International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

FOUNDATION MODELS BEYOND LANGUAGE: A COMPREHENSIVE STUDY OF MULTIMODAL, AGENTIC, AND RETRIEVAL-AUGMENTED ARCHITECTURES FOR REAL-WORLD DECISION MAKING

Komal Saxena[1*], Mohit Ranjan Panda[2], A. Anthony Raj[3], Dr Deepak Vidhate[4], Elangovan Guruva Reddy[5], Prabir Kumar Das[6], Mayank Saini[7]

[1*]Associate Professor, AIIT, Amity University, NOIDA, U.P., India MAIL ID: ksaxena1@amity.edu Orchid ID: https://orcid.org/0000-0001-5070-8355
[2]Associate Professor, School of Computer Engineering, KIIT deemed to be University, Orchid ID - https://orcid.org/0000-0002-4816-478X, Email id - mohit.pandafcs@kiit.ac.in
[3]Assistant Professor, Panimalar Engineering College, Email ID: rajaloyola16@gmail.com Orcid ID: 0000-0001-5417-4856
[4]Professor & HOD IT, Dr Vithalrao Vikhe Patil College of Engineering, Ahilyanagar Email ID- dvidhate@yahoo.com
[5] Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Vijayawada – 522302, Guntur District, Andhra Pradesh.. Email: gurugovan@gmail.com, eguruvareddy@kluniversity.in ORCID ID: https://orcid.org/0000-0002-8490-6189
[6]Assistant Professor, Department of Computer Science & Engineering (CS&DS) Brainware University, Barasat. Personal_mail_id- pbrdas82@gmail.com, ORCID iD: 0009-0005-7680-2194
[7]Pre-final year student, Master's of Computer Application, Galgotias University, Greater Noida, Uttar Pradesh 203201 Email ID: mayanksainikht@gmail.com

Abstract

Foundation models have transformed artificial intelligence by demonstrating remarkable adaptability across a range of tasks. However, language-only approaches remain insufficient for real-world decision-making, where context requires perception across modalities, adaptive agency, and dynamic grounding in external knowledge. This study critically examines the evolution of foundation models beyond language by analyzing three emerging paradigms: multimodal models, agentic architectures, and retrieval-augmented systems. A systematic and analytical review was undertaken to explore how each paradigm addresses the limitations of traditional language-based models. Multimodal models expand the perceptual capacity of artificial intelligence through the integration of text, vision, and structured data. Agentic models move beyond passive output generation to autonomous reasoning and planning, supported by memory augmentation and external tool use. Retrieval-augmented models reduce hallucination and increase reliability by linking parametric knowledge with external databases. Comparative synthesis reveals that each paradigm contributes unique strengths but also introduces challenges related to scalability, safety, interpretability, and evaluation. The study highlights the theoretical alignment of these paradigms with cognitive functions such as perception, reasoning, memory, and action. It further emphasizes the need for ethical and regulatory safeguards to address bias, transparency, and human–artificial intelligence collaboration. The findings suggest that the future of foundation models lies in the integration of multimodality, agency, and retrieval, paving the way for unified decision-making systems that are both technically advanced and socially responsible.

Keywords: Foundation models, Multimodal artificial intelligence, Agentic architectures, Retrieval-augmented generation, Decision-making systems, Artificial intelligence ethics

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1. Introduction

The progress in artificial intelligence (AI) has been driven mostly by the creation of foundation models, which are general-purpose systems, and can be fine-tuned to a variety of tasks via transfer learning [1]. As a result of advances in deep neural network architecture, and training them at scale, foundation models, especially large language models (LLMs), have become a topic of discussion in computer science, industry, and society. This processing capacity to handle large volumes of data and produce coherent and context-sensitive outputs has made them useful in natural language processing and translation as well as decision-support systems and in the development of creative content [2]. This has changed the image of AI, as no longer a specialized problem-solving system but as a generalizable and scalable framework capable of serving a wide range of fields.

The popularity of LLMs and other foundation models is constrained, even at the level of addressing real-world reasoning and decision-making, although their applications are becoming increasingly popular. Research has pointed out that these models are not very effective when it comes to instances where actions require the generation of text, summarization and understanding of semantics but fail to provide more profound understanding of the human being and the contextual meaning [3]. The real world situation can hardly be reduced to a purely linguistic expression. They require combination of various modalities like vision, audio, structured information, and sensory information and also need reasoning abilities that transcend beyond passively predicting. Such a mismatch highlights one of the key research gaps: language-only models are very strong but cannot be used to make important and responsible decisions in dynamic and unpredictable situations [4].

The weaknesses of language-based models are especially apparent in high-stakes contexts including healthcare, finance, law, and autonomous systems in which correct decisions depend upon multimodal information and the understanding of context. Although under some conditions, LLMs can demonstrate emergent capabilities, e.g. in-context learning, abstraction or reasoning [5], they are not always trustworthy. Their appearance is very sensitive to scale and exposure to data, which causes unpredictable behaviors. Moreover, since scaling laws have been shown to improve as a function of model size and training data, but do not necessarily imply interpretability and generalizable reasoning [6]. Therefore, excessive use of scaling as the approach towards intelligence creates technical and ethical threats.

The other important constraint is the lack of transferability of the knowledge of LLM to new fields. Transfer learning has been identified as a very strong paradigm to ensure that pretrained features are used in different tasks [7], however, its effectiveness in text-based models does not necessarily transfer to multimodal and decision-intensive models. As one example, although LLMs can be trained to handle new tasks using prompt engineering or fine-tuning, they do not have the innate capability to base their reasoning on external data sources or sensory modalities. This limits their performance in situations where perception, action and dynamic knowledge integration is very important.

Researchers have come to suggest that not all of the most sophisticated capabilities of LLMs are deliberate, but emerge as a result of training [8]. Such emergent behaviors as few-shot learning and complex reasoning are creating opportunities and challenges. On the one hand, they show how much foundation models can generalize in unforeseen directions by being untapped. On the downside, the vagaries of such emergent characters inhibit their use in areas where reliable and consistent traits are required. Emergent abilities on their own cannot support real world decision-making systems without well-developed grounding, validation and integration of multimodal input.

In order to overcome these problems, it is informative to re-examine the domain of cognitive architectures, which have long attempted to replicate human-like reasoning by incorporating perception, memory and action [9]. Although LLMs and other foundation models offer a level of scale and language proficiency never before seen, they do not have the organized processes of autonomy, planning, and contextual adaptation that cognitive systems studies focus on. To bridge these paradigms a more detailed conceptualization of the ways in which foundation models can be scaled

Received: August 14, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

up to multimodal, agentic, augmented with retrieval is required. This integration is more of a conceptual rather than a technical nature and is more of a transition between passive models of prediction and active systems of decision support.

The gaps in the current approaches have also been emphasized by recent surveys of reasoning in foundation models [10]. Although these models are making steps in the right direction in terms of structured reasoning, they fail to provide transparency, accountability and grounding on real-world data. Their thinking styles are more likely to be statistical pattern-matching than actual thinking. This constraint highlights the need to develop architectures that can produce responses but can also reason in the same way that human beings make decisions.

It is against this background that the current research paper seeks to investigate systematically foundation models outside the language domain and in specific three crucial areas, namely, multimodal integration, agentic architectures, and retrieval-augmented systems. All these paradigms are separate but complementary to the failures of the models of language only. Multimodal models seek to admit more context with the aid of information on multiple sources; agentic models are concerned with autonomy, planning and action, and retrieval-augmented systems seek to resolve the issue of grounding by dynamically relating models with external knowledge repositories. A combination of these paradigms is an indication of the direction that the foundation models are taking to the real world.

The article has a two-fold contribution. This synthesizes the existing body of knowledge in multimodal, agentic, and retrieval-augmented paradigms and critically examines their strengths, limitations, and implications to decision-making. Second, it suggests a conceptual framework of integrating these paradigms, providing a roadmap of moving foundation models to reliable, trustworthy and scalable decision-making platforms. By engaging both technical insights and societal considerations, this article positions itself at the intersection of AI research, cognitive science, and applied decision-making.

Objectives of the Study:

1. To critically review and synthesize the role of multimodal, agentic, and retrieval-augmented paradigms in overcoming the limitations of language-only foundation models
2. To propose a conceptual framework that integrates these paradigms into a unified pathway for real-world decision-making with foundation models

2. Theoretical Foundations

Theoretical foundations underpinning the evolution of foundation models highlight why they are increasingly being extended beyond language-only systems to multimodal, agentic, and retrieval-augmented paradigms. These foundations rest on principles of scaling, emergent abilities, transfer learning, and cognitive-inspired models of decision-making. Together, they provide a conceptual lens for understanding both the strengths and limitations of current approaches, while also explaining why more advanced paradigms are essential for real-world decision support.

2.1 Evolution from LLMs to Multimodal and Agentic Systems

Large language models (LLMs) represent a breakthrough in the use of large-scale pretraining for natural language tasks. The success has seen them being widely adopted in various applications however their functionality is limited by dependence on textual inputs only. The same way that simple medical conditions cannot be perceived through individual symptoms but need to be assessed in a multidimensional manner [6], the decision-making tasks in AI require the synthesis of various sources of data. The rise of multimodal architectures as the next stage in the evolution of linguistic systems can be seen as the acknowledgement that the real-life world needs more sophisticated modes of perception and contextual processing [7].

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In the same way, transition to agentic systems represents a loss of passive prediction in favor of autonomy, flexibility and engagement with dynamic environments. Similar to the clinical research that emphasizes patient-reports and experiences in addition to biological indicators [8], AI researchers are finding it clear that models should not be passive, but proactive in terms of planning and tool utilization. This trend can be visualized as a continuum of more and more context-grounded and decision-making sophisticated systems as Figure 1 illustrates the development of LLMs into multimodal models, agentic systems, and retrieval-augmented architectures.

Figure 1. Evolutionary pathway of foundation models, progressing from LLMs to multimodal, agentic, and retrieval-augmented systems

2.2 Core Principles: Scaling Laws, Emergent Abilities, and Transfer Learning

Conceptual models of foundation models are based on three principles that are interrelated. The former is scaling laws, which include demonstrating predictably better performance with an increase in model size, dataset scale, and compute resources. This association has been mathematically explained as:

$$L(N, D) \propto N^{-\alpha} + D^{-\beta}$$

where $L(N, D)$ denotes the loss function, N is the number of model parameters, D the dataset size, and α, β represent scaling exponents [9]. While scaling delivers measurable gains, it also creates diminishing returns, making structural innovations more important than brute-force expansion.

The second is emergent abilities, in which new abilities are developed at scale without programmed instructions. These are few-shot reasoning, abstraction and in-context adaptation. Emergent behaviors are unpredictable, although they are promising which brings the question of reliability. Their randomness is similar to variability of the human quality-of-life outcome in complex situations, where the patterns emerge but cannot always be generalized [10].

The third principle is transfer learning which allows making use of learned representations to be reused across tasks so as to offer adaptability. This allows the efficient use of pretrained features to novel domains, but is also susceptible to transfer of bias. Table 1 is a summary of these three theoretical principles, their descriptions, and how they can be applied in the real world when making decisions.

Table 1. Core theoretical principles of foundation models and their implications for decision-making

Principle	Description	Implication for Decision-Making
Scaling Laws	Predictable improvements in performance with more parameters and data [9]	Efficiency gains, but diminishing returns and resource concerns
Emergent Abilities	Unprogrammed capabilities arising at scale [9], [10]	New reasoning opportunities, but limited reliability
Transfer Learning	Reuse of learned features across tasks [10]	Adaptability across domains, but risk of bias transfer

By synthesizing scaling, emergence, and transfer learning, it becomes evident that foundation models are theoretically powerful, but require further refinements to achieve dependable decision support in practice.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.3 Decision-Making as a Cognitive Process

Human decision-making is not linear or restricted to language; it involves an interplay of perception, reasoning, memory, and action. Foundation models restricted to text inputs lack perceptual grounding, preventing them from integrating multimodal signals effectively. Multimodal models address this by incorporating vision, audio, and structured data, thereby expanding situational awareness. Reasoning remains a challenge, as models often engage in statistical pattern-matching rather than logical inference. Similarly, memory both episodic and semantic remains underdeveloped in most current systems. Finally, action requires not only prediction but also the capacity for planning and adaptive engagement with dynamic environments.

Figure 2 illustrates these four interconnected components perception, reasoning, memory, and action positioned as the cognitive foundations of decision-making. The diagram demonstrates how each component interacts with the others, creating a feedback loop necessary for adaptive intelligence.

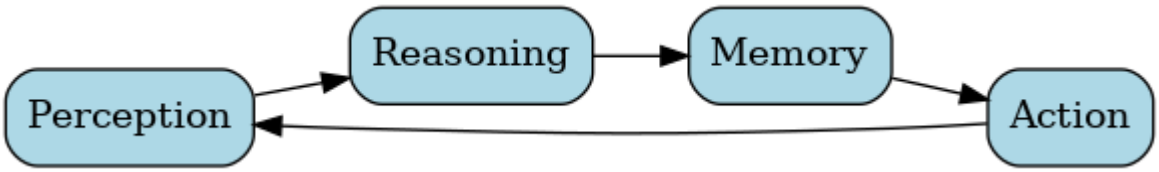

Figure 2. Components of cognitive decision-making perception, reasoning, memory, and action as a conceptual framework for extending foundation models

By aligning foundation model architectures with this cognitive framework, researchers can identify specific gaps, such as the absence of grounded perception in LLMs or the limited memory in current agentic systems, and target them through multimodal, agentic, and retrieval-augmented approaches.

2.4 Reason For Paradigms Are Suitable for Real-World Decision Support

Real-world decision-making requires more than predictive accuracy; it demands robustness, contextual grounding, and adaptability. Multimodal models are suited to these challenges because they integrate diverse streams of data, akin to how multidimensional assessments provide richer understanding in population studies [11]. Agentic models introduce autonomy and planning, allowing systems to adapt to uncertainty and interact with dynamic environments. Retrieval-augmented systems bridge foundation models with knowledge repositories, which overcomes the weaknesses of parametric memory and gives the foundation model grounding in verifiable data.

Collectively, these paradigms answer the shortcomings of language-only models directly by incorporating perception, reasoning, memory and action into their designs. This synthesis makes them not just as new technical developments but also as practical decision support structures in areas where reliability, transparency and responsibility are critical.

3. Paradigms Beyond Language

The shortcomings of language-only models have inspired the creation of novel paradigms extending foundation models to the worlds of perception, action and integration of external knowledge. Such paradigms multimodal foundation models, agentic models and retrieval-augmented architectures are the most notable milestones towards real-life decision support. They have their own contributions and their convergence is an indication of a route towards strong, credible, and adaptive systems.

3.1 Multimodal Foundation Models

Architectures

Multimodal models are an expansion of large language models, where information of various modalities, including vision, audio, video, and structured data, is used. They allow more contextual

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

grounding through finding correspondence among modalities, which makes it possible to reason outside of the text. Recent advances, such as multimodal transformers [12], contrastive pretraining frameworks like CLIP [13], and domain-specific models such as Med-Flamingo in medical imaging [14], illustrate the diversity of architectures. Notably, Gemini has introduced architectures capable of processing text, vision, and code simultaneously [15]. These approaches reflect a paradigm shift from single-modality language models to systems that emulate human perception by unifying sensory input. Table 2 summarizes representative multimodal architectures, their modalities, and application domains.

Table 2. Representative multimodal foundation models and their applications

Model/Framework	Modalities	Domain/Application	Reference
Multimodal Foundation Model	Text + Vision + Audio	General intelligence	[12]
CLIP	Text + Vision	Robust image-text retrieval, robustness	[13]
Med-Flamingo	Text + Vision	Medical diagnostics (radiology, pathology)	[14]
Gemini	Text + Vision + Code	Broad AI assistance	[15]
Vision-Language Models	Text + Vision	Robotics, perception	[16]

This table highlights how multimodal architectures differ in scope, ranging from specialized healthcare applications to general-purpose models, reflecting both scalability and domain specificity. Multimodal models have demonstrated their utility across healthcare, robotics, finance, and customer interaction. In healthcare, Med-Flamingo supports diagnosis by combining clinical text and radiology images [14]. In robotics, hybrid multimodal approaches facilitate navigation and perception [17]. In finance, multimodal data streams (e.g., reports + market signals) enhance decision-making reliability. These applications illustrate that multimodality directly contributes to situational awareness in high-stakes environments.

Despite progress, multimodal models face challenges of alignment, where different modalities may not map seamlessly to a shared latent space [12]. Data imbalance across modalities also limits generalization [16]. Evaluation remains difficult, as no universal benchmarks exist for multimodal reasoning. Figure 3 illustrates the multimodal architecture pipeline, showing how inputs from text, vision, and structured data are fused in shared latent spaces before downstream decision tasks.

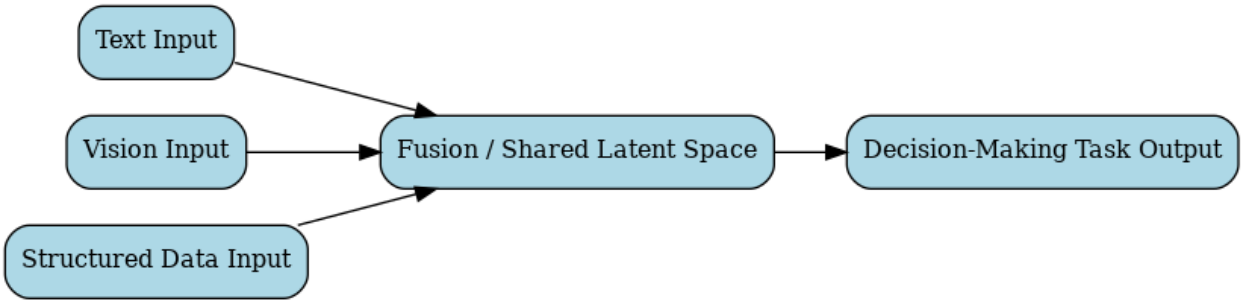

Figure 3. General pipeline of a multimodal foundation model, integrating text, vision, and structured data into shared latent representations for decision-making

3.2 Agentic Models

Active Decision-Making Architectures

The foundation models are extended to agentic models, which are models of active action, with planning, reasoning, and the use of tools. This paradigm is an indication of the increasing need of dynamically adaptable systems. Examples of this shift include architectures like agentic LLM architectures [18], compositional planning models [19] and tool-augmented systems (TALM) architectures [20]. More recent developments are memory-augmented architectures like MemInsight [21] which give episodic and semantic recall to agents. The architecture of an agentic model is

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

illustrated in Figure 4, which illustrates the interaction of planning, tool integration, and memory to make autonomous decisions.

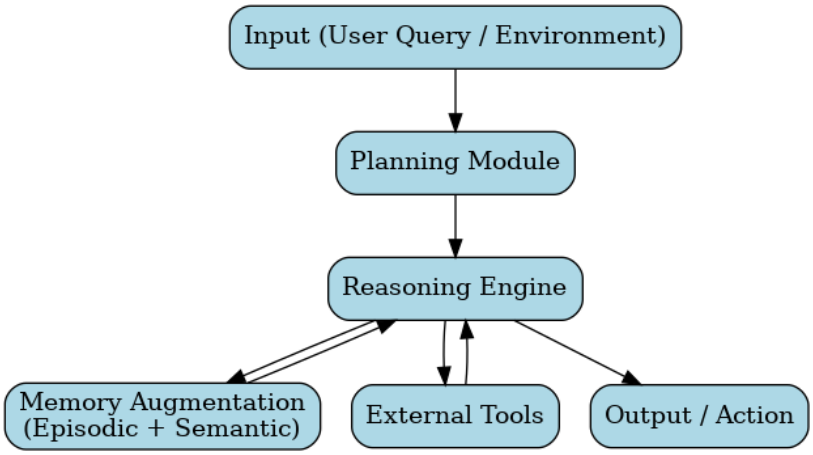

Figure 4. Conceptual architecture of agentic foundation models, integrating planning, external tools, and memory augmentation for active decision-making

The agentic models are also being used in areas like autonomous robotics where the reinforcement learning is combined with the foundation models of adaptive control [22]. Generative AI agents of autonomous machines draw attention to opportunities and risks in safety-critical situations [23]. Scientific discovery and policy design based on simulation is also performed using these systems. Planning and adjusting agentic models bring them a step closer to models that can be autonomous in the real world. The hope of agentic models is restrained by safety and responsibility issues [22]. One of the basic issues is to make sure that autonomous decisions are in line with human values. Scalability is a factor, as well: memory-augmented and tool-augmented systems require large amounts of computing resources [21]. Lastly, interpretability has not been developed yet, and actions produced by agentic systems are frequently not transparent to human control [18].

3.3 Retrieval-Augmented Models

Architectures

Retrieval-augmented generation (RAG) represents a powerful strategy for grounding foundation models in external knowledge bases [24]. These architectures combine parametric knowledge (stored within model weights) with dynamic retrieval pipelines. Advances include interactive evaluation frameworks like Kieval [25], integration with vector databases [26], and long-term memory systems such as Think-in-Memory [27]. Retrieval-enhanced editing methods allow LLMs to update knowledge dynamically for tasks like multihop question answering [28].

The retrieval mechanism is often formalized as:

$$P(y \mid x) = \sum_{k=1}^{K} P(y \mid x, r_k) \cdot P(r_k \mid x)$$

where x is the input query, y the generated output, and r_k retrieved knowledge passages. This illustrates how outputs are conditioned jointly on input and external retrieval [24].

RAG-based systems have been applied in knowledge-intensive NLP tasks [24], compliance analysis, and scientific literature-based reasoning. Vector database integrations extend decision-making in business intelligence [26]. Memory-augmented retrieval has also improved long-term context handling [27]. Table 3 summarizes retrieval-augmented architectures and their contributions.

Table 3. Retrieval-augmented approaches and their key contributions

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Approach	Contribution	Reference
Retrieval-Augmented Generation (RAG)	Combines parametric and non-parametric memory for knowledge-intensive tasks	[24]
Kieval	Knowledge-grounded evaluation framework	[25]
Vector Database + LLM	Enhances retrieval for decision tasks	[26]
Think-in-Memory	Enables long-term recall and reasoning	[27]
Retrieval-Enhanced Editing	Supports knowledge updating and QA	[28]

This table highlights how retrieval mechanisms expand the horizon of foundation models by addressing the hallucination problem and enabling dynamic knowledge updates.
Retrieval-augmented systems introduce challenges of latency and scalability, as real-time retrieval requires efficient infrastructure [26]. The quality of external knowledge also directly affects reliability, raising concerns about misinformation [25]. While retrieval reduces hallucination, it does not eliminate it entirely [28]. Figure 5 illustrates the RAG pipeline, showing how queries are processed through retrieval, grounding, and generation modules.

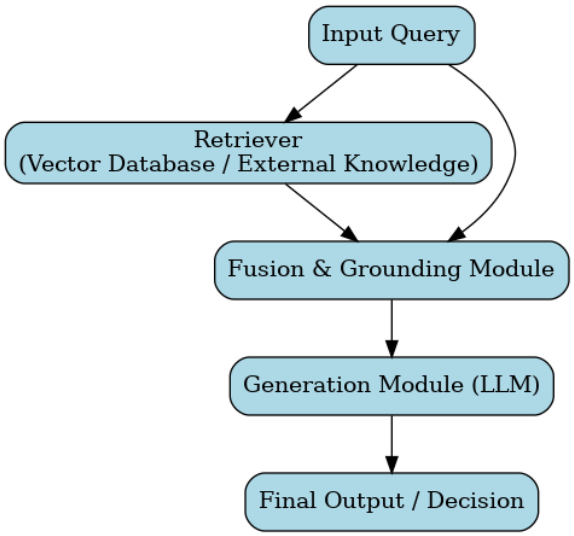

Figure 5. General pipeline of a retrieval-augmented model, showing interaction between input queries, external databases, and generation modules

3.4 Synthesis of Paradigms

The three paradigms multimodality, agency, and retrieval address complementary weaknesses of language-only models. Multimodality expands perception, agentic models introduce autonomy, and retrieval architectures provide grounding. Together, they outline a path toward robust foundation models capable of real-world decision-making.

4. Comparative Analysis and Discussion

Multimodal, agentic, and retrieval-augmented foundation models present both opportunities and trade-offs in efficiency, scalability, and reliability, making it possible to make more effective decisions. A comparative lens helps to better appreciate the complementary and challenging nature of these paradigms in relation to each other and also unveil theoretical implications, practical applications, and research questions.

4.1 Cross-Comparison: Strengths, Weaknesses, and Trade-Offs

The three paradigms multimodal, agentic, and retrieval-augmented have their own strengths and weaknesses. Multimodal models offer enhanced perception through combination of multiple streams of data although they have problems of synchronization and resource intensity [29]. The agentic

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

models are extensions of the foundation models to autonomy and adaptive reasoning, but they are also associated with interpretability and safety issues [30]. Retrieval-augmented architectures enhance grounding based on external sources of knowledge with the weak point of reliability and latency of retrieval [31]. Table 4 will compare these paradigms with regard to their strengths, their weaknesses and trade-offs.

Table 4. Comparative strengths, weaknesses, and trade-offs of multimodal, agentic, and retrieval-augmented models

Paradigm	Strengths	Weaknesses / Challenges	Trade-Offs	Reference
Multimodal Models	Rich perception, information fusion [29], [33]	High compute cost, modality alignment issues	Accuracy vs scalability	[29], [33]
Agentic Models	Autonomy, adaptive planning [29], [32]	Interpretability, safety concerns	Flexibility vs reliability	[29], [32]
Retrieval-Augmented	Grounded outputs, reduces hallucination [30]	Latency, dependency on external data quality	Knowledge depth vs response speed	[30]

This table highlights how no paradigm provides a complete solution in isolation, making integration an attractive strategy for balancing perception, autonomy, and grounding.

4.2 Theoretical Implications: Toward Cognitive-Like Architectures

From a theoretical perspective, the convergence of paradigms echoes long-standing efforts in cognitive systems research to model perception, memory, reasoning, and action in unified frameworks [32]. Multimodal architectures resemble human sensory integration, while agentic models emulate planning and adaptive decision-making. Retrieval-augmented frameworks provide externalized memory, functioning as an analogue to knowledge retrieval in human cognition. Together, these paradigms suggest an emerging cognitive-like architecture where each paradigm maps to a distinct cognitive function [33]. Figure 6 illustrates this theoretical alignment, mapping multimodal perception, agentic reasoning, and retrieval-based memory into a cognitive architecture inspired by human decision-making processes.

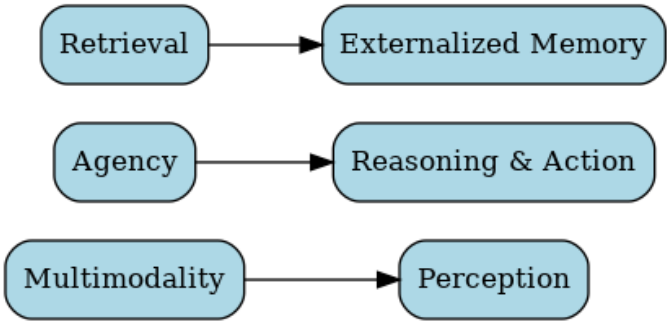

Figure 6. Theoretical alignment of paradigms with cognitive functions: multimodality as perception, agency as reasoning and action, and retrieval as externalized memory

This framing suggests that the integration of paradigms may approximate cognitive architectures that are both adaptive and generalizable, providing a foundation for artificial decision-making systems closer to human intelligence.

4.3 Practical Implications: Deployment in Healthcare, Finance, Robotics, and Policy

The practical value of these paradigms becomes most evident in domain-specific applications. In healthcare, multimodal systems enhance diagnostic accuracy by fusing radiology images with textual medical notes, though trade-offs arise between unified large-scale models and localized fine-tuned

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

approaches for highly specialized tasks [30]. Retrieval-augmented methods further support evidence-based diagnostics by linking decisions to medical literature.

In finance, retrieval-augmented systems enable real-time integration of structured databases and unstructured market reports, mitigating hallucination risks. Multimodality is particularly relevant for sentiment analysis, combining textual and visual market indicators [33]. In robotics, agentic models integrate reinforcement learning and planning modules to support autonomous navigation and interaction. Wireless agentic AI frameworks show promise for combining multimodal semantic perception with retrieval pipelines to enable adaptive robotic control [29].

Policy-making and governance benefit from hybrid models that combine multimodal situational awareness (e.g., text, geospatial imagery) with retrieval of legal and regulatory frameworks, supported by agentic planning mechanisms [32]. These deployments, however, highlight the trade-off between accuracy and interpretability, a recurring challenge across domains.

4.4 Limitations of Current Research

Despite progress, current research remains limited by boundaries in evaluation, scalability, and integration. Benchmarks for assessing AI remain fragmented, with studies highlighting inconsistencies in evaluation standards [31]. Multimodal benchmarks often lack cross-domain generalizability, while agentic benchmarks fail to account for safety and interpretability. Retrieval-based evaluation is challenged by the difficulty of tracing knowledge provenance.

Scalability poses another limitation. Multimodal and agentic systems are resource intensive in terms of compute and data requirements, and present a sustainability issue [29]. Even though retrieval-augmented approaches are efficient in memory, they also create latency which limits their use in real-time systems [30]. The combination of the paradigms is more theoretical than practical because most of the deployed systems are restricted to two-paradigm combinations, but not multimodal-agentic-retrieval architecture.

5. Ethical, Societal, and Regulatory Implications

In extending foundation models beyond language to multimodal, agentic, and retrieval-augmented systems, the pressing nature of the need to ensure that their impacts on ethics, society, and regulation are considered is critical. As these models become more sophisticated, they present the challenges of bias, governance, trustworthiness and interpretability. This section critically analyses these challenges and explains frameworks that would make sure that decision-making AI develops responsibly.

5.1 AI Ethics in Multimodal Models

Multimodal models combine vision, text, audio and structured information making it difficult to ethically control. In contrast to unimodal systems, multimodal systems encounter the problem of propagation of bias across modalities, which makes the notions of fairness and accountability a concern [34]. Ethical principles that apply particularly to multimodal AI emphasize openness in the data curation, cross-modal reasoning interpretability and elucidable decision pipelines. Such frameworks contend that the design process needs to have ethical protection embedded in it, as opposed to coming in after the fact. The general ethical risk map of multimodal models (Figure 7) indicates the points where bias may be introduced during the processes of data collection, model training, and inference.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

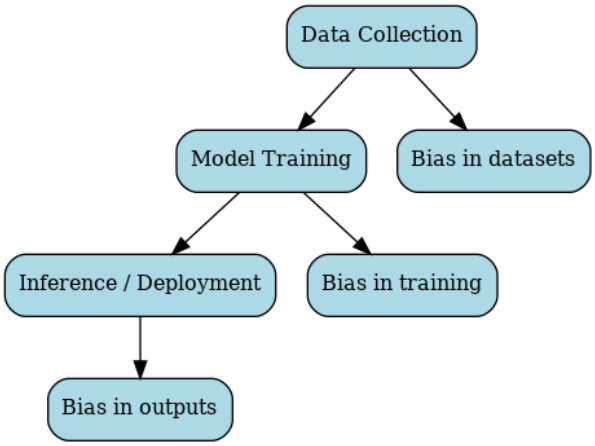

Figure 7. Ethical risk map in multimodal AI, showing bias entry points across data, training, and inference pipelines

This highlights that ethics in multimodal systems requires a proactive approach, balancing technical design with societal accountability.

5.2 Bias in Foundation Models

Prejudice is also a structural issue in foundation models because of the size and diversity of training data [35]. Biased predictions are disproportionately relevant to vulnerable groups in healthcare and other policy applications. Prejudice does not only destroy fairness, but it also diminishes the reliability of decision-making products. To solve these problems, both technical solutions like bias detection and mitigation and regulatory controls should be used to control the outputs of models to ethical standards. Table 5 provides a summary of bias types that are typical of foundation models and the implications in society.

Table 5. Categories of bias in foundation models and their societal implications

Bias Category	Example Manifestation	Societal Consequence	Reference
Demographic Bias	Overrepresentation of specific groups	Discrimination in healthcare, hiring	[35]
Domain Bias	Unequal performance across tasks	Reduced reliability in critical domains	[35]
Cultural Bias	Insensitivity to linguistic/cultural diversity	Marginalization of underrepresented groups	[35]

The persistence of bias suggests that fairness must remain a central component of both design and governance.

5.3 AI Governance Frameworks

The need for structured governance frameworks is now widely recognized [36]. Governance addresses how systems are developed, monitored, and regulated across their lifecycle. Proposed frameworks stress principles such as accountability, transparency, and human oversight. These frameworks are not only regulatory tools but also technical guides for model deployment, ensuring that the ethical dimensions of AI remain central to its practical application. The integration of governance models across industries will determine whether AI technologies can scale responsibly.

5.4 Regulation of Autonomous Decision-Making Systems

The agentic foundation models which can think and act independently present special regulatory challenges [37]. Their application to such areas of life as healthcare, robotics, and transportation can be held responsible because mistakes can lead to on-the-job accidents. Such regulation should then focus on safety auditing, accountability attribution and explainability. Furthermore, agentic systems

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

are dynamic in nature and thus, regulations should be flexible to changing risks. Figure 8 presents a regulatory life cycle of autonomous systems, including training the models and deploying them, as well as ongoing audits.

Figure 8. Regulatory lifecycle of autonomous decision-making systems, including pre-deployment auditing, deployment oversight, and post-deployment monitoring

This lifecycle approach underscores that regulation cannot be a one-time effort but must remain continuous and adaptive.

5.5 Trustworthy AI and Transparency

The concept of AI trustworthiness goes beyond technical reliability to include transparency, strength and acceptance in society [38]. A reliable system is not only correct but also readable and in accordance with human values. Transparency is especially important in the context of decision-making; the stakeholders should be knowledgeable of how the results are obtained. This involves giving explainable descriptions of what model prediction is and making sure the decision-making pipelines are auditable. Credible AI means that technical protections are appropriate to the expectations of the culture and society and thus have legitimacy.

5.6 Human–AI Collaboration Ethics

Implementation of AI into decision-making presents some ethical concerns of human-AI cooperation. Ethical collaboration models propose that AI is not to substitute human judgment; rather, it should supplement it [39]. This necessitates that systems be made to be complementary, so that human decision-makers do not lose control and power. The dependency risks are also ethical issues, which involve over-dependence on AI to the point that human agency is weakened. Responsible collaboration models highlight the creation of interfaces enabling human interpretation of AI outputs and taking action.

5.7 Explainability and Interpretability in AI

One of the most problematic areas of ensuring ethical AI is explainability [40]. Trust cannot be built without being interpretable and accountability is impossible to find. Psychological accounts emphasize that explainability should be subject to human cognitive demands, and be explicable, not in technical terms. Explainability methods are model distillation, visualization, and natural language explanation. These approaches should be made in such a way that the explanations are technical and cognitive. Some of the most important dimensions of explainability in the context of decision-making AI are listed in Table 6.

Table 6. Dimensions of explainability and interpretability in AI

Dimension	Description	Ethical Relevance	Reference
Transparency	Clarity of decision pathways	Builds trust	[40]
Accountability	Traceability of decision-making steps	Ensures oversight	[40]
Cognitive Alignment	Explanations suited to human understanding	Enhances usability	[40]

It demonstrates that the interpretability is not a pure technical issue, but the ethical one that requires an orientation to the human values and cognition.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Ethical, social, and regulatory considerations are the determining factors in application of foundation models in the real world decision making. With the multimodal ethics to mitigation of bias, governance system, regulation, credibility, collaboration and interpretability, all aspects bring out the interaction of technology and society. Tackling these implications involves not only technical innovation, but also robust interdisciplinary interactions in the fields of ethics, law, policy and cognitive science.

6. Future Directions and Conclusion

The history of the development of foundation models beyond language paves a path to the integration of multimodality, agency, and retrieval in the form of unified architectures. All these paradigms overcome a particular limitation of the traditional models based only on language: multimodality allows perception of various inputs, agency allows adaptive reasoning and autonomous action, and retrieval allows grounding in credible external knowledge. The combination of these abilities is pointing towards context-aware and decision-capable systems, which will be the next generation of artificial intelligence. Another key milestone in this direction is the emergence of new standards that can be used to assess the effectiveness of decision making in its entirety. Existing benchmarks are more likely to focus on single modalities, or simple reasoning tasks, but another approach that needs to be implemented in new benchmarks is to evaluate systems by their capacity to combine perception, reasoning, memory, and action in dynamic and complex environments. Trustworthiness, transparency, and adaptability should also be measured by such benchmarks as evaluation should be consistent with the real-world needs. These models will probably be put to test in areas that traditional AI has failed in such as frontier applications. An example of this is climate modeling, which involves the synthesis of multimodal streams of data, such as satellite data or the policy report. The analysis of policies at the global level requires basing on massive and dynamic knowledge bases but with agentic reasoning to simulate and predict. Healthcare crises also require evidence-based decision-making that is quick and multimodal and also retrieves the changing clinical guidelines and autonomous triage assistance. These stakes-high areas demonstrate the need to have single-purpose architectures that can be used with certainty in the face of uncertainty. Over the long term, such advances will open the way to Artificial General Decision-Making Systems (AGDMs) models that are not only capable of processing information but also are flexible, transparent, and ethically sound enough to make decisions in a wide range of fields. Although AGDMs are still aspirational, the fact that they are being worked on points out the significance of interdisciplinary cooperation. The input of computer science, cognitive science, ethics, law and public policy will be needed to make sure that these systems are not only technically superior but also socially accountable. Finally, foundation models will have the future of going beyond language to unified, cognitively inspired architectures. This vision will necessitate new standards, pioneer applications and cross-disciplinary collaborations, which will make the decision-making AI grow as a scientific and social innovation.

References:

1. Chen, P. Y., & Liu, S. (2025). *Introduction to Foundation Models*. Springer Nature.
2. Sindhu, B., Prathamesh, R. P., Sameera, M. B., & KumaraSwamy, S. (2024, May). The evolution of large language model: Models, applications and challenges. In *2024 international conference on current trends in advanced computing (ICCTAC)* (pp. 1-8). IEEE.
3. Cuskley, C., Woods, R., & Flaherty, M. (2024). The limitations of large language models for understanding human language and cognition. *Open Mind*, *8*, 1058-1083.
4. Wallace, R. (2018). AI in the Real World. In *Carl von Clausewitz, the Fog-of-War, and the AI Revolution: The Real World Is Not A Game Of Go* (pp. 1-45). Cham: Springer International Publishing.

Received: August 14, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5. Berti, L., Giorgi, F., & Kasneci, G. (2025). Emergent abilities in large language models: A survey. *arXiv preprint arXiv:2503.05788*.
6. Rosenfeld, J. S. (2021). Scaling laws for deep learning. *arXiv preprint arXiv:2108.07686*.
7. Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., & Tekinerdogan, B. (2019). Analysis of transfer learning for deep neural network based plant classification models. *Computers and electronics in agriculture*, *158*, 20-29.
8. Havlík, V. (2025). Why are LLMs' abilities emergent?. *arXiv preprint arXiv:2508.04401*.
9. Thórisson, K., & Helgasson, H. (2012). Cognitive architectures and autonomy: A comparative review. *Journal of Artificial General Intelligence*, *3*(2), 1.
10. Sun, J., Zheng, C., Xie, E., Liu, Z., Chu, R., Qiu, J., ... & Li, Z. (2023). A survey of reasoning with foundation models. *arXiv preprint arXiv:2312.11562*.
11. Manzoor, M. A., Albarri, S., Xian, Z., Meng, Z., Nakov, P., & Liang, S. (2023). Multimodality representation learning: A survey on evolution, pretraining and its applications. *ACM Transactions on Multimedia Computing, Communications and Applications*, *20*(3), 1-34.
12. Fei, N., Lu, Z., Gao, Y., Yang, G., Huo, Y., Wen, J., ... & Wen, J. R. (2022). Towards artificial general intelligence via a multimodal foundation model. *Nature Communications*, *13*(1), 3094.
13. Fang, A., Ilharco, G., Wortsman, M., Wan, Y., Shankar, V., Dave, A., & Schmidt, L. (2022, June). Data determines distributional robustness in contrastive language image pre-training (clip). In *International conference on machine learning* (pp. 6216-6234). PMLR.
14. Moor, M., Huang, Q., Wu, S., Yasunaga, M., Dalmia, Y., Leskovec, J., ... & Rajpurkar, P. (2023, December). Med-flamingo: a multimodal medical few-shot learner. In *Machine Learning for Health (ML4H)* (pp. 353-367). PMLR.
15. Team, G., Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., ... & Blanco, L. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.
16. Bordes, F., Pang, R. Y., Ajay, A., Li, A. C., Bardes, A., Petryk, S., ... & Chandra, V. (2024). An introduction to vision-language modeling. *arXiv preprint arXiv:2405.17247*.
17. Aisha, L. (2024). Hybrid AI Models for Multimodal Data Analysis in Robotics, Medical Imaging, and Customer Experience.
18. Sypherd, C., & Belle, V. (2024). Practical considerations for agentic llm systems. *arXiv preprint arXiv:2412.04093*.
19. Ajay, A., Han, S., Du, Y., Li, S., Gupta, A., Jaakkola, T., ... & Agrawal, P. (2023). Compositional foundation models for hierarchical planning. *Advances in Neural Information Processing Systems*, *36*, 22304-22325.
20. Parisi, A., Zhao, Y., & Fiedel, N. (2022). Talm: Tool augmented language models. *arXiv preprint arXiv:2205.12255*.
21. Salama, R., Cai, J., Yuan, M., Currey, A., Sunkara, M., Zhang, Y., & Benajiba, Y. (2025). Meminsight: Autonomous memory augmentation for llm agents. *arXiv preprint arXiv:2503.21760*.
22. Jabbour, J., & Janapa Reddi, V. (2024, October). Generative AI agents in autonomous machines: A safety perspective. In *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design* (pp. 1-13).
23. Moroncelli, A., Soni, V., Shahid, A. A., Maccarini, M., Forgione, M., Piga, D., ... & Roveda, L. (2024). Integrating reinforcement learning with foundation models for autonomous robotics: Methods and perspectives. *arXiv preprint arXiv:2410.16411*.
24. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, 9459-9474.

Received: August 14, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

25. Yu, Z., Gao, C., Yao, W., Wang, Y., Ye, W., Wang, J., ... & Zhang, S. (2024). Kieval: A knowledge-grounded interactive evaluation framework for large language models. *arXiv preprint arXiv:2402.15043*.
26. Jing, Z., Su, Y., & Han, Y. (2025, February). When large language models meet vector databases: A survey. In *2025 Conference on Artificial Intelligence x Multimedia (AIxMM)* (pp. 7-13). IEEE.
27. Liu, L., Yang, X., Shen, Y., Hu, B., Zhang, Z., Gu, J., & Zhang, G. (2023). Think-in-memory: Recalling and post-thinking enable llms with long-term memory. *arXiv preprint arXiv:2311.08719*.
28. Shi, Y., Tan, Q., Wu, X., Zhong, S., Zhou, K., & Liu, N. (2024, October). Retrieval-enhanced knowledge editing in language models for multi-hop question answering. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management* (pp. 2056-2066).
29. Liu, G., Liu, Y., Zhang, R., Du, H., Niyato, D., Xiong, Z., ... & Jamalipour, A. (2025). Wireless agentic ai with retrieval-augmented multimodal semantic perception. *arXiv preprint arXiv:2505.23275*.
30. Wu, Z., Zhang, L., Cao, C., Yu, X., Liu, Z., Zhao, L., ... & Liu, T. (2025). Exploring the trade-offs: Unified large language models vs local fine-tuned models for highly-specific radiology nli task. *IEEE Transactions on Big Data*.
31. Reuel, A., Hardy, A., Smith, C., Lamparth, M., Hardy, M., & Kochenderfer, M. J. (2024). Betterbench: Assessing ai benchmarks, uncovering issues, and establishing best practices. *Advances in Neural Information Processing Systems*, *37*, 21763-21813.
32. Kandel, A., & Langholz, G. (2020). *Hybrid architectures for intelligent systems*. CRC press.
33. Zhang, C., Yang, Z., He, X., & Deng, L. (2020). Multimodal intelligence: Representation learning, information fusion, and applications. *IEEE Journal of Selected Topics in Signal Processing*, *14*(3), 478-493.
34. Patkar, U. C., Pitrubhakta, V., Sutar, S., Arora, A., Jaiswal, R., Bhardwaj, S., & Patil, U. S. (2025, January). Ethical Framework for Multimodal AI Systems. In *2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET)* (pp. 1-6). IEEE.
35. Czum, J., & Parr, S. (2023). Bias in foundation models: primum non nocere or caveat emptor?. *Radiology: Artificial Intelligence*, *5*(6), e230384.
36. Almeida, V., Mendes, L. S., & Doneda, D. (2023). On the development of AI governance frameworks. *IEEE Internet Computing*, *27*(1), 70-74.
37. Osborne, M., Hawkins, R., & McDermid, J. (2022, June). Analysing the safety of decision-making in autonomous systems. In *International Conference on Computer Safety, Reliability, and Security* (pp. 3-16). Cham: Springer International Publishing.
38. Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: a review. *ACM computing surveys (CSUR)*, *55*(2), 1-38.
39. Boni, M. (2021). The ethical dimension of human–artificial intelligence collaboration. *European View*, *20*(2), 182-190.
40. Broniatowski, D. A., & Broniatowski, D. A. (2021). *Psychological foundations of explainability and interpretability in artificial intelligence* (Vol. 4, p. 00). US Department of Commerce, National Institute of Standards and Technology.