangen. Dafür schreibt man gewünschte Eingabe als „Positiver_Prompt“.

► Führe diese Zelle aus, bis dir eines der Bilder in der Grundstruktur gefällt!

 **Positiver_Prompt:** "plant whisperer, teenager girl, super power, drawing, artwork, fantasy hero character portrait *(z.B. Firestarter, Plant Whisperer, Wind Rider usw.)"

Negativer_Prompt: "Insert text here"

Orientierungsskala: 

Anzahl_Inferenzschritte: 

[Show code](#)


Beim „Bildanpassung“ geht es um Bearbeitung eines spezifisches Seeds¹¹ von vorher generierten Bilder. Seed soll als „Bild“ eingegeben werden, wobei Seeds sind wie folgt nummeriert:



1	2
3	4


¹¹ Siehe Glossar auf s. 3

- Wähle das Bild aus, welches dir am besten gefallen hat und mache Änderungen!

 **Bild:** 1

Positive_Prompt_Ergänzung: " Insert text here

Negative_Prompt_Ergänzung: " Insert text here

Orientierungsskala: 

Anzahl_Inferenzschritte: 

Weitere Einstellungen wie Orientierungsskala und Anzahl_Inferenzschritte sind auch in Glossar auf Seite 3 beschrieben.

Für das Bildgenerierung Werte beim Orientierungsskala können beliebig ausgewählt werden, jedoch eine gut balancierte Wert wäre 6.5.

Anzahl Inferenzschritte ist auch eine flexible Einstellung, die in eine direkte Verbindung mit dem Prompt steht. Um eine geeignete Bildqualität für ein kompliziertes Prompt zu gewährleisten braucht das Modell in der Regel mehr Schritte als für ein einfaches Prompt. Hier ist es immer am besten selbst zu experimentieren um eine für bestimmte Prompt passende Wert zu finden, jedoch als Startwert wird eine Zahl zwischen 50 und 100 eine gute Wahl darstellen.

7. Sprachgenerierung

7.1. Theoretischer Teil

Die Stimme einer Person wird durch eine Reihe von Faktoren bestimmt, darunter die Form des Mundes, des Rachens und der Nase, die Größe und Form der Stimmbänder und die Art und Weise, wie Luft durch die Stimmbänder gepresst wird. Diese Faktoren können alle durch das Tortoise TTS-Modell simuliert werden.

Um eine Voice-Cloning mit Tortoise TTS durchzuführen, benötigt das Modell eine ausreichend große Menge an Audioaufnahmen der Stimme der Person, die geklont werden soll. Diese Aufnahmen werden verwendet, um das Modell zu trainieren, die charakteristischen Merkmale der Stimme zu erkennen.

Sobald das Modell trainiert ist, kann es verwendet werden, um neue Audiodateien mit der Stimme der geklonten Person zu generieren. Diese Audiodateien können dann verwendet werden, um die Person in Videos oder Audioaufnahmen zu imitieren.


Der Prozess des Voice-Cloning mit Tortoise TTS ist noch relativ neu, aber er wird sich in Zukunft wahrscheinlich weiterentwickeln. Mit der Verbesserung der Technologie wird es möglich sein, noch realistischere und überzeugendere Voice-Clones zu generieren.

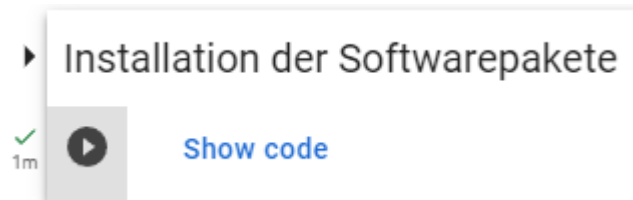
Es ist wichtig zu beachten, dass Voice-Cloning mit Tortoise TTS nicht perfekt ist. Die generierten Audiodateien sind möglicherweise nicht identisch mit der Originalstimme. Es ist auch möglich, dass die generierten Audiodateien erkennbar als geklont sind.

Trotz dieser Einschränkungen kann Voice-Cloning mit Tortoise TTS ein nützliches Werkzeug für eine Vielzahl von Zwecken sein. Es kann verwendet werden, um Audioaufnahmen von Personen zu erstellen, die nicht mehr sprechen können, oder um Personen zu imitieren, die in Videos oder Audioaufnahmen nicht persönlich erscheinen können.

7.2. Praktischer Teil

Gegebene Vorlage für die Sprachgenerierung lässt innerhalb wenigen Minuten alle nötige Vorbereitungen wie laden von Bibliotheken vornehmen und mit der Sprachgenerierung anfangen.

Laden Sie bitte KIM_Sprachgenerierung_v2.ipynb in Google Colab hoch. Der erste Schritt ist „Installation der Softwarepakete“. Dafür kann man auf „Play-Button“  klicken. Dann muss man ca. 1 Minute warten, bis alle nötigen Bibliotheken hochgeladen sind. Inzwischen kann man danach folgende Hinweise lesen. Wenn alles geladen ist, steht ein grünes Häkchen daneben.




Danach braucht man drei Stimmeaufnahmen.

- **Spreche hier den wütenden Text ein und höre dir das Ergebnis noch einmal an.**

 [Show code](#)


```
array([0, 0, 0, ..., 1, 1, 0], dtype=int16)
```

Aufzeichnung läuft... Klick mich nochmal um die Aufnahme zu stoppen.

Dabei ist es wichtig zu beachten, dass eine Aufnahme beginnt sofort sobald „Play-Button“  angeklickt wird. Um die Aufnahme zu beenden man muss auf **Aufzeichnung läuft... Klick mich nochmal um die Aufnahme zu stoppen.** klicken. Die Aufnahme wird dann automatisch gespeichert.

Als letztes kommt das Generierungsschritt. Hier man kann das gewünschte Text schreiben und generieren lassen. Wichtig dabei ist: 1. Text soll relativ kurz sein (1-2 Sätze), sonst wird es ewig generiert; 2. Text soll auf Englisch geschrieben werden, da das Modell generiert mit Englischer Aussprache (Aufnahmen trotzdem auf Deutsch gemacht werden können).

- Hier spricht dein KI-Synchrone Sprecher deinen Text ein!

 **Text:** " Maya Clarke became the superheroine Blaze, igniting justice wherever she went. She is determined to bring safety to a world in need.
Emotion: wütend
Qualität: fast
[Show code](#)

Einstellung „Qualität“ wird nicht empfohlen zu ändern, weil das Modell auch mit „fast“ mehr als passende für den Zweck des Workshops Qualität liefert. Außerdem eine höhere Qualitätseinstellung wird das Generierungszeit spürbar erhöhen.

8. Musikgenerierung

8.1. Theoretischer Teil

AudioCraft ist ein KI-basiertes Musikgenerierungstool, das von Meta entwickelt wurde. Es besteht aus drei separaten Modellen:

- **MusicGen:** Dieses Modell verwendet einen großen Datensatz von Musikstücken, um Musik aus Textanweisungen zu generieren.
- **AudioGen:** Dieses Modell verwendet einen großen Datensatz von Soundeffekten, um realistische Audiosignale wie das Bellen eines Hundes oder Schritte auf einem Holzboden zu generieren.
- **EnCodec:** Dieses Modell ist ein Decoder, der die Audiosignale von MusicGen und AudioGen in eine hochwertige Audiodatei umwandelt.

Um eine Musikdatei mit AudioCraft zu generieren, müssen Sie zunächst eine Textanweisung erstellen. Diese Anweisung kann so einfach wie "Spiele eine traurige Melodie" oder so komplex wie "Spiele eine Rockballade mit einem treibenden Beat und einer emotionalen Gesangsmelodie" sein.

Sobald Sie die Textanweisung erstellt haben, wird sie von MusicGen verarbeitet. MusicGen verwendet dann seinen Datensatz von Musikstücken, um eine Musikdatei zu generieren, die Ihrer Anweisung entspricht. Die generierte Musikdatei wird dann von AudioGen und EnCodec verarbeitet, um eine hochwertige Audiodatei zu erstellen.

AudioCraft ist noch in der Entwicklung, aber es hat bereits die Fähigkeit, eine Vielzahl von Musikgenres zu generieren, von klassischer Musik bis hin zu Popmusik. Es kann auch verwendet werden, um kreative Musikstücke wie Soundscapes und Ambient-Musik zu generieren.

AudioCraft ist ein vielseitiges Werkzeug, das Musikern und Kreativen neue Möglichkeiten eröffnet. Es hat das Potenzial, die Art und Weise, wie wir Musik hören und kreieren, grundlegend zu verändern.

Dabei ist es sehr wichtig die richtige Prompts¹² einzugeben. Diese kann man sowohl auf Englisch als auch auf Deutsch eingeben. Wenn Workshopteilnehmern Tipps für Prompt-Engineering brauchen, dann kann man folgende verwenden:

- **Sei spezifisch.** Je genauer deine Anweisungen sind, desto wahrscheinlicher ist es, dass AudioCraft ein Musikstück generiert, das deinen Wünschen entspricht. Vermeide allgemeine Formulierungen wie "Spiele eine schöne Melodie". Stattdessen kannst du etwas sagen wie "Spiele eine traurige Melodie in Moll mit einem langsamen Tempo".
- **Verwende Schlüsselwörter.** Schlüsselwörter können AudioCraft helfen, das richtige Genre und die richtige Stimmung für dein Musikstück zu finden. Zum Beispiel könnte das Schlüsselwort "rockig" AudioCraft dazu bringen, ein Rockstück zu generieren, während das Schlüsselwort "atmosphärisch" AudioCraft dazu bringen könnte, eine Soundscape zu generieren.

¹² Siehe Glossar auf s. 3


- **Sei kreativ.** AudioCraft ist ein mächtiges Werkzeug, das dir hilft, neue und innovative Musik zu erstellen. Experimentiere mit verschiedenen Ideen und du wirst sehen, was du kreieren kann.

Hier sind einige Prompt Beispiele:

- crazy EDM, heavy bang
- classic reggae track with an electronic guitar solo
- lofi slow bpm electro chill with organic samples
- rock with saturated guitars, a heavy bass line and crazy drum break and fills
- earthy tones, environmentally conscious, ukulele-infused, harmonic, breezy, easygoing, organic instrumentation, gentle grooves
- epic music with heavy bass, large-scale orchestral arrangement, soaring melodies, dynamic rhythms, impactful percussion


8.2. Praktischer Teil

Gegebene Vorlage für die Musikgenerierung lässt innerhalb einige Minuten alle nötige Vorbereitungen wie laden von Bibliotheken vornehmen und mit der Musikgenerierung anfangen.

Laden Sie bitte KIM_Musik_v3.ipynb in Google Colab hoch. Der erste Schritt ist „Installation der Softwarepakete“. Dafür kann man auf „Play-Button“  klicken. Dann muss man ca. 2 Minuten warten, bis alle nötigen Bibliotheken hochgeladen sind. Inzwischen kann man danach folgende Hinweise lesen. Wenn alles geladen ist, steht ein grünes Häkchen daneben.

Ab sofort kann man mit Generierung anfangen. Dafür schreibt man gewünschte Eingabe als „Prompt“. „Länge“ wird in Sekunden angegeben.

► Generiere Musik mit deiner Beschreibung und spiele sie ab!

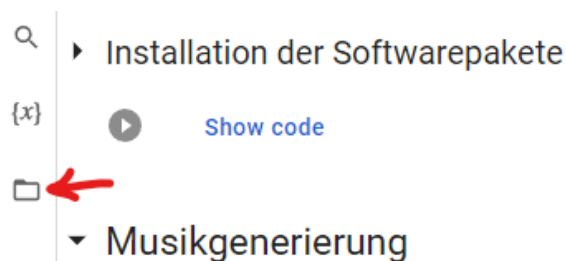
 Die Generierungszeit beträgt ungefähr **das Doppelte** der Zeit der unten angegebenen **Länge** generierten Audiotrack (in Sekunden)

Länge:

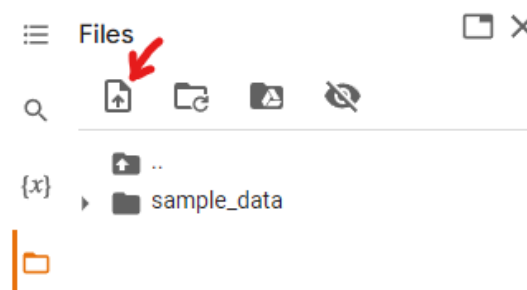
Prompt: "

[Show code](#)

Für eine Generierung mit Beispielmelodie man muss diese erst hochladen. Dafür ist es nötig Filemanager zu öffnen.



Wenn Zeichen „Verzeichnis“ angeklickt wird, sieht man den Filemanager. Danach kann das Beispielmelodie geladen werden.



Man muss nur noch den Namen von Beispielmelodie als Parameter eingeben (mit dem Fileformat, z.B. bach_kurz.mp3) und generieren.

► Generiere Musik mit deiner Beispielmelodie und Beschreibung und spiele sie ab!

▶ Die Generierungszeit beträgt ungefähr **das Dreifache** der Zeit der unten angegebenen **Länge** generierten Audiotrack (in Sekunden)

Länge: 

Prompt: " rock

Beispielmelodie: " bach_kurz.mp3

[Show code](#)

9. Videogenerierung

9.1. Theoretischer Teil

Stable Diffusion Videos ist ein generatives KI-Modell, das hochauflösende Videos aus Textbeschreibungen generieren kann. Es basiert auf dem Prinzip der Diffusionsmodelle, die Bilder oder Videos schrittweise aus einem Rauschen herausgenerieren.

Das Stable Diffusion Videos-Modell wurde von Forschern von Nvidia entwickelt und wird auf einem großen Datensatz von Videos trainiert. Das Modell lernt, wie es Videos aus einem latenten Raum generieren kann. Der latente Raum ist eine abstraktere Darstellung des Videos, die es dem Modell ermöglicht, kreativere und flexiblere Videos zu generieren.

Bei der Videogenerierung mit Stable Diffusion Videos wird der Text zunächst in den latenten Raum des Modells übertragen. Das Modell verwendet dann den latenten Raum, um ein Video zu generieren. Das Video wird dann schrittweise aus dem Rauschen heraus generiert.

Die Ausgabe von Stable Diffusion Videos kann sehr realistisch und natürlich aussehen. Das Modell ist in der Lage, eine Vielzahl von Videogenres zu generieren, von Actionfilmen bis hin zu romantischen Komödien.

Stable Diffusion Videos ist ein beeindruckendes Modell, das viele neue Möglichkeiten für die Videogenerierung eröffnet. Es hat das Potenzial, die Art und Weise, wie wir Videos erstellen und konsumieren, grundlegend zu verändern.

Hier sind einige weitere Details zu den Arbeitsabläufen von Stable Diffusion Videos:


- Textbeschreibungen: Die Textbeschreibungen können sehr detailliert sein, um das Video genau zu spezifizieren. Sie können auch eher allgemein sein, um dem Modell mehr Spielraum zu geben.
- Latent Space: Der latente Raum ist eine abstraktere Darstellung des Videos, die es dem Modell ermöglicht, kreativere und flexiblere Videos zu generieren. Es wird durch einen großen Datensatz von Videos trainiert, um die charakteristischen Merkmale von Videos zu lernen.
- Diffusionsprozess: Der Diffusionsprozess ist der Prozess, bei dem das Video aus dem Rauschen heraus generiert wird. Es ist ein schrittweiser Prozess, bei dem das Modell das Video schrittweise klarer macht.
- Qualität: Die Qualität der generierten Videos hängt von der Größe des Datensatzes, auf dem das Modell trainiert wird, und von den Einstellungen des Diffusionsprozesses ab.

Stable Diffusion Videos ist noch in der Entwicklung, aber es hat bereits das Potenzial, eine Revolution in der Videogenerierung zu bewirken.

In diesem Workshop wird das Modell mit sehr einfache Einstellungen benutzt was eine sehr kleine Bildauflösung mit einer befriedigenden Qualität liefert, dafür aber in noch ertragbare Zeit.


9.2. Praktischer Teil

Gegebene Vorlage für die Videogenerierung lässt innerhalb einige Minuten alle nötige Vorbereitungen wie laden von Bibliotheken vornehmen und mit der Videogenerierung anfangen.

Laden Sie bitte KIM_Video_v3.ipynb in Google Colab hoch. Der erste Schritt ist „Installation der Softwarepakete“. Dafür kann man auf „Play-Button“  klicken. Dann muss man ca. 2 Minuten warten, bis alle nötigen Bibliotheken hochgeladen sind. Inzwischen kann man danach folgende Hinweise lesen. Wenn alles geladen ist, steht ein grünes Häkchen daneben.

Ab sofort kann man mit Generierung anfangen. Dafür beschreibt man als „Prompt_1“ das Startbild und als „Prompt_2“ das Endbild. „Länge“ wird in Sekunden angegeben. Das Modell generiert dann beliebig viele Bilder (frames per second - fps) pro Sekunde und fügt diese zu einem Video zusammen.

► Generiere ein Animationsvideo nach deiner Beschreibung!

 Länge:

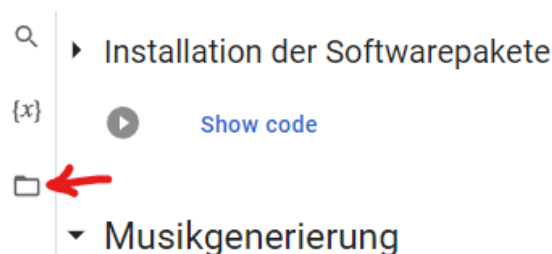
Prompt_1:

Prompt_2:

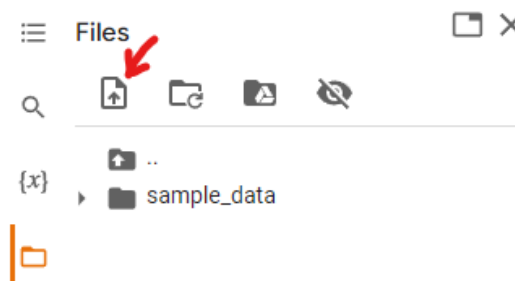
fps:

[Show code](#)

Für Musikvideogenerierung muss man noch ein Song hochzuladen. Dafür ist es nötig Filemanager zu öffnen.



Wenn das Zeichen „Verzeichnis“ angeklickt wird, sieht man den Filemanager. Danach kann ein Song geladen werden.



Man muss nur noch den Namen des Songs als Parameter für „Song“ eingeben (mit dem Fileformat, z.B. bach.mp3) und generieren.

10. Textgenerierung

Textgenerierung-KI-Modelle funktionieren alle auf ähnliche Weise. Sie werden an einem riesigen Datensatz aus Text und Code trainiert. Dieser Datensatz enthält Bücher, Artikel, Codes und andere Arten von Text.

Das Modell verwendet dann diese Daten, um zu lernen, wie Text funktioniert. Es lernt, wie Wörter zusammenhängen, wie Sätze aufgebaut sind und wie Text verschiedene Arten von Bedeutung vermitteln kann. Mit diesem Wissen kann das Modell dann neuen Text generieren.

Wenn Sie ein Textgenerierung-KI-Modell bitten, Text zu generieren, beginnt es damit, einen Satz zu erstellen (siehe Abbildung unten). Es verwendet seine Kenntnisse über Text, um einen Satz zu erstellen, der sinnvoll und zusammenhängend ist. Dann verwendet es seine Kenntnisse über die Sprache, um den Satz zu verbessern. Es fügt möglicherweise neue Wörter hinzu, um den Satz flüssiger zu machen, oder es ändert die Reihenfolge der Wörter, um den Satz klarer zu machen.

Das Modell wiederholt diesen Prozess, bis es einen Text erstellt hat, der Ihren Anforderungen entspricht.

Es gibt viele verschiedene Arten von Textgenerierung-KI-Modellen. Einige Modelle sind besser in der Generierung von kreativen Textformaten wie Gedichten, Codes, Skripten, Musikstücken, E-Mails, Briefen usw., während andere besser in der Generierung von informativem Text sind, wie Zusammenfassungen von Sachthemen oder Antworten auf Fragen.

OL Ich muss ein Vortrag zum Thema "Multimedia generierung mittels Künstliche Intelligenz" halten, hier ist das Plan:
Was ist die Idee (und warum finden wir Sie gut)?
1. SchülerInnen lernen mit generativer künstlicher Intelligenz (KI) Medien zu generieren

- Motivation informatiknahe Studiengänge aufzugreifen
- Aufzeigen zukunftsfähiger Berufsbilder (Prompt-Engineer)
- Anpassbarkeit der Thematik auf die Zielgruppe


2. SchülerInnen lernen die (aktuellen) Möglichkeiten und Limitationen generativer Modelle kennen

- Entwicklung von Medien- und KI-„Literacy“
- Hinterfragen einer von KI-geprägten Gesellschaft

3. SchülerInnen lernen wie generative Modelle funktionieren

- Bezugsherstellung zu informatiknahen Studiengängen
- „Entmystifizierung“ von generativer KI

Schreib mir bitte ein Script für ungefähr 5 Minuten

 Guten Tag, meine Damen und Herren,

ich freue mich, heute hier zu sein, um Ihnen über ein faszinierendes Thema zu berichten: die Multimedia-Generierung mittels künstlicher Intelligenz. In meinem Vortrag werde ich Ihnen die Idee hinter diesem Thema erläutern und warum wir sie für äußerst wertvoll halten. Zudem werde ich Ihnen einen Einblick in die aktuellen Möglichkeiten und Limitationen generativer Modelle geben und Ihnen zeigen, wie Schülerinnen und Schüler lernen können, solche Modelle zu verstehen und anzuwenden.

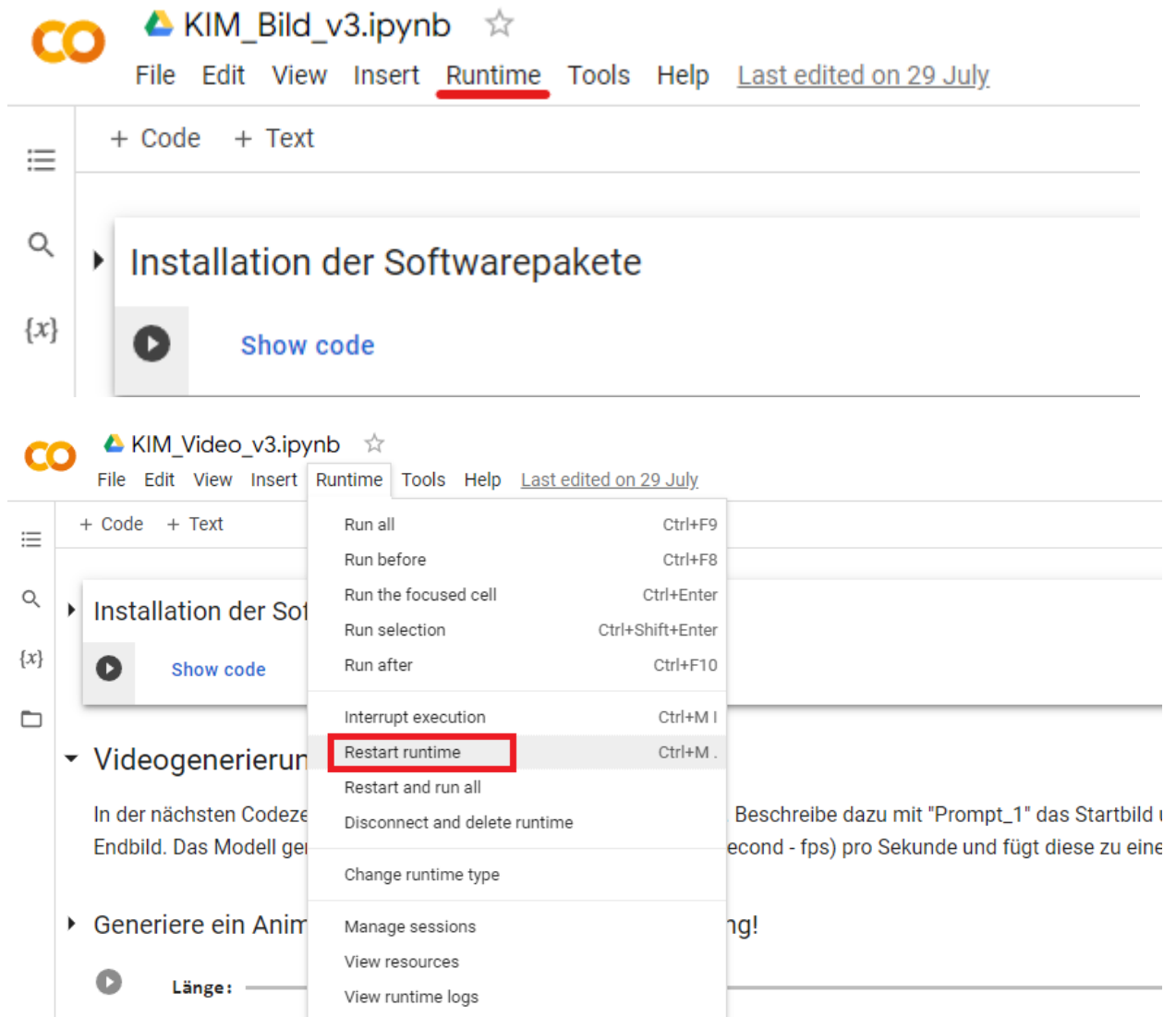
Die Idee hinter der Verwendung von generativer künstlicher Intelligenz, oder KI, zur Multimedia-Generierung in Schule Regenerate response nen möchten wir die Motivation für informatiknahe Studiengänge üllern wecken. Durch den Einsatz

Send a message...

11. Bekannte Fehler(meldungen)

11.1. „Out of memory“

Falls das RAM ist voll nach einer Generierung, man muss "Runtime - Restart runtime" nutzen, und dann wieder die Bibliotheken herunterladen und danach kann man generieren.



11.2. „Too many sessions“

Passiert immer, wenn man mehr als ein Colab-Notebook öffnet.

Too many sessions

You have too many active sessions. Terminate an existing session to continue. If you are interested in using Colab with more sessions, you may want to take a look at [Colab Pro](#).

Cancel [Manage sessions](#)

Hier klickt man auf „Manage sessions“ und dann auf „Terminate other sessions“.

12. Fazit

Im Rahmen dieses Schülerworkshops haben die Teilnehmerinnen und Teilnehmer die Möglichkeit, sich mit verschiedenen KI-Modellen vertraut zu machen. Dabei haben sie die Gelegenheit, die Potenziale und Risiken von KI kennenzulernen und sich eigene Gedanken zu den möglichen Anwendungen zu machen.

12.1. Beurteilung der Modelle

- **Bildgenerierung:** sehr gut. Bestimmt das beste Teil des Workshops, die ausgezeichnete Lösungen für die Ziele des Workshops bietet. Außerdem macht sie den Teilnehmer erfahrungsgemäß am meisten Spaß. Wenn nur ein Teil zur Vorstellung gewählt werden muss, dann ist es sicher diese.
- **Textgenerierung:** gut – sehr gut (je nach Modell). Das einzige Teil in dem die Modelle qualitativ mit der Bildgenerierung mithalten und oft sogar gewinnen können. Kann gute Lösungen für alle Ziele des Workshops anbieten.
- **Sprachgenerierung:** gut. Qualität ist in meisten Fällen sehr überzeugend. Jedoch dauert es am meisten bis man das Ergebnis sich anhören kann. Auch für Erfüllung der Ziele des Workshops ist diese Teil bestimmt nicht das beste.
- **Musikgenerierung:** befriedigend. Ein gut funktionierendes Prompt zu finden ist nicht so einfach wie beim Bildgenerierung. Auch für Erfüllung der Ziele des Workshops ist diese Teil bestimmt nicht das beste.
- **Videogenerierung:** ausreichend. Qualität vor allem in Vergleich zum Bildgenerierung ist in meisten Fällen eher mangelhaft, dazu noch dauert die Generierung länger. Auch für Erfüllung der Ziele des Workshops ist diese Teil bestimmt nicht das beste.

12.2. Einige konkrete Erkenntnisse, die die Teilnehmerinnen und Teilnehmer aus dem Workshop mitnehmen können

- KI-Modelle sind in der Lage, sehr realistische und überzeugende Inhalte zu generieren.
- KI-Modelle können jedoch auch verwendet werden, um irreführende oder manipulative Inhalte zu erstellen.
- Es ist wichtig, sich der Risiken von KI bewusst zu sein und Maßnahmen zu ergreifen, um sie zu minimieren.
- KI-Modelle können von jedem entwickelt und trainiert werden.
- Schülerinnen und Schüler haben die Möglichkeit, KI-Modelle zu verwenden, um ihre eigenen Ideen und Projekte zu realisieren.
- Die Teilnehmerinnen und Teilnehmer können diese Erkenntnisse nutzen, um sich aktiv mit KI auseinanderzusetzen und ihre eigene Meinung zu diesem wichtigen Thema zu bilden.

13. Ausblick

Die in diesem Workshop vorgestellten Technologien befinden sich noch in der Entwicklung, aber sie haben das Potenzial, die Art und Weise, wie wir mit Medien interagieren, grundlegend zu verändern. In Zukunft sind folgende Weiterentwicklungen möglich:

- **Bereitstellung der Modelle über API:** Die Modelle können über eine API bereitgestellt werden, sodass sie von anderen Anwendungen und Diensten verwendet werden können. Dies würde es ermöglichen, die Modelle für eine Vielzahl von Zwecken zu verwenden, z. B. für die Erstellung von personalisierten Medien, die Übersetzung von Sprachen oder die Erstellung von neuen Formen der Kunst und Unterhaltung.
- **Entwicklung einer eigenen Nutzeroberfläche:** Die Modelle können mit einer eigenen Nutzeroberfläche versehen werden, sodass sie von Menschen ohne Programmierkenntnisse verwendet werden können. Dies würde es ermöglichen, die Modelle einem breiteren Publikum zugänglich zu machen.
- **Austausch der Modelle:** Die Modelle können ausgetauscht werden, sodass Benutzer die Modelle verwenden können, die ihren Bedürfnissen am besten entsprechen. Dies würde es ermöglichen, die Modelle an die neuesten Entwicklungen anzupassen und den Benutzern die Möglichkeit zu geben, die Modelle zu wählen, die die besten Ergebnisse für ihre Aufgaben erzielen.
- **Zusätzliche Unterlagen zur Erklärung:** Zusätzliche Unterlagen zur Erklärung können bereitgestellt werden, um den Benutzern zu helfen, die Funktionsweise der Modelle zu verstehen. Dies würde es den Benutzern ermöglichen, die Modelle effektiver zu verwenden und die besten Ergebnisse zu erzielen.
- **Generieren einer kombinierten Medieneinheit (Video mit Subtitle, Audio, etc.):** Die Modelle können verwendet werden, um eine kombinierte Medieneinheit zu generieren, die aus einem Video, Untertiteln, Audio und anderen Medienformaten besteht.