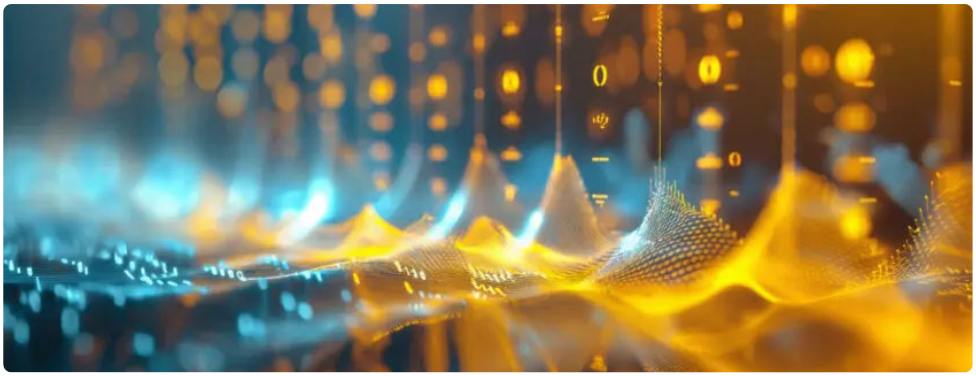
SocietyByte
BFH-Magazin für die Humane Digitale Transformation

Musik auf Abruf: Wie KI mit wenigen Stichworten ganze Songs komponiert

Von Yannis Schmutz (BFH Technik & Informatik) | 0 Kommentare



Die Fähigkeiten von generativer KI wortgewandte Texte und eindrucksvolle Bilder zu erzeugen ist längst allgemein bekannt. Doch neben diesen Errungenschaften zeichnet sich auch im Bereich der Musik eine bemerkenswerte Entwicklung ab. So kann generative KI nun ganze Songs komponieren; und zwar lediglich mit der Eingabe weniger Stichwörter.

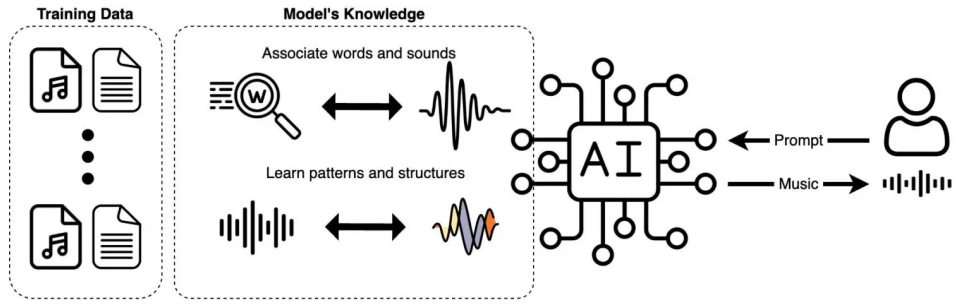
KI-Technologien zur Erstellung von Melodien oder Liedern sind als generative Musikmodelle bekannt und übersetzen einen vom Benutzer eingegebenen Text in eine einzigartige Musikkomposition. Die erzeugten Klänge spiegeln dabei die Stimmung, den Stil oder sogar bestimmte Details wider, die im Text beschrieben werden [1] [#_ftn1] . Dies ist ein grosser Fortschritt in der Art und Weise, wie KI die menschliche Kreativität unterstützen kann. Ob es um die Erstellung von Hintergrundmusik für Videos, die Komposition thematischer Klanglandschaften für Videospiele oder die Personalisierung von Liedern geht, die textabhängige Musikgenerierung gewinnt zunehmend an Bedeutung.

Schauen wir uns ein Beispiel an. Zur Beschreibung der Musik nutze ich die folgenden Stichwörter: *“Simple melodic house track”.* Die verwendete Text-zu-Musik-Software [2] [#_ftn2] spuckt daraufhin innert Sekunden ein dreiminütiges Lied aus. Hier hören Sie einen kurzen Ausschnitt des geschaffenen Werkes:

00:00 00:00

Wie funktioniert die textabhängige Musikerzeugung?

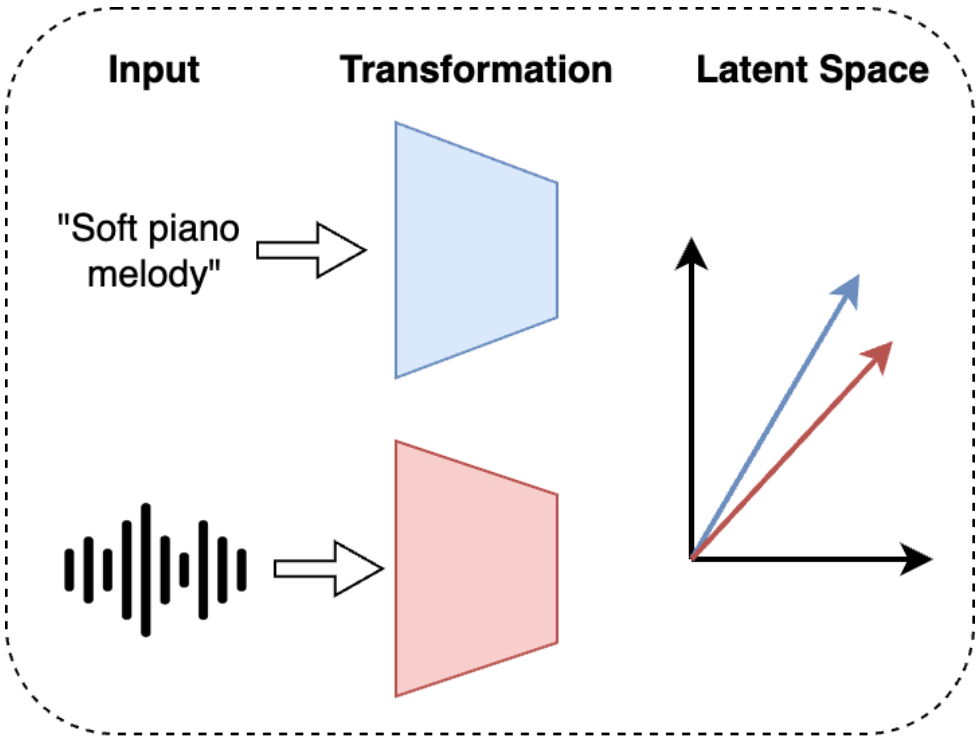
Das obige Beispiel zeigt, wie ein einsatzbereites Musikmodell im Grundsatz funktioniert. Der Benutzer formuliert stichwortartig den gewünschten Sound, worauf die KI die Beschreibung in passende Musik umsetzt. Doch um eine solche Verknüpfung zwischen Text und Audio herstellen zu können, muss sie erst lernen Schlüsselelemente wie Genre, Stimmung, Instrumente und Tempo im Text zu identifizieren. Dieser Lernprozess findet während dem “Training” des Modelles statt. Generative Musikmodelle werden anhand grosser Mengen von Musikdaten trainiert, die aus Paaren von Liedern und deren Schlüsselelemente in Textform bestehen. Während des Trainings lernt das Modell einerseits welche Wörter mit welchen Klängen assoziiert sind. Also beispielsweise, dass bestimmte Begriffe, wie «ruhig» oder «schnell», mit spezifischen Klangmustern, wie sanften Melodien oder schnellen Rhythmen, übereinstimmen. Andererseits erwirbt es die Fähigkeit deren Muster und Strukturen zu erkennen, um so kohärente Lieder konstruieren zu können.



Generative Music Principle

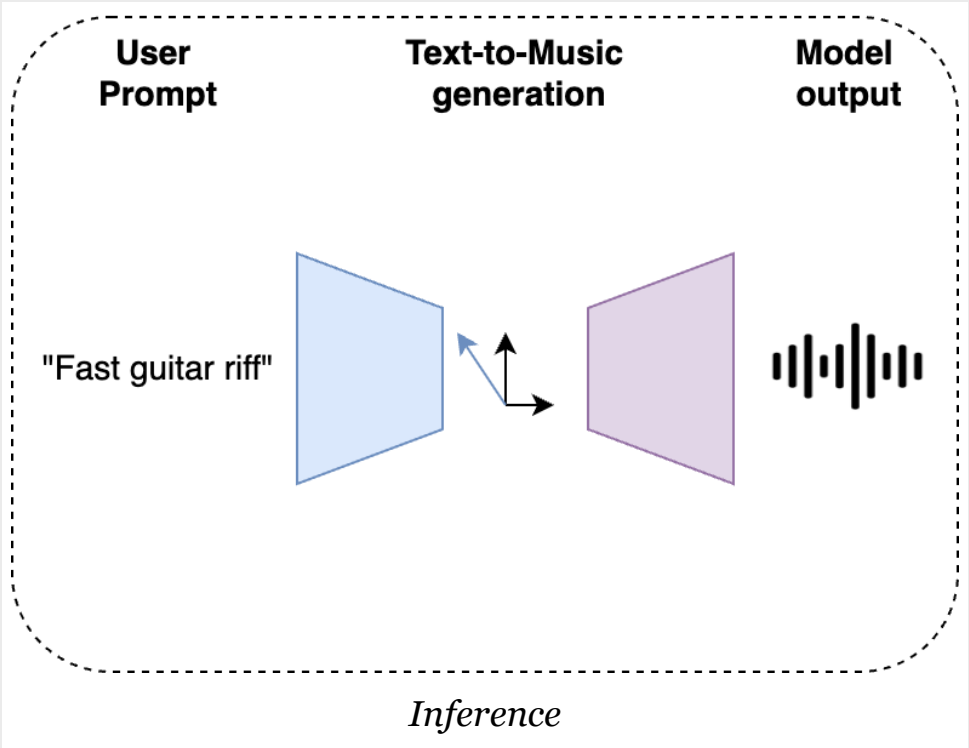
Doch wie kann eine Assoziation von Text und Klang überhaupt erfolgen; schliesslich handelt es sich dabei um verschiedene Modalitäten, die entsprechend unterschiedlich repräsentiert werden? Genau das ist der springende Punkt. Ein generatives Musikmodell bildet die eingegebenen Textbeschreibungen und Audiodaten so ab, dass diese in gleicher Weise interpretiert und somit verglichen werden können. Konkret werden beide Modalitäten in denselben latenten Raum – eine mathematische Darstellung – transformiert.

Wir können uns dies vereinfacht so vorstellen: Nach der Transformation werden beide Eingaben durch einen entsprechenden Pfeil dargestellt. Zeigen die beiden Pfeile in eine ähnliche Richtung, so ist der Text und der Klang ähnlich. Anhand der unzähligen Trainingspaaren lernt die KI, wie sie die Transformation justieren muss muss, um die Pfeile eines Paares ähnlich auszurichten. Dabei repräsentiert die Richtung eines Pfeils ebenfalls die Attribute die Schlüsselelemente des Textes und Attribute des Klangs.



Latent Space

Zudem ist das Modell in der Lage, einen Pfeil zurück in ein Audiosignal zu verwandeln. Wenn wir ein trainiertes generatives Musikmodell nutzen, geben wir lediglich die Textbeschreibung ein. Durch die Transformation entsteht ein Pfeil, der sowohl als Text wie auch als assoziierten Klang interpretiert werden kann und somit dessen Attribute repräsentiert. Durch die Rückwandlung des Text-Pfeils entsteht so ein Klang, der zu der eingegebenen Beschreibung passt.



Beliebte generative Musikmodelle

Meta's MusicGen [3] [#_ftn3]

00:00 00:00 [javascript:void(0);]

Riffusion [4] [#_ftn4]

00:00 00:00 [javascript:void(0);]

Suno [2] [#_ftn2]

00:00 00:00 [javascript:void(0);]

Die Beispiele wurden mit demselben Beschrieb generiert: *"An upbeat deep house song from the 1980s with lush jazz-funk chords and touches of soul music"*

Ethische und urheberrechtliche Überlegungen zur generativen Musik

Textbasierte Musikmodelle bieten zwar spannende Möglichkeiten, werfen aber auch wichtige ethische und rechtliche Fragen auf, insbesondere in Bezug auf das Urheberrecht [5] [#_ftn5] . Diese Modelle werden auf umfangreichen Datenbeständen bestehender Musik trainiert, von denen einige urheberrechtlich geschützt sein können [6] [#_ftn6] . Daher ist es denkbar, dass KI-generierte Musik versehentlich wesentliche Elemente geschützter Werke nachahmen könnte, was zu einer möglichen Verletzung des Urheberrechts führen würde [7] [#_ftn7] .

In ethischer Hinsicht hat die KI-gestützte Kreation sowohl transformative als auch disruptive Auswirkungen auf Musikschaffende [8] [#_ftn8] . Sie vermag Musikern zu helfen ihre Produktivität zu steigern und ihre Kreativität anzuregen. Im Umkehrschluss könnte dies die Beschäftigungsmöglichkeiten für Kreative verringern. Gleichzeitig sind generative Musikmodelle für ihr Training auf menschlich erzeugte Audiodaten angewiesen, um qualitativ hochwertige Ausgaben zu ermöglichen. Ohne Künstler gibt es also auch keine generative künstliche Intelligenz. Daher ist ein Gleichgewicht zwischen den Vorteilen der KI und einer fairen Vergütung und Anerkennung für menschliche Künstler bei der Weiterentwicklung dieser Technologie von entscheidender Bedeutung.

Referenzen

1 [#_ftnref1] Bengesi, Staphord, et al. «Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers.» *IEEE Access* (2024).

2 [#_ftnref2] Suno. Suno AI. https://suno.com/. Accessed Sept. 12, 2024.

3 [#_ftnref3] Copet, Jade, et al. «Simple and controllable music generation.» *Advances in Neural Information Processing Systems* 36 (2024).

4 [#_ftnref4] Forsgren, Seth, and Hayk Martiros. «Riffusion – Stable Diffusion for Real-Time Music Generation.» *Riffusion*, 2022, https://riffusion.com/about. [https://riffusion.com/about] Accessed Sept. 12, 2024.

5 [#_ftnref5] Deng, Junwei, Shiyuan Zhang, and Jiaqi Ma. «Computational Copyright: Towards A Royalty Model for Music Generative AI.»

6 [#_ftnref6] Peukert, Christian, and Margaritha Windisch. «The economics of copyright in the digital age.» *Journal of Economic Surveys* (2024).

7 [#_ftnref7] Henderson, Peter, et al. «Foundation models and fair use.» *Journal of Machine Learning Research* 24.400 (2023): 1-79.

8 [#_ftnref8] Lin, Tsen-Fang, and Liang-Bi Chen. «Harmony and algorithm: Exploring the advancements and impacts of AI-generated music.» *IEEE Potentials* (2024).

AUTHOR: YANNIS SCHMUTZ

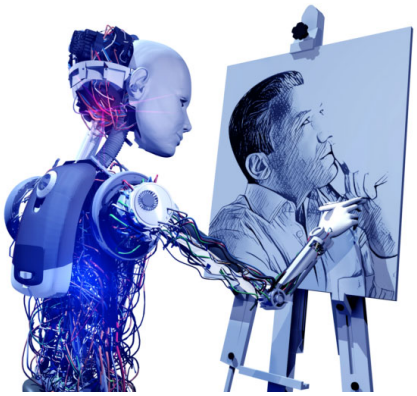
Yannis Schmutz ist wissenschaftlicher Mitarbeiter am Generative AI Lab der Berner Fachhochschule. Seine Forschungsschwerpunkte sind im Bereich der Audio- und Bildgenerierung sowie der Deep Learning basierte Wetterrekonstruktion.

Posts from Yannis Schmutz

Create PDF

Ähnliche Beiträge

Diffusion Models: Ein neuer Horizont in der Bilderzeugung



Generative AI – was ist das und was kann sie bereits?

0

COMMENTS