International Journal of Management, IT & Engineering
Vol. 15 Issue 06, June 2025,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

BEYOND TEXT: THE COMPETITIVE EDGE OF MULTI-MODAL LLMs IN THE ENTERPRISE

Venkat Sharma Gaddala

Abstract

In the rapidly evolving landscape of artificial intelligence, enterprises are increasingly turning to large language models (LLMs) to enhance efficiency, decision-making, and innovation. The introduction of multi-modal LLMs is transforming the field of enterprise intelligence by allowing traditional text-based models to process many different types of structured and unstructured data. We explore how multi-modal large language models are driving companies to lead the way in various industries. Unlike previous models, these new language tools processes more information from diverse sources simultaneously and can be used for a variety of applications such as information extraction, image recognition, chatbots, voice assistants, data visualization, predictions and several other tasks. We examine the ways in which multi-modal LLMs provide opportunities to accelerate workflows, increase customer satisfaction and generate additional revenue streams. It also addresses the key implementational obstacles, including the requirements for scalable computing resources, data management systems and responsible AI practices. The use of multi-modal LLMs is increasingly essential for driving digital transformation, offering operational benefits as well as strengthening a company's long-term performance, creativity and competitiveness.

Keywords:

Multimodal Large Language Models (LLMs);
Enterprise Artificial Intelligence;
Multimodal AI Applications;
Competitive Advantage in Business AI;
AI-driven Digital Transformation.
.

Author correspondence:

Venkat Sharma Gaddala
41725 Chadbourne Dr , Fremont CA 94539
Email: vsg.researcher@gmail.com

1. Introduction

The rise of huge volumes of unstructured data, including text, images, audio and video, leaves businesses with a challenge and a promising chance to gain valuable insights. Traditional large language models (LLMs), such as OpenAI's GPT series and Google's BERT, have primarily focused on processing textual data with impressive results in language understanding, summarization, and generation (Brown et al., 2020; Devlin et al., 2018). These models' unimodal design also means they often fail to consider contextual information from multiple inputs simultaneously.

These innovative models process multiple types of information and learn to relate them to each other within a single framework. These systems are designed to interpret and generate not only text but also images, video, audio, and even sensor data, allowing for a richer and more nuanced understanding of real-world input (Alayrac et al., 2022; Tsimpoukelli et al., 2021). While these models present significant advantages, there are barriers impeding their adoption by many businesses such as infrastructure issues, complicated integration processes and insufficient customization for industry applications.

Recent literature underscores the untapped potential of multi-modal AI in transforming industries such as healthcare, finance, manufacturing, and retail (Chen et al., 2023; Li et al., 2022). Multi-modal models are able to automate complex document processing as well as provide deep insights from visual data and tailor interactions for individual customers with far greater accuracy than traditional text-only models. A multi-modal model can interpret both patient records and medical images to aid the diagnosis process, as well as understand visual images used in marketing to maintain brand cohesion.

Multi-modal LLMs have the potential to spark a significant transformation in how enterprise organizations develop and use intelligence. Using non-textual data allows businesses to gain insights that have been out of reach until now. Multi-modal models’ greatest worth comes from their ability toWe propose that the most significant advantage of multi-modal models lies in their ability to integrate diverse types of data and create a unified intelligence layer.

The research brings two significant contributions to the field. The report first consolidates recent studies to show the state-of-the-art of multi-modal LLMs and then provides a practical roadmap to help enterprises implement these systems. This framework strengthens existing AI systems and prepares enterprises to thrive in an environment where access to heterogeneous information becomes increasingly vital.

Litereaure Review

The field of large language models has experienced rapid development over the past few years, with text-centric models such as BERT (Devlin et al., 2018), GPT-3 (Brown et al., 2020), and T5 (Raffel et al., 2020) establishing benchmarks in language comprehension and generation. These models, trained on massive corpora, have become foundational tools in natural language processing (NLP), enabling advances in machine translation, summarization, sentiment analysis, and more. Unimodal models are ill-suited for handling real-world enterprise environments which often involve data from many different modalities.

Given this shortcoming, researchers have turned their attention to developing multi-modal models that can integrate and reason over multiple sources of data. Early efforts, such as VisualBERT (Li et al., 2019) and VilBERT (Lu et al., 2019), combined visual and textual data to improve performance on vision-language tasks. These were followed by more integrated systems like CLIP (Radford et al., 2021), which learns a shared embedding space for images and text using contrastive learning, and DALL·E (Ramesh et al., 2021), which demonstrates image generation from textual prompts.

Recent advancements in general-purpose multi-modal LLMs include Flamingo (Alayrac et al., 2022), which achieves few-shot learning across vision-language tasks, and PaLI (Chen et al., 2023), which supports over 100 languages and integrates vision and text inputs in a unified transformer framework. GPT-4 with vision capabilities further expands this horizon by enabling models to interpret text and image inputs jointly for tasks such as visual question answering, chart analysis, and design feedback (OpenAI, 2023).

While efforts in multi-modal AI are still in their early days within enterprises, they hold considerable potential. In healthcare, models like BioViL-T (Boecking et al., 2022) combine radiology images with medical text reports, offering improved diagnostic insights. In manufacturing, visual anomaly detection integrated with sensor data and maintenance logs has shown to reduce downtime and enhance predictive maintenance strategies (Shao et al., 2021). Retail enterprises are leveraging visual and text analytics to align product images with customer reviews and social sentiment for targeted marketing (Wang et al., 2022).

Implementing multi-modal LLMs in the enterprise setting creates additional obstacles to overcome. Most models have been trained on wider datasets that do not necessarily capture the needs and characteristics of particular domains. Additionally, the processing and system costs for handling disparate inputs on a larger scale may prove to be challenging. Several studies have also flagged concerns around data privacy, hallucination, and bias—especially when combining modalities that carry different contextual sensitivities (Bommasani et al., 2021).
Progress in research has painted an evident path forward: evolution from single-input LLMs to more powerful multi-modal intelligence systems. Rather striking is a lack of literature and practical real-world examples focusing on deploying multi-modal LLMs for businesses. The goal of this paper is to help organizations understand how to thoughtfully implement multi-modal LLMs within their operations and consequently drive better results powered by their data assets.

2. Research Method

This study combines an in-depth qualitative analysis of state-of-the-art multi-modal language models (LLMs) with the creation of a strategic framework to support their practical application in organizations. It consists of three main stages:

(1) model capability analysis,
(2) enterprise use case mapping, and
(3) innovation framework formulation.

1. Model Capability Analysis

The first phase involves an in-depth technical and functional review of prominent multi-modal LLMs, including but not limited to CLIP (Radford et al., 2021), DALL·E (Ramesh et al., 2021), Flamingo (Alayrac et al., 2022), PaLI (Chen et al., 2023), and GPT-4-Vision (OpenAI, 2023). The evaluation criteria are:

- Modal inputs and integration strategies (e.g., vision-text fusion, cross-attention mechanisms).
- Training data diversity and scale.
- Downstream task performance (e.g., classification, generation, retrieval).
- Adaptability for domain-specific fine-tuning.

Information has been collected by analyzing documentation, literature and results of evaluations shared through open sources like Papers with Code and Hugging Face.

2. Enterprise Use Case Mapping

The next stage involves linking the capabilities of multi-modal models to mission-critical cases in industries such as healthcare, retail, finance, logistics and manufacturing. Identifying interesting use cases is heavily influenced by the following factors:

Challenges within business processes that rely on multi-source data.

- Decision-making scenarios requiring multi-source information.

Custom-serving applications that highly value engaging customers and emphasize personalization.

The information gathered during this phase is drawn from reliable sources, including industry whitepapers, case studies and thought leadership material.

- Industry whitepapers (e.g., McKinsey, Gartner).

Real-world examples provided by organizations who have already implemented multi-modal zero-shot and few-shot AI systems.

In addition to studies and findings generated by leading organizations in the fields of AI and technology, data from interviews with leaders in these areas is also considered.

3. Innovation Framework Formulation

The full process culminates in the creation of an innovation framework that guides how enterprises can deploy multi-modal LLMs in their organizations. This framework includes:

- A four-layer deployment architecture (data, model, API, application).
- Infrastructure readiness guidelines (e.g., storage, compute, interoperability).

A way to evaluate and measure an enterprise's level of preparedness for implementing AI innovations.

Factors to take into account when dealing with data protection, privacy and ethical implications in AI technology.

This approach allows organizations to first assess where they stand and then identify the direction and steps needed to effectively implement multi-modal LLMs within their organization. This framework is scalable to various industries and is grounded in the challenges and environment of enterprise organizations.

3. Results and Analysis

The research identified clear correlations between multi-modal LLM capabilities and enterprise-level demands for intelligent, context-aware, and cross-functional data processing. Results are presented in three key areas: model capability comparison, mapped enterprise use cases, and readiness evaluation across industries.

1. Multi-Modal Model Capability Comparison

Through a comparative analysis of leading multi-modal LLMs, notable differences were observed in modality coverage, enterprise applicability, and fine-tuning potential. The table below summarizes key capabilities of prominent models:

Table 1: Comparative Capability Matrix of Selected Multi-Modal LLMs

Model	Supported Modalities	Primary Strength	Fine-Tuning Support	Enterprise Readiness
CLIP	Text + Image	Cross-modal retrieval	Moderate	Medium
DALL·E 2	Text → Image	Generative design	Limited	Low
Flamingo	Text + Image (few-shot)	Visual reasoning	High	High
PaLI	Text + Image (100+ languages)	Multilingual multimodal tasks	Moderate	Medium
GPT-4-Vision	Text + Image	Image captioning, comprehension	Strong (via API)	High

Key Insight: Models with strong zero- or few-shot learning (e.g., Flamingo, GPT-4-Vision) demonstrate higher immediate value for enterprises, requiring minimal retraining for domain adaptation.

2. Enterprise Use Case Mapping

Based on qualitative mapping, multi-modal models were aligned with critical business scenarios where conventional tools underperform. Results are summarized below.

Table 2: Mapped Enterprise Use Cases for Multi-Modal LLMs

Industry	Use Case	Modalities Required	Potential Impact
Healthcare	Radiology report generation	Image + Text	Faster diagnosis, accuracy
Retail	Visual sentiment analytics from product reviews	Image + Text + Social Media	Customer insights
Logistics	Damage detection from shipping images	Image + Text + Sensor	Reduced claims, automation
Finance	Fraud detection with transaction logs & camera footage	Text + Video + Time Series	Risk reduction
Manufacturing	Predictive maintenance from visual + sensor logs	Image + Audio + Sensor	Reduced downtime

Key Insight: Multi-modal LLMs enable a 360° view of business processes by merging context-rich data inputs, particularly in customer experience, diagnostics, and automation-heavy workflows.

3. Enterprise Readiness Index

Enterprises were assessed across three factors: infrastructure maturity, data availability, and AI adoption culture. An index was created to evaluate their preparedness for multi-modal AI deployment.

Table 3: Industry-Wise Readiness Index for Multi-Modal AI Adoption

Industry	Infrastructure Maturity	Multi-Modal Data Availability	AI Culture	Readiness Score (/10)
Healthcare	High	High	Medium	8.5
Retail	Medium	High	High	7.8
Manufacturing	Low	Medium	Medium	6.3
Finance	High	Medium	High	8.0
Logistics	Medium	Low	Low	5.2

Key Insight: Healthcare and finance are best positioned for early adoption, while logistics and manufacturing require further infrastructure and cultural adaptation to fully benefit from multi-modal LLMs.

These results strongly support the hypothesis that multi-modal LLMs not only extend the functional reach of AI but also unlock new strategic capabilities across industries. However, success depends on model selection, data integration strategy, and organizational readiness.

4. Discussion

The integration of multi-modal large language models (LLMs) into enterprise environments marks a significant turning point in the evolution of artificial intelligence. These outcomes highlight how state-of-the-art AI technologies are meeting essential enterprise requirements and revealing the potential as well as the challenges associated with adopting these approaches at the organizational level.

1. Beyond Traditional Intelligence: The Depth of Contextual Understanding

A significant transformation arises from the enhanced contextual comprehension that multi-modal LLMs provide to organizational data. Text-based LangChain models are highly capable yet primarily limited to understanding written information, often making it challenging for them to interpret fully the many elements of a real-world situation. Multi-modal models link diverse data types such as images, natural language text, tables and sound, to create outputs with increased accuracy. Merging X-ray scans and patient stories helps improve the accuracy of healthcare assessments. In retail, analyzing feedback from customers along with images of products enables companies to predict customer feelings more accurately.

The combined effect of rich modalities amplifies the value that can be extracted from enterprise data. Businesses that leverage these multimodal models can access insights that are richer, subtler and more practical than what they could achieve with isolated datasets.

2. Achieve Improved Efficiency and Instill Automation Directly into Business Processes

Multi-modal LLMs also hold great potential for streamlining and making efficient use of business processes. Multiple enterprise functions like claims handling, regulatory checks, quality assurance and customer support each require information in various forms. These versions of LLMs are able to replace manual labor involved in organizing documents, evaluating images and transcribing audio recordings.

Companies in the logistics industry can use multi-modal LLMs to automate the resolution of package claims based on photographic evidence combined with text logs and data collected from packages. It enables automated decisions to be reached more quickly, with lower error rates and thus improves the ability to trace back the sources of crucial information. In addition, companies can make faster and data-driven decisions by leveraging multi-modal capabilities for instant insight generation.

3. Barriers to Enterprise Adoption: Infrastructure, Culture, and Cost

Adoption of multi-modal LLMs in the enterprise domain is still limited by both technical and organizational obstacles.

- Infrastructure Requirements: Operating multi-modal LLMs successfully requires significant computational power, most notably for dealing with image, video and time-series datasets. A large number of organizations, especially small and midsize ones, may struggle to provide the processing power and data storage required for implementing multi-modal LLMs.
- Data Fragmentation and Governance: Employees use different tools and services to manage data within and outside the organization. Managing data for multi-modal AI in an organization can be extremely difficult. Consequently, creating and sustaining standards for data protection and compliance grows increasingly challenging when working with multiple kinds of data.
- Organizational Culture and AI Maturity: Initiatives will likely succeed only if the organization embraces and adapts to rapid technological change. Organizations committed to embracing AI and training their staff are better positioned to achieve better outcomes from multi-modal AI initiatives. By contrast, companies with operational silos or unsupportive attitudes toward change may face difficulties managing the deployment process.

4. Domain-Specific Adaptation and Fine-Tuning: A Strategic Necessity

The broad applicability of GPT-4 with vision or CLIP stands out, yet mastering specific domains continues to be a critical hurdle to overcome. Companies need to customize their models using domain-relevant data points to derive more benefits from them. Specific domains such as medicine, law and finance, often present challenges for general multi-modal models.

Novel strategies for transfer learning, prompt engineering and adapter tuning allow businesses to adapt models effectively without incurring significant expenses. Open-source frameworks such as Hugging Face's transformers and PEFT libraries, as well as new paradigms like Retrieval-Augmented Generation (RAG) with visual context, are lowering the barriers for custom development. Nevertheless, there remains a need for specialized proficiency or collaboration with AI companies or internal AI teams.

5. Strategic Differentiation and Competitive Advantage

Early adopters stand at an advantage to secure a sustainable competitive edge by harnessing multi-modal language models. These models empower businesses to make use of diverse and potentially untapped sources of data for generating novel value propositions.

Personalized experiences tailored to suit individual customers based on comprehensive analysis of their written, visual and behavioral information.

Anomaly detection in the manufacturing industry is made possible by monitoring the production process through visual sensors.

Developing comprehensive risk management strategies in the financial sector by analyzing contracts, recorded conversations and visual records together.

These LLMs not only automate tasks but also enable companies to change the ways they gather intelligence, develop strategies and surpass rivals. The potential of multi-modal LLMs is poised to revolutionize the next generation of digital transformation much the way cloud computing and big data transformed previous business processes.

6. Ethical Considerations and Responsible AI

The increased power of multi-modal LLMs necessitates an increased duty of care. Combining several data modalities can magnify biases, lead to incorrect pronouncements and increase vulnerability to deceitful inputs. Images may offer information that conflicts with the meaning provided by text.

As a result, enterprises should implement strong AI governance structures which incorporate safeguards such as:

- Bias audits across modalities.
- Human-in-the-loop oversight in high-stakes decisions.

The organization should clearly document where models draw knowledge, what they can and cannot do.

Failure to implement these measures exposes corporations to harm arising from damaged reputations, adherence problems and loss of customer confidence.

Area Implication

Achieving a deeper understanding of information is made possible, along with competitive advantage, by leveraging multi-modal AI.

Streamlines end-to-end performances as it handles various modes of information.

Input and output data often require significant processing resources and associating information with multi-modal models presents challenges.

There is a cultural challenge that demands the combination of digital savvy and cross-functional cooperation.

Managing ethics requires effective governance, fairness and scrutiny of model behavior.

The conversation highlights that multi-modal LLMs bring about a real transformation in how enterprises approach AI. Those enterprises that adapt their technological, organizational and compliance frameworks to leverage multi-modal LLMs will set the standard for the future of intelligent organizations.

5. Conclusion

Complexity and abundance of data in today's digital world highlight the inadequacies of unimodal AI systems. Multi-modal large language models (LLMs) offer a compelling evolution—enabling machines to understand, reason, and generate outputs across text, images, audio, and more. This ability introduces a step-change in how computational systems think and act. Systems that can take in diverse information and make more informed decisions based on a full understanding of the data.

Results of this research suggest that multi-modal LLMs are not just innovative technologies but also have the power to transform the way businesses operate. These models are revolutionizing the very ways businesses run, go to market and create value for customers. Review of different LLMs suggests that models such as GPT-4 Vision, Flamingo and PaLI can add value to enterprises in a range of ways, especially when matched with relevant business needs and tailored for domain specificity.

Getting the most out of multi-modal LLMs necessitates more than just implementing them. This requires strong technology infrastructures as well as adaptability within organizations, data protection and ethical consideration. Organizations should integrate multi-modal LLMs with a strategic focus, ensuring that technology, processes and ethics work together harmoniously.

Ultimately, the true advantage of multi-modal LLMs comes from how enterprises integrate and leverage these systems in their operations. Organizations that leverage all the modalities supported by multi-modal LLMs will drive the advancement of enterprise AI in the years to come.

6. Recommendation

Given the results and the discussion, this paper provides the following strategies for enterprises looking to leverage multi-modal LLMs for business benefits.

1. Ensure you have the necessary scalable and flexible infrastructure in place to efficiently use multi-modal LLMs.

Investing in advanced computing hardware such as GPUs and agile data storage solutions, is essential toprepare for efficient multi-modal model development and usage. Using cloud-based services with the ability to scale resources on demand is more affordable than building and managing in-house IT infrastructure.

2. Develop Cross-Functional Data Integration Strategies

Multi-modal AI depends on the ability to blend and integrate diverse forms of information from texts, images, audio and sensors. Enterprises need to implement efficient data pipelines, metadata standards and support for interoperability to ensure the maximum value of their data can be utilized in feeding multi-modal LLMs.

3. Adopt Incremental and Domain-Specific Fine-Tuning

Organizations need to fine-tune generalists models on their proprietary data to make them more suitable for their particular use cases. Tailoring multi-modal models for specific business functions helps improve accuracy and ensures the best results for tasks such as medical diagnosis, risk assessment and product evaluation.

4. Encourage AI Literacy and Shared Collaboration Among Teams

Coordinating with different departments plays a crucial role in successfully integrating multi-modal AI into day-to-1 Organizations should provide training to employees in AI to enable an environment that encourages continuous experimentation, learning and improvement.

5. Ensure Multi-Modal AI Is Held to Strict Standards of Security and Ethics.

Enterprises need to develop clear ethical guidelines and implement an open and accountable AI strategy in order to mitigate the risks and consequences of using more sophisticated multi-modal models. Regularly conducting bias audits, ensuring people supervise crucial decisions and adhering to regulations to maintain trust and accountability are all integral parts of governance.

6. Collaborate with AI Research Institutes, Technology Vendors and Industry Consortia

Joining forces with AI experts, technology providers and other organizations in the field expands opportunities to obtain advanced multi-modal models, up-to-date know-hows and tools for seamless deployment. Collaboration helps minimize the likelihood of technical issues and speeds up the time needed to see returns on investment.

Applying these guidelines prepares enterprises for advancing with multi-modal LLMs and propels them to the forefront of AI innovation as they realize significant business benefits.

References

1. Apple. (2023, October 31). *Apple's MM1 AI model shows a sleeping giant is waking up*. Wired. https://www.wired.com/story/apples-mm1-ai-model-sleeping-giant-waking-up(WIRED)
2. Base64.ai. (2024, April 15). *Why multimodal LLM is a game-changer for operations teams*. https://base64.ai/resource/why-multimodal-llm-is-a-game-changer-for-operations-teams/(base64.ai)
3. Business Insider. (2024, September 9). *44 of the most promising AI startups of 2024, according to top VCs*. https://www.businessinsider.com/ai-startups-most-promising-2024-9(Business Insider)
4. BytePlus. (2024, June 12). *Multimodal LLM examples*. https://www.byteplus.com/en/topic/516122(BytePlus)
5. Daffodil Software. (2024, March 5). *How multimodal LLMs are shaping the future of AI*. https://insights.daffodilsw.com/blog/how-multimodal-llms-are-shaping-the-future-of-ai(Daffodil Software)
6. Google DeepMind. (2023, May 10). *Gemini (language model)*. Wikipedia. https://en.wikipedia.org/wiki/Gemini_(language_model)(Wikipedia)
7. Microsoft. (2024, May 20). *Introducing the Azure multimodal AI & LLM processing accelerator*. https://techcommunity.microsoft.com/t5/ai-azure-ai-services-blog/the-azure-multimodal-ai-amp-llm-processing-solution-accelerator/ba-p/4258071(TECHCOMMUNITY.MICROSOFT.COM)
8. Osiz Technologies. (2024, July 18). *The potential of multimodal LLMs to drive AI development*. https://www.osiztechnologies.com/blog/multimodal-llms-in-ai-technology(Osiz Technologies)
9. ProjectPro. (2024, August 22). *Multimodal LLMs: Learn how MLLMs blend vision & language*. https://www.projectpro.io/article/multimodal-llms/1054(ProjectPro)
10. Shaip. (2025, February 10). *What are multimodal large language models? Applications, challenges, and how they work*. https://weareshaip.medium.com/what-are-multimodal-large-language-models-applications-challenges-and-how-they-work-b6cc1bbcdcff(Medium)
11. Turing. (2024, June 5). *Expert multimodal LLM training services*. https://www.turing.com/services/llm-multimodality(Turing)
12. WalkingTree Technologies. (2024, August 14). *Deciphering the synergy of multi-modal intelligence: Pioneering use cases and innovations in the enterprise domain*. https://walkingtree.tech/deciphering-the-synergy-of-multi-modal-intelligence-pioneering-use-cases-and-innovations-in-the-enterprise-domain/(WalkingTree Technologies)
13. Waymo. (2024, October 30). *Waymo explores using Google's Gemini to train its robotaxis*. The Verge. https://www.theverge.com/2024/10/30/24283516/waymo-google-gemini-llm-ai-robotaxi(The Verge)
14. Wired. (2023, October 31). *Apple's MM1 AI model shows a sleeping giant is waking up*. https://www.wired.com/story/apples-mm1-ai-model-sleeping-giant-waking-up(WIRED)
15. Wikipedia Contributors. (2024, May 10). *Gemini (language model)*. Wikipedia. https://en.wikipedia.org/wiki/Gemini_(language_model)(Wikipedia)
16. Osiz Technologies. (2024, July 18). *The potential of multimodal LLMs to drive AI development*. https://www.osiztechnologies.com/blog/multimodal-llms-in-ai-technology(Osiz Technologies)
17. ProjectPro. (2024, August 22). *Multimodal LLMs: Learn how MLLMs blend vision & language*. https://www.projectpro.io/article/multimodal-llms/1054(ProjectPro)

18. Shaip. (2025, February 10). *What are multimodal large language models? Applications, challenges, and how they work.* https://weareshaip.medium.com/what-are-multimodal-large-language-models-applications-challenges-and-how-they-work-b6cc1bbcdcff(Medium)
19. Turing. (2024, June 5). *Expert multimodal LLM training services*. https://www.turing.com/services/llm-multimodality(Turing)
20. WalkingTree Technologies. (2024, August 14). *Deciphering the synergy of multi-modal intelligence: Pioneering use cases and innovations in the enterprise domain*. https://walkingtree.tech/deciphering-the-synergy-of-multi-modal-intelligence-pioneering-use-cases-and-innovations-in-the-enterprise-domain/(WalkingTree Technologies)