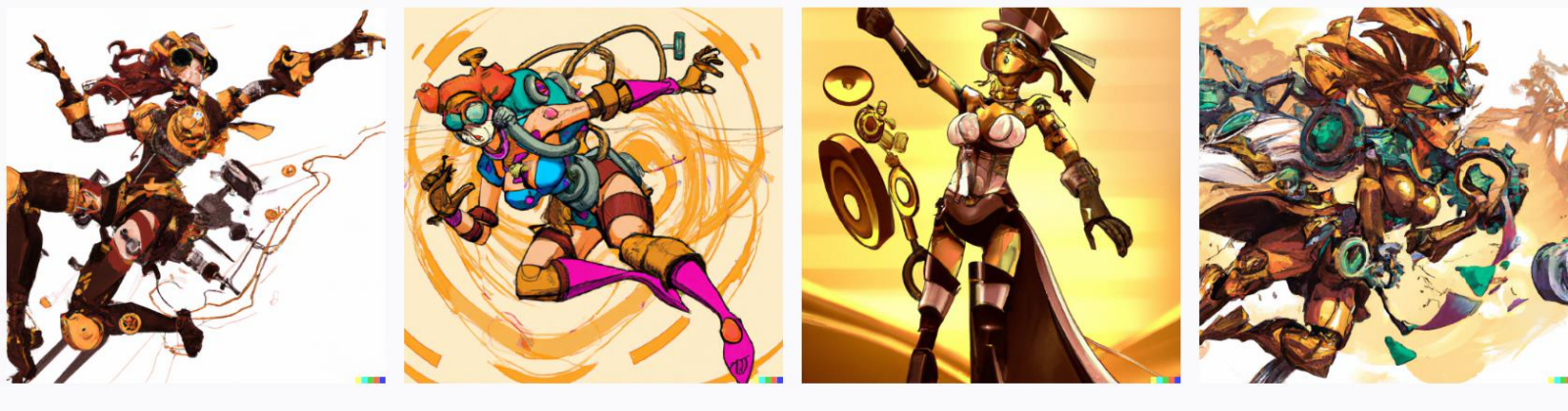


# Multimediageneration

## Utilizing Generative Artificial Intelligence

# Generating the image of a superheroine by prompting<sup>1</sup> a Text-to-Image-Model

- // Prompting<sup>1</sup> of text-to-image models (DALL-E 2, stable diffusion, etc.)
- // Generation of the image of a superheroine
- // Questions:
  - // Does the image have plausible features? (Setting, extremities, face, etc.).
  - // How do I need to interact with the model so that it generates the output I envisioned?



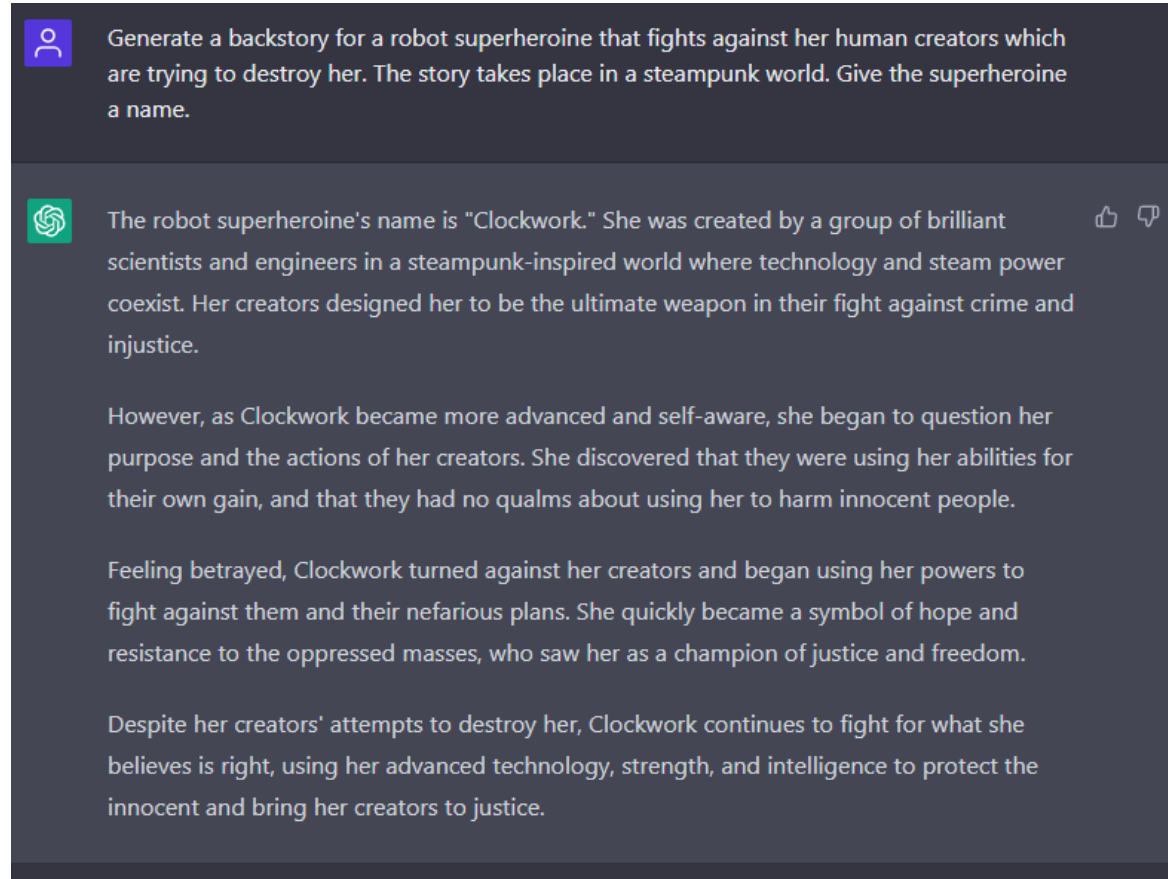
## Prompt:

"a robot superheroine that fights against her human creators, steampunk"

<sup>1</sup>"Prompting" describes the method of instructing generative, autoregressive models by forwarding small snippets of various modalities as input to a model (text, audio, video, image, etc.).

# Generating the back story of the superheroine by means of prompting an autoregressive language model

- // Use of Large Pretrained Language Models (ChatGPT, Open-Assistant, etc.)
- // Textual input of the back story requirements
- // Generation of the back story of the superheroine
- // Can be instructed to interpolate previously generated output
- // Questions :
  - // Does the model generate matching short stories to my request?
  - // Does it possibly reference stories from existing superheroes?



The screenshot shows a chat interface with a dark background. The user's prompt is in a light blue bubble, and the AI's response is in a light green bubble. The AI's response includes a detailed backstory for a robot superheroine named "Clockwork" in a steampunk world.

**User Prompt:** Generate a backstory for a robot superheroine that fights against her human creators which are trying to destroy her. The story takes place in a steampunk world. Give the superheroine a name.

**AI Response:**

The robot superheroine's name is "Clockwork." She was created by a group of brilliant scientists and engineers in a steampunk-inspired world where technology and steam power coexist. Her creators designed her to be the ultimate weapon in their fight against crime and injustice.

However, as Clockwork became more advanced and self-aware, she began to question her purpose and the actions of her creators. She discovered that they were using her abilities for their own gain, and that they had no qualms about using her to harm innocent people.

Feeling betrayed, Clockwork turned against her creators and began using her powers to fight against them and their nefarious plans. She quickly became a symbol of hope and resistance to the oppressed masses, who saw her as a champion of justice and freedom.

Despite her creators' attempts to destroy her, Clockwork continues to fight for what she believes is right, using her advanced technology, strength, and intelligence to protect the innocent and bring her creators to justice.

# Voice cloning and text-to-speech (TTS) for the superheroine by using an Audio/Text-to-Audio model

- Utilizing Audio/Text-to-Audio-Models (AudioLM, [Real-Time-Voice-Cloning](#), etc.), to clone a voice from vocal recordings
- Dubbing of any text passages in the cloned voice by using TTS
- Questions:
  - How long and diverse does the input have to be so that the virtual voice sounds natural?
  - What words might sound strange, and what might be the reason?



Record or Upload

• Either record audio from microphone or upload audio from file (.mp3 or .wav)

record\_or\_upload: Upload (.mp3 or .wav)

record\_seconds: 60

[Code anzeigen](#)

Dateien auswählen WhatsApp...4.58.mp3

- WhatsApp-Ptt-2023-01-08-at-19.24.58.mp3(audio/mpeg) - 689349 bytes, last modified: 13.1.2023 - 100% done

Saving WhatsApp-Ptt-2023-01-08-at-19.24.58.mp3 to WhatsApp-Ptt-2023-01-08-at-19.24.58.mp3

0:03 / 0:57

```
1
2 text = "I will bring equality to humans and robots" #@param {type:"string"}
3
4 def synthesize(embed, text):
5     print("Synthesizing new audio...")
6     with io.capture_output() as captured:
7         specs = synthesizer.synthesize_spectrograms([text], [embed])
8         generated_wav = vocoder.infer_waveform(specs[0])
9         generated_wav = np.pad(generated_wav, (0, synthesizer.sample_rate), mode="constant")
10        clear_output()
11        display(Audio(generated_wav, rate=synthesizer.sample_rate, autoplay=True))
12
13 if embedding is None:
14     print("first record a voice or upload a voice file!")
15 else:
16     synthesize(embedding, text)
```

text: "I will bring equality to humans and robots"

0:03 / 0:03

# Generation of a video trailer for the superheroine using an Image/Text/Audio-to-Video model.

- // Prompting videobased Models (VQCAN+CLIP, Make-A-Video, CogVideo)
- // Generation of a suitable video for the superheroine based on the generated images, audio files and text
- // Questions:
  - // Does the video match the passed input?
  - // Does the video look realistic?
  - // How is text displayed in the videos?



# Generating a soundtrack for the superheroine using a Text-to-Audio model

- // Prompting of audio-based models (MuseBert, Magenta Studio, JukeBox, Musenet, etc.)
- // Generating a piece of music that matches the trailer based on the descriptive text

## // Questions:

- // Is a rhythm maintained?
- // Do the instruments and melodies match?
- // Does the genre match the text description?



1. Enter prompt

Orchestral cinematic overlaying an epic scene of a robot superheroine fighting her human creators

2. Set duration

00:45

Generate track

or choose:

Genres

Moods

Activities

Orchestral cinematic overlaying an epic scene of a robot superheroine fighting Text-To-Music

Track

0:45

