

Nachstehender Text ist eine
Leseprobe aus AV-DIALOG 4/2024

Der **AV-DIALOG** erscheint viermal im Jahr und ist eine reine Mitgliederzeitschrift.

Regelmäßige Rubriken sind u. a.

- Gestaltung von AV-Produktionen
- Technik
- Berichte (von Veranstaltungen)

Weitere Leseproben finden Sie auf www.av-dialog-magazin.de

Über neue Leser (und Mitglieder) freuen sich der Verein und die Redaktion.

Kontakt über: heftredaktion@av-dialog.de

Eine unterhaltsame Lektüre wünscht

Klaus Fritzsche
(Chefredakteur)

Musik mit KI generieren

Komponieren per Prompt

Von Norbert Ott

Mit KI-Werkzeugen lassen sich nicht nur Texte und Bilder generieren. Auch für die Komposition von Musik gibt es spezialisierte Tools. Ein Überblick über den Stand der Technik.

Wer musiziert denn da?

Dieser Punkt ist wohl jedem AV-Autor bekannt: Die AV-Grundidee ist solide ausgearbeitet und der Teil „Vision“ mit den ersten Bildsequenzen zusammengestellt. Jetzt kommt der „Audio“-Teil mit Musik und ggf. Geräuschen. Spätestens jetzt wird es, zumindest bei mir, oft etwas zähflüssig. Verlockend ist der Gedanke, einfach ein paar eingängige Melodien unterzulegen – fertig! Am Ende sieht das AV-Werk zwar ganz nett aus, aber der berühmte „Funke“ springt sicher nicht über. Das Werk ist nicht rund, weil etwas Wesentliches fehlt. Fesselnde Bilder allein reichen bekanntlich nicht aus, erst eine hierzu korrespondierende akustische Umsetzung ergibt eine stimmige Einheit. Musik und Geräusche unterstreichen die gewünschte Stimmung – mindestens, denn oft ist es erst der „richtige Ton“, der die beabsichtigte emotionale Wirkung erzielt. Wir AV-Autoren sind also gut beraten, auch in diesen erfolgsrelevanten Teil unserer Produktion einen angemessenen Aufwand zu investieren.

Glücklich kann sich schätzen, wer an dieser Stelle nur mal eben schnell ins Audioarchiv greift und ein Werk zur Hand hat,

das zur gewünschten Stimmung und von der Länge her passt. Meine Realität aber besteht oft aus vielen Stunden mühsamer Suche, heutzutage etwas erleichtert durch das Medium Internet mit seinen Suchwerkzeugen. Viele bis sehr viele Interpreten und ihre Werke müssen angehört werden, um am Ende hoffentlich das ideale, d. h. stimmige Paar für Bild und Musik zu finden. Eine Prozedur, die für jede Passage unserer AV-Show erneut durchlaufen werden muss.

Es stellt sich die Frage, ob dieser zwar interessante, aber sehr zeitintensive Prozess nicht noch etwas optimiert werden könnte. Gibt es da nicht ein KI-Tool? Hinsichtlich der Geräusche hat das bereits vorgestellte KI-Tool *Elevenlabs* vor einiger Zeit nachgebessert. Per Prompt den Wunsch beschreiben und dann aus vier Varianten auswählen, das klappt schon ganz gut. Aber wie sieht es mit kurzen Musikpassagen oder gar ganzen Musikstücken aus?

Und welche Überraschung: es findet sich eine Vielzahl von Anbietern, die uns versprechen, mit ihrem Tool am Beginn einer großen Komponistenkarriere zu stehen. Sehr verlockend, denn schon die Schulzeit

hat mir die persönlichen Grenzen meiner musikalischen Begabung aufgezeigt. Musik hören mit größtem Vergnügen, nicht aber selbst komponieren. Also begeben wir uns erwartungsvoll auf eine Reise in die neue Welt der KI-Musik. Ein paar Klicks später wird klar, dass derzeit zwei Tools „den Ton angeben“ – SUNO.AI und UDIO. Diese beiden sollen auch hier im Mittelpunkt stehen.

SUNO.AI und UDIO

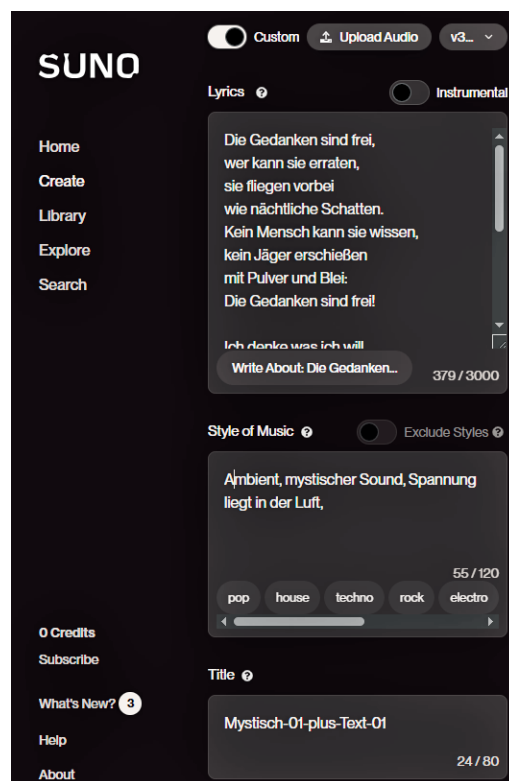
Bevor es losgehen kann, ist die übliche Registrierung erforderlich, und wie so oft im Bereich der KI steht uns ein (mengenmäßig begrenzter) kostenloser Einstieg zur Verfügung. Dazu mengen- und funktionserweiterte Abonnements (ab ca. 10 bis 12 Euro pro Monat). Erfreulich ist, dass bereits die kostenlosen Versionen einen weitreichenden Einstieg ermöglichen.

Also, los geht's. Auf der Startseite erwartet uns quasi als Inspiration eine Fülle bereits generierter Stücke inklusive der zugrundeliegenden Prompts, anhand derer man ein erstes Gefühl dafür entwickeln kann, welche Eingaben die KI zu welchen Ergebnissen verarbeitet, denn auch im Bereich der KI-Musik führt der Weg zum Ziel standardmäßig über eine Text- bzw. Prompterfassung.

Welche Aussagen beschreiben unsere musikalische Zielvorstellung? Wesentlich an dieser Stelle ist mindestens die Angabe des gewünschten Musikstils bzw. Genres, also Folk, Pop, Jazz, Easy Listening, Singer-Songwriter, Blues etc. Ergänzt werden sollte die Angabe um Aussagen zur Stimmung (melancholisch, fröhlich, episch etc.). Wer möchte, kann noch Angaben zur Instrumentierung hinzufügen. Hier ist also der Platz, um unsere Wünsche treffend zu

formulieren. SUNO erlaubt an dieser Stelle etwas knappe 120 Zeichen. UDIO hingegen unterstützt u. a. durch eine sehr hilfreiche dynamische Vorschlagsliste von inhaltlich passenden Promptbegriffen.

Es erscheint naheliegend, zusätzlich den Namen eines Musikers oder Komponisten einzugeben, z. B. Ludovico Einaudi, da dessen Werke und Stil bekannt sind. Die Folge: eine Fehlermeldung. Die Betreiber der KI scheuen Rechtsstreitigkeiten und unterbinden daher die Angabe von noch lebenden, aber auch bereits verstorbenen Komponisten oder Musiker. UDIO hilft an dieser Stelle, indem es ergänzend zu unse-



Die Oberfläche des KI-Tools SUNO

rem Komponistenwunsch entsprechende Stil-Charakterisierungen des Sounds vorschlägt, das ist sehr praktisch.

Bevor wir nun unser erstes musikalisches Werk in Auftrag geben, müssen wir uns entscheiden, ob wir als Ergebnis ein Instrumentalstück oder doch einen Gesangstitel erwarten. Derzeit dominiert jedenfalls das Genre der Instrumentalmusik im audiovisuellen Bereich. Mit einem einfachen Schalter verbannen wir den Gesang aus dem Ergebnis.

Ein Klick auf CREATE, und die KI-Tools legen los. Es dauert nur wenige Augenblicke, bis wir zwei leicht unterschiedliche Ergebnisse zur Begutachtung vorliegen haben. Apropos Ergebnis: Technisch gesehen erhalten wir MP3-Stereo-Dateien. Bei SUNO mit 192 kBit/Sek., UDIO liefert 320 kBit/Sek., im kostenpflichtigen Abo auf Wunsch auch WAV-Dateien.

Ziemlich schnell wird klar, dass detailliertere Prompt-Angaben, garniert mit musikalischen Fachbegriffen, passendere Ergebnisse liefern, wenngleich das Thema „Wunscherfüllung“ insgesamt noch deutlich Luft nach oben hat. Das Verhalten der KI kennen wir bereits, es braucht etwas Geduld, um die Sprache der KI-Tools zu treffen, um sich mit stetig angepassten Prompts und einigen weiteren Versuchen iterativ dem gewünschten Ergebnis anzunähern.

Doch zurück zur Entscheidung Gesangstück oder nicht? Schauen wir hier genauer hin, wird es äußerst interessant, denn auf Wunsch nehmen die Tools in einem separaten Feld (SUNO: Custom-Mode) auch von uns selbst erstellte Liedtexte (Lyrics) an. Wer möchte, strukturiert das Werk mit sogenannten Meta-Tags feiner (Intro, Ver-

se/Strophe, Chorus/Refrain, Solo-Passagen und Outro als Beispiel). Im Prompt bestimmen wir zusätzlich z. B. die Ausprägung des Gesangs (Mann, Frau, Chor).

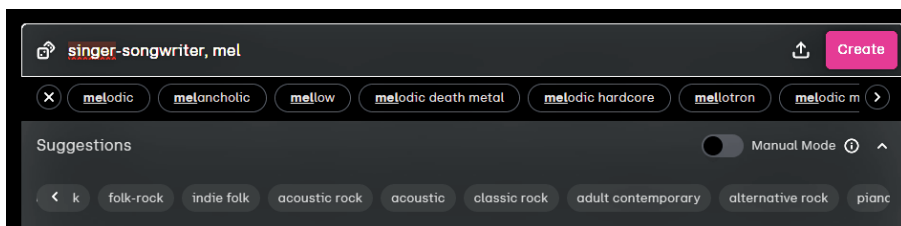
Damit öffnet sich eine wirklich große Tür zu einem Heidenspaß! Um Ihnen einen Eindruck zu vermitteln, habe ich mir u. a. einmal deutsche Kinderlieder vorgenommen. Einige Beispiele finden Sie in den Links am Ende des Artikels.

Wie wäre es, ein oder mehrere Lieder mit eigenen, thematisch passenden Texten für unser nächstes AV-Projekt zu verwenden? Bisher undenkbar, ab jetzt durchaus in greifbare Nähe gerückt. Denken Sie mal kurz darüber nach, ein wirklich kreatives Feld für unser schönes Hobby. Auch hier ein Beispiel unter den Links.

Ein Blick auf die Ergebnisse offenbart leichte Unterschiede vor allem bei der Länge der „Werke“. Im ersten Durchlauf produziert SUNO bis zu vier Minuten Musik, wobei die generierten Stücke in der Regel „rund“, d. h. mit Intro etc. erstellt werden. Je nach Länge der Songtexte auch mal kürzer. UDIO füllt seine Produktionsstufen (aktuell „nur“ 0:32 bzw. 2:10 Min.) immer komplett aus, ggf. auch einmal musikalisch nicht ganz so vorteilhaft.

Entspricht unsere erste „Ein-Klick-Produktion“ noch nicht unseren Vorstellungen, stehen uns diverse Werkzeuge zur Verfügung (z. T. erst in den kostenpflichtigen Abonnements).

Beginnen wir mit einer „musikalischen Erweiterung“. Beide Kandidaten bieten die Möglichkeit, eine weitgehend stimmige Ergänzung (EXTEND) vorzunehmen, z. B. durch Einfügen eines Intros, Outros oder Fortsetzung/Ersetzung ab einer bestimmten Stelle.



Auswahlmöglichkeiten im KI-Tools UDIO

UDIO ermöglicht zusätzlich sogenannte InPaintings, d. h. direkte Änderungen an beliebiger Stelle: Markieren, Änderung vorgeben (z. B. eine einzelne Liedtextzeile anpassen oder ein Instrumentensolo einfügen etc.), und die KI versucht eine oft erfolgreiche Umsetzung.

Interessant ist auch die STEMS-Funktion (derzeit UDIO, für SUNO angekündigt). Das bedeutet, dass Gesang und einzelne Instrumentenspuren, konkret Bass, Drum und Others, getrennt an uns übergeben werden. Der Toningenieur in uns erhält damit die Freiheit, im Nachhinein den Mix in Teilen neu zu erstellen, oder z. B. die Stimmen komplett wegzulassen, um doch noch ein reines Instrumentalstück zu generieren, das kreativ auch als musikalische Klammer verwendet werden kann. Gut, dafür gibt es auch heute schon diverse separate, oft KI-basierte Tools, Stichwort Karaoke.

Mein Favorit ist eine der neuesten Erweiterungen – die Möglichkeit des Audio-Uploads.

In SUNO wird dieses kurze Musik- oder auch nur Rhythmusbeispiel (6 bis 60 Sek.) als Ausgangspunkt für eine Komposition verwendet, um nach dieser vorgegebenen Struktur plus ergänzenden Promptangaben zum gewünschten Genre einen kompletten Song zu generieren (funktioniert nicht mit urheberrechtlich geschützter Musik!). Ein

weiteres Anwendungsbeispiel: Die Vorgabe besteht aus einigen A-cappella-Strophen (die eigene Stimme?). Wenn man nun einen ganz anderen Text dazu schreibt, versucht SUNO aus der Stimme und dem neuen Text einen Song zu generieren. Auch der umgekehrte Weg, also ein Instrumentalstück (Upload) mit einer Gesangsspur über den Lyrics-Prompt zu versehen, funktioniert; sehr beeindruckend. UDIO erlaubt es, unseren Upload an jeder beliebigen Stelle zu ergänzen. Intro, einfach den Mittelteil fortsetzen oder ein Outro hinzufügen, alles ist möglich. Außerdem wird auch urheberrechtlich geschütztes Material als Upload akzeptiert – vorerst noch.

Zusammenfassend kann man sagen, dass beide KI-Musik-Tools atemberaubende Ergebnisse liefern und viel Spaß machen. Wenn unsere Ziele jedoch spezifischer sind, kommen wir nicht umhin, einige spezielle musikalische Fachbegriffe zu verwenden. Insgesamt glänzt SUNO schon bei der allerersten Generierung mit „ausgereifteren“ Ergebnissen. UDIO hat mir als Laien zwar subjektiv besser gefallen (abwechslungsreichere musikalische Umsetzung), aber nach der ersten Generierung sind häufiger Nachbearbeitungen notwendig, um einen kompletten Song zu erhalten. Dafür erlaubt UDIO wesentlich tiefere, nachträgliche Eingriffe, richtet sich also gefühlt etwas

mehr an Musiker mit deren Nachbearbeitungs-Workflow. Für beide Tools gilt: Etwas ärgerlich ist die Tatsache, dass erneute Generierungen mit dem gleichen Prompt oft inhaltlich völlig anders ausfallen inklusive veränderter Gesangsstimme. Die Lösung: Über den Umweg der EXTEND-Funktion oder aufwendiger eines UPLOADs gelangen auch Variationen einer bereits weitgehend passenden musikalischen Grundidee.

Insgesamt gilt auch hier, dass die KI-Weiterentwicklungen uns in naher Zukunft sicher noch viel Überraschendes beschern werden. Beide Tools erzeugen schon heute in kürzester Zeit wahre Wunderwerke – probieren Sie es aus!

Rechtliches

Was wäre ein Artikel über Musik ohne das Thema „Rechte“! Beginnen wir mit dem Urheberrecht. Hier verhält es sich ähnlich wie z. B. bei den Bildgeneratoren. Das Urheberrecht setzt eine persönliche (menschliche) geistige Schöpfung voraus (§ 2 Abs. 2 UrhG). Diese ist nach derzeitiger Rechtsauffassung auch im Bereich der KI-Musik nicht gegeben, d. h. es besteht kein Urheberrecht, die Ergebnisse sind damit rechtfrei. Aber Achtung, auch Songtexte sind urheberrechtlich geschützt. Also nicht einfach bestehende Texte kopieren und als Lyrics für die KI verwenden. Bezüglich der Nutzungsrechte erlauben die kostenlosen Angebote eine nicht-kommerzielle Nutzung. Die kostenpflichtigen Abos beinhalten sogar das Recht zur uneingeschränkten kommerziellen Nutzung. Dies auch dann, wenn das Abonnement zwischenzeitlich gekündigt wurde. Insgesamt also für uns erfreulich und transparent.

KI mischt unsere ganze Welt auf, auch

die Musikindustrie. Die ersten Klagen sind erwartungsgemäß bei den KI-Anbietern angekommen, Stichwort Trainingsdaten. Unbestritten ist, dass die Luft für Komponisten von sog. „Kaufhausmusik“ dünner und es auch sonst mehr oder weniger große Veränderungen geben wird. Wie diese aussehen werden, wird die Zukunft zeigen. Die KI wird schon nicht die Charts stürmen, meinen Sie? Dann sollten Sie sich mal „Verknallt in einen Talahon“ vom Künstler Butterbro anhören. Dieser KI-Song (der Text wurde noch „menschlich“ erstellt) schaffte es bereits in die deutschen Top 50. Noch weiter kam „Zo Zomer“ von John de Koning, der es in den Niederlanden auf Platz 3 schaffte! Die ersten KI-Songs in den Charts, aber sicher nicht die letzten.

Vielleicht fühlt sich der eine oder andere Leser ermutigt, uns an dieser Stelle über seine Erfahrungen z. B. bezüglich der Verwendung von Fachbegriffen zu berichten. Wir würden uns freuen.

Komme ich mit KI also schneller ans Ziel? Nüchtern betrachtet hängt das wohl von zwei Faktoren ab, der „Festigkeit“ meiner Vorstellungen und der Anzahl der dafür notwendigen Iterationen. Und dann ist da aber noch der Spaßfaktor, und der ist mit der neuen Technik wirklich enorm. Probieren Sie es aus! □

Links

Suno: suno.com/create

Udio: ud.io

„Verknallt in einen Talahon“ von Butterbro:
youtube.com/watch?v=o0aG3S3iD6Y

Zo Zomer von John De Koning:
youtube.com/watch?v=G5-gwJjotJI&t=6s

Alle Links und Ton-Beispiele finden Sie auch auf av-dialog-magazin.de (Web-Links)