MARKETLOGIC

Everything an insights professional needs to know about gen AI (but doesn't)

The must-know gen AI lexicon for insights professionals

How this AI glossary can help you

This resource is designed to empower insights professionals like yourself with the essential generative (gen) AI concepts and terms to confidently advocate for the best AI insights solution for your business.

This guide will equip you to influence IT teams and decision-makers who may not understand the unique benefits of AI solutions tailored to market insights. With the right language and expertise, you'll feel better prepared to navigate conversations on what sets a great AI insights solution apart — and help enhance your organization's market and customer knowledge across the entire business.

Use this resource to elevate your know-how in market insights, stay ahead in the AI-driven landscape, and bolster your case for cutting-edge, purpose-built AI for insights, such as DeepSights™.

Generative AI (Gen AI)

Generative AI is a type of artificial intelligence designed to create new content or data by learning from large datasets of existing content. It operates by analyzing complex patterns, relationships, and "rules," which it then applies to create something new based on the patterns it learned.

Large Language Models (LLMs)

Large Language Models (LLMs) are a type of generative AI trained on large amounts of text data to learn patterns, context, and nuances in language. In its raw form, an LLM essentially functions as an advanced word predictor, similar to a highly sophisticated autocomplete system. LLMs predict the next word in a sequence based on the probability calculated from the words that precede it.

Hallucinations: Because LLMs are word-predictors, they can generate new text that is grammatically correct and contextually relevant, even if it's not necessarily factually correct. AI like ChatGPT can produce information that is incorrect, misleading, or entirely fabricated, despite appearing plausible. This is because the LLM is trying to generate coherent responses based on patterns in data rather than verified facts.

Generative Pre-trained Transformer (GPT)

Generative Pre-trained Transformer (GPT) refers to a family of LLMs that use “transformer” architecture. The transformer architecture of GPT models allows the models to process and predict language patterns more effectively than traditional LLMs through its use of self-attention mechanisms, pre-training on massive datasets, and fine-tuning for conversational tasks. GPT models are large-scale and trained for advanced human-like text generation, whereas LLMs can be small-scale and fine-tuned for non-generative tasks like categorization, sentiment analysis, etc.

Chatbot: ChatGPT is a type of chatbot powered by GPT models developed by the company OpenAI. While ChatGPT uses the underlying GPT architecture, it’s specifically optimized for maintaining conversations, answering questions, and assisting with various tasks in a more interactive manner. Unlike simpler chatbots that might be programmed with fixed responses, ChatGPT a uses large language model that has been fine-tuned for a chatbot experience, to generate more dynamic and contextually relevant replies.

DeepSights™ leverages powerful LLM models using APIs

When you ask AI assistant DeepSights™ a question, it rapidly makes several API calls to OpenAI’s powerful GPT (LLM) models. It does this through Microsoft Azure’s OpenAI service, which gives DeepSights™ the security capabilities of Microsoft Azure while also being able to leverage OpenAI’s GPT models in its proprietary process.

But in order to answer your market research questions to the highest degree of accuracy and based on your own knowledge assets, DeepSights™ uses a technique called RAG (Retrieval-Augmented Generation).

Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) is a technique that combines an information retrieval system with a text-generation model. The retrieval system first fetches relevant documents or data snippets from a knowledge base, which are then used for the text-generation model to produce an answer. RAG is particularly useful in tasks like question answering, where the ability to refer to specific information significantly enhances performance; it allows the model to produce more informed, accurate, and contextually relevant responses.

Embeddings

Embeddings are numerical representations of words, sentences, or entire documents. Series of numbers capture semantic relationships, which are then mapped out in a high-dimensional vector space — where similar meanings are situated closer to each other. Embeddings are crucial in RAG for converting both queries and documents into vectors, which are stored in vector databases, and used for similarity search and contextual understanding.

Vector databases

Vector databases are specialized databases designed to handle and query high-dimensional vectors like embeddings. Vector databases allow for efficient searches by measuring distances of similar meanings between vectors. This transformation allows models to more effectively compare and retrieve semantically related content because similar pieces of content embeddings are located closer together.

LLM-based reflection

LLM-based reflection refers to the process of using LLMs like GPT to analyze and generate insights about text or ideas. It involves leveraging the model's ability to understand context to explore or critique concepts, assumptions, or arguments. This can help deepen understanding, highlight nuances, or offer alternative perspectives.

DeepSights™ — a RAG tool customized for the market insights context

DeepSights™ RAG process vectorizes (via embeddings) all your documents and integrated source content, then it stores that in a vector database. When you ask DeepSights™ a question, your question is also embedded/vectorized into the vector database so that the model can identify and retrieve the most semantically similar evidence to your query.

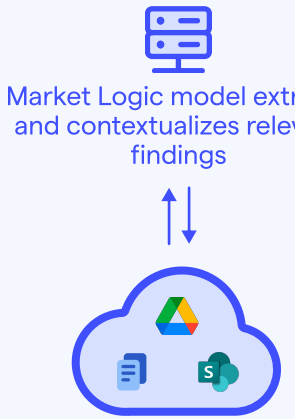
But insights professionals can't risk poor quality answers or hallucinations, so DeepSights™ goes a step further. Before generating an answer, DeepSights™ employs an LLM-based evidence filtration and re-ranking, sort of like a double-check, to ensure that only the most relevant information is passed on to the answer-generation component. This step also ensures that the evidence meets an acceptable threshold to eliminate hallucinations.

The relevant evidence is then packaged in a prompt and sent to the LLM to produce a highly-relevant and useful answer for the business-insights context. All cited sources are visible and verifiable with a simple hover over them.

DeepSights RAG

An overview of the DeepSights™ Retrieval Augmented Generation (RAG) flow – enabled by gen AI tech:

1 Ingestion phase



In the ingestion phase, DeepSights™ uses best-in-class technologies and **LLMs to ingest and embed customer data.** These embeddings are stored in a vector database.

2 Question asked

When a **user asks DeepSights a question**, the question itself is embedded, to facilitate semantic search of the Knowledge Estate.

3 Vector search

Vector search then identifies the most relevant pieces of evidence.

4 Smart receiver

DeepSights then applies additional evidence filtering and re-ranking to further refine the **vector search results** and ensure high evidence quality.

5 Response

All of the relevant evidence chunks are then **packaged and passed to an LLM, where the DeepSights answer to the user's question is crafted**.

Using a specially built gen AI assistant for insights, rather than generic solutions, means the output is based only on relevant data from the start — resulting in market answers to your questions that you can trust, and minimizing hallucinations.
To see DeepSights™ in action, contact us for a demo.

Application Programming Interface (API)

An Application Programming Interface (API) is a set of protocols that allows different software programs to talk to each other. It sets the rules for how they can ask for and share information. This means developers can use features from other apps or services in their own software. While not specific to Gen AI in particular, APIs are a cornerstone of modern web applications, so we have included them in our glossary to help tie together the pieces. Although there are many LLMs available, OpenAI was one of the first to make an effective API for their GPT model. That's why many developers, including Microsoft, use OpenAI's GPT API.

DeepSights™ watch-outs feature helps you find contradictions

DeepSights™ watch-outs feature uses an LLM-based reflection process in which it can reflect on its answer and point out contradictions in evidence, older sources, lack of coverage in certain geographies or demographics, etc.

About Market Logic

MARKETLOGIC

Market Logic is a market leading SaaS provider of insights management solutions. Our award-winning AI-enabled insights management platform DeepSights™ allows insights teams to equip business decision-makers with trusted insights at scale and speed. Since 2006, we've helped hundreds of consumer-focused brands across the globe to transform into insights-driven businesses. Market leaders such as Unilever, Vodafone, and Tesco are driving innovation and making smarter market moves with the support of Market Logic.

www.marketlogicsoftware.com
info@marketlogicsoftware.com

Europe

Franklinstrasse 28,
10587 Berlin,
Germany

North America

223 W. Jackson
Blvd, Suite 525,
Chicago, IL
60606, USA

2024 © Market Logic AG. The information contained in these documents may not be used, published or redistributed without the prior written consent of Market Logic Software AG, Franklinstraße 28, 10587 Berlin, VAT Nr. DE 249354497