Music Production with Artificial Intelligence

Martin Gasser

Introducing myself

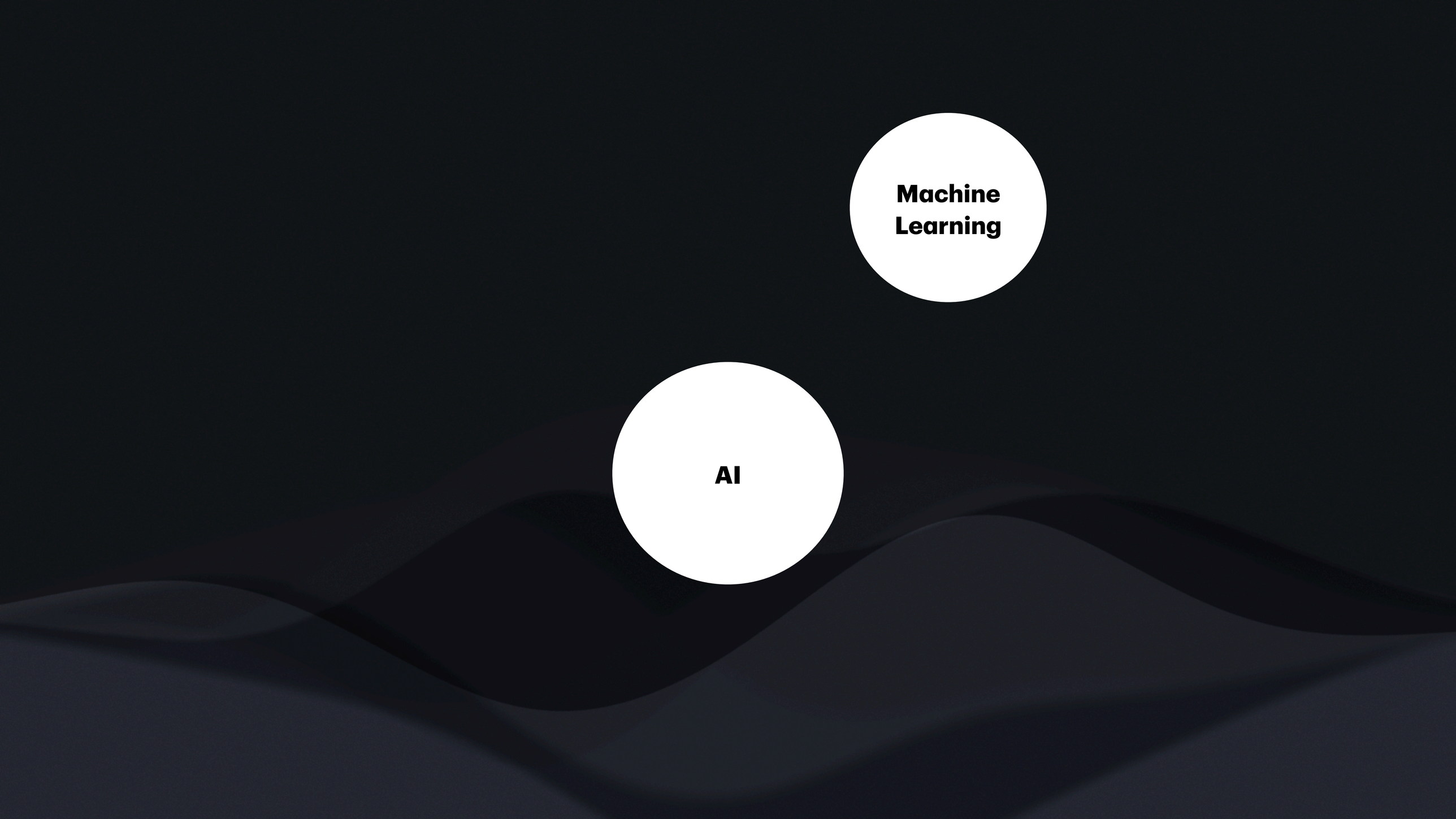
- Interdisciplinary background (Computer Science, Music)
- Musician / media artist
- Currently working mostly for Muse Group (Ultimate Guitar, MuseScore,...)

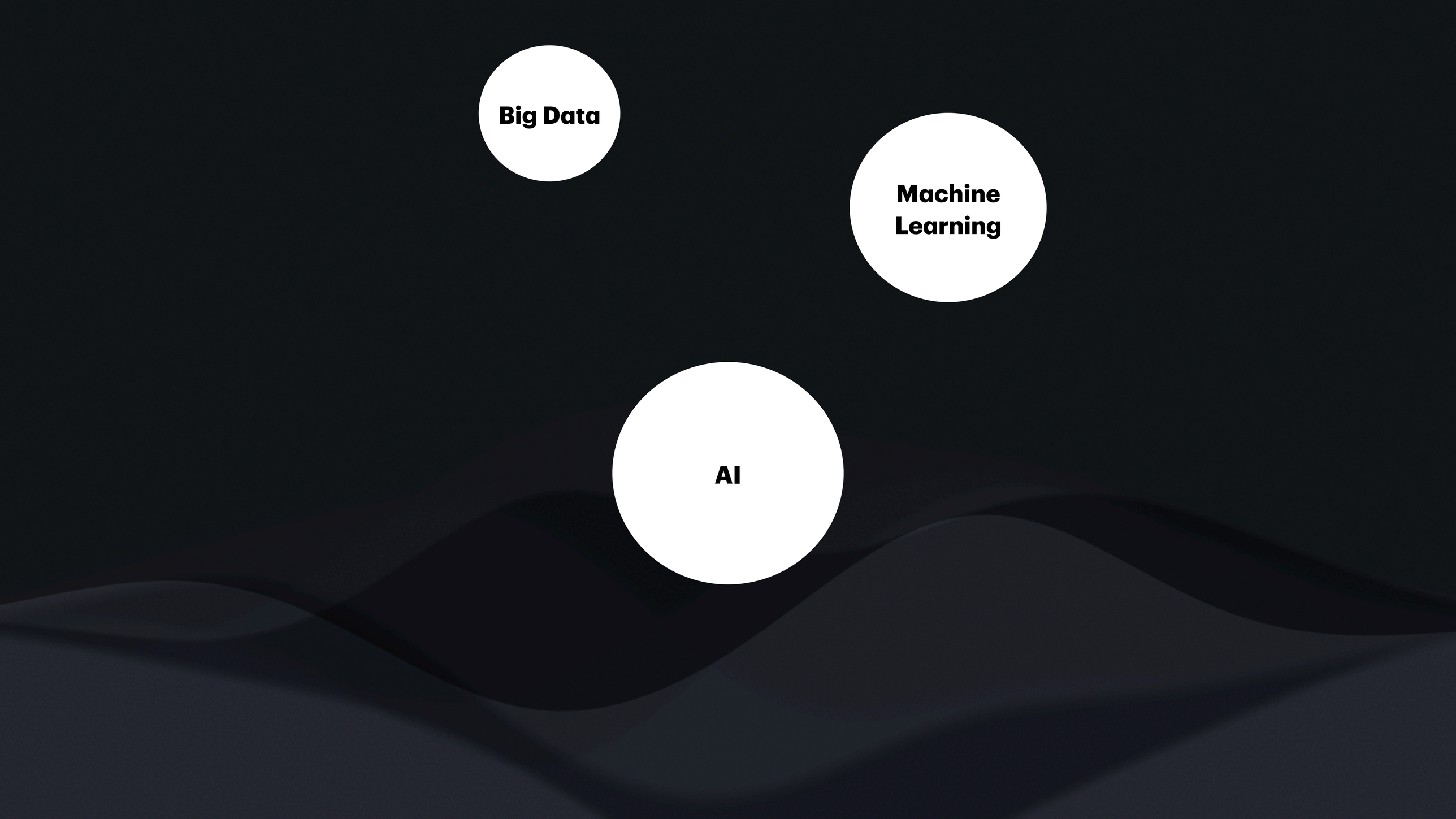
- https://martingasser.com

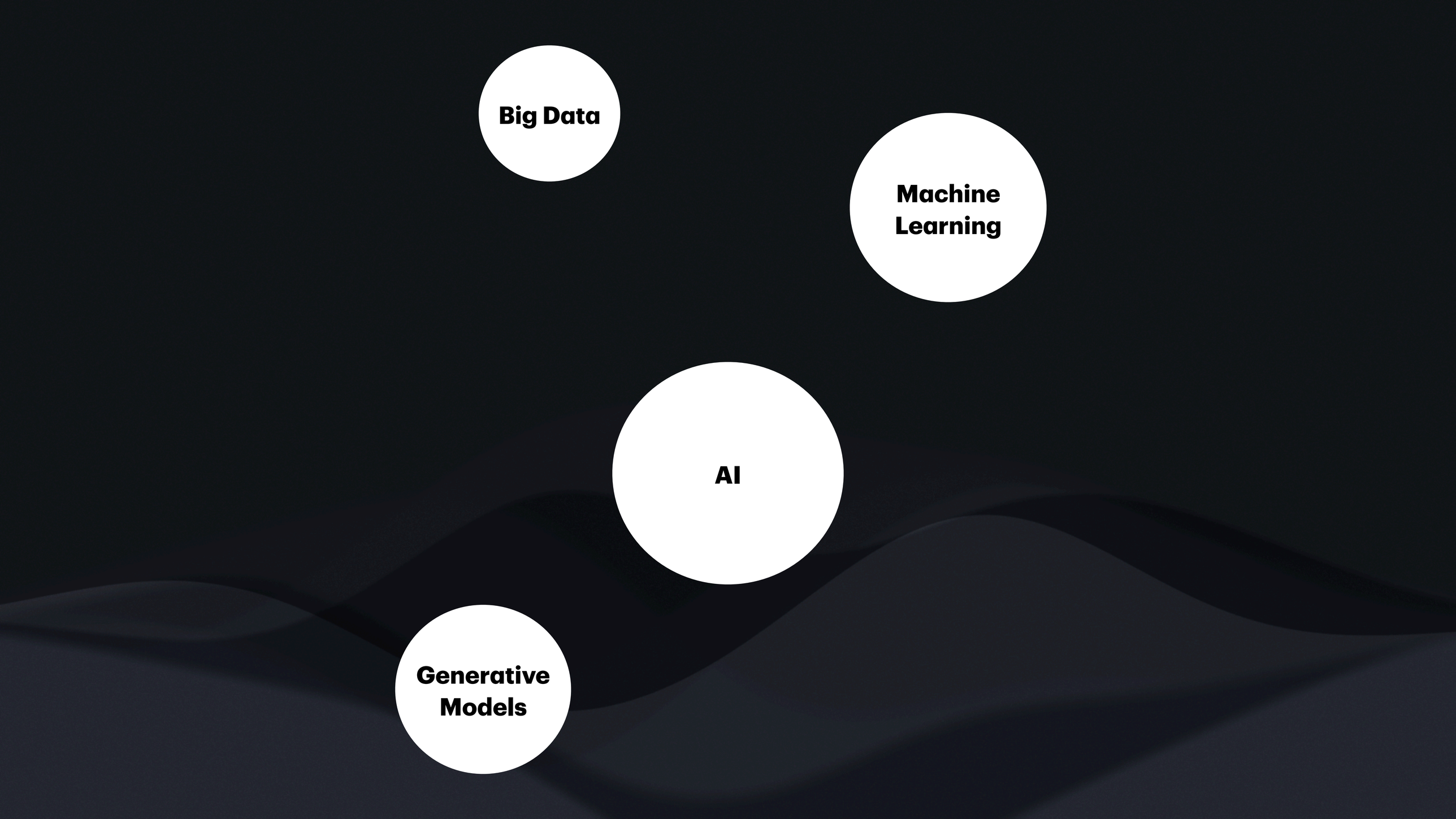
Motivation

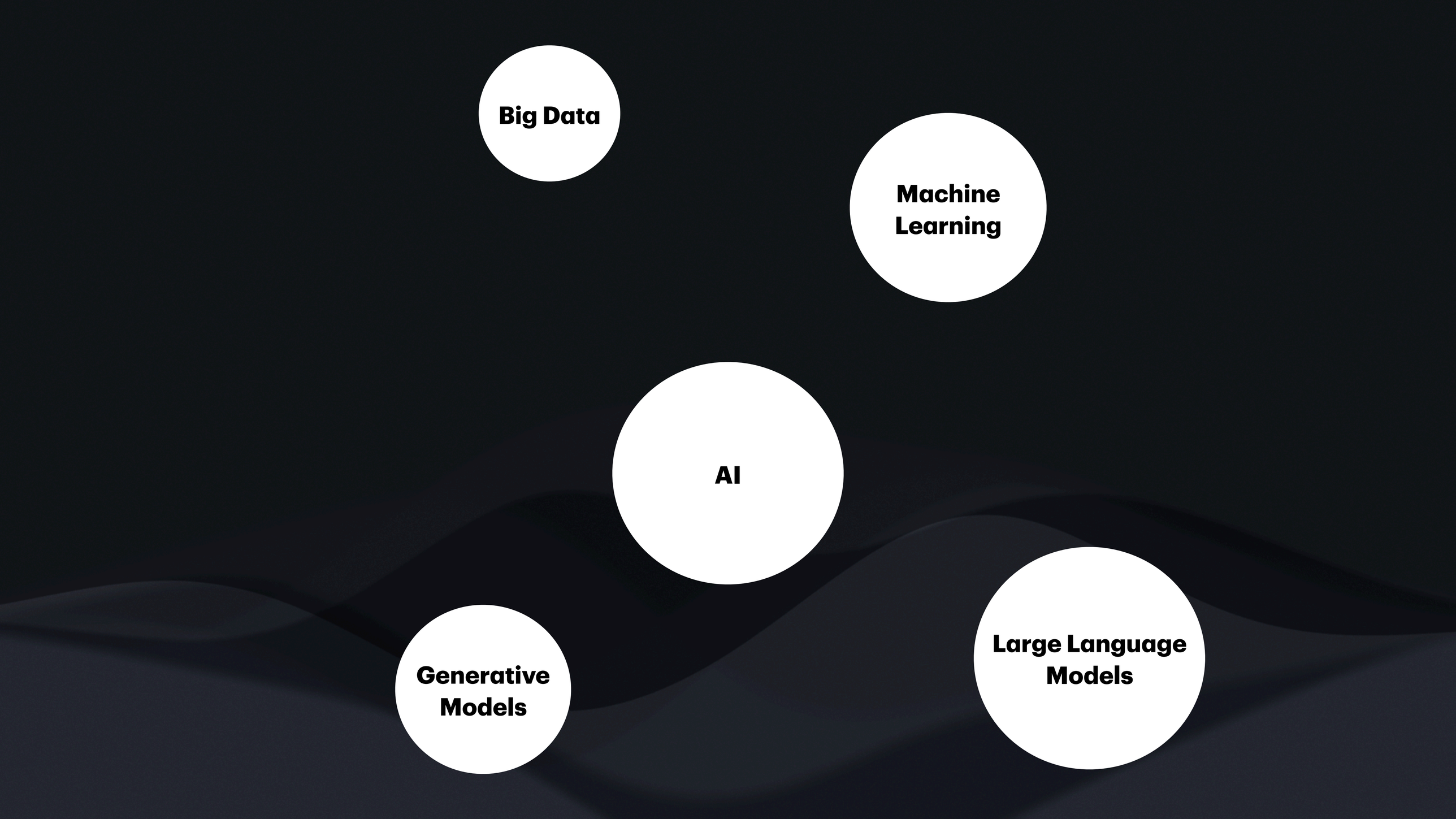
- How can computers understand/produce music?
- Which tasks in a music production workflow can be accomplished by AI systems?
- What are the ramifications of widespread use of AI systems?
- Which AI-based music creation tools are available?
- Hands-on exploration of tools

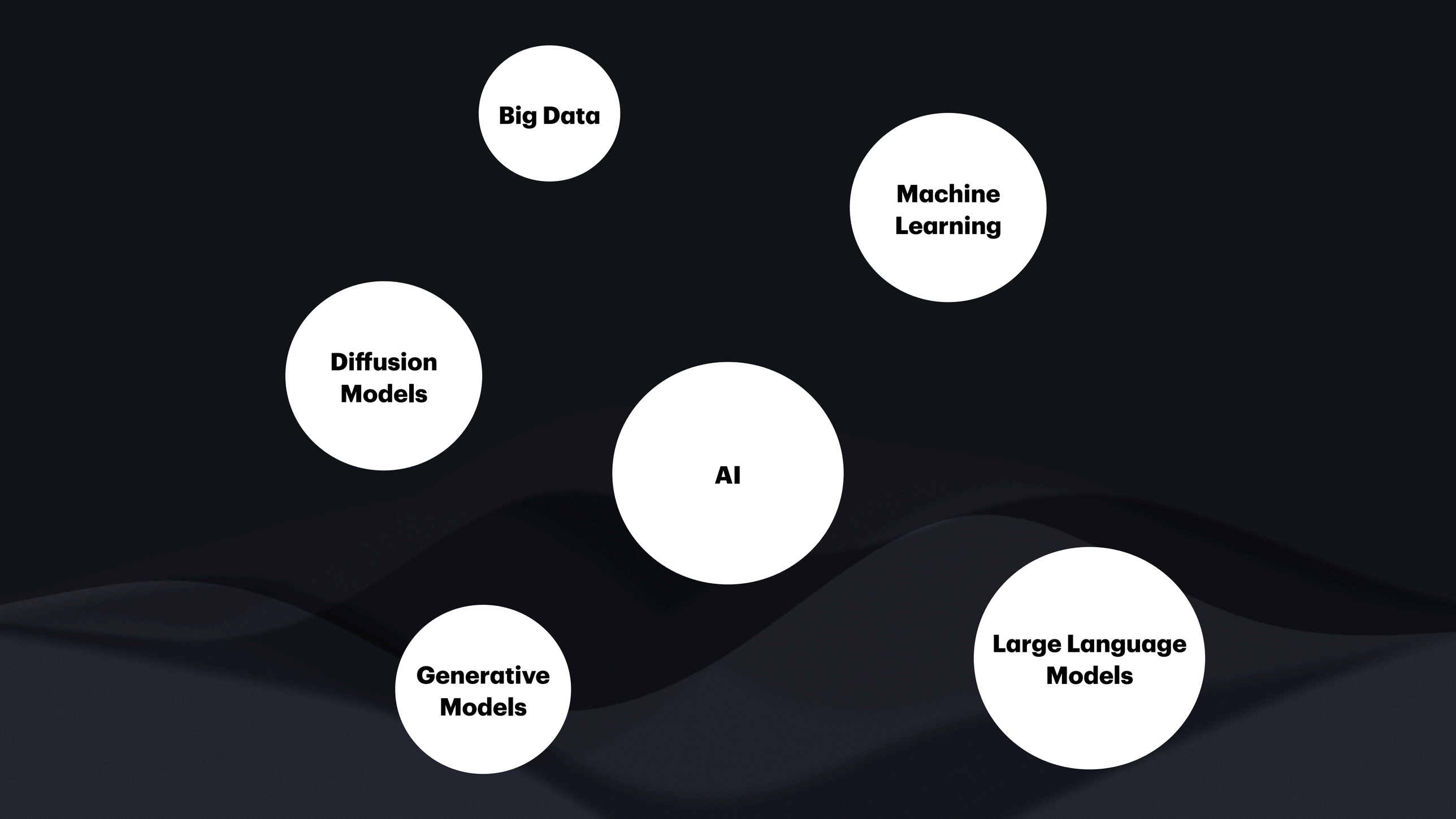
AI


Machine Learning
AI


Big Data
Machine Learning
AI


Big Data
Machine Learning
AI
Generative Models

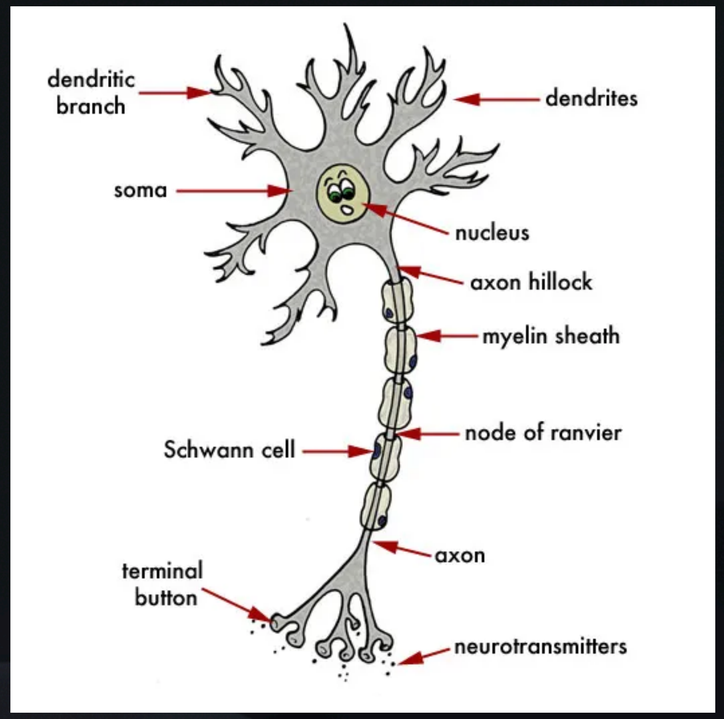

Big Data
Machine Learning
AI
Generative Models
Large Language Models

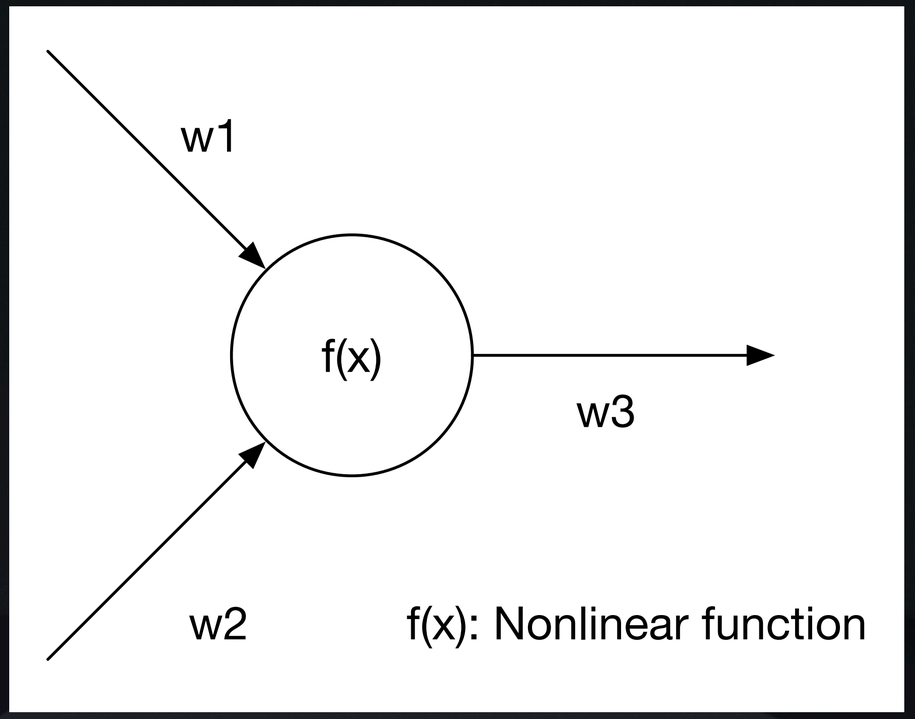

Big Data
Machine Learning
Diffusion Models
AI
Generative Models
Large Language Models

“Naive” Artificial Intelligence

Create a clone of the human mind by meticulously copying the human brain

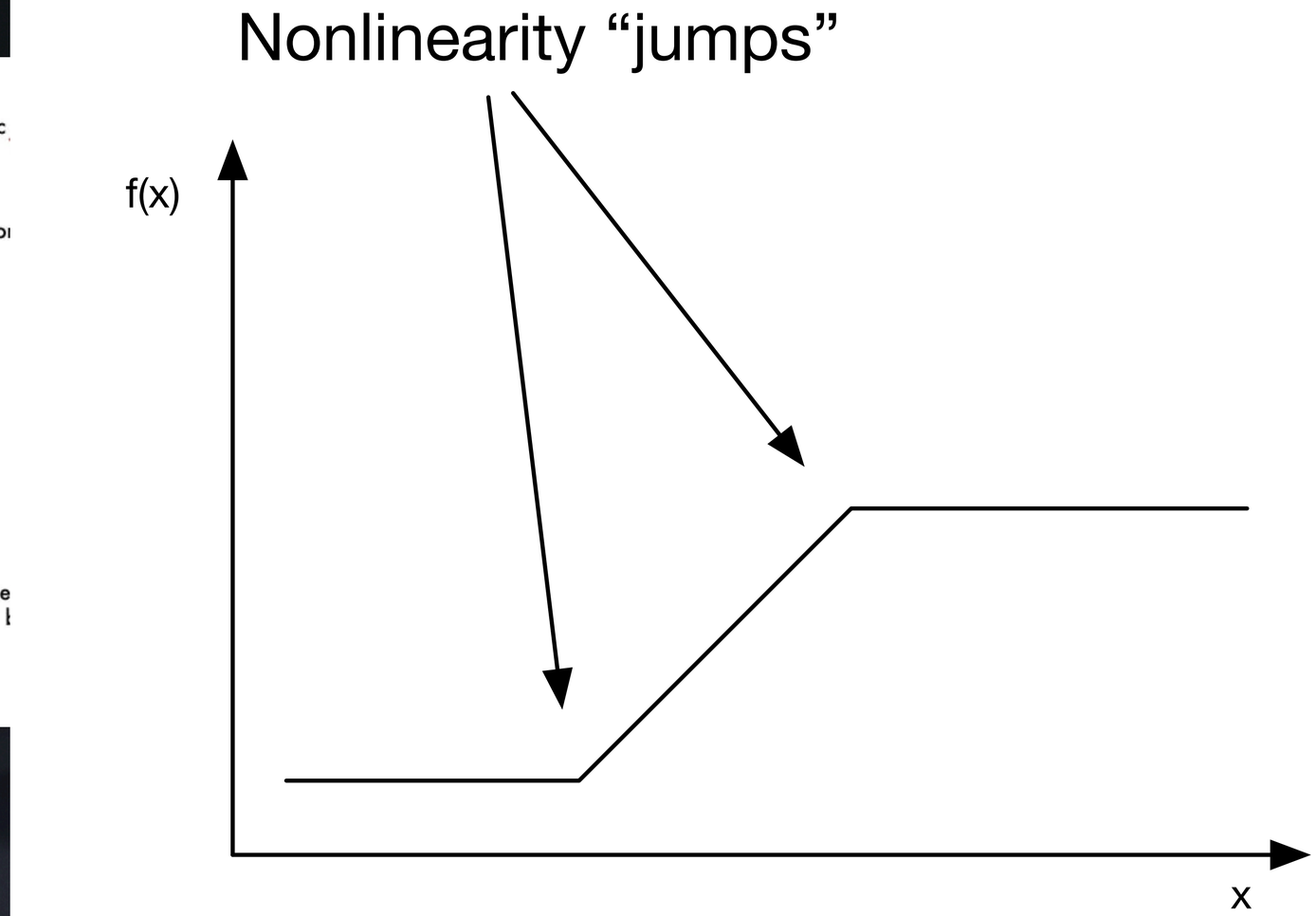
Artificial Neural Networks





f(x) simulates behaviour of cell nucleus (action potential)

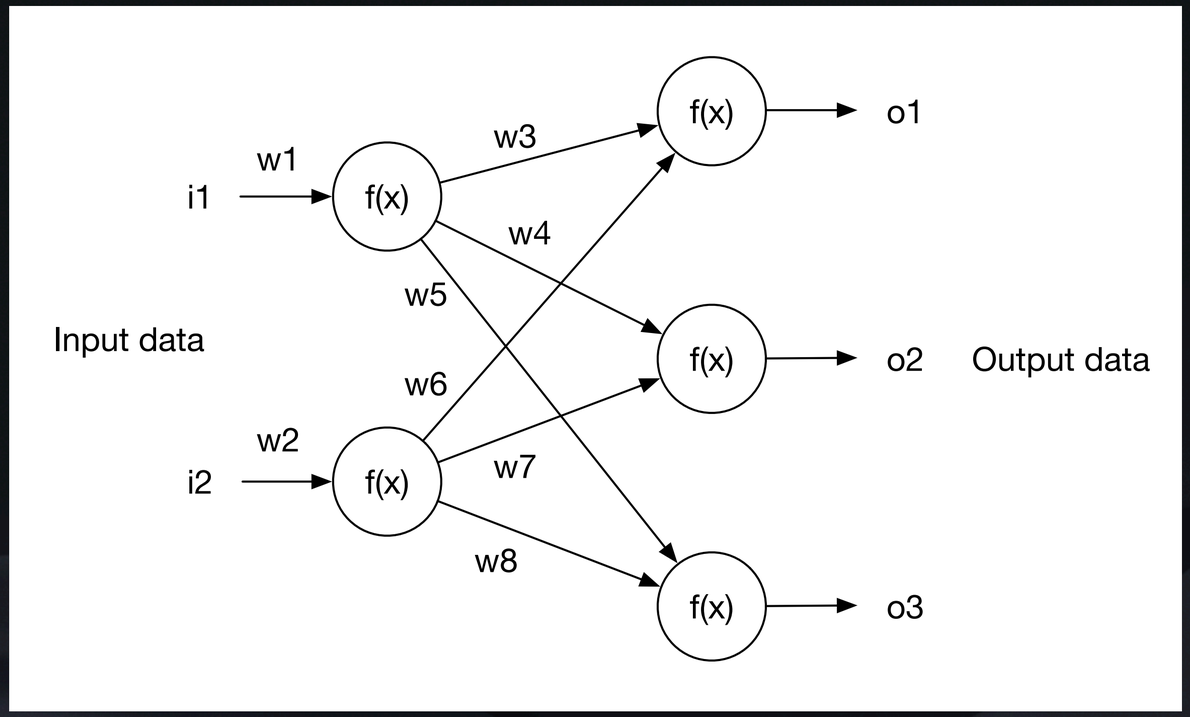
Artificial Neural Networks



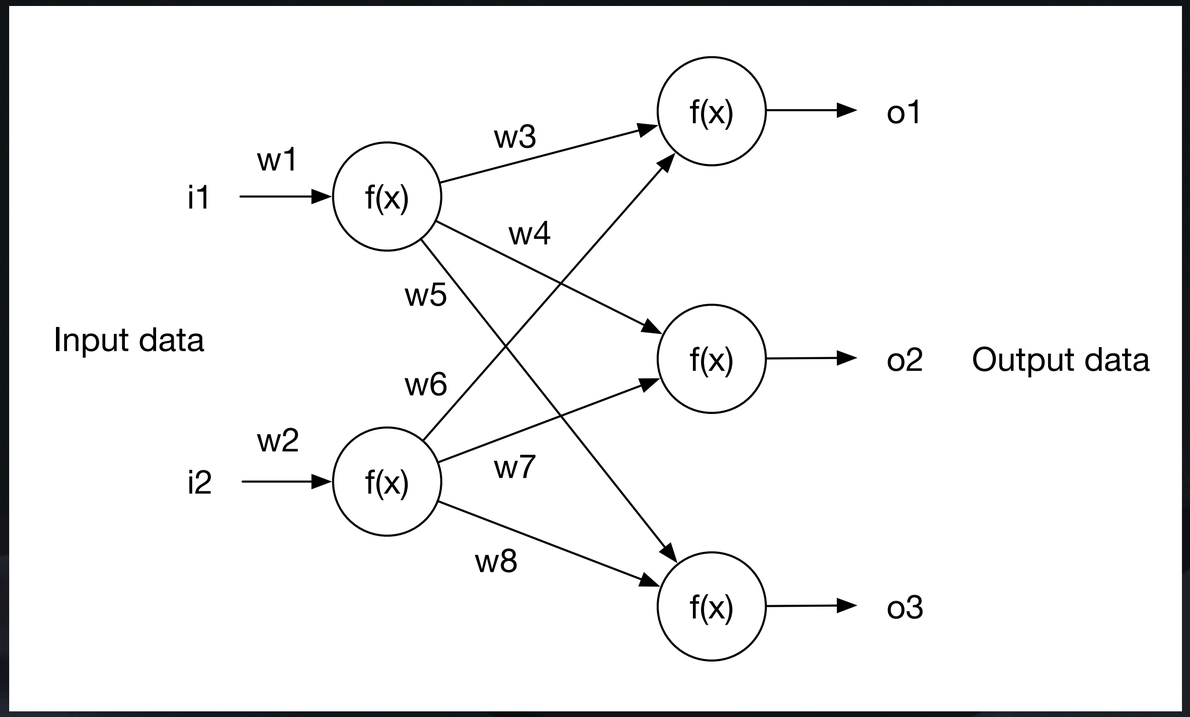
nction

f(x) simulates behaviour of cell nucleus (action potential)

Operation of neural networks (NN)

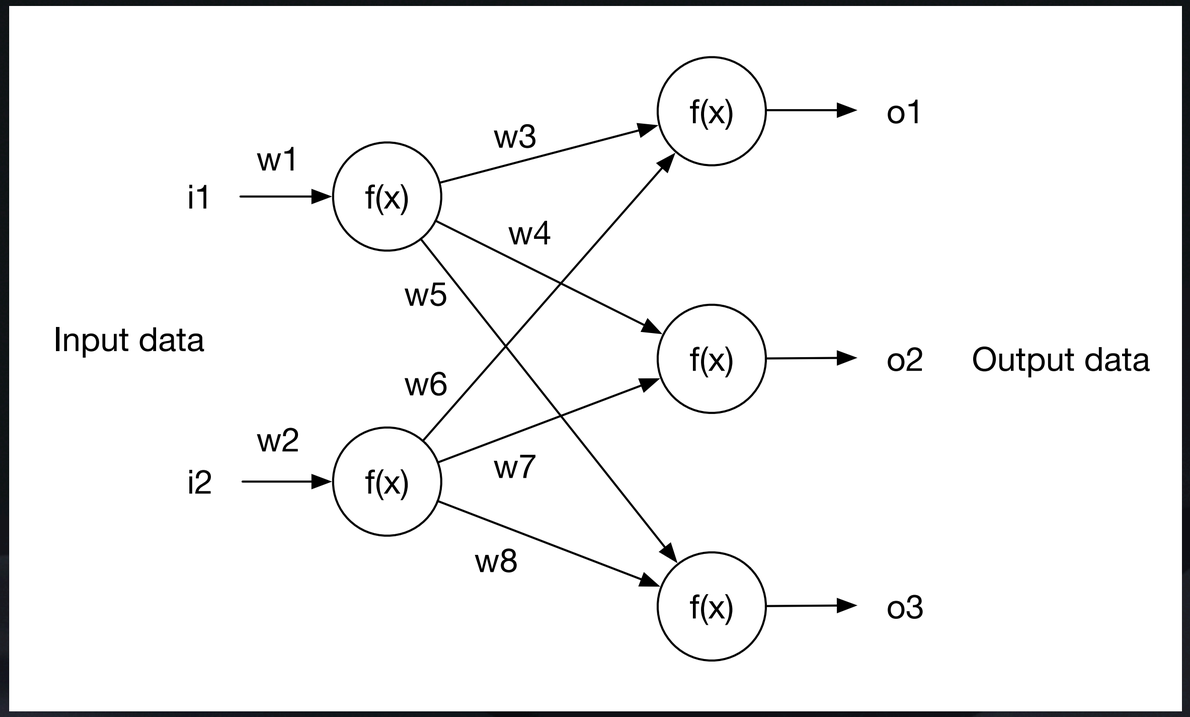


Operation of neural networks (NN)



$$o_1 = f(w_3\, f(w_1 i_1) + w_6\, f(w_2 i_2))$$
$$o_2 = f(w_4\, f(w_1 i_1) + w_7\, f(w_2 i_2))$$
$$o_3 = f(w_5\, f(w_1 i_1) + w_8\, f(w_2 i_2))$$

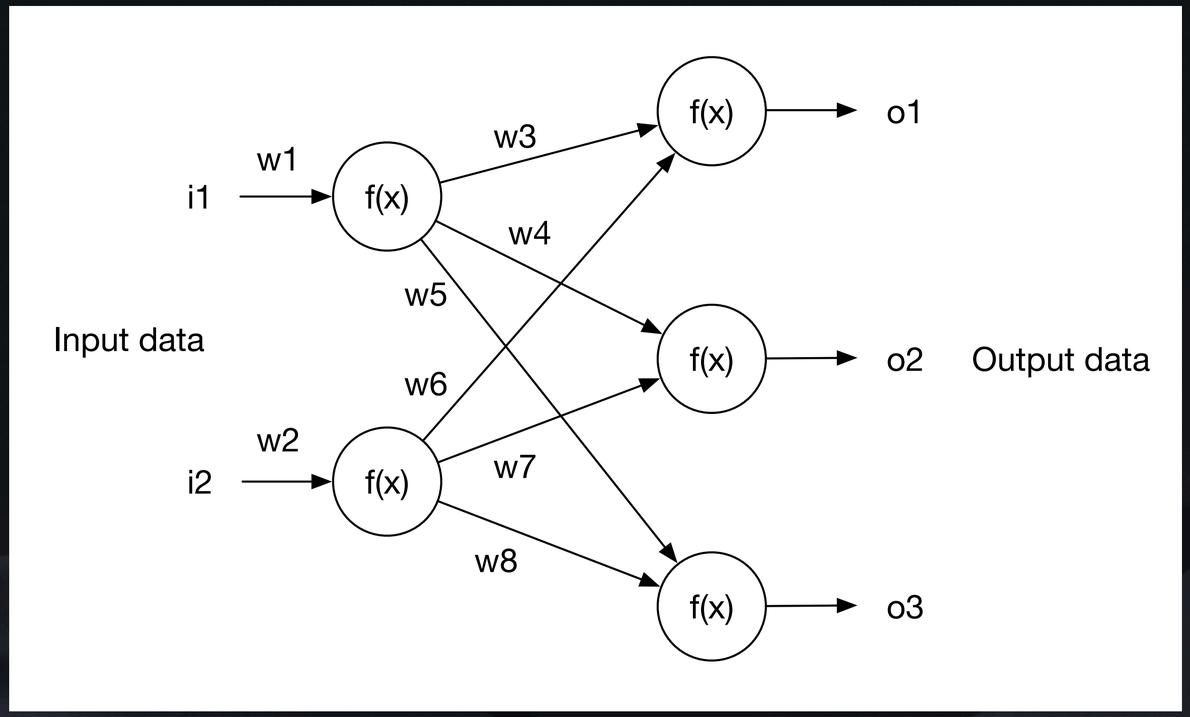
Operation of neural networks (NN)



$o_1 = f(w_3\, f(w_1 i_1) + w_6\, f(w_2 i_2))$
$o_2 = f(w_4\, f(w_1 i_1) + w_7\, f(w_2 i_2))$
$o_3 = f(w_5\, f(w_1 i_1) + w_8\, f(w_2 i_2))$

NN training:
➡ known input and output data
➡ modify weights accordingly

Operation of neural networks (NN)



$o_1 = f(w_3\, f(w_1 i_1) + w_6\, f(w_2 i_2))$
$o_2 = f(w_4\, f(w_1 i_1) + w_7\, f(w_2 i_2))$
$o_3 = f(w_5\, f(w_1 i_1) + w_8\, f(w_2 i_2))$

NN training:
➡ known input and output data
➡ modify weights accordingly

NN inference:
➡infer output data from input data

Problems...

Problems...

Training algorithms?

Problems...

Training algorithms?

Data?

Problems...

Training algorithms?

Data?

Computing resources?

Problems...

AI winter

Consequences

Neural Networks were **not taken seriously** for a long time

Dominance of

- Statistical methods
- Classical pattern recognition
- Rule-based systems

Big Bang

Big Bang

Cheap and powerful GPU’s

Big Bang

Cheap and powerful GPU's

+

Big Data

Big Bang

Cheap and powerful GPU's

+

Big Data

=

Deep Learning Revolution

Deep Learning timeline

- 2012: Deep Learning outperforms classical image recognition algorithms
- 2014: Invention of Generative Adversarial Networks
- 2015: Google Deep Dream, StyleGAN
- 2017: Attention mechanism, transformers
- 2018: GPT, Large Language Models (LLM's)
- 2022: Stable Diffusion, ChatGPT
- ....what's next?

AI in music (1)

Music Information Retrieval: Use AI/ML technology to analyze and organize music

Examples:

- Music Recommendation Systems (Spotify, Apple Music...)
- Music teaching (MuseClass, MakeMusic,...)
- Music transcription (Basic Pitch)

Music Recommendation

Find **similar** music

Similarity

- Based on textual descriptions / metadata
- Based on what other people listened to
- Based on what the music sounds like

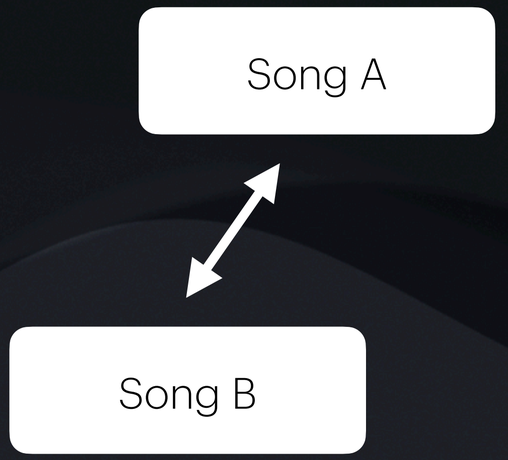
Song A

Music Recommendation

Find **similar** music

Similarity

- Based on textual descriptions / metadata
- Based on what other people listened to
- Based on what the music sounds like

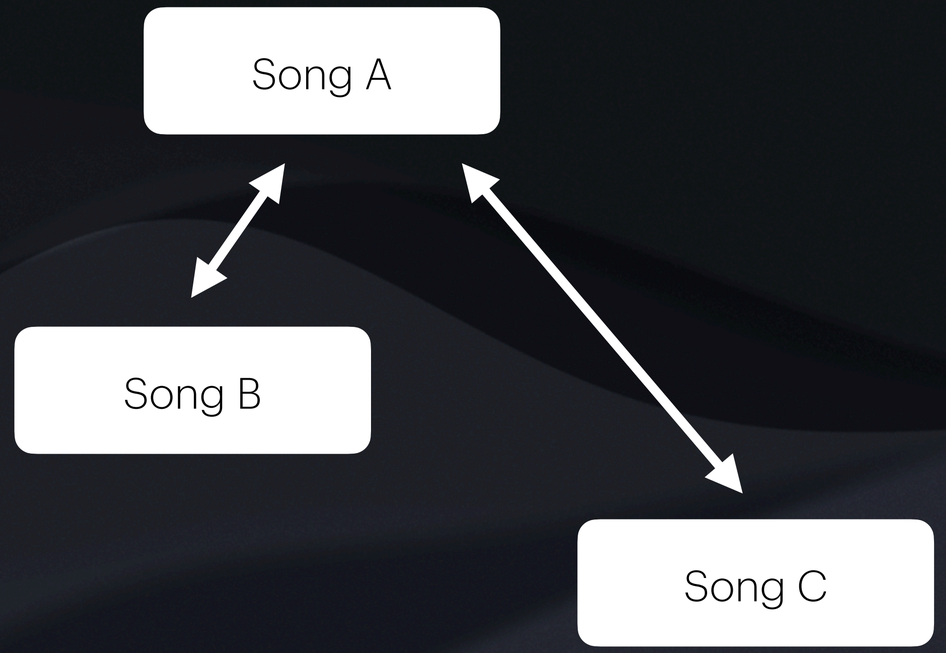


Music Recommendation

Find **similar** music

Similarity

- Based on textual descriptions / metadata
- Based on what other people listened to
- Based on what the music sounds like

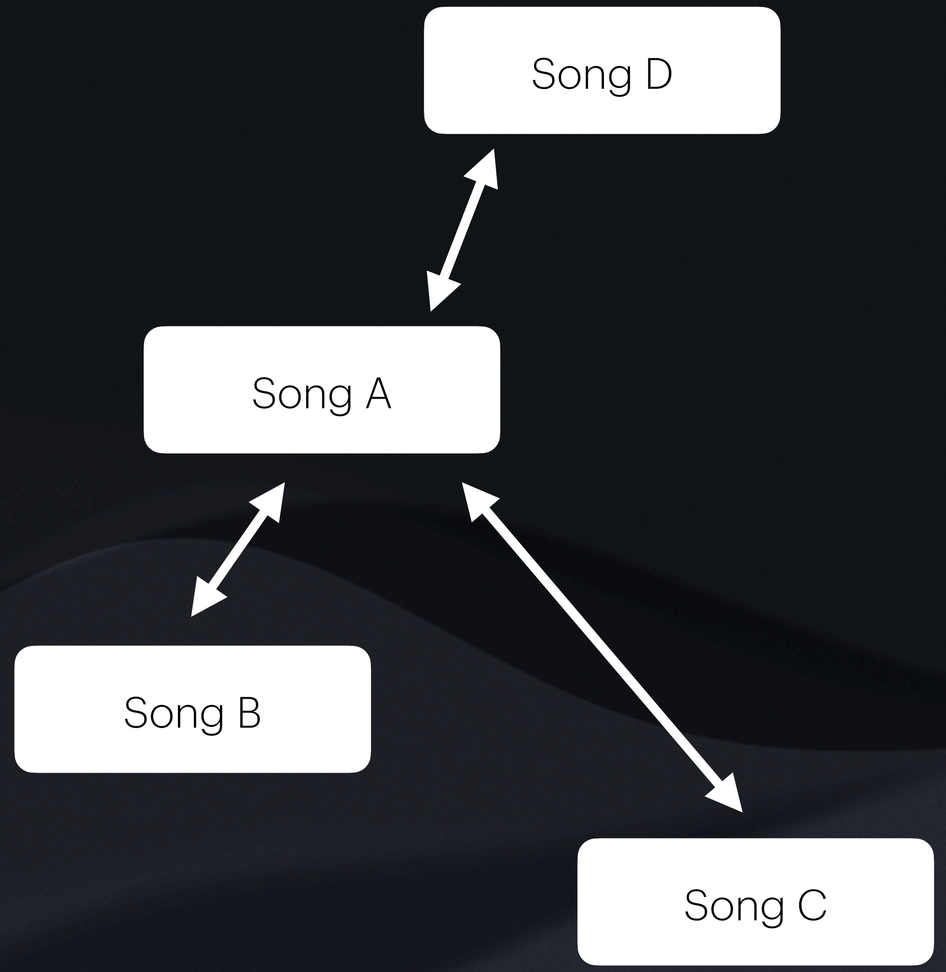


Music Recommendation

Find **similar** music

Similarity

- Based on textual descriptions / metadata
- Based on what other people listened to
- Based on what the music sounds like



Music teaching

Computer-supported music teaching

Teachers can distribute exercises to students on a web platform

Students can use an app to practice, app provides immediate automatic feedback after practice sessions

Music teaching

Computer

Teachers c

Students c ıfter
practice se

Music transcription

Transform music from audio to “notation”

Upload audio

Get back a MIDI file

MIDI file can be subsequently modified and used in other applications

Music transcription

Spotify R&D

Transforn

Upload a

Get back

MIDI file c

AI in music (2)

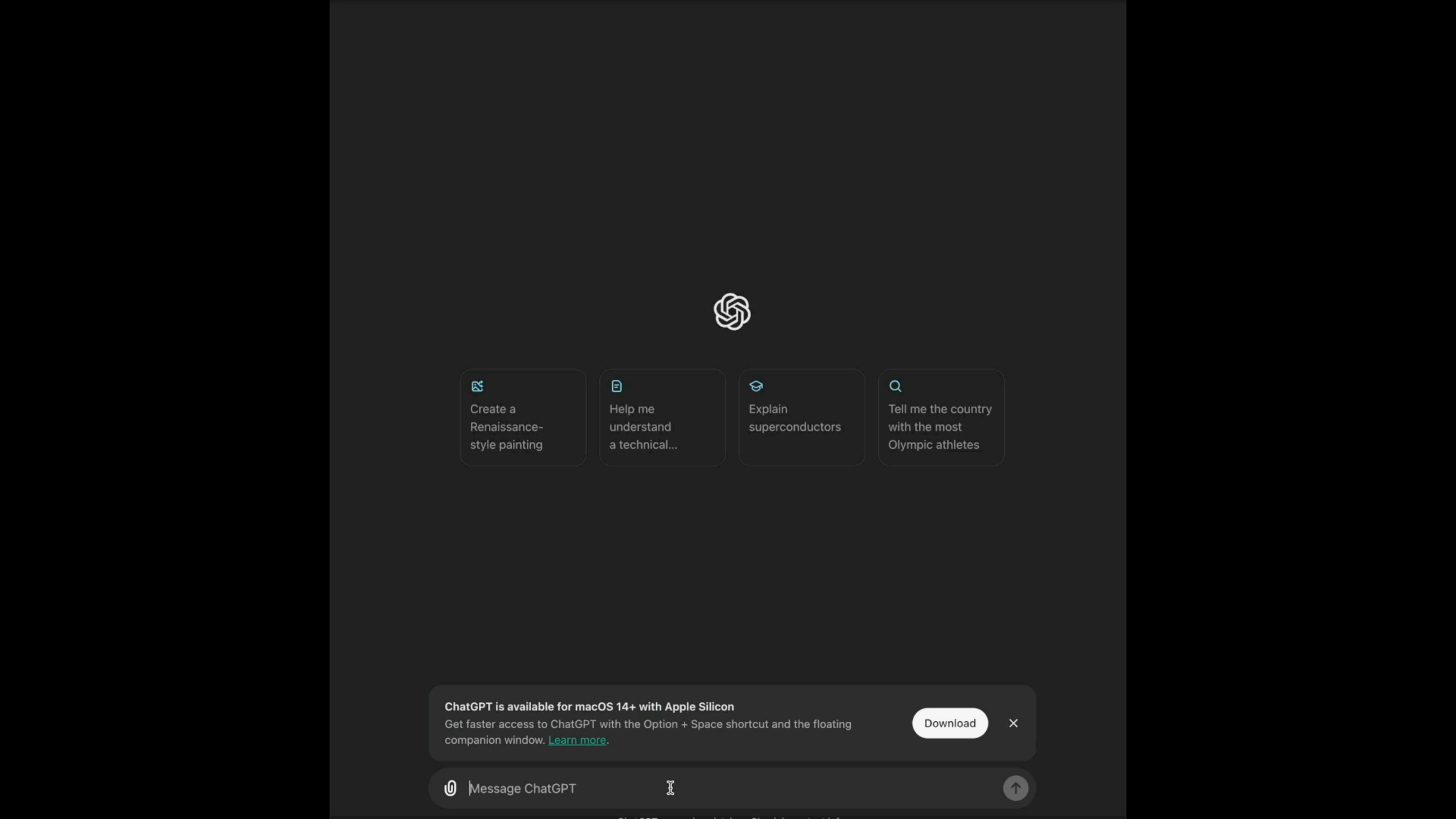
Generative systems/composition assistants

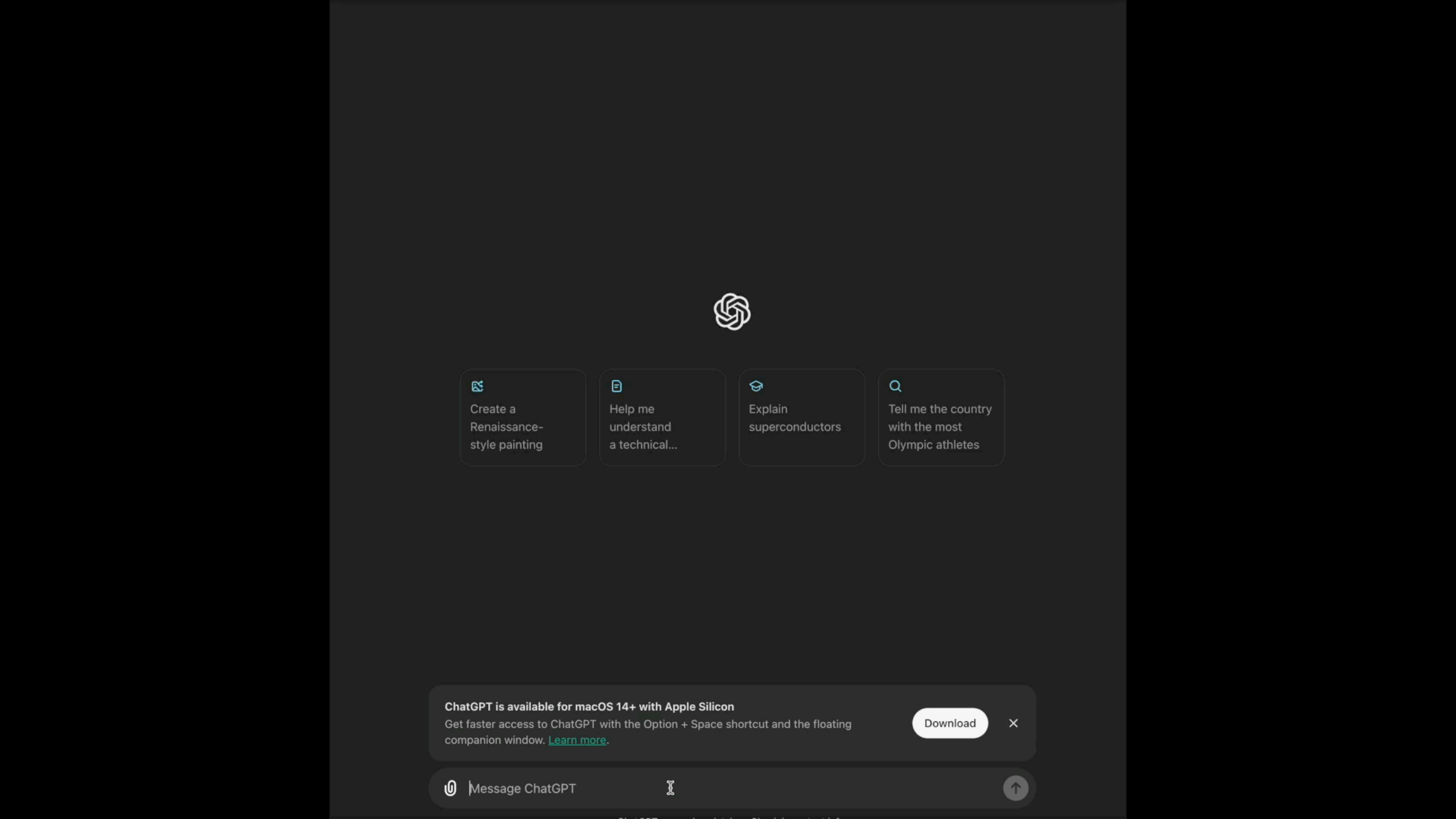
Commercial DAW's start integrating AI features (e.g., Apple Logic Pro)

ChatGPT can write (simple) chord progressions and melodies

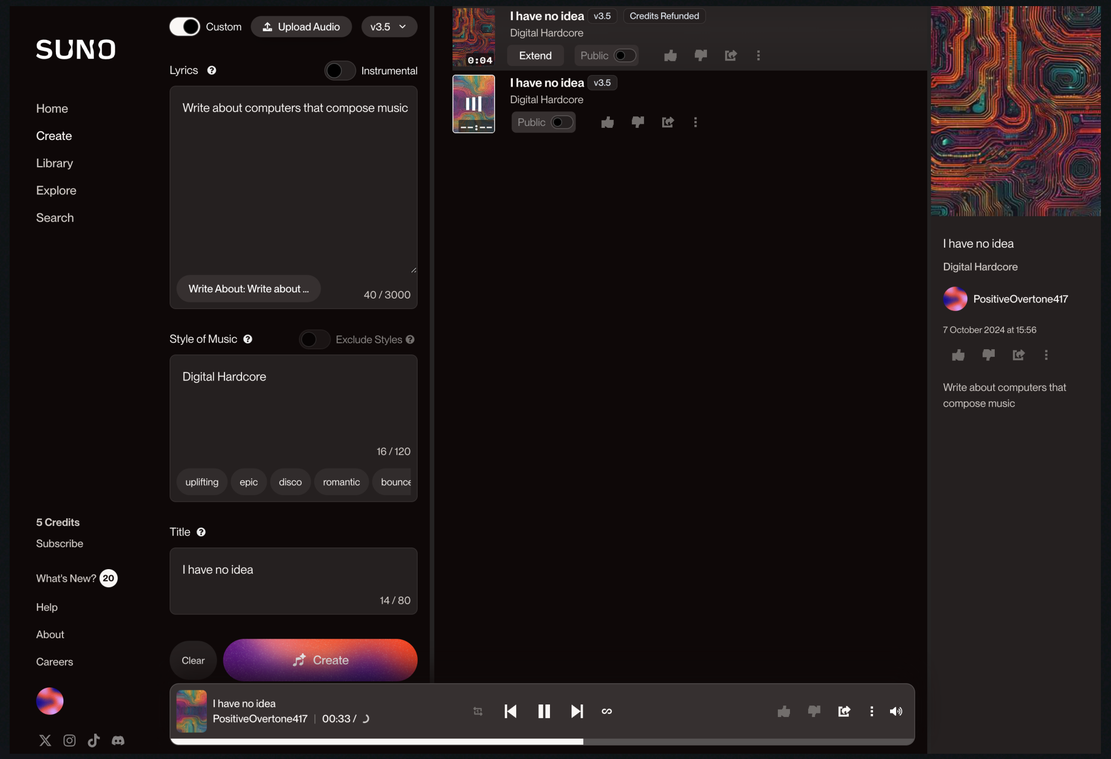
Text to music/sound: suno.ai, Stable Audio

Mixing/mastering assistants: LANDR, Cryo Mix


Create a Renaissance-style painting
Help me understand a technical...
Explain superconductors
Tell me the country with the most Olympic athletes
ChatGPT is available for macOS 14+ with Apple Silicon
Get faster access to ChatGPT with the Option + Space shortcut and the floating companion window. Learn more.
Download
Message ChatGPT


Create a Renaissance-style painting
Help me understand a technical...
Explain superconductors
Tell me the country with the most Olympic athletes
ChatGPT is available for macOS 14+ with Apple Silicon
Get faster access to ChatGPT with the Option + Space shortcut and the floating companion window. Learn more.
Download
Message ChatGPT

suno.ai



Generate music from textual descriptions

Consists of

- a **large language model** for understanding user prompts
- a **diffusion model** for music generation

High-quality output

Can even generate **convincing human voices**

suno.ai

Tips:

Use **custom mode** to get more control over lyrics and song structure

Try to let ChatGPT **write lyrics** for a song

Use **structural tags** like [Intro], [Verse], [Refrain] for lyrics

Make **cover songs** (style “Indie Folk” and “Dance Rock”)

See https://sunoaiwiki.com/

suno.ai

Tips:

Use **custom mode** to get more control over lyrics and song structure

Try to let ChatGPT **write lyrics** for a song

Use **structural tags** like [Intro], [Verse], [Refrain] for lyrics

Make **cover songs** (style “Indie Folk” and “Dance Rock”)

See https://sunoaiwiki.com/

suno.ai

Tips:

Use **custom mode** to get more control over lyrics and song structure

Try to let ChatGPT **write lyrics** for a song

Use **structural tags** like [Intro], [Verse], [Refrain] for lyrics

Make **cover songs** (style “Indie Folk” and “Dance Rock”)

See https://sunoaiwiki.com/

suno.ai

Tips:

Use **custom mode** to get more control over lyrics and song structure

Try to let ChatGPT **write lyrics** for a song

Use **structural tags** like [Intro], [Verse], [Refrain] for lyrics

Make **cover songs** (style “Indie Folk” and “Dance Rock”)

See https://sunoaiwiki.com/

suno.ai

Tips:

Use **custom mode** to get more control over lyrics and song structure

Try to let ChatGPT **write lyrics** for a song

Use **structural tags** like [Intro], [Verse], [Refrain] for lyrics

Make **cover songs** (style “Indie Folk” and “Dance Rock”)

See https://sunoaiwiki.com/

suno.ai

Tips:

Use **custom mode** to get more control over lyrics and song structure

Try to let ChatGPT **write lyrics** for a song

Use **structural tags** like [Intro], [Verse], [Refrain] for lyrics

Make **cover songs** (style “Indie Folk” and “Dance Rock”)

See https://sunoaiwiki.com/

suno.ai

Tips:

Use **custom mode** to get more control over lyrics and song structure

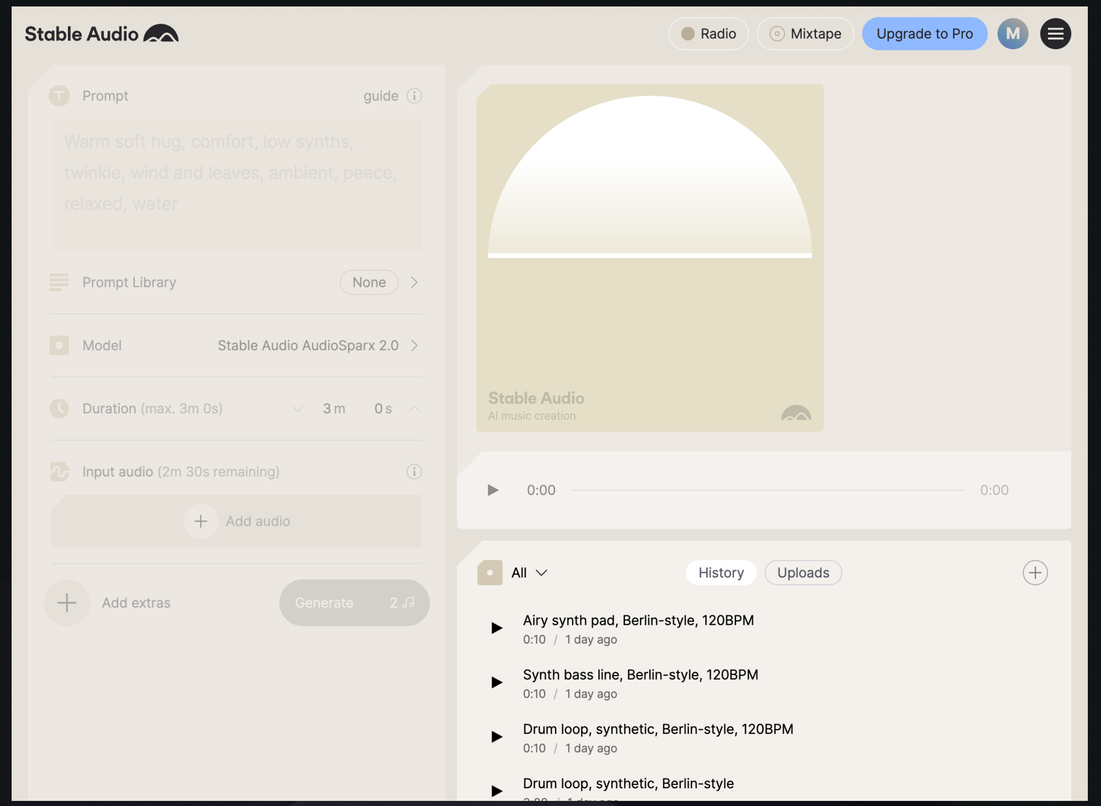
Try to let ChatGPT **write lyrics** for a song

Use **structural tags** like [Intro], [Verse], [Refrain] for lyrics

Make **cover songs** (style “Indie Folk” and “Dance Rock”)

See https://sunoaiwiki.com/

Stable Audio



Based on **published research**

Transparent with respect to which training data was used

Free/open version can be **self-hosted** and used to build custom applications

Stable Audio

Tips:

More suitable to generate individual **stems** (e.g., a drum track or a bass line)

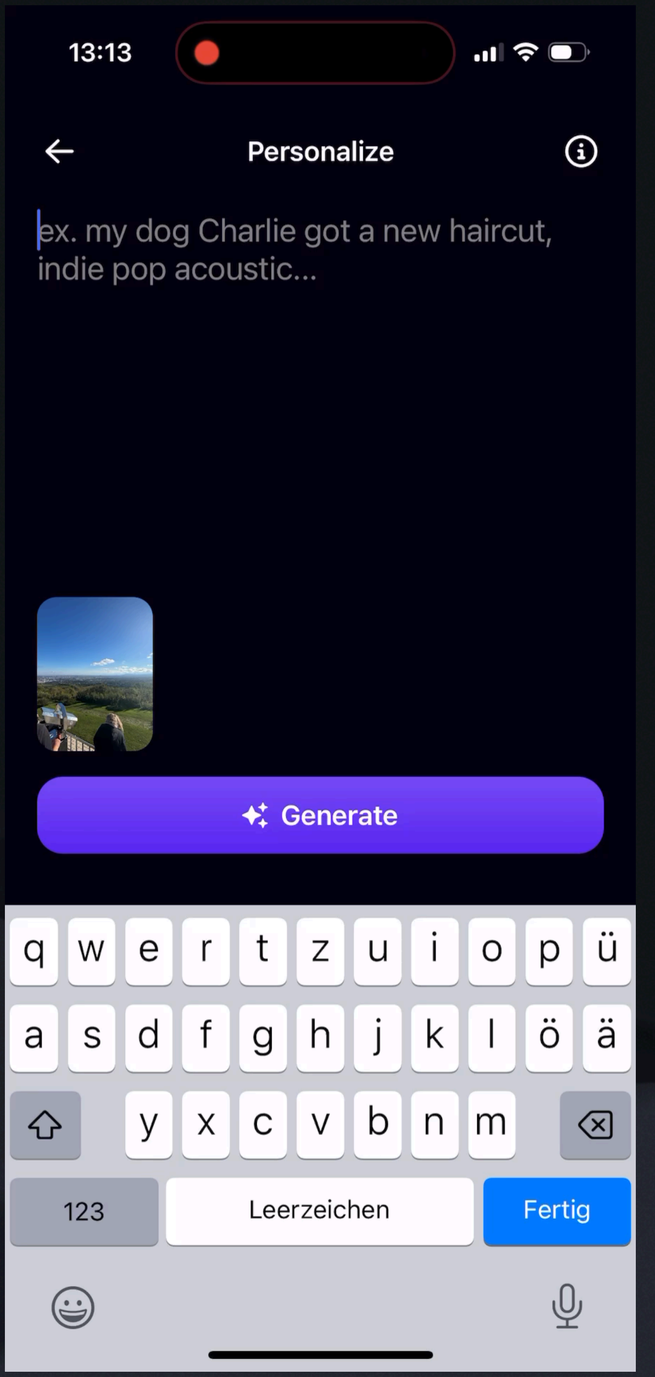
Also very useful for **ambience** and **sound effects**

See https://www.stableaudio.com/user-guide/

Riffusion

Fun way to create lyrics and music based on text prompts and/or photos

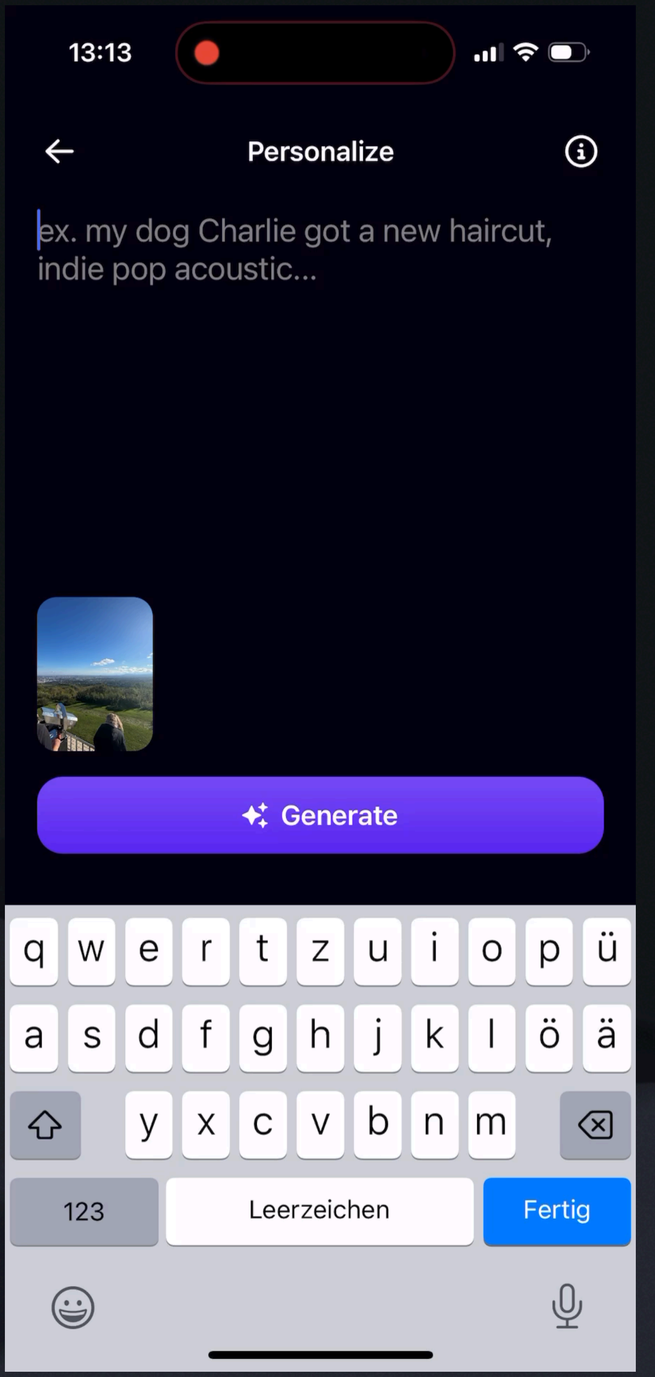
Available as free iOS/Android app



Riffusion

Fun way to create lyrics and music based on text prompts and/or photos

Available as free iOS/Android app



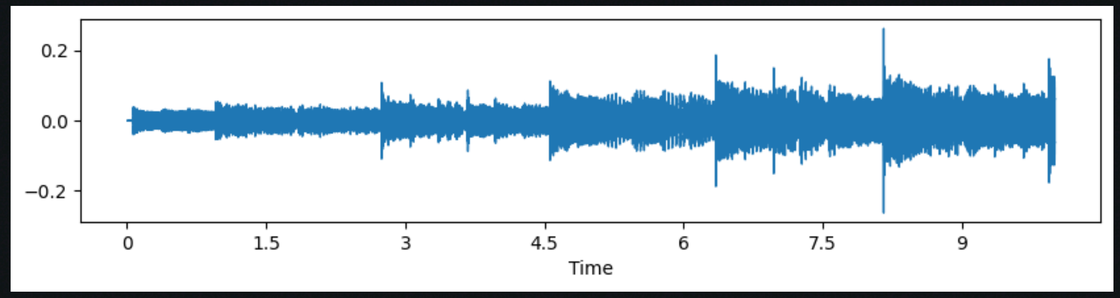
AI in music (3)

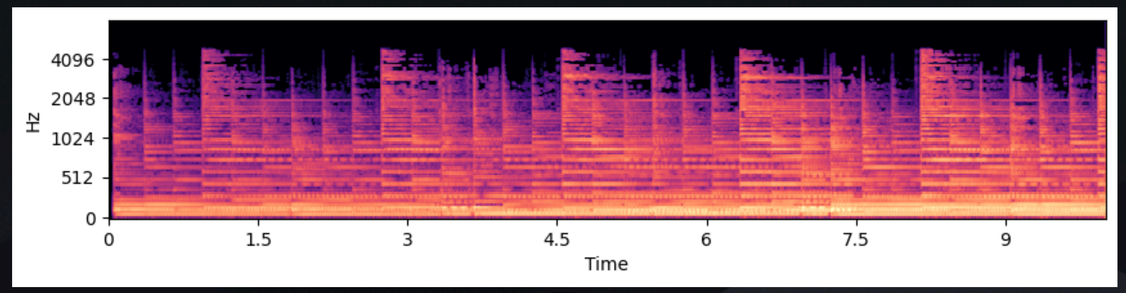
Source separation/track splitting: Spleeter (Colab)

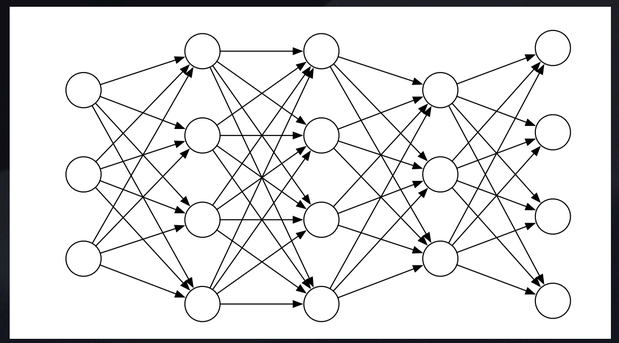
Speech synthesis: murf.ai, OpenAI TTS

Speech recognition: OpenAI Whisper (HF space)

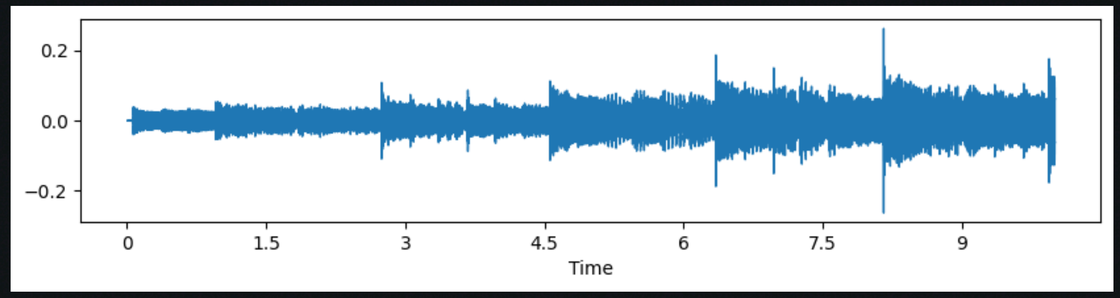
How does it work (analysis)?

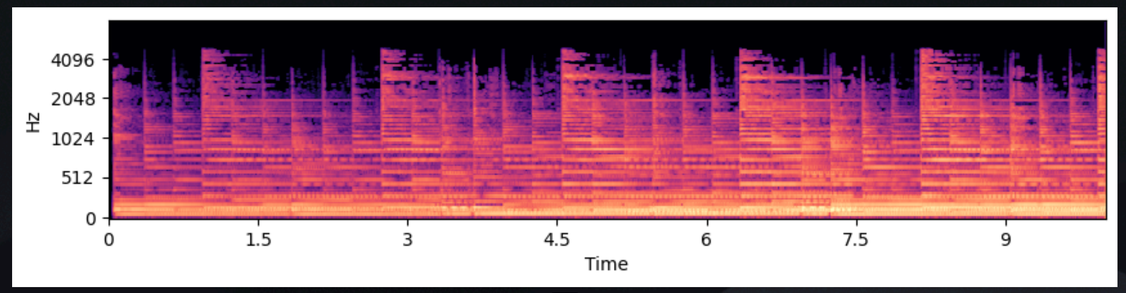


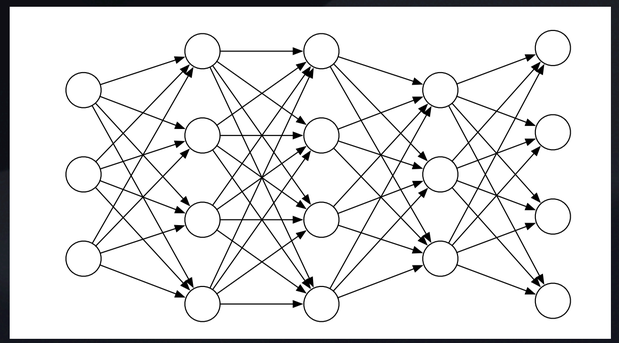




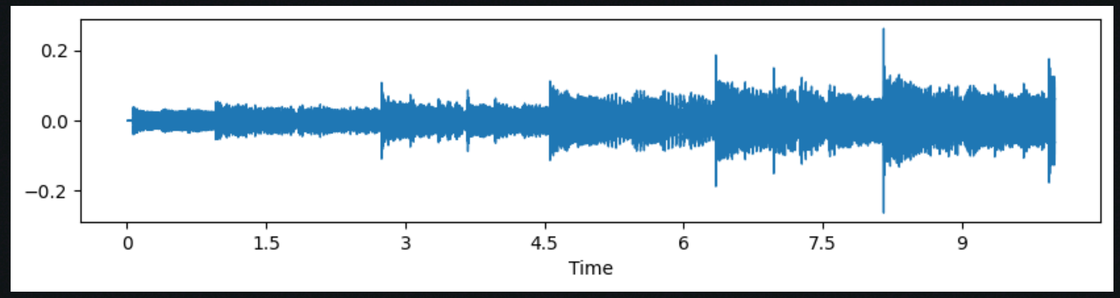
How does it work (analysis)?

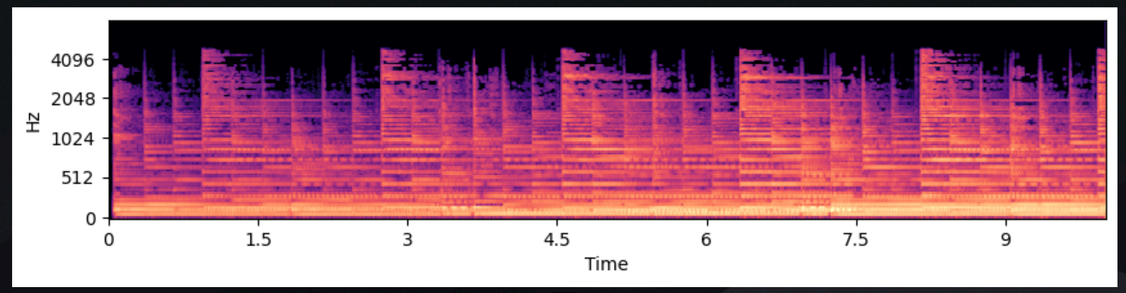


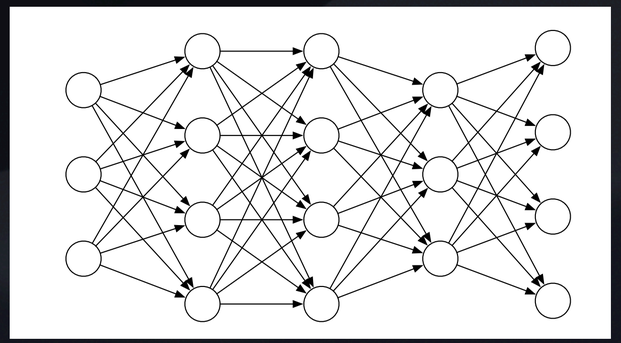




How does it work (analysis)?

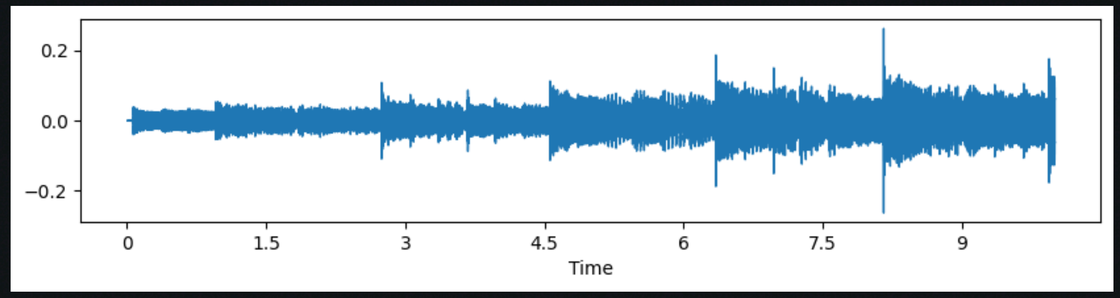


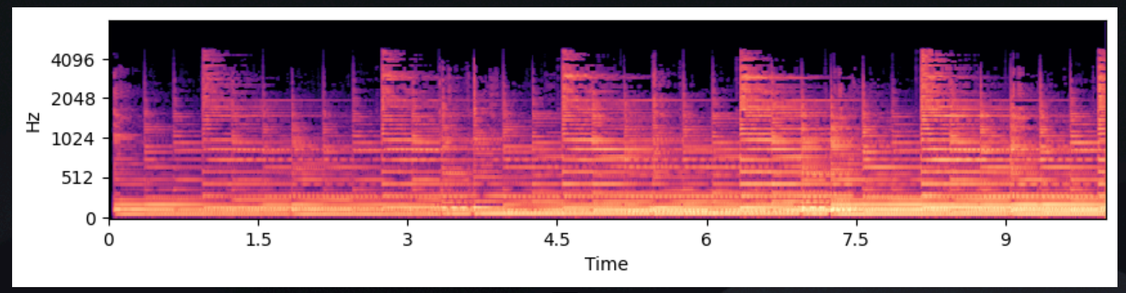


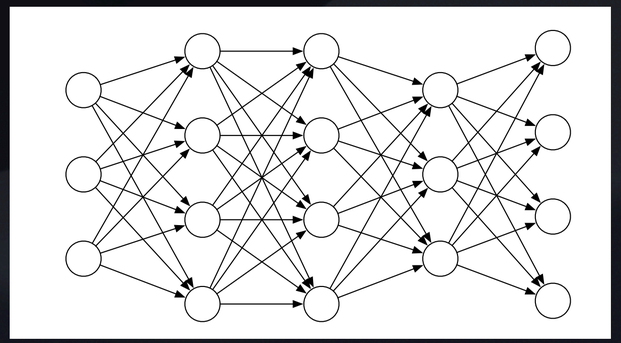


0,2 Jazz

How does it work (analysis)?



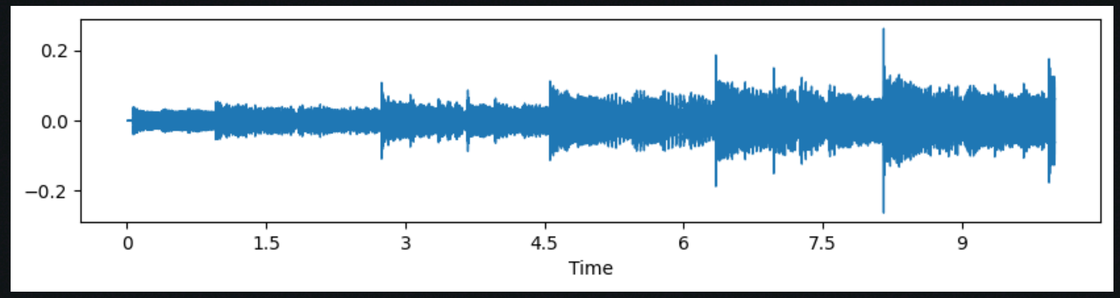


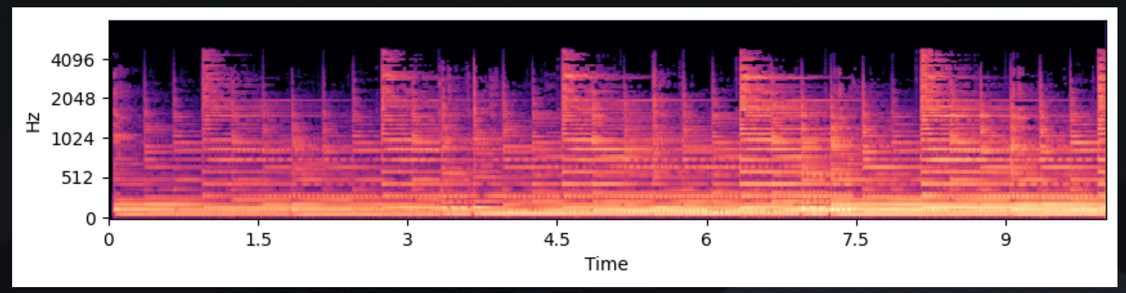


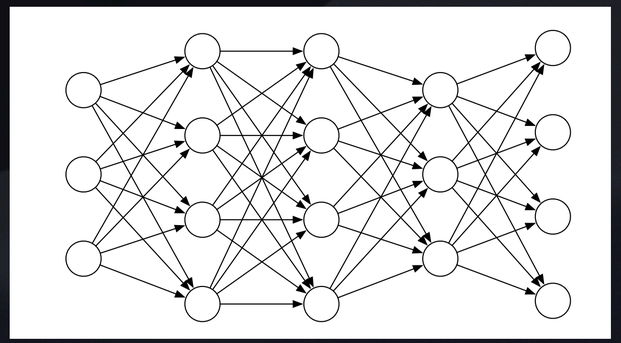
0,2 Jazz

0,7 Rock

How does it work (analysis)?





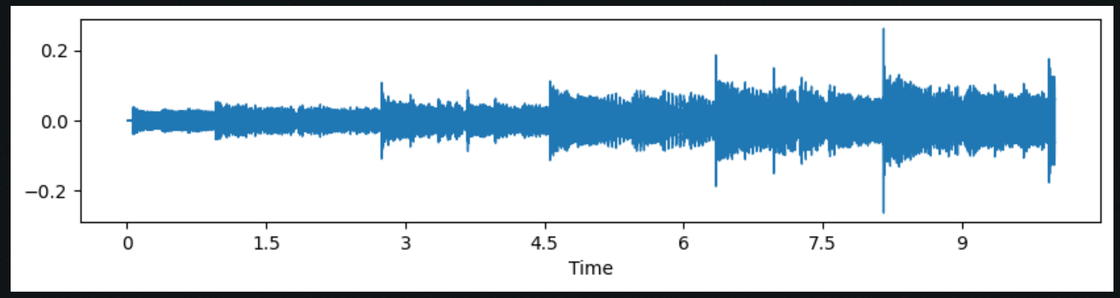


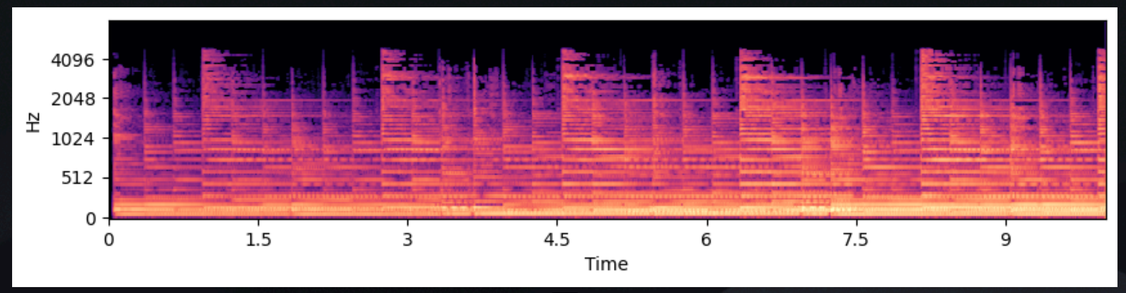
0,2 Jazz

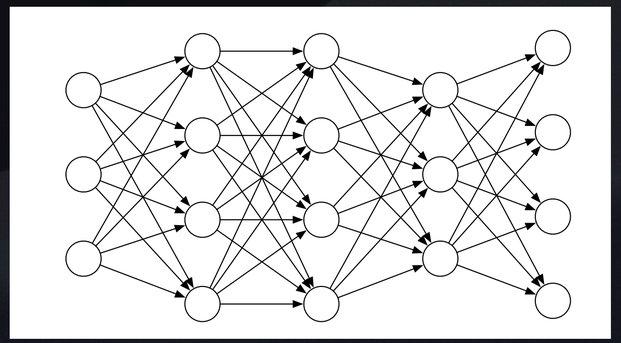
0,7 Rock

0,05 Heavy Metal

How does it work (analysis)?







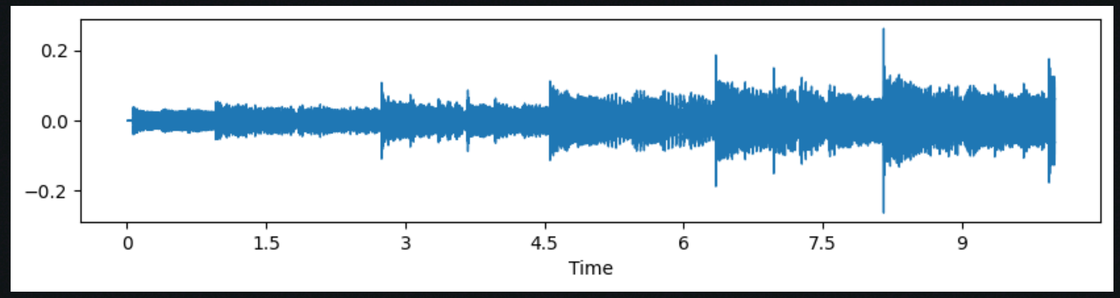
0,2 Jazz

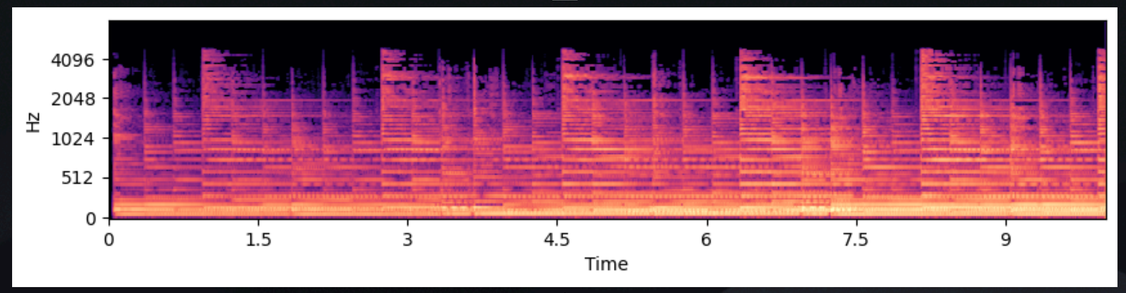
0,7 Rock

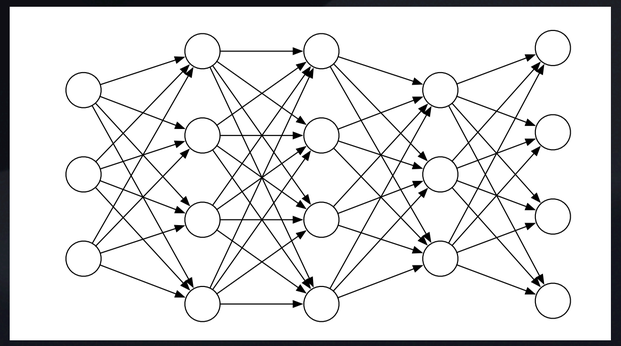
0,05 Heavy Metal

0,05 Classic

How does it work (generation)?







0,2 Jazz

0,7 Rock

0,05 Heavy Metal

0,05 Classic

Generative architectures

Variational autoencoder: Learns to compress media into a **latent space** and reconstruct images/text/audio/...

Generative Adversarial Network: Actually two neural networks

One learns to distinguish fake from real content

The other learns to trick the first one

Diffusion models: Learns how to transform pure noise into meaningful content (state-of-the-art approach)

Real Life-deployments

Consist of **billions** of parameters (weights)

- GPT-4o: ~200 billions
- Stable Audio Open ~1,3 billion parameters

Lots of data needed for training

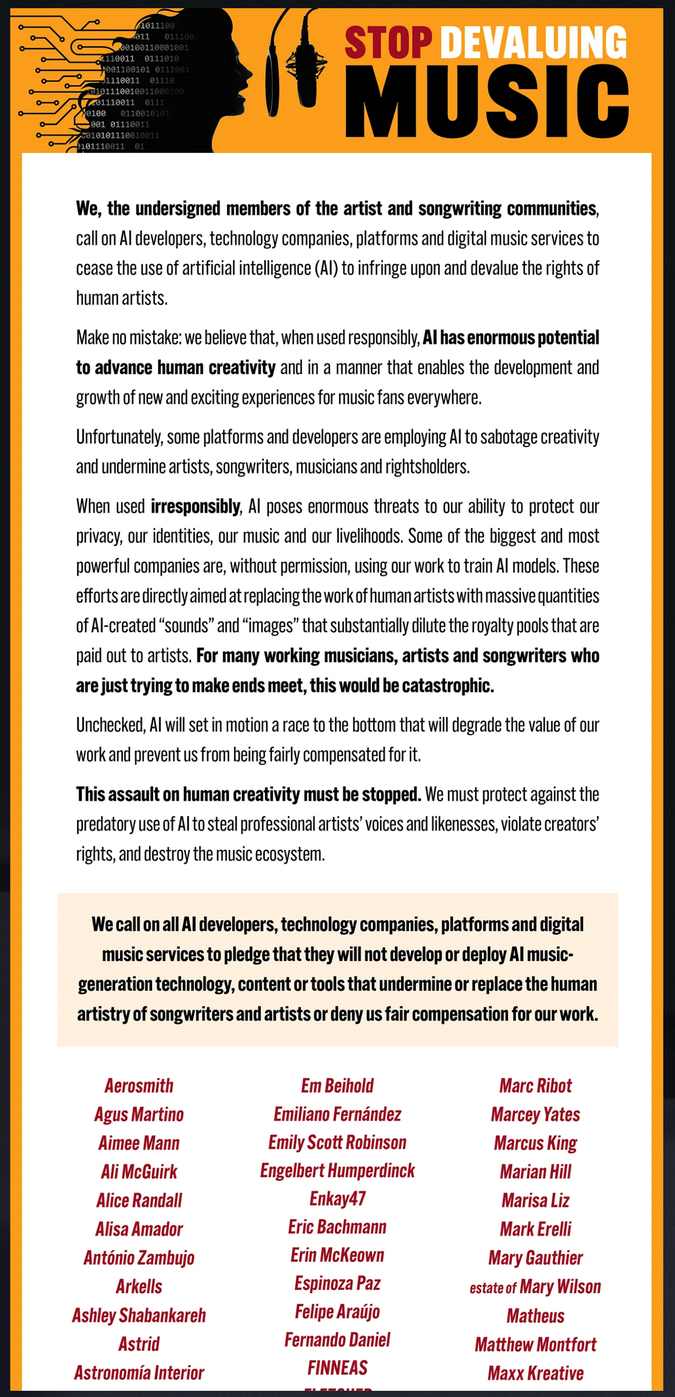
Where does the training data originate?

- Legal/ethical consequences

Open questions

How to **compensate** for production of training data?

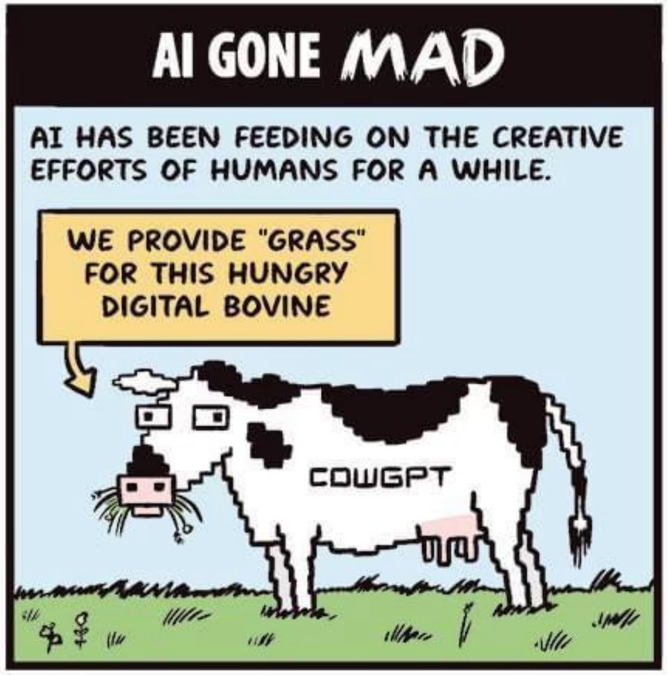
If generative systems are **trained on generated** data…?

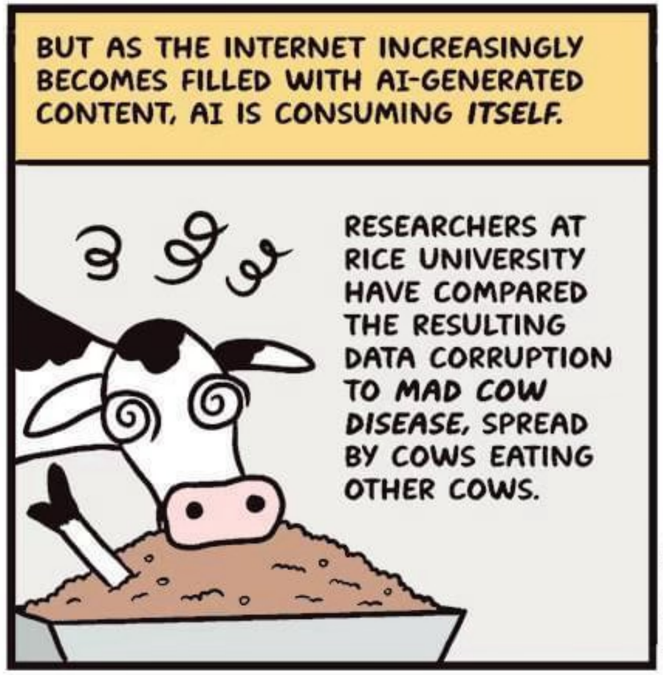


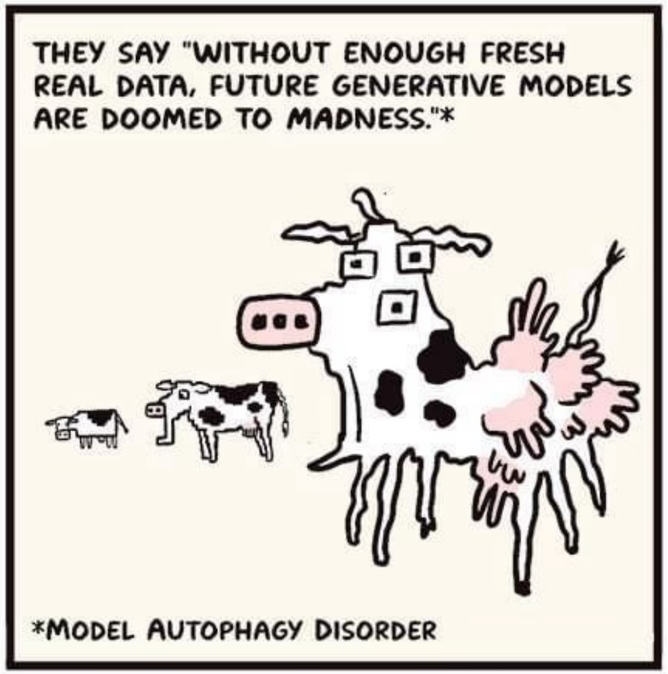
Open q

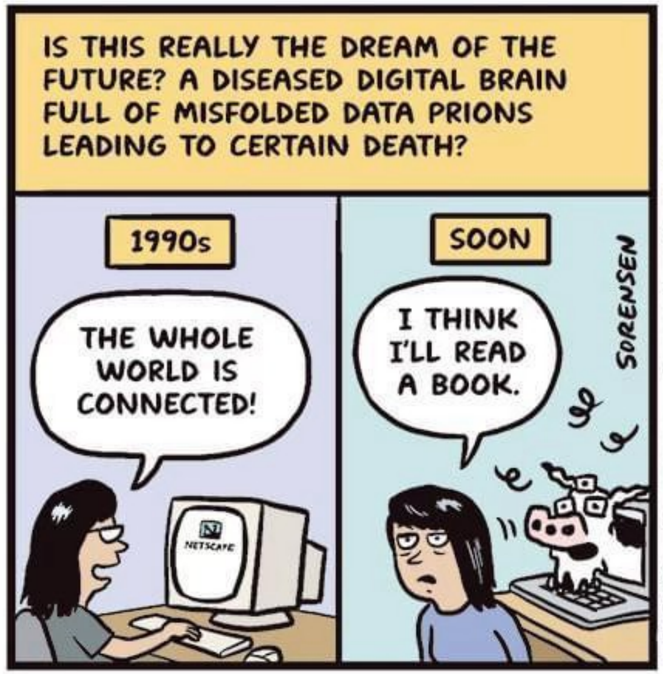
How to **compens**

If generative syste









© 2024 Jen Sorensen - Be a subscriber! jensorensen.com/subscribe

ned members of the artist and songwriting communities, s, technology companies, platforms and digital music services to tificial intelligence (AI) to infringe upon and devalue the rights of

e believe that, when used responsibly, **AI has enormous potential n creativity** and in a manner that enables the development and xciting experiences for music fans everywhere.

platforms and developers are employing AI to sabotage creativity ts, songwriters, musicians and rightsholders.

nsibly, AI poses enormous threats to our ability to protect our es, our music and our livelihoods. Some of the biggest and most are, without permission, using our work to train AI models. These imed at replacing the work of human artists with massive quantities ls" and "images" that substantially dilute the royalty pools that are **For many working musicians, artists and songwriters who nake ends meet, this would be catastrophic.**

et in motion a race to the bottom that will degrade the value of our from being fairly compensated for it.

man creativity must be stopped. We must protect against the o steal professional artists' voices and likenesses, violate creators' he music ecosystem.

developers, technology companies, platforms and digital s to pledge that they will not develop or deploy AI music- ology, content or tools that undermine or replace the human riters and artists or deny us fair compensation for our work.

Em Beihold
Emiliano Fernández
Emily Scott Robinson
Engelbert Humperdinck
Enkay47
Eric Bachmann
Erin McKeown
Espinoza Paz
Felipe Araújo
Fernando Daniel
FINNEAS

Marc Ribot
Marcey Yates
Marcus King
Marian Hill
Marisa Liz
Mark Erelli
Mary Gauthier
estate of Mary Wilson
Matheus
Matthew Montfort
Maxx Kreative

Who's using it?

Holly Herndon / PROTO

We had about six months of boring results before we started to get interesting results. The spoken part of "Birth," which is trained on my voice, was the first time we were like, "You can hear the logic of the neural network at work." AI is a combination of processing power and data sets. That's why the Chinese government, and companies like Google and Facebook, have the most sophisticated AI models.

Full interview: https://www.thefader.com/2019/05/21/holly-herndon-proto-ai-spawn-interview

W

Ho

Who’s using it?

The Beatles / Now and Then

- Originally written by John Lennon
- Never recorded
- Lennon’s voice was extracted from a low-quality demo tape and improved with ML technology

W

Th

Link collection

Basic Pitch - https://basicpitch.io

Suno - https://suno.ai, https://sunoaiwiki.com/

Stable Audio - https://stableaudio.com/, https://www.stableaudio.com/user-guide/

Stable Audio Open - https://huggingface.co/spaces/artificialguybr/Stable-Audio-Open-Zero

Riffusion - https://www.riffusion.com/

Spleeter Colab - https://colab.research.google.com/github/deezer/spleeter/blob/master/spleeter.ipynb

OpenAI Text-to-Speech - https://platform.openai.com/playground/tts

OpenAI Whisper - https://huggingface.co/spaces/openai/whisper

Amped Studio - https://app.ampedstudio.com/