

Einfach, weil's wichtig ist.

ERGO

The Next Generation of Generative Artificial Intelligence

Progress and Multimodal Applications
in the Insurance Industry



„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, Bill Gates, co-founder of Microsoft

In the ERGO Innovation Lab, our daily work focuses on discovering new technologies, derive potential use cases from and rigorously assess their value for the entire Group. Generative Artificial Intelligence (GenAI) has rapidly established itself as a transformative technology on a global scale. What was once the exclusive domain of experts is now accessible to a much broader audience, thanks to rapid technological advancements. This shift is fundamentally changing how we interact with technology, opening up entirely new possibilities for using complex systems in a simple and effective way. At ERGO, GenAI has now also become an integral part of our operations. The first pilot projects, such as our internal ERGO GPT, have been successfully implemented and are now scaling up – more than 30,000 of our employees have access to this technology.

Simultaneously, GenAI technology has evolved at a rapid pace. Many of today's models are no longer limited to text-based inputs and outputs; they now process diverse data types—such as images and audio—concurrently and at impressive speeds. Milestones like the OpenAI Spring Update and Google I/O in May have highlighted new possibilities for the insurance industry. At the ERGO Innovation Lab, we actively track these advancements and have already tested multimodal GenAI technologies across various use cases. This has

enabled us to swiftly gain valuable insights, which are now being integrated throughout the ERGO Group.

The insurance industry has the potential to benefit from a range of applications, such as enhancing the customer experience by streamlining claims processing through photo and video analysis for loss assessment. Beyond this, these advancements are driving the optimisation of internal processes. In practice, AI goes beyond generating letters or emails; it can analyse data from multiple sources, complete forms, and address customer inquiries more efficiently, reducing routine workloads. Additionally, the technology strengthens risk assessment by detecting potential weather-related risks early, providing valuable insights for informed decision-making. Strategic decisions also benefit from multimodal GenAI, enabling faster, data-backed market analysis to identify new business opportunities.

In our previous white paper, 'ChatGPT and speech models: An introduction with reference to the insurance industry' (May 2023)[1], we examined basic applications of unimodal, text-based speech models. This white paper investigates capabilities and functionalities of current GenAI models, with particular focus on multimodality. We illustrate how this technology can be applied within the insurance sector, while exploring both its opportunities and challenges.

Table of Contents

The Next Evolutionary Stage of Artificial Intelligence	4
An insight into the way multimodal generative AI functions	7
Potential Applications of Multimodal GenAI in the Insurance Industry	10
Challenges and Limitations of New Generative AI Models	19
Outlook: Why the Insurance Industry Must Take Action Now	21
References	24

The Next Evolutionary Stage of Artificial Intelligence

A new generation of generative AI models (GenAI models) is ready to fundamentally change our world. It is not only the remarkable speed and performance of these models that stand out, but also their ability to seamlessly handle multiple modalities—such as text, images, audio, and video—at the same time. This section delves into the key capabilities of modern GenAI models, highlighting their advanced context processing, enhanced speech comprehension, and groundbreaking innovations within each modality.

A uniform model for seamless multimodal interactions

Modern GenAI models can process multiple input modalities—such as text, images, audio, and video—within a single system, a capability that was previously unattainable⁴. This enables connections between modalities, leading to precise, context-rich results. Multimodality reduces information loss and simplifies interaction. Moreover, significant advancements have been made in the processing of individual modalities, which we will explore briefly below.

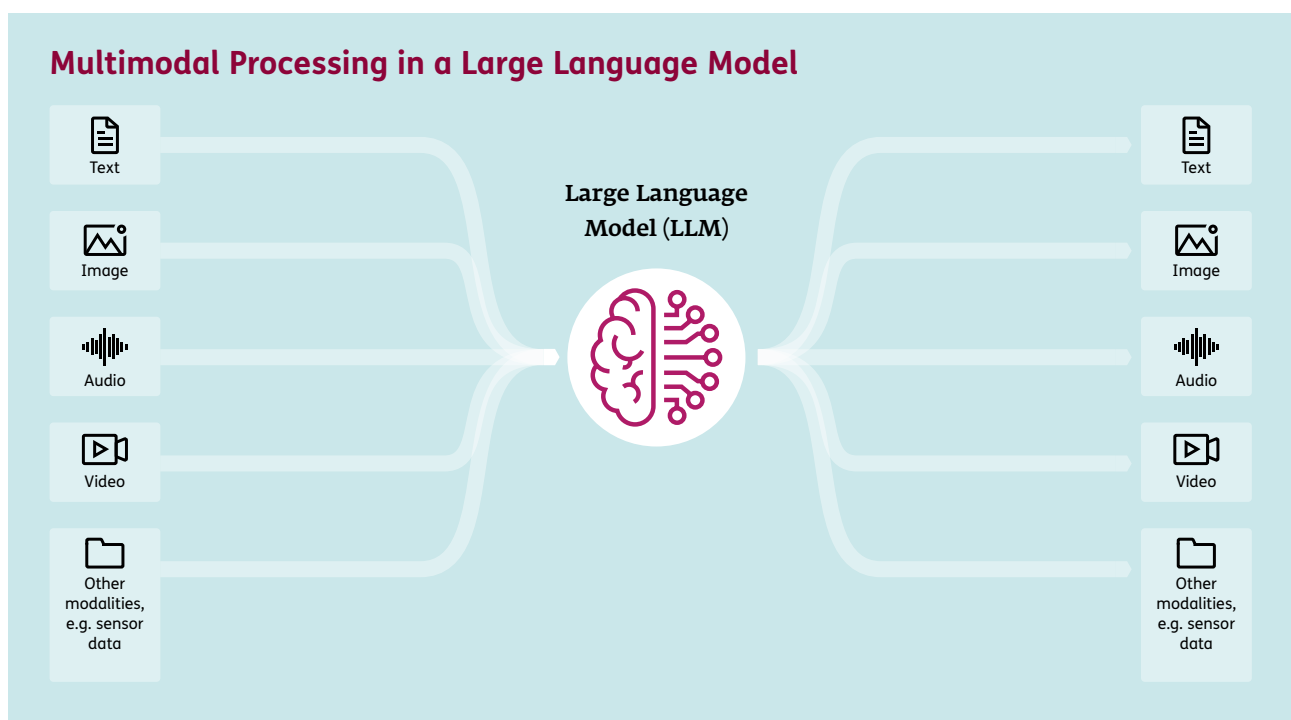
Developments in the area of different modalities

Recent advancements in **text processing** have greatly expanded the capabilities of GenAI models. They now

comprehend not only continuous text but also complex formats, including tables, diverse document types (e.g., PDFs, Word files, presentations), forms, HTML files, and other web content, as well as mathematical and scientific formulas. This versatility makes them highly adaptable for a wide range of applications.

Modern architectures also incorporate technologies such as Retrieval-Augmented Generation (RAG), to search for relevant information in databases or in the internet, providing models with relevant data to formulate more accurate responses.

Speech processing and synthesis capabilities in GenAI models have advanced considerably. The ‘Barge-in’ function allows users to interrupt and adapt speech interactions in real time, for example, to change the narration style or voice. This flexibility makes communication more natural and intuitive, improving the user experience. The models also offer multilingual support with seamless language switching, and the ability to clone specific voices and humanize speech enables more lifelike interactions. By adjusting audio elements like pitch, tempo, and emotion, users can create the desired atmosphere, tailoring interactions to specific contexts and applications. This approach enhances the model's ability to generate context-aware answers, even for complex queries.



⁴ RAG (Retrieval-Augmented Generation): Combines content from external sources with GenAI to create context-based results.

Modern GenAI models unlock new possibilities in image processing by combining analysis and generation within a single system. This allows for seamless integration of text into images, such as automatically generating subtitles or captions. Image generation has also been refined for greater precision, enabling users to provide detailed instructions on aspects like color, composition, and lighting. Additionally, many GenAI models include post-processing tools for further refinement of generated visuals. This is particularly advantageous in scenarios where realistic product images or compelling visuals are crucial.

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

Demis Hassabis, CEO, DeepMind

In **video processing**, quality and realism of AI-generated content has significantly improved. Current GenAI models can generate longer, more complex video formats with smooth transitions and improved visual quality. Similarly to image generation, content in video generation can be produced with high precision by defining specific parameters such as camera angle, lighting, and effects. This opens new possibilities in media production, advertising, customer communications and wherever visual content forms are needed.

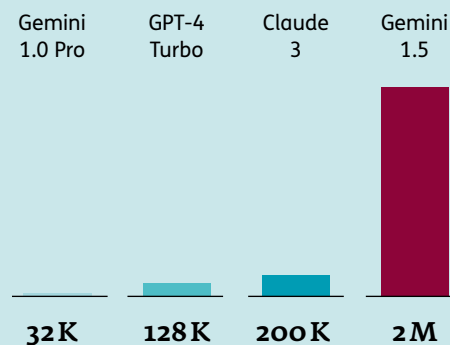
In addition to processing text, image, video, and audio data, multimodal GenAI models can also handle less common formats. These include **sensor and motion data**, which is, for example, particularly important in the insurance industry for loss prevention, or in the detection of faulty water pipes. However, these are special use cases and therefore outside the scope of this white paper.

Expanded context window

The context comprises all the information available to AI for generating its responses, including user input, additional context information and previous dialogues. The size of the context window plays a crucial role in the performance of a GenAI model. A larger context window enhances the coherence and relevance of responses by drawing on more extensive information and details from previous interactions.

Enhanced storage capacity allows GenAI models to recall content from previous dialogues and personalise their

Context Window Size of various GenAI Models (in Token)⁷



Context window size is usually stated in tokens. One token represents one unit of text, e.g. a word or a character. Current GenAI models such as Google Gemini 1.5 Pro support expanded context windows of up to two million tokens⁵ while GPT-4o can process 128,000 tokens⁶ and allow GenAI models to process vast amounts of data simultaneously. This is equivalent to two hours of video, 22 hours of audio, or up to 1.4 million words. These extensive context windows allow a vast amount of content to be processed and analysed simultaneously, including long documents, videos, and audio sequences.

responses. In this way, it contributes to improved communication consistency and better customisation by continuous adaptation to individual user characteristics. This enables efficient, human-like and empathetic interactions.

Human response times and enhanced interactivity

One of the most striking features of modern multimodal GenAI models is their ability to react to users almost in real time. With an average latency of 232 milliseconds, these models react faster than humans⁸. This enables fluent, natural communication with the AI and improves the customer experience significantly, particularly in longer, interactive dialogues.

A further function, which takes advantage of this fast response time is the 'Barge-in' function. As outlined above in the section on audio modality, this function allows users to interrupt and adapt an ongoing speech interaction. This adaptation capability is particularly important in scenarios requiring smooth integration of human and artificial intelligence, such as customer service. In addition to the processing time of the GenAI model, latency due to the validation of the generated content also plays a crucial role.

„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

Improved language comprehension and multilingual capabilities

The latest generation of language models exhibits exceptional language comprehension, performing strongly against established benchmarks like the Massive Multitask Language Understanding (MMLU). MMLU evaluates their ability to tackle complex linguistic tasks, ranging from simple text processing to advanced comprehension challenges. High scores on these benchmarks underscore the capacity of GenAI models to handle a wide array of linguistic complexities.

Moreover, modern models support multiple languages and can switch between them effortlessly. This functionality is especially valuable for companies operating in multilingual markets⁸, as it facilitates seamless communication across language barriers.

An insight into the way multimodal generative AI functions

Gaining a clear understanding of how multimodal Generative AI (GenAI) functions begins with grasping its foundational mechanisms. Though these concepts may appear intricate at first, a simple analogy can shed light on the underlying principles.

Picture, for instance, an opera performance:

- **Image:** On the stage, a singer in an elaborate costume moves around an extravagant stage which reflects the mood of the storyline.
- **Audio:** The singer performs an emotive aria accompanied by the orchestra, heightening the scene's drama.
- **Text:** Subtitles, which translate the text of the song so that the audience can follow the storyline more easily.

Each modality contributes a unique element: the image provides the visual impression, the soundtrack enhances the scene's drama, and the text ensures understanding for non-native speakers. A multimodal AI model combines these elements by learning how image, audio and text fit together to make a complete scene. The AI can then later generate similar content or precisely describe what is taking place on stage.

This interplay of modalities highlights the importance of multimodal GenAI understanding these connections. But how exactly does multimodal GenAI work?

How does multimodal GenAI work? A simple explanation in five steps

Step 1: Data collection and pre-processing

The AI collects text, images, audio files, and videos. In the first step, data cleansing removes irrelevant elements from the raw data, such as automatic signatures at the end of emails. For texts, this might be irrelevant text parts, such as automatic signatures at the end of emails.

After cleansing, the data is standardized into a uniform format. For text, AI employs specialized techniques such as Word2Vec or BERT to identify key terms and meanings, a process known as "tokenization." In contrast, images

undergo procedures like scaling to a consistent size or adjusting color values to ensure uniform brightness and contrast. Audio data is divided into frequency and time segments, often by converting it into spectrograms, allowing patterns in sound to be detected during further analysis. As for video, it is first separated into visual and audio components. The visual frames are processed similarly to static images, while the audio track is handled in the same way as individual audio files.

Step 2: Feature extraction and data representation

Once the data has been pre-processed, **feature extraction** takes place, during which the AI identifies key characteristics within the data and converts them into numerical representations known as **embeddings**. These embeddings serve as a universal framework across different modalities, allowing the AI to process and analyze diverse data types while uncovering intricate relationships and patterns.

For this process, specialized techniques are employed for each modality:

- In the case of **texts**, AI keywords and meanings are identified through specialised processes such as **Word-2Vec^b** and then transformed into a multidimensional space. Words with similar meanings, or ones which frequently occur together, are positioned close to each other in this multidimensional space. For instance, in this space, words like "ice" and "cold" may be positioned closely because they frequently occur in similar contexts. Advanced models like **BERT** or **GPT^c** go a step further by interpreting the meaning of words within the context of entire sentences or paragraphs, enabling a deeper understanding of their nuances. This allows for more comprehensive and sophisticated analysis.
- For **images**, convolutional neural networks (**CNNs**)^d are employed to extract visual features such as shapes and colors. For instance, AI can identify the shape and color of a car within an image and translate this visual information into numerical sequences that represent the image, enabling the AI to process and understand it.

^b Word2Vec (Word2Vector): Technology that converts words into numerical sequences, in order to capture their meaning and the context.

^c GPT (Generative Pre-trained Transformer): An advanced language model which generates human-like text, as it comprehends the context and meaning of words within sentences and paragraphs.

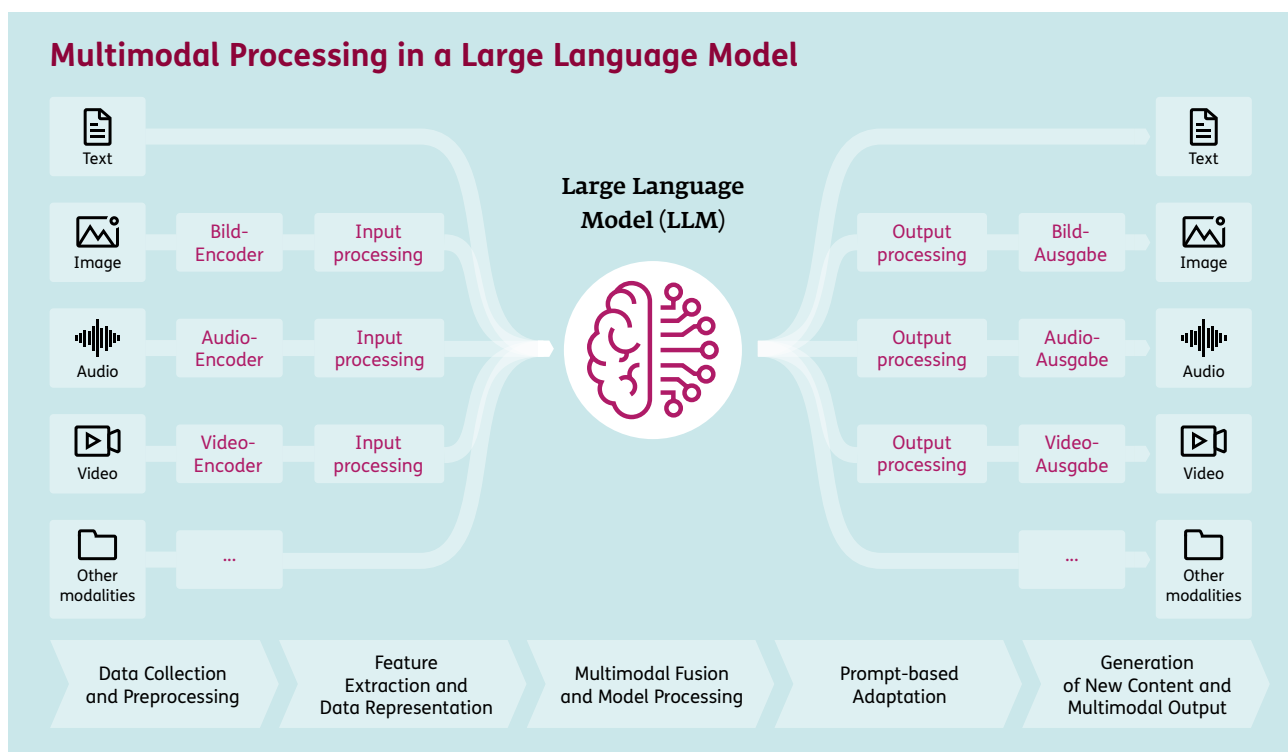
^d CNN (Convolutional Neural Network): A special type of artificial neuronal network specialised in processing image data.

- **Audio data** is transformed into spectrograms, which visually depict frequencies and volume over time. Techniques like **MFCCs^e (Mel-Frequency Cepstrum Coefficients)** then enable AI to identify key acoustic features such as pitch and volume, converting them into numerical sequences. For example, this process allows AI to analyze the pitch and volume of a musical instrument, capturing its unique sound patterns.
- For **videos**, AI integrates both visual and audio data, employing techniques such as **recurrent neural networks (RNNs)^f** to analyze temporal sequences. This allows AI to interpret dynamic patterns, such as gestures and speech in a video conference, capturing the flow and interaction over time.
- **Early fusion:** In this approach, data of all modalities is merged immediately after capture and processed together. It is particularly suited for real-time applications like safety monitoring, where rapid data processing is critical.
- **Intermediate fusion:** Here, fusion happens during the processing phase, after partial analysis of the data. This method is commonly applied in communication services, allowing AI to extract subtle cues, such as mood or intent, through simultaneous analysis of text, tone, and visual signals.
- **Late fusion:** In this stage, each data type is analyzed separately, and the results are combined at the end. This method is often used in scientific research, where detailed analysis of each data source is essential for a thorough and precise evaluation.

Step 3: Multimodal fusion and model processing

After extracting relevant features from various data types and converting them into a universal representation, the AI merges them through a process called multimodal fusion, combining information to enhance analysis and decision-making. Fusion occurs in three stages: early, intermediate, or late fusion:

The choice of fusion method depends on the application area and also on whether the focus is on rapid processing (early fusion), in-depth context analysis (intermediate fusion) or precise individual analysis (late fusion). Each approach offers distinct advantages, allowing systems to be tailored to specific performance requirements.



The processing of different formats through a Large Language Model (LLM)⁹ (modified)

^e MFCC (Mel-Frequency Cepstrum Coefficients): A technology for the analysis and representation of audio frequencies.

^f RNN (Recurrent Neural Network): A neuronal network for the processing of sequential data by utilising prior information, ideal for analysis of text, audio and video.

In addition, transformer models and attention mechanisms play a key role. They enable the AI to subsequently filter out the relevant information, such as key scenes or core messages, from the merged data sources and carry out specific analyses.

Step 4: Prompt-based alignment

At this stage, the fine-tuning takes place in the data processing based on specific user instructions (prompts). Based on these prompts, GenAI deduces which information is particularly relevant and how this should be represented, for example:

- **Content format alignment:** Depending on the required output modality, the AI can generate text, images, audio or video in a specific format
- **Stylistic alignment:** The AI can further adapt the style of the output to correspond to the prompt specification, such as the tone of the content in the case of text (e.g. formal, informal, technical, creative)

- **Content-related specification:** The AI delivers precise and detailed information based on specific questions or instructions
- **Customisation:** The AI offers individually customised content or recommendations which are tailored to the requirements or preferences of the user

Step 5: Generation of new content and multi-modal output

Following the analysis and fusion of the data, the AI uses special models to generate new content. Two frequently used models are **variational autoencoders (VAEs^g)** and **generative adversarial networks (GANs^h)**. VAEs create new content by recognizing patterns in the existing data and generating similar outputs based on these insights. GANs, however, operate with a unique dual structure: a generator that produces new data, and a discriminator that evaluates its authenticity. This dynamic interplay between the two leads to highly realistic and refined generated content.

^g Variational Autoencoder (VAE): A machine learning model which aims to analyse complex data and to produce similar new data, by learning the essential features of the original data.

^h Generative Adversarial Network (GAN): A machine learning model, consisting of a generator, which generates new data, and a discriminator, which checks the authenticity of this data, in order to improve the quality of the content generated.

Potential Applications of Multimodal GenAI in the Insurance Industry

This section offers an overview of potential applications for multimodal GenAI models in the insurance industry. We begin by examining the core functions shared across all modalities, followed by a focus on the practicalities of effective prompting. Specific use cases will be presented to illustrate these core functions in action.

This includes a focus on specific application areas within the insurance industry, such as the creation of competitive analyses, video-based estimation of household insurance values, and the development of multimodal content. The goal is to demonstrate how multimodal

GenAI models can enhance internal processes while also creating new opportunities for customer interaction.

Core functions of multimodal GenAI

The table below highlights the diverse core functions enabled by multimodal GenAI models, showcasing the extensive capabilities achieved through the integration of various data modalities—such as text, images, and audio—within the insurance industry. This overview provides a foundation for understanding the vast potential of multimodal GenAI and its wide range of applications.

Core functions of multimodal GenAIⁱ

Core function	Description	Example
Creation	The generation of new content, e.g. text, images, audio and videos	Automated creation of marketing materials such as TikTok videos, automated content creation for blogs
Simulation	Simulation and imitation across modalities	Voice imitation or mimicking visual styles
Transfer	Transferring information between modalities, e.g. image-to-text transformation	Transferring images into written image descriptions
Modification	Adaptation and modifying of existing content	Customising marketing content based on customer segments
Improvement	Optimising quality and style	Enhancing a mock-up based on specific design principles
Reduction	Summarising and condensing information	Generating summaries from long videos or transcribing audio to text
Analysis	Analysing and evaluating content	Verifying whether text descriptions match photographs of damage
Synchronisation	Aligning content across modalities	Creating audio content for a video sequence
Organisation	Structuring and organizing information	Generating to-do lists or schedules
Explanation	Clarifying complex issues	Simplified explanations for non-specialists
Prediction	Analysing patterns to predicting trends	Forecasting market trends
Customisation	Adapting content to individual needs	Personalized diet or fitness plans

ⁱ This table expands on the core functions outlined in the previous whitepaper (ChatGPT und Sprachmodelle: Eine Einführung mit Blick auf die Versicherungsbranche) and has been further refined to include aspects of multimodality.

Executing core functions through effective prompting

After exploring the core functions of multimodal GenAI models and their potential, this section addresses the practical aspects of effective prompting.

Well-structured prompts that consider at least role, task definition, context, greatly enhance the accuracy of AI-generated outputs. Furthermore, each modality has unique parameters, as detailed in the table below.

At the ERGO Innovation Lab, we've observed that initial prompt results often require further refinement to meet desired specifications. An iterative approach is frequently necessary, where users provide feedback to the AI to adjust the output. For instance, after receiving the initial result, a user might request, 'Please use a more formal tone,' prompting the AI to revise its response accordingly.

This iterative process has proven essential for achieving precise, high-quality outcomes.

Practical application examples in the insurance industry

Drawing on our ongoing research and extensive expertise at the ERGO Innovation Lab, we now explore practical applications of multimodal GenAI in the insurance industry. There are numerous opportunities for its deployment, with the potential to enhance process efficiency, optimise customer service, and foster innovative product development. The following prompt examples illustrate how GenAI can be strategically applied to achieve these goals. **For clarity, the prompts have been distilled to their essentials and the outputs significantly condensed, allowing a focused demonstration of the core functions and the technology's capabilities.**

Overview of relevant prompt elements and parameters, grouped according to output modality^j

Prompt element	Description	Text	Image	Audio	Video ^k
Role definition	Defines the perspective from which the AI provides responses	Expert role (marketing expert, strategy consultant), style (formal, informal), tone (e.g. factual, emotional), communication style (e.g. narrating, argumentative)	n/a	Speaker type (e.g. age, gender), style (soothing, enthusiastic), pitch	n/a
Task clarification	Specifies the core function and details the objective	Text type (e.g. press release, report, storyline), structure, aim (e.g. inform, persuade)	Style (e.g. modern, traditional), perspective (e.g. subjective, objective), image type (e.g. photography, illustration), degree of detail (e.g. sketchy, detailed)	Representation of specific scenes which are described in the script	Representation of specific scenes which are described in the script, style (e.g. animation, live action),
Context specification	Adjusts output to suit the target group or context	Target group (e.g. customers, employees), context (e.g. training, marketing)	Details of environment (e.g. indoors, outdoors), focus (e.g. product, person), lighting (e.g. natural, artificial)	Ambient sounds (e.g. street, nature), recording location (e.g. studio, on location), speaker's role (e.g. narrator, character)	Camera perspective (e.g. subjective, objective), camera work (e.g. static, dynamic), image detail (e.g. close-up, long shot), lighting (e.g. dramatic, natural), dynamic (e.g. calm, hectic)
Output format	Specifies the format type and structure	Text structure (e.g. continuous text, list, table, code), style (e.g. narrative, descriptive)	Output size (e.g. poster, thumbnail)	Spoken speech (voice-over, description etc.), music, sound effects, sounds	Format (e.g. 16:9, square), post-processing (e.g. transitions, colour filters, effects)

^j This list illustrates the prompt elements and parameters as examples and does not claim to be exhaustive.

^k In this table, the focus when considering videos is on the visual aspects of the moving images, and not on the soundtrack.

Analysis of textual and visual web content
and subsequent report generation

Competitor analysis for a new motor insurance policy



You are an experienced market analyst in an insurance company. Draw up a brief competitor analysis for insurance policies. The report should describe the major competitors, briefly describe their product range and pricing and contain a recommendation for the positioning of the new motor insurance policy on the market. Use publicly available information and competitors' branch reports and websites. The report should contain an introductory continuous text and the competitors' characteristics, as well as suggestions for the new product positioning, in the form of a list.

Role Task Context Format



Brief report on the competitor analysis for motor insurance policies

Section 1: Market overview • Text: Short introduction to the market for motor insurance policies, including the key trends and market players.

Section 2: Competitor analysis. Brief description of the major competitors, their product range and pricing.

- **Competitor 1:** Overview of the main features of their products and price structure
- **Competitor 2:** Overview of the main features of their products and price structure
- **Competitor 3:** Overview of the main features of their products and price structure

Section 3: Recommendations on positioning. Brief recommendation on the positioning of the new motor insurance policy, based on the analysis of the competitors.

- **Suggestions for product differentiation**
- **Suggestions for a pricing strategy**
- **Suggestions for marketing approaches**



For a thorough analysis, it is advisable to refine the prompt by specifying relevant competitor features, sources, and clearly outlining the expectations for the new positioning.

Video analysis and report generation

Video-based consultation for assessing the insured value of household contents



You are a household contents insurance specialist. The attached video contains a recording of a flat. Please list in abbreviated form the relevant items in the flat and estimate the value of each object. Make a note also of items that are generally found in a household, but which are not pictured in the video. Recommend a plausible amount insured based on the video, and assuming that the other items are also in the flat.

Files: Video of a flat tour

Role
 Task
 Context
 Format



Description of the items in the flat:

1. **Television (approx. 55-inch):** €800
2. **Television stand:** €150
3. **Sofa (corner sofa):** €1,000
4. **Coffee table:** €200
5. **Dining table with 4 chairs:** €400
6. **Computer (monitor, keyboard, mouse):** €600
7. **Display cabinet with glasses and decoration:** €300
8. **Air purifier:** €100
9. **Standard lamp:** €50
10. **Decorative items, (e.g. model aeroplane, cuddly toys):** €100
11. ...

Estimated value of the items in the flat:

Total value: 9.700 €

Other items which are generally found in a household:

1. **Kitchen:**
 - Microwave: 100 €
 - Kitchen utensils and crockery: 300 €
2. **Bedroom:**
 - Bedside table: 200 €
3. ...

Total value of the additional items: 2.300 €

Recommended insured amount for the whole household: **€12,700**



To generate more accurate price estimates, specific brands and models of items shown in the video can be identified. By recording and cataloging these details, precise market values can be assessed, resulting in a more exact determination of the insured amount.

UX/UI analysis of websites

Evaluation of the user-friendliness and design of a landing page and improvement measures derived from this



You are an experienced UX/UI designer and you have been tasked with analysing the landing page of an insurance company (www.fiktive-versicherung.de) and making suggestions for improvements derived from this. The aim of the analysis is to evaluate the user-friendliness and the visual design of a landing page and develop practical suggestions for improvement. The landing page is used by existing and potential customers to obtain information about insurance policies, notifications of claims and the products and services offered. Set out the results of the analysis and the suggestions for improvement in the form of a list.

Role Task Context Format



Navigation structure: The landing page navigation has a solid basis, in that the main menu items are clearly listed. This helps users to find their way around the site. However, accessibility could potentially be improved, in particular for users who want to quickly find specific information or functions. One or two adjustments could significantly enhance the user experience.

Visual design: The current design of the landing page is successful in achieving an impression of a trustworthy company. This is important for strengthening customer trust. Updating the colour palette would ensure that the design better reflects the insurance company's brand identity and would at the same time improve readability and visual appeal.

Structure of content: The information on the landing page is comprehensive and gives visitors all the key details. Restructuring and better formatting of the text could make the information even easier to access, enabling users to find the information that they are looking for more quickly (...)

Measures to improve the UX/UI:

- 1. Improvements to the navigation structure:** Introduction of 'Quick Links' to the most frequently used functions and information, such as 'Compare Policies' or 'Request a Quote', directly on the home page or in the main menu. This would make it easier for users to access the key areas and make the site more user-friendly.
- 2. Updating the visual design:** Aligning the colour palette and fonts with the insurance company's corporate identity. This could act to refresh the visual elements, make the landing page appear more modern and unified and, at the same time, strengthen the brand identity.
- 3. Implementation of a clear information hierarchy and visual structure:** To improve the accessibility of the comprehensive information on the landing page, a clear information hierarchy should be established. This can be achieved by the introduction of headings, subheadings and bullet points (...)



Many GenAI models are already able to access websites in real time and, for example, analyse and compare competitors' websites, or examine SEO aspects such as keywords, meta tags and website architecture, in order to improve visibility and ranking in search engines.

Creation of image material

Generation of images for designing a landing page for nursing care insurance



Create a photo-realistic image in landscape format with the following content: An elderly person with wrinkles, wearing a cosy jumper, is sitting on a soft sofa in a living room. A carer in uniform is sitting next to her, holding her hand and talking to her in a friendly manner, while the other lady smiles happily and looks attentive. The room is comfortably furnished with a large couch, a wooden coffee table, a bookcase and several potted plants. Sunlight is streaming through the window and bathing the living room in a warm, atmospheric light, giving a feeling of care and security.

Task Context Format



Detailed descriptions generally result in significantly better images than short prompts. Numerous tools are now available to help optimise and enhance prompts for image creation, to achieve better image quality.

Some image generators, such as Midjourney, support the use of specific parameters at the end of a prompt, thereby enabling more precise control of the image design. These parameters are prefaced with two dashes, followed by their name and value: --ar (aspect ratio), --stylise (degree of stylisation), --weird (degree of surreality), --no (elements or styles to be excluded).

Creation of audio content

Creation of a post for a mini-podcast series on insurance information

A concept and a clear topic structure are required to create a podcast. Each episode will be prepared by means of an individual prompt, and the script generated will then be set to music using text-to-speech technology.

Topics, script and background music can also be developed by GenAI models, with the whole production process being executed efficiently and cost-effectively.



You are an insurance expert and host a podcast which provides listeners with information about different insurance topics. Create a script for a podcast episode which provides an answer to a frequently asked insurance question. The episode should be written in basic English which is easy to understand, and should contain practical tips or recommendations to help listeners make informed decisions. The episode should be 8-10 minutes long. Deliver a complete script for the podcast episode which gives a clear introduction to the topic, explains relevant examples and scenarios and, at the end, provides practical tips or recommended actions.

Role Task Context Format



Podcast episode: The advantages of your third-party insurance policy and current offers

'Welcome to today's episode of 'Insurance Made Easy'! Today, we're going to talk about a topic which is relevant for many of us: third-party insurance. What exactly is this, and why should we all have it? Imagine that, while visiting your aunt, you break a valuable vase. Or your child accidentally spills blackcurrant juice on the carpet at your friend's house. These are scenarios where third-party liability insurance comes into play...'

Section 1: Introduction and the meaning of third-party insurance

'We'll begin by talking about what third-party insurance actually is. Third-party insurance is one of the most important insurance policies for private individuals. It protects you from financial consequences in cases where you might unintentionally have caused damage to another person, whether damage to property or physical injury. In Germany, a person who causes damage to another is liable under law. This means that, in the worst-case scenario, you might be liable and have to pay very high costs. This is where third-party insurance comes in (...)'



Google LM Notebooks can be used to automatically generate and synthesize podcast scripts, allowing language models to analyze topics and create relevant content. This simplifies the production of well-structured and informative episodes with minimal effort.

Creation of marketing videos

Creating a Video Scene for a Pet Health Insurance Marketing Campaign

To develop a complete marketing video, the first step is to create a concept and storyline. Each scene of the film will be described with a specific prompt, requiring multiple prompts for the full video creation. Once all the scenes are generated, they can be assembled into the final marketing video. Some models also allow for simultaneous scene creation, including transitions, making the process even more efficient. The generated scenes can then be dubbed with audio as needed.

The storyline, script, and audio dubbing for the marketing video can also be produced by additional GenAI models.

This means that the entire production process—from concept development to the final campaign video—can be efficiently handled using cost-effective GenAI technologies.

The example below was created using the AI video platform Runway, demonstrating a scene for pet health insurance for dogs. Two different approaches for GenAI video generation are illustrated: text-to-video and text-to-image-to-video.

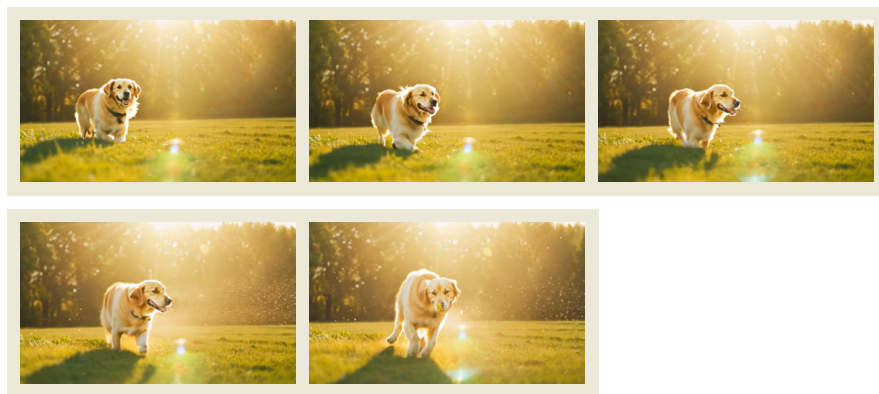


Prompt version 1 (text-to-video):

Wide-angle camera, slow motion: A golden retriever runs through a park on a bright summer's day

Role Task Context Format

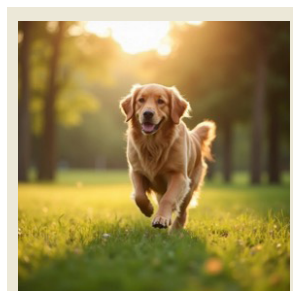
Output:



Prompt version 2 (text-to-image-to-video):

A golden retriever runs through a park on a bright summer's day

Output:





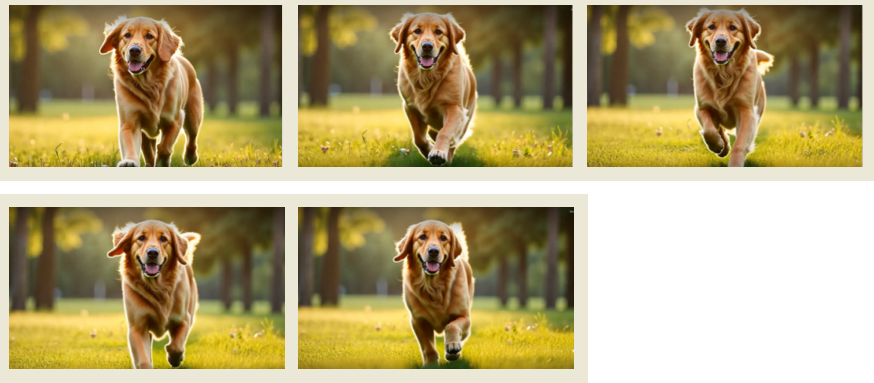
Prompt 2.2 (image to video):

Wide-angle camera, slow motion: A golden retriever runs through a park on a bright summer's day

Role Task Context Format



Output:



Our work at the ERGO Innovation Lab has demonstrated that video quality improves when not relying solely on text prompts. Instead, combining high-quality, pre-generated AI images with text prompts produces significantly better results (as shown in prompt version 2).

In our experience, video generators often face challenges in depicting natural, non-verbal interactions between individuals. Moreover, inconsistencies frequently emerge in the portrayal of physical elements, such as the accurate rendering of shadows and light reflections.

Challenges and Limitations of New Generative AI Models

In the preceding sections, we outlined multiple promising applications of multimodal GenAI. However, these new technologies come with challenges, ranging from technological hurdles to regulatory, legal, trust-related, ethical, and societal issues^{10, 11, 12}.

Technological challenges

Data accuracy and ‘hallucinations’

Like all GenAI models, multimodal GenAI can generate “hallucinated” content - incorrect or invented information that is not present in the training data. This can lead to false or unexpected responses, which are particularly problematic in sensitive domains such as insurance. For instance, a court of arbitration in British Columbia recently ruled against Air Canada, after a chatbot provided incorrect information about airline ticket refunds¹³. Thus, it is crucial that generated content is validated, by a human (a ‘human in the loop’) or through established regulations and hallucination detection tools.

Processing power and costs

Multimodal data processing requires substantial computational power, which can lead to higher costs. Companies therefore need to carefully plan how they access the models and how they intend to use them. Independent model training remains expensive. A more economical alternative is to use pre-trained models, mainly in cloud-based services. These enable flexible scaling of the required processing power, without costly investment in hardware.

The cost of using pre-trained, cloud-based AI models has decreased significantly, with providers like OpenAI reducing prices and offering mini versions which are more cost-effective and resource-efficient. Companies should carefully calculate processing costs and explore the most effective options for scaling AI usage in a cost-efficient and resource-conscious manner.

Regulatory and legal aspects

Copyright aspects

Copyright issues in multimodal GenAI, where image and audio elements are combined, pose unique challenges, as several copyright lines can be affected or infringed. Direct comparison between AI-generated content and original

data is often difficult, leading to legal uncertainties. For instance, this applies to images containing elements protected under copyright, or to audio content where voices are imitated—such as in the case of OpenAI, which used a voice resembling that of Scarlett Johansson¹⁴. To avoid infringement, it is therefore pertinent to carry out careful legal checks when using multimodal GenAI systems. Some providers, like Adobe, also offer an IP release protections for AI-generated content as their models are trained on proprietary or copyright-free data.

Regulation and compliance

The use of multimodal GenAI is increasingly subject to regulatory requirements, particularly in high-risk sectors like insurance. The EU AI Act outlines strict standards for the use of AI in areas such as risk assessment and claims processing, requiring compliance with transparency, fairness, and data protection rules.

Transparency and user acceptance

Explainability and trust

The lack of transparency in AI decisions represents a significant challenge in industries where trust and transparency are of paramount importance, such as insurance. To address this issue, it is essential to equip models with Explainable AI (XAI) technologies. These technologies facilitate a clear understanding of the decision-making process, enabling users to comprehend how GenAI models arrive at their recommendations. This, in turn, enhances trust and user acceptance. It is crucial to display the input parameters that inform a decision and to illustrate how alterations in these parameters could influence the outcome.

Acceptance and training

Introducing GenAI solutions requires not only technological, legal, and safety considerations but also employee buy-in. The success of multimodal GenAI in organizations depends on user-friendly interfaces and focused training programs. A study by Bitkom¹⁵ found that 48% of German industrial companies see a lack of expertise as the primary barrier to adopting GenAI. Companies should implement structured training programs to help employees build skills and understand how to effectively use these technologies, fostering a culture of continuous learning and adaptation..

Ethical and societal implications

Prejudices and fairness

AI models can adopt biases from the training data, leading to unfair or biased decisions, as UNESCO research has shown¹⁶. This includes gender stereotypes or racist biases, which can be particularly problematic in industries like insurance, where equitable treatment of customers is essential. Companies must actively monitor and adjust their training data to ensure fair and non-discriminatory AI usage.

Deep fakes, disinformation and ethics

The rise of GenAI brings serious challenges in terms of security and ethics. The growing demand for cybersecurity and fact-checking measures^{17, 18} highlights the need to safeguard sensitive data and ensure the integrity of generated content.

Companies must invest in advanced security technologies and establish ethical standards to guide the responsible use of GenAI. Comprehensive regulations can help build trust in these systems among employees, customers, and stakeholders^{19, 20, 21}.

Summary

In summary, the potential of multimodal GenAI is vast, but it comes with a range of challenges that must be addressed. These include ensuring data accuracy and managing computational costs, as well as navigating legal frameworks and promoting trust through transparency and fairness. To mitigate these risks, companies must take proactive steps. Furthermore, ethical considerations, particularly around bias and deepfakes, will continue to shape how GenAI is implemented responsibly in industries like insurance, by further expanding the use of these technologies and continuously working on their optimization.

Outlook: Why the Insurance Industry Must Take Action Now

Prognoses on the future impact of multimodal GenAI models

The insurance industry is facing a technological transformation, driven by the integration of the new GenAI models. These technologies have the potential to optimise existing processes and generate new business models. Insurance companies operate data-driven business models, making them ideal candidates for GenAI adoption. In contrast to other industries, the impact of this technology will be particularly noticeable, as it provides efficient support for the processing and analysis of large quantities of text.

The text modality is likely to have the greatest impact on the industry, as a large proportion of the interactions and data processing is carried out using text-based formats. However, there are also numerous applications where multimodality will play a key role in the future – in the analysis of images and videos documenting damage, for example, or in the provision of virtual avatars for personalised customer service.

Democratisation of knowledge and skills

A feature of the years ahead will be a democratisation of knowledge, made possible by almost unlimited and constantly available access to knowledge²². This trend will significantly influence how people access information^{23, 24}.

Insurance companies will benefit from this democratisation by improving employees' access to valuable information and training. This will enhance skills development within the company and improve customer consultations through faster information access.

Increase in the productivity of knowledge workers

GenAI will significantly increase the productivity of knowledge workers²⁵ and change their working processes in the long term. Studies show that around 75% of the productivity increases from generative AI are in the areas of customer service, sales & marketing, software development and research & development²⁶.

GenAI enables knowledge workers to improve efficiency and focus on creative, strategic tasks, delivering greater value for companies. This shift empowers companies to leverage human expertise more effectively, securing a competitive advantage in an increasingly AI-driven world.

Hyper-personalisation as the new standard

Whether in terms of customising marketing messages in real time or the designing of individual insurance products based on customer needs, there are many possibilities which offer significant improvements to the customer experience. Through the deployment of GenAI, insurance companies can offer customised solutions to suit requirements and preferences of each customer.

Integration in robots, machines and vehicles

The integration of multimodal GenAI into machines and robots will make the interactions with these more human-like²⁷ and increase their capabilities, both in the execution of physical and cognitive tasks²⁸. This technology will not only revolutionise the industry; it will also set new benchmarks in safety and efficiency, whether in factories²⁹, on building sites or in autonomous vehicles³⁰.

For insurance companies, integrating GenAI into robots, machines, and vehicles presents numerous opportunities. GenAI-guided systems could reduce the risk of accidents through improved safety mechanisms, leading to fewer loss events and insurance claims. Furthermore, data from these intelligent systems will enable precise risk analysis and customised insurance policies, contributing to the development of dynamic premium models which reflect the actual terms of use and behaviour of machines and vehicles.

Exponential acceleration of innovation and development cycles

Accelerating innovation cycles with GenAI-generated ideas and rapid implementation will be crucial³¹. This will enable companies to react with more agility to market changes and continuously improve their products, leading to a competitive advantage in an increasingly dynamic economic environment.

Strategic Recommendations for insurance companies

The integration of multimodal GenAI offers a wealth of opportunities to fundamentally transform business models, processes and the customer experience. To fully harness the potential of GenAI, insurance companies must focus on their data strategy and invest in the right infrastructure and talent.

Developing new business opportunities and adapting the product portfolio

Insurance companies should regularly assess the impact of GenAI on their business model and services to identify long-term market changes and devise appropriate strategies. For example, a reduction in accident risk through autonomous vehicles might decrease the demand for conventional motor insurance. At the same time, new preventative approaches could be developed due to the deployment of AI-based technologies such as sensors and cameras, thereby transforming the existing product portfolio in the direction of technology-driven preventative initiatives.

Strategic alliances and partnerships

Strategic alliances and partnerships are essential for accessing advanced technologies and expertise. Insurance companies should collaborate with leading technology companies and AI start-ups so that they can benefit from their innovation and expertise. Through collaborating with branch initiatives and platforms, best practices can be exchanged, standards developed and common challenges arising from the implementation of GenAI addressed.

Adapting data strategy and operations

One of the core challenges in integrating GenAI is adapting the data strategy. Insurance companies need to ensure that they have top-level data integration and data quality. This requires the development and implementation of systems which can seamlessly integrate and consolidate data from different internal and external sources. This includes data sources that may have been previously overlooked, such as documents with continuous text, which were not traditionally considered 'data'. The evaluation and preparation of these sources is vital in order to have the optimum range of data available.

Implementing real-time data processing technology is crucial to fully exploit GenAI applications like real-time damage assessment and direct customer communication.

Investment in the AI infrastructure

To achieve a scalable and efficient GenAI environment, insurance companies should ensure investment in their technological infrastructure, both in the short and long term. Short-term investment is necessary to guarantee the requisite flexibility and enable a fast response to pre-

vailing market requirements. This includes the implementation of robust cloud platforms and advanced software solutions, as well as selected AI hardware, if necessary. At the same time, it is imperative to secure long-term scalability in order to adapt to the ever-changing needs of the market. These strategies will allow insurance companies to fully harness GenAI's potential across their processes and services.

Change management and promoting a culture of innovation

Implementing GenAI requires a comprehensive change management strategy across insurance companies. This process should aim to prepare the entire organisation for the use of AI-supported processes and technologies. This includes fostering a culture of continuous improvement and openness to change.

A successful change management process should also incorporate the democratisation aspects of multimodal GenAI, with the company ensuring that all employees have access to the relevant tools and resources to make their work more efficient and effective. Equal digital participation fosters an inclusive company culture and maximises the potential of GenAI across the organisation.

Recruitment of tech talent and employee training

In addition to recruiting AI experts, ongoing and comprehensive training of employees is a key factor in ensuring the successful deployment of GenAI models in the company. Companies should make it possible for employees to expand their skills and competence in handling GenAI, ensuring use of the latest developments and best practices. This can be achieved through targeted training initiatives, workshops and mentoring, not only to impart technical skills, but also to promote understanding of the strategic benefits of GenAI.

Creation of a general regulatory and ethical framework

A further aspect is the development of guidelines for the ethical use of GenAI. Insurance companies must ensure that the technology is implemented responsibly. This requires compliance with new regulations, such as the EU AI Act, and the company's own guarantee that the GenAI initiatives comply with all legal requirements.

By implementing clear guidelines and ethical standards, insurance companies can ensure the responsible use of GenAI while aligning with their customers' expectations for innovation and reliability, to ultimately build trust and provide transparency in an evolving ai-driven landscape.

Summary

Multimodal GenAI has the potential to fundamentally transform the way we work with data – not only in the insurance industry, but also in many other sectors. In this white paper, we have examined diverse areas of application – from the transformation of the customer experience to the optimisation of internal processes and through to data-led support for strategic decisions. One thing is clear: we are only at the start of this development. This technology is developing at an incredible pace – almost daily, we hear reports of further progress and innovative GenAI applications.

While there are promising opportunities, the challenges should not be underestimated. These range from the

adaptation of data strategies through to investment in infrastructure and talent development. Companies that act now and implement this technology will gain a significant competitive edge.

Our experiences in the ERGO Innovation Lab show that the implementation of multimodal GenAI technologies opens up a wealth of opportunities. Our collaborations in specialised sectors in the prototypes projects have already given us valuable insights and shown how practical solutions can be achieved. With this knowledge, we would like to contribute to making the insurance industry fit for the future. Now is the ideal time to take decisive steps toward long-term success by broadening the application of these technologies and consistently refining them.

Authors

Georgina Neitzel

Head of Innovation Lab
ERGO Group AG

Nicolas Konnerth

Head of Conversational AI
ERGO Group AG

Acknowledgements

This whitepaper has been shaped by the valuable support, constructive ideas, and insightful feedback of many contributors. We would like to sincerely thank everyone who provided expert input, thorough proofreading, and thoughtful suggestions, which played a key role in the development and improvement of this document. Your contributions have brought fresh perspectives and added depth to important aspects, enriching the overall content of this whitepaper.

References

- 1 „OpenAI Spring Update,“ (Online). Available: <https://openai.com/index/spring-update/>. (Accessed on May 28, 2024).
- 2 G. I. . i. u. 1. minutes. (Online). Available: <https://www.youtube.com/playlist?list=PLOU2XLXmsIJZaqcQHCNrDtoaEgvZORMk>. (Accessed on May 28, 2024).
- 3 „Hello GPT-4o,“ (Online). Available: <https://openai.com/index/hello-gpt-4o/>. (Accessed on May 28, 2024).
- 4 „Gemini Pro - Our best model for general performance across a wide range of tasks,“ (Online). Available: <https://deepmind.google/technologies/gemini/pro/>. (Accessed on May 28, 2024).
- 5 „Introducing GPT-4o: our fastest and most affordable flagship model,“ (Online). Available: <https://platform.openai.com/docs/models/gpt-4o>. (Accessed on May 28, 2024).
- 6 OpenAI, „GPT-4 is OpenAI’s most advanced system, producing safer and more useful responses,“ (Online). Available: <https://openai.com/index/gpt-4/>. (Accessed on May 30, 2024).
- 7 OpenAI, „Hello GPT-4o,“ (Online). Available: <https://openai.com/index/hello-gpt-4o/>. (Accessed on May 30, 2024).
- 8 W. e. al., „NEXt-GPT: Any-to-Any Multimodal LLM,“ 13 September 2023. (Online). Available: <https://arxiv.org/pdf/2309.05519>.
- 9 L. M. e. al., „On the Challenges and Opportunities of Generative AI,“ 2024. .
- 10 „Leveraging generative AI in Europe: The opportunities and challenges,“ McKinsey. (Online). (Accessed on June 6, 2024).
- 11 McKinsey, „GenAI in der Fertigung angekommen – nächste Welle der Automatisierung mit großem Potenzial,“ (Online). Available: <https://www.mckinsey.com/de/news/presse/2024-04-22-hannover-messe-genai>. (Accessed on June 6, 2024).
- 12 S. Online. (Online). Available: <https://www.spiegel.de/wirtschaft/unternehmen/air-canada-chatbot-verspricht-kunden-irrtuemlich-rueckerstattung-airline-muss-zahlen-a-0af54651-fbb7-4d8f-ab01-862a8e723ac9>. (Accessed on June 10, 2024).
- 13 „Arup lost \$25mn in Hong Kong deepfake video conference scam,“ Financial Times, (Online). Available: <https://www.ft.com/content/b977e8d4-664c-4ae4-8a8e-eb93bdf785ea>. (Accessed on May 28, 2024).
- 14 „Fake News war gestern. Die Ära der Fake Reality beginnt,“ Spiegel Online, (Online). Available: <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>. (Accessed on May 29, 2024).
- 15 M. Kinsey, „As gen AI advances, regulators—and risk functions—rush to keep pace,“ (Online). Available: <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/as-gen-ai-advances-regulators-and-risk-functions-rush-to-keep-pace>. (Accessed on May 28, 2024).
- 16 pwc, „Trusted AI: Ethics and responsible AI in Generative Systems,“ (Online). Available: <https://www.pwc.de/en/digitale-transformation/generative-ai-artificial-intelligence/the-genai-building-blocks/the-ethical-imperative-in-ai-development.html>. (Accessed on May 28, 2024).
- 17 H. B. Review, „Navigating the New Risks and Regulatory Challenges of GenAI,“ (Online). Available: <https://hbr.org/2023/11/navigating-the-new-risks-and-regulatory-challenges-of-genai>. (Accessed on May 28, 2024).
- 18 „bitcom,“ 6 2024 Mai. (Online). Available: <https://www.bitkom.org/Presse/Presseinformation/Industrie-Zukunft-KI-Einsatz>.
- 19 Gartner, „Generative AI Can Democratize Access to Knowledge and Skills,“ (Online). Available: <https://www.gartner.com/en/articles/generative-ai-can-democratize-access-to-knowledge-and-skills>. (Accessed on May 29, 2024).
- 20 Google, „Generative AI in Search: Let Google do the searching for you,“ (Online). Available: <https://blog.google/products/search/generative-ai-google-search-may-2024/>. (Accessed on May 30, 2024).
- 21 T. W. Post, „AI is changing Google search: What the I/O announce-ment means for you,“ (Online). Available: <https://www.washingtonpost.com/technology/2023/05/10/google-search-ai-io-2023/>. (Accessed on May 30, 2024).
- 22 W. t. E. i. A. i. 2024, „Stanford University Human-Centered Artificial Intelligence,“ (Online). Available: <https://hai.stanford.edu/news/what-expect-ai-2024>. (Accessed on May 28, 2024).
- 23 McKinsey, „The economic potential of generative AI The next productivity frontier,“ 2023. .
- 24 A. Velinov, „How Multimodal AI Will Transform Human-Machine Interaction Forever,“ (Online). Available: <https://www.linkedin.com/pulse/how-multimodal-ai-transform-human-machine-interaction-alex-velinov-t5kff/>. (Accessed on May 30, 2024).
- 25 W. Knight, „Google DeepMind’s Demis Hassabis Says Gemini Is a New Breed of AI,“ (Online). Available: <https://www.wired.com/story/google-deepmind-demis-hassabis-gemini-ai/>. (Accessed on May 30, 2024).
- 26 Forbes, „The Future Of Generative AI: 6 Predictions Everyone Should Know About,“ (Online). Available: <https://www.forbes.com/sites/bernardmarr/2024/03/05/the-future-of-generative-ai-6-predictions-everyone-should-know-about/>. (Accessed on May 29, 2024).
- 27 Forbes, „Sensing Success: OpenAI, Anthropic And 40+ Others Leverage Multimodal AI,“ (Online). Available: <https://www.forbes.com/sites/digital-assets/2024/05/28/sensing-success-openai-anthropic-and-40-others-leverage-multimodal-ai/>. (Accessed on May 28, 2024).
- 28 „From design thinking to Autonomous Innovation,“ Board of Innovation, (Online). Available: <https://www.boardofinnovation.com/blog/introducing-autonomous-innovation/>. (Accessed on May 29, 2024).
- 29 „ChatGPT und Sprachmodelle: Eine Einführung mit Blick auf die Versicherungsbranche,“ (Online). Available: <https://www.ergo.com/de/Newsroom/Medieninformation/2023/20230612-Whitepaper-ChatGPT>.
- 30 „Our next-generation model: Gemini 1.5,“ (Online). Available: <https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/#context-window>. (Accessed on May 28, 2024).
- 31 IBM, „The most important AI trends in 2024,“ (Online). Available: <https://www.ibm.com/blog/artificial-intelligence-trends/>. (Accessed on May 28, 2024).