

»Wir erleben gute und schlechte Tage mit KI«

EIN GESPRÄCH MIT NIC BUCHANAN*

Was kann KI?

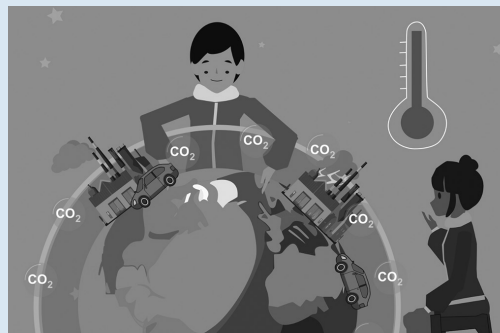
Buchanan: Es scheint, dass KI fast alles kann, was wir Menschen auch können. Wir stoßen zwar an gewisse Grenzen, aber ich denke, die Probleme werden in naher Zukunft gelöst werden, und die Technologie wird reibungslos funktionieren. KI lernt immer noch dazu und jede Woche kommen neue Entwicklungen und neue Software hinzu. Ich denke also, dass sie praktisch alles kann.

Sie arbeiten an der Umsetzung von *Das Lexikon der Klimakrise der Initiative For Us – No Planet B!*, in der der Off-Text für möglichst viele Sprachen per KI übersetzt werden soll. Mit welchen Tools haben Sie bisher gearbeitet?

Buchanan: Die Reise begann mit großer Begeisterung und Enthusiasmus, die auch weiterhin bestehen. Es hat allerdings eine Weile gedauert, die richtige Software zu finden und nicht einfach der Werbung und dem Hype um das neueste, teuerste Programm zu folgen. Diverse Tools versprechen viel, lösen ihre Versprechungen aber nicht ein. Wir kauften beispielsweise Lizenzen für Programme, die versprachen, alles sei möglich, doch wir scheiterten mit ihnen dann doch oft, weil von Softwareseite Probleme auftraten. Besonders zu Beginn der Reise war es sehr schwierig herauszufinden, welche Sprachsoftware sich tatsächlich für TV und Social Media eignet. Die Suche wurde dann viel einfacher, als feststand, dass die Beiträge des Lexikons ausschließlich für eine Online-Nutzung angedacht waren, keine Sendequalität

Das Lexikon der Klimakrise

Einfache und ansprechende Animationen (max. 1,30 Min.) erklären grundlegende Begriffe der globalen Erwärmung. Die Animationen wurden von Teams aus verschiedenen Ländern erstellt (z. B.: CO₂-Fußabdruck und Fast Fashion (Brasilien) oder Fossile Brennstoffe und CO₂ (Südafrika, s. Abb.)) und werden in zahlreichen Sprachen online zur Verfügung gestellt.



haben mussten und auch die Audio-settings nicht das höchste Niveau haben mussten. Diese Einschränkung öffnete viele Türen für einfachere, aber zweckmäßige Software.

Auch bei visueller generativer KI, insbesondere bei Animationen, lautet die erste Frage, wie das Endprodukt genutzt wird. Dementsprechend wählt man dann die Software. Seit dem Start von *Das Lexikon der Klimakrise* sind nun fast 2 Jahre vergangen und ich finde das Projekt weiterhin faszinierend und liebe diese Reise.

Können Sie uns Schritt für Schritt die wichtigsten Etappen Ihrer Reise mit KI und den verwendeten Programmen erläutern?

Buchanan: Zu Beginn war die Text-zu-Video-Software *HeyGen* das

angesagte Programm. Die Firma hatte Verträge mit Hollywood-Schauspieler*innen abgeschlossen, um deren Stimmen verwenden zu können. Wir versuchten, einfache Erklärvideos mit *HeyGen* zu vertonen, aber das funktionierte nicht. Nach mehreren Versuchen landeten wir bei *ElevenLabs*, das sich für uns wie der »Heilige Gral der Software« anfühlte. Dennoch mussten wir in den letzten 6 Monaten 2 weitere Male die Software wechseln und nutzen jetzt *Resemble*, das ich vor einem Jahr noch nicht einmal kannte. Wir müssen quasi ständig am Ball bleiben, Entwicklungen beobachten und uns neu aufstellen. Und obwohl *Resemble* momentan unser Favorit ist, könnte es passieren, dass wir irgendwann wieder zu *HeyGen*

oder *ElevenLabs* zurückkehren werden. Zum Glück ist die Programmierung nicht unsere Aufgabe. Es gibt Profis, die versuchen, unsere Bedürfnisse zu verstehen. Wir haben Workshops mit verschiedenen Softwarefirmen abgehalten, um zu erklären, was wir benötigen. Dieser Austausch ist für beide Seiten wertvoll. Nur wenn sie unsere Ziele verstehen, können sie Lösungen finden, zum Beispiel wenn wir eine deutsche Stimme in eine portugiesische oder spanische Stimme umwandeln möchten.

Was waren die größten Herausforderungen?

Buchanan: Eine große Herausforderung sind »Halluzinationen«, die man bei allen KI-Systemen findet. Halluzinationen bedeutet, dass ein Programm sich selbst

nicht versteht. Im Englischen nennen wir es oft »brain fart«, ein Aussetzer, der jedem oder jeder passieren kann, wenn man müde ist. Ebenso wird die KI gelegentlich »müde«. Diese Halluzinationen führten beim *Lexikon der Klimakrise* beispielsweise dazu, dass wir 30 Sekunden mit der eingekauften KI-Stimme erhielten und die Software dann plötzlich zur Originalstimme zurücksprang.

Es ist ungut, nicht zu wissen, wie und warum diese Fehler auftreten. Selbst die Software-Entwickler*innen konnten es manchmal nicht erklären. Wir vermuteten zunächst, dass die Internetverbindung und -geschwindigkeit in Afrika eine Rolle spielte, aber die Entwickler*innen meinten, dies würde nur die Geschwindigkeit verlangsamen, nicht aber die Leistung beeinträchtigen. Das heißt, wir erleben gute und schlechte Tage mit KI.

Die 2 größten Herausforderungen sind somit, die Qualität zu halten und mit Halluzinationen umzugehen. Diese sind zwar weniger geworden, aber immer noch vorhanden.

Auf welche Probleme stießen Sie bei den Übersetzungen?

Buchanan: Halluzinationen zeigten sich auch bei den Übersetzungen. *ElevenLabs* erwies sich für interne Übersetzungen als recht gut geeignet, mit nur wenigen Fehlern bei der Umwandlung englischer Audios ins Deutsche. Das war ein Pluspunkt dieser Software. *Resemble* dagegen greift auf *Google Translate* zurück und fügt manchmal Wörter oder Sätze hinzu, die so nicht im Skript stehen.

Wie definieren Sie die Qualität von KI-generierten Stimmen und Übersetzungen?

Buchanan: Es ist tatsächlich ein Paradox: Wir nutzen KI, um eine menschliche Stimme zu erzeugen. Wir wünschen uns den menschlichen Klang in dieser KI-generierten Stimme, wollen oder können aber aus Zeit- und Kostengründen keine menschlichen Stimmen verwenden. Wir durchlaufen

quasi den ganzen Prozess der Arbeit am Computer und mit verschiedener Software, um am Ende eine möglichst authentische menschliche Stimme zu bekommen – das ist eigentlich ein bizarrer Weg, den wir hier gehen.

Da ich selbst nur 2 Sprachen spreche, kann ich die Qualität der Übersetzungen nicht bewerten, sondern kann nur auf den Klang der Stimme achten und darauf, ob eine gewisse Emotion in der Darbietung überkommt. Die nächste Generation von KI-Stimmensoftware wird es sicherlich möglich machen, Wärme und verschiedene Emotionen wie Wut in eine Stimme einfließen zu lassen, ähnlich wie es bei Sprachaufnahmen mit Menschen möglich ist. Doch letztlich hängt alles vom Sound-Editing ab und davon, wie stark wir die Audios, ob von Menschen gesprochen oder computergeneriert, optimieren können. Wie gesagt, es ist eine bizarre Reise von Mensch zu Computer und zurück zum Menschen.

Arbeiten Sie auch bei den Animationen mit visueller generativer KI?

Buchanan: Ja. Es gibt viele Programme, die aus Text-Prompts visuelle Inhalte generieren, die dann wiederum in kurze Motion Clips umgewandelt werden können.

Visuelle KI bietet 2 Möglichkeiten: Wir können zum einen die Figuren aus einer bereits existierenden Animationsendung in die KI einspeisen und ihr auftragen, diese Figuren verschiedene Aktionen ausführen zu lassen, wie zum Beispiel einen Baum hochzuklettern oder zu laufen. Zum anderen kann man die KI völlig neue Charaktere und Settings entwerfen lassen.

Ich habe dabei jedoch feststellen müssen, dass die Technologie für Animationen der Sprach-KI noch hinterherhinkt, auch wenn man manchmal beeindruckende Demos in Hollywood-Qualität sieht, die in irgendeiner Hinterhofgarage produziert wurden. Kurzum: Die Technologie wächst, entwickelt sich weiter und verbessert sich, allerdings ist sie kein Wundermittel für alle unsere Probleme.

Wird es nächstes Jahr eine KI-basierte Animation geben?

Buchanan: Ich glaube, das wird es. Wir haben uns im Team entschieden, bei der Ideenentwicklung auf von Menschenhand gezeichnete Concept Art zu setzen und diese dann von der KI ausarbeiten zu lassen. Animationen haben, wie jedes Filmmaterial, eine Bildfrequenz von 25 Bildern pro Sekunde. Wenn wir an die alten Disney-Zeiten zurückdenken, in denen jedes Bild manuell gezeichnet und die Bewegungen durchgeblättert wurden, sieht man ganz deutlich, dass KI hier wirklich viel helfen kann, weil man eben nicht mehr 25 Bilder pro Sekunde zeichnen muss. Man zeichnet ein Bild und die KI animiert die nächsten 5 Sekunden in weniger als einer Minute. Es funktioniert nicht immer perfekt, da KI, wie bereits erwähnt, oftmals halluziniert, aber für eine*n menschliche*n Animator*in wäre dies eine Aufgabe für mehrere Tage gewesen – und auch hier wären wahrscheinlich Fehler passiert, die in weiteren Arbeitstagen hätten korrigiert werden müssen.

Die Arbeit mit KI ist kein kreativer Prozess, aber KI ist ein Werkzeug, das wir nutzen müssen, wenn wir ohne große Hollywood-Budgets unsere Kreativität ausdrücken und etwas ins Fernsehen oder auf die Kinoleinwand bringen wollen. KI bietet uns die Chance, selbst mit kleinen Budgets großartige, qualitativ hochwertige Programme für Kinder zu produzieren und sie durch KI-generierte Sprachübersetzungen einer globalen Zielgruppe zugänglich zu machen. ■

ANMERKUNG

¹ <https://prixjeunesse.de/share-pool/for-us-no-planet-b/>; https://izi.br.de/english/publication/television/37_2024_E/What_we_can_do.pdf

* Nic Buchanan ist Geschäftsführer des 2D Animationsstudios Pix-comm, Südafrika.

