

Die Gegenüberstellung menschlicher Kreativität und generischer Musikproduktion

Esther Fee Reinhardt

Über Musikgeschmack

Menschen mögen Bekanntes, da Bekanntes oft mit Emotionen verbunden ist und diese wieder kurzzeitig hervorholt und zum Schwingen bringt. Außerdem sind wir Energieminimierer¹ und Bekanntes muss nicht mit so viel Energie analysiert und verarbeitet werden, wie Neues.

Ist ein akustisches Erlebnis so neu, dass wir kaum Anknüpfungspunkte zu etwas bereits Erfahrenem finden, verwerfen wir das Gehörte meist als »Das ist gar nicht mein Stil!« oder »Ist das überhaupt Musik?«.

Der Anteil an Neuem darf größer sein, je mehr unterschiedliche Musikstile wir kennen und je besser wir gelernt haben, dass sich die Anstrengung lohnt, das Gehörte zu abstrahieren und nach Anknüpfungspunkten im Gedächtnis zu suchen.

Zur leichten Musikunterhaltung sollte der Anteil von Neuem daher eher klein sein und um diesen Zweck zu erfüllen, könnte der Algorithmus hinter generativer künstlicher Intelligenz (KI) nicht geschickter erdacht worden sein. Er zieht sein gesamtes kreatives Potenzial aus den zur Verfügung gestellten Trainingsdaten, also ausschließlich aus wiederkehrenden Mustern mit der höchsten Wahrscheinlichkeit. Was so profan klingt, ist der menschlichen Kreativität jedoch sehr ähnlich.

Kreativität und das menschliche Selbstverständnis

Warum lesen wir seit Beginn 2023 ständig Beiträge über Kreativität, obwohl eine Technik im Bereich des Machine Learnings, die mit dem Perceptron in ihren Grundzügen in den

¹ Die Autorin besteht entgegen den Manuskriptvorgaben auf der Verwendung des generischen Maskulinums. Sie fühlt sich selbst der Gruppe der Wissenschaftler und Künstler zugehörig.

1950er Jahren beschrieben wurde (Rosenblatt 1958), jetzt nur schneller geht und mehr Trainingsdaten zur Verfügung hat?²

Als die erste bekannte algorithmische Komposition – die Illiac Suite – durch ihre menschlichen Geburtshelfer Hiller und Issacson 1959 das Licht der Welt erblickte, war der Aufschrei nicht vergleichbar (Hiller/Issacson 1959), auch nicht, als es in den 1990ern möglich war, mit der Software Band-in-a-Box (Gannon 1991) neue Solos über durch den Benutzer erstellte Akkordfolgen in unterschiedlichen Stilen generieren zu lassen. In diesen Fällen wurde menschliches Expertenwissen, das durch Erfahrung und Analysen gewonnen wurde, in Regeln festgehalten und durch einen Computer zur Anwendung gebracht. Dieses regelbasierte maschinelle Lernen (auch »Symbolic Artificial Intelligence (AI)« oder »Good old-fashioned AI«) baute auf dem Expertenwissen auf, und die abgeleiteten Ergebnisse waren nachvollziehbar.

Im Gegensatz dazu sind die Fähigkeiten von generativer KI unvorhersehbar und wir können uns nicht wie in den obigen Beispielen anzeigen lassen, welche Regeln und Schlüsse ein neuronales Netz aus seinen Trainingsdaten gezogen hat. Was wir sehen, sind scheinbar kreative Ergebnisse, von denen wir – so muss man es sagen – doch überrascht wurden.

So wurde uns deutlich vor Augen geführt, dass was wir als Kreativität empfinden, extrem viel mit digitalisierbarer Datenverarbeitung zu tun hat, zu der es den aktiven Menschen fast gar nicht benötigt – dieser Umstand bringt das menschliche Selbstverständnis durcheinander.

Die Menschheit ging bereits mindestens zweimal notgedrungen durch eine Spezifizierungsphase. Als Dampfmaschinen, die keinen Schlaf, Essen oder Lohn benötigen, die Arbeit von Arbeitern übernahmen, die ihren Wert in der Gesellschaft bis dato oft über ihre Leistungsfähigkeit definierten, blieben für die Menschen die Qualitäten des Wissens und der Logik oder – grob zusammengefasst – die Intelligenz. Als Maschinen plötzlich Berechnungen wesentlich schneller und präziser als wir ausführen konnten, musste sich die Spezies Mensch wieder auf eine andere Tugend, zu der nur Menschen in der Lage wären, berufen.

Und hier stehen wir und lesen wir Texte, schauen auf Bilder oder hören Musik, die bis vor kurzem nur durch die Anwendung menschlicher Kreativität erschaffen werden konnten, jetzt aber scheinbar aus dem Nichts und mit minimalem Aufwand vor unseren Augen entstehen. Die Definition von Intelligenz hat durch die maschinelle Konkurrenz profitiert und mit neuen Facetten, wie der emotionalen Intelligenz, an Profil gewonnen. Dort gibt es Aspekte, die besser, gut oder gar nicht von Rechenprozessen erfüllt werden können. Und bei der Kreativität?

2 Es soll nicht unerwähnt bleiben, dass die Lösung des Vanishing Gradient Problems (Hochreiter 1991) und andere algorithmische Verbesserungen auch ihren Teil zu dem Fähigkeitsprung von KI beigetragen haben.

Kreativität als Methode

Kreativität beginnt mit der Wahrnehmung, gefolgt von einem Filter, der entscheidet, welche Eindrücke es in unsere Aufmerksamkeitssphäre schaffen und was davon wie bewusst abgespeichert wird. Ein Beispiel: Wir würden nicht nur die Szene eines Baumes auf einem Hügel in der Sonne abspeichern, sondern auch die einzelnen Details wie die Blattform, die Farben, den Geruch, die Wärme oder die Rauigkeit der Rinde. Wobei wir womöglich den Farben unseren Fokus schenken und den Geruch nur bei einem besonderem Auslöser wieder aktiv im Gedächtnis finden können. Wenn wir kreativ tätig werden, schöpfen wir aus diesem Vorrat und kombinieren Abgespeichertes in unterschiedlicher Granularität. Daher sind Aufmerksamkeit und Kreativität auch stark miteinander verknüpft.

So entstehen bei Kompositionen mit sehr grober Granularität von bereits Gehörtem neue Musikstücke, die anderen Stücken in Sound, ganzen Akkordfolgen, Rhythmen oder anderen musikalischen Dimensionen sehr ähnlich sind. Wir können von der Verwendung großer Eindrucks-Moleküle sprechen. Während Kompositionen, die uns vollkommen neuartig und noch nie dagewesen erscheinen, aus ganz fein zerlegten Eindrücken – sozusagen fast aus Eindrucks-Atomen – zusammengesetzt wurden. Ein starkes Beispiel hierfür ist die Verwendung einzelner Sinustöne anstelle von Naturtönen³ oder ganzer Frequenzbereiche wie bei Stockhausen (Stockhausen 1956).

Die Auswahl der Moleküle unserer Eindrücke und das Kombinieren derer entspricht ziemlich genau dem, was in einem variational autoencoder⁴ (Kingma&Welling 2013) geschieht. Bei der KI kommt in den Entscheidungsprozessen bei der Auswahl und der Kombination die Wahrscheinlichkeit für ein Element in Verbindung mit einem *tag* oder *keyword* aus den gelernten Daten als Kriterium zum Einsatz. Dieser Teil von Kreativität ist – quod erat demonstrandum – technisch machbar.

Also sind Maschinen doch kreativ? Ja und Nein.

Erweitern wir die Kreativitätsdefinition um die Fähigkeit, die Auswahl der Moleküle und die Kombination so zu gestalten, dass sich nicht das Wahrscheinlichste ergibt, sondern etwas, das ein besonders treffendes Abbild einer persönlichen Emotion oder Vorstellung ist. Diese persönliche zugrundeliegende Vorstellung, mit der Ergebnisse immer wieder abgeglichen werden, ist auch der Grund, warum Maschinen in einem Aspekt den Menschen weit hinter sich lassen. Algorithmen werden nicht durch Angst gehemmt und erstellen in wenigen Sekunden nicht nur ein, sondern oft eine Vielzahl an Ergebnissen.

Keine Emotionen haben zu können, bedeutet auch, keine Angst davor zu haben, dass zukünftige Rezipienten das Ergebnis kritisieren, das so oft mit der inneren Vorstellung abgeglichen wurde, die wiederum stark mit der eigenen Identität verbunden ist. Menschliche Kreativität wird aber nicht nur durch die Angst der Urteile anderer

3 Naturtöne sind aus einzelnen Sinusschwingungen zusammengesetzt.

4 Ein variational autoencoder (VAE) ist eine Netzwerkarchitektur, auf der viele generativen KI Systeme aufbauen.

gehemmt, sondern auch durch eigene Selbstzweifel. Besonders die Frage, ob man dem inneren Bild überhaupt gerecht wird, kann den Kreativprozess immer wieder zum Stocken bringen. Das angstfreie Schaffen von Algorithmen ist Menschen eigentlich nur im Flow, im höchsten Modus der Kreativität, möglich.

So ergibt sich folgende vollständigere Definition: Kreativität ist die Zerlegung von Wahrgenommenem, Auswahl und Neukombination⁵ mit der Fähigkeit, einen Abgleich eines (Zwischen-)Ergebnisses mit einem inneren Thema – einer Emotion oder Vorstellung – vorzunehmen. Die Gegenüberstellung mit dem inneren Thema erfordert wiederum die Befähigung, Gefühle empfinden zu können und ein Bewusstsein.

Aber, da wir doch schon über die Natur des allgemeinen Musikgeschmacks gelesen haben, stellt sich die Frage: Ist diese Tiefe von Kreativität mit der dazu benötigten Emotionalität tatsächlich notwendig, um Musik zu erschaffen? Oder ist das nur ein verzweifelter Wunsch der Menschen, sich doch noch als Krone der Schöpfung zu behaupten?

Musik und Emotionen

Um zu diskutieren, wie viel Emotion in Musik eingebracht werden muss, führen wir den relativen »minimal emotion needed (MEN-) value« ein. Dieser geht über Musik hinaus und gibt allgemein an, wie viel Emotion mindestens notwendig ist, um eine bestimmte Funktion zu erfüllen. Diese Funktionen rangieren von Fahrstuhlmusik zu improvisierten Live-Konzerten. Betrachten wir ein paar Randfälle:

Die Fahrstuhlmusik dient dazu, Stille zu vermeiden, sodass Fahrstuhlfahrer sich nicht unter Druck gesetzt fühlen, ein Gespräch beginnen zu müssen und um die Wahrscheinlichkeit für eventuell aufkommende Klaustrophobie zu verringern. So kann die Fahrstuhlmusik etwa ein großer Jazz Standard mit hochkarätigen Interpreten sein, muss es aber nicht. Zu diesem Zweck ist ein generiertes Stück, in dem melodische Solos über typischen Jazzakkordfolgen automatisiert und beliebig erstellt werden, mehr als ausreichend. Je typischer, desto besser, da es dem Hörer, der keine akustischen Überraschungen erlebt, dann keine Aufmerksamkeit und somit keine Energie abverlangt.

Einen höheren MEN-value hat Musik für das folgende Szenario: ein Jingle für einen beliebigen (in unserem Beispiel nicht künstlerischen) Podcast. Funktion: ein kurzes Motiv, das zur Wiedererkennung des Podcasts dient. So könnte man sich ohne jegliche musikalische Kenntnisse 50 mögliche Jingles erzeugen lassen und bräuchte nur den eigenen Geschmack, um den emotionalen Abgleich zu machen.

Weit auf der anderen Seite der Skala befinden sich musikalische Erlebnisse, die die Funktion der tiefen empathischen Verbundenheit zwischen Menschen innehaben. Wer etwa einmal Joe Cocker auf der Bühne gesehen hat, weiß, dass alleine das Betrachten eines Künstlers, der emotional tief in sein Werk abtaucht und dies durch starken Ausdruck in Klang und Körpersprache transportieren kann, auf den Betrachter wie ein empathischer Katalysator wirkt. (vgl. Iomey 2007)

Also lässt sich sagen: Es ist technisch möglich, dass jede Musik, die eine Funktion erfüllen soll, die keinen menschlichen Künstler mit Leidenschaft, Empathie und Intention

5 Bis hierhin technisch realisierbar

benötigt, von einem KI-System ersetzt werden kann. Und das trifft auf viele Szenarien der heutigen Musikproduktion zu, bei denen der Hauptfokus der Hörer nicht auf der Schöpfungstiefe eines Musikstücks liegt: Musik für Werbung, Akkordbegleitung zum Üben, Hintergrundmusik im Restaurant, zum Tanzen oder Feiern animierende Musik und viele andere.

Wie bereits erwähnt, ist in vielen Fällen der Zweck sogar besser erfüllt, je weniger Neuartiges im Musikstück zu finden ist. Was übrig bleibt, ist Kunst.

Die neuen Rollen von KI

Mit den technischen Fähigkeiten von generativer KI ist es möglich, Musik zu erzeugen, die sich für viele Aufgabenbereiche ausreichend eignet und die sich nicht mehr von rein menschlichen Werken unterscheiden lässt. Je mehr Bedeutung ein Musikstück durch seine Funktion aber haben soll, desto wichtiger wird die Fähigkeit der menschlichen Emotion und Empathie im Hintergrund.

Was heißt das nun für den heutigen Musiker? Sich gänzlich von der Musik abwenden, weil KI bald alles übernimmt? Oder sollte man sich als »ernsthafter Künstler« von KI abgrenzen, um weiterhin authentisch zu bleiben? Um sich dieser Frage anzunähern, müssen wir mindestens drei unterschiedliche Rollen betrachten, die KI im kreativen Schaffen einnehmen kann:

- *KI als one-klick-wonder* mit einem beliebigen oder lapidaren menschlichen Input, so etwa bei einem 3-Sekunden-Jingle für den Podcast über eine Südamerika Reise.
- *KI als Inspiration und Werkzeug*: Die Konfrontation mit einem potentiellen Endergebnis führt zu einer inneren Stellungnahme des Künstlers, der einen Abgleich mit seiner inhärenten Vorstellung machen muss. Diese Vorstellung wird aber fortwährend weiterentwickelt durch die Wahrnehmung der generierten Ausgabe. Diese basiert selbst auf den Werken aller Künstler, die ins Training eingeflossen sind, und dem Prompt des aktiven Künstlers. So kann man von Inspiration und Rückinspiration sprechen. Ein besonderer Vorteil von der Mensch-KI-»Co-Creativity« ist, dass durch die Interaktion mit KI anstelle eines Menschen, Kreativen Ängste genommen werden. Und da echte Kreativität nicht in einer Umgebung der »vorschnellen Bewertung« entstehen kann, kann sogar der Eindruck eines Kreativitätsbeschleunigers entstehen, obwohl es einfach keinen Hinderer gibt. Ein Nachteil jedoch ist, dass kreative Synergie, wie bei der Mensch-Mensch-Co-Kreativität, nur durch eine gemeinsame Begeisterung für eine geteilte Vision möglich wird. Diese emotionale Synchronizität entfällt bei einer Mensch-Maschinen-Co-Kreativität.
- *KI zum Modalitätstransfer*: Ein fertig trainiertes künstliches neuronales Netz ist eine Black-Box für den menschlichen Verstand und nicht regelbasiert formulierbar – auch wenn es großartige Ergebnisse liefert. Dieser Nachteil hat aber auch ein gewaltiges Potenzial. So ist es für eine KI möglich, schwer oder nicht formulierbare Zusammenhänge zu lernen, die in einem Datensatz vorhanden sind, beispielsweise auch über die Grenzen einzelner Modalitäten hinaus. Hier ein paar Beispiele: Bild

zu Klang, Gestik und Gesichtsausdruck zu Intensität (beispielsweise bei der Analyse eines Dirigenten), Gefühl in Wort zu Gefühl in Gesichtsausdruck.

Diese dritte Rolle ist die noch nicht so häufig genutzte *Superfähigkeit* von KI, die auch in der menschlichen Kreativität neue Türen aufstoßen kann.

Kann KI Kunst?

Wenn KI, wie im letzten Beispiel genannt, auch in der Lage ist, Emotionen zu suggerieren, gibt es dann noch Grenzen? Kann KI Kunst?

In der Rolle der Inspirationsunterstützerin und als Übersetzerin von einer Ausdrucksform in eine andere wird KI als Werkzeug verwendet, um eine menschliche Vision umzusetzen. Interessant ist es, wie ein Werk zu beurteilen ist, das ohne künstlerische Intention entstanden ist.

Nach Tasos Zembylas (1997) obliegt die Definition von Kunst einem stetigen Wandel, der gerade durch solch disruptive Technologiesprünge, wie dem durch generative KI, befeuert wird. So kommen wir nicht umhin, uns damit auseinander zu setzen, inwiefern die Persönlichkeit eines Künstlers für den Kunstbegriff doch erforderlich ist. Menschen sind in allererster Instanz soziale Wesen. Daher ist uns die Verbundenheit zu anderen Menschen und der gemeinsame Austausch ein sehr hohes Ziel. Dies ist die Wurzel unserer Empathie und unseres Interesses, herauszufinden, was unser Gegenüber innerlich bewegt.

Das ist auch der Grund, warum wir an Kunst, die durch einen Menschen geschaffen wurde, ein anderes Interesse hegen als an einem Zufallsprodukt, selbst wenn wir dieses akustisch oder optisch nicht mehr von einem von Menschen gemachten Werk unterscheiden könnten.

Wie zu Beginn erwähnt, ist der Mensch auch ein Energieminimierer, und weil der Versuch, sich vollen Herzens auf etwas einzulassen und Verständnis für ein Kunstwerk zu erlangen, emotionale Öffnung und Energie verlangt, müssen wir davon überzeugt sein, dass sich eine Intention und damit eine Emotion hinter der Fassade verbirgt, die wir bei erfolgreicher Synchronizität erleben können.

Denn worin besteht der Grund, sich etwas hinzugeben, das ohne Hingabe erschaffen wurde?

Literatur

- Gannon, P. (1991). *Band-in-a-Box*. PG Music. Inc., Hamilton, Ontario, 1999.
- Hiller, L. A., & Isaacson, L. M. (1979). *Experimental Music; Composition with an electronic computer*. Greenwood Publishing Group Inc.
- Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen. *Diploma, Technische Universität München*, 91(1), 31.
- Kingma, D. P. & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

- Lomey (2007): Joe Cocker – You Are so Beautiful (Nearly Unplugged) [Youtube], <https://www.youtube.com/watch?v=wLDmslyGmGI>
- Rosenblatt, Frank (1958): The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, in: *Psychological Review*, vol. 65, no. 6, S. 386
- Stockhausen, Karlheinz (1956): *Studie II.*, Michigan: Universal Edition
- Vaswani, Ashish, et al. (2017): Attention Is All You Need, in: *Advances in Neural Information Processing Systems*, vol. 30
- Zembylas, Tasos (1997): *Kunst oder Nichtkunst: Über die Bedingungen und Instanzen ästhetischer Beurteilung*, Wien: WUV-Universitätsverlag

