

# KI im Musikbereich: Von einer emotionalen Debatte zu einer nüchternen Betrachtung

---

Christine Bauer

Austrian Composers Day 2025



**ORF TV-Programm**

**Künstliche Musik - Die KI-Revolution im Pop**

**mdr**

Startseite Sachsen Sachsen-Anhalt

**Künstliche Intelligenz**

03:34

**KÜNSTLICHE INTELLIGENZ**

**Wie Mensch und Maschine gemeinsam Kunst schaffen**

30. September 2024, 05:00 Uhr

**ZEIT ONLINE**

Aus der Serie **Natürlich intelligent**

**KI-Kunst**

**Ist KI-Musik Kunst? Egal!**

Unsere KI-Kolumnistin ist ergriffen von einem Song über den Ukraine-Krieg, generiert von einer KI. Und sieht darin neue Möglichkeiten – für uns alle.

## Künstliche Intelligenz: Übernehmen Fake-Künstler bald den Musikmarkt?

**DERSTANDARD**

MUSIKWELT

### Die neue Problemzone der Musikbranche heißt künstliche Intelligenz

Die Unterhaltungskonzerne sind in Aufruhr. Gerade wurde ein durch KI entstandener Fake-Song von Drake aus dem Netz genommen. Musikmanager Hannes Eder sieht das gelassen: Die Branche sei selbst schuld

**CHIP**

### Ohrwurm per künstlicher Intelligenz: Sind Ihre Lieblingslieder bald nur noch fake?

Wenn KI komponiert

### Geht künstliche Intelligenz den Popstars an den Kragen?

Sie legt Drake Worte in den Mund und lässt Oasis wieder singen: KI-generierte Musik schürt Ängste. Doch zum Hitparaden-Star taugt die Maschine nicht. Dafür hat sie andere Fähigkeiten.

• **hhvmag**

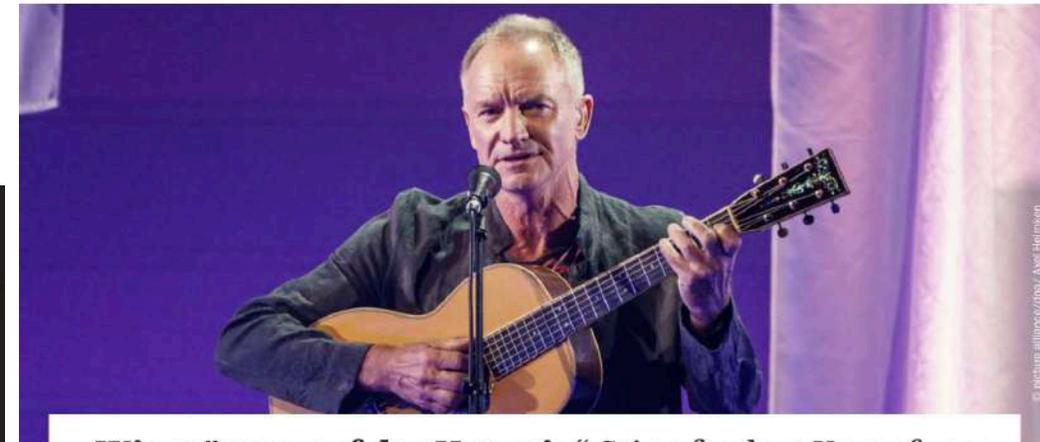
### Verändert künstliche Intelligenz die zeitgenössische Musik?

## Künstliche Intelligenz und Musik: Chance oder Risiko?

1996 starb die Soul-Sängerin Eva Cassidy an einer Krebserkrankung. Mit Hilfe von unveröffentlichten Studio-Aufnahmen wurde sie posthum zu einer Berühmtheit. In diesem März veröffentlichte ihr Label ein neues Album - arrangiert von einer künstlichen Intelligenz. Ist das ethisch zu vertreten? Bedeutet künstliche Intelligenz eine Gefahr für die menschliche Kreativität?

Journalist	Moritz Wulf
Land	Frankreich
	Deutschland
Jahr	2023

**TAGESSPIEGEL**



### „Wir müssen auf der Hut sein“ Sting fordert Kampf von Musikern gegen Künstliche Intelligenz

Auch in die Musik hat KI längst Einzug gehalten. Rockstar Sting mahnt seine Kollegen nun eindringlich zur Vorsicht: „Die Bausteine der Musik gehören uns, den Menschen.“

Von Sven Lemkemeyer  
18.05.2023, 19:12 Uhr

# Technologische Entwicklungen waren seit jeher eine treibende Kraft für den Musiksektor.

1935  
Tonband

1977  
Synclavier

1975  
Musik-  
video

1978  
Drum-  
Computer

1983  
MIDI

1999  
Napster

2001  
iTunes

2009  
Streaming

Einzelne Akteure waren auf unterschiedliche Art und Weise beflügelt oder betroffen.

2023  
Künstliche  
Intelligenz

# Typische Reaktion



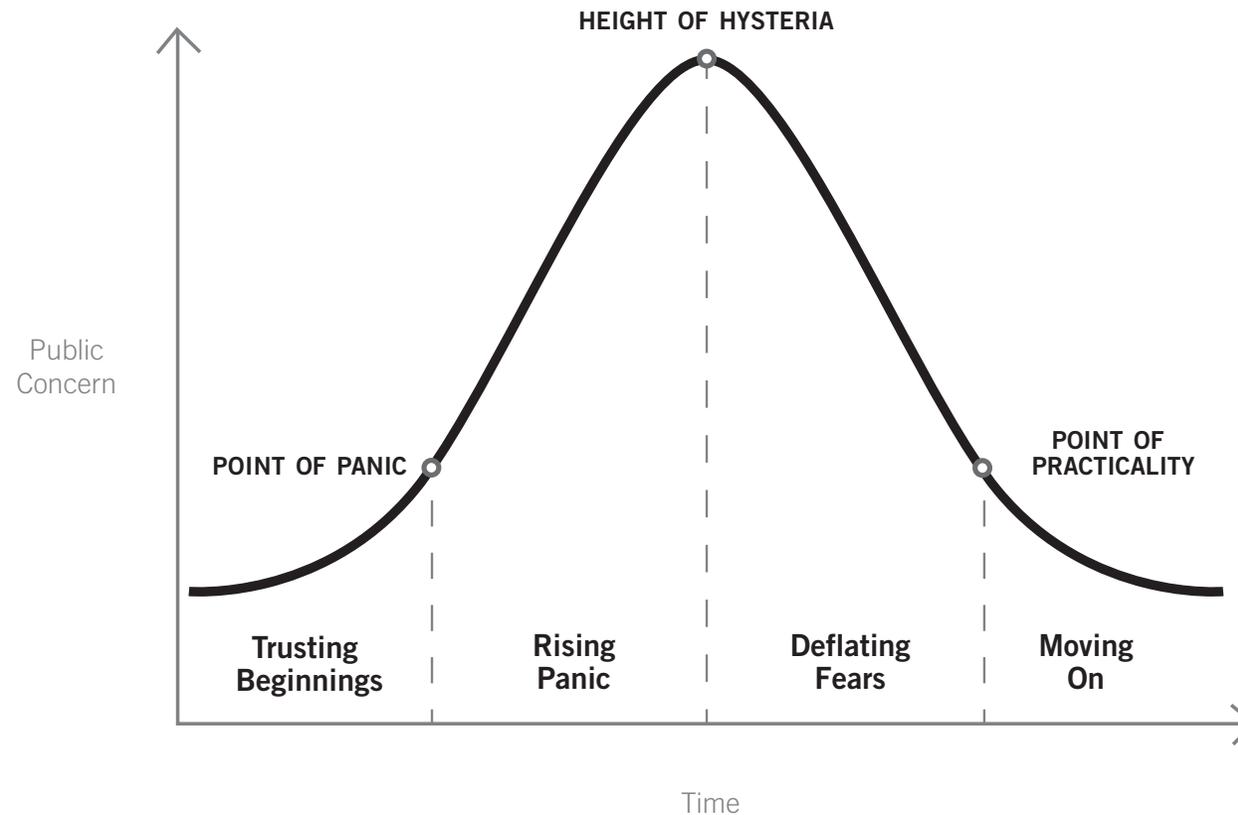
Image by master1305 on Freepik

Panik

Alles soll beim Alten bleiben.  
Das Neue muss weg!

# Technologiepanik—The same old story

Tech Panic Cycle

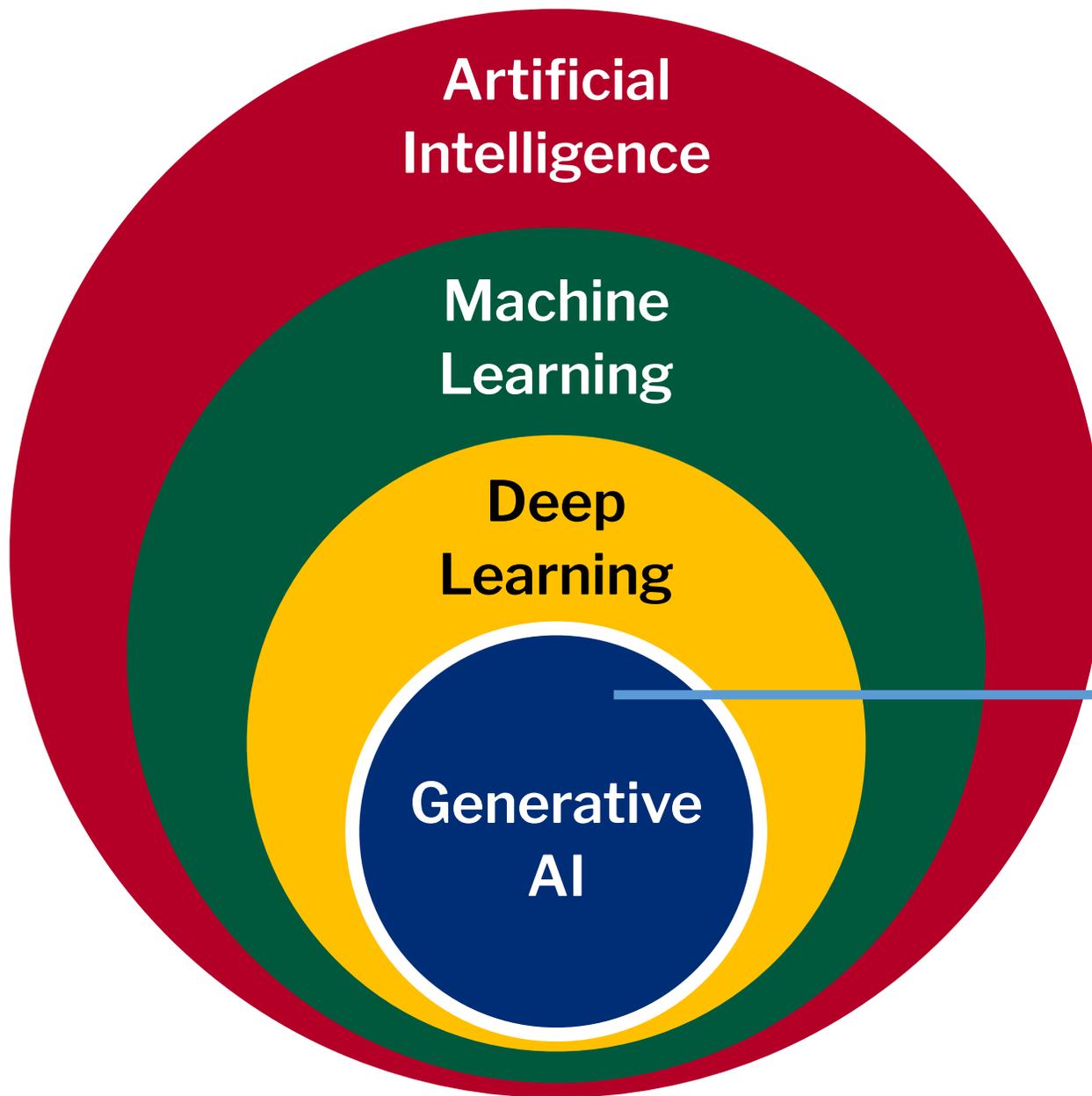


<https://itif.org/publications/2015/09/10/privacy-panic-cycle-guide-public-fears-about-new-technologies/>



**Es sind viele Einzelschritte notwendig!**

**KI kann und wird auf vielfältige Weise  
eingesetzt.**



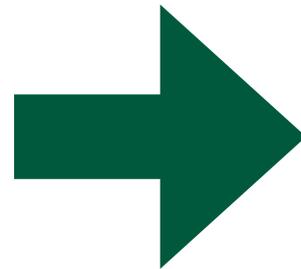
Erstellung von **Modellen**, die Inhalte generieren können, die bestehenden **Daten ähneln**.

**Modell** = vereinfachtes Abbild der Wirklichkeit;  
Nachbildung der wesentlich Eigenschaften; Weglassen der “unwesentlichen” Aspekte

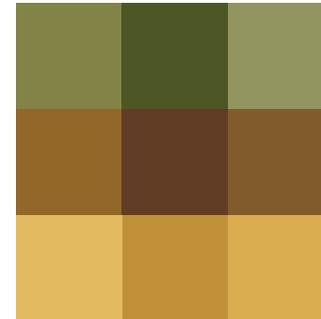
# Beginnen wir mit Mona Lisa



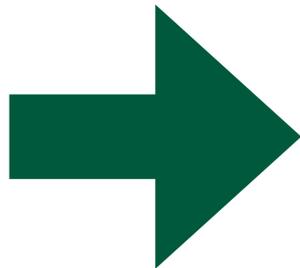
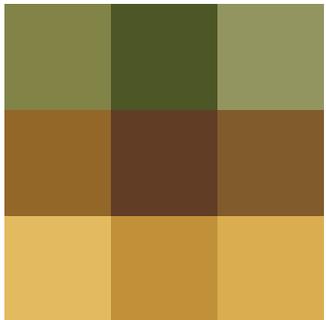
[https://en.wikipedia.org/wiki/Mona\\_Lisa#/media/File:Mona\\_Lisa,\\_by\\_Leonardo\\_da\\_Vinci,\\_from\\_C2RMF\\_retouched.jpg](https://en.wikipedia.org/wiki/Mona_Lisa#/media/File:Mona_Lisa,_by_Leonardo_da_Vinci,_from_C2RMF_retouched.jpg)



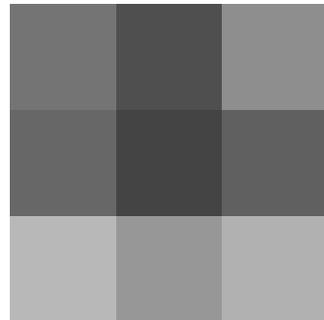
Mona Lisa  
auf Pixel-Ebene



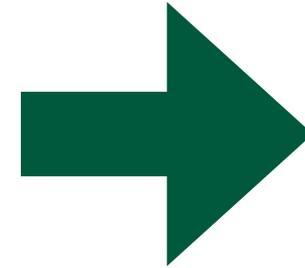
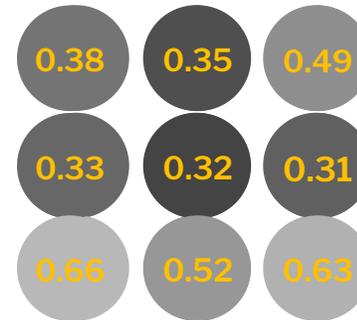
Mona Lisa  
auf Pixel-Ebene



Auf Pixel-Ebene in  
Graustufen



Jedem Mona-Lisa-  
Pixel wird der  
entsprechende  
Graustufenwert  
zugeordnet



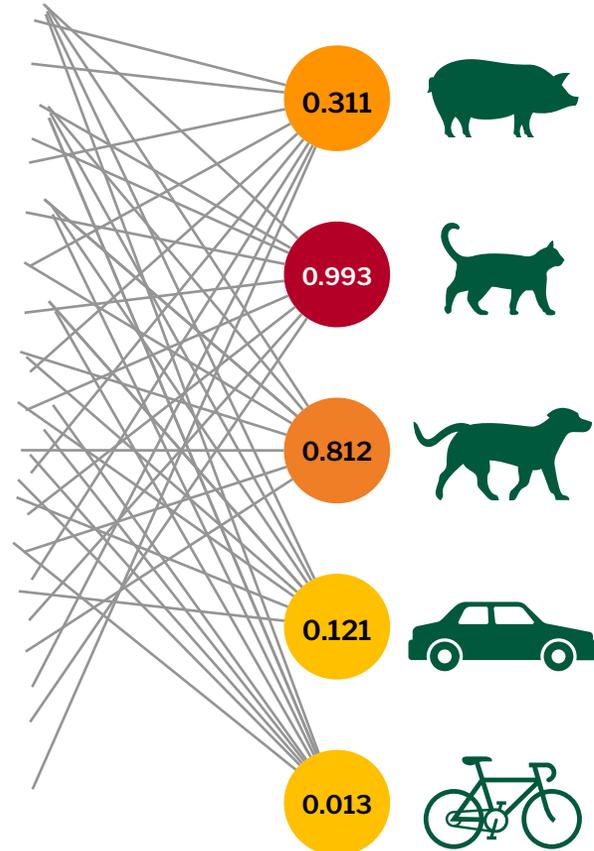


# Manche Repräsentationen sind einander ähnlicher (= näher) als andere

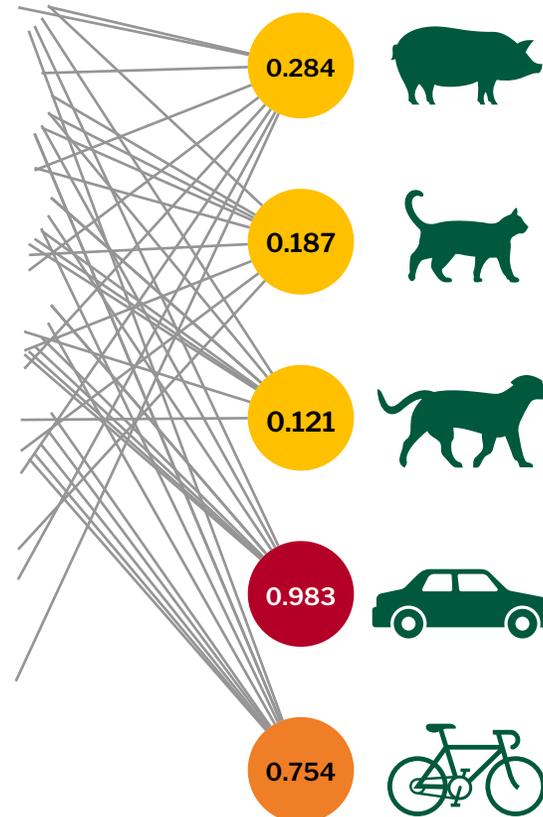


# Embedding-Vektoren

Embedding-Vektor für Katze 



Embedding-Vektor für Auto 

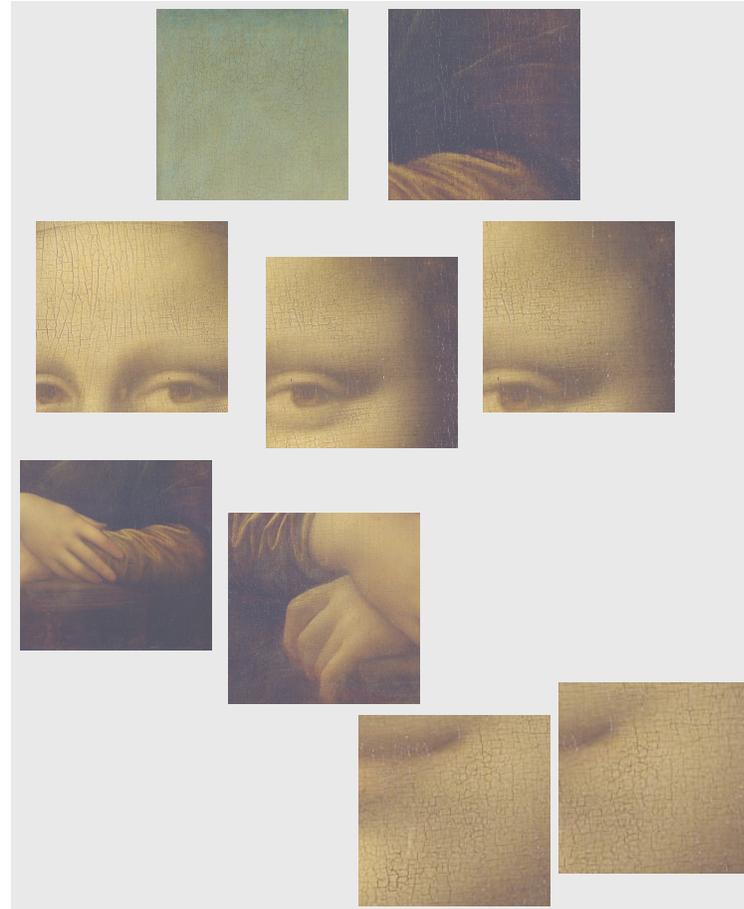


Quelldaten  
zB aus dem  
Web

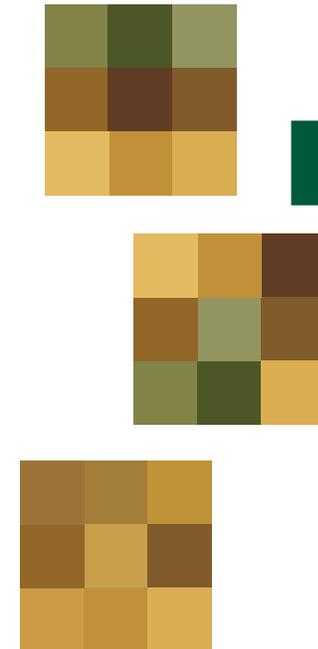


[https://en.wikipedia.org/wiki/Mona\\_Lisa#/media/File:Mona\\_Lisa,\\_by\\_Leonardo\\_da\\_Vinci,\\_from\\_C2RMF\\_retouched.jpg](https://en.wikipedia.org/wiki/Mona_Lisa#/media/File:Mona_Lisa,_by_Leonardo_da_Vinci,_from_C2RMF_retouched.jpg)

“Schnipsel” auf Pixelebene und  
einige größere Teile verteilt



Auf Pixel-Ebene:  
Für Mensch idR nicht  
interpretierbar



## Anreicherung mit Annotationen



## Analyse auf Patterns (zB gemeinsames Auftreten)



Bild: Pixabay/Gerd Altmann

## Modell



Bild: Pixabay/Gerd Altmann

zB Wahrscheinlichkeit, dass  
der nächste Pixel grün ist

# Das resultierende Modell enthält Wahrscheinlichkeiten, numerische Darstellung, Annotationen als Text

Das Modell enthält **keine** Quelldaten



Bild: Pixabay/Gerd Altmann

# Textgenerierende Systeme

<i>To tell cats from dogs,</i>
<i>To tell cats from dogs, you</i>
<i>To tell cats from dogs, you have</i>
<i>To tell cats from dogs, you have to</i>
<i>To tell cats from dogs, you have to know</i>
<i>To tell cats from dogs, you have to know what</i>
<i>To tell cats from dogs, you have to know what they</i>
<i>To tell cats from dogs, you have to know what they're</i>
<i>To tell cats from dogs, you have to know what they're looking</i>
<i>To tell cats from dogs, you have to know what they're looking for</i>
<i>To tell cats from dogs, you have to know what they're looking for.</i>

# Large Language Models (LLMs): Bei jedem Schritt wird ein Wort hinzugefügt

Text: “The best thing about AI is its ability to”  
Analyse aller Textdaten.

*The best thing about AI is its ability to*

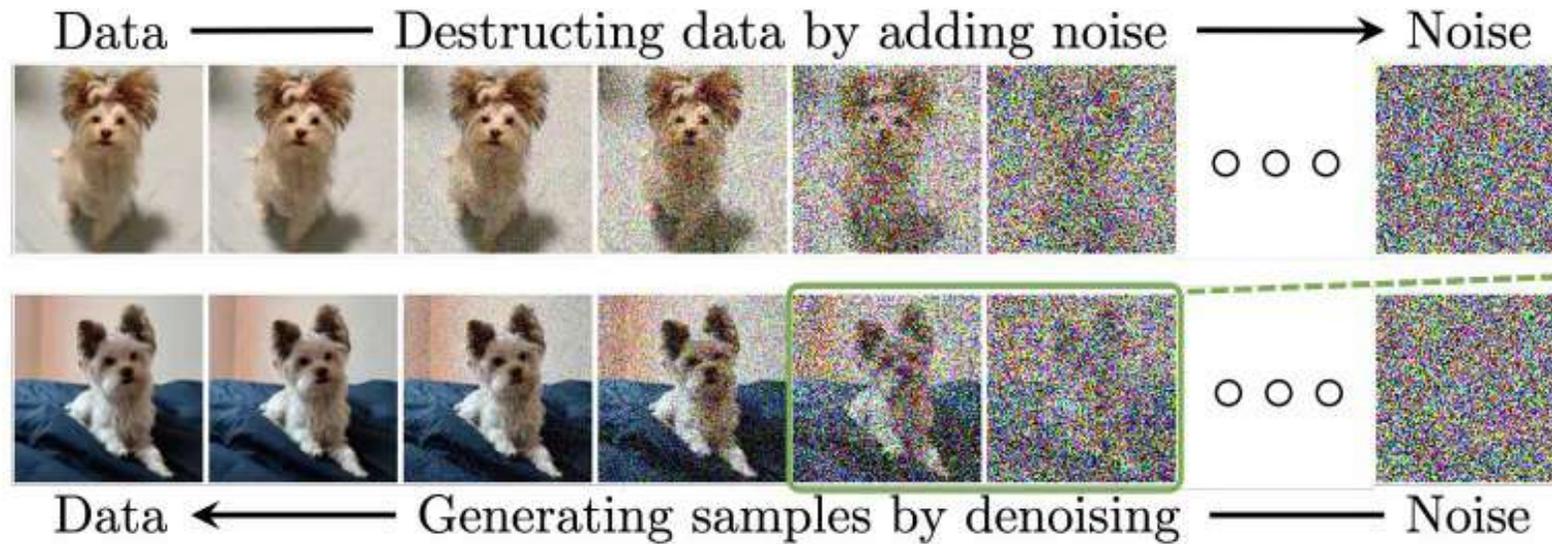
learn	4.5%
predict	3.5%
make	3.1%
understand	3.1%
do	2.9%

“learn” ist  
das nächste Wort mit  
der höchsten  
Wahrscheinlichkeit  
(Temperature=0)

Manchmal soll  
zufällig "Nicht-Top"-  
Wort gewählt werden  
(zB Temperature=0.8)

<https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>

# Diffusion models



Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang (2023). Diffusion Models: A Comprehensive Survey of Methods and Applications. ACM Computing Survey 56(4), Article 105 (April 2024). DOI: 10.1145/3626235

“Rauschen” (=Noise) wird schrittweise hinzugefügt.

Training Schrittweise Rekonstruktion aus dem Rauschen.

Modell = “Entrauschungsfunktion”

Generation: Anwenden des Modells auf anderen Input

# Diffusion models



Stefanie Meisl – “Schiele’s Ghost”

# KI in der Musikproduktion

- Effekte, Signal Processing: Rauschunterdrückung, Hall-Reduktion (De-Reverb)
- Transkription (Audio-to-Text oder Audio zu Musiknotation)
- Beats, Sounds
- Source Separation: Trennung einzelner Tonspuren aus Gesamtaufnahme
- Automatisches Mixing
- ...



# (Echtzeit-)Autotune



<https://www.antarestech.com/products/auto-tune/artist>



<https://www.antarestech.com/community/antares-announces-auto-tune-artist-next-generation-upgrade-auto-tune-live>

# KI zur Kreation

- Ideation: Inspiration, Ko-Kreation
- Generierte Musik: Melodien, Akkordprogressionen
- Generierte Lyrics
- Voice Cloning: Imitation oder Nachbildung einer bestimmten Stimme

**Stable Audio** 

 **ChatGPT**

# Beim Stichwort KI denken wir meist an Generierende KI—sehr oft an Deep Fakes



Guetta verwendete “geklonte” Eminem-Stimme in Live-Set

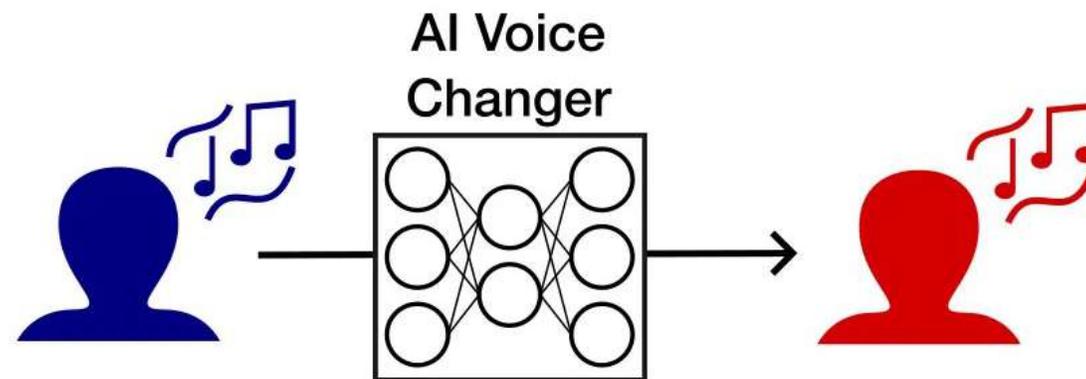
<https://www.youtube.com/shorts/98WTwSnkoas>

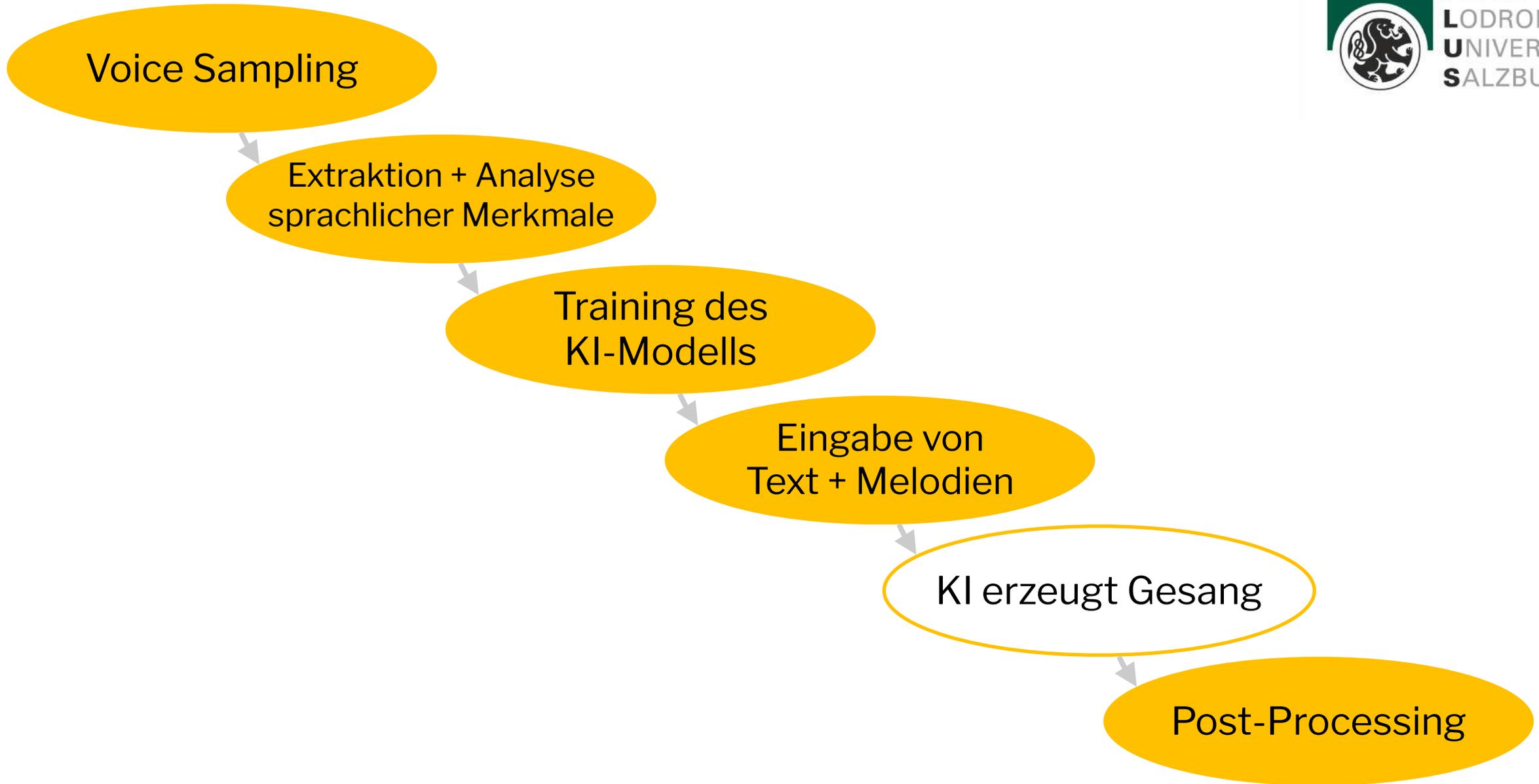
Spezifisch handelt es sich hier um  
**Music Deep Fake**

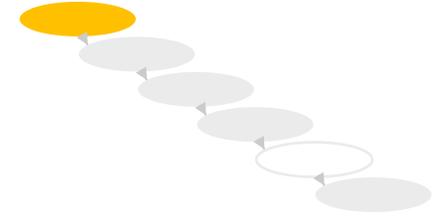
# Music Deep Fakes

Zentrales Element: **Voice Conversion** (Stimmkonvertierung)

Umwandlung einer Quellstimme in eine Zielstimme  
(bei gleichbleibendem sprachlichem Inhalt—unabhängig von Gesang/Musik)

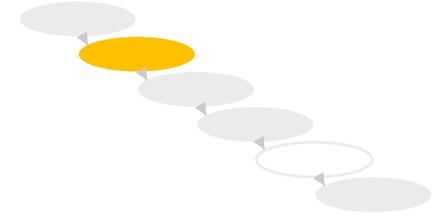






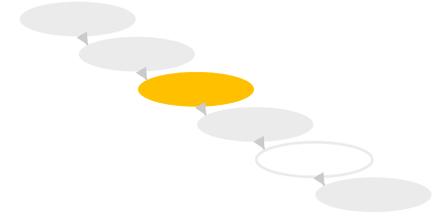
# Voice Sampling

- Audioaufnahmen einer *Zielperson* (Sprache, Gesang)
  - Je vielfältiger und klarer die Aufnahmen, desto besser
  - zB Interviews, Studioaufnahmen, Konzerte
- Ziel: die gesamte Bandbreite an Lauten, Betonungen und Sprechweisen erfassen



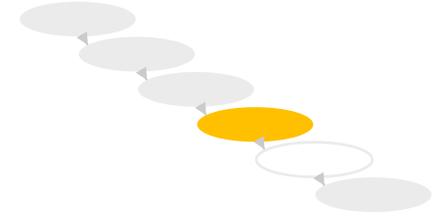
# Extraktion + Analyse sprachlicher Merkmale

- Aufnahme wird in Phoneme zerlegt (kleinste Lauteinheiten)
- Analyse von Klangfarbe, Intonation, Tempo etc
  
- Ziel: Identifikation besonderer Eigenschaften der Stimme (zB Akzent, Intonation, Rhythmus, Pausen)



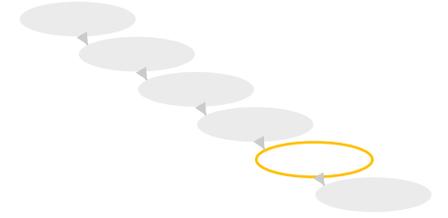
# Training des KI-Modells

- Neuronale Netze lernen die Stimme anhand der extrahierten Merkmale nachzubilden
  - zB Klangfarbe, Intonation, Rhythmus und Artikulation
- Ziel: Modell wird darauf trainiert, diese Merkmale in neuen Texten korrekt umzusetzen
- idR große Datenmengen für glaubwürdige Ergebnisse nötig



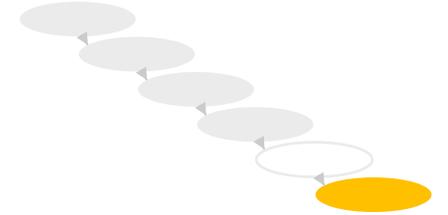
# Eingabe von Text + Melodien

- Text-to-Speech: KI spricht oder singt eingegebenen Text
- Singing Voice Synthesis (SVS): Text + Melodie → Gesang
  
- Kontrolle über Tempo, Tonhöhe, Ausdruck ggf möglich



# KI erzeugt Gesang

- Modell erzeugt neue Sprache/Gesang in der geklonten Stimme
  - dh Modell simuliert Stimmverhalten der gelernten Stimme
- Ausgabe klingt wie echte Stimme, wurde aber künstlich erzeugt



# Post-Processing

- Nachbearbeitung mit Effekten: EQ, Reverb, Auto-Tune etc
- Feinabstimmung verbessert Natürlichkeit, Betonung und Klangtreue
  
- Ziel: Natürlichkeit und professionelle Tonqualität

# KI-basierte Technologie für Marketing und Distribution in der Musikbranche

- Erstellung von Zusatzmaterialien
  - Pressetexte, Covers, Zusammenfassungen
- *Musikvideos*
- *Transkription*
- Distribution
  - Kundensegmentierung, Targeting
  - Empfehlungssysteme: zur richtigen Zeit am richtigen Ort an die richtige Person

# KI-basierte Technologie für musikalische Bildung

- Transkription
- Live-Tracking beim Üben + Analyse
- Empfehlung von Musiknoten oder Übungen (zB passend zu Spielniveau)



Ersetzen von Menschen durch KI



Unterstützung von Menschen durch KI

Replace **tasks** not humans.

# Gemeinsam nach einer Lösung suchen.

- Komplette Abwehr von neuer Technologie ist eine Einbahnstraße:
  - Nicht gegen die Technologie, sondern **mit** der Technologie.
- Alte Mittel funktionieren bei neuen Entwicklungen nur begrenzt:
  - Stichwort Urheberrecht bzw Lizenzsystem (die aktuelle Struktur, *nicht* die Existenz per se!)
- Weitblick ist gefragt!
  - Langfristige Implikationen für Marktakteure berücksichtigen

# KI im Musikbereich: Von einer emotionalen Debatte zu einer nüchternen Betrachtung

---

Christine Bauer

Austrian Composers Day 2025

