

Einfach, weil's wichtig ist.

ERGO

Die nächste Generation Generativer Künstlicher Intelligenz

Fortschritte und multimodale Anwendungen
in der Versicherungsbranche



„Generative AI has the potential to change the world in ways that we can't even imagine. It has the power to create new ideas, products, and services that will make our lives easier, more productive, and more creative.“

Bill Gates, Mitbegründer von Microsoft

Im ERGO Innovation Lab befassen wir uns täglich damit, neue Technologien zu entdecken, ihr Potenzial zu bewerten und sie bei Eignung für die gesamte Gruppe nutzbar zu machen. Generative Künstliche Intelligenz (GenAI) hat sich dabei in kürzester Zeit weltweit als transformative Technologie etabliert. Was früher nur Experten vorbehalten war, ist heute durch rasante technologische Fortschritte für eine breite Öffentlichkeit zugänglich. Diese Entwicklung verändert grundlegend, wie wir mit Technologie arbeiten und eröffnet völlig neue Möglichkeiten, komplexe Systeme einfach und effektiv zu nutzen. Auch bei ERGO ist GenAI inzwischen ein fester Bestandteil unserer Arbeit. Erste Pilotprojekte, wie unser internes ERGO GPT, sind erfolgreich umgesetzt und befinden sich bereits in der Skalierungsphase – mittlerweile haben mehr als 30.000 unserer Mitarbeiterinnen und Mitarbeiter Zugriff auf diese Technologie.

Gleichzeitig hat sich die GenAI-Technologie rasant weiterentwickelt. Viele aktuelle Modelle sind nicht nur auf rein textuelle Ein- und Ausgabe beschränkt, sondern können inzwischen auch multimodal agieren. Das heißt, sie verarbeiten nicht nur Texte, sondern auch Bilder, Videos und andere Datenformate gleichzeitig und in atemberaubender Geschwindigkeit. Diese Fortschritte, die auf Veranstaltungen wie dem OpenAI Spring Update¹ und der Google I/O² im Mai dieses Jahres eindrucksvoll demonstriert wurden, eröffnen auch für die Versicherungsbranche neue Möglichkeiten. Im ERGO Innovation Lab verfolgen wir diese Entwicklungen kontinuierlich und haben bereits multimodale GenAI-Technologien für verschiedene Anwendungsfälle getestet. Dabei konnten wir in

kurzer Zeit wertvolle Erfahrungen sammeln, die wir nun aktiv in die ERGO Gruppe integrieren.

Die Anwendungsfälle für Versicherungsunternehmen sind vielfältig: Dazu gehört zum Beispiel die Verbesserung des Kundenerlebnisses durch eine schnelle und nutzerfreundliche Schadensregulierung, basierend auf der Analyse von Fotos oder Videos von Schadensfällen. Darüber hinaus führt die neue Technologie zu einer Optimierung interner Prozesse. Konkret bedeutet dies, dass KI nicht nur Briefe oder E-Mails automatisch vorschreibt, sondern auch Daten aus verschiedenen Quellen analysiert, Formulare vorausfüllt und Kundenanfragen schneller bearbeitet, um Routineaufgaben zu vereinfachen und den Arbeitsaufwand zu reduzieren.

Darüber hinaus verbessert die Technologie die Risiko einschätzung, indem sie beispielsweise frühzeitig potenzielle Gefahren in Wetterdaten erkennt und so die Entscheidungsfindung maßgeblich unterstützt. Auch strategische Entscheidungen profitieren von multimodaler GenAI, beispielsweise durch schnellere und fundiertere Marktanalysen zur Identifikation neuer Geschäftsfelder.

In unserem ersten Whitepaper „ChatGPT und Sprachmodelle: Eine Einführung mit Blick auf die Versicherungsbranche“ (Mai 2023)³ haben wir die grundlegenden Anwendungsmöglichkeiten unimodaler, textbasierter Sprachmodelle untersucht. Dieses Whitepaper befasst sich nun mit den Fähigkeiten und Funktionalitäten der neuesten Generation von GenAI-Modellen mit besonderem Fokus auf Multimodalität. Wir demonstrieren anhand von Beispielen, wie diese Technologie im Versicherungskontext und darüber hinaus eingesetzt werden kann und welche Chancen und Herausforderungen damit verbunden sind.

Inhaltsverzeichnis

Die nächste Evolutionsstufe Künstlicher Intelligenz	4
Ein Einblick in die Funktionsweise multimodaler Generativer KI	7
Einsatzmöglichkeiten multimodaler GenAI-Modelle in der Versicherungsbranche	10
Aktuelle Herausforderungen und Grenzen neuer Generativer KI-Modelle	19
Ausblick: Die Versicherungsbranche muss aktiv werden	21
Quellenverzeichnis	24

Die nächste Evolutionsstufe Künstlicher Intelligenz

Eine neue Generation Generativer AI-Modelle (GenAI-Modelle) steht bereit, um unsere Welt grundlegend zu verändern. Diese Modelle überzeugen nicht nur durch höhere Geschwindigkeit und Leistungsfähigkeit, sondern auch durch die Fähigkeit zur parallelen Verarbeitung mehrerer Modalitäten, darunter Texte, Bilder, Audio und Video. Das vorliegende Kapitel beleuchtet diese und weitere herausragende Eigenschaften aktueller GenAI-Modelle. Dazu gehören außerdem umfassende Kontextverarbeitung, und ein verbessertes Sprachverständnis sowie wesentliche Neuerungen der einzelnen Modalitäten.

Ein einheitliches Modell für nahtlose, multimodale Interaktionen

Aktuelle GenAI-Modelle sind in der Lage, mehrere Eingabemodalitäten wie Text, Bild, Audio und Video in einem einzigen System zu verarbeiten – eine Fähigkeit, die bisher nicht möglich war^a. Sie ermöglicht eine Herstellung von Zusammenhängen zwischen den unterschiedlichen Modalitäten, was zu präziseren und kontextreicheren Ergebnissen führt. Im Gegensatz zu unimodalen Systemen, die für jede Modalität separate Modelle benötigten, minimiert die Multimodalität Informationsverluste und vereinfacht die Interaktion. Diese Entwicklung ermöglicht neue Anwendungsfelder, da die Modelle in der Lage sind, komplexere Daten simultan zu interpretieren.

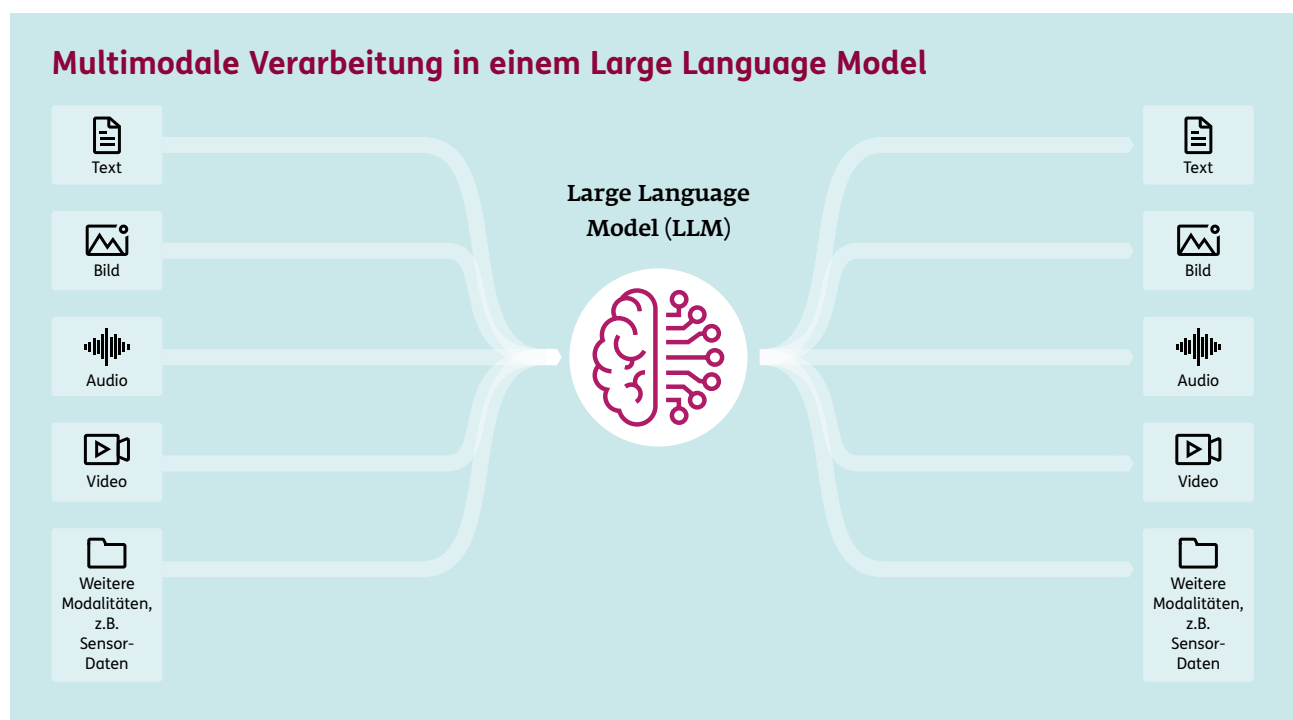
Daneben wurden auch Fortschritte innerhalb der Verarbeitung der spezifischen Modalitäten erzielt, welche im Folgenden kurz beleuchtet werden.

Entwicklungen im Bereich der verschiedenen Modalitäten

Die neuesten GenAI-Modelle haben sich auch bei der **Textverarbeitung** deutlich weiterentwickelt. Sie sind nicht mehr nur in der Lage, einfachen Fließtext zu verstehen, sondern können auch komplexe Textformate wie Tabellen, umfangreiche Dokumente (z. B. PDFs, Word-Dateien oder Präsentationen), Formulare, HTML und andere Webinhalte sowie mathematische und wissenschaftliche Formeln verarbeiten, und sind daher besonders vielseitig in ihrer Anwendung.

Moderne Architekturen beinhalten zusätzlich Technologien wie Retrieval-Augmented Generation (RAG^a), das heißt, sie sind in der Lage, relevante Informationen in einer Datenbank oder im Internet zu suchen, die den Modellen relevante Informationen aus Datenbanken oder dem Internet für die Formulierung einer Antwort zur Verfügung stellen.

In der Verarbeitung von Audio-Daten haben GenAI-Modelle ihre Fähigkeiten in der Sprachverarbeitung und -syn-



^a RAG (Retrieval-Augmented Generation): Kombiniert Inhalte externer Quellen mit GenAI zur Erstellung kontextbasierter Ergebnisse.

these erheblich erweitert. Die „Barge-In“-Funktion erlaubt es Nutzerinnen und Nutzern nach wie vor, Sprachinteraktionen in Echtzeit anzupassen und zu unterbrechen, beispielsweise, um den Erzählstil oder die Stimme zu ändern. Diese Flexibilität trägt dazu bei, die Kommunikation insgesamt natürlicher und intuitiver zu gestalten und verbessert damit die Benutzererfahrung. Darüber hinaus unter-

„Multimodal AI systems have the potential to solve complex problems that have so far been beyond our reach.“

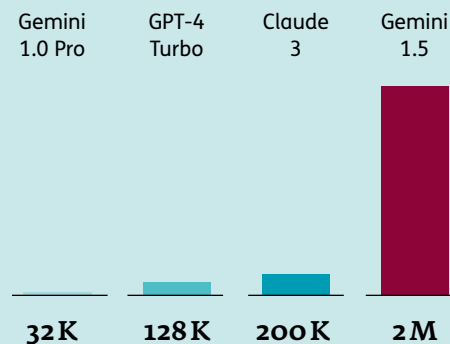
Demis Hassabis, CEO of DeepMind

stützen neue Architekturen Mehrsprachigkeit, sodass während eines Dialogs nahtlos zwischen verschiedenen Sprachen gewechselt werden kann, um eine breite Zielgruppe zu erreichen und sprachliche Barrieren zu überwinden. Durch die präzise Steuerung von Audioeigenschaften wie Tonhöhe, Tempo und Emotion können Nutzer die gewünschte Atmosphäre, Stimmung oder Emotion erzeugen, was die Anpassung an verschiedene Kontexte und Anwendungen erleichtert.

Auch in der **Bildverarbeitung** eröffnen moderne GenAI-Modelle neue Möglichkeiten. Die Analyse und Generierung von visuellen Inhalten kann nun in einem Modell erfolgen, sodass beispielsweise Text direkt in Bilder integriert werden kann. Ebenso können auch automatisch erstellte Untertitel oder Beschriftungen generiert werden. Des Weiteren wurde die Präzision der Bildgenerierung weiter optimiert. Nutzerinnen und Nutzer können detaillierte Anweisungen zur Bildgestaltung geben, beispielsweise hinsichtlich Farbwahl, Komposition und Lichtverhältnissen. Zudem bieten viele GenAI-Modelle Nachbearbeitungswerkzeuge, mit denen sich die generierten Bilder weiter anpassen lassen. Dies ist überall dort von Vorteil, wo realistische Produktbilder oder ansprechende Visualisierungen eine wichtige Rolle spielen.

Im Bereich **Video** haben sich die Qualität und der Realismus von KI-generierten Inhalten signifikant verbessert. Moderne GenAI-Modelle sind in der Lage, längere und komplexere Videoformate mit flüssigen Übergängen und hoher visueller Qualität zu erstellen. Ähnlich wie bei der Bildgenerierung, können Inhalte bei der Videogenerierung bereits sehr präzise erzeugt werden. Dazu können Nutzerinnen und Nutzer spezifische Parameter wie Kamerawinkel, Lichtverhältnisse oder Effekte vorab genau definieren. Damit eröffnen sich neue Möglichkeiten in Medienproduktion, Werbung, in der Kundenkommunikation und überall dort, wo visuelle Inhalte zentral sind.

Kontextfenstergröße verschiedener GenAI-Modelle (in Token)⁷



Die Größe von Kontextfenstern wird in der Regel in Token angegeben. Ein Token stellt eine Texteinheit dar, beispielsweise ein Wort oder ein Zeichen. Aktuelle GenAI-Modelle wie Google Gemini 1.5 Pro unterstützen erweiterte Kontextfenster bis zu einem Umfang von zwei Millionen Token⁵, während GPT-4o 128.000 Token⁶ verarbeiten kann. Dies entspricht entweder zwei Stunden Videomaterial, 22 Stunden Audioaufnahmen oder bis zu 1,4 Millionen Wörtern – mehr als alle sieben Harry Potter Bände zusammen. Diese umfangreichen Kontextfenster ermöglichen die simultane Verarbeitung und Analyse umfangreicher Inhalte, einschließlich langer Dokumente, Videos und Audiosequenzen.

Multimodale GenAI-Modelle können aber nicht nur Text-, Bild-, Video- und Audiodaten verarbeiten, sondern noch weitere, teilweise weniger verbreitete Formate. Dazu zählen beispielsweise **Sensor- und Bewegungsdaten**, welche insbesondere für die Versicherungsbranche in der Schadenprävention von Bedeutung sind, beispielsweise in der Detektion von defekten Wasserleitungen. Hierbei handelt es sich jedoch um Spezialfälle, weshalb sie in diesem Whiptaper nicht näher betrachtet werden.

Erweitertes Kontextfenster

Der Kontext umfasst alle Informationen, die einer KI zur Generierung von Antworten zur Verfügung stehen. Dazu gehören unter anderem Benutzerinput, zusätzliche Kontextinformationen sowie vorherige Dialoge. Die Größe des Kontextfensters hat somit einen unmittelbaren Einfluss auf die Leistungsfähigkeit eines GenAI-Modells. Ein größeres Kontextfenster verbessert die Kohärenz und Relevanz der Antworten, da zur Generierung des Outputs mehr zusätzliches Wissen sowie Details aus früheren Interaktionen herangezogen werden können.

Diese umfangreichen Kontextfenster ermöglichen die simultane Verarbeitung und Analyse umfangreicher Inhalte, einschließlich langer Dokumente, Videos und Audiosequenzen.

Die erweiterte Speicherfähigkeit ermöglicht es GenAI-Modellen zudem, sich an frühere Gesprächsinhalte erinnern

„The leap forward with multimodal language models is akin to teaching machines not just to read and write, but to see and hear as well. This multi-sensory approach enables a richer interaction with technology, paving the way for applications that were previously unimaginable, from advanced customer service bots to sophisticated diagnostic tools in medicine.“

Andrew Ng, Founder of Coursera and deeplearning.ai

und Antworten individuell an Nutzervorlieben anzupassen. Dadurch wird eine konsistente Kommunikation sowie eine bessere Personalisierung gewährleistet, da die KI sich kontinuierlich an individuellen Nutzercharakteristika anpassen kann. Dies ermöglicht eine effiziente, menschliche und einfühlsame Interaktion.

Menschliche Reaktionszeit und erweiterte Interaktivität

Zu den bemerkenswertesten Eigenschaften moderner multimodaler GenAI-Modelle zählt ihre Fähigkeit, nahezu in Echtzeit auf Nutzereingaben zu reagieren. Die durchschnittliche Latenzzeit beträgt etwa 232 Millisekunden, somit reagieren diese Modelle schneller als ein Mensch⁸. Dies ermöglicht eine flüssige und natürliche Kommunikation mit der KI und verbessert die Benutzererfahrung erheblich, insbesondere in längeren, interaktiven Dialogen.

Eine weitere Funktion, die von der schnellen Reaktionszeit profitiert, ist die „Barge-In“-Funktion. Wie bereits im Abschnitt zur Audiomodalität beschrieben, ermöglicht diese Funktion den Nutzern, laufende Sprachinteraktionen flexibel zu unterbrechen und anzupassen. Diese Anpassungsfähigkeit ist insbesondere in Kontexten von Bedeu-

tung, in denen eine nahtlose Integration menschlicher und künstlicher Intelligenz erforderlich ist, wie in der Kundenbetreuung. Neben der Verarbeitungszeit des GenAI Models spielt hier insbesondere auch die Latenz durch Validierung der erzeugten Inhalte eine wichtige Rolle.

Verbessertes Sprachverständnis und Mehrsprachigkeit

Die Sprachmodelle der neuesten Generation überzeugen durch ein hohes Niveau im Sprachverständnis. In anerkannten Benchmarks wie beispielsweise dem Massive Multitask Language Understanding (MMLU) schneiden aktuellen GenAI Modelle sehr gut ab. Der MMLU misst die Fähigkeit, komplexe sprachliche Aufgaben zu bewältigen – von einfacher Textverarbeitung bis hin zu anspruchsvollen Verständnisfragen. Die hohen Punktzahlen verdeutlichen, dass aktuelle GenAI-Modelle in der Lage sind, eine Vielzahl sprachlicher Herausforderungen erfolgreich zu meistern.

Darüber hinaus beherrschen aktuelle Modelle eine Vielzahl an Sprachen und können nahtlos zwischen ihnen wechseln. Diese Eigenschaft macht sie besonders nützlich für Unternehmen, die in mehrsprachigen Märkten operieren⁸ und ermöglicht damit eine Kommunikation über Sprachgrenzen hinweg.

Ein Einblick in die Funktionsweise multimodaler Generativer KI

Um die Funktionsweise der multimodalen GenAI zu verstehen, ist es hilfreich, sich mit ihren grundlegenden Mechanismen vertraut zu machen. Auch wenn diese auf den ersten Blick komplex erscheinen, lassen sich die Grundprinzipien anhand eines einfachen Beispiels veranschaulichen.

Stellen wir uns eine Opernaufführung vor:

- **Bild:** Auf der Bühne bewegt sich die Sängerin in einem kunstvoll gestalteten Kostüm durch ein aufwendiges Bühnenbild, welches die Stimmung der Handlung widerspiegelt.
- **Ton:** Die Sängerin singt eine gefühlvolle Arie, begleitet vom Orchester, das die Dramatik der Szene verstärkt.
- **Text:** Untertitel, die den gesungenen Text übersetzen, damit das Publikum der Handlung besser folgen kann.

Jede dieser Modalitäten liefert nur einen Teil des Ganzen: Das Bild vermittelt visuelle Eindrücke, die Tonspur ergänzt die Handlung mit Musik und Gesang, während der Text den Inhalt für diejenigen verständlich macht, die die Sprache nicht sprechen. Ein multimodales KI-Modell verknüpft diese Elemente, indem es lernt, wie Bild, Ton und Text zusammenpassen, um eine vollständige Szene zu erzeugen. So kann die KI später ähnliche Inhalte generieren oder genau beschreiben, was auf der Bühne passiert.

Dieses Zusammenspiel der verschiedenen Modalitäten zeigt, wie wichtig es für die multimodale GenAI ist, diese Zusammenhänge zu verstehen. Nun stellt sich die Frage, wie multimodale GenAI genau funktioniert.

Wie funktioniert multimodale GenAI? Eine einfache Erklärung in fünf Schritten

Schritt 1: Datenerfassung und Vorverarbeitung

Der erste Schritt sind Datenerfassung und Vorverarbeitung. Hierbei sammelt die KI unterschiedliche Datenarten – Texte, Bilder, Audiodateien und Videos. In einem ersten Schritt werden diese Daten zunächst **bereinigt**, um unre-

levante oder störende Elemente vorab aus den Rohdaten zu entfernen. Bei Texten sind dies beispielsweise irrelevante Textteile, etwa automatische Signaturen am Ende von E-Mails.

Nun müssen die bereinigten Daten in ein einheitliches Format gebracht werden. Bei **Texten** bedeutet dies, dass sie in kleinere Einheiten, wie Sätze oder Wörter zerlegt werden. Man nennt dieses Verfahren Tokenisierung. Bei **Bildern** hingegen erfolgt unter anderem eine Skalierung auf eine einheitliche Größe oder eine Anpassung der Farbwerte, um Konsistenz in Helligkeit und Kontrast zu erreichen. **Audiodateien** werden in Frequenz- und Zeitabschnitte zerlegt, beispielsweise durch die Umwandlung in Spektrogramme^b, was in der späteren Verarbeitung die Erkennung von Mustern im Klang ermöglicht. Bei **Videos** werden zunächst visuelle und auditive Inhalte separiert, anschließend erfolgt eine Extraktion der Bilder in einzelne Frames. Für die Bilder erfolgt nun eine ähnliche Vorverarbeitung wie bei statischen Bildern, während die Audiospur wie bei reinen Audiodateien analysiert wird.

Schritt 2: Merkmalsextraktion und Datenrepräsentation

Nachdem die Daten vorverarbeitet sind, erfolgt die **Merkmalsextraktion**, bei der die KI wichtige Merkmale in den Daten identifiziert und in numerische Darstellungen, sogenannte **Embeddings**, umwandelt. Diese Embeddings sind eine Art modalitätsübergreifende, universelle Repräsentation, die es der KI ermöglicht, verschiedenen Datentypen zu verarbeiten und zu analysieren, um komplexe Zusammenhänge zu erkennen.

Für diesen Prozess gibt es für jede Modalität spezielle Techniken:

- Bei **Texten** bedeutet dies, dass die KI Schlüsselwörter und Bedeutungen mit Hilfe spezieller Verfahren wie **Word2Vec**^c identifiziert und in einem multidimensionalen Raum, bzw. Zahlenfolgen, übersetzt. Wörter mit ähnlichen Bedeutungen oder solche, die häufig zusammen vorkommen, liegen in diesem Raum näher beieinander. Zum Beispiel könnten 'Eis' und 'kalt' in ähnlichen Regionen dieses Raumes zu finden sein, weil sie oft in einem ähnlichen Kontext verwendet werden. Fortgeschrittene Modelle wie **BERT** oder **GPT**^d können

^b Spektrogramm: Visuelle Darstellung, die die Frequenzzusammensetzung eines Signals über die Zeit.

^c Word2Vec (Word2Vector): Technik, die Wörter in Zahlenfolgen umwandeln, um ihre Bedeutung und den Kontext zu erfassen.

darüber hinaus die Bedeutung der Wörter im Kontext eines ganzen Satzes oder Absatzes interpretieren und damit die Bedeutung von Wörtern im gesamten Textkontext verstehen, wodurch sie tiefergehende Textanalysen durchführen können.

- Bei **Bildern** werden Techniken wie **CNNs^e** verwendet, um visuelle Merkmale wie Formen oder Farben zu extrahieren. Zum Beispiel kann die KI in einem Bild die Formen und Farben eines Autos erkennen und diese visuellen Informationen in Zahlenfolgen umwandeln, die das Bild repräsentieren und für die KI verständlich sind.
- **Audiodaten** werden zunächst in ein visuelles Format namens Spektrogramm umgewandelt, welches die Frequenzen und Lautstärken über die Zeit zeigt. Techniken wie **MFCs^f (Mel-Frequenz-Cepstrum-Koeffizienten)** helfen der KI anschließend, wichtige akustische Merkmale wie Tonhöhe und Lautstärke zu erkennen und in Zahlenfolgen umzuwandeln. Ein Beispiel wäre die Analyse der Tonlage und Lautstärke eines Musikinstruments, um seine Klangmuster zu identifizieren.
- Bei **Videos** kombiniert die KI Bild- und Audioinformationen und nutzt Technologien wie **Recurrent Neural Networks (RNNs^g)**, um zeitliche Abfolgen zu analysie-

ren, wie beispielsweise bei der Analyse von Gesten und Sprache in einem Videokonferenzgespräch.

Schritt 3: Multimodale Fusion und Modellverarbeitung

Nachdem die KI die relevanten Merkmale aus den verschiedenen Datentypen extrahiert und diese jeweils in eine universelle Repräsentation übersetzt hat, werden die Daten nun in einem Verfahren, der multimodalen Fusion, zusammengeführt. Diese Fusion kann auf drei verschiedenen Stufen erfolgen:

- **Frühe Fusion:** Hierbei werden alle Daten direkt nach der Erfassung zusammengeführt und gemeinsam verarbeitet. Diese Methode eignet sich besonders für Echtzeitanwendungen wie Sicherheitsüberwachung, bei denen eine Echtzeitverarbeitung entscheidend ist.
- **Intermediäre Fusion:** Die Fusion findet während des Verarbeitungsprozesses statt, nachdem die Daten teilweise analysiert wurden. Diese Methode wird beispielsweise in Kommunikationsdiensten verwendet, um subtile Hinweise wie Stimmungen oder Absichten aus der gleichzeitigen Analyse von Text, Ton und visuellen Hinweisen zu extrahieren.

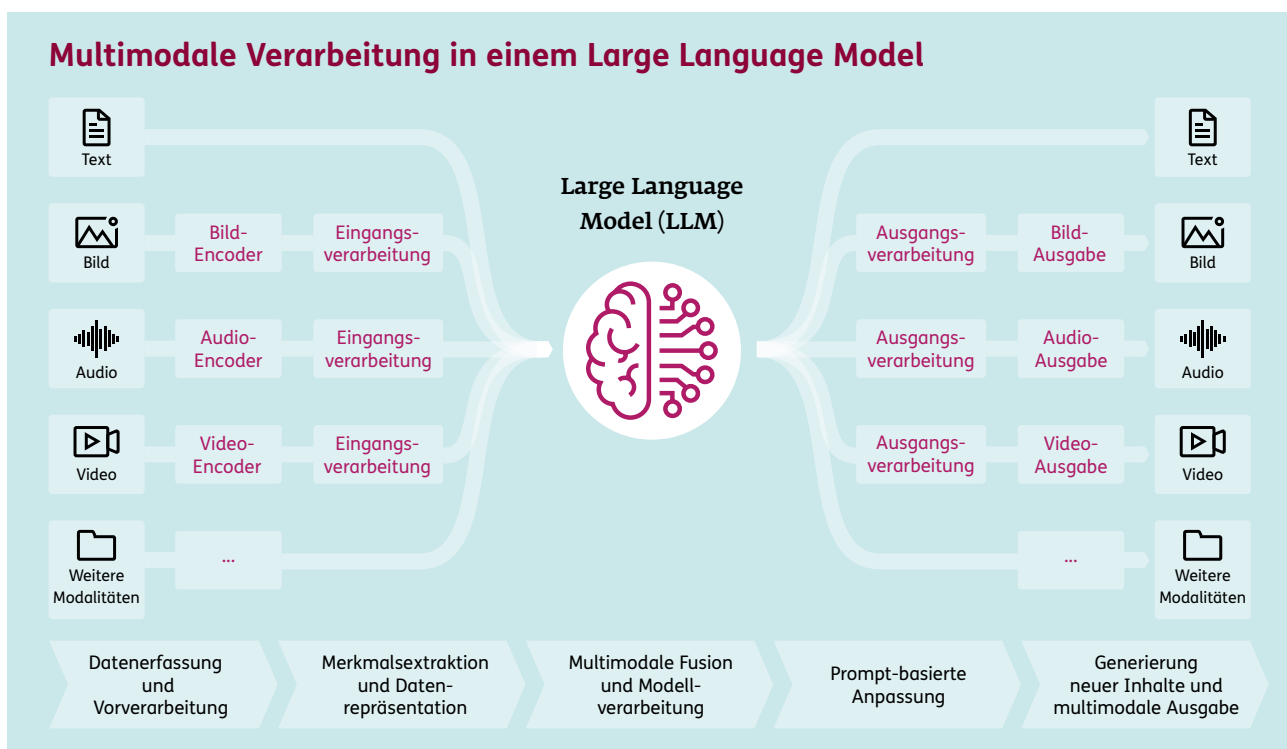


Abbildung: Verarbeitung verschiedener Formate durch ein Large Language Model (LLM)⁹ (modifiziert)

^d GPT (Generative Pre-trained Transformer): Fortgeschrittenes Sprachmodell, das menschenähnlichen Text generiert, indem es den Kontext und die Bedeutung von Wörtern innerhalb von Sätzen und Absätzen versteht.

^e CNN (Convolutional Neural Network): Ein spezieller Typ von künstlichem neuronalen Netz, der auf die Verarbeitung von Bilddaten spezialisiert ist.

^f MFCC (Mel-Frequenz-Cepstrum-Koeffizienten): Eine Technik zur Analyse und Darstellung von Audiofrequenzen.

^g RNN (Recurrent Neural Network): Ein neuronales Netzwerk zur Verarbeitung von sequenziellen Daten durch Nutzung vorheriger Informationen, ideal für Text-, Audio- und Videoanalysen.

- **Späte Fusion:** Jede Datenart wird zunächst individuell analysiert und die Ergebnisse werden am Ende zusammengeführt. Diese Methode wird häufig in der wissenschaftlichen Forschung eingesetzt, wo eine detaillierte Analyse jeder Datenquelle erforderlich ist, um eine präzise und umfassende Auswertung zu gewährleisten.

Die Wahl der Fusionsmethode hängt vom Anwendungsfall ab und davon, ob der Fokus auf schneller Verarbeitung (frühe Fusion), tiefgehender Kontextanalyse (intermediäre Fusion) oder präziser Einzelanalyse (späte Fusion) liegt.

Darüber hinaus spielen Transformer-Modelle und Aufmerksamkeitsmechanismen eine zentrale Rolle. Sie ermöglichen es der KI, anschließend relevante Informationen, wie Schlüsselszenen oder Kernbotschaften, aus den zusammengeführten Datenquellen herauszufiltern und gezielt zu analysieren.

Schritt 4: Prompt-basierte Anpassung

An dieser Stelle erfolgt in der Datenverarbeitung die Feinabstimmung basierend auf spezifischen Nutzeranweisungen (Prompts). Aus diesen Prompts leitet die GenAI ab, welche Informationen besonders relevant sind und wie sie diese darstellen soll, beispielsweise

- **Anpassung des Inhaltsformats:** Je nach gewünschter Ausgabemodalität kann die KI Text, Bilder, Audio oder Video in einem bestimmten Format generieren

- **Stilistische Anpassungen:** Die KI kann entsprechend der Prompt-Vorgabe den Stil der Ausgabe weiter anpassen, wie beispielsweise bei Texten den Ton des Inhalts (z. B. formell, informell, technisch oder kreativ)

- **Inhaltliche Präzisierung:** Die KI liefert präzise und detaillierte Informationen basierend auf spezifischen Fragen oder Anweisungen

- **Personalisierung:** Die KI bietet individuell zugeschnittene Inhalte oder Empfehlungen, die auf die Bedürfnisse oder Präferenzen der Nutzerin oder des Nutzers abgestimmt sind

Schritt 5: Generierung neuer Inhalte und multimodale Ausgabe

Nachdem die Daten analysiert und zusammengeführt wurden, nutzt die KI spezielle Modelle, um neue Inhalte zu erstellen. Zwei häufig verwendete Modelle sind **Variational Autoencoders (VAEs^h)** und **Generative Adversarial Networks (GANsⁱ)**. VAEs erzeugen neue Inhalte, indem sie die Muster in den vorhandenen Daten erkennen und darauf aufbauend ähnliche Inhalte erstellen. GANs funktionieren etwas anders: Sie bestehen aus zwei Teilen – einem Generator, der neue Daten erzeugt, und einem Diskriminator, der prüft, wie echt diese Daten wirken. Diese beiden Teile arbeiten im Wettbewerb miteinander, was dazu führt, dass die final generierten Inhalte besonders realistisch werden.

^h Variational Autoencoder (VAE): Ein maschinelles Lernmodell, das darauf abzielt, komplexe Daten zu analysieren und ähnliche, neue Daten zu erzeugen, indem es die wesentlichen Merkmale der Originaldaten lernt.

ⁱ Generative Adversarial Network (GAN): Ein maschinelles Lernmodell, bestehend aus einem Generator, der neue Daten erstellt, und einem Diskriminator, der die Echtheit dieser Daten prüft, um so die Qualität der generierten Inhalte zu verbessern.

Einsatzmöglichkeiten multimodaler GenAI-Modelle in der Versicherungsbranche

Das folgende Kapitel gibt einen umfassenden Überblick über die Anwendungsmöglichkeiten multimodaler GenAI mit Blick auf die Versicherungsbranche. Zunächst werden grundlegende, modalitätsübergreifende Kernfunktionen erläutert. Anschließend wenden wir uns der Praxis und damit dem Prompting selbst zu und werden die Kernfunktionen anhand konkreter Anwendungsfälle illustrieren.

Dabei wird auf spezifische Anwendungsbereiche in der Versicherungsbranche eingegangen, wie beispielsweise die Erstellung von Wettbewerbsanalysen, die Video-basierte Abschätzung von Hausratsversicherungssummen sowie die Erstellung multimodaler Inhalte. Ziel ist es,

aufzuzeigen, wie multimodale GenAI-Modelle interne Prozesse verbessern und darüber hinaus neue Möglichkeiten in der Kundeninteraktion schaffen können.

Kernfunktionen Multimodaler Generativer KI

Die folgende Tabelle zeigt die vielfältigen Kernfunktionen, die durch den Einsatz multimodaler GenAI-Modelle möglich werden. Diese Funktionen spiegeln das breite Spektrum an Fähigkeiten wider, welche die Integration verschiedener Datenmodalitäten wie Text, Bild und Audio in der Versicherungsbranche bietet. Diese Übersicht dient als Grundlage, um die Potenziale und vielseitigen Anwendungen von multimodaler GenAI zu erfassen.

Kernfunktionen multimodaler GenAI^j

Kernfunktion	Beschreibung	Beispiel
Kreieren	Generierung neuer Inhalte, z. B. Texte, Bilder, Audio, und Videos	Generierung von Werbematerial wie beispielsweise TikTok Videos, automatisierte Content-Erstellung für Blogs
Simulieren	Simulation und Nachahmung in verschiedenen Modalitäten	Stimmimitation oder Imitation visueller Stile
Transferieren	Übertragung von Informationen zwischen Modalitäten, z. B. Bild-zu-Text-Transformation	Umwandlung eines Bildes in einen beschreibenden Text
Modifizieren	Anpassung und Veränderung bestehender Inhalte	Personalisierung von Marketinginhalten basierend auf Kundensegmenten
Verbessern	Optimierung von Qualität und Stil	Verbesserung eines Mock-ups basierend auf bestimmten Designprinzipien
Reduzieren	Zusammenfassung und Verdichtung von Informationen	Erstellung von Zusammenfassungen aus langen Videos, Audio zu Text-Transkriptionen
Analysieren	Analyse und Deutung von Inhalten	Überprüfen, ob Textbeschreibungen mit Schadensfotos übereinstimmen
Synchronisieren	Modalitätsübergreifende Synchronisation von Inhalten	Erstellung von Audioinhalten zu einer Videosequenz
Organisieren	Strukturierung und Anordnung von Informationen	Erstellen von To-do-Listen oder Zeitplänen
Erklären	Erläuterung komplexer Sachverhalte	Vereinfachte Erklärungen für Laien
Vorhersagen	Analyse von Mustern zur Vorhersagen von Trends	Vorhersage von Markttrends
Personalisieren	Anpassung von Inhalten an individuelle Bedürfnisse	Personalisierte Essens- oder Fitnesspläne

^j Diese Tabelle baut auf der Kernfunktions-Übersicht aus unserem vorherigen Whitepaper [3] auf und wurde weiterentwickelt, um Aspekte der Multimodalität zu berücksichtigen.

Ausführung der Kernfunktionen durch effektives Prompten

Nachdem die Kernfunktionen multimodaler GenAI-Modelle und ihre breite Anwendbarkeit beleuchtet wurden, folgen nun die praktischen Aspekte des effektiven Promptens. Dieser Abschnitt vermittelt, wie durch klar strukturierte Anweisungen die Leistungsfähigkeit dieser Systeme optimal genutzt werden kann, um die oben beschriebenen Kernfunktionen effektiv zu nutzen. Klare, strukturierte Prompts, die Rollen, Aufgabenstellungen, Kontext und spezifische Kriterien beachten, verbessern die Präzision der KI-Ergebnisse, darüber hinaus gibt es modalspezifische Parameter. Diese werden in der folgenden Tabelle zusammen mit den Prompting-Grundsätzen erläutert:

Aus unserer täglichen Arbeit im ERGO Innovation Lab wissen wir, **dass das erste Ergebnis eines Prompts häufig noch nicht den gewünschten Anforderungen entspricht**. Daher ist es oft notwendig, die Ausgabe in einem iterativen Prozess weiter zu optimieren. Diese Iterationen erfolgen meist in Dialogform, bei dem der Benutzer der KI kontinuierlich Feedback gibt, um die Ausgabe weiter

anzupassen. So kann es beispielsweise nach der ersten Ausgabe Anweisungen wie „Bitte verwende eine formellere Sprache“ geben, woraufhin die KI die Ausgabe entsprechend anpasst. Unsere Erfahrung zeigt, dass dieser iterative Ansatz ein entscheidender Erfolgsfaktor ist, um präzise und qualitativ hochwertige Ergebnisse zu erzielen.

Praktische Anwendungsbeispiele in der Versicherungsbranche

Basierend auf unserer aktuellen Forschung und umfangreichen Expertise im ERGO Innovation Lab werfen wir nun einen Blick auf die praktischen Anwendungen der multimodalen GenAI in der Versicherungsbranche. Hier gibt es vielfältige Einsatzmöglichkeiten, die das Potenzial haben, Prozesse effizienter zu gestalten, den Kundenservice zu optimieren und innovative Produkte zu entwickeln. Die folgenden Prompt-Beispiele veranschaulichen, wie GenAI gezielt dafür eingesetzt werden kann. **Aus Gründen der Übersichtlichkeit wurden die Prompts auf das Wesentliche reduziert und die Ergebnisse stark verkürzt, um die Kernfunktionen und den Nutzen der Technologie hervorzuheben.**

Übersicht relevanter Promptelemente und Parameter, gruppiert nach Ausgabemodalität^k

Prompt-Element	Beschreibung	Text	Bild	Audio	Video ^l
Rollen- definition	Bestimmt, aus welcher Perspektive die KI antwortet und wie die Inhalte präsentiert werden	Expertenrolle (Marketing-Experte, Strategieberater), schriftlicher Stil (formell, informell), Tonart (z. B. sachlich, emotional), Kommunikationsart (z. B. erzählend, argumentativ)	n.a.	Sprechertyp (z.B. Geschlecht, Alter) Sprecherstil (z.B. beruhigend, enthusiastisch), Stimmlage	n.a.
Aufgaben- klärung	Definiert, welche Kernfunktion die KI erfüllen soll, und beschreibt diese detailliert und legt ggf. fest, welche spezifischen Ziele erreicht werden sollen	Texttyp (z. B. Pressemitteilung, Bericht, Storyline), Struktur, Ziel (z.B. informieren, überzeugen)	Stil (z. B. modern, traditionell), Perspektive (z. B. subjektiv, objektiv), Bildart (z.B. Fotografie, Illustration), Detailgrad (z. B. skizzenhaft, detailliert)	Darstellung bestimmter Szenen, welche im Skript beschrieben werden	Darstellung bestimmter Szenen, welche im Skript beschrieben werden, Stil (z.B. Animation, Live-Action),
Kontext- spezifikation	Passt die Ausgabe an den spezifischen Kontext und die Bedürfnisse der Zielgruppe an	Zielgruppe (z.B. Kunden, Mitarbeiter), Kontext (z. B. Schulung, Marketing)	Umgebungsdetails (z. B. Innen, Außen), Fokus (z.B. Produkt, Person), Licht (z.B. natürlich, künstlich)	Umfeldgeräusche (z.B. Straße, Natur), Aufnahmeort (z.B. Studio, vor Ort), Sprecherrolle (z. B. Erzähler, Charakter)	Kameraperspektive (z.B. subjektiv, objektiv), Kameraführung (z.B. statisch, dynamisch), Bildausschnitt (z.B. Nahaufnahme, Totale), Beleuchtung (z.B. dramatisch, natürlich), Dynamik (z.B. ruhig, hektisch)
Ausgabe- format	Festlegung des Formats und der Struktur bzw. Templates, angepasst an die Modalität	Textstruktur (z.B. Fließtext, Liste, Tabelle, Code), Stil (z. B. narrativ, deskriptiv)	Ausgabegröße (z.B. Poster, Thumbnail)	Gesprochene Sprache (Voice-over, Beschreibungen usw.), Musik, Soundeffekte, Geräusche	Format (z. B. 16:9, quadratisch), Nachbearbeitung (z. B. Übergänge, Farbfilter, Effekte)

^k Diese Liste illustriert beispielhaft Promptelemente und -parameter und erhebt keinen Anspruch auf Vollständigkeit.

^l In dieser Tabelle liegt der Schwerpunkt bei Videos auf den visuellen Aspekten der bewegten Bilder, nicht auf der Audiospur.

Analyse von textuellen und visuellen Webinhalten und Berichtserstellung

Wettbewerbsanalyse für eine neue Kfz-Versicherung



Du bist ein erfahrener Marktanalyst in einer Versicherungsgesellschaft. Erstelle eine kurze Wettbewerbsanalyse für Versicherungen. Der Bericht sollte die wichtigsten Wettbewerber, deren Produktangebote und Preisgestaltung kurz beschreiben und eine Empfehlung zur Positionierung der neuen Kfz-Versicherung auf dem Markt enthalten. Nutze öffentlich zugängliche Informationen und Branchenberichte und Webseiten der Wettbewerber. Der Bericht soll einen einleitenden Fließtext und die Wettbewerbermerkmale sowie Vorschläge für die neue Produktpositionierung als Liste enthalten.

Rolle
 Aufgabe
 Kontext
 Format



Kurzbericht zur Wettbewerbsanalyse für Kfz-Versicherungen

Kapitel 1: Marktübersicht • Text: Kurze Einführung in den Markt für Kfz-Versicherungen, einschließlich der wichtigsten Trends und Marktteilnehmer.

Kapitel 2: Analyse der Wettbewerber. Kurze Beschreibung der wichtigsten Wettbewerber, ihrer Produktangebote und Preisgestaltung.

- **Wettbewerber 1:** Übersicht über die Hauptmerkmale der Produkte und die Preisstruktur
- **Wettbewerber 2:** Übersicht über die Hauptmerkmale der Produkte und die Preisstruktur
- **Wettbewerber 3:** Übersicht über die Hauptmerkmale der Produkte und die Preisstruktur

Kapitel 3: Empfehlungen zur Positionierung. Kurze Empfehlung zur Positionierung der neuen Kfz-Versicherung, basierend auf der Analyse der Wettbewerber.

- **Vorschläge für Produktdifferenzierung**
- **Vorschläge für die Preisstrategie**
- **Vorschläge für Marketingansätze**



Für eine umfassende Analyse empfiehlt es sich, den Prompt weiter zu spezifizieren und relevante Wettbewerbsmerkmale, Quellen und spezifische Erwartungen an die neue Positionierung genauer zu definieren.

Videoanalyse und Berichtserstellung

Video-basierte Beratung zur Abschätzung der Hausratsversicherungssumme



Du bist ein Experte im Thema Hausratsversicherung. Das beigefügte Video enthält die Aufnahme einer Wohnung. Bitte liste stichwortartig relevante Gegenstände der Wohnung und schätze den Wert dieser Gegenstände. Vermerke außerdem weitere Gegenstände, die in einem Haushalt normalerweise vorhanden sind, aber nicht in dem Video zu finden sind. Empfehle eine plausible Versicherungssumme auf Basis des Videos und unter der Annahme, dass die weiteren Gegenstände ebenfalls im Haushalt vorhanden sind.

Dateien: Videoaufnahme einer Wohnungstour

Rolle
 Aufgabe
 Kontext
 Format



Beschreibung der Gegenstände in der Wohnung:

1. **Fernseher (ca. 55 Zoll):** 800 €
2. **Fernsehtisch:** 150 €
3. **Sofa (Ecksofa):** 1.000 €
4. **Couchtisch:** 200 €
5. **Esstisch mit 4 Stühlen:** 400 €
6. **Computer (Monitor, Tastatur, Maus):** 600 €
7. **Vitrine mit Gläsern und Dekoration:** 300 €
8. **Luftreiniger:** 100 €
9. **Stehlampe:** 50 €
10. **Dekorationsartikel (z.B. Modellflugzeug, Plüschtiere):** 100 €
11. ...

Geschätzter Wert der Gegenstände in der Wohnung:

Gesamtwert: 9.700 €

Weitere Gegenstände, die in einem Haushalt normalerweise vorhanden sind:

1. **Küche:**
 - Mikrowelle: 100 €
 - Küchenutensilien und Geschirr: 300 €
2. **Schlafzimmer:**
 - Nachttische: 200 €
3. ...

Geschätzter Wert der weiteren Gegenstände: 2.300 €

Empfohlene Versicherungssumme für den gesamten Haushalt: 12.700 €



Um präzisere Preisabschätzungen zu machen, können spezifische Marken und Modelle der Gegenstände im Video identifiziert werden. Indem diese Details erfasst und benannt werden, lassen sich genauere Marktwerte ermitteln, was zu einer genaueren Schätzung der Versicherungssumme führt.

UX/UI Analyse von Webseiten

Bewertung von Nutzerfreundlichkeit und Design einer Landingpage und Ableiten von Verbesserungsmaßnahmen



Du bist ein versierter UX/UI-Designer und hast die Aufgabe, die Landingpage einer Versicherung (www.fiktive-versicherung.de) zu analysieren und Verbesserungsvorschläge abzuleiten. Ziel der Analyse ist es, die Benutzerfreundlichkeit und das visuelle Design der Landingpage zu bewerten und konkrete Verbesserungsvorschläge zu entwickeln. Die Landingpage wird von bestehenden und potenziellen Kunden genutzt, um Informationen über Versicherungspolizen, Schadenmeldungen und die angebotenen Produkte und Dienstleistungen zu erhalten. Gib Analyseergebnisse und Verbesserungsvorschläge als Liste aus.

Rolle
Aufgabe
Kontext
Format



Navigationsstruktur: Die Navigation der Landingpage zeigt eine solide Grundlage, indem die Hauptmenüpunkte klar benannt sind. Dies hilft den Nutzern, sich auf der Seite zurechtzufinden. Es besteht jedoch Potenzial, die Zugänglichkeit weiter zu verbessern, insbesondere für Nutzer, die schnell zu spezifischen Informationen oder Funktionen gelangen möchten. Ein paar Anpassungen könnten die Nutzererfahrung deutlich steigern.

Visuelles Design: Das aktuelle Design der Landingpage schafft es, eine vertrauenswürdige Atmosphäre zu vermitteln. Dies ist wichtig, um das Vertrauen der Kunden zu stärken. Mit einer Aktualisierung der Farbpalette könnte das Design noch besser zur Markenidentität der Versicherung passen und gleichzeitig die Lesbarkeit und visuelle Attraktivität erhöhen.

Inhaltliche Struktur: Die Informationen auf der Landingpage sind umfassend und bieten den Nutzern alle notwendigen Details. Durch eine Umstrukturierung und bessere Formatierung der Texte könnten die Informationen noch leichter zugänglich gemacht werden, wodurch Nutzer schneller das finden, was sie suchen (...)

Maßnahmen zur Verbesserung von UX/UI:

- 1. Verbesserung der Navigationsstruktur:** Einführung von Quick-Links zu den am häufigsten genutzten Funktionen oder Informationen wie „Versicherungen vergleichen“ oder „Angebot anfordern“ direkt auf der Startseite oder im Hauptmenü. Dies würde den Nutzern den Zugang zu wichtigen Bereichen erleichtern und die Benutzerfreundlichkeit steigern.
- 2. Modernisierung des visuellen Designs:** Anpassung der Farbpalette und Schriftarten an die Corporate Identity der Versicherung. Dies könnte durch eine Erfriechung der visuellen Elemente erreicht werden, was die Landingpage moderner und einheitlicher wirken lässt und gleichzeitig die Markenidentität stärkt.
- 3. Umsetzung einer klaren Informationshierarchie und visuellen Strukturierung:** Um die Zugänglichkeit der umfassenden Informationen auf der Landingpage zu verbessern, sollte eine klare Informationshierarchie etabliert werden. Dies kann durch die Einführung von Überschriften, Unterüberschriften und Bullet-Points erreicht werden (...)



Viele GenAI-Modelle können bereits in Echtzeit aus Webseiten zugreifen und beispielsweise Webseiten von Mitbewerbern analysieren und vergleichen oder SEO-Aspekte wie Keywords, Meta-Tags und Webseitenstruktur überprüfen, um die Sichtbarkeit und das Ranking in Suchmaschinen zu verbessern.

Erstellung von Bildmaterial

Generierung von Bildern für die Gestaltung einer Landingpage einer Pflegeversicherung



Erstelle ein fotorealistisches Bild im Querformat, mit folgendem Inhalt: Eine ältere Person mit Falten, die einen gemütlichen Pullover trägt, sitzt auf einem weichen Sofa in einem Wohnzimmer. Eine Pflegerin, gekleidet in Berufskleidung, sitzt daneben, hält ihre Hand und führt ein freundliches Gespräch, während ihr Gegenüber freundlich lächelt und aufmerksam wirkt. Der Raum ist gemütlich eingerichtet mit einer großen Couch, einem hölzernen Couchtisch, einem Buchregal und ein paar Topfpflanzen. Sonnenlicht strömt durch das Fenster und taucht das Wohnzimmer in ein warmes, stimmungsvolles Licht, das ein Gefühl von Geborgenheit und Fürsorge vermittelt.

Aufgabe Kontext Format



Grundsätzlich liefern sehr detaillierte Bildbeschreibungen deutlich bessere Bilder als kurze Prompts. Es gibt mittlerweile zahlreiche Tools, die Prompts zur Bildgenerierung optimieren und anreichern, um eine bessere Bildqualität zu erreichen.

Einige Bildgeneratoren, wie Midjourney, unterstützen die Verwendung bestimmter Parameter am Ende eines Prompts und ermöglichen damit eine feinere Kontrolle über das Bild-design. Diese Parameter werden mit zwei Strichen eingeleitet gefolgt von ihrem Namen und ihrem Wert: --ar (Seitenverhältnis), --stylize (Grad der Stilisierung), --weird (Grad der Surrealität), --no (exkludierte Elemente oder Stile).

Erstellung von Audioinhalten

Erstellung eines Beitrags für eine Mini-Podcast Serie über Versicherungswissen

Für die Erstellung eines Podcasts ist ein Konzept und eine klare Themenstruktur notwendig. Jede Episode wird durch einen individuellen Prompt vorbereitet und die generierten Skripte anschließend durch Text-to-Speech-Technologie vertont.

Themen und musikalische Untermalung können ebenfalls durch GenAI-Modelle entwickelt werden, wodurch der gesamte Produktionsprozess effizient und kosteneffektiv abläuft.



Du bist ein Experte für Versicherungen und Gastgeber eines Podcasts, der Hörer über verschiedene Versicherungsthemen aufklärt. Erstelle ein Skript für eine Podcast-Episode, die eine häufig gestellte Frage zu einem Versicherungsthema beantwortet. Die Episode sollte in einfacher, verständlicher Sprache gehalten sein und praktische Tipps oder Empfehlungen enthalten, die den Hörern helfen, informierte Entscheidungen zu treffen. Die Länge der Episode sollte zwischen 8-10 Minuten liegen. Liefere ein vollständiges Skript für die Podcast-Episode, das eine klare Einführung in das Thema bietet, relevante Beispiele und Szenarien erklärt und abschließend praktische Tipps oder Handlungsempfehlungen gibt.

Rolle

Aufgabe

Kontext

Format



Podcast-Episode: Die Vorteile Ihrer Haftpflichtversicherung und aktuelle Angebote

„Willkommen bei unserer heutigen Episode von 'Versicherungswissen leicht erklärt'! Heute sprechen wir über ein Thema, das für viele von uns relevant ist: die Haftpflichtversicherung. Was genau ist das und warum sollte jeder von uns eine haben? Stellen Sie sich vor, Sie zerbrechen beim Besuch Ihrer Tante eine wertvolle Vase. Oder ihr Kind kippt versehentlich Kirschsafte auf den Teppich ihrer Freunde. In solchen Situationen kommt die Haftpflichtversicherung ins Spiel...“

Abschnitt 1: Einführung und Bedeutung der Haftpflichtversicherung

„Beginnen wir damit, was eine Haftpflichtversicherung überhaupt ist. Die Haftpflichtversicherung ist eine der wichtigsten Versicherungen für Privatpersonen. Sie schützt Sie vor den finanziellen Folgen, wenn Sie unbeabsichtigt jemandem Schaden zufügen – sei es ein Sachschaden oder ein Personenschaden. In Deutschland ist derjenige, der einem anderen Schaden zufügt, gesetzlich verpflichtet, für den Schaden zu haften. Das bedeutet, dass Sie im schlimmsten Fall für sehr hohe Kosten aufkommen müssten. Hier setzt die Haftpflichtversicherung an (...)



Google LM Notebooks können verwendet werden, um automatisch Podcast-Skripte zu generieren und sprachlich zu synthetisieren, indem Sprachmodelle Themen analysieren und passende Inhalte erstellen. Dies vereinfacht die Produktion von gut strukturierten und informativen Episoden mit minimalem Aufwand.

Erstellung von Werbevideos

Erstellung einer Videoszene für ein Hundekrankenversicherungs-Werbevideo

Für die Erstellung eines kompletten Werbefilms ist es zunächst notwendig, ein Videokonzept und eine Storyline zu entwerfen. Jede Szene des Films wird anschließend durch einen Prompt beschrieben; somit sind mehrere Prompts für die Erstellung des gesamten Videos erforderlich. Nachdem alle Szenen generiert wurden, werden diese zusammengesetzt, um den vollständigen Werbefilm zu erzeugen. Manche Modelle erlauben die simultane Erstellung aller Szenen inklusive Übergänge, was die Erstellung noch effizienter macht. Anschließend kann dieser auch auditiv unterlegt werden.

Storyline, Drehbuch und auditive Untermalung des Werbefilms können ebenfalls durch zusätzliche GenAI-Mo-

delle entwickelt werden. So wird der gesamte Produktionsprozess – von der Idee über das Drehbuch bis hin zur fertigen Videokampagne – durch den Einsatz neuer GenAI-Technologien effizient und kosteneffektiv unterstützt.

Der im Folgenden präsentierte Videoclip wurde mit der AI-Video-Plattform Runway erstellt. Er zeigt exemplarisch eine Szene für die Hundekrankenversicherung. Im Folgenden werden insgesamt zwei Ansätze für die GenAI Videogenerierung demonstriert (Text-zu-Video und Text-zu-Bild-zu-Video).

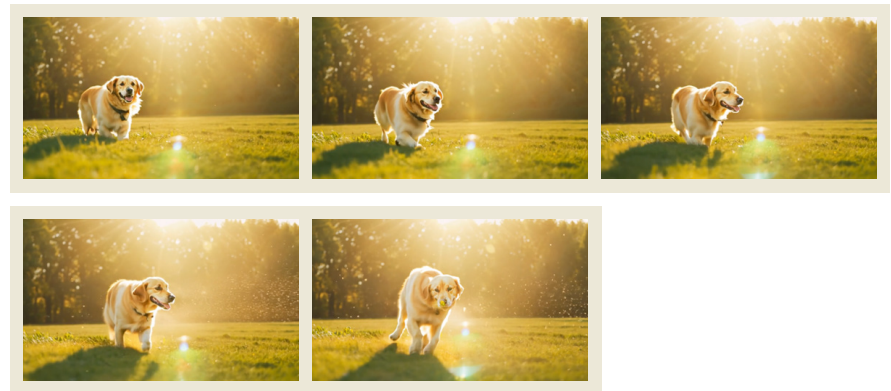


Prompt-Variation 1 (Text zu Video):

Weitwinkel Kamera, Zeitlupe: Ein Golden Retriever rennt durch einen Park an einem strahlenden Sommertag

Rolle
 Aufgabe
 Kontext
 Format

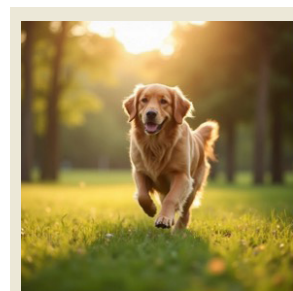
Output:



Prompt-Variation 2 (Text zu Bild zu Video):

Ein Golden Retriever rennt durch einen Park an einem strahlenden Sommertag

Output:





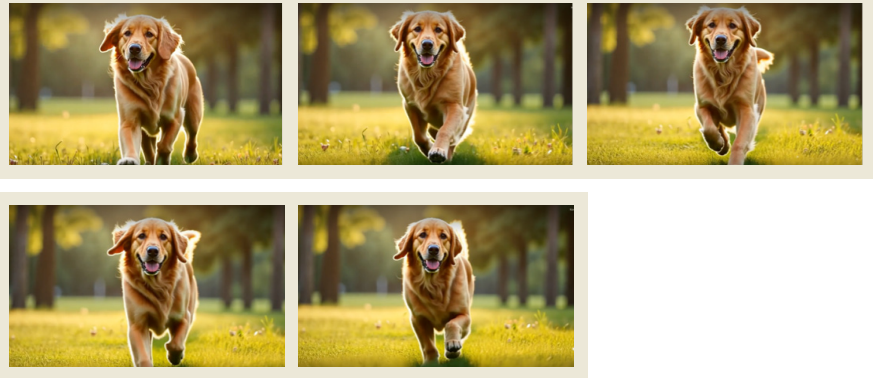
Prompt-Variation 2.2 (Bild zu Video):

Weitwinkel Kamera, Zeitlupe: Ein Golden Retriever rennt durch einen Park an einem strahlenden Sommertag

Rolle
 Aufgabe
 Kontext
 Format



Output:



Aus unserer umfangreichen Arbeit im ERGO Innovation Lab zeigt sich, dass die Videoqualität oft besser ist, wenn nicht ausschließlich auf Textprompts gesetzt wird. Stattdessen erzielt die Kombination eines vorab generierten, qualitativ hochwertigen AI-Bildes und Textprompt deutlich bessere Ergebnisse (Prompt-Variation 2).

Unsere Erfahrung zeigt, dass Videogeneratoren oft noch Schwierigkeiten haben, natürliche Bewegungen sowie non-verbale Interaktionen zwischen Menschen realistisch darzustellen. Weitere Schwächen sind oftmals die physikalische Konsistenz, wie beispielsweise die korrekte Darstellung von Schatten und Lichtreflexionen.

Aktuelle Herausforderungen und Grenzen neuer Generativer KI-Modelle

In den vergangenen Kapiteln wurde ersichtlich, dass sich eine Vielzahl von Anwendungsfällen für multimodale GenAI mit vielversprechenden Vorteilen ableiten lässt. Gleichzeitig sind jedoch auch neue Herausforderungen zu bewältigen, die sich aus technologischen, regulatorischen und rechtlichen, vertrauensbezogenen sowie ethischen und gesellschaftlichen Aspekten ergeben^{10, 11, 12}.

Technologische Herausforderungen

Datenkorrektheit und Halluzinationen

Wie alle GenAI-Modelle kann auch multimodale GenAI halluzinierte Inhalte erzeugen. Dabei handelt es sich um fehlerhafte, erfundene Inhalte, welche nicht in den Trainingsdaten enthalten sind. Sie können zu falschen oder unerwarteten Antworten führen, was in sensiblen Bereichen wie Versicherungen problematisch ist. Ein Beispiel: Ein Schiedsgericht in British Columbia verurteilte Air Canada, weil ein Chatbot falsche Informationen zur Rückerstattung eines Flugtickets gegeben hatte¹³. Eine Validierung generierter Inhalte durch Menschen („Human-in-the-Loop“), festgelegte Regeln oder Halluzinations-Checker ist daher aktuell unerlässlich.

Rechenleistung und Kosten

Die Verarbeitung multimodaler Daten ist mit einem hohen Rechenaufwand verbunden, was zu entsprechenden hohen Kosten führt. Unternehmen müssen daher sorgfältig planen, auf welche Art und Weise sie auf die Modelle zugreifen und wie sie diese einsetzen möchten. Das eigenständige Training von Modellen ist nach wie vor mit sehr hohen Kosten und erheblichem Aufwand verbunden. Eine wirtschaftlichere Alternative stellt die Nutzung vortrainierter, meist Cloud-basierter Modell-Dienste dar. Diese ermöglichen eine flexible Skalierung der benötigten Rechenleistung, ohne dass teure Hardwareinvestitionen erforderlich sind.

Die Kosten für die Nutzung solcher vortrainierter, Cloud-basierter KI-Modelle sind in jüngster Zeit deutlich gesunken. Anbieter wie OpenAI haben die Preise für die Nutzung bestimmter Modelle erheblich gesenkt, darüber hinaus gibt mittlerweile auch kostengünstigere, ressourcenschonende Mini-Varianten, welche insgesamt weniger Rechenleistung benötigen. Unternehmen sollten die Rechenleistungskosten genau kalkulieren und die ver-

schiedenen Zugriffs- und Nutzungsmöglichkeiten sorgfältig abwägen, um kosten- und ressourcenschonend die Vorteile multimodaler GenAI effizient zu nutzen.

Regulatorische und rechtliche Fragestellungen

Urheberrechtliche Fragestellungen

Das Urheberrecht stellt im Bereich der multimodalen GenAI, in der Bild- und Audioelemente kombiniert werden, eine besondere Herausforderung dar, da mehrere Urheberrechtslinien berührt oder verletzt werden könnten. Der direkte Vergleich von AI-generierten Inhalten mit den Originaldaten ist oft nicht ohne weiteres möglich, was zu rechtlichen Unsicherheiten führt. Dies betrifft beispielsweise Bilder, die einzelne Elemente enthalten könnten, die urheberrechtlich geschützt sind, sowie Audioinhalten, bei denen Stimmen imitiert werden, wie im Fall der Nutzung einer an Scarlett Johansson angelehnten Stimme durch OpenAI¹⁴. Um Urheberrechts- oder Persönlichkeitsrechtsverletzungen zu vermeiden, ist daher eine sorgfältige rechtliche Prüfung bei der Nutzung von multimodalen GenAI-Systemen von besonderer Relevanz. Darüber hinaus bieten manche Anbieter wie Adobe eine IP-Haftungsfreistellung für generierte Bilder, da deren Modelle auf gemeinfreien oder eigenen Inhalten trainiert wurden.

Regulierung und Compliance

Die Nutzung multimodaler GenAI ist mit steigenden regulatorischen Anforderungen verbunden, insbesondere in der Versicherungsbranche. Der EU AI Act legt klare Vorgaben für den Einsatz solcher Technologien fest, insbesondere für Hochrisiko-Anwendungen wie die Risikobewertung und Schadensabwicklung. Diese Anwendungen unterliegen strengen Vorschriften hinsichtlich Transparenz, Fairness und Datenschutz.

Nachvollziehbarkeit und Förderung der Nutzerakzeptanz

Erklärbarkeit und Vertrauen

Die Entscheidungen von KI-Modellen sind oft schwer nachvollziehbar. Im Versicherungsbereich, wo Vertrauen und Transparenz entscheidend sind, sollten Modelle mit erklärbaren KI-Techniken (XAI: Explainable AI) ausgestat-

tet werden. Diese Techniken machen den Entscheidungsprozess transparenter und ermöglichen es den Nutzern, die Logik hinter den Empfehlungen der KI besser zu verstehen, was wiederum Vertrauen und Akzeptanz gegenüber GenAI Technologien stärkt. Dazu gehört das Aufzeigen der Eingabeparameter, die zu einer bestimmten Entscheidung geführt hat bzw. welchen Einfluss eine Änderung dieser Parameter auf die Entscheidung hätte.

Akzeptanz und Schulung

Bei der Einführung von GenAI-Lösungen müssen neben technologischen, rechtlichen und sicherheitsrelevanten Aspekten vor allem die Mitarbeitenden einbezogen werden, um Akzeptanz zu schaffen. Der Erfolg von multimodalen GenAI-Modellen in Unternehmen hängt von benutzerfreundlichen Schnittstellen sowie gezielten Aus- und Weiterbildungsangeboten ab. Eine Studie der bitkom¹⁵ zeigt, dass 48 Prozent der deutschen Industrieunternehmen fehlendes Know-how als Haupthindernis für die Integration von GenKI sehen. Unternehmen sollten daher Weiterbildungsprogramme und Supportstrukturen einführen, um die Kompetenzen der Mitarbeiterinnen und Mitarbeiter zu auszubauen.

Ethische und gesellschaftliche Implikationen

Vorurteile und Fairness

KI-Modelle können unbewusste Vorurteile aus den Trainingsdaten übernehmen, was zu unfairen oder verzerrten Entscheidungen führen kann, wie eine UNESCO Studie zeigte¹⁶. Dazu zählen unter anderem Geschlechtsstereotype und rassistische Klischees. Diese Problematik stellt eine bedeutende Herausforderung dar, besonders im Hinblick auf die Gleichbehandlung aller Kundinnen und Kunden. Um Diskriminierung zu reduzieren, müssen Unter-

nehmen ihre Trainingsdaten sorgfältig überwachen und anpassen, um sicherzustellen, dass die Nutzung der KI möglichst fair und diskriminierungsfrei ist.

Deep Fakes, Desinformation und Ethik

Der Einsatz von GenAI bringt auch neue Herausforderungen mit sich, insbesondere im Bereich der Sicherheit und Ethik. Der wachsende Bedarf an Cybersecurity-Maßnahmen und Faktenprüfungsdiensten^{17, 18} untermauert, dass der Schutz sensibler Daten und die Gewährleistung der Integrität generierter Inhalte immer relevanter werden.

Unternehmen sollten daher nicht nur in fortschrittliche Sicherheitsmaßnahmen investieren, sondern auch ethische Standards und umfassende Regulierungen einführen, um den verantwortungsvollen Einsatz von GenAI zu unterstützen und das Vertrauen von Mitarbeiterinnen und Mitarbeitern, Kundinnen und Kunden sowie allen relevanten Stakeholdern in diese Technologien zu fördern^{19, 20, 21}.

Fazit

Das Potenzial von multimodaler GenAI ist enorm, allerdings sind mit dessen Realisierung auch eine Reihe von Herausforderungen verbunden, die es zu bewältigen gilt. Zu den wesentlichen Herausforderungen zählt die Sicherstellung der Datengenauigkeit, das Management der Rechenkosten sowie das Navigieren durch rechtliche Rahmenbedingungen. Von besonderer Bedeutung ist dabei die Förderung von Vertrauen durch Transparenz und Fairness. Um diese Risiken zu mindern, müssen Unternehmen proaktive Maßnahmen ergreifen. Zudem werden ethische Überlegungen, insbesondere in Bezug auf Voreingenommenheit und Deepfakes, weiterhin beeinflussen, wie GenAI in Branchen wie der Versicherungswirtschaft verantwortungsvoll implementiert wird.

Ausblick: Die Versicherungsbranche muss aktiv werden

Prognosen für den künftigen Einfluss multimodaler GenAI-Modelle

Die Versicherungsbranche steht vor einer technologischen Transformation, die durch die Integration neuer GenAI-Modelle angetrieben wird. Diese Technologien haben das Potenzial, bestehende Prozesse zu optimieren, und darüber hinaus auch neue Geschäftsmodelle zu kreieren. Da Versicherungsunternehmen ein datenbasiertes Geschäftsmodell verfolgen, sind sie prädestiniert für den Einsatz von GenAI. Im Vergleich zu anderen Branchen wird der Einfluss dieser Technologie hier besonders spürbar sein, da sie die Verarbeitung und Analyse großer Textmengen effizient unterstützt.

Vermutlich wird die Modalität Text den größten Einfluss auf die Branche haben, da ein Großteil der Interaktionen und der Datenverarbeitung in textbasierter Form erfolgt. Dennoch gibt es zahlreiche Anwendungsfälle, in denen die Multimodalität zukünftig eine entscheidende Rolle spielen wird, wie beispielsweise die Analyse von Bildern und Videos zur Schadensdokumentation oder die Bereitstellung virtueller Avatare zur personalisierten Kundenberatung.

Demokratisierung von Wissen und Fähigkeiten

Die kommenden Jahre werden geprägt sein von einer Wissensdemokratisierung, ermöglicht durch den nahezu unbegrenzten und ständig verfügbaren Zugang zu Wissen²². Dieser Trend wird großen Einfluss darauf haben, wie und wo Menschen Informationen suchen^{23, 24} und fördert eine Kultur des kontinuierlichen Lernens.

Auch Versicherungsunternehmen profitieren von dieser Demokratisierung, da sie ihren Mitarbeiterinnen und Mitarbeitern besseren Zugang zu wertvollen Informationen und Schulungsmaterialien ermöglichen können. Dies fördert nicht nur die Kompetenzentwicklung innerhalb des Unternehmens, sondern verbessert auch die Beratung und den Service für Kundinnen und Kunden durch schnellere Informationsbeschaffung.

Steigerung der Produktivität von Wissensarbeitern

GenAI wird die Produktivität von Wissensarbeitern signifikant steigern²⁵ und deren Arbeitsweise nachhaltig verändern. Studien zeigen, dass etwa 75% der Produktivitätssteigerung von Generativer KI in den Bereichen Kundenservice, Marketing und Vertrieb, Softwareentwicklung, Forschung und Entwicklung zu verzeichnen sind²⁶.

Die Integration von GenAI in diese Schlüsselbereiche ermöglicht es Wissensarbeitern, ihre Effizienz zu steigern, und sich auf strategischere und kreativere Aufgaben zu konzentrieren und dadurch einen höheren Wert für ihre Unternehmen zu generieren.

Hyper-Personalisierung als neuer Standard

Ob in der Anpassung von Marketingbotschaften in Echtzeit oder der individuellen Gestaltung von Versicherungsprodukten basierend auf Kundenbedürfnissen, die Möglichkeiten sind vielfältig und versprechen eine signifikante Verbesserung der Kundenerfahrung. Durch den Einsatz von GenAI können Versicherungsunternehmen maßgeschneiderte Lösungen anbieten, die den spezifischen Bedürfnissen und Präferenzen jedes Kunden gerecht werden.

Integration in Roboter, Maschinen und Fahrzeugen

Die Integration von multimodale GenAI in Maschinen und Roboter wird die Interaktion mit diesen Geräten menschenähnlicher gestalten²⁷ und ihre Fähigkeiten erweitern, sowohl in der Ausführung physischer als auch kognitiver Aufgaben²⁸. Diese Technologie wird nicht nur die Industrie revolutionieren, sondern auch neue Maßstäbe in der Sicherheit und Effizienz setzen, sei es in Fabriken²⁹, Baustellen oder autonomen Fahrzeugen³⁰.

Für Versicherungsunternehmen eröffnet die Integration von GenAI in Roboter, Maschinen und Fahrzeuge mehrere Chancen. GenAI-gesteuerte Systeme könnten das Risiko von Unfällen durch verbesserte Sicherheitsmechanismen reduzieren, was zu niedrigeren Schadensfällen und Versicherungsansprüchen führt. Zudem ermöglichen Daten aus diesen intelligenten Systemen präzisere Risikoanalysen und maßgeschneiderte Versicherungspolice, was zur Entwicklung dynamischer Prämienmodelle beiträgt, die den tatsächlichen Nutzungsbedingungen und dem Verhalten der Maschinen und Fahrzeuge entsprechen.

Exponentielle Beschleunigung von Innovations- und Entwicklungszyklen

Ein weiterer entscheidender Faktor wird die Beschleunigung von Innovationszyklen durch GenAI-generierte Ideen und deren schnelle Umsetzung sein³¹. Dies ermöglicht es Unternehmen, agiler auf Marktveränderungen zu reagieren und ihre Produkte kontinuierlich zu verbessern, was zu einem Wettbewerbsvorteil in einer zunehmend dynamischen Wirtschaftsumgebung führt.

Handlungsempfehlungen für Versicherungsunternehmen

Die Integration multimodaler GenAI bietet immense Chancen, Geschäftsmodelle, Prozesse und Kundenerlebnisse grundlegend zu transformieren. Um das volle Potenzial von GenAI auszuschöpfen, müssen Versicherungsunternehmen ihren Fokus auf ihre Datenstrategie legen und in die richtige Infrastruktur und Talente investieren.

Erschließung neuer Geschäftsfelder und Anpassung des Produktportfolios

Unternehmen sollten die Auswirkungen von GenAI auf ihr Geschäftsmodell und ihre Dienstleistungen konstant analysieren, um langfristige Marktveränderungen zu erkennen und geeignete Maßnahmen abzuleiten. Beispielsweise könnte eine Verringerung des Unfallrisikos durch autonomes Fahren die Nachfrage nach traditionellen Kfz-Versicherungen reduzieren. Gleichzeitig können durch den Einsatz von KI-basierten Technologien wie Sensoren und Kameras neue Präventionsansätze entwickelt werden, die das bestehende Produktportfolio in Richtung eines technologiegetriebenen Präventionsansatzes transformiert.

Strategische Allianzen und Partnerschaften

Strategische Allianzen und Partnerschaften sind entscheidend, um Zugang zu modernster Technologie und Expertise zu erhalten. Versicherungsunternehmen sollten mit führenden Technologieunternehmen und KI-Startups zusammenarbeiten, um von deren Innovationskraft und Know-how zu profitieren. Durch Kooperationen mit Brancheninitiativen und -plattformen können Best-Practices ausgetauscht, Standards entwickelt und gemeinsame Herausforderungen angegangen werden, die durch die Implementierung von GenAI entstehen könnten.

Anpassung der Datenstrategie und -operationen

Eine der zentralen Herausforderungen bei der Integration von GenAI ist die Anpassung der Datenstrategie. Versicherungsunternehmen müssen sicherstellen, dass ihre Datenintegration und -qualität auf höchstem Niveau ist. Dies erfordert die Entwicklung und Implementierung von Systemen, die Daten aus verschiedenen internen und externen Quellen reibungslos integrieren und konsolidieren können. Dazu gehören insbesondere auch bisher eher vernachlässigte Datenquellen wie Fließtextdokumente, welche bis dato eher nicht als „Daten“ betrachtet wurden. Die Evaluation und Aufbereitung dieser Quellen ist unerlässlich, um die bestmögliche Bandbreite an Daten nutzen zu können.

Ein weiterer wichtiger Schritt ist die Implementierung von Technologien zur Echtzeit-Datenverarbeitung, um das

Potential von GenAI-basierten Applikationen bestmöglich auszuschöpfen, wie beispielsweise Schadensbewertungen in Echtzeit und in der direkten Kundenkommunikation.

Investition in KI-Infrastruktur

Um eine skalierbare und effiziente GenAI-Umgebung zu schaffen, sollten Versicherungsunternehmen sowohl kurz- als auch langfristige Investitionen in ihre technologische Infrastruktur tätigen. Kurzfristige Investitionen sind notwendig, um die erforderliche Flexibilität zu gewährleisten und schnell auf aktuelle Marktanforderungen reagieren zu können. Dazu gehört die Implementierung robuster Cloud-Plattformen und fortschrittlicher Softwarelösungen sowie gegebenenfalls ausgewählte KI-Hardware. Gleichzeitig gilt es, die langfristige Skalierbarkeit sicherzustellen, um den sich ständig ändernden Marktanforderungen gerecht zu werden. Diese duale Strategie ermöglicht es Versicherungsunternehmen, das volle Potenzial von GenAI in ihren Prozessen und Dienstleistungen ausschöpfen.

Change-Management und Förderung der Innovationskultur

Die Einführung von GenAI erfordert einen umfassenden Change-Management-Prozess innerhalb der Versicherungsunternehmen. Dieser Prozess sollte darauf ausgerichtet sein, die gesamte Organisation auf die Nutzung von GenAI-gestützten Prozessen und Technologien vorzubereiten. Dazu gehört auch die Förderung einer Kultur der kontinuierlichen Verbesserung, des Lernens und der Offenheit gegenüber Veränderungen zu etablieren.

Ein erfolgreicher Change-Management-Prozess sollte auch die Demokratisierung von multimodaler GenAI berücksichtigen, indem sicherstellt wird, dass alle Mitarbeiter Zugang zu entsprechenden Tools und Ressourcen haben, um ihre Arbeit effizienter und effektiver zu gestalten. Diese gleichberechtigte digitale Teilhabe fördert eine integrative Unternehmenskultur und steigert das Potenzial von GenAI im gesamten Unternehmen.

Rekrutierung von Tech-Talent und Mitarbeiterschulungen

Neben der Rekrutierung von KI-Experten ist eine kontinuierliche, umfassende Weiterbildung von Mitarbeiterinnen und Mitarbeitern entscheidend, um den erfolgreichen Einsatz von GenAI-Technologien im Unternehmen sicherzustellen. Unternehmen sollten es ihren Mitarbeitern ermöglichen, ihre Fähigkeiten und Kompetenzen im Umgang mit GenAI zu erweitern, um die neuesten Entwicklungen und Best Practices zu nutzen. Dies kann durch gezielte Schulungsmaßnahmen, Workshops und Mentoring erreicht werden, die nicht nur technische Fähigkeiten, sondern auch ein Verständnis für die strategischen Vorteile von GenAI vermitteln.

Schaffung von regulatorischen und ethischen Rahmenbedingungen

Ein weiterer wichtiger Aspekt ist die Entwicklung von Richtlinien zur ethischen Nutzung von GenAI. Versicherungsunternehmen müssen sicherstellen, dass die Technologie verantwortungsbewusst eingesetzt wird. Dies erfordert die Einhaltung neuer Vorschriften, wie beispielsweise dem EU AI Act, und die eigenverantwortliche Sicherstellung, dass GenAI-Initiativen allen gesetzlichen Anforderungen entsprechen.

Durch geeignete Maßnahmen und klare Richtlinien können Versicherungsunternehmen sicherstellen, dass die Nutzung von GenAI ethischen Standards entspricht und gleichzeitig den Anforderungen und Erwartungen ihrer Kundinnen und Kunden an innovative Versicherer gerecht wird.

Fazit

Multimodale GenAI hat das Potenzial, die Art und Weise, wie wir mit Daten arbeiten, grundlegend zu transformieren – nicht nur in der Versicherungsbranche, sondern in zahlreichen weiteren Sektoren. In diesem Whitepaper haben wir bereits vielfältige Anwendungsbereiche beleuchtet – von der Transformation des Kundenerlebnis-

ses über die Optimierung interner Prozesse bis hin zur datengetriebenen Unterstützung strategischer Entscheidungen. Eins ist jedoch klar: Wir stehen erst am Anfang dieser Entwicklung. Diese Technologie entwickelt sich mit enormer Geschwindigkeit weiter, fast täglich erreichen uns Meldungen über Fortschritte und innovative GenAI-Anwendungen.

Die vielversprechenden Möglichkeiten sind da, dennoch dürfen die Herausforderungen nicht unterschätzt werden. Diese reichen von der Anpassung der Datenstrategie bis hin zu Investitionen in Infrastruktur und Talententwicklung. Unternehmen, die jetzt handeln und diese Technologie strategisch einsetzen, sichern sich einen entscheidenden Wettbewerbsvorteil.

Unsere Erfahrungen im ERGO Innovation Lab zeigen: Die Implementierung multimodaler GenAI-Technologien eröffnet eine Fülle von Möglichkeiten. Unsere Zusammenarbeit mit den Fachbereichen in ersten Prototypen liefert bereits wertvolle Einblicke und zeigt, wie praxisnahe Lösungen aussehen können. Mit diesen Erkenntnissen möchten wir dazu beitragen, die Versicherungsbranche zukunftsfähig zu gestalten. Jetzt ist ein guter Zeitpunkt, die nächsten Schritte zu gehen und die Weichen für nachhaltigen Erfolg zu stellen, indem wir den Einsatz dieser Technologien weiter ausbauen.

Autoren

Georgina Neitzel

Head of Innovation Lab
ERGO Group AG

Nicolas Konnerth

Head of Conversational AI
ERGO Group AG

Danksagung

Dieses Whitepaper ist durch die wertvolle Unterstützung, konstruktive Ideen und das fachliche Feedback zahlreicher weiterer Mitwirkender entstanden. Wir möchten uns daher herzlich bei allen bedanken, die durch ihren wertvollen fachlichen Input, sorgfältiges Korrekturlesen und konstruktives Feedback maßgeblich zur Entstehung und Verbesserung dieses Whitepapers beigetragen haben. Durch Ihr Engagement konnten wertvolle neue Impulse gesetzt und wesentliche Aspekte vertieft werden, was dieses Whitepaper bereichert hat.

Quellenverzeichnis

- 1 Open AI, „OpenAI Spring Update“. <https://openai.com/index/spring-update/> (Zugriff am 28. Mai 2024).
- 2 Google, <https://io.google/2024/> (Zugriff am 28. Mai 2024).
- 3 J. Sievert, N. Konnerth, „ChatGPT und Sprachmodelle: Eine Einführung mit Blick auf die Versicherungsbranche“. <https://www.ergo.com/de/Newsroom/Medieninformation/2023/20230612-Whitepaper-ChatGPT>.
- 4 OpenAI, „Hello GPT-4o“. <https://openai.com/index/hello-gpt-4o/> (Zugriff am 28. Mai 2024).
- 5 Google, „Gemini Pro - Our best model for general performance across a wide range of tasks“. <https://deepmind.google/technologies/gemini/pro/> (Zugriff am 28. Mai 2024).
- 6 OpenAI, „Introducing GPT-4o: our fastest and most affordable flagship model“. <https://platform.openai.com/docs/models/gpt-4o> (Zugriff am 28. Mai 2024).
- 7 OpenAI, „GPT-4 is OpenAI’s most advanced system, producing safer and more useful responses“. <https://openai.com/index/gpt-4/> (Zugriff am 30. Mai 2024).
- 8 OpenAI, „Hello GPT-4o“, <https://openai.com/index/hello-gpt-4o/> (Zugriff am 30. Mai 2024).
- 9 S. Wu et al., „NEX-T-GPT: Any-to-Any Multimodal LLM“, <https://arxiv.org/pdf/2309.05519> (Zugriff am 01. Juni 2024).
- 10 L. Manduchi et al., „On the Challenges and Opportunities of Generative AI“, arXiv:2403.00025, 2024.
- 11 McKinsey, „Leveraging generative AI in Europe: The opportunities and challenges“, <https://www.mckinsey.com/featured-insights/lifting-europes-ambition/leveraging-generative-ai-in-europe-the-opportunities-and-challenges> (Zugriff am 06. Juni 2024).
- 12 McKinsey, „GenAI in der Fertigung angekommen – nächste Welle der Automatisierung mit großem Potenzial“, <https://www.mckinsey.com/de/news/presse/2024-04-22-hannover-messe-genai> (Zugriff am 06. Juni 2024).
- 13 Spiegel Online, „Chatbot verspricht Fluggast irrtümlich Rückerstattung – Airline muss zahlen“, <https://www.spiegel.de/wirtschaft/unternehmen/air-canada-chatbot-verspricht-kunden-irrtuemlich-rueckerstattung-airline-muss-zahlen-a-0af54651-fbb7-4d8f-jb01-862a8e723ac9> (Zugriff am 10. Juli 2024).
- 14 ZDF Heute, „Scarlett Johansson legt sich mit OpenAI an“, <https://www.zdf.de/nachrichten/panorama/scarlett-johansson-chatgpt-stimme-100.html> (Zugriff am 21. Mai 2024).
- 15 bitcom, „Industrie sieht in KI die Zukunft – aber zögert beim Einsatz“, <https://www.bitkom.org/Presse/Presseinformation/Industrie-Zukunft-KI-Einsatz> (Zugriff am 06. Mai 2024).
- 16 Zeit Online, „KI-Anwendungen bedienen laut Unesco Geschlechterstereotype“, <https://www.zeit.de/digital/2024-03/unesco-ki-sexismus-geschlechter> (Zugriff am 06. Mai 2024).
- 17 Financial Times, „Arup lost \$25mn in Hong Kong deepfake video conference scam“, <https://www.ft.com/content/b977e8d4-664c-4ae4-8a8e-eb93bdf785e9> (Zugriff am 28. Mai 2024).
- 18 Spiegel Online, „Fake News war gestern. Die Ära der Fake Reality beginnt“, <https://www.spiegel.de/netzwelt/netzpolitik/tiktok-plus-ki-fake-news-war-gestern-die-aera-der-fake-reality-beginnt-sascha-lobo-kolumne-a-d358aa2e-86cb-49cd-9317-62c23f97a6cf> (Zugriff am 29. Mai 2024).
- 19 McKinsey, „As gen AI advances, regulators—and risk functions—rush to keep pace“, <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/as-gen-ai-advances-regulators-and-risk-functions-rush-to-keep-pace> (Zugriff am 28. Mai 2024).
- 20 pwc, „Trusted AI: Ethics and responsible AI in Generative Systems“, <https://www.pwc.de/en/digitale-transformation/generative-ai-artificial-intelligence/the-genai-building-blocks/the-ethical-imperative-in-ai-development.html> (Zugriff am 28. Mai 2024).
- 21 Harvard Business Review, „Navigating the New Risks and Regulatory Challenges of GenAI“, <https://hbr.org/2023/11/navigating-the-new-risks-and-regulatory-challenges-of-genai> (Zugriff am 28. Mai 2024).
- 22 Gartner, „Generative AI Can Democratize Access to Knowledge and Skills“, <https://www.gartner.com/en/articles/generative-ai-can-democratize-access-to-knowledge-and-skills> (Zugriff am 29. Mai 2024).
- 23 Google, „Generative AI in Search: Let Google do the searching for you“, <https://blog.google/products/search/generative-ai-google-search-may-2024/> (Zugriff am 30. Mai 2024).
- 24 Washington Post, „AI is changing Google search: What the I/O announcement means for you“, <https://www.washingtonpost.com/technology/2023/05/10/google-search-ai-io-2023/> (Zugriff am 30. Mai 2024).
- 25 Shana Lynch, „What to expect in 2024“, <https://hai.stanford.edu/news/what-expect-ai-2024> (Zugriff am 28. Mai 2024).
- 26 McKinsey, „The economic potential of generative AI The next productivity frontier“, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier> (Zugriff am 30. Mai 2024).
- 27 A. Velinov, „How Multimodal AI Will Transform Human-Machine Interaction Forever“, <https://www.linkedin.com/pulse/how-multimodal-ai-transform-human-machine-interaction-alex-velinov-t5kff/> (Zugriff am 30. Mai 2024).
- 28 W. Knight, „Google DeepMind’s Demis Hassabis Says Gemini Is a New Breed of AI“, <https://www.wired.com/story/google-deep-mind-demis-hassabis-gemini-ai/> (Zugriff am 30. Mai 2024).
- 29 Forbes, „The Future Of Generative AI: 6 Predictions Everyone Should Know About“, <https://www.forbes.com/sites/bernard-marr/2024/03/05/the-future-of-generative-ai-6-predictions-everyone-should-know-about/> (Zugriff am 29. Mai 2024).
- 30 Forbes, „Sensing Success: OpenAI, Anthropic And 40+ Others Leverage Multimodal AI“, <https://www.forbes.com/sites/digital-assets/2024/05/28/sensing-success-openai-anthropic-and-40-others-leverage-multimodal-ai/> (Zugriff am 28. Mai 2024).
- 31 Board of Innovation, „From design thinking to Autonomous Innovation“, <https://www.boardofinnovation.com/blog/introducing-autonomous-innovation/> (Zugriff am 29. Mai 2024).