Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag



Themenkurzprofil Nr. 93
Oktober 2026

KI-gestützte Simultanübersetzung

Simone Ehrenberg-Silies

In Kürze

KI-gestützte Simultanübersetzung kann verändern, wie Menschen in internationalen Kontexten miteinander kommunizieren. Sie ermöglicht kommunikativen Austausch, ohne dass man von seiner Mutterin eine Fremdsprache wechseln oder die Sprache des Gegenübers beherrschen muss. Anbieter solcher Systeme sind überwiegend die großen Technologieunternehmen aus den USA und China. Mit DeepL spielt jedoch auch ein europäischer Akteur eine bedeutende Rolle.

Die technische Entwicklung begann mit Modellen, die nur wenige Sprachen unterstützten. Heute wird an mehrsprachigen Übersetzungssystemen auf der Grundlage großer Sprachmodelle geforscht. An Bedeutung in Forschung und Entwicklung gewinnen zudem die multimodale Übersetzung, also die parallele Verarbeitung von Sprachsignalen und etwa Lippenbewegungen, sowie Ansätze, die kommunikative Absicht zu erfassen anstatt nur wortgetreu zu übersetzen. Zudem wird eine effiziente Verbindung menschlicher Verdolmetschung mit KI-gestützter Simultanübersetzung relevanter.

Die am Markt verfügbaren Systeme können bei Weitem noch keine flüssige, reibungslose Kommunikation von Personen mit unterschiedlichen Muttersprachen ermöglichen. Dem stehen unter anderem eine teilweise eingeschränkte Übersetzungsqualität und begrenzte Sprachabdeckung entgegen.

KI-gestützte Simultanübersetzung könnte die gesellschaftliche Teilhabe von Menschen verbessern. Je nach Ausgestaltung der konkreten Anwendung stellen sich für solche Systeme auch Fragen des Datenschutzes und des Schutzes der Vertraulichkeit des gesprochenen Wortes.

Welche mittel- bis langfristigen Folgen die Verwendung KI-gestützter Simultanübersetzungssysteme auf Kommunikation, Spracherwerb oder für das Berufsbild und die Beschäftigungssituation von Dolmetscher/innen haben könnte, lässt sich nicht verlässlich abschätzen.

Hintergrund und Entwicklungsstand

Ende der 1980er Jahre wurden die ersten Sprachübersetzungssysteme entwickelt, die akustische Sprachsignale in fremdsprachigen Text übertragen konnten. In derselben Entwicklungsphase wurden zudem die ersten Systeme zur Übersetzung von gesprochener Sprache in gesprochene Zielsprache (Speech-to-Speech Translation) realisiert (Sperber/Paulik 2020, S. 7410 f.).

Kaskadenmodelle

Frühere Systeme zur Übersetzung von gesprochener Sprache in gesprochene Zielsprache basierten auf Kaskadenmodellen. Diese bestehen typischerweise aus drei aufeinanderfolgenden Modulen: automatische Spracherkennung (Automatic Speech Recognition), maschinelle Übersetzung und Sprachsynthese (Text-to-Speech – TTS) (Sperber/Paulik 2020, S. 7409 f.).

Die Kaskadenmodelle weisen allerdings drei zentrale Schwächen auf: Durch die stufenweise Übersetzung können sich Fehler zwischen den einzelnen Modulen fortpflanzen (Error Propagation). Fehler, die in einer frühen Verarbeitungsstufe entstehen, werden an die nachfolgenden Komponenten weitergegeben und können sich dort verstärken. Wird beispielsweise ein gesprochenes Wort von der automatische Spracherkennung falsch erkannt, basiert die maschinelle Übersetzung auf dieser fehlerhaften Eingabe. Dies kann schließlich zu einer inkorrekten Übersetzung und damit zu einer fehlerhaften Sprachausgabe führen. Das war vor allem bei frühen Systemen ein Problem, die interlinguabasierte maschinelle Übersetzung mit Parsern, Programmen zur Analyse der Satzstruktur, verwendeten (Sperber/Paulik 2020, S. 7410). Bei der interlinguabasierten maschinellen Übersetzung wird der Ausgangssatz nicht direkt Wort für Wort übersetzt, sondern in eine sprachunabhängige Zwischenpräsentation, die Interlingua, übertragen. Zu diesem Zweck analysieren Parser die Grammatik der erkannten Sätze. Parser haben jedoch Schwierigkeiten, mit fehlerhaften oder grammatikalisch unvollständigen Ausgaben der automatischen Spracherkennung umzugehen (Sperber/Paulik 2020, S. 7410).

Ein weiterer Nachteil ist, dass automatische Spracherkennungs- und maschinelle Übersetzungsmodelle mit unterschiedlichen Daten trainiert werden. Während automatische Spracherkennungsmodelle mit Sprachdaten trainieren, trainieren maschinelle Übersetzungsmodelle mit Textdaten, die Interpunktion enthalten. In der Kaskade muss das maschinelle Übersetzungsmodul jedoch Text verarbeiten, der aus gesprochener Sprache generiert wurde und in dem daher Interpunktion fehlt. Auch dies kann zu fehlerhaften Übersetzungen führen (Sperber/Paulik 2020, S. 7410 f.). Übersetzungskorpora aus der gleichen Domäne (gesprochene Sprache), die gesprochene Eingaben mit gesprochenen Ausgaben verknüpfen, etwa Audioaufnahmen von Verdolmetschungssituationen oder manuelle Übersetzungen aus Daten der automatischen Spracherkennung, sind aufgrund von zu geringem Datenumfang, begrenzter Sprachabdeckung und hohen Kosten keine verbreitete Trainingsalternative (Sperber/Paulik 2020, S. 7410 f.).

Ein weiterer Nachteil ist, dass bei der Sprachübersetzung mit Kaskadenmodellen prosodische Informationen, etwa Betonungen, Pausen, Rhythmus, weitgehend verloren gehen (Sperber/Paulik 2020, S. 7412).

Neben den genannten Nachteilen weisen Kaskadenmodelle jedoch einen zentralen Vorteil auf: Ihre Modularität trägt dazu bei, dass Optimierungen einzelner Module, beispielsweis der automatischen Spracherkennung, die Leistungsfähigkeit des gesamten Sprachübersetzungssystems verbessern können (Sarim et al. 2025, S. 2).

Beispiele für Kaskadenmodelle sind die beiden Forschungsprototypen IBM Mastor und Verbmobil (Sarim et al. 2025, S. 5). IBM Mastor wurde während des Irakkriegs 2003 (IBM o. J.) im Rahmen von Programmen der Defense Advanced Research Project Agency (DARPA), einer Bundesbehörde des US-amerikanischen Verteidigungsministeriums zur Förderung sicherheitsrelevanter Technologien, entwickelt. Ziel war es, ein echtzeitfähiges Übersetzungssystem von gesprochener Sprache in gesprochene Zielsprache für ressourcenarme Sprachen zu erforschen, das auf tragbaren Endgeräten wie Laptops oder PDAs funktionieren sollte. IBM Mastor sollte vor allem die Kommunikation des militärischen und medizinischen Personals sowie von Mitarbeiter/innen humanitärer Hilfsorganisationen mit Einheimischen bei Auslandeinsätzen unterstützen (Gao et al. 2006, S. 53). Als ressourcenarme Sprachen werden Sprachen bezeichnet, für die nur unzureichende digitale Daten für das Vortraining von Sprachmodellen vorliegen (Meyer/Buys 2026, S. 1). Im Vergleich zu ressourcenstarken Sprachen, etwa Englisch und Spanisch, sind sie zudem weniger erforscht und werden seltener unterrichtet (Magueresse et al. 2020, S. 1).

Verbmobil gilt als Pionierprojekt für Übersetzungstechnologien. Unter Führung des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) erforschte ein Verbund aus 31 internationalen Partnern aus Wissenschaft und Wirtschaft zwischen 1993 und 2000 die Grundlagen der Übersetzung gesprochener Sprache in gesprochene Zielsprache in nahezu Echtzeit; auf den Ergebnissen konnten Entwicklungen wie DeepL aufbauen. Forscher/innen aus dem Verbmobil-Projekt gingen später zu globalen Techunternehmen und arbeiteten beispielsweise an Google Translate mit (Lernende Systeme. Die Plattform für künstliche Intelligenz o. J.).



Direkte Übersetzung von gesprochener Sprache in gesprochene Zielsprache

Mit Translatotron stellte Google 2019 eines der ersten Systeme für die direkte Übersetzung gesprochener Sprache in gesprochene Zielsprache vor (Jia et al. 2019). Im Unterschied zu den Kaskadenmodellen übersetzen direkte Übersetzungssysteme von einer Ausgangs- in eine Zielsprache ohne Texterzeugung als Zwischenschritt (Jia et al. 2019, S. 1). Die Architektur von Translatotron basiert auf einem neuronalen Netzwerk, das lernt, Sprachspektrogramme (Audiodaten) der Ausgangssprache mit Sprachspektrogrammen der Zielsprache zu verknüpfen. Dabei analysiert ein Encoder die eingehenden Sprachsignale, während ein Decoder die korrespondierenden Spektrogramme der Zielsprache generiert. Diese werden anschließend mithilfe eines Vocoders in Audiosignale umgewandelt (Jia et al. 2019, S. 1 f.). Um eine hohe Übersetzungsqualität zu erreichen, wurde das Modell auf die Erkennung von Lauten der Ausgangs- und Zielsprache, etwa langes i oder kurzes i, hin trainiert (Jia et al. 2019, S. 2).

Da direkte Übersetzungssysteme lediglich einen einzigen Dekodierungsschritt durchlaufen, während Kaskadenmodelle in der Regel drei umfassen, benötigen sie geringere Rechenkapazitäten und weisen eine geringere Inferenzlatenz auf. Zudem bleiben charakteristische Merkmale der Ausgangssprache, etwa Stimme, Emotionen und Prosodie, erhalten und werden in die synthetisierte Zielsprachausgabe übertragen. Auch Wörter, die keiner Übersetzung bedürfen, etwa Personennamen, können aufgrund der Konditionierung des Modells auf die Ausgangssprache natürlich in die Übersetzung der Zielsprache integriert werden, da die Aussprache der Personennamen einfach nachgebildet werden kann (Jia et al. 2019, S. 1). Ein Kaskadenmodell hingegen könnte Schwierigkeiten haben, den Personennamen als solchen zu erkennen, ihn falsch schreiben und infolgedessen eine fehlerhafte Sprachsynthese generieren.

Ein wesentlicher Nachteil direkter Übersetzungssysteme liegt in der begrenzten Verfügbarkeit geeigneter Trainingsdaten: Für ein vollständig überwachtes Ende-zu-Ende-Training werden große Mengen an Sprachein-/-ausgabepaaren benötigt. Entsprechende Datenkorpora sind jedoch nur eingeschränkt verfügbar. Im Gegensatz dazu sind Paralleltextpaare für maschinelle Übersetzung oder Sprach-Text-Paare für automatische Spracherkennung oder Sprachsynthese, wie sie für das Training von Kaskadenmodellen benötigt werden, deutlich leichter zugänglich (Jia et al. 2019, S. 1). Obwohl Translatotron eine gute Übersetzungsqualität erreicht, bleibt seine Leistung hinter der von Kaskadenmodellen zurück (Jia et al. 2019, S. 4).

KI-Modelle mit allgemeinem Verwendungszweck

KI-Modelle mit allgemeinem Verwendungszweck (General Purpose AI) wie SeamlessM4T von Meta stellen die nächste Entwicklungsstufe von Übersetzungssystemen von gesprochener Sprache in gesprochene Zielsprache dar.

Als KI-Modelle mit allgemeinem Verwendungszweck werden leistungsfähige KI-Modelle bezeichnet, die – anders als für einen spezifischen Anwendungsfall entwickelte Modelle – für vielfältige Anwendungen adaptiert werden können. Die Grundlage hierfür ist das Training mit sehr großen Datenmengen aus unterschiedlichen Quellen (Bommasani et al. 2021; Helmholtz Zentrum München o. J.). SeamlessM4T aus dem Jahr 2023 ist ein einheitliches, mehrsprachiges und multimodales Übersetzungssystem mit dem folgenden Leistungsspektrum:

- Übersetzung gesprochener Sprache in gesprochene Zielsprache: Übertragung von 101 Ausgangssprachen in 36 Zielsprachen
- gesprochene Sprache-zu-Text-Übersetzung: Übertragung von 101 Ausgangssprachen in 96 Zielsprachen
- Text-zu-Text-Übersetzung: Unterstützung von 96 Sprachen
- Text-zu-gesprochene-Sprache-Übersetzung: Übersetzung von 96 Ausgangssprachen in 36 Zielsprachen (Seamless Communication Team 2025, S. 587)

Während bisherige Systeme zur Übersetzung gesprochener Sprache in gesprochene Zielsprache auf ressourcenreiche Sprachen fokussieren und im Wesentlichen lediglich eine unidirektionale Übersetzung von einer Ausgangssprache ins Englische für wenige Sprachpaare unterstützen – jedoch nicht umgekehrt –, deckt SeamlessM4T die wichtigsten europäischen, asiatischen und afrikanischen Sprachen ab inklusive ressourcenarmer Sprachen, etwa Swahili und die südasiatischen Sprachen Bengalisch, Kannada, Telugu und Tamil (Seamless Communication et al. 2023, S. 14 f.; Seamless Communication

Team 2025, S. 587). Das Entwicklerteam von SeamlessM4T klassifizierte alle Sprachen mit weniger als 500 Stunden überwachter Trainingsdaten als ressourcenarm (Communication et al. 2023, S. 31). SeamlessM4T kann nichtenglische Sprachrichtungen übersetzen, ohne dass es mit konkreten Übersetzungsbeispielen für diese Sprachen gezielt trainiert worden ist. Auch die Übersetzung von Text in gesprochene Zielsprache kann SeamlessM4T ohne vorheriges Training für diesen spezifischen Anwendungsfall erfolgreich durchführen (Seamless Communication Team 2025, S. 587).

Die Grundlage von SeamlessM4T ist ein Datensatz von 470.000 Stunden gesprochenen Übersetzungen (SeamlessAlign) – das entspricht in etwa 53 Jahren. Zur Herstellung dieses vergleichsweise sehr umfangreichen Datensatzes setzten die Entwickler/-innen das Verfahren SONAR (Sentence-level Multimodal and Language-Agnostic Representations) ein. Mit diesem konnten Bedeutungsähnlichkeiten in unterschiedlichen Sprachen automatisch erkannt und einander zugeordnet werden. Da das Ergebnis automatischer Identifizierung und Zuordnung jedoch fehlerhaft sein kann, wurde anschließend eine Teilmenge des Datensatzes mit zusätzlichen Trainingsdaten verknüpft, deren Ergebnisse von Menschen erstellt oder geprüft bzw. von bereits vorhandenen KI-Modellen erzeugt wurden (Seamless Communication Team 2025, S. 587).

Bewertet nach BLEU (Bilingual Evaluation Understudy), einer Metrik zur automatischen Bewertung maschineller Übersetzungen, erreicht SeamlessM4T 8 % höhere Scores im Bereich Sprache-zu-Text-Übersetzung und 23 % höhere Scores bei Sprache-zu-Sprache-Übersetzung im Vergleich zu führenden Kaskadenmodellen (Seamless Communication Team 2025, S. 587). Überdies wurde die Robustheit von SeamlessM4T, also die Toleranz des Systems gegenüber Hintergrundgeräuschen und Sprecherunterschieden, wie Stimmlage, Akzent und Sprechgeschwindigkeit, mit dem großen mehrsprachigen Modell Whisper-Large-V2 von OpenAI, verglichen. Bezogen auf Hintergrundgeräusche schnitt SeamlessM4T 42 % robuster und bezogen auf Sprecherunterschiede 66 % robuster als Whisper-Large-V2 ab (Seamless Communication Team 2025, S. 591).

Hintergrundgeräusche erschweren die Erkennung und damit die Verarbeitung von Sprachsignalen. Bei Übersetzungssystemen, die zusätzlich Stimme, Prosodie und Emotionen übertragen sollen, besteht zusätzlich die Herausforderung, dass die Hintergrundgeräusche möglicherweise zu den Ausdrucksmerkmalen des Sprechenden gezählt und fälschlicherweise in der Zielsprache wiedergegeben werden (Hwang et al. 2024, S. 15524 f.).

Ein Problem bei KI-Übersetzungen ist hinzugefügte Toxizität, also dass Beleidigungen, Herabsetzungen, Sexismus, Rassismus in der Übersetzung vorkommen, die nicht in der Ausgangssprache vorhanden sind. Hinzugefügte Toxizität entsteht beispielsweise, wenn Trainingsdaten verzerrt sind, also neutrale Begriffe häufig im Kontext mit abwertenden Bezeichnungen auftreten oder ein neutraler Begriff ähnlich klingt wie ein abwertendes Wort, sodass die Spracherkennung inkorrekt ist.

Bei SeamlessM4T wurde versucht, durch spezielle Filter- und Trainingsmethoden hinzugefügte Toxizität zu reduzieren (Seamless Communication Team 2025, S. 591 f.).

Eine weitere Schwierigkeit bei maschineller Übersetzung ist Gender Bias. Gender Bias kann in zwei Formen auftreten:

- In der Zielsprache verwendet das System die männliche Form, obwohl in der Ausgangssprache keine Zuordnung von Geschlechtern erkennbar ist (Seamless Communication Team 2025, S. 592).

- Die Übersetzungsqualität hängt davon ab, welche Geschlechter im Satz erwähnt werden (Seamless Communication Team 2025, S. 592), also ob der Satz „Die Ärztin leistet gute Arbeit." oder der Satz „Der Arzt leistet gute Arbeit." übersetzt werden soll. Ursächlich für die mangelnde Übersetzungsqualität kann beispielsweise die Unterrepräsentation weiblicher oder nicht-binärer Formen in den Trainingsdaten sein oder die überwiegende Verknüpfung eines Begriffs mit einem bestimmten Geschlecht in den Trainingsdaten (zum Beispiel „Erzieherin" kommt häufiger vor als „Erzieher" oder „Arzt" häufiger als „Ärztin"). Zwar konnte bei -SeamlessM4T die Übersetzungsqualität im Vergleich zu anderen Modellen verbessert werden, nicht aber die Bevorzugung eines Geschlechts (Seamless Communication Team 2025, S. 592).

Neben Herausforderungen wie hinzugefügter Toxizität und Gender Bias wird die Leistungsfähigkeit maschineller Übersetzungssysteme auch von systembedingten Faktoren eingeschränkt. Bei KI-gestützten Simultanübersetzungssystemen besteht ein Trade-off zwischen Qualität und Latenz, da mit der Übersetzung begonnen werden muss, während gleichzeitig noch Input aufgenommen wird. Ein zu frühes Übersetzen kann zu Fehlern führen, wohingegen ein zu langes Abwarten die Latenz erhöht. Besonders herausfordernd kann hier die Verarbeitung langer Sprachströme sein (Liu et al. 2024, S. 1 f.).

Simultanübersetzungssysteme auf dem Markt

Am Markt sind zurzeit im Wesentlichen vier verschiedene Kategorien von Simultanübersetzungssystemen verfügbar: in Videokonferenzsysteme integrierte Lösungen, videokonferenzsystemübergreifende Anwendungen, Produkte für die Übersetzung von Online- und Präsenzveranstaltungen sowie Systeme für persönliche, informelle Gesprächssituationen.



- Ein Beispiel für eine in Videokonferenztools integrierte Lösung ist Google Meet Speech Translation. Die Funktion unterstützt ausschließlich Übersetzungen zwischen Englisch und Französisch, Deutsch, Hindi, Italienisch, Portugiesisch und Spanisch. Die Sprachausgabe erfolgt in einer synthetisierten Stimme, die der Stimme der sprechenden Person ähneln soll. Pro Meeting kann nur ein Sprachpaar übersetzt werden, also zum Beispiel Englisch–Französisch oder Spanisch–Englisch (Google o. J.). Die gleichzeitige Übersetzung von Französisch oder Spanisch sprechenden Teilnehmer/innen ins Englische ist bislang nicht möglich (Google o. J.), wird aktuell aber erprobt (Weerasinghe/Lu 2026). Zudem ist die Nutzung der Übersetzungsfunktion zurzeit auf 90 Minuten pro Meeting begrenzt. Google weist selbst auf einige Schwierigkeiten hin, die bei der Übersetzung auftreten können, unter anderem Grammatik- oder Übersetzungsfehler, eine fehlerhafte Wiedergabe des grammatischen Geschlechts von Substantiven sowie unerwartete Akzente in der Sprachausgabe (Google o. J.).

- Das 2017 in Köln gegründete Unternehmen DeepL bietet eine Desktop-App an, mit der sich Videokonferenzen in Microsoft Teams, Zoom Meetings und Google Meet übersetzen lassen. Sie soll unabhängig vom eingesetzten Videokonferenzsystem ein einheitliches Übersetzungserlebnis ermöglichen. Die App unterstützt die Übersetzung gesprochener Sprache sowohl in Text für mehr als 40 Sprachen als auch in gesprochene Zielsprache für 30 Sprachen. Die gleichzeitige Übersetzung unterschiedlicher Sprachen im selben Meeting ist möglich. Grundlage sind neue Sprachmodelle, die Merkmale der Stimme, den Tonfall, das Sprechtempo und den emotionalen Ausdruck der Sprecher/innen in die Übersetzung übertragen sollen. Dadurch soll die synthetische Sprachausgabe natürlicher klingen (DeepL o. J.a; Floemer 2026).

- KUDO bietet sowohl KI-basierte Sprachübersetzung als auch professionelles menschliches Dolmetschen für Präsenzveranstaltungen sowie Online- und Hybridformate an. Die KI-gestützte Übersetzung steht für mehr als 70 Sprachen zur Verfügung und wird als gesprochene Übersetzung und/oder übersetzte Liveuntertitel ausgegeben. Bei Präsentationen und vergleichbaren Eins-zu-vielen-Situationen, in denen eine Person zu einem mehrsprachigen Publikum spricht, bietet KUDO eine KI-gestützte Simultanübersetzung in nahezu Echtzeit an. In Gesprächssituationen mit mehreren Sprecher/innen mit unterschiedlichen Sprachen kommen dagegen ausschließlich übersetzte Liveuntertitel zum Einsatz, um Latenzen möglichst gering zu halten. Ob eine KI-gestützte Übersetzung oder menschliches Dolmetschen geeigneter sind, richtet sich nach dem Kommunikationsanlass, der Art und Dauer der Veranstaltung und den benötigten Sprachen (Kudo o. J.).

- Mit Apples AirPods lassen sich Gespräche in ausgewählten Sprachen – Chinesisch, Deutsch, Englisch, Französisch, Italienisch, Japanisch, Koreanisch, Portugiesisch und Spanisch (Apple 2025) – live übersetzen. Voraussetzungen für die Nutzung sind kompatible AirPods (AirPods 4 mit aktiver Geräuschunterdrückung, AirPods Pro 2 und 3) sowie ein Apple-Intelligence-fähiges iPhone mit dem Betriebssystem iOS 26.2 oder

neuer. Vor Beginn des Gesprächs müssen Ausgangs- und Zielsprache ausgewählt und die entsprechenden Sprachmodelle auf das iPhone geladen werden. Die Übersetzung wird über die AirPods mit einer synthetisierten Stimme ausgegeben. Wechseln Gesprächspartner/innen unvermittelt die Sprache oder kommen Personen hinzu, die eine andere als die zuvor ausgewählte Sprache sprechen, kann das System hierauf nicht flexibel reagieren (von Thun 2025).

Für Teilnehmer/innen einer internationalen Konferenz, die sich etwa beim Kaffeetrinken miteinander austauschen möchten, sind die bislang verfügbaren Produkte allenfalls ein Hilfsmittel. Insbesondere ressourcenarme Sprachen, Gespräche mit Personen, die unterschiedliche Sprachen sprechen, Hintergrundgeräusche sowie hinzugefügte Toxizität und geschlechtsspezifische Verzerrungen können einer reibungslosen Konversation im Weg stehen.

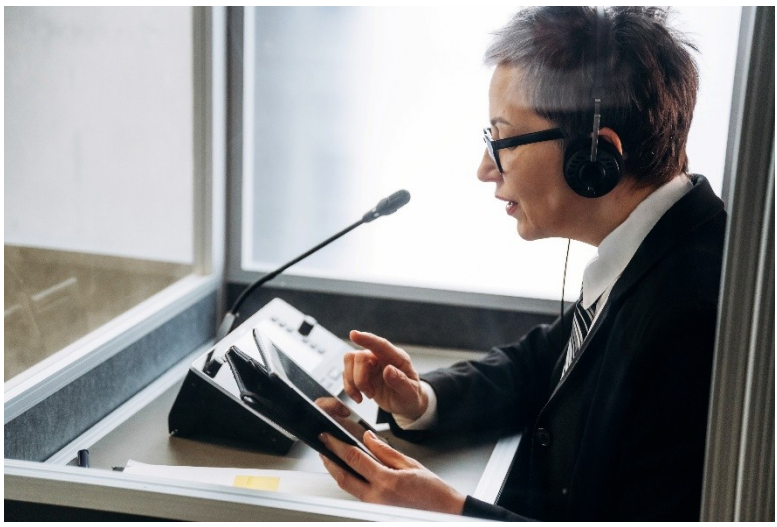
Entwicklungstrends

Aktuelle Entwicklungstrends knüpfen an diesen Grenzen an und adressieren beispielsweise die Verbesserung von Spracherkennung und des Kontext- und Bedeutungsverständnisses maschineller Systeme sowie die effiziente Verbindung maschineller Übersetzung mit menschlichem Dolmetschen.

In einem Forschungsstrang werden beispielsweise die Erfassung und die Verarbeitung von visuellen Signalen wie Lippenbewegungen zusätzlich zu Sprachsignalen verfolgt. Diese Herangehensweise könnte vor allem die automatische Spracherkennung in lauten Umgebungen erleichtern (Hu et al. 2023, S. 15213 f.). In einem weiteren Forschungsstrang geht es darum, dass maschinelle Übersetzungen in Zukunft nicht nur Wörter und Satzstrukturen korrekt wiedergeben können, sondern eine tiefere inhaltliche Übersetzung ermöglichen, also etwa kommunikative Absichten mit übertragen (Ward et al. o. J.).

Künftig dürfte sich der Trend zunehmend in Richtung einer aufgabenspezifischen Arbeitsteilung und Kooperation zwischen Mensch und Maschine entwickeln, um die jeweiligen Stärken beider Seiten zu kombinieren und die jeweiligen Schwächen möglichst auszugleichen.

KI-Systeme zeichnen sich insbesondere durch eine hohe Verfügbarkeit, Skalierbarkeit und Kosteneffizienz aus. Zudem können sie mehrere Sprachen gleichzeitig abdecken, wohingegen menschliche Übersetzer/innen und Dolmetscher/innen zumeist auf wenige Sprachkombinationen spezialisiert sind. Demgegenüber fehlen den KI-Systemen bislang ein tiefes kontextuelles Sprachverständnis sowie die Sensibilität für kulturelle Besonderheiten. Sie haben Schwierigkeiten beim Verstehen von Ironie und können Emotionen nicht interpretieren. Während eine KI beim Simultandolmetschen dichte Textpas-



sagen Wort für Wort übersetzt, kann ein Mensch sinnvoll komprimieren (Dolmetschwerk.com o. J.).

Ein weiterer Vorteil bezogen auf das Simultandolmetschen besteht darin, dass menschliche Simultandolmetscher/innen transparent damit umgehen, wenn sie sich unsicher sind, ob ihre Verdolmetschung korrekt ist. Im Unterschied dazu liefern KI-Systeme stets eine scheinbar eindeutige Übersetzung ab, auch wenn diese inhaltlich fehlerhaft oder unsicher ist (Della Torre 2025).

Darüber hinaus unterliegen professionelle Dolmetscher/innen Vertraulichkeits- und Verschwiegenheitsverpflichtungen, wohingegen bei KI-Systemen Datenschutz- und Vertraulichkeitsrisiken bestehen können (Dolmetschwerk.com o. J.). In sensiblen Bereichen, etwa der Diplomatie, könnte daher weiterhin menschlichen Übersetzer/innen und Dolmetscher/innen der Vorzug gegeben werden.

Gleichzeitig können KI-Systeme Simultandolmetscher/innen unterstützen, indem sie beispielsweise Zahlen erfassen und übertragen oder passende Fachtermini aus hinterlegten Glossaren vorschlagen und auf Bildschirmen in den Dolmetscherkabinen anzeigen. Die Simultandolmetscher/innen können diese Vorschläge dann unmittelbar in ihre Verdolmetschung einbeziehen. Besonders bei Zahlen ist eine solche Unterstützung sinnvoll, da deren korrekte Verdolmetschung unter Zeitdruck anspruchsvoll und fehleranfällig ist (Brockhues 2025).

Gesellschaftliche und politische Relevanz

Die Schätzungen zum aktuellen und zukünftigen Marktvolumen für die Übersetzung gesprochener Sprache in gesprochene Zielsprache zeigen sehr unterschiedliche Ergebnisse. Dies könnte unter anderem darauf zurückzuführen sein, dass den Marktanalysen unterschiedliche Definitionen des Analyse-

gegenstands zugrunde liegen und/oder verschiedene Prognosemodelle verwendet worden sind. Definitionen und Prognosemodelle werden häufig nicht explizit oder nur unzureichend offengelegt. Beispielsweise sah ResearchAndMarkets.com den Markt für Simultanübersetzungstools im Jahr 2024 bei einem Volumen von 524,3 Mio. US-Dollar. Für die kommenden Jahre wird ein durchschnittliches jährliches Wachstum von 7,4 % erwartet, sodass das Marktvolumen im Jahr 2030 auf 803,4 Mio. US-Dollar nach Einschätzung von ResearchAndMarkets.com anwachsen könnte. Als zentrale Treiber dieser Entwicklung sieht ResearchAndMarkets.com die zunehmende technologische Reife der angebotenen Systeme, die wachsende Internationalisierung (Reisen, Migration, globale Geschäftstätigkeiten) sowie eine stärkere Fokussierung auf Inklusion und Zugänglichkeit, zum Beispiel im Gesundheits- und Bildungswesen (ResearchAndMarkets.com 2025). Mordor Intelligence (o. J.) berechnet für das Jahr 2025 ein Marktvolumen von 690 Mio. US-Dollar und prognostiziert 762 Mio. US-Dollar für das Jahr 2026. Im Vergleich zu ResearchAndMarkets.com erwartet es zwischen 2026 und 2031 eine höhere jährliche Wachstumsrate von 10,4 %. Demgemäß würde das Marktvolumen im Jahr 2031 1,25 Mrd. US-Dollar umfassen (Mordor Intelligence o. J.). Mordor Intelligence hat sechs zentrale Treiber und vier wichtige Hemmnisse für das Marktwachstum identifiziert (Tabelle).

Tabelle Treiber und Hemmnisse

Treiber	Hemmnisse
Verbreitung intelligenter Sprachassistenten	Dialekt- und Code-Switching-Genauigkeitslücken
Wachstum des internationalen Tourismus und des grenzüberschreitenden E-Commerce	Datenschutz- und Datensicherheitsbedenken
5G-Rollouts ermöglichen Low-Latency-Cloud-Inferenz	hohe Kosten für Echtzeitübersetzungshardware auf dem Gerät
Big-Tech-Investitionen in Sprach-KI-Ökosysteme	soziokultureller Widerstand gegen maschinell vermittelten Dialog
industrielle IoT-Nachfrage nach mehrsprachiger Sprachsteuerung	
Telemedizineinführung für Minderheitensprachen	

Quelle: Mordor Intelligence o. J.

Der Markt wird als mäßig konzentriert bewertet mit Branchenführern aus den USA und China. 48 % des Umsatzes im Jahr 2024 entfielen auf Microsoft, Google, Amazon, Meta und Baidu (Mordor Intelligence o. J.).

In gesellschaftlicher Hinsicht können KI-gestützte Übersetzungstools dazu beitragen, dass Menschen, insbesondere Mitarbeiter/innen von Unternehmen, in globalen und internationalen Kontexten leichter in ihrer jeweiligen Muttersprache miteinander interagieren. Insbesondere könnten sie die World Readiness von Personen erhöhen, denen die Ressourcen für den Erwerb einer Fremdsprache fehlen. World Readiness beschreibt die Fähigkeit, aus instrumentellen (Schule, Ausbildung und Beruf) und kulturellen Gründen (Verständigung und Handlungsfähigkeit in einer globalisierten Welt) in einer Fremdsprache zu kommunizieren (Seamless Communication Team 2025, S. 592). Anbieter wie DeepL (o. J.b) verweisen darauf, dass ihre Tools dazu beitragen können, Mitarbeiter/innen stärker auf Basis ihres Fachwissens und weniger aufgrund vorhandener Sprachkenntnisse einstellen zu können. In der Folge könnte die Relevanz von Fremdsprachenkenntnissen in künftigen Qualifikationsprofilen von Arbeitssuchenden und Beschäftigten abnehmen. Zugleich könnten Benachteiligungen von Personen reduziert werden, die nicht mit vergleichsweise weit verbreiteten Weltsprachen wie Englisch oder Spanisch aufgewachsen sind und bislang in internationalen Kontexten häufig auf eine dieser Sprachen ausweichen mussten.

Multimodale, sprachgestützte Systeme wie SeamlessM4T können darüber hinaus die Teilhabe von Menschen erhöhen, die nicht lesen und/oder schreiben können, sowie von Menschen mit Sehbeeinträchtigung bzw. blinden Personen (Seamless Communication Team 2025, S. 592).

Zu berücksichtigen ist jedoch, dass Spracherwerb zumeist nicht ausschließlich einer auf Verwertung ausgerichteten Logik folgt. Vielmehr leistet er auch einen wichtigen Beitrag zur interkulturellen Verständigung, da Sprachschüler/innen nicht nur Vokabeln und Grammatik erlernen, sondern auch Einblicke in die Geschichte, die Gesellschaft und die Kultur von Ländern erhalten, in denen die jeweilige Fremdsprache gesprochen wird.

Wie zuvor bereits angedeutet, können KI-gestützte Übersetzungstools auch das Berufsbild von Übersetzer/innen und Simultandolmetscher/innen verändern.

Grundsätzlich könnte KI-gestützte Simultanübersetzung mit zunehmender technischer Reife überall dort eingesetzt werden, wo Menschen mit in einer fremden Sprache gesprochenen Inhalten konfrontiert sind: im Alltag, im Beruf ebenso wie auf Reisen, bei einem Theaterbesuch im Ausland oder im Kino. Bislang werden Übersetzungen nicht immer in der eigenen Muttersprache angeboten: Filme werden lediglich in ausgewählten Sprachen synchronisiert und bisweilen nur untertitelt. Selbst in internationalen Gremien wird nicht immer eine Verdolmetschung in alle vertretenen Sprachen angeboten. Mehrsprachi-

ge Übersetzungssysteme könnten hier Abhilfe schaffen. Welche längerfristigen Auswirkungen eine breite Nutzung solcher Systeme auf Kommunikation, Sprachkompetenzen und interkulturelle Verständigung hat, ist allerdings unklar, da hierzu kaum Studien existieren und entsprechend keine gesicherten Erkenntnisse vorliegen.

Mit der Übersetzung und Verarbeitung gesprochener Sprache sowie der Nachbildung individueller Stimmmerkmale werden auch Fragen des Datenschutzes und des Schutzes der Vertraulichkeit des gesprochenen Wortes aufgeworfen. Relevant sind in diesem Zusammenhang vor allem folgende rechtliche Vorgaben:

- Artikel 9 Datenschutz-Grundverordnung verbietet die Verarbeitung biometrischer Daten zur eindeutigen Identifizierung einer natürlichen Person.
- § 201 Abs. 1 Nr. 1 Strafgesetzbuch stellt das Aufnehmen des nicht öffentlich gesprochenen Wortes auf einen Tonträger unter Strafe.
- § 8 Telekommunikation-Digitale-Dienste-Datenschutz-Gesetz untersagt die Herstellung und die Bereitstellung auf dem Markt von Geräten, mit denen unbemerkt Aufnahmen gemacht werden können, weil sie wie Alltagsgegenstände aussehen.

Erfolgt die Aufnahme bzw. Übersetzung nach wirksamer Einwilligung, kann eine Grundlage für einen rechtmäßigen Einsatz geschaffen sein. Problematisch stellt sich die Situation dar, wenn Äußerungen einer Person ohne deren explizites Einverständnis beispielsweise über Kopfhörer übersetzt werden. Entscheidend für die Zulässigkeit in diesem Fall dürfte sein, ob bei dem technischen Design des verwendeten Übersetzungssystems tatsächlich von einer Aufnahme gemäß dem Wortlaut des Gesetzes gesprochen werden kann oder ob lediglich ein technisches Zwischenspeichern im Arbeitsspeicher vorliegt. Ohne eine nähere Betrachtung der Technik lässt sich dies nicht pauschal beurteilen (Fuchs 2026; Klaas/Basar 2025).

Mögliche vertiefte Bearbeitung des Themas

Dieser Überblick über Entwicklungen im Bereich KI-gestützter Simultanübersetzung zeigt, dass erste Produkte bereits auf dem Markt sind, technisch aber noch eine ganze Reihe von Herausforderungen besteht, unter anderem bezogen auf den Umgang mit Hintergrundgeräuschen, den Trade-off zwischen Qualität und Latenz, hinzugefügte Toxizität, Gender Bias und die Sprachabdeckung. Zudem besteht Forschungsbedarf hinsichtlich der Messung der Übersetzungsqualität. Die gemeinhin verwendete Metrik zur automatischen Bewertung maschineller Übersetzungen BLEU steht in der Kritik (Fantinuoli/Prandi 2021). Eine verbesserte Qualitätsmessung dürfte dazu beitragen, die Entwicklung besserer Übersetzungssysteme voranzutreiben. Für internationale Gremien und Parlamente dürfte perspektivisch insbesondere von Interesse sein zu beobachten, ob und in welcher Weise sich KI-gestützte Simultanübersetzung auf das Angebot an gedolmetschten Sprachen und die Qualität der Übersetzung auswirken wird. Ebenso ist für diese Institutionen von Belang, wie Datenschutz- und Vertraulichkeitsfragen technisch gelöst werden können. Hier bestehen auch Optionen für eine Zusammenarbeit zwischen Innovatoren und staatlichen Stellen, die potenziell über einen großen Erfahrungs- und Datenschatz verfügen und durch eine Beteiligung auf eine verantwortungsvolle Entwicklung hinwirken könnten.

Mögliche Auswirkungen KI-gestützter Simultanübersetzungssysteme auf unterschiedliche Branchen, stark betroffene Berufsprofile – etwa Fremdsprachentrainer/innen und -lehrer/innen sowie Übersetzer/innen und Dolmetscher/innen –, Spracherwerb, interkulturelle Verständigung, Inklusion sowie Datenschutz- und Missbrauchsaspekte könnten perspektivisch vertiefend in einer Szenariostudie untersucht werden, wenn sich eine weitere Verbreitung dieser Systeme durch zunehmende Technologiereife abzeichnet.

Literatur

- Apple (2025): Live Übersetzung mit AirPods kommt in die EU. https://www.apple.com/de/newsroom/2025/11/live-translation-on-airpods-expands-to-the-eu/ (11.9.2026)
- Bommasani, R. et al. (2021): On the Opportunities and Risks of Foundation Models. Stanford
- Brockhues, A. (2025): Besser dolmetschen mithilfe von KI. Deutschlandfunk Nova, https://www.deutschlandfunknova.de/beitrag/kuenstliche-intelligenz-besser-dolmetschen-mit-hilfe-von-ki (15.9.2026)
- DeepL (o. J.a): Live translations for virtual meetings. https://www.deepl.com/en/products/voice/deepl-voice-for-meetings (13.10.2026)
- DeepL (o. J.b): Sprachübersetzung in Echtzeit. https://www.deepl.com/de/products/voice (29.6.2026)
- Della Torre, S. (2025): Unterhaltung in Fremdsprache: Was taugt die Echtzeit-Übersetzung? SRF, https://www.srf.ch/wissen/kuenstliche-intelligenz/ki-sprachmodelle-

unterhaltung-in-fremdsprache-was-taugt-die-echtzeit-uebersetzung (25.6.2026)

- Dolmetschwerk.com (o. J.): KI-Dolmetschen bei Live-Events: geht das? https://www.dolmetschwerk.com/ki-dolmetschen (25.6.2026)

- Fantinuoli, C.; Prandi, B. (2021): Towards the evaluation of automatic simultaneous speech translation from a communicative perspective. In: Federico, M. et al. (Hg.): Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021). Stroudsburg, S. 245–254

- Floemer, A. (2026): DeepL Voice: Echtzeit-Übersetzung mit originaler Stimmcharakteristik. Heise online, https://www.heise.de/news/DeepL-Voice-Neue-Sprachmodelle-bewahren-individuelle-Stimmen-bei-Uebersetzungen-11454033.html (13.10.2026)

- Fuchs, T. (2026): Endet die Privatsphäre im öffentlichen Raum? Debatte um Smart Glasses. Tagesspiegel Background, https://background.tagesspiegel.de/digitalisierung-und-ki/briefing/endet-die-privatsphaere-im-oeffentlichen-raum (13.10.2026)

- Gao, Y. et al. (2006): IBM MASTOR SYSTEM: Multilingual Automatic Speech-to-speech Translator. In: Bouillon, P. et al. (Hg.): Proceedings of the First International Workshop on Medical Speech Translation. New York. Association for Computational Linguistics. o. O., S. 53–56

- Google (o. J.): Gemini in Meet. Learn about Speech Translation. https://support.google.com/meet/answer/16221730?hl=en&co=GENIE.Platform%3DDesktop&oco=0 (15.9.2026)

- Helmholtz Zentrum München (o. J.): Foundation Models. https://www.helmholtz-munich.de/newsroom/entdecke-unsere-themen/foundation-models (15.9.2026)

- Hu, Y. et al. (2023): Hearing Lips in Noise: Universal Viseme-Phoneme Mapping and Transfer for Robust Audio-Visual Speech Recognition. In: Rogers, A. et al. (Hg.): Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Stroudsburg, S. 15213–15232

- Hwang, M.-J. et al. (2024): Textless Acoustic Model with Self-Supervised Distillation for Noise-Robust Expressive Speech-to-Speech Translation. In: Findings of the Association for Computational Linguistics, S. 15524–15541, https://doi.org/10.48550/arXiv.2406.02733

- IBM (o. J.): Machine-aided translation. A century-long quest to streamline communications across languages. https://www.ibm.com/history/machine-aided-translation (15.9.2026)

- Jia, Y. et al. (2019): Direct speech-to-speech translation with a sequence-to-sequence model. In: ISCA (Hg.): Interspeech 2019. International Speech Communication Association. o. O. S. 1123–1127, https://www.isca-archive.org/interspeech_2019/jia19_interspeech.pdf

- Klaas, A.; Basar, E. (2025): KI-Transkription und § 201 StGB. Wessing & Partner Rechtsanwälte mbB, https://www.unternehmensstrafrecht.de/ki-transkription-und-%C2%A7-201-stgb/ (13.10.2026)

- Kudo (o. J.): KUDO AI Speech Translator. https://kudo.ai/solutions/kudo-ai-speech-translator/ (15.9.2026)

- Lernende Systeme. Die Plattform für künstliche Intelligenz (o. J.): Verbmobil. Deutsche Akademie für Technikwissenschaften, https://www.plattform-lernende-systeme.de/erfolge-der-deutschen-ki-forschung-projekt/verbmobil.html (29.6.2026)

- Liu, X. et al. (2024): Recent Advances in End-to-End Simultaneous Speech Translation. In: IJCAI (Hg.): Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24). International Joint Conferences on Artificial Intelligence. o. O., S. 8142–8150

- Magueresse, A. et al. (2020): Low-resource Languages: A Review of Past Work and Future Challenges. o. O., https://doi.org/10.48550/arXiv.2006.07264

- Meyer, F.; Buys, J. (2026): The Geometry of Low-Resource Language Representations. o. O.

- Mordor Intelligence (o. J.): Speech To Speech Translation Market Size & Share Analysis - Growth Trends and Forecast (2026-2031). https://www.mordorintelligence.com/industry-reports/speech-to-speech-translation (29.6.2026)

- ResearchAndMarkets.com (2025): Speech-to-Speech (S2S) Translation Strategic Business Analysis Report 2025: Global Market to Surpass $800 Million by 2030 - Integration with Smart Devices and IoT Expands Opportunities. Business Wire, https://www.businesswire.com/news/home/20250324835335/en/Speech-to-Speech-S2S-Translation-Strategic-Business-Analysis-Report-2025-Global-Market-to-Surpass-%24800-Million-by-2030---Integration-with-Smart-Devices-and-IoT-Expands-Opportunities (22.6.2026)

- Sarim, M. et al. (2025): Direct Speech to Speech Translation: A Review. o. O. https://doi.org/10.48550/arXiv.2503.04799
- Seamless Communication et al. (2023): SeamlessM4T: Massively Multilingual & Multimodal Machine Translation. o. O. https://doi.org/10.48550/arXiv.2308.11596
- Seamless Communication Team (2025): Joint speech and text machine translation for up to 100 languages. In: Nature 637(8046), S. 587–593, https://doi.org/10.1038/s41586-024-08359-z
- Sperber, M.; Paulik, M. (2020): Speech Translation and the End-to-End Promise: Taking Stock of Where We Are. In: Jurafsky, D. et al. (Hg.): Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online. Stroudsburg, S. 7409–7421
- von Thun, F. (2025): Live-Übersetzung mit den AirPods: So funktioniert der Dolmetscher in der Praxis. CHIP, https://www.chip.de/news/apple/airpods-erhalten-endlich-live-uebersetzung-wie-funktioniert-sie-in-der-praxis_d1c02806-81b0-47fe-b787-57aae334a179.html (29.6.2026)
- Ward, N. G. et al. (o. J.): The Interspeech 2026 Challenge on Transfer of Pragmatic Intent in Speech-to-Speech Translation. o. O.
- Weerasinghe, A.; Lu, T. (2026): Fluid, natural voice translation with Gemini 3.5 Live Translate. News from Google, https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-live-3-5-translate/ (13.10.2026)

Herausgeber
Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag (TAB)

Bildnachweis
Supatman/iStock (S. 1); Ventz/iStock (S. 3);
imaginima/iStock (S. 5); -standret/iStock (S. 6)

KI-Hinweis
Zur Erstellung dieses Textes wurden AitisiGPT und ChatGPT Business eingesetzt. Die Unterstützung umfasste insbesondere die Recherche, die Erstellung von Textzusammenfassungen sowie die sprachliche und strukturelle Überarbeitung des Textentwurfs. Die Nutzung erfolgte im Rahmen der TAB-Leitlinien für generative KI. Die endgültige Verantwortung für die inhaltliche Richtigkeit, die kritische Reflexion und die Interpretation der Ergebnisse liegt bei der Autorin.

ISSN: 2629-2874
DOI: 10.5445/IR/1000198767

Horizon SCANNING

Das Horizon-Scanning ist Teil der Foresight-Aktivitäten des TAB und wird vom Institut für Innovation und Technik (iit) in der VDI/VDE Innovation + Technik GmbH durchgeführt.
www.tab-beim-bundestag.de/horizon-scanning