

# Hype Cycle for Natural Language Technologies, 2023

Published 27 July 2023 - ID G00790908 - 106 min read

By Analyst(s): Bern Elliot, Adrian Lee, Gabriele Rigon

Initiatives: [Artificial Intelligence](#); [Evolve Technology and Process Capabilities to Support D&A](#)

A broad range of advances in artificial intelligence is enabling innovative approaches and advances in the field of natural language technologies. This report will assist IT leaders in assessing how and where these new opportunities and methods can best be applied.

## More on This Topic

This is part of an in-depth collection of research. See the collection:

- [2023 Hype Cycles: Deglobalization, AI at the Cusp and Operational Sustainability](#)

## Analysis

### What You Need to Know

*This document was revised on 22 August 2023. The document you are viewing is the corrected version. For more information, see the [Corrections](#) page on gartner.com.*

Natural language technology (NLT) encompasses technologies and methods that enable human-language-based communication between humans and systems, as well as the analysis of those communications. NLT includes tools, services and products in the areas of:

- Natural language understanding (NLU)
- Natural language generation (NLG)
- Text analytics
- Conversational use interfaces
- Language knowledge graphs and vector databases
- Machine translation
- Text summarization
- Speech technology and neural symbolic language models

New capabilities are emerging, some leveraging foundation models and generative AI. These new capabilities, combined with existing methods and workflows, are enabling significantly improved functionality.

IT leaders should understand how the technologies outlined in this report can be composed to produce useful and differentiating capabilities. Often, the value emerges when the language techniques are combined and integrated with other business applications and processes.

The following Hype Cycles cover related technologies and trends:

- [Hype Cycle for Artificial Intelligence, 2023](#)
- [Hype Cycle for Data Science and Machine Learning, 2023](#)

- [Hype Cycle for Customer Service and Support Technologies, 2023](#)

## The Hype Cycle

This Hype Cycle focuses on innovations where generative AI and large language models (LLM) are enabling significant advances in natural language capabilities.

NLT is not a single market or industry, but rather a broad set of capabilities. Functionality is sometimes delivered as either:

- A stand-alone, commoditized capability
- Multiple functions that are combined to deliver a targeted solution or platform
- Functionality embedded in existing solutions

The fragmentation in the NLT market is largely due to four factors:

- Rapid evolution of generative AI methods, especially those related to LLMs and ChatGPT, which have created significant disruption and uncertainty in the NLT market, while also creating significant market interest.
- Many NLTs are broad solution enablers. As a result, functionality may be purchased as a general tool, or it may be embedded in applications targeting specific markets.
- Many emerging NLT capabilities disrupt solutions that are already established in the market but are based on an earlier generation of technologies. These emerging solutions may both disrupt preexisting markets and enable new markets.
- Regional language differences, which create fragmented NLT markets.

Due to the hype around the rapid and recent advances enabled by generative AI methods, take care to separate hype from what is actually delivering consistent results. While most IT leaders select NLT solutions from vendors that fit specific needs and use cases, there is a trend toward developing in-house language skills that can act as a resource to internal users.

New technology entrants this year:

- Vector databases

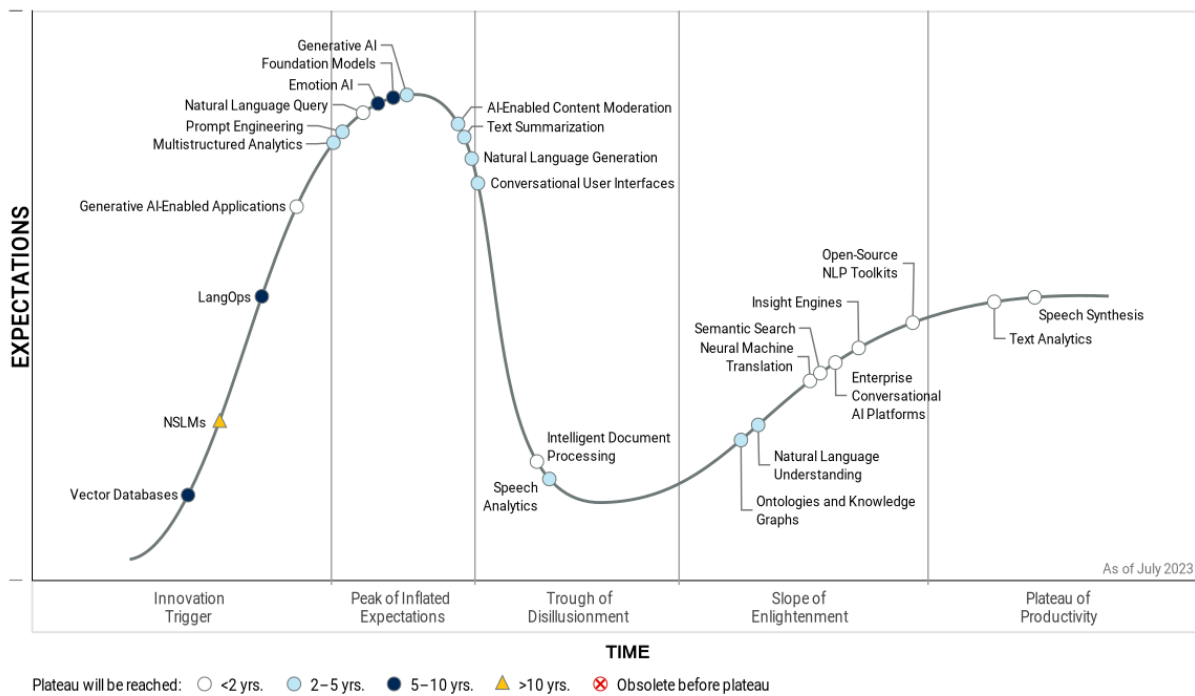
- Prompt engineering
- Generative AI
- AI-enabled content moderation
- Conversational UIs

Consolidated:

- Virtual assistants and chatbots were incorporated into the conversational UIs profile. These profiles were increasingly redundant.

**Figure 1: Hype Cycle for Natural Language Technologies, 2023**

## Hype Cycle for Natural Language Technologies, 2023



## The Priority Matrix

The Priority Matrix maps the benefit rating for each innovation profile (IP) against the amount of time each IP requires to achieve mainstream adoption. The benefit rating provides a general indicator of the IP's potential, but the rating may not be applicable to all organizations. As a result, IT leaders should identify which of the IPs offer significant potential benefits to their own organization based on their own use cases. This can then guide investment decisions. IPs offering more significant near-term benefits should be carefully considered, as they can offer both strategic and tactical benefits. Innovations with longer-term benefits should be considered if they offer strategic value.

As the Priority Matrix illustrates, many of the IPs reviewed in this report offer relatively near-term benefits. This reflects the tremendous impact that new AI processes have on NLT methods. Conversational AI platforms represent a particular area of significant exploration and investment among enterprises. Adoption has been assisted by the availability of application development tools. Apart from neural machine translation (NMT), the tools in many of the other innovations outlined in this Hype Cycle are not very mature. As a result, IT leaders should consider how to approach solution development when they review these areas.

**Table 1: Priority Matrix for Natural Language Technologies, 2023**

(Enlarged table in Appendix)

Benefit ↓	Years to Mainstream Adoption			
	Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Transformational	Enterprise Conversational AI Platforms Neural Machine Translation	Conversational User Interfaces Generative AI Natural Language Understanding	Emotion AI Foundation Models	NSLMs
High	Generative AI-Enabled Applications Insight Engines Intelligent Document Processing Natural Language Query Open-Source NLP Toolkits Semantic Search	Multistructured Analytics Natural Language Generation Ontologies and Knowledge Graphs Prompt Engineering	LangOps Vector Databases	
Moderate	Speech Synthesis Text Analytics	AI-Enabled Content Moderation Speech Analytics Text Summarization		
Low				

Source: Gartner (July 2023)

## Off the Hype Cycle

### Dropped:

- **Adaptive PEMT** — This functionality is now incorporated into broader translation management system suites.
- **Multilingual models** — This technology was rendered redundant with LLM methods.
- **Chatbots** — This functionality is now incorporated into the conversational UIs profile.
- **Natural language query** — This technology was rendered redundant with LLM methods.
- **Virtual assistants** — This functionality is now incorporated into the conversational UIs profile.

- **Text autoclassification** — This functionality has been incorporated as a capability of LLMs, and is no longer delivered as a stand-alone function.

## On the Rise

### Vector Databases

Analysis By: Arun Chandrasekaran, Radu Miclaus

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

#### Definition:

Vector databases store numerical representations of data. In such databases, each point is represented by a vector with a fixed number of dimensions, which can be compared via mathematical operations, such as distance measures. Vector databases are commonly used in machine learning (ML) solutions, where vectors represent data features/attributes, such as text embeddings. Storing these vectors in a database enables users to search for similar data points with low latency.

#### Why This Is Important

Vector databases serve such use cases as similarity search and product recommendation. Rapid innovation in generative AI and adoption of AI foundation models have spawned interest in vector databases. When customers adopt generative AI models, vector databases store the embeddings that result from the model training. Storing vector embeddings representing the model training, the database can do a similarity search, which matches a prompt (the question) with specific or similar vector embedding.

#### Business Impact

Businesses thrive by delivering differentiated customer experience (CX). Generative AI is increasingly embedded in applications to empower the human-machine symbiosis, and organizations need scalable and accelerated ways to build and support these applications long term. Vector databases are an important back-end service that allows businesses to future-proof and scale their generative-AI-enabled applications. These drive business value through customer engagement and adoption.

#### Drivers

- **Popularity of vector embeddings:** With the rise of AI foundational models, embeddings have become the cornerstone for semantic search. Hence, they are the working inputs for training large foundation models.



- **Performance and scalability needs:** The applications looking to embed generative methods that use embeddings-based models need back-end services that can respond with low latency to high concurrency requests (prompts) and responses (completions) for generative AI use cases.
- **Service architecture:** Because most applications are built on service-based architectures, vector databases are ideally presented to applications as services that communicate with the interface via APIs.
- **Hybrid implementation of retrieval and generative models:** Vector databases are optimal for both semantic search (retrieval based on vector similarity) and generative inference through foundation models. This hybrid combination of models drives the need for optimized vector databases, because both generative and retrieval are used together for grounding of facts.
- **Developer focus:** Developers of new applications are driving the demand for vector databases by presenting use cases that cannot scale without the ability for embeddings to be stored in an optimized structure for high-throughput production applications.

## Obstacles

- Enterprises lack an understanding of what vector databases do and the unique use cases they enable.
- Vector databases are superspecialized databases that may cause challenges around data migration and integration and limited extensibility across use cases.
- Most vector databases are delivered as cloud-managed service — the complexity of deploying, configuring and operating them outside cloud environments requires deep technical skills and know-how.
- Vector databases can be expensive to implement, given the newness of the technology and lack of industry skills to deploy and manage it.
- The vector database market is nascent and populated mostly by startups, which may not have extensive experience working with enterprise clients, as well as unproven product market fit.

## User Recommendations

- Determine whether your functional requirements can be satisfied by incumbent vendors that can support the storage and retrieval of vector embeddings – you may not always need a purpose-built vector database.
- Prioritize developer experience, ecosystem integration, use case fit, reliability and performance as important selection criteria, and validate them thoroughly via a POC process.
- Select managed, cloud-based vector databases as deployment modes, unless you have stringent requirements and deep technical skills for an on-premises, self-managed deployment mode.
- Conduct internal training and education on the appropriate use cases for vector databases, how to leverage their true potential, and effective ways to optimize their deployment and maximize their value.

## Sample Vendors

Couchbase; Croma; Elastic; Google; Pinecone Systems; Qdrant; Redis; Weaviate; Zilliz

## Gartner Recommended Reading

[Innovation Insight for Artificial Intelligence Foundation Models](#)

[Quick Answer: What Is GPT-4?](#)

[Executive Pulse: AI Investment Gets a Boost From ChatGPT Hype](#)

[How Large Language Models and Knowledge Graphs Can Transform Enterprise Search](#)

## NSLMs

Analysis By: Van Baker

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

**Definition:**

Neural symbolic language models (NSLMs) are a form of composite AI that combines deep neural networks' learning abilities with symbolic models' reasoning abilities to create a new class of language-processing model. NSLMs are designed to transcend the inductive bias of neural network language models (NNLMs) by adding symbolic reasoning that can address the issues NNLMs face. These include spatial or temporal properties in language that can be represented by logical or mathematical expressions.

**Why This Is Important**

Deep neural networks can provide advanced natural language technology (NLT) capabilities such as large language models (LLMs), but still have limitations. Processing language via symbolic models using symbolic constructs has delivered results, but progress has slowed. NSLMs' combination of the two approaches has produced results that show benefits over LLMs alone. These benefits can also be seen to some degree by combining LLMs with knowledge graphs.

**Business Impact**

Neural symbolic models can improve the performance of standard LLMs and facilitate new solutions, assuming advances are repeatable. Further disruption of language-processing technologies are expected as we have seen strong advances in the LLM market as shown by ChatGPT. The practicality of training larger and larger models is limited, and combinations with symbolic approaches may be explored. The investment required in this market may force small vendors to rethink strategies and choose to partner with hyperscalers to remain competitive.

## Drivers

- Demand for language AI technology has exploded due to the launch of ChatGPT.
- Natural language processing technology continues to enable increasingly complex solutions. These incorporate transactional capabilities and automated decision making and workflow solutions, such as intelligent document processing.
- The creation of very large language models has had a significant impact on language processing. In addition to advancing the performance of natural language processing, they have shown abilities to answer questions about subjects they have not been trained on. Language technology continues to evolve at a rapid pace, and new approaches, such as NSLMs, promise continued improvement.
- Composite AI techniques have started to show remarkable results in addressing complex business and engineering problems.
- Neuro-symbolic models have already demonstrated the capability to outperform state-of-the-art deep learning models in domains such as image and video reasoning (see [Neuro-Symbolic AI: An Emerging Class of AI Workloads and Their Characterization](#) from Cornell University).
- Some evidence suggests that neuro-symbolic AI systems can be trained with significantly less data than what is required for other methods (see [Neuro-Symbolic AI](#), from MLearning.ai).

## Obstacles

- Neural symbolic models are a fundamentally different approach that requires two dramatically different approaches to AI model building.
- Combining deep machine learning and symbolic models may prove to be very challenging, and the different camps within the AI community add to the challenge.
- Renewed advances from very large language models have reinvigorated the deep learning approach to language AI solutions.
- The lack of ModelOps methodologies is a challenge for these models.
- NSLMs are likely to come from the large cloud providers, which may constrain access to and availability of solutions that use these models.

## User Recommendations

- Assume that many natural language solutions will keep evolving rapidly, with a range of different technical approaches to addressing language solutions.
- Expect volatility among language AI vendors to be high, as technical advantages will emerge that disrupt language solutions.
- Plan for continued instability in the natural language processing market. This will fuel deployment of rapidly evolving language-processing-based solutions, especially modular or composite solutions that enable swapping of core services.
- Assume that language-focused AI models will increasingly be applied to complex tasks and not be limited to simple solutions, as advances in language technologies are likely to deliver substantial improvements in language understanding.

## Sample Vendors

IBM; Microsoft; OpenAI

## LangOps

Analysis By: Bern Elliot, Soyeb Barot

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

### Definition:

Enterprise natural language operations (LangOps) are practices that support the management of the full life cycle of the language models and solutions implemented in enterprise architectures. This includes the development and curation of training data, artificial intelligence (AI) models and semantic data (e.g., knowledge graphs), as well as the continuous delivery of retrained models and functionality integrated with relevant business processes and applications.

## Why This Is Important

A growing number of natural language-based solutions are being used broadly across enterprises. Users want to customize and adapt capabilities based on their business requirements. To be successful, enterprise leaders must operationalize their approaches with the implementation of natural language technologies (NLTs) and techniques, to enable repeatability and reduce technical debt. This practice is called LangOps.

## Business Impact

Enterprise LangOps operate natural language technologies — enabling enterprises to integrate these technologies across enterprise applications, improving efficiency, reusability of assets and scalability of deployments. LangOps may initially be limited to single NLT areas, such as text analytics, conversational platforms, large language models or translation. However, synergies across multiple NLT areas will drive language centers of excellence (COEs) and cross-functional LangOps streams.

## Drivers

- Rapid increases in NLT by business areas. The business requirements are becoming more complex, and the underlying use of NLT is proliferating across increasingly broad business domain areas.
- The increasing need for sophisticated data and modeling practices, along with the volume of data. There is also demand for sharing training data and data-handling tools for semistructured and unstructured content.
- AI ModelOps and XOps practices that are not always applicable or adaptable to language areas. This drives a need for specialized streams in some language areas.
- Natural language solutions' need to draw from diverse technology and business skills. LangOps enables these diverse participant efforts to be organized and focused.
- The desire to share best practices for ingesting, managing, storing, governing, and monetizing large sets of unstructured data — for example, transcribed call recordings, specialized terminology libraries, and the use of vector data bases, as well as ontologies and knowledge graphs.
- Cross-functional uses for language technology, including model customization or fine-tuned language models. These might be used across multiple areas, including marketing, customer service, and websites.

## Obstacles

- **Immature practices and methods:** The overall maturity of many emerging NLTs is accompanied by immaturity of the best practices for managing the full solution life cycle. LangOps will need to fit with ModelOps, XOps and emerging GenAI practices. Implementing human in the loop (HITL) practices for training may be nonstandard to Ops practices.
- **Language applications span different NLT platforms:** This causes inconsistencies for LangOps across platforms. LangOps approaches are sometimes focused on specific use cases for NLTs, such as translation and localization.
- **Organizational issues:** The cross-functional nature of LangOps requires people with different skills and reporting responsibilities to work together. This poses challenges that can be complicated by the potential for conflicting “turf” and budget control issues. For example, a single language solution may be used by different areas of the business. Previously, separate solutions were handled by each business area.

## User Recommendations

- Develop an informal community of interest that includes business users of language technology as well as AI, generative AI, data science, machine learning (ML) and language experts. For many organizations, thinking about language technology as an interconnected area is a new concept. This community will assist in socializing this concept.
- Define a strategic enterprise NLT roadmap. This will start by viewing language initiatives as part of a broader portfolio, not as discrete projects. Allow projects to advance at their own pace, but look for how and where synergies across the organization will be useful. Provide guidelines for operationalizing NLT projects and the management of metadata and semantic data.
- Enlist the support of CxO-level sponsors as part of the planning. This will allow diverse business areas and groups to work together to reduce “turf” conflicts.
- Focus initially on language areas in which technical support to improve operations is already in use.

## Sample Vendors

IBM; Microsoft, Unbabel; Veritone

## Gartner Recommended Reading

[Market Guide for AI-Enabled Translation Services](#)

[Tool: Vendor Identification for Natural Language Technologies](#)

[ChatGPT Research Highlights](#)

[Innovation Insight for Generative AI](#)

[Best Practices for the Responsible Use of Natural Language Technologies](#)

## Generative AI-Enabled Applications

Analysis By: Radu Miclaus, Arun Chandrasekaran

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

### Definition:

Generative AI-enabled applications use generative AI for user experience (UX) and task augmentation to accelerate and assist the completion of a user's desired outcomes. When embedded in the experience, generative AI offers richer contextualization for singular tasks like generating and editing text, code, images and other multimodal output. As an emerging capability, process-aware generative AI agents can be prompted by users to accelerate workflows that tie multiple tasks together.

### Why This Is Important

Fast-moving advances in foundation models drive generative AI-enabled applications, which have the potential to democratize the workforce. Since applications can now be enabled with generative AI capabilities that process and provide output in human consumable modalities (text, images, sound, etc.), the use cases will permeate a wide spectrum of domains and skill sets within the knowledge workforce, reimagining how enterprises think of scale and productivity.



## Business Impact

Generative AI chatbots/agents/co-pilots within applications will target time-consuming, manual-prone and repetitive tasks, such as knowledge discovery, summarization and contextualization, software engineering and coding, graphic and video design, and workflow design and execution. With these tools at their disposal, knowledge workers and creatives will sustain new learning curves toward innovative ways to scale businesses. Businesses not making use of these tools will struggle to compete.

## Drivers

- **Fast advancement of foundation models:** Foundation models like GPT are advancing at an accelerated rate. There is a movement toward democratizing foundation models via open-sourcing variations, such as Meta AI (Large Language Model Meta AI [LLaMa]) or BigScience Large Open-science Open-access Multilingual Language Model (BLOOM).
- **Wider range of applications:** Among others, the most common pattern for generative AI-embedded capabilities today is text-to-X, which democratizes the access for knowledge workers to what used to be specialized tasks via prompt engineering using natural language. For example, **text-to-text** supports knowledge discovery, summarization and contextualization in communication applications across the enterprise. **Text-to-code** is emerging as developer processes get augmented through “pair programming” with AI co-pilots directly into the coding experience, with use cases ranging across the software development life cycle. **Text-to-image/video (image-to-image)** applies when applications from graphics design to video editing and full video generation see generative capabilities added both by traditional technology players as well as new startups. **Text-to-process/workflow** is emerging as generative AI agents enable users to use text and voice to generate workflows and generative tasks together in cohesive domain-specific applications. **Text-to-multimodal** supports the building of high-fidelity avatars, or digital objects that have image, sound and narrative/text modalities, as an example of multimodal application in metaverse and gaming.
- **Domain specialization:** Specialization on top of foundation models is extending into domain-specific refinement, as well as refinement based on internal/private/licensed knowledge bases and process definitions for enterprises.
- **Acceptance into professional life:** Consumers are pulling the generative AI-enabled applications into their professional life.
- **Computation cost optimization:** The computational innovations for training and inference are focusing on optimizing and refining the cost structures across the entire software stack (infrastructure, methodologies and integrations).

## Obstacles

- **Security, consumer privacy and enterprise intellectual property (IP) protection concerns:** A large number of inquiries from potential buyers of generative capabilities are concerned with the wide umbrella of trust and security. While large hyperscale vendors and startups are racing to make generative AI services enterprise-ready, in the short- to midterm, there will still be a lack of regulation and appropriate adaptable oversight.
- **Accuracy and veracity of outputs:** Hallucinations and inaccuracy will continue to be a concern for generative AI.
- **Fear around automation and job replacement:** Human nature brings a blend of excitement and fear around widespread adoption.
- **Learning curves and uncertainty:** As generative AI technology evolves, there is confusion about the implementation that is right for enterprises, how quickly the market is evolving and the lack of skills on transformers available in the market.
- **Regulation:** While currently lagging, regulations will follow and may increase the friction in innovation speed and adoption.

## User Recommendations

- **Seek technology providers that can offer vertical specialization:** Vendors who will accelerate the refinement and adoption of generative AI capabilities in the context of vertical and business processes of the enterprise should be prioritized in evaluation for existing and future needs.
- **Use enterprise-ready technologies:** For enabling and embedding generative AI in applications, (a hybrid build-and-buy approach), prioritize research into the roadmaps of enterprise-ready generative AI services with a focus on addressing the privacy, security and IP protection needs of the enterprise.
- **Encourage steady growth:** Challenge knowledge workers to engage in new learning curves, and improve or redesign business processes to respond to this disruption.

## Sample Vendors

Adobe; AgentGPT; Amazon; Anthropic; Google; Hugging Face; Inflection; Microsoft; OpenAI; Salesforce

## Gartner Recommended Reading

[Innovation Insight for Generative AI](#)

[Emerging Tech: Generative AI Needs Focus on Accuracy and Veracity to Ensure Widespread B2B Adoption](#)

[Quick Answer: How Can You Manage Trust, Risk and Security for ChatGPT Usage in Your Enterprise?](#)

[Innovation Insight for ML-Powered Coding Assistants](#)

[Quick Answer: Will Machine-Learning-Generated Code Replace Developers?](#)

## At the Peak

### Multistructured Analytics

Analysis By: David Pidsley, Stephen Emmott, Tim Nelms, Anthony Mullen

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

#### Definition:

Multistructured analytics describes techniques applied to unlocking the value of wide data – the spectrum of multistructured information (structured, semi-structured, and unstructured) of any formats, including language (text/audio) and vision (image/video), sourced internally or externally. Multistructured analytics decomposes meaningful features of human-generated data for DSML modeling. It composes machine-generated data into context-enriched analysis for expert decision making.

#### Why This Is Important

D&A leaders must bridge the gap between the analytics of today and the context-enriched analysis for decision makers to uncover unique insights. Exponential growth in the spectrum of multistructured information sources/formats requires techniques beyond traditional, structured, transactional or relational data. To get a grip on business complexity, harness multistructured analytics to support expert decision making with richer situational awareness, augment business workflows and automate decisions.

#### Business Impact

Across industries/business functions, organizations accelerate their application of multistructured analytics to wide data sources/formats to reduce costs, address new uncertainties, drive growth and enable innovation in

the use cases for analytics. Multistructured analytics of audio/video streams in sales and marketing can identify behaviors and sentiments of customers and influencers across channels for new insight, experience optimization, real-time dynamic pricing and competitive intelligence.

#### Drivers

- Decisions are more complex, with more stakeholders and choices than two years ago. Scenarios need context-sensitive evaluation, beyond individual events, using multidimensional models of real-world uncertainties. Lacking the right variety of data stifles this.
- By 2025, 70% of organizations will shift their focus from big to wide data, providing more context for analytics. Internal (digital workplace) and external (business ecosystem) data sources continue to increase silos, leaving hidden intelligence for competitive advantage. Data marketplaces and exchanges make larger, pretrained and more diverse data assets widely available.
- Organizations are adopting multistructured analytics to move beyond storing content, to extract meaningful features and insights. By 2023, over 80% of organizations will use some form of computer vision to analyze images and videos.
- All forms of wide data can now be processed. Document topics can be tagged, speech transcribed, imaged environments annotated, emotions predicted from video, gauges digitized, opening new doors for analytics, data science and machine learning (DSML) and AI.
- Transformer models (via BERT and GPT techniques), advanced text analytics and deep learning have been a catalyst for linguistic and visual analysis. By 2025, AI for video, audio, vibration, text, emotion and other content analytics will trigger major innovations and transformations in most global enterprises.
- Improved price/performance ratio of cloud AI developer services has made experimenting accessible and scalable.
- Multistructured analytics enriches structured data with categorization and tagging. Analytics and BI and DSML platforms are adding multistructured analytics and graph capabilities so that by 2025, context-driven analytics and AI models will replace 60% of existing models built from traditional data sources, consolidating a mix of analytics solutions.

## Obstacles

- Although different forms of (text) content analytics have been deployed for years, many organizations avoid leveraging multistructured information due to limited competencies, specialized tools and their perceived difficulty fueled by confusion around terminology.

- While the tools now exist to deploy multistructured analytics uniting the spectrum of multistructured information to model complex context, it's a shift in the way D&A teams undertake data modeling especially in natural language processing.
- The difficulty of combining techniques (composite AI) to handle specific formats/sources — like deep learning for videos, symbolic algorithms for text analytics, and knowledge graphs — is a challenge.
- Data sourcing, quality and privacy are common challenges that can be cost prohibitive for large datasets. Finding suitable data for a specific use case can be difficult and require governance.
- The market for multistructured analytics tools is fragmented and will likely require leveraging multiple vendors, increasing costs.

## User Recommendations

- Leverage multistructured analytics for richer situation awareness and expert decision support.
- Conduct proof of value/pilots and understand the data, technical and organizational gaps.
- Apply text analytics for supply chain optimization, image analytics for diagnostic maintenance, video analytics for conferences and audio analytics for fraud prevention.
- Provide context-enriched analysis for decision makers by applying multistructured analytics to multistructured information.
- Explore multistructured analytics capabilities and roadmaps of vendors, including insight engines for text content and cloud AI developer services for image, video and audio analytics.
- Engage startups and hyperscale cloud providers for innovation.
- Estimate your compute and storage needs to train/run effective ML models that leverage multistructured information.
- Invest in taxonomy/ontology skills to accelerate the refinement and automation of information tagging/classification.
- Revise data collection, management and integration practices to take advantage of multistructured analytics.

## Sample Vendors

Amazon Web Services; Databricks; Elastic; Google; IBM; Microsoft; OpenAI

## Gartner Recommended Reading

[Use Multistructured Analytics for Complex Business Decisions](#)

[Quick Answer: What Are the Short-Term and Midterm Implications of ChatGPT for Data and Analytics?](#)

[Buyer's Selection Spotlight: Insight Engines](#)

[Magic Quadrant for Cloud AI Developer Services](#)

[Working With Semistructured and Unstructured Datasets](#)

## Prompt Engineering

**Analysis By:** Frances Karamouzis, Afraz Jaffri, Jim Hare, Arun Chandrasekaran, Van Baker

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Emerging

### Definition:

Prompt engineering is the discipline of providing inputs, in the form of text or images, to generative AI models to specify and confine the set of responses the model can produce. The inputs prompt a set that produces a desired outcome without updating the actual weights of the model (as done with fine-tuning). Prompt engineering is also referred to as “in-context learning,” where examples are provided to further guide the model.

### Why This Is Important

Prompt engineering is the linchpin to business alignment for desired outcomes. Prompt engineering is important because large language models (LLMs) and generative AI models in general are extremely sensitive to nuances and small variations in input. A slight tweak can change an incorrect answer to one that is usable as an output. Each model has its own sensitivity level, and the discipline of prompt engineering is to uncover the sensitivity through iterative testing and evaluation.



## Business Impact

Prompt engineering has the following business impacts:

- **Performance:** It helps improve model performance and reduce hallucinations.
- **Business alignment:** It allows subject data scientists, subject matter experts and software engineers to steer foundation models, which are general-purpose in nature, to align to the business, domain and industry.
- **Efficiency and effectiveness:** Alternative options, such as building a model from scratch or fine-tuning, can be much more complex, drive longer time to market and be more expensive.

## Drivers

- **Balance and efficiency:** The fundamental driver for prompt engineering is it allows organizations to strike a balance between consuming an “as is” offering versus pursuing a more expensive and time-consuming approach of fine-tuning. Generative AI models, and in particular LLMs, are pretrained, so the data that enterprises want to use with these models cannot be added to the training set. Instead, prompts can be used to feed content to the model with an instruction to carry out a function.
- **Process or task-specific customizations or new use cases:** The insertion of context and patterns that a model uses to influence the output generated allows for customizations for a particular enterprise or domain, or regulatory items. Prompts are created to help improve the quality for different use cases — such as domain-specific question answering, summarization, categorization, and so on — with or without the need for fine-tuning a model, which can be expensive or impractical. This would also apply to creating and designing new use cases that utilize the model’s capability for image and text generation.
- **Validation and verification:** It is important to test, understand and document the limits and weaknesses of the models to ensure a reduced risk of hallucination and unwanted outputs.

## Obstacles

- **Embryonic nature of the discipline:** Prompt engineering processes and roles are either unknown or enterprises have a low level of understanding and experience. Gartner webinar polling data (over 2,500 responses; see [Executive Pulse: AI Investment Gets a Boost From ChatGPT Hype](#)) revealed that approximately 60% of respondents self-reported that they had not heard of prompt engineering. And 90% of those same respondents revealed that their organization did not currently have prompt engineers.
- **Role alignment:** Data scientists are critical to understanding the capabilities and limits of models, and to determine whether to pursue a purely prompt-based or fine-tuning-based approach (or combination of approaches) for customization. The ultimate goal is to use machine learning itself to generate the best prompts and achieve automated prompt optimization. This is in contrast to an end user of an LLM who concentrates on prompt design to manually alter prompts to give better responses.
- **Lack of business alignment:** There is often a lack of consensus on prompt engineering's business approach, as well as agreed-upon standards, methodology and approaches. This has led to fierce debates on the value of prompt engineering and how to establish governance.
- **Risk:** Beyond the early stages of awareness and understanding, the biggest obstacle may be that prompt engineering is focused on verification, validation, improvement and refinement; however, it's not without risk. Prompt engineering is not the panacea to all of the challenges. It helps to manage risk, not remove it completely. Errors may still occur, and potential liability is at stake.

## User Recommendations

- Rapidly build awareness and understanding of prompt engineering in order to quickly start the journey of shape-shifting the appropriate prompt engineering discipline and teams.
- Build critical skills across a number of different team members that will synergistically contribute critical elements. For example, there are important roles for data scientists, business users, domain experts, software engineers and citizen developers.
- Communicate and cascade the message that prompt engineering is not foolproof. Rigor and diligence need to permeate and work across all the enterprise teams to ensure successful solutions.

## Sample Vendors

FlowGPT; HoneyHive; LangChain; PromptBase; Prompt Flow; PromptLayer

## Gartner Recommended Reading

[Quick Answer: How Will Prompt Engineering Impact the Work of Data Scientists?](#)

[Quick Answer: What Impact Will Generative AI Have on Search?](#)

[Accelerate Adoption of Generative AI by Offering an FMOps- or a Domain-Specific Partner Ecosystem](#)

[Glossary of Terms for Generative AI and Large Language Models](#)

## Natural Language Query

Analysis By: David Pidsley, Rita Sallam

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

**Definition:**

Natural language query (NLQ) allows business users to query information using business terms typed into a search box or chatbot, or via voice. Vendors' techniques differ in analytical complexity of queries, data volumes and types supported. These keyword searches translate terms into natural language questions using natural language processing technologies and LLM like ChatGPT. Some support querying structured data, and others enable semantic search of multistructured information.

**Why This Is Important**

- Business users need to make faster data-driven decisions and get context-enriched analysis that includes reasoning about location and time-sensitive situations.
- Despite significant advances in the usability of the point-and-click visual-based analytics, business intelligence (BI) platforms and other knowledge bases, traditional access paradigms are still too hard for most business users.
- Flattening the learning curve for BI platform users enables adoption by the remaining two-thirds of employees in organizations that do not use them.

**Business Impact**

NLQ drives adoption by nontechnical users, offering the ability to ask questions to gain insights, overcoming resistance to visual-based self-service analytics interfaces. NLQ is an increasingly important interface for analytic content development and consumption in data-driven decision making accessible to those unfamiliar with SQL. For data pipelines to enable multistructured analytics across a spectrum of structured data and unstructured content, NLQ can unify a multiexperience user interface.

## Drivers

- Foundation models like BERT, large language models (LLMs) and ChatGPT see NLQ repositioned at the Peak of Inflated Expectations and a high benefit rating with less than 2 year time to plateau.
- Generative AI hype is accelerating NLQ capabilities with advanced text analytics and deep learning as catalysts of natural language technologies, including natural language generation (NLG) and NLQ. They enable two-way communication between the human questioner and the machine-generated answer based on the data.
- Demand for generative D&A is substantial with the substantial increase in entrants in 2023. Established ABI platform vendors responded to ChatGPT by improving support for and innovations in NLQ, which is a well-established critical capability of the platforms. Adoption continues to grow as NLQ awareness, availability and solution capabilities improve.
- Orchestration of the entire analytics workflow will increasingly become NLQ-driven and used to manage the analytics and application development activities.
- Augmented analytics capabilities make the analytics consumer of tomorrow a power user by today's standards. Most analytics consumers enter the data story workflow when viewing content that has been created from prepared components and existing data visualizations. Their interaction is typically followed by NLQ or conversational analytics.
- NLQ is becoming central to personalized, consumer-oriented user experiences that combine augmented analytics or automated insights into automated data stories, scenario analysis and conversational analytics. Analytics collaboration enables NLQ engines to learn from team-usage preferences.
- Increasingly mobile workforces using handheld devices and voice interfaces need NLQ to interpret geospatial questions and immediately deliver location-based answers and business insights as a best-fit map visualization. Geospatial analytics and algorithm advances enable NLQ to deliver geospatial reasoning of distance, route calculations and analytics about entities near, farther than or within a certain proximity or boundary, based on business-defined regions or geocoded reference data.

## Obstacles

- Limitations in real-time type ahead search-bar suggestions can frustrate users, reduce usefulness and hinder adoption. Some users may not understand the implicit structure of underlying data, rendering queries uninterpretable by the NLQ parser.
- Unindexed datasets often hinder bringing search into an ABI platform. The effort/costs to map/model wide data are high, although generative AI is enabling NLQ of unstructured data to expand the scope and enable multistructured analytics.
- A substantial variety exists in the analytical complexity of queries, NLQ reasoning, support for suggestions for the next questions to ask, NLG to explain findings and support for large data volumes, structured and formats.
- Poor support of spoken languages beyond English, limited domain and industry ontologies, difficulty in configuration, and the need to be predefined in advance means optimizing NLQ implementations often requires customizing the platform and curating synonyms.
- Consistency is lacking for where users can ask questions across platforms and where implementations embed NLQ into the decision making or business process.

## User Recommendations

- Help users adopt NLQ for decision making and orchestrating workflows.
- Promote NLQ-specific data literacy training for augmented consumers, business analysts and analytics developers.
- Assess the NLQ roadmaps of vendors and augmented analytics startups.
- Prioritize vendors based on how and what a platform learns (from activate metadata for personalization) via a proof of concept with real data and users.
- Evaluate how NLQ fits into analytics solution architectures. Involve IT in evaluation, data preparation and deployment of ABI platforms.
- Support multiple use cases with multiexperience UIs including evaluating enterprise conversational AI platforms.
- Invest in design thinking on dialogue flows and in competencies to connect conversational analytics to the ecosystem of APIs; for example, ABI platforms and insight engines that enable semantic search and analyzing results sets of wide data with multistructured analytics.

## Sample Vendors

ConverSight; iGenius; Pyramid Analytics; Qlik; Tellius; ThoughtSpot

## Gartner Recommended Reading

[Magic Quadrant for Analytics and Business Intelligence Platforms](#)

[Is Your Business Intelligence Enabling Intelligent Business?](#)

[Quick Answer: What Are the Short-Term and Midterm Implications of ChatGPT for Data and Analytics?](#)

[Magic Quadrant for Insight Engines](#)

[Magic Quadrant for Enterprise Conversational AI Platforms](#)

## Emotion AI

**Analysis By:** Annette Zimmermann

**Benefit Rating:** Transformational

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Emerging

### Definition:

Emotion artificial intelligence (AI) technologies (also called affective computing) use AI techniques to analyze the emotional state of a user (via computer vision, audio/voice input, sensors and/or software logic). Emotion AI can initiate responses by performing specific, personalized actions to fit the mood of the customer.

### Why This Is Important

Emotion AI is considered transformational as it turns human behavioral attributes into data that will have a large impact on human-machine interface (HMI). Machines will become more “humanized” as they can detect sentiments in many different contexts. Furthermore, applying deep learning to computer vision or audio-based systems to analyze emotions in real time has spawned new use cases for customer experience enhancements, employee wellness and many other areas.

## Business Impact

Contact centers use voice analysis and natural language processing (NLP)-based algorithms to detect emotions in voice conversations, in personal chat conversations and chatbots. Computer vision (CV)-based emotion AI has already been used for more than a decade in market research with neuromarketing platforms that test users' reactions toward products. In addition, we see the technology expanding to other verticals and use cases, i.e., healthcare (diagnostic), sales enablement and employee wellness.

## Drivers

The increasing number of use cases we have identified indicates an increase in commercialization as emotion AI finds applicability in new domains:

- One of the drivers for detecting emotions/states is the need for a system to act more sympathetically. For instance, it creates anthropomorphic qualities for personal assistant robots (PARs) and virtual beings, making them appear more “human.” This “emotional capability” is an important element in enhancing the communication and interaction between users and a PAR.
- This can be an empathic avatar or an emotion-detection-enabled chatbot. A person's daily behavior, communication and decisions are based on emotions — our nonverbal responses in a one-to-one communication are an inseparable element from our dialogues and need to be considered in the human-machine interface (HMI) concept.
- Combinatorial technology solutions such as computer-vision-based and audio analytics, or language-based and computer vision, enable customer experience enhancements.
- Strongest adoption is currently happening in the context of contact centers where voice-based emotion analysis supports multiple use cases such as real-time analysis on voice conversations, emotion detection in chat conversations, emotional chatbots and more.
- Market research and neuromarketing tools are continuously leveraging emotion detection in various user scenarios including focus groups and product testing. Vendors have been extending their offerings toward remote/online interviews during 2020 — due to the pandemic.
- In the creation of virtual beings in customer service or other consumer-facing scenarios, emotional responses are a critical element.



- As the metaverse unfolds, virtual beings will play an important role as business models evolve and the entire ecosystem of this new digital world emerges.

## Obstacles

- Privacy concerns are the main obstacle to rapid adoption in the enterprise. This is especially a concern in real-life situations (vs. lab/research environments) for both consumer-facing (e.g., monitoring emotions in a retail environment via cameras) and employee-facing situations. Research environments like product testing have the advantage that the emotion AI is used for this specific purpose and the user (product tester) is fully aware that their emotions are being captured to improve usability or other features.
- Bias: When using facial expression analysis, models are likely to be retrained in different geographies to get the system to detect the different nuances present due to different cultural backgrounds.
- Variation across modalities. Certain emotions can be better detected with one technology mode than with another. For instance, “irony” can be detected using voice-based analysis while this is close to impossible to detect with facial expression analysis.

## User Recommendations

- Review vendors’ capabilities and reference cases carefully. As the market is currently very immature, most vendors are focused on two or three use cases in two or three industries. At the same time, identifying and processing human emotion is currently a gray area, especially in the EU. The EU Commission has started an initiative to review the ethical aspects of AI technologies, and emotion AI will certainly be part of this debate.
- Enhance your customer analytics and behavioral profiling by applying emotion AI technologies bringing your customer experience strategy to the next level.
- Be use-case-driven. The use case will determine the emotion AI technology to be used and vendor selection.
- Appoint responsibility for data privacy in your organization — a chief data privacy officer or equivalent.
- Work with your vendor on change management in order to avoid user backlash due to sensitive data being collected.

## Sample Vendors

Behavioral Signals; Cogito; DAVI; Intelligent Voice; kama.ai; MorphCast; Soul Machines; Superceed; Symanto; Uniphore

## Gartner Recommended Reading

[Competitive Landscape: Emotion AI Technologies](#)

[Emerging Tech: Computer Vision, Voice Analysis and CGI Evolve Into Emotionally Intelligent Virtual Beings](#)

[Tool: Vendor Identification for Natural Language Technologies](#)

## Foundation Models

Analysis By: Arun Chandrasekaran

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

### Definition:

Foundation models are large-parameter models that are trained on a broad gamut of datasets in a self-supervised manner. They are mostly based on transformer or diffusion deep neural network architectures and will potentially be multimodal in the near future. They are called foundation models because of their critical importance and applicability to a wide variety of downstream use cases. This broad applicability is due to the pretraining and versatility of the models.

### Why This Is Important

Foundation models are an important step forward for AI due to their massive pretraining and wide use-case applicability. They can deliver state-of-the-art capabilities with higher efficacy than their predecessors. They've become the go-to architecture for NLP, and have also been applied to computer vision, audio and video processing, software engineering, chemistry, finance, and legal use cases. Primarily text-based, large language models (LLMs) are a popular subset of foundation models. ChatGPT is based on one (GPT-4).

## Business Impact

With their potential to enhance applications across a broad range of natural language use cases, foundation models will have a wide impact across vertical industries and business functions. Their impact has accelerated, with a growing ecosystem of startups building enterprise applications on top of them. Foundation models will advance digital transformation within the enterprise by improving workforce productivity, automating and enhancing CX, and enabling rapid, cost-effective creation of new products and services.

## Drivers

Foundation models:

- **Require only limited model customization to deliver effective results.** Foundation models can effectively deliver value through prebuilt APIs, prompt engineering or further fine-tuning. While fine-tuning may deliver the best value because of customization possibilities, the other two options are less complex.
- **Deliver superior natural language processing.** The difference between these models and prior neural network solutions is stark. The large pretrained models can produce coherent text, code, images, speech and video at a scale and accuracy not possible before.
- **Enable low-friction experimentation.** The past year has seen an influx of foundation models, along with smaller, pretrained domain-specific models built from them. Most of these are available as cloud APIs or open-source projects, further reducing the time and cost to experiment.
- **Have accelerated AI innovation with massive model sizes.** Examples include OpenAI's GPT-4; Google's AI's PaLM; Google DeepMind's Gopher and Chinchilla; Meta AI's LLaMA; and Alibaba's M6. In addition, companies such as Hugging Face, Stability AI and EleutherAI have open-sourced their models.

## Obstacles

Foundation models:

- **Do not deliver perfect results.** Although a significant advance, foundation models still require careful training and guardrails. Because of their training methods and black-box nature, they can deliver unacceptable results or hallucinations. They also can propagate downstream any bias or copyright issues in the datasets.

- **Require appropriate skills and talent.** As with all AI solutions, the end result depends on the skills, knowledge and talent of the trainers, particularly for prompt engineering and fine-tuning.
- **Expand to impractical sizes.** Large models are up to billions or trillions of parameters. They are impractically large to train for most organizations because of the necessary compute resources, which can make them expensive and ecologically unfriendly.
- **Concentrate power.** These models have been mostly built by the largest technology companies with huge R&D investments and significant AI talent, resulting in a concentration of power among a few large, deep-pocketed entities. This situation may create a significant imbalance in the future.

## User Recommendations

- **Create a strategy document** that outlines the benefits, risks, opportunities and execution plans for these models in a collaborative effort.
- **Plan to introduce foundation models into existing speech, text or coding programs.** If you have any older language processing systems, moving to a transformer-based model could significantly improve performance. One example might be a text interpretation, where transformers can interpret multiple ideas in a single utterance. This shift in approach can significantly advance language interfaces by reducing the number of interactions.
- **Start with models that have superior ecosystem support,** have adequate enterprise guardrails around security and privacy, and are more widely deployed.
- **Explore new use cases,** such as natural language inference, sentiment analysis or natural-language-based enterprise search, where the models can significantly improve both accuracy and time to market.
- **Designate an incubation team** to monitor industry developments, communicate the art of the possible, experiment with BUs and share valuable lessons learned companywide.

## Sample Vendors

Alibaba Group; Amazon; Baidu; Cohere; Google; Hugging Face; IBM; Microsoft; OpenAI; Stability AI

## Generative AI

Analysis By: Svetlana Sicular, Brian Burke

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

### Definition:

Generative AI technologies can generate new derived versions of content, strategies, designs and methods by learning from large repositories of original source content. Generative AI has profound business impacts, including on content discovery, creation, authenticity and regulations; automation of human work; and customer and employee experiences.

### Why This Is Important

Generative AI exploration is accelerating, thanks to the popularity of Stable Diffusion, Midjourney, ChatGPT and large language models. End-user organizations in most industries aggressively experiment with generative AI. Technology vendors form generative AI groups to prioritize delivery of generative-AI-enabled applications and tools. Numerous startups have emerged in 2023 to innovate with generative AI, and we expect this to grow. Some governments are evaluating the impacts of generative AI and preparing to introduce regulations.

### Business Impact

Most technology products and services will incorporate generative AI capabilities in the next 12 months, introducing conversational ways of creating and communicating with technologies, leading to their democratization. Generative AI will progress rapidly in industry verticals, scientific discovery and technology commercialization. Sadly, it will also become a security and societal threat when used for nefarious purposes. Responsible AI, trust and security will be necessary for safe exploitation of generative AI.

## Drivers

- The hype around generative AI is accelerating. Currently, ChatGPT is the most hyped technology. It relies on generative foundation models, also called “transformers.”
- New foundation models and their new versions, sizes and capabilities are rapidly coming to market. Transformers keep making an impact on language, images, molecular design and computer code generation. They can combine concepts, attributes and styles, creating original images, video and art from a text description or translating audio to different voices and languages.
- Generative adversarial networks, variational autoencoders, autoregressive models and zero-/one-/few-shot learning have been rapidly improving generative modeling while reducing the need for training data.
- Machine learning (ML) and natural language processing platforms are adding generative AI capabilities for reusability of generative models, making them accessible to AI teams.
- Industry applications of generative AI are growing. In healthcare, generative AI creates medical images that depict disease development. In consumer goods, it generates catalogs. In e-commerce, it helps customers “try on” makeup and outfits. In manufacturing, quality inspection uses synthetic data. In semiconductors, generative AI accelerates chip design. Life sciences companies apply generative AI to speed up drug development. Generative AI helps innovate product development through digital twins. It helps create new materials targeting specific properties to optimize catalysts, agrochemicals, fragrances and flavors.
- Generative AI reaches creative work in marketing, design, music, architecture and content. Content creation and improvement in text, images, video and sound enable personalized copywriting, noise cancellation and visual effects in videoconferencing.
- Synthetic data draws enterprises’ attention by helping to augment scarce data, mitigate bias or preserve data privacy. It boosts the accuracy of brain tumor surgery.
- Generative AI will disrupt software coding. Combined with development automation techniques, it can automate up to 30% of the programmers’ work.

## Obstacles

- Democratization of generative AI uncovers new ethical and societal concerns. Government regulations may hinder generative AI research. Governments are currently soliciting input on AI safety measures.
- Hallucinations, factual errors, bias, a black-box nature and inexperience with a full AI life cycle preclude the use of generative AI for critical use cases.
- Reproducing generative AI results and finding references for information produced by general-purpose LLMs will be challenging in the near term.
- Low awareness of generative AI among security professionals causes incidents that could undermine generative AI adoption.
- Some vendors will use generative AI terminology to sell subpar “generative AI” solutions.
- Generative AI can be used for many nefarious purposes. Full and accurate detection of generated content, such as deepfakes, will remain challenging or impossible.
- The compute resources for training large, general-purpose foundation models are heavy and not affordable to most enterprises.
- Sustainability concerns about high energy consumption for training generative models are rising.

## User Recommendations

- Identify initial use cases where you can improve your solutions with generative AI by relying on purchased capabilities or partnering with specialists. Consult vendor roadmaps to avoid developing similar solutions in-house.
- Pilot ML-powered coding assistants, with an eye toward fast rollouts, to maximize developer productivity.
- Use synthetic data to accelerate the development cycle and lessen regulatory concerns.
- Quantify the advantages and limitations of generative AI. Supply generative AI guidelines, as it requires skills, funds and caution. Weigh technical capabilities with ethical factors. Beware of subpar offerings that exploit the current hype.
- Mitigate generative AI risks by working with legal, security and fraud experts. Technical, institutional and political interventions will be necessary to fight AI's adversarial impacts. Start with data security guidelines.
- Optimize the cost and efficiency of AI solutions by employing composite AI approaches to combine generative AI with other AI techniques.

## Sample Vendors

Adobe; Amazon; Anthropic; Google; Grammarly; Hugging Face; Huma.AI; Microsoft; OpenAI; Schrödinger

## Gartner Recommended Reading

[Innovation Insight for Generative AI](#)

[Emerging Tech Roundup: ChatGPT Hype Fuels Urgency for Advancing Conversational AI and Generative AI](#)

[Emerging Tech: Venture Capital Growth Insights for Generative AI](#)

[Emerging Tech: Generative AI Needs Focus on Accuracy and Veracity to Ensure Widespread B2B Adoption](#)

[ChatGPT Research Highlights](#)



## AI-Enabled Content Moderation

Analysis By: Gabriele Rigon, Bern Elliot

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

### Definition:

AI-enabled content moderation (AICM) is the technology that automates the identification and/or removal of inappropriate user-generated content on online platforms, such as social media or chats, and on organization-internal channels like instant messaging products. AICM should review content with limited or no human intervention to ensure it meets the standards defined in the platform's guidelines and to foster healthy interaction practices.

### Why This Is Important

Content moderation is key to enterprise trust and safety. Use cases for this technology comprise channels through which customers interact with organizations, and employees interact with each other. Human moderation struggles to keep up with the increase of user-generated content, especially on online platforms. AICM can automate part of the job. Virtually all content can be screened and classified by AICM, which enables moderation at a much larger scale compared with manual approaches.

### Business Impact

- AICM can enhance enterprises' capability to mitigate liability and reputation risks at scale in real time.
- Since AICM works without requiring hundreds of human moderators to view and manually moderate content, it is then enabling significant reduction of labor costs and improved efficiency.
- Human moderators are exposed to all sorts of inappropriate content, which may cause mental illness such as burnout. AICM can reduce such exposure and positively impact the workforce's mental well-being.

## Drivers

- **Volumes:** Handling the volumes of user-generated content manually, especially in online communities, requires up to thousands of human moderators. Businesses need to alleviate pressure on the workforce, reduce costs and process content at scale. AICM can meet these needs by enabling the analysis of enormous volumes of data systematically and fast.
- **Use cases:** AICM is often associated with online platforms, and the primary audience for this technology is marketing. However, some offerings can integrate with enterprise-internal messaging products to monitor employee communications. AICM is then potentially valuable in all use cases in which malicious user-generated content needs to be intercepted, such as controlling the spread of misinformation. Vendors also started providing AICM services via API to filter interactions with large language models (LLMs) via conversational user interfaces. AICM can also help identify harmful content in training data when building ML models.
- **Additional capabilities:** Beyond the identification of noncompliant content, additional capabilities found in AICM products allow monitoring healthy, engaging user behaviors, which can spread and bring benefits to online communities. AICM solutions can also help qualify leads from personalized social media campaigns and high-volume social channels.
- **Underlying tooling:** Emerging AI techniques for natural language technologies (NLTs), and LLMs in particular, are being applied in innovative ways to process large-scale data sources.
- **Public awareness:** The general public and authorities' awareness of online bullying and abuse is increasing. Regulations are evolving to address risks entailed by such harmful behaviors online. For example, the Online Safety Bill proposed by the U.K. Parliament goes in this direction. AICM provides the tooling to improve screening and prevention of abuse online and to mitigate liability risks.

## Obstacles

- Enterprise-online trust and safety has not been a priority for many organizations and is a relatively new concept. While each potential enterprise customer represents a large number of users in their network, a limited number of enterprises need to monitor very high volumes of interaction among a large user population.
- AICM models need to be customized to attain an acceptable level of accuracy in specific use cases and domains. Words have different meanings and nuances in different communities, jargon and dialects. Lack of customization options may impact the overall accuracy of AICM models.
- Threats evolve to adapt to technology trying to stop them, so bad actors are constantly trying to find new ways to fool AICM (e.g., creating their own codes).
- Toxic content is moving beyond text and may include images and video. Currently, the data processing capabilities required to overcome this challenge are beyond the scope of text-based AICM solutions.

## User Recommendations

- Moderate content by reviewing forums and social media where customers and employees are interacting directly. Doing so can ensure the online safety of participants (such as from harassment).
- Define a role responsible for online-employee health, trust and safety, whose responsibility includes the use of AICM tools to manage large-scale social interactions.
- Leverage AICM tools to augment human moderators, not to replace them. Human supervision is key to validate decisions recommended by AICM, especially when it comes to taking actions based on the detection of potential harmful behaviors. Favor solutions that embed/integrate with human-in-the-loop (HITL) components.
- Improve the accuracy of the AICM models by customizing and continuously optimizing them based on the evolution of the user-generated content being monitored.

## Sample Vendors

Amazon; Cohere; Meta; Microsoft; Nightfall; OpenAI; Pattr; Spectrum Labs; TaskUs

## Gartner Recommended Reading

[Cool Vendors in Natural Language Technology for Processing Enormous Volumes of Unstructured Data](#)

[Best Practices for the Responsible Use of Natural Language Technologies](#)

[Tool: Vendor Identification for Natural Language Technologies](#)

## Text Summarization

Analysis By: Shubhangi Vashisth, Stephen Emmott

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

### Definition:

Text summarization is a range of techniques that produce short textual summaries representing longer or multiple texts. Its principal purpose is to reduce the time and effort required to get insight from content, either by signaling the value of reading the source(s) or by delivering value directly in the form of the summary.

### Why This Is Important

Text summarization benefits not only humans but also nonhuman entities, such as insight engines, where summaries are used to enhance indexing and knowledge management; in prompt engineering; and more. Appropriate use of the technology can reduce task time and effort, improving access to insight for all employees of an organization. Thus, text summarization can reduce the time required for decisions, actions and ideation.

### Business Impact

Text summarization is valuable across many use cases:

- The principal use remains real-time, reader-led summarization of found/shared documents (transcriptions, meeting summaries, etc.).
- The functional domains likely to benefit most from text summarization are legal, finance, HR, supply chain and research. Government and education benefit too.

- Integration with insight engines, knowledge graphs and content management systems is rising.
- It generates short summaries of customer interactions as part of agent assist products in conversational AI systems.

## Drivers

Key drivers of text summarization technology, which dates back to the 1950s, include:

- Large language models (LLMs) have massively democratized summarization, resulting in higher demand from end users, who want to apply the technology to a broad spectrum of use cases.
- Text summarization reduces the employee time and effort needed to acquire insight from content, leading to faster decision making.
- Approaches to AI herald abstraction, including transformers, have advanced. The abstraction approach has become a viable way to apply text summarization at the domain, foundation and situation levels of the application portfolio.

Two approaches drive text summarization:

- **Extraction:** Statistically significant words, phrases and sentences are extracted, leaving the rest. Extraction is comparable to, but falls short of, skim-reading the source. Of the two text summarization approaches, extraction remains the dominant form.
- **Abstraction:** Supervised learning is used to map salient words and phrases from the source to novel sentences in the target.

## Obstacles

Although the technology continues to improve in performance, it falls short of human-derived summarization for the following reasons:

- **Quality:** The extraction approach is well-established but produces summaries that are further removed from the quality of human-generated summaries.
- **Contextualization:** The abstraction approach has emerged in recent years and produces summaries that are closer to human-generated summaries. However, it requires training in context and is therefore more challenging and costly to apply.

- **Omission of key information:** Missing vital information from a summary reduces the quality of insight needed from the source document, adversely impacting analysis, decision and action. Such risks are intolerable in many vertical industries and functional domains, such as compliance within banking or research within life sciences and healthcare.
- **Hallucination and errors:** Content produced by text summarization techniques that leverage LLMs should not be considered entirely reliable and error-free. There are concerns around the trust and explainability of the models as well.

## User Recommendations

- List the business functions and use cases that require text summarization capabilities, and check if these capabilities are embedded within existing applications. Text summarization is often deployed as a component of the applications that are specific to, or customized for, these situational uses.
- Leverage such techniques for contexts in which the risk and exposure level of the summaries is low and some form of human validation/postediting is performed.
- Review and, if needed, fact-check the information included in the content produced by text summarization techniques. Select products that offer explainability and traceability.
- Approach the technology as a prepackaged business capability that can be reused across the application portfolio. This prepackaged solution will help application leaders looking for an abstraction-based approach.

## Sample Vendors

Amazon Web Services; Datamatics; expert.ai; Google; IBM; iManage; InMoment (Lexalytics); Microsoft; Salesforce (Tableau)

## Gartner Recommended Reading

[Market Guide for Text Analytics](#)

[Tool: Vendor Identification for Natural Language Technologies](#)

[Best Practices for the Responsible Use of Natural Language Technologies](#)

[Innovation Insight for Generative AI](#)

## Natural Language Generation

Analysis By: Bern Elliot

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Emerging

### Definition:

Natural language generation (NLG) solutions automatically convert structured and unstructured data into text-based narratives. This is achieved either through rule-based methods, which have been in use for over 15 years, or the one based on recent large language model (LLM) and generative AI methods. Both approaches have strengths and weaknesses; however, they can also be combined into hybrid solutions.

### Why This Is Important

NLG solutions improve understanding and operational efficiencies by making it easier to generate content or appraise, via summary or extraction, large or complex material and data. Recent advances in generative AI have enabled significantly more fluid and creative language generation. However, those methods may also introduce erroneous content, an unacceptable risk in many use cases. Leading solutions combine rule-based and generative-AI-based approaches, enabling selectable control over which method is used.

### Business Impact

NLG supports multiple productivity-enhancing use cases by augmenting human editors and writers. It can increase the speed at which textual information can be produced and shared, reducing cost and time to market for new content on multiple channels. It also allows applications to communicate with users via conversational UIs in a more fluid and natural manner, such as by supporting interpretation of complex information like analytic reports. Generative AI functionality has expanded the scope, variability and creativity of what can be generated.

### Drivers

The most common uses cases for NLG driving adoption fall into several categories:

- **Enhance understanding of business analytics:** For instance, integrating NLG functionality with existing analytics and business intelligence (BI) and data science initiatives.
- **Article-type short summaries:** For instance, writing summaries or analysis of business data, financial data, wealth management information, personalized marketing copy or sports — perhaps in conjunction with abstractive text summarization technology.
- **Conversation responses:** For instance, writing personalized communications to customers via email or text.
- **Easing data access:** For instance, writing short, prose-based product descriptions as per database product information. These might then be posted as a reply to website information requests.
- **Generating variants of outbound messaging and marketing copy:** The last four years have seen growth in the number of short-form NLG specialist vendors.

Emerging, complex use cases include:

- **The combination of NLG with automated pattern/insight detection and self-service data preparation:** This can drive the user experience of next-generation augmented analytics platforms. Users have varying degrees of analytics skills to correctly interpret and act on statistically significant relationships in visualization. This use case could also expand the benefits of advanced analytics to a wider audience of business users, as well as make existing analysts and data scientists more efficient.
- **Tighter integration with BI workflows and experiences:** Context-based narration will reinforce mobile BI use cases, where a lack of screen space is a major impediment to information consumption. It will also expand the use of conversational analytics that combine natural language query (NLQ), chatbots and NLG via virtual personal assistants.
- **Complementing conversational experiences:** Conversational solutions, including virtual assistants, will be able to use NLG methods to enable more complex and natural-sounding interactions.



## Obstacles

- The rule-based NLG solutions are mature but are limited in the range of language and complexity of content types that they can generate. However, their output is deterministic and accurate, which is required for many applications, and hence, their use cannot be eliminated.
- Generative-AI-based NLG solutions are emerging, with best practices for their architecture and use still being defined. Additionally, the potential introduction of inaccurate content makes them unsuitable for many use cases. While generation can be “grounded” or based on a narrow set of information, it only reduces, and cannot completely eliminate, the likelihood of errors.
- Hybrid approaches, while promising, are new and may be more complex to use as they involve two different underlying techniques.
- Generative AI approaches require a foundation model which is difficult to develop. Many will rely on existing foundation models, including Generative Pre-trained Transformer (GPT). Usage may incur significant expenses.

## User Recommendations

- Be aware of a solution’s maturity, particularly in terms of its ability to deliver hybrid functionality.
- Be aware of the platform data integration and preparation requirements, the platform’s self-learning capabilities, and the upfront set-up and configuration required.
- Define the languages that need to be supported, the extent of narration, the degree of story automation and control supported, and the accuracy requirements of the findings and narration.
- Investigate and understand potential drawbacks relating to multilingual user scenarios, as NLG requires specific libraries for each language in use. Additionally, industry-specific use cases need to be considered carefully with respect to jargon, tone and specialized ontologies.
- Identify how NLG could be attractive to organizations wishing to make their analytics, BI solutions and other classes of visual information accessible to visually impaired audiences (for instance, to comply with the U.S. Americans with Disabilities Act and similar mandates in other countries).

## Sample Vendors

Arria NLG; AX Semantics; Marlabs; neuroflash; OpenAI; Retresco; Salesforce (Narrative Science); ThoughtSpot; Yseop

## Gartner Recommended Reading

[ChatGPT Research Highlights](#)

[Magic Quadrant for Enterprise Conversational AI Platforms](#)

[Magic Quadrant for Cloud AI Developer Services](#)

[Magic Quadrant for Analytics and Business Intelligence Platforms](#)

## Sliding into the Trough

### Conversational User Interfaces

**Analysis By:** Gabriele Rigon, Stephen Emmott, Van Baker, Bern Elliot, Frank O'Connor

**Benefit Rating:** Transformational

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Adolescent

#### Definition:

Conversational user interfaces (CUIs) are human-computer interfaces that enable natural language interactions for the purpose of fulfilling a request, such as answering a question or completing a task. The sophistication of a CUI can vary from understanding basic queries to handling complex multiturn dialogs, so CUIs range from Q&A bots to more advanced virtual assistants (VAs). CUIs fundamentally shift the interaction medium from traditional point-and-click to natural-language-driven.

#### Why This Is Important

UIs provide direct control between the user and the applications they are operating. In a CUI, this responsibility shifts from application-specific controls to conversational controls, and the CUI is determining the intent and acting upon it. This makes CUIs more widespread as agent (acting) UIs for software, devices and the Internet of Things. AI-enabled CUIs can provide a single, intuitive, common interface to multiple application functions across the entire organization.

#### Business Impact

Training, onboarding, escalations, productivity, empowerment and responsibility all change with CUIs and need to be embraced as part of CUI projects. AI-enabled CUIs can dramatically standardize and improve the usability of a variety of applications across all business functions, such as CRM, the digital workplace and ERP, hence improving efficiency. They can also benefit customer experience when used to automate support in the form of self-service chatbots or VAs.

#### Drivers

- **Users' expectations and generative AI:** Users increasingly expect to be able to hold conversations with and ask natural language questions of the applications they use. CUIs are beginning to complement or even replace traditional interfaces in a variety of applications, such as search and insight engines, business intelligence platforms and productivity software, such as document and spreadsheet applications. The trend toward the enablement of interactions in natural language between users (customers and employees) and software has been significantly accelerated by the hype around generative AI and ChatGPT.
- **Conversational AI platforms:** The underlying technology supporting custom-developed CUIs (like chatbots and VAs) built on top of conversational AI platforms (CAIPs) has matured significantly in the last few years. Vendors are investing in core AI technologies, such as large language models (LLMs), to improve components such as natural language understanding. They are also expanding their capabilities to support broader use cases beyond self-service chatbots and toward broader B2C and B2E automation.
- **Search:** CUIs will be increasingly used for knowledge search and retrieval based on document ingestion. Some technologies driving this include LLM-enabled enterprise applications, such as Microsoft 365 Copilot, as well as ChatGPT-like Q&A chatbots and LLM-powered VAs. This is also causing the market to be flooded with dedicated add-ons and even new vendors.
- **Multimodal interactions:** Generative AI methods are increasing the availability of multimodal interactions, such as those based on images, videos, audio and other sensory data. As a matter of fact, beyond text, voice is emerging as a primary modality of interaction between users and CUIs. This can add a powerful enhancement to the communications. Multimodality can solve some of the problems of the current generation of LLMs. Multimodal language models will also unlock new applications that were impossible with text-only models.

## Obstacles

- Developing CUIs is intrinsically complex and requires more effort than graphical UIs. More sophistication has to be built into VAs' conversational capabilities to deal with a range of users and edge cases. CUIs' predictions about users' intents can be wrong, so the CUI designer has to keep ambiguity in mind.
- Lack of CUI personality, poor accuracy and conversational design, as well as unreliability of answers generated by LLMs, can affect user sentiments negatively and, as a consequence, adoption and ROI.

- CUIs are available from many sources, whether offered by applications, CAIPs or through separate augmentation. For example, transactional conversational AI use cases require capabilities that only platforms can provide. Q&A scenarios may also be supported by architectures primarily leveraging search and LLMs. Understanding the sophistication and the limitations of these and other approaches is not trivial. This may lead buyers to choose the wrong tooling and many CUIs to fail.

## User Recommendations

- Treat CUIs as transformative, and plan on them becoming the dominant interaction model between users and applications.
- Prioritize the requirements of your custom CUIs in terms of sophistication, integration and control. Do not underestimate the risks of building CUIs that do not meet enterprise-grade performance, accuracy and security standards.
- Develop your strategy for consolidation upon one or few conversational AI platforms or approaches, avoiding challenges that derive from the proliferation of CUIs deployed by different business units in different regions.
- Educate stakeholders around benefits and limitations of generative-AI-enabled CUIs, and encourage well-informed employees to experiment with such CUIs.
- Prepare for new roles and skills in the enterprise. Dialogue designers and AI trainers, for example, are needed to enable custom CUI initiatives. Citizen developers will acquire prompt engineering and model management skills to leverage generative-AI-enabled CUIs effectively.

## Sample Vendors

Amelia; Avaamo; Cognigy; Google; IBM; Kore.ai; Omilia; OneReach.ai; OpenAI

## Gartner Recommended Reading

[Magic Quadrant for Enterprise Conversational AI Platforms](#)

[Critical Capabilities for Enterprise Conversational AI Platforms](#)

[Competitive Landscape: Conversational AI Platform Providers](#)

[Emerging Tech Roundup: ChatGPT Hype Fuels Urgency for Advancing Conversational AI and Generative AI](#)

## Innovation Insight for Generative AI

### Intelligent Document Processing

Analysis By: Shubhangi Vashisth, Stephen Emmott, Anthony Mullen

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Adolescent

#### Definition:

Intelligent document processing (IDP) solutions extract data to support automation of high-volume, repetitive document processing tasks and to provide analysis and insight. IDP uses natural language technologies and computer vision to extract data from structured and unstructured content, especially from documents, to support automation and augmentation.

#### Why This Is Important

IDP is increasingly important to create operational efficiencies in business processes that need to extract information from semistructured and unstructured data as part of automation of any workflow. Such tasks are routine, repetitive and primarily dependent on human effort. IDP caters to a wide variety of use cases — from digitization initiatives to more complex processes such as document-centric taxation processing and pension fund management. IDP is one of the technologies within a spectrum of technologies that enable hyperautomation.

#### Business Impact

IDP can benefit the business by:

- Reducing the human labor needed to process documents and improving document-based workflows
- Extracting relevant data from different input formats for further analysis, validation and/or automation
- Preprocessing unstructured data for analysis
- Automating document and email classification and extraction

- Enabling discovery and insight
- Creating workflows to support process automation or integration with existing automation solutions

## Drivers

Key drivers include:

- The desire to consolidate document processing across multiple applications into one component.
- The need to ingest data from diverse sources and formats (e.g., PDFs, images) and extract information from it.
- Pressure to improve the accuracy and efficiency of extraction and automation processes.
- The desire to leverage generative AI for document classification and data extraction. Many vendors are already contextualizing large language models (LLMs), such as BERT and GPT-3, for specific industries and use cases.
- Enhanced capabilities to denoise and preprocess semistructured and unstructured data. Many vendors are expanding with additional capabilities for intelligent content processing (ICP) to process various content types, such as video, audio and images.
- Support for additional capabilities, such as document classification, metadata extraction, knowledge graphs, search and natural language question answering.
- Increased leverage of human in the loop (HITL) training methods to simplify adoption, ease deployment and continuously improve automation accuracy.

Examples of use cases span many enterprise departments and vertical industries, including:

- **Accounts payable/receivable:** Processing of invoices, purchase orders, payments, expense reports and receipts.
- **Healthcare:** Processing of medical forms.
- **Banking and financial services:** Processing of loan applications, driver licenses and other collateral; customer onboarding; environmental, social and governance (ESG) initiatives; and compliance.

- **Government:** Processing of forms, driver licenses, passports and other IDs.
- **Manufacturing:** Processing of equipment maintenance records, RFPs, business contracts and operating agreements.
- **HR:** Employee onboarding, travel and expenses.

## Obstacles

- **Complex, consolidating markets:** The market has a competitive vendor landscape, with dedicated solutions and offerings from adjacent technology markets. These markets include insight engine vendors, OCR vendors, RPA vendors, cloud providers and, increasingly, service providers. Selecting the right solution gets tricky, as vendors offer overlapping capabilities and differentiation is low.
- **Integration challenges:** Many organizations already have either a homegrown solution or an existing IDP tool, but are looking for enhanced features, such as sophisticated text analytics, to cater to wider use cases and growing business needs. However, a single tool may not be able to cater to all requirements, and integration complexity makes it challenging to have multiple tools.
- **Category bleed, which confuses buyers:** With semantic platforms, insight engines, RPA and conversational AI vendors all offering IDP-like solutions to interpret and mine document form factors, buyers may not feel compelled to purchase an additional format-specific (document) solution.



## User Recommendations

- Evaluate the entire business process to understand where and how IDP solutions can be integrated. Treat IDP as a component that integrates with other platforms/applications.
- Adopt industry- and/or business-domain-focused solutions for a quick time to implementation.
- Align with stakeholders on accuracy and efficiency baselines for the process.
- Investigate the difference between placed-framed extraction and semantic-framed extraction. The former is not IDP, and only the latter can scale to unstructured content. If you want to use IDP as a launchpad for broader handling of semistructured and unstructured data, evaluate insight engine or semantic AI platforms that offer IDP along with other services.
- Design the HITL validation process either by leveraging internal sources or by outsourcing the task to the IDP solution provider.
- Discuss specialized requirements, such as the ability to process documents in entirely new formats, data preprocessing needs and SLAs around processing time.
- Compare the ease of integration of new tools, if looking to complement the capabilities of an existing solution.

## Sample Vendors

Alkymi; Altilia; Applica; DocDigitizer; Eigen Technologies; Hyperscience; Infrd; Indico Data; Kofax; OpenText

## Gartner Recommended Reading

[Infographic: Understand Intelligent Document Processing](#)

[Market Guide for Intelligent Document Processing Solutions](#)

[Quick Answer: How to Prioritize Requirements in the RFP for Intelligent Document Processing](#)

[Tool: RFP for Intelligent Document Processing](#)

[Intelligent Document Processing Growth Opportunities: Top Strategies for Tech CEOs](#)

## Speech Analytics

Analysis By: Steve Blood

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

### Definition:

Speech analytics is a combination of phrase-matching, phonetic indexing and transcription technologies used to extract contextual insights from recorded and real-time voice streams. Speech analytics surfaces insights such as topics, categories and emotional engagement in conversations. AI and machine learning improve categorization and accuracy of analysis. Use cases range from customer service to sales and marketing.

### Why This Is Important

Rapid advancements in speech-to-text technologies and natural language technology in recent years, including the use of AI techniques such as machine learning, have improved categorization and analysis accuracy. This is enabling organizations to be more accurate and confident when using the insights surfaced from analyzing voice conversations for compliance, training, analytics and collaboration initiatives.

### Business Impact

Speech analytics can be used to provide insights into:

- Employee performance and compliance by analyzing conversations for keywords, acoustics and talkover to highlight areas where further training could be required.
- Emotional analysis to provide feedback on customer and employee sentiment, individually or in aggregate.
- Behavioral analytics to understand more about people's personalities, authentication and credibility.
- Insights on conversational dynamics, topics and flows.

## Drivers

- A common use case for speech analytics is found in the customer service department. These use cases focus on mining of recorded calls to surface insights into quality of customer interactions, customer satisfaction and sentiment, opportunities for employee training, and process improvements.
- Just as sentiment analysis is used in determining customer satisfaction in the front office, it could also be used to determine employee sentiment as part of a broader voice of the employee initiative. Analysis of conversations in the enterprise will greatly increase the ability to monitor compliance and risk, identify areas of improvements, and streamline automation of processes.
- Real-time speech analytics is emerging for use cases with sales and complex customer service interactions. Guidance applications are analyzing real-time conversations, determining intent and offering next-best action and coaching support to sales and customer service reps.

## Obstacles

- Speech analytics ownership and procurement is fragmented across multiple operations in organizations, with little coordination for selecting a primary vendor to meet multiple business use cases. This will hamper broader scale and adoption of speech analytics as they remain discrete projects.
- Speech-to-text capabilities are progressing rapidly, but understanding intent and surfacing useful insights from analysis require investment in people and time. A continual update of the speech library with relevant data (new products, competitors, campaigns) is essential to uncover useful insights across the organization. Without this maintenance, the insights will become less accurate and of less use to the organization.
- Monitoring and analysis of conversations in the enterprise can have a positive impact on operational performance. But, if not used positively, the perception of “big brother” listening to and analyzing what people say can lead to disengagement of associates and distrust of employers.

## User Recommendations

- Conduct pilot projects to evaluate whether speech analytics can provide demonstrably valuable insights into customer sentiment, employee quality and operational excellence. Align speech analytics with other customer service analytics initiatives such as text analytics, which is inherently more mature.
- Experiment with meeting transcription and analytics to tease out the value of surfacing insights from a series of internal meetings.
- Favor the SaaS deployment model with short-term licensing commitments or consumption-based billing to limit financial exposure as part of a trial. Ensure your data privacy and security requirements are maintained through trial into production.
- Create policies and an ethics board with employee representation for fair use of conversational data.
- Engage HR councils and workers' representatives to ensure the thin line of monitoring and analysis is not overstepped. Look for positive use-case examples of analytics already in use in the organization.

## Sample Vendors

Amazon; Balto; CallMiner; Gong; MeetRecord; NICE; Uniphore; ZoomInfo

## Gartner Recommended Reading

[Market Guide for Speech Analytics Platforms](#)

[Use-Case Prism: Artificial Intelligence for Customer Service](#)

[Getting Started With Technology for the Connected Rep](#)

[Boost CX by Adopting Real-Time Speech Analytics](#)

[Use Analytics to Improve Customer Service Experience and Productivity](#)

## Climbing the Slope

### Ontologies and Knowledge Graphs

Analysis By: Stephen Emmott, Anthony Mullen

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Early mainstream

#### **Definition:**

Ontologies and knowledge graphs enable users to model a set of entities plus their properties and relationships in a particular domain. They support the development of a consistent terminology and allow complex relationships to be represented. They are often used to abstract away from underlying schemas, and can be seen as a flexible knowledge network with broad applicability across many natural language technology use cases.

#### **Why This Is Important**

Ontologies and the knowledge graphs (KGs) they define are methods to represent data and knowledge that is human- and machine-readable, standardized, easily shareable, and abstracted from the underlying storage mechanism. They help align data models across applications serving multiple purposes, and enable data's location, retrieval and use, increasing the actual value of data. This includes content, which is often stored as unstructured data, rendering it challenging to process without metadata.

#### **Business Impact**

As investment and dependence on natural language technology (NLT) increases, especially in the context of generative AI, ontologies and knowledge graphs will become foundational for semantic modeling. Better integrations will enable applications to use and reuse data and provide consistent labeling for software development and data and analytics. This supports broad application to use cases, multiagent systems, and more performant AI systems using composite AI.

#### **Drivers**

- Language automation is driving usage. Conversational AI, insight engines, and intelligent document processing are just some of the submarkets driving the broader language automation market. While some vendors may weight structural roles (for example, ontologies and KGs) and computational roles (for example, deep neural networks) differently, we see aggregate growth of ontologies and KGs both by vendors and clients. Major drivers — like conversational AI, especially in the wake of ChatGPT — have given ontologies and KGs renewed prominence as an alternative to traditional, intent-based approaches.
- Generative AI, specifically prompt engineering and model training, rely on the identification, retrieval, and processing of data, which has elevated the importance and prominence of metadata.
- Augmenting and automating metadata management relieves the burden on humans working alone to manage ontologies and KGs through the use of machine learning to support their creation, maintenance and tuning. Large language models (LLMs) in particular can automate many aspects of the creation and management of ontologies and metadata, promising to scale metadata management to organizational demand such as D&A catalogs.
- Accelerated and broader use in data management. Ontologies and knowledge graphs offer a flexible knowledge network, enabling broader use in both metadata management and data fabrics.
- New approaches to composite AI in NLT. Many semantic platforms have pivoted to integrate symbolic approaches (such as ontologies and KGs) and subsymbolic approaches (such as deep neural networks) over recent years. This has improved NLT performance.
- Improvement in ModelOps and AI orchestration techniques for graphs. New techniques for distributed architecture and federated learning provide greater capacity to manage billions of triples, and allow for secure data collaboration.

## Obstacles

- Vendors using ontologies and/or knowledge graphs often don't make the information architecture directly available to buyers. Vendors should make them available as an asset, in a marketplace or exchange, rather than a hidden mechanic for the end users they serve.
- Most buyers have really just begun to develop their AI data pipelines, and therefore don't realize the benefits of ontologies and knowledge graphs.

- Although LLMs provide opportunities to apply metadata as well as manage it, the perception of ChatGPT as a panacea is obscuring the essential role that metadata plays in the use and creation of LLMs.
- Substantial effort is needed. As domains expand, the complexity of the graph increases, and without a good UI they become too unwieldy to manage. Developing ontologies and instantiating knowledge graphs from these requires distinct expertise that is challenging to find, whether through recruitment, training, or professional services.

## User Recommendations

- Establish ownership of metadata management — including ontologies and knowledge graphs — typically within the remit of data and analytics leaders, delegated to appropriate role/team, and integral to data management strategies.
- Review what already exists, both internally and externally, to see if any large-scale ontologies are available for the industry/domains relevant to applications and use cases. Also, examine ontology vendors and their wider offerings.
- Ensure easy access to data assets for software engineers and data scientists. Represent data and product catalogs and services as ontologies and graphs to enable richer analytics and support collaborations with partners.
- Counter employee churn and retiring experts by developing ontologies and knowledge graphs to model tacit and implicit knowledge of employees. Connect human-in-the-loop enrichment of semantic assets to knowledge workers across business units.

## Sample Vendors

Expert.ai; Franz; Ontotext; SciBite; semantha; Semantic Web Company; Smartlogic; Synaptica; TAIGER; Talkmap

## Gartner Recommended Reading

[AI Design Patterns for Knowledge Graphs and Generative AI](#)

[Adopt a Data Semantics Approach to Drive Business Value](#)

[Tool: Vendor Identification for Natural Language Technologies](#)

[Magic Quadrant for Insight Engines](#)

## Natural Language Understanding

Analysis By: Bern Elliot

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

### Definition:

Natural language understanding (NLU) is a subtopic of the natural language processing (NLP) field of artificial intelligence. NLU focuses primarily on interpreting or understanding the text, and it is typically based on matching the parsed input to an underlying knowledge model or structured ontology. Recent large language models (LLMs) have been used to augment and sometimes replace the existing NLU methods.

### Why This Is Important

NLU encompasses a broad range of applications, from simple understanding tasks like interpreting short direct commands, to complex understanding tasks like comprehending newspaper articles or maintaining a humanlike conversation. NLU functionality, even the basic ones, offers enterprises useful ways of directing the actions of a business application, appliance or device based on the natural language inputs received from users, and the intent they entail.

### Business Impact

There is significant and widespread demand for systems that can understand or interpret natural language, and that can interact with people in a conversational style. Applications that leverage NLU include chatbots, virtual assistants, text summarization and text content analysis. Other uses include smart vehicles, machinery, and consumer “intelligent” devices and appliances. In many cases, the effectiveness of NLU will determine the overall satisfaction with the application or appliance.

### Drivers

The fundamental driver for NLU improvements is a more accurate identification of a user’s intent:



- People have multiple ways of expressing the same thing. Conversely, people may use the same words to express different meanings. The NLU functionality will examine the parsed elements of the text and allow an application to interpret what was meant or intended based on its underlying model.
- LLMs have played a role in enhancing the ability to identify user intents and are often used to complement other NLU approaches. It is likely that, as LLM models and usage matures, they will increasingly become the primary method for interpreting user intents.

NLUs, that increasingly leverage LLM functionality, play a significant role in the key functions and applications below:

- **Virtual assistants:** NLU enables the chatbot to identify the intent of the user's input, and, as needed, to extract key entities from that.
- **Text categorization and classification:** NLU enables systems to analyze and assign text input into predefined categories. Examples include spam filters, document classification and script compliance.
- **Automatic text summarization:** NLU can play a role in creating summaries of longer text sections.
- **Question answering (QA) and semantic parsing:** QA and search systems leverage several of the above outlined functions to create a natural language interaction.
- **Content moderation:** NLU can assist in detecting offensive/inappropriate/harmful content and moderate interactions are increasingly needed.
- **Sentiment analysis:** NLU helps to identify and measure the sentiment behind an opinion or context.

## Obstacles

- **Complexity:** While some NLU objectives are simple, many are complex, which becomes an obstacle to successful usage of NLU. Factors driving complexity include the use of large-scale vocabularies, grammar, ontologies and models. This is one of the reasons why LLMs are increasingly seen as a useful option to traditional NLU.
- **Customization:** Many NLU implementations require customization, including the use of custom training datasets. These datasets may take time to develop.
- **Evolving technology:** The optimal techniques for implementing NLU continue to evolve. While methods, such as tree graph analysis are well-established and newer methods based on LLMs are emerging. The tooling and level of complexity required by each application significantly vary for each use case.
- **Bundling:** NLUs are often bundled within a chatbot platform and are managed via an integrated developer environment. As a result, they are often not separately reviewed.

## User Recommendations

- **Examine the architecture of the NLU module.** Each NLU solution is intimately tied to its NLP environment, as well as to the application it supports, the model(s) it supports and the data that is in the model. It is critical that enterprise technical teams understand how these interrelated areas come together within the NLU module.
- **Determine the types of NLU modules that best fit your models.** In some cases, it may be useful to have NLU modules that can be interchanged, support multiple underlying models or offer preexisting data for the models.
- **Select based on language and use case.** Evaluate the NLU options based on the specific types of language and types of use cases that enterprise technical teams will need to support. Favor composite-AI approaches that optimize accuracy, transparency and costs, which is relevant when third-party LLMs are becoming part of the mix.
- **Anticipate that LLMs will displace or radically change existing NLU methods and pipelines.**

## Sample Vendors

Amazon Web Services; IBM; NLTK; Rasa

## Gartner Recommended Reading

[Magic Quadrant for Enterprise Conversational AI Platforms](#)

[Critical Capabilities for Enterprise Conversational AI Platforms](#)

[Applying AI — A Framework for the Enterprise](#)

[ChatGPT Research Highlights](#)

## Neural Machine Translation

**Analysis By:** Gabriele Rigon, Bern Elliot

**Benefit Rating:** Transformational

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Early mainstream

### Definition:

Neural machine translation (NMT) is the application of deep neural network models to automate the translation of text and speech from one language to another. Use cases that exploit NMT include synchronous and asynchronous translation of documents (for example, reports, blog posts and chatbot answers) and multimedia content, including for real-time dubbing and subtitling.

### Why This Is Important

NMT enables higher quality translations than previous machine translation approaches and it can enhance market access for content and services across linguistic barriers. It can support use cases for content translation that was previously too expensive or complex, or not readily available. NMT is also being used to augment current human translation methods, allowing the translator to focus on quality via postediting machine-translated text, or on more difficult translation tasks.

## Business Impact

Due to the lower costs of NMT compared to human translation, the most significant impact of this technology has been to enable new applications by creating access to a greater audience for digital content. This includes the benefits NMT entails for global labor arbitrage and for the ability to work with linguistically diverse staff. Other use cases include real-time synchronous scenarios, where speed is critical, and the augmentation of human translators for improved efficiency.

## Drivers

- Core drivers for NMT are productivity and operational efficiency, associated with significantly reduced costs over human translation, and the speed of translation.
- There are significant quality improvements that NMT offers over previous techniques, especially when vendors provide options to customize and adapt models. This enables a broad set of use cases across many business areas, such as text translation for documents, e-commerce and chat forums. It also offers multimedia translation as in the case, for example, of training seminars and entertainment.
- On top of preexisting NMT approaches, large language models (LLMs) such as GPT-4 or BLOOMZ work surprisingly well and sometimes outperform traditional NMT models in the case of bilingual translations. Their effectiveness in handling multilingual machine translation, which involves a massive number of languages at the same time, is being tested.
- As NMT matures, secondary effects of higher quality and near-real-time outputs are also driving new markets. NMT is being integrated with a variety of applications and platforms. These include website content management systems, IT service management (ITSM), contact center and CRM platforms, and conversational AI platforms – aiming to support chatbot localization.
- NMT is being integrated with multiple adjacent translation technologies, such as computer assisted translation (CAT) tools in a process called postedit of machine translation (PEMT). Quality evaluation solutions are being developed to measure the accuracy of NMT outputs. NMT-selector hub services are offered to route translation requests to the best fit NMT provider. Translation management systems (TMS) and the language services provider (LSP) market are shifting to adapt to NMT innovation. Website localization methods are using NMT for novel, proxy-based dynamic approaches. There are markets for data for customization of NMT models according to verticals and dialects.

## Obstacles

- NMT still has quality limitations and so requires careful fitting of use cases — especially in emerging or specialist domains, such as legal or medical fields.
- NMT without postediting will not fit the requirements for use cases where high accuracy is a critical requirement, such as medical and legal translation.
- Availability and quality of NMT models varies significantly between language pairs and among industries, and there's still a deficit for low resource languages.
- The processes and best practices associated with effectively using NMT are still nascent in products and enterprise preparedness.
- Some enterprises worry about privacy and confidentiality of data sent to cloud-based NMT models. This applies to NMT solutions that are cloud hosted, and not those used in the context of regulated on-premises deployments for industries such as government and healthcare.

## User Recommendations

- Review where enterprise processes could be improved with NMT. This may involve modifying existing processes that require a high degree of human intervention.
- Assess where and how translation service providers are using NMT to improve their current service offerings.
- Collect translated text and supporting materials such as glossaries, which will be useful for customizing language models. This includes determining the format of the data so that it is usable for NMT model training.
- Evaluate NMT solutions based on your target language pairs and text samples related to your actual documents and use cases. Not all NMTs are equally effective in all languages and industries.
- Evaluate the benefits of using different NMT vendors for different language pairs and industries in order to deliver the optimal NMT strategy.

## Sample Vendors

AppTek; DeepL; Google; Pangeanic; RWS; Smartling; SYSTRAN; Translated; TransPerfect; XTM International

## Gartner Recommended Reading

[Market Guide for AI-Enabled Translation Services](#)

[Toolkit for AI-Enabled Translation Services — Vendor Profiles for Market Guide](#)

[Cool Vendors in Translation Technology](#)

[Maverick\\* Research: Machine Translation Can Save the World's Languages From Extinction](#)

[Best Practices for Localizing Your Chatbot Initiative](#)

[Tool: Vendor Identification for Natural Language Technologies](#)

## Semantic Search

Analysis By: Stephen Emmott

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

### Definition:

Semantic search processes natural language to capture the relationships between words in addition to words themselves. By processing “things, not strings,” entities such as people, organizations and products are revealed along with their attributes and relationships. This serves to better mediate between intent and outcome within a given context, improving relevance and accuracy.

### Why This Is Important

Semantic search amplifies performance by indexing words and their relationships to one another. This enables the processing of what is represented by content and data, as well as the context and intent of natural language queries, thereby improving the accuracy and relevance of experiences compared to keyword-based search. Such an approach impacts all types of search, especially domain-specific, and it can represent a critical component of generative AI pipelines in a variety of design patterns.

## Business Impact

Semantic search can:

- Elevate employee experience by connecting employees to one another, or to information, based on their expertise, activities or need for knowledge in the digital workplace.
- Enhance customer experience by supporting product selection and purchase, or self-help before or after sale in digital commerce.
- Decompose natural language into data for analysis and automated processing.
- Build context for prompt engineering, elevating relevance and accuracy, which facilitates the adoption of generative AI.

## Drivers

The adoption of semantic search is driven by:

- The continued need for more relevant experiences in the context of digital commerce, enterprise applications, analytics and automation. For instance, semantic search is considered a necessary prerequisite for conversational search applications such as conversational commerce.
- The capabilities and expectations of generative AI models, and the application of these models, as demonstrated by ChatGPT. Semantic search can be a key component of mature and complex design patterns that leverage large language models through retrieval augmented generation. Semantic search enables meaning to be processed reliably in the workplace.
- The capability to extract data from content including audiovisual sources, thereby extending the reach of search to a wide range of data sources and formats, enabling multistructured analytics. Semantic search also enables natural language query in analytics and business intelligence platforms.
- The need to support discovery in the context of multiagent systems.

- Advances in vector-based embedding to provide an effective, math-based alternative to rule-based approaches to processing the semantics of natural language. Such approaches are delivering capability in a way that is easily attainable and scalable for enterprises. This is also leading to advances in knowledge graphs, enabling explicit representation of the entities and concepts that words represent, their attributes and relationships.
- The emergence of composite AI and the ability to marshal multiple AI approaches such as rules and machine learning to optimize capabilities, for example, analyzing semantics across languages and harmonizing search results for multilingual datasets.
- The availability of data exchanges and marketplaces for ontologies and other semantic assets. In addition, the entry barrier is low due to availability of open-source platforms and toolkits such as txtai.

## Obstacles

- Vendors are working to accommodate foundational models, and their use through prompt engineering, into their products.
- While foundational models are flexible, rule-based approaches require further development — meaning those products that utilize a composite approach tend to be limited to specific languages.
- There is a requirement for continuous commitment to the development and optimization of models, rules and data structures. Specific use cases mean fewer customers, resulting in models and rules that are at lower levels of maturity. Use is typically related to private, confidential or commercial sources, which results in learning and adaptation that cannot be shared beyond individual customers.
- Semantic models need to be developed, whether from scratch or derived from existing models available externally. Professional services are often required to get started and continue, especially as insufficient data governance and management lead to poor-quality data.



## User Recommendations

- Use insight engines or digital commerce search as the platform for building applications that utilize semantic search. Where applicable, consider vector databases as alternative or complement. Start by reviewing your existing insight engine's capabilities to ensure they support semantic search, and factor it in as a requirement when selecting new products.
- Use semantic search as an integral part of your generative AI initiatives, to facilitate grounding in the context of prompt engineering.
- Link your search and insight activities with your other natural language technology initiatives and in the context of your organization's wider data fabric. Seek rationalization and consolidation with natural language query solutions for structured data where appropriate.
- Test the performance of semantic search in the business context. Engage subject matter experts to contribute — ideally independently and proactively — to the maintenance of semantic search capabilities.

## Sample Vendors

Access Innovations; Cohere; expert.ai; Glean Technologies; IBM; Mindbreeze; Ontotext; Semantic Web Company; Yext

## Gartner Recommended Reading

[Magic Quadrant for Insight Engines](#)

[Critical Capabilities for Insight Engines](#)

[Quick Answer: What Impact Will Generative AI Have on Search?](#)

[AI Design Patterns for Large Language Models](#)

[How Large Language Models and Knowledge Graphs Can Transform Enterprise Search](#)

## Enterprise Conversational AI Platforms

Analysis By: Gabriele Rigon, Bern Elliot

Benefit Rating: Transformational

**Market Penetration:** More than 50% of target audience

**Maturity:** Mature mainstream

**Definition:**

Enterprise conversational AI platforms (ECAIPs) are a strategic product for enterprises that need multiple, scalable and sophisticated chatbots and virtual assistants. These platforms provide a variety of capabilities with a no-code toolset for business users to build and maintain chatbots and virtual assistants. There is an incredibly wide array of use cases, including coverage for customer service, contact center automation, ITSM and HR scopes.

**Why This Is Important**

ECAIPs comprise a collection of technologies that enable centralized implementation and operations across multiple chatbot initiatives. They provide enterprise-grade, no-code tooling to support both Q&A and transactional use cases, and to create custom conversational interfaces to interact with business applications in general. Such platforms include capabilities to improve business outcomes when building, testing, deploying and optimizing AI-enabled conversational interfaces.

**Business Impact**

Automating and augmenting human-based interactions in the enterprise has tremendous business impact. Among the benefits brought by chatbots, AI-enabled conversational interfaces and specific capabilities offered by ECAIPs are human capacity augmentation, customer and employee-experience enhancement, cost savings, insightful analytics based on conversational data. ECAIPs allow enterprises to centralize assets and capabilities and to democratize access to design, development and monitoring of CAI implementations.

## Drivers

- **Generative AI:** The hype around generative AI and ChatGPT is generating an increasing amount of interest in conversational AI capabilities in the market. Conversational AI platforms offer a center of gravity that an organization can build a chatbot strategy around.
- **Diversity and abundance of offerings:** The number of vendors is still staggeringly large, at more than 2,000 worldwide. However, Gartner estimates that only about 150 of these vendors can be considered to be possible strategic choices for the major use cases.
- **Voice capabilities:** The consolidation of sophisticated voice capabilities, including conversation analytics and voice biometrics capabilities, have made large-scale contact center automation viable, with huge potential for savings and great customer experience.
- **Broader scope of use cases and capabilities:** There is still value in specialist offerings — vendors that specialize in industry, domain or even tasks. However, over the long term, the market will be dominated by larger horizontal platforms that offer industry, domain and task components on top of a versatile platform core, to address the needs of broader use cases such as contact center automation and employee-facing use cases that span CAI but also intelligent document processing and robotic process automation.

## Obstacles

- **Immature buyers:** Many buyers remain immature or have little experience and rely on outdated advice and practices, resulting in simplistic and tactical applications that generate poor adoption.
- **Diversity of approaches:** End users are still exposed to a variety of products, deployment methodologies, pricing structures, ambiguous scopes of work, and the inability to quantify the value of platforms and implementations.
- **Scaling:** The major challenge is scaling the implementations from proofs of concept or initial deployments to strategic and operationalized deployments that bring tangible business value.
- **Generative AI:** The increasing availability of embedded generative and conversational AI capabilities and user interfaces (UIs) in enterprise and consumer products, such as insight engines or productivity applications, will also erode the utility of traditional chatbots in nontransactional, Q&A-based interactions.

## User Recommendations

- Avoid technical debt by planning and investing in options for integrating with LLMs and generative AI capabilities, choosing platforms built on a modular architecture that allows for flexibility in choosing embedded LLM or external generative AI services.
- Enable your initiative by choosing enterprise-ready platforms that support your core requirements around use-case scope, language support, as well as no-code options that allow nondevelopers and non-data-scientists to do build, maintenance and evolution tasks.
- Mitigate risks of low adoption and noncompliance by choosing platforms that provide capabilities and accelerators aligned to your domain (use-case) needs coupled with continuous optimization tooling. Focus on monitoring and enhancing the overall UX and conversational aspects as well as security and privacy-protection skills.
- Ensure the scalability of your investment by strategically choosing platforms that support multiple use cases, multiple bots and multiple roles within the enterprise, and that have strong orchestration and synergies between the required integrations.

## Sample Vendors

Amelia; Avaamo; Cognigy; Google; IBM; Kore.ai; Omilia; OneReach.ai

## Gartner Recommended Reading

[Magic Quadrant for Enterprise Conversational AI Platforms](#)

[Critical Capabilities for Enterprise Conversational AI Platforms](#)

[Choosing the Right Conversational AI Platform](#)

[Quick Answer: How to Prioritize Requirements in the RFP for Conversational AI Tool: RFP for Enterprise Conversational AI](#)

## Insight Engines

Analysis By: Stephen Emmott

Benefit Rating: High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Early mainstream

**Definition:**

Insight engines apply relevancy methods to discover, analyze, describe and organize content and data. They enable the interactive or proactive delivery or synthesis of information to people, and data to machines, in the context of their respective business moments.

**Why This Is Important**

The growth of content and data — both internal and external to organizations — makes it challenging for employees and customers to gain the insight they need for better decisions and action. Insight engines solve this by combining search with composite AI. Applications built on insight engines' capabilities (which includes semantic search and "grounding" generative AI) can retrieve and synthesize the information required for insight; and the data required for analysis and automation.

**Business Impact**

Insight engines serve employees in the context of enterprise applications, analytics and automation. Insight engines are platforms that marshal natural language technologies for a range of prebuilt and custom-made applications. They apply across all domains (business functions and industry verticals), but especially localized use cases, such as customer support. Experiences are delivered in the form of separate touchpoints, embedded in third-party applications, or as data feeds.

## Drivers

- The bundling of search within the digital workplace application suites provided by Microsoft 365 and Google Workspace provides a foundational capability by default.
- The increase in the quantity and variety of data (including content and procedural knowledge) within organizations has given rise to “wide data” (including videos, podcasts, JSON, PDFs, tables and records).
- The need for custom-made applications serving localized use cases is providing deeper context and amplifying relevance.
- The need to provide better user experiences with less friction, by embedding information in the applications where it is needed – bringing orientation, decisions and actions into one place or tool.
- Developments in semantic search, which drive the capability of insight engines, is enabling wide data to be synthesized as information throughout the digital workplace.
- Programmatic access to content (and other data sources) for various applications to support digital transformation (for example, automation).
- The need for a natural language technology platform upon which multiple insight applications can be developed and delivered cohesively and according to aspiration.
- SaaS deployment models and packaged capabilities that are speeding up the time to value for insight engines in certain market segments.

## Obstacles

- Although the impact of insight engines is high, the benefits can be difficult to articulate and quantify as impact is often misunderstood and undervalued. Relevance varies across vendors, and client deployments, making it challenging to measure and optimize.
- Custom development is required, expanding the time, skills and cost needed to attain higher aspirations for the technology. The range of connectors to data sources is constrained and varies across vendors as does the range of prebuilt integrations into third-party applications, which is often limited.
- The complexity of insight engines means they can be challenging for proof-of-concept initiatives, and usually entail a long-term commitment. Shifting to semantic search entails a commitment to natural language technologies, necessitating the continuous development of language models and rules. Generative AI techniques such as large language models (LLMs) are lowering the entry barrier to enable a variety of use cases supported by insight engines.

## User Recommendations

- Establish whether the search functions included in digital workplace application suites are sufficient and shape your portfolio of search and insight applications to ensure you can address use cases across enterprise application, analytics and automation use cases.
- Place employees at the center to focus the purpose of insight engines by using personas to identify the use cases and applications used to conduct work, and the sources of data from which the information they need is retrieved or synthesized.
- Ensure that insight engines are viewed as integral to your wider natural language initiatives, and consider consolidating these technologies as an insight engine. Align expectations. At the highest level of maturity, insight engines synthesize information proactively in the context of work, but more typically, they offer interactive retrieval.
- Choose an implementation partner that can accelerate your deployment and reap value from the investment.

## Sample Vendors

Elastic; Expert.ai; IBM; Microsoft; Mindbreeze; OpenText; Sinequa; Squirro

## Gartner Recommended Reading

[Magic Quadrant for Insight Engines](#)

[Critical Capabilities for Insight Engines](#)

[Buyers' Selection Spotlight: Insight Engines](#)

[Quick Answer: What Impact Will Generative AI Have on Search?](#)

[Use Multistructured Analytics for Complex Business Decisions](#)

## Open-Source NLP Toolkits

Analysis By: Adrian Lee

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

### Definition:

Open-source natural language processing (NLP) toolkits enable end users and commercial companies to review, modify or design the source code and models for their own purposes, mostly free of charge. Open-source NLP toolkits address the problems of analyzing and processing large, unstructured natural language corpora by means of structured libraries to facilitate the support for applications that require NLP tools and tasks, such as text analytics or virtual assistants.

### Why This Is Important

Open-source NLP toolkits are widely adopted and expected to grow in usage by strong market interest in generative AI outputs. Current releases enable more advanced NLP tasks that incorporate large language models to provide value by optimizing business processes and operations. Applicable use cases are in business intelligence (text analytics), customer service (improving customer satisfaction and increasing engagement) and employee support (productivity applications and knowledge base management).



## Business Impact

Open-source NLP toolkits enable:

- Capabilities for summarization, translation, Q&A or search in use cases for text and sentiment analysis.
- Rapid prototyping of conversational agents with virtual assistants (VAs) as front ends for leveraging NLP.
- Combination with NLP toolkits and solutions as a test environment to advance NLP accuracy and outputs.
- Enterprises to gain control over the application to develop customized solutions.
- Conversational AI platform and other vendors with NLP elements to develop their solutions without costly ML investments.

## Drivers

- Continued growth in the adoption of NLP capabilities and heightened interest in generative AI outputs is lowering the barrier to entry for enterprises building solutions with open-source NLP toolkits to integrate into their digital products.
- Enterprises need to customize their integrations of NLP with their applications and domain-specific use cases.
- There is a lack of availability of customization of custom-made and packaged NLP integrated development environment (IDE) solutions without high professional services fees.
- NLP toolkits are a mature NLP foundation with broad user communities that are widely accepted by enterprises and conversational AI providers to enable the most common modes of linguistic analysis to help machines understand the text.
- Core features of open-source NLP toolkits position them to be suitable, lower-cost platforms for prototyping and piloting conversational agents or natural language applications and appeals to new or smaller organizations.
- A significant proportion of open-source NLP toolkits come with some pretrained intent models, multiple languages and test suites to accelerate the project delivery for end users.

## Obstacles

- NLP toolkits are often focused toward experimental research and projects combining commercial solutions. Prototypes may not make it into production or be scaled up for commercial use.
- Enterprises that build their own products leveraging NLP toolkits may find it challenging to keep pace with the rapid evolution of NLP as it requires custom data and in-house AI and ML skills that require mid-to-long-term funding.
- IT leaders typically consider managed service providers that fulfill domain-specific needs and prebuilt intent models to accelerate deployment.
- Limitations can exist where end users do not augment toolkits with suitable ML algorithms and language model optimization to improve the performance of the NLP output.
- Using NLP toolkits is strictly iterative and requires customization and integrations with enterprise applications to deliver business benefits.
- Differences exist between conversational AI providers and open-source platforms in ontologies, taxonomies and domain specificity to drive contextual natural language understanding.

## User Recommendations

- Implement open-source NLP toolkits to build an internal NLT stack by starting with a suitable toolkit to match existing IT infrastructure or to incorporate into a sellable digital product.
- Possess skilled in-house resources of data scientists and AI technology engineers for NLP accuracy and model fine-tuning.
- Evaluate a long-term and more costly initial IT investment by using open-source NLP toolkits in order to benefit subsequently from lower total costs of ownership.

## Sample Vendors

Amazon; Baidu; Google; Hugging Face; NLTK Project; OpenAI; PyTorch; Rasa Technologies; spaCy

## Gartner Recommended Reading

[Cool Vendors in Conversational and Natural Language Technology](#)

Selecting Conversational AI Solutions for Chatbot and Virtual Assistant Initiatives

Emerging Tech Roundup: ChatGPT Hype Fuels Urgency for Advancing Conversational AI and Generative AI

## Entering the Plateau

### Text Analytics

Analysis By: Shubhangi Vashisth, Stephen Emmott

Benefit Rating: Moderate

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

#### Definition:

Text analytics is the process of deriving business insight from structured and unstructured textual data. This process can include determining and classifying the subjects of texts, summarizing texts, extracting key entities from texts, and identifying the tone or sentiment of texts.

#### Why This Is Important

Text analytics products extract meaning and context from vast volumes of textual content that cannot be processed by humans. These extracts can then be used to derive insights and action, either within the context of the product or by other products with access to the data. Large language models (LLMs) and related applications (e.g., ChatGPT) are fueling further interest in this area, with capabilities like written content augmentation and creation, classification, summarization, and natural language question answering.

#### Business Impact

Text analytics, when combined with other analytics capabilities, such as intelligent document processing (IDP) and search, can benefit the organization in the following areas:

- Preprocessing, identifying and analyzing unstructured data for compliance, risk and fraud detection
- Automated document matching and classification
- Discovery and insight from vast amounts of data that can't be analyzed manually
- Metadata and information architecture development
- Sentiment, voice of the customer (VoC)/voice of the employee (VoE) and emotion detection

## Drivers

Key drivers include:

- A surge in the volume of textual data, especially from sources other than traditional “documents” (such as instant messages, emails and automatically extracted metadata)
- Interest in LLMs, fueled further by ChatGPT, which has wide applicability across a diverse set of use cases and empowers all kinds of end users, both technical and nontechnical, to derive insights from existing data
- The desire to complement insights gleaned from analysis of structured numerical data with text-based facts and hidden intents and insights for more robust predictive modeling
- Advancements in nonsymbolic (machine-learning-based) techniques, which enable products to operate against a wider range of languages

Text analytics uses different combinations of technologies to drive solutions to different business use cases:

- **Healthcare:** Analyzing medical records by mapping key medical terms into a graph
- **Insurance:** Identifying fraudulent claims by analyzing the narratives and identifying common individuals across claims
- **Finance:** Gaining insights into investments by monitoring public information sources and social media
- **Legal:** Supporting contract review by extracting key terms and obligations from complex contracts
- **Retail:** Monitoring product pricing across markets
- **Marketing:** Monitoring brand loyalty and sentiment by analyzing social media feeds and customer feedback
- **Law enforcement:** Performing forensic analysis of a body of documents by identifying key subjects and dates and developing a chain of events
- **Digital publishing:** Identifying related articles and developing a summary relevant to an article in progress

## Obstacles

Several factors hinder the emergence of more pervasive, easy-to-use business solutions for text analytics:

- The differentiation between the many overlapping vendors is too nuanced for those organizations without in-house expertise. The capabilities also overlap with adjacent technology areas, such as insight engines.
- Embedded text analytics is a challenge for stand-alone text analytics. Stand-alone text analytics offers generic capabilities that may address basic enterprise requirements at a broad level, but that are not deep enough for specialized or complex use cases.
- Although text analytics products are easier to use, it is still challenging to incorporate solutions into an organization's wider digital platform, given the diversity of use cases and specialist skills needed to utilize and gain benefit.
- Most organizations lack a strategy to deal with unstructured data. The approach to select tools for point solutions adds to the problem of tool sprawl.
- Training the solutions for specialized use cases is also a barrier to adoption.

## User Recommendations

- **Position text analytics as a natural language technology (NLT)** in the context of internal discussions to correctly frame its role in augmentation and automation.
- **Identify and prioritize use cases** that text analytics can address. Create an enterprise text analytics strategy. Tie these initiatives with master metadata management projects for enterprisewide standardization.
- **Review the text analytics market** to acquaint yourself with its vendors, products and capabilities.
- **Start with prepackaged products designed for business users** to administer for well-established use cases, such as VoC. Cloud-based text analytics packages are another good way to experiment and enable easy adoption.
- **Select products based on how well they suit specific business scenarios** and their ability to integrate with other applications that work with unstructured data.
- **Allow a realistic lead time to recruit text analytics talent.** Consider working with a third-party analytics service provider for text analytics initiatives.

## Sample Vendors

Amazon Web Services; Amenity Analytics; Bitext Innovations; Google; IBM; InMoment (Lexalytics); Megaputer Intelligence; Microsoft; Qualtrics; SavantX

## Gartner Recommended Reading

[Artificial Intelligence Primer for 2023](#)

[Market Guide for Text Analytics](#)

[Tool: Vendor Identification for Natural Language Technologies](#)

[Innovation Insight for Generative AI](#)

[Best Practices for the Responsible Use of Natural Language Technologies](#)

## Speech Synthesis

Analysis By: Bern Elliot

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

### Definition:

Speech synthesis is the artificial production of human speech. Current methods synthesize voice using a variety of modeling approaches, including AI-based language and voice models. Enhancements allow models to mimic specific speech patterns of individuals by tuning the synthesis model based on recordings. Recent advances in generative AI are enabling additional capabilities, improving the ability to mimic specific voices with limited training.

### Why This Is Important

High-quality speech synthesis allows vastly improved user experiences at a fraction of the cost of a human-recorded alternative. The current deep-learning-based approach uses acoustic models to produce waveforms and generate speech that is highly accurate. This approach was only developed in 2016. The rapid advancement of this technique is enabling new and useful applications.

## Business Impact

Speech synthesis is most used in the following business areas and applications:

- High volumes of generated speech (e.g., real-time streaming multimedia content)
- Frequent updates to content or where audio-visual material is dynamically assembled (e.g., training material)
- Where imitation or mimicry of characteristics is desirable (e.g., reading books aloud)
- Hands-free environments with audio output
- To make avatars or animations more realistic



## Drivers

- **Reading and communication aids for the blind and visually impaired.**
- **Improved user experiences with speech-enabled virtual assistants, chatbots and personal assistants;** modernization of contact center interactive voice response systems; clearer and more engaging narration of news or information drawn from data, such as sports or business events.
- **Better-quality and less-expensive audio narratives,** including audio books, such as read-aloud children's stories with fun voices, with interesting possible applications to e-learning activities.
- **Match voice personas to audiences.** Target audiences based on both content and voice characteristics in website and other commercial audio information. This can be coupled with visual personas such as avatars.
- **Hands free.** Improved usability of hands-free voice controls and interaction with voice assistants.
- **Markup expressivity.** Voice controls and markup to more effectively draw listeners' attention to specific details.
- **Gaming immersion.** More engaging characters and dialogues offered in computer games.
- **Media and entertainment industry.** Fast production of video and audio contents, such as audiobooks and multilingual training material.
- **Conversationally enabling large language models (LLMs).** On top of recent approaches for speech synthesis enabled by language models themselves, the intense interest in LLMs, and applications enabling LLMs such as ChatGPT, and similar styles of interaction, has raised interest in enabling speech-based interactions, including in consumer segments.

## Obstacles

- **Ethical issues.** Recent models can mimic a specific individual's speech to the point that it is very difficult to identify it as a fake (aka deepfake). While there are legitimate uses for mimicry, speech synthesis also poses ethical issues where it may be used to manipulate or deceive listeners.
- **Risk of taking digital recordings out of context.** While this is a risk, the introduction of "digital watermarks" makes it possible to better identify when generated speech has been taken out of context and/or used without proper licensing.
- **The "uncanny valley" experience.** Users may find overly realistic virtual assistants "creepy."
- **Cost and time for custom-made voices.** Custom speech synthesis models take expertise and time to develop and tune.

## User Recommendations

- Business leaders should review their current applications that use speech and consider how this low-cost approach to speech generation, which combines low cost with high quality, can be best leveraged. However, planners should evaluate solutions before committing, because quality and latency issues may make some solutions less desirable than others.
- Leaders and those responsible for risk and security should determine where there may be exposures to voice mimicking. This can include situations where significant actions are taken solely based on verbal commands via phone. Personnel, clients and employees should be warned of the risk and enlisted in identifying situations at risk for this sort of fraud.
- Leaders should not lock-in their applications to vendors generating voice. To be able to choose between different providers of voice and to reduce technical debt of natural language technology solutions, ensure that either you or your vendor allow voice generation services to be easily interchanged.

## Sample Vendors

Acapela Group; Amazon; DeepZen; Descript; IBM; LOVO; Microsoft; ReadSpeaker (rSpeak); SoundHound; Verbio Technologies

## Gartner Recommended Reading

[Best Practices for the Responsible Use of Natural Language Technologies](#)

[Critical Capabilities for Enterprise Conversational AI Platforms](#)

[Tool: Vendor Identification for Natural Language Technologies](#)

## Appendixes

See the previous Hype Cycle: [Hype Cycle for Natural Language Technologies, 2022](#)

## Hype Cycle Phases, Benefit Ratings and Maturity Levels

**Table 2: Hype Cycle Phases**

(Enlarged table in Appendix)

<i>Phase</i> ↓	<i>Definition</i> ↓
<i>Innovation Trigger</i>	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
<i>Peak of Inflated Expectations</i>	During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the innovation is pushed to its limits. The only enterprises making money are conference organizers and content publishers.
<i>Trough of Disillusionment</i>	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
<i>Slope of Enlightenment</i>	Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the innovation's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.
<i>Plateau of Productivity</i>	The real-world benefits of the innovation are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.
<i>Years to Mainstream Adoption</i>	The time required for the innovation to reach the Plateau of Productivity.

Source: Gartner (July 2023)

Table 3: Benefit Ratings

Benefit Rating ↓	Definition ↓
Transformational	Enables new ways of doing business across industries that will result in major shifts in industry dynamics
High	Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise
Moderate	Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings

Source: Gartner (July 2023)

**Table 4: Maturity Levels**

(Enlarged table in Appendix)

<i>Maturity Levels</i> ↓	<i>Status</i> ↓	<i>Products/Vendors</i> ↓
<i>Embryonic</i>	In labs	None
<i>Emerging</i>	Commercialization by vendors Pilots and deployments by industry leaders	First generation High price Much customization
<i>Adolescent</i>	Maturing technology capabilities and process understanding Uptake beyond early adopters	Second generation Less customization
<i>Early mainstream</i>	Proven technology Vendors, technology and adoption rapidly evolving	Third generation More out-of-box methodologies
<i>Mature mainstream</i>	Robust technology Not much evolution in vendors or technology	Several dominant vendors
<i>Legacy</i>	Not appropriate for new developments Cost of migration constrains replacement	Maintenance revenue focus
<i>Obsolete</i>	Rarely used	Used/resale market only

Source: Gartner (July 2023)

**Document Revision History**[Hype Cycle for Natural Language Technologies, 2022 - 11 July 2022](#)[Hype Cycle for Natural Language Technologies, 2021 - 21 July 2021](#)[Hype Cycle for Natural Language Technologies, 2020 - 6 July 2020](#)**Recommended by the Authors**

Some documents may not be available as part of your current Gartner subscription.

[Understanding Gartner's Hype Cycles](#)[Tool: Create Your Own Hype Cycle With Gartner's Hype Cycle Builder](#)[Applying AI — A Framework for the Enterprise](#)[Magic Quadrant for Enterprise Conversational AI Platforms](#)[ChatGPT Research Highlights](#)[Magic Quadrant for Insight Engines](#)

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.

Table 1: Priority Matrix for Natural Language Technologies, 2023

Benefit	Years to Mainstream Adoption			
↓	Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Transformational	Enterprise Conversational AI Platforms Neural Machine Translation	Conversational User Interfaces Generative AI Natural Language Understanding	Emotion AI Foundation Models	NSLMs
High	Generative AI-Enabled Applications Insight Engines Intelligent Document Processing Natural Language Query Open-Source NLP Toolkits Semantic Search	Multistructured Analytics Natural Language Generation Ontologies and Knowledge Graphs Prompt Engineering	LangOps Vector Databases	
Moderate	Speech Synthesis Text Analytics	AI-Enabled Content Moderation Speech Analytics Text Summarization		
Low				



Source: Gartner (July 2023)

Table 2: Hype Cycle Phases

Phase ↓	Definition ↓
<i>Innovation Trigger</i>	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
<i>Peak of Inflated Expectations</i>	During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the innovation is pushed to its limits. The only enterprises making money are conference organizers and content publishers.
<i>Trough of Disillusionment</i>	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
<i>Slope of Enlightenment</i>	Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the innovation's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.
<i>Plateau of Productivity</i>	The real-world benefits of the innovation are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.
<i>Years to Mainstream Adoption</i>	The time required for the innovation to reach the Plateau of Productivity.

Phase ↓

Definition ↓

Source: Gartner (July 2023)

Table 3: Benefit Ratings

Benefit Rating ↓	Definition ↓
Transformational	Enables new ways of doing business across industries that will result in major shifts in industry dynamics
High	Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise
Moderate	Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings

Source: Gartner (July 2023)

Table 4: Maturity Levels

Maturity Levels ↓	Status ↓	Products/Vendors ↓
Embryonic	In labs	None
Emerging	Commercialization by vendors Pilots and deployments by industry leaders	First generation High price Much customization
Adolescent	Maturing technology capabilities and process understanding Uptake beyond early adopters	Second generation Less customization
Early mainstream	Proven technology Vendors, technology and adoption rapidly evolving	Third generation More out-of-box methodologies
Mature mainstream	Robust technology Not much evolution in vendors or technology	Several dominant vendors
Legacy	Not appropriate for new developments Cost of migration constrains replacement	Maintenance revenue focus
Obsolete	Rarely used	Used/resale market only

Source: Gartner (July 2023)