



Vertragsbedingungen KI

Empfehlungen zur Vertragserstellung für Sprachsynthesen zum Einsatz in generativer KI vom Verband Deutscher Sprecher:innen
www.sprecherverband.de

Aktualisiert Anfang 2024

Da die Synthetisierung von Stimmen und der Einsatz dieser mittels KI zum aktuellen Zeitpunkt beinahe unreguliert ist, sodass die Rechte der Beteiligten nicht eindeutig geschützt sind (z.B. Persönlichkeitsrecht, Urheberrecht, Vergütungsanspruch, Haftung uvm.), warnt der VDS seine Mitglieder eindringlich vor einem sorglosen Umgang mit derlei Produktionen.

Keine Gagenliste kann den vollständigen Verlust der Marktwirksamkeit eines/r Sprecher:in und der Selbstbestimmung über die eigene Stimme abbilden.

Aus diesem Grund gibt der VDS zum aktuellen Zeitpunkt keine Gagenempfehlungen für dieses Segment heraus.
Ausgenommen hiervon sind klassische TTS Sprachsynthesen.

Da der Einsatz von KI und synthetischen Stimmen nicht aufzuhalten ist und auch manch Gutes mit sich bringt (z.B. Barrierefreiheit), engagiert sich der VDS und insbesondere die AG Sprache & KI auch im internationalen Zusammenschluss der United Voice Artists für eine Regulierung dessen.

Nachfolgend finden Sie die Bedingungen, welche bei Aufträgen zur Sprachsynthese für die Nutzung in generativer KI unbedingt eingehalten, vor der Unterzeichnung einer NDA schriftlich bestätigt und vor der Aufnahme in einem Vertrag schriftlich festgehalten werden sollten:

Vertragsbedingungen

1. Vergütung

Die Vergütung sollte angemessen sein. Eine Vergütung ist angemessen, sofern sie folgende Faktoren berücksichtigt:

- Zeitinvestition und Arbeitsaufwand für die Aufnahme und etwaige Nachbearbeitung
- Verwertungsrechte (Dauer, Medien, Produkte/Programme, Ort)
- ggf. laufende Lizenzgebühren für Einzelauswertungen bei einzeln buchbarer Stimmennutzung (z.B. Portal, bei welchem einzeln Stimmen für einen bestimmten Nutzungszweck eingekauft werden können, z.B. für Telefonansagen, Werbespots usw.)
- Abdeckung etwaiger Exklusivität für den Verwertungszeitraum und damit Berufsverbot des/der Sprecher:in
- Abdeckung der entstandenen Überpräsenz und des Stimmverbrauchs des/der Sprecher:in während und nach dem Verwertungszeitraum

2. Urheberrecht & Leistungsschutzrecht

Es sollte klar definiert sein, wer die Rechte an der synthetisierten Stimme besitzt und wie sie genutzt werden darf. In Deutschland sind Sprecher:innen durch das Urheberrecht und das Leistungsschutzrecht geschützt.

3. Datenschutz

Der/die Kund:in muss sicherstellen, dass die persönlichen Daten des/der Sprecher:in geschützt sind. In Deutschland ist der Datenschutz durch die Datenschutz-Grundverordnung (DSGVO) streng geregelt.

Die für die Synthese erforderlichen Daten sowie daraus resultierende Informationen dürfen zu keiner Zeit außerhalb Deutschlands übertragen oder verarbeitet werden. Das heißt, die Server, auf welchen die KI gehostet wird und ebenso die Server, auf welchen der Algorithmus liegt und auf denen der Klon verarbeitet wird, sollten in Deutschland stehen, um einem Missbrauch vorzubeugen und angemessenen Datenschutz zu gewährleisten.

4. Persönlichkeitsrecht - Das Recht an der eigenen Stimme

Der/die Sprecher:in ist durch das Persönlichkeitsrecht, welches das Recht an der eigenen Stimme umfasst, besonders geschützt. Dieses Recht kann nicht übertragen werden, auch wenn es sich um eine Synthetisierung der eigenen Stimme handelt. Der/die Sprecher:in sollte somit eine Verwertung der Aufnahmen, die über den ursprünglich vereinbarten Zweck hinausgeht, ablehnen können, auch wenn hierfür eine zusätzliche Vergütung angeboten wird.

5. Spezifische Produkte/Programme

Der Vertrag sollte genau angeben, für welche Produkte oder Programme die synthetisierte Stimme verwendet werden darf und in welchem Umfang diese Nutzer:innen zur Verfügung gestellt wird. Dies könnte eine bestimmte Software, ein bestimmtes Gerät, eine bestimmte Plattform oder eine bestimmte Dienstleistung sein. Es kann auch eine uneingeschränkte Nutzung vereinbart werden. Eine eindeutige Definition ist jedoch notwendig um die Vergütung und Nutzungsrechte (s.o.) angemessen festlegen zu können.

6. Ausschluss bestimmter Inhalte & Missbrauch

Der/die Sprecher:in sollte das Recht haben, bestimmte Arten von Inhalten auszuschließen. Dies könnte beispielsweise die Verwendung seiner/ihrer Stimme für politische, religiöse, erotische oder andere kontroverse Inhalte betreffen.

Weiterhin sollte ein etwaiger Missbrauch der Stimme seitens des/der Kund:in verboten werden und von dieser/m auf eigene Kosten strafrechtlich verfolgt werden.

Der/die Kund:in sollte hierfür die Sorgfaltspflicht tragen und den/die Sprecher:in bei Kenntnisnahme des Missbrauchs oder Verwendung in ausgeschlossenen Inhalten umgehend informieren müssen.

7. Zustimmung zu neuen Verwendungen

Wenn das Unternehmen die synthetisierte Stimme für ein neues Produkt, Programm, Funktion oder einen veränderten Nutzerkreis verwenden möchte, welche nicht im ursprünglichen Vertrag aufgeführt sind, sollte es die Zustimmung der Sprecher:innen schriftlich als Vertragsergänzung einholen müssen.

8. Qualität der Synthese

Der/die Sprecher:in könnte daran interessiert sein, wie gut die synthetisierte Stimme seine/ihre natürliche Stimme nachahmt und wie gut diese Sprache umgesetzt (z.B. Phonetik, Sprachmelodie, Betonung, Ausdruck). Eine schlechte Synthese könnte seinen/ihren Ruf zusätzlich schädigen. Der/dem Sprecher:in sollte das Recht zugestanden werden, die Synthese nach Fertigstellung abzunehmen und vor der Veröffentlichung eine qualitativ treffendere und realitätsgetreuere Sprachsynthese der eigenen Stimme zu verlangen.

9. Vertragsbedingungen

Die Vertragsbedingungen sollten klar und verständlich sein und alle relevanten Punkte abdecken. In Deutschland ist es besonders wichtig, dass der Vertrag den Anforderungen des Bürgerlichen Gesetzbuches (BGB) entspricht.

10. Anwalt & Anwaltskosten

Der/dem Sprecher:in sollte ein ausreichender zeitlicher Vorlauf gewährt werden, um den Vertrag durch eine/n selbst gewählte:n Anwält:in prüfen zu lassen. Eine anwaltliche Prüfung und Rechtsberatung bei Verträgen zu Sprachsynthese ist unabdinglich. Die Kosten hierfür sollten von der/dem Kund:in getragen werden.

11. Vertragssprache

Der Vertrag sollte auf Deutsch verfasst sein, damit der/die Sprecher:in ihn vollständig verstehen kann.

12. Gerichtsstand

Der im Vertrag angegebene Gerichtsstand muss Deutschland sein.

13. Haftungsausschluss

Der/die Sprecher:in sollte von jeglicher Haftung für die mit seiner/ihrer synthetisierten Stimme hergestellten Inhalte freigesprochen werden. Sämtliche Haftung liegt bei dem/der Kund:in. Diese:r sollte hierfür eine Haftpflichtversicherung abgeschlossen haben.

Fragen

Folgende von NAVA (National Association Of Voice Actors) definierten Fragen sollten von Kund:innen, die Stimmen synthetisieren wollen, den Sprecher:innen beantwortet werden können:

Fragen an Tech-Kunden

Frage	Was bedeutet das?	Was wäre der Nutzen?
Wird die Stimme vom Kunden zur Synthese und Erstellung abgeleiteter Werke verwendet? Wird die Stimme zur Erstellung abgeleiteter Werke durch Dritte verwendet? Wenn ja, werden abgeleitete Werke ausgestrahlt, kommerzialisiert oder vertrieben?	Werden neue Inhalte mit meiner Stimme vom Kunden oder einer anderen Person erstellt und verkauft? Wenn ja, werde ich dafür entschädigt?	Standard-TTS, Erstellung digitaler Stimmen und Personen, visuelle multimodale Interaktion, künstliche Intelligenz, virtuelle Realität oder andere realitätsbasierte animierte Bilder, Webcasting, Werbung, Performance, Rundfunk, Nachrichtenübertragung, Hörbücher, intelligenter Kundenservice, intelligente Geräteinteraktion, Wir-Medien-Synchronisation usw. (diese Liste ist endlos, aber ein guter Anfang für die Erstellung einer Liste von Arten abgeleiteter Werke).
Wird die Stimme verwendet, um bestehende Sprachdienstmodelle zu erstellen, zu verbessern oder zu erweitern?	Wird meine Stimme verwendet, um bestehende Sprachtechnologien unabhängig von KI zu verbessern?	Spracherkennung und -generierung, einschließlich Sprachtranskription (STT), Text-to-Speech, Sprachübersetzung, Unternehmensbürosoftware, cloudbasierte neuronale Netze und Datenanalyse
Wird die Stimme synthetisiert und/oder mit anderen Stimmen kombiniert (abgeleitete Stimmen) oder wird die Stimme in ihrem aktuellen Timbre geklont?	Wird die Stimme wie ich klingen? Wird aus meiner und anderen Stimmen eine neue Stimme entstehen?	Standard-TTS, Erstellung digitaler Stimmen und Personen, visuelle multimodale Interaktion, künstliche Intelligenz, virtuelle Realität oder andere realitätsbasierte animierte Bilder, Webcasting, Werbung, Performance, Rundfunk, Nachrichtenübertragung, Hörbücher, intelligenter Kundenservice, intelligente Geräteinteraktion, Wir-Medien-Synchronisation usw.

Frage	Was bedeutet das?	Was wäre der Nutzen?
Werden diese Sprachdaten zum Training der KI verwendet?	Wird meine Stimme die zukünftige KI-Technologie verbessern?	Maschinelles Lernen, KI-Training
Werden diese Sprachdaten als primäre Identität der KI verwendet?	Wird meine Stimme die Hauptstimme der KI sein und als KI dargestellt werden?	Diverse
Gibt es funktionale Einschränkungen dieser KI oder ist die KI nicht generativ? Bsp.: eine Empfehlungsmaschine, ein Informationskiosk oder ein anderer begrenzter Zweck?	Gibt es Grenzen für die Möglichkeiten dieser KI und gibt es eine Spezialität oder einen spezifischen Verwendungszweck für diese KI?	Diverse
Ist die KI generativ?	Kann die KI bei Aufforderung selbständig und ohne Einschränkungen neue Inhalte erstellen?	Diverse
Können sich diese Nutzungsbedingungen ändern?	Kann diese Vereinbarung nach Ermessen des Kunden geändert werden?	Diverse
Hat der/die Sprecher:in das Recht, die Genehmigung zu widerrufen?	Kann ich meine Stimme in Zukunft aus diesen Technologien entfernen oder besitzt der Kunde meine Stimme ohne Einschränkungen?	Diverse

Fragen an Indie-Kunden

Gibt es einen Teil dieses Projekts, der KI-generiert ist? Skripte, digitale Assets, Stimme, Musik usw.?

Gibt es Elemente dieses Projekts, die Sie als NFTs verkaufen möchten?

Wären Sie bereit, eine einfache Vereinbarung zu unterzeichnen, die besagt, dass meine Stimme nicht von KI geklont oder zum Trainieren von KI verwendet wird?

Bedenken

Um spezifische Bedenken zu adressieren, könnten Sprecher:innen und Kund:innen folgende Maßnahmen ergreifen:

Verlust der Kontrolle über die eigene Stimme

Um dies zu vermeiden, sollte der Vertrag klar definieren, wie und wo die synthetisierte Stimme verwendet werden darf. Der/die Sprecher:in sollte das Recht haben, bestimmte Verwendungen auszuschließen und die Zustimmung für neue Verwendungen zu verlangen.

Verlust der Einzigartigkeit

Um die Einzigartigkeit der Stimme des/der Sprechers/Sprecherin zu bewahren, könnte der Vertrag eine Klausel enthalten, die die Verbreitung der synthetisierten Stimme begrenzt. Beispielsweise könnte die Verwendung auf bestimmte Produkte oder Dienstleistungen beschränkt werden.

Verlust von zukünftigen Einnahmen

Um dies zu vermeiden, könnte der/die Sprecher:in eine laufende Lizenzgebühr für die Nutzung der synthetisierten Stimme und die Zeit danach verlangen. Dies würde sicherstellen, dass der/die Sprecher:in weiterhin Einnahmen aus der Verwendung und dem Reputationsverlust seiner/ihrer Stimme erhält.

Datenschutz und Sicherheit

Um die persönlichen Daten des/der Sprechers/Sprecherin zu schützen, sollte der/die Kund:in strenge Datenschutz- und Sicherheitsmaßnahmen einhalten. Der Vertrag sollte auch Klauseln enthalten, die den Missbrauch der synthetisierten Stimme verbieten.

Qualität der Synthese

Um sicherzustellen, dass die synthetisierte Stimme von hoher Qualität ist, sollte der/die Sprecher:in mit einem seriösen und erfahrenen Unternehmen zusammenarbeiten, das fortschrittliche Sprachsynthesetechnologien verwendet.

Rechtliche Bedenken

Um rechtliche Bedenken zu adressieren, sollte der/die Sprecher:in rechtlichen Rat einholen. Der Vertrag sollte die Rechte des/der Sprechers/Sprecherin an seiner/ihrer Stimme respektieren und schützen, einschließlich seiner/ihrer Urheberrechte, Leistungsschutzrechte und Persönlichkeitsrechte.

Ethik und Moral

Um ethische und moralische Bedenken zu adressieren, sollte der/die Sprecher:in das Recht haben, die Verwendung seiner/ihrer Stimme für bestimmte Zwecke auszuschließen. Der/die Kund:in sollte die ethischen und moralischen Werte des/der Sprechers/Sprecherin respektieren und seine/ihre Zustimmung für die Verwendung seiner/ihrer Stimme in bestimmten Kontexten einholen.

Durch die Zusammenarbeit und offene Kommunikation können Sprecher:innen und Kund:innen sicherstellen, dass die Synthese von Stimmen auf eine Weise erfolgt, die die Rechte und Interessen aller Beteiligten respektiert und somit Überraschungen oder langwierigen Rechtsstreitigkeiten und Schadensersatzforderungen vorbeugt.