Agentic AI: A Comprehensive Survey of Architectures, Applications, and Future Directions

Mohamad Abou Ali[1, 3, 4] and Fadi Dornaika[*1, 2]

[1]*University of the Basque Country*, [2]*IKERBASQUE*, [3]*Lebanese International University (LIU)*, [4]*The International University of Beirut*,

mohamad.abouali01@liu.edu.lb, fadi.dornaika@ehu.eus

Abstract

Agentic AI represents a transformative shift in artificial intelligence, but its rapid advancement has led to a fragmented understanding, often conflating modern neural systems with outdated symbolic models—a practice known as *conceptual retrofitting*. This survey cuts through this confusion by introducing a novel **dual-paradigm framework** that categorizes agentic systems into two distinct lineages: the **Symbolic/Classical** (relying on algorithmic planning and persistent state) and the **Neural/Generative** (leveraging stochastic generation and prompt-driven orchestration). Through a systematic PRISMA-based review of 90 studies (2018–2025), we provide a comprehensive analysis structured around this framework across three dimensions: (1) the theoretical foundations and architectural principles defining each paradigm; (2) domain-specific implementations in healthcare, finance, and robotics, demonstrating how application constraints dictate paradigm selection; and (3) paradigm-specific ethical and governance challenges, revealing divergent risks and mitigation strategies. Our analysis reveals that the choice of paradigm is strategic: symbolic systems dominate safety-critical domains (e.g., healthcare), while neural systems prevail in adaptive, data-rich environments (e.g., finance). Furthermore, we identify critical research gaps, including a significant deficit in governance models for symbolic systems and a pressing need for hybrid neuro-symbolic architectures. The findings culminate in a strategic roadmap arguing that the future of Agentic AI lies not in the dominance of one paradigm, but in their intentional integration to create systems that are both *adaptable* and *reliable*. This work provides the essential conceptual toolkit to guide future research, development, and policy toward robust and trustworthy hybrid intelligent systems.

Keywords— Agentic AI, artificial intelligence, systematic review, neural architectures, symbolic AI, multi-agent systems, AI governance, neuro-symbolic AI

---

[*]Corresponding author

1

1 Introduction

The field of Artificial Intelligence (AI) is undergoing a paradigm shift from the development of passive, task-specific tools toward the engineering of autonomous systems that exhibit genuine agency. Modern agentic AI systems [1, 2] are defined by capabilities such as proactive planning, contextual memory, sophisticated tool use, and the ability to adapt their behavior based on environmental feedback. These systems operate not as mere solvers but as collaborative partners, capable of dynamically perceiving complex environments, reasoning about abstract goals, and orchestrating sequences of actions—either independently or as part of a sophisticated multi-agent ecosystem [3, 4].

To establish a precise conceptual foundation, we distinguish between the field's core concepts. An *AI Agent* (or a *Single-Agent System*) is a self-contained autonomous system designed to accomplish a goal. It operates primarily in isolation, though it may interact with tools and APIs. Its agency is defined by its *autonomy*, *proactivity*, and its ability to complete a task from start to finish independently.

For example, a single, powerful *LLM-based (Large Language Model-based)* agent tasked with "Write a full project proposal for a new mobile app" would autonomously break down the task, conduct research, write the sections, and format the final document.

In contrast, *Agentic AI* is the broader field and architectural approach concerned with creating systems that exhibit agency. Crucially, this often involves the orchestration of *Multi-Agent Systems (MAS)*, where multiple specialized agents work together, coordinating and communicating to solve problems that are too complex for a single agent.

For example, an Agentic AI system designed for the same task would employ a team of specialized agents: a *Project Manager Agent* to break the goal into tasks, a *Researcher Agent* to gather market data, a *Writer Agent* to draft content, and a *Quality Assurance Agent* to review the output. Their collaborative workflow is the embodiment of Agentic AI.

In summary, one can conceptualize an *AI Agent* as a single, sophisticated worker, while *Agentic AI* represents the principle of leveraging agency, frequently by architecting and managing an entire team of such workers.

This rapid evolution, however, has led to a fragmented and often anachronistic understanding of the field. A critical issue identified in prior reviews is *conceptual retrofitting*—the misapplication of classical symbolic frameworks (e.g., Belief–Desire–Intention (BDI) [5], *perceive–plan–act–reflect (PPAR)* loops [6, 7]) to describe modern systems built on *large language models (LLMs)* [8], which operate on fundamentally different principles of stochastic generation and prompt-driven orchestration. This practice obscures the true operational mechanics of LLM-based agents [9, 10, 11, 12] and creates a false sense of continuity between incompatible architectural paradigms, whether applied to a single complex agent or a coordinated MAS.

This paper addresses these gaps by first establishing a clear historical context (Figure 1), which delineates the evolution of AI through five distinct but overlapping eras.
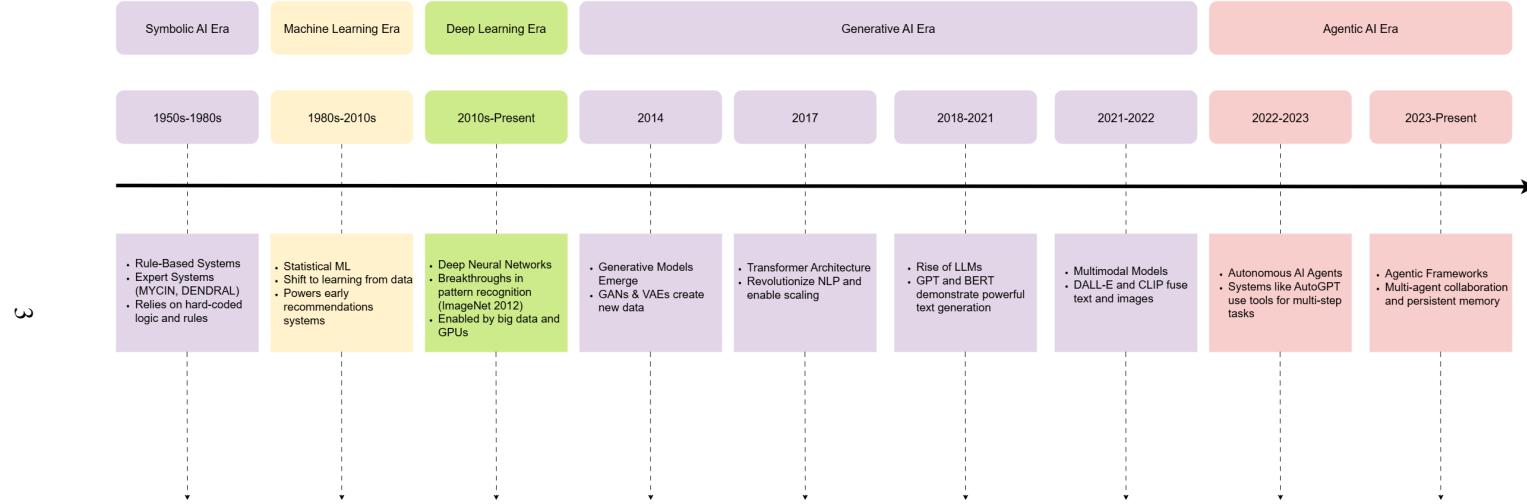
Figure 1: Historical Evolution of AI Paradigms: This timeline charts the key breakthroughs and eras in AI, from early symbolic systems to the modern agentic era. It highlights the Transformer architecture as the pivotal enabling technology for large language models (LLMs), which in turn powered the generative AI revolution and provided the substrate for contemporary agentic systems.

The **Symbolic AI Era (1950s–1980s)** [13] established the foundational ambition of artificial intelligence, grounded in logic and explicit human knowledge. This period was dominated by rule-based systems and expert systems such as MYCIN and DENDRAL [14], which operated on carefully hand-crafted symbolic rules. Intelligence was conceived as a top-down, deductive process, representing the purest form of the symbolic paradigm.

The **Machine Learning (ML) Era (1980s–2010s)** [15, 16, 17] marked a pivotal shift away from hard-coded logic toward systems that could learn from data. While still heavily dependent on human-engineered features, this period introduced statistical ML models such as Support Vector Machines and decision trees, which powered applications ranging from classification to recommendation. It was a transitional stage that moved the field away from pure symbolism but still lacked the automated feature learning that would define subsequent eras.

The arrival of the **Deep Learning Era (2010s–Present)** [18, 19, 20, 21, 22] was catalyzed by the confluence of increased compute power and large datasets. Deep neural networks, including convolutional and recurrent architectures, enabled systems to automatically learn hierarchical representations from raw data. This era revolutionized pattern recognition in vision, speech, and text, breaking longstanding barriers in perception. Yet, despite their power, these models largely functioned as sophisticated pattern classifiers rather than autonomous agents.

Out of this foundation emerged the **Generative AI Era (2014–Present)** [23, 24, 25, 26, 27], fueled by advances in generative modeling. Early breakthroughs such as Generative Adversarial Networks were soon eclipsed by the introduction of the Transformer architecture in 2017, which enabled the scaling of large language models (LLMs) such as GPT and BERT. These systems moved beyond perception to generation, producing coherent text, code, and media. In doing so, they provided the essential substrate—a powerful, general-purpose statistical reasoner—that made modern agentic AI feasible.

Finally, the **Agentic AI Era (2022–Present)** represents the current frontier, where the generative capabilities of LLMs are harnessed for action and autonomy. This era is characterized by the rise of AI agents [28, 29, 30] such as AutoGPT, which can pursue goals through planning and tool use. Increasingly, these agents evolve into multi-agent systems [31, 32, 33, 34, 35], exemplified by frameworks like CrewAI and AutoGen, where specialized roles and orchestrated collaboration enable teams of agents to tackle complex problems. In contrast to the algorithmic deliberation of the symbolic paradigm, this stage is defined by the neural paradigm, where agency emerges from the stochastic orchestration of generative models.

This chronological progression provides essential context but also reveals a critical conceptual schism. The agentic AI era is not simply a linear descendant of symbolic AI but is instead built upon a completely different architectural foundation. To address this, we introduce a novel conceptual framework (Figure 2) designed to prevent retrospective conflation by clearly distinguishing the symbolic and neural lineages of agentic AI. This dual-axis taxonomy provides the unified lens necessary to rigorously analyze the field's theoretical underpinnings, architectural innovations, and practical deployments.

The journey to modern agentic AI is best understood through its historical progression, as detailed in Figure 1. This evolution moved from the deterministic, rule-based systems of the symbolic era through the data-driven revolutions of machine learning and deep learning, culminating in the transformative advent of large language models (LLMs) [36, 37] and generative AI.

However, a chronological account is insufficient for analytical rigor. The central challenge in current discourse is the conceptual retrofitting of modern, neural agentic architectures into the frameworks of the symbolic era. To resolve this, we propose a dual-paradigm taxonomy in Figure 2. This framework categorizes agentic systems along two independent dimensions: their **Architectural Paradigm** (Symbolic vs. Neural) and their **Degree of Agency & Coordination** (Single-Agent vs. Multi-Agent). This model is designed not to show evolution, but to provide

a clear analytical structure for classification and comparison, ensuring systems are evaluated on their own operational terms.
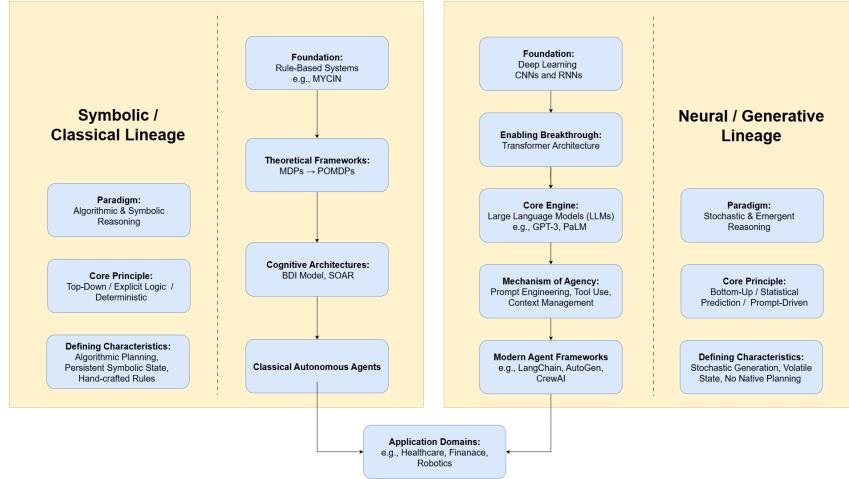

Figure 2: Conceptual Framework of Agentic AI's Dual Lineages. This taxonomy resolves conceptual retrofitting by distinguishing the Symbolic/Classical lineage (left), defined by algorithmic planning and persistent state, from the Neural/Generative lineage (right), defined by stochastic generation and prompt-driven orchestration. While both paradigms target similar applications, their underlying mechanisms are fundamentally incompatible. This framework provides the analytical structure for this survey.

This review is structured around this framework to synthesize three critically interconnected layers:

The first layer encompasses the **Theoretical Foundations**, including core principles of autonomy and agency [38], and decision-making models like Markov Decision Processes (MDPs) and Partially Observable MDPs (POMDPs) [39, 40]. It is crucial to note that these models provide a theoretical language for describing agency that originated in the *Symbolic paradigm*, but modern systems *implement* these concepts in entirely new ways.

The second layer analyzes **Architectural Frameworks**, focusing on the modern infrastructures powering the *Neural paradigm*. We examine systems like LangChain [41], AutoGen, and CrewAI, which achieve agency through mechanisms like prompt chaining, conversation orchestration, and dynamic context management—a clear departure from the symbolic planning of the classical lineage.

The third layer investigates **Application Domains**, exploring the practical deployment of agentic systems across fields such as healthcare [42], finance [43], scientific discovery [44], and legal reasoning [45]. Our framework allows us to map these applications to the appropriate paradigm and analyze their unique implementation challenges.

1.1 Current Surveys Gaps and Contributions

The current discourse on agentic AI suffers from the conceptual retrofitting illustrated in Figure 2. Classical AI frameworks, such as the BDI model or perceive–plan–act–reflect (PPAR)

loops, are often rhetorically applied but are fundamentally mismatched to the stochastic, non-symbolic, and context-driven nature of LLM-based agents [5]. Furthermore, existing reviews are often narrow in scope, lacking empirical comparisons or integrated governance insights. As summarized in Table 1, current literature leaves substantial gaps in understanding the field's current state.

Table 1: Summary of Prior Surveys on Agentic AI

Reference	Focus	Key Contributions	Limitations
Plaat et al. (2025) [8]	Agentic LLMs	Reasoning-Acting-Interacting taxonomy	Limited empirical validation; no evolutionary context
Schneider (2025) [46]	GenAI to Agentic shift	Conceptual framework for autonomy	No performance metrics; ignores architectural mechanisms
Acharya et al. (2025) [47]	Foundational methods	Combined RL with cognitive architectures	Scalability not addressed; overlooks LLM-based paradigms
Gridach et al. (2025) [44]	Scientific discovery	Tools for autonomous research workflows	No governance discussion; isolated application view
Hosseini & Seilani (2025) [48]	Enterprise strategy	Agentic design for organizational alignment	Lack of technical depth; no architectural analysis
Ozman (2025) [49]	Business operations	Systematic review methodology	Missing benchmark comparisons; no unifying framework

This review directly addresses these limitations through four integrated contributions:

1. **A Novel Dual-Paradigm Taxonomy:** We introduce and employ the framework in Figure 2 as our primary analytical tool, explicitly distinguishing symbolic and neural lineages to prevent conceptual retrofitting and enable accurate system classification.

2. **Architectural Clarification:** We demystify the operational principles of modern neural frameworks (Section 4), explaining how they achieve agency through mechanisms like prompt chaining and conversation orchestration, rather than symbolic planning.

3. **Empirical Mapping:** We conduct a systematic PRISMA-based literature review of 90 studies, categorizing them using our dual-paradigm framework to trace research trends and evaluate architectures by their appropriate standards.

4. **Governance Anchoring:** We embed ethical, accountability, and alignment challenges within each paradigm of our taxonomy to ensure that safety considerations are discussed in the correct technological context (Section 7).

1.2 Structure of the Paper

To guide the reader through our analysis, the paper is structured to logically develop the argument for a dual-paradigm understanding of Agentic AI. We begin by establishing the necessary theoretical context in Section 2, which explores the foundations of agency and introduces our core taxonomic framework. Section 3 then details the systematic methodology underpinning our literature review.

The subsequent sections apply this framework to analyze the field: Section 4 reviews key architectural frameworks through our taxonomic lens, and Section 5 examines how different application domains influence paradigm selection. Section 6 presents a comprehensive paradigm-aware taxonomy of the literature, serving as a foundational reference and key output of our review. Section 7 investigates the paradigm-specific nature of ethical and governance challenges, leading directly into Section 8, which outlines the critical research gaps identified by our analysis.

The final sections synthesize our findings and look forward. Section 9 then charts an actionable research roadmap toward hybrid intelligence, building directly upon both the identified gaps and our stated contributions. Finally, Section 10 provides a final synthesis of our findings and their implications for the field.

This structure is designed to first equip the reader with the necessary conceptual tools, then systematically analyze the landscape, and conclude by synthesizing the insights into a coherent vision for the future of Agentic AI.

2 Theoretical Foundations: Mapping the Dual Lineages of Agentic Intelligence

The architectural history of agentic AI is not a linear progression but a branching into two distinct paradigms, as defined by our conceptual framework (Figure 2). This section delineates the theoretical and cognitive groundwork for both the **Symbolic/Classical** and **Neural/Generative** lineages, clarifying their foundational principles and highlighting the paradigm shift that separates them.

2.1 Core Principles of Autonomy and Agency

The conceptual language for describing agency originated within the symbolic paradigm. The foundational constructs of *autonomy* and *agency* are essential for both lineages, though they are implemented in fundamentally different ways. Autonomy refers to a system's ability to operate independently, free from direct human intervention, whereas agency encapsulates the notion of goal-directed behavior that incorporates intention, contextual awareness, and decision-making capabilities [50, 38]. Agentic AI synthesizes these traits by initiating tasks, dynamically ranking goals, monitoring progress, and adjusting behavior through feedback loops [51].

These mechanisms parallel human executive functions such as planning, inhibition, and cognitive flexibility. They provide the high-level descriptive framework for intelligent behavior, which both symbolic and neural systems aim to achieve through divergent mechanisms.

2.2 The Symbolic Lineage: Algorithmic Decision-Making

The symbolic lineage is characterized by explicit logic, algorithmic planning, and deterministic or probabilistic models. Its evolution provides the theoretical bedrock for pre-LLM autonomous systems.

2.2.1 Markov Decision Processes (MDPs)

MDPs provide the mathematical scaffolding for modeling environments with full state information [52, 53], a hallmark of early symbolic and classical statistical AI. An MDP is defined by a tuple (S, A, P, R), representing states, actions, transition probabilities, and rewards. These systems operate effectively in deterministic, rule-based domains but lack the capacity for robust reasoning under uncertainty, anchoring them firmly in the symbolic paradigm.

2.2.2 Partially Observable MDPs (POMDPs)

POMDPs extend MDPs by introducing probabilistic *belief states* to handle environments where the agent has incomplete information [54, 55]. This was a key advancement, allowing symbolic agents to infer hidden states through observation and enabling more adaptive behavior. However, as illustrated in Figure 3, this is still a form of algorithmic state estimation. The significant computational overhead of belief tracking limits their scalability and real-world application [56, 57], a fundamental constraint of the symbolic approach.
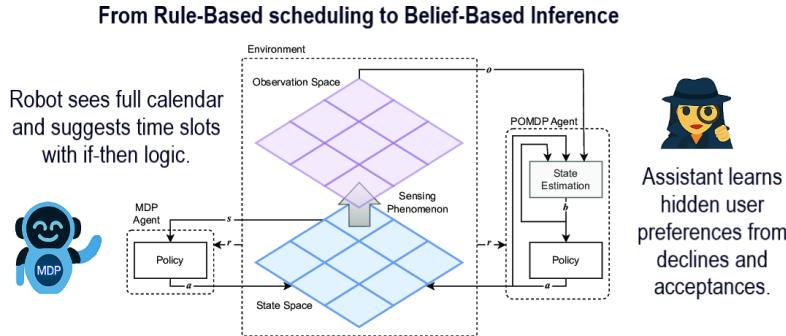

Figure 3: Classical symbolic reasoning: Comparison between a rule-based MDP scheduler (left) and a belief-based POMDP assistant (right). The MDP agent relies on explicit calendar states and deterministic policies, while the POMDP agent infers hidden user preferences from behavioral feedback. Both represent the symbolic paradigm's approach to decision-making.

2.2.3 Cognitive Architectures: BDI and SOAR

Cognitive architectures like Belief-Desire-Intention (BDI) and SOAR represent the pinnacle of the symbolic paradigm's attempt to engineer agency. They explicitly model internal states and processes, as summarized in Table 2. These systems directly implement a perceive-plan-act-reflect loop using symbolic representations, making them powerful but brittle and difficult to scale to complex, real-world environments. Their relationship to human cognitive functions is a direct, top-down mapping of symbolic logic.

Table 2: Mapping Human Cognitive Functions to Symbolic AI Modules

Component	Human Function	Symbolic AI Parallel
Belief Module	Working Memory	Symbolic Knowledge Base / World Model
Desire Module	Motivation	Goal Stack / Utility Function
Intention Module	Executive Control	Action Policy / Planner
Meta-cognition Layer	Self-reflection, Error Monitoring	Monitor / Replan Loop

2.3 The Neural Lineage: Statistical Learning and Emergent Reasoning

The neural lineage is built on a foundation of statistical learning from data, culminating in the generative capabilities of large language models (LLMs). Its progression is marked by a move away from explicit logic toward emergent, stochastic behavior.

2.3.1 Deep Reinforcement Learning (DRL)

Deep Reinforcement Learning (DRL) represents a critical transition. It scales learning to high-dimensional inputs (like images and text) using neural networks [58, 59]. DRL agents learn policies directly from data, moving away from hand-crafted symbolic rules. Methods such as PPO allow for fine-grained behavioral optimization [60, 61]. As shown in Figure 4, advancements like meta-DRL introduced generalization across tasks, a precursor to the adaptability required for modern agency. DRL is a bridge, using neural networks to learn the policies that symbolic systems would have to be explicitly programmed with.

2.3.2 The LLM Substrate and The Paradigm Shift

The emergence of Large Language Models (LLMs) was not an evolution but a revolution that created the new neural paradigm. LLMs provided a powerful, general-purpose substrate for reasoning based on statistical prediction in a high-dimensional space of concepts. This enabled a fundamental architectural shift from designing cognitive agents to orchestrating generative pipelines.
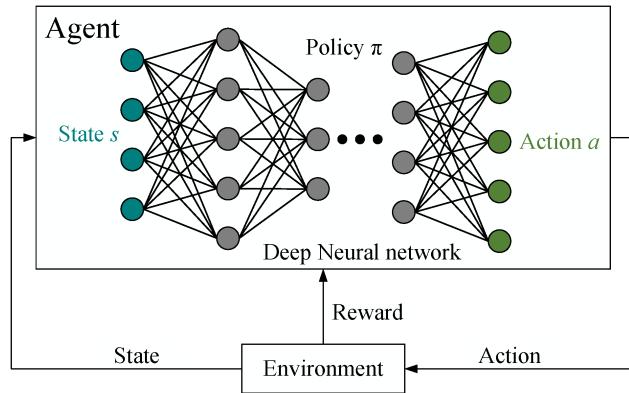
Frameworks like LangChain, AutoGen, and CrewAI do not implement symbolic PPAR loops or BDI architectures. They represent a new paradigm of **LLM Orchestration**, where pre-trained models act as central executives that coordinate tasks through fundamentally different mechanisms, as detailed in Table 3.

This shift marks the definitive break from the symbolic tradition. Agency in the neural paradigm is an emergent property of prompt-driven orchestration, not a product of internal symbolic logic. The evolution of a personal assistant, depicted in Figure 5, culminates in this new architecture.

2.4 Multi-Agent Orchestration: The Pinnacle of the Neural Paradigm

The most advanced manifestation of the neural paradigm is multi-agent orchestration. Frameworks like AutoGen [67] and LangGraph [81] coordinate diverse, modular agents through struc-

Deep Reinforcement Learning (DRL)

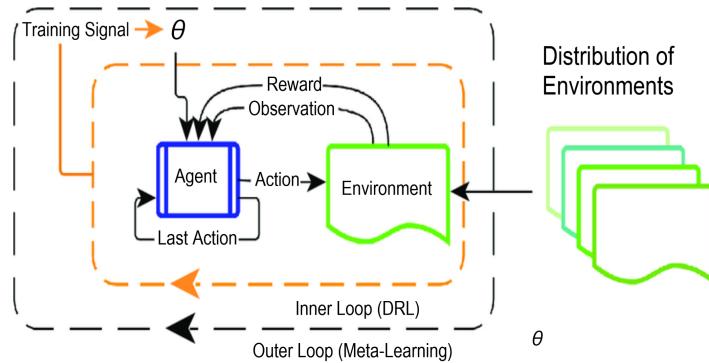
DRL with Meta-Learning

Figure 4: The shift toward learned behavior: Architectural contrast between vanilla DRL (single-task optimization) and meta-DRL (dual-loop generalization). The latter improves adaptability across tasks through meta-optimization loops, moving from explicit programming toward learned, emergent capabilities.
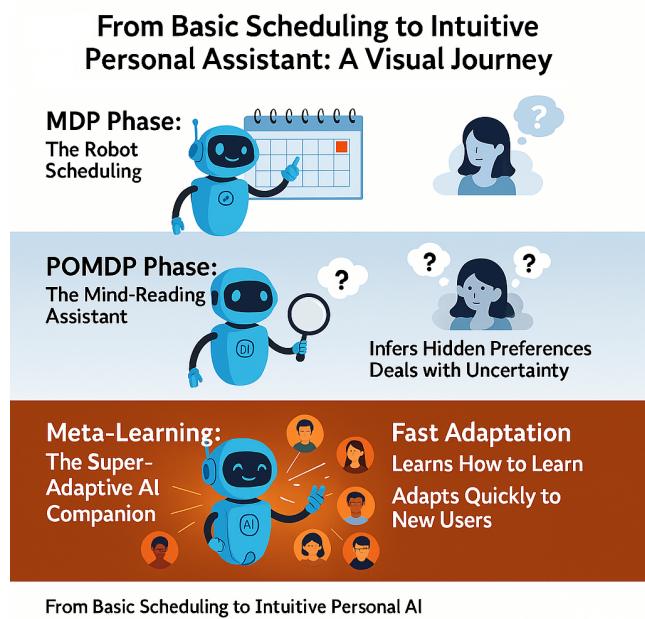
Figure 5: The journey from symbolic to neural agency: The evolution of a personal assistant from a deterministic rule-based (MDP) system, to an uncertainty-aware (POMDP) system, and finally to a modern LLM-orchestrated agent. This journey bridges the two paradigms, ending with a system that exhibits intelligent behavior through entirely different mechanisms.

Table 3: Orchestration Mechanisms of Modern Neural Agentic Frameworks

Framework	Primary Mechanism	Functional Paradigm and Representative Applications
LangChain [41, 62, 63, 64, 65]	Prompt Chaining	Orchestrates linear sequences of LLM calls and API tools. Replaces symbolic planning with stochastic generation of next steps. Applications: Multi-step workflow automations, automated medical reporting [66].
AutoGen [67, 68]	Multi-Agent Conversation	Facilitates structured dialogues between collaborative LLM agents. Replaces monolithic control with emergent problem-solving through conversation. Applications: Collaborative task solving, economic research coordination [69].
CrewAI [70, 71]	Role-Based Workflow	Assigns roles and goals to a team of agents, managing their interaction workflow. Replaces centralized scheduling with dynamic, role-driven process management. Applications: Market analysis and risk modeling [43].
Semantic Kernel [72, 73, 74]	Plugin/Function Composition	Connects LLMs to pre-written code functions ("skills"). Replaces integrated actuation with stochastic planning of plugin sequences. Applications: Breaking down high-level user intents into executable skills.
LlamaIndex [75, 76, 77, 78]	Retrieval-Augmented Generation (RAG)	Provides sophisticated data connectors and indexing. Replaces internal symbolic knowledge bases with on-demand, external context retrieval. Applications: Financial sentiment analysis [79], enhancing information retrieval for research [80].

tured communication protocols. As visualized in Figure 6, an orchestrator (often an LLM itself) acts as a context manager and task router, assessing the overall goal and dynamically assigning

specialized subtasks to other agents.

This architecture achieves scalability and complex problem-solving not through a single agent's cognitive complexity, but through the emergent intelligence of a well-orchestrated system. It is the culmination of the neural lineage, firmly establishing the new orthodoxy of LLM-driven pipelines and completing the paradigm shift from the symbolic AI tradition.
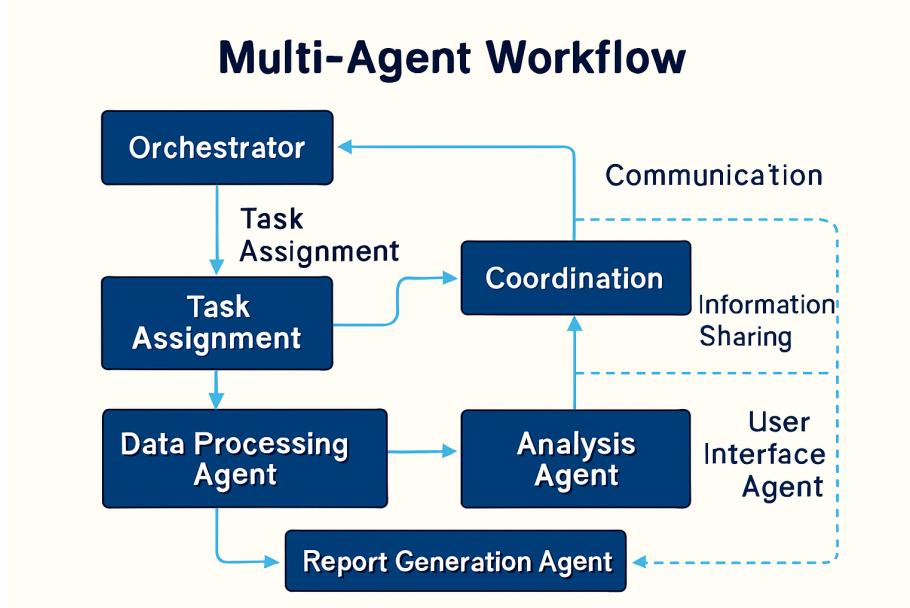

Figure 6: The architecture of the neural paradigm: Multi-Agent Orchestration in modern AI systems. This schematic illustrates the operational paradigm of neural systems. A central orchestrator (e.g., an LLM) manages a dynamic workflow of specialized agents through structured messaging and context management. Functionality emerges from prompt routing and API tool use, explicitly replacing the symbolic perceive-plan-act-reflect loop.

3   Methodology

A rigorous and transparent methodology is essential for constructing a comprehensive review that captures the dual paradigms of Agentic AI. This section outlines the systematic process used to identify, evaluate, and synthesize literature, with a specific focus on categorizing works according to the symbolic and neural lineages defined in our conceptual framework (Figure 2). It follows established review protocols to ensure reproducibility while accounting for the field's rapid evolution.

3.1   Review Design

This study adopts the **PRISMA 2020 framework** (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [82, 83], guiding all stages from search strategy to synthesis. The

methodology is designed to capture and distinguish between the symbolic/classical and neural/generative lineages of agentic AI research across computer science, cognitive psychology, robotics, and ethics.

Objectives: This systematic review aims to provide a comprehensive analysis of agentic AI systems through the following specific research objectives:

1. To identify, classify, and synthesize literature based on the dual architectural paradigms (Symbolic vs. Neural) of Agentic AI.

2. To examine the evolution of capabilities, applications, and performance metrics within and across each paradigm.

3. To analyze governance frameworks and ethical challenges, contextualizing them within their respective architectural paradigms.

4. To highlight paradigm-specific research gaps and propose informed future directions based on the synthesized evidence.

3.2 Data Sources and Search Strategy

A multi-database search strategy was employed to identify literature across both historical symbolic and modern neural agentic AI research. Sources included: IEEE Xplore, ACM Digital Library, arXiv, SpringerLink, ScienceDirect, and Google Scholar.

The search strategy employed a structured set of keyword clusters designed to comprehensively capture the core concepts associated with both architectural paradigms. To represent the **Symbolic/Classical** lineage, targeted terms included foundational concepts such as "Cognitive architectures," "BDI agent," "SOAR," "POMDP," "symbolic planning," and "multi-agent systems" (in its traditional sense). Conversely, the **Neural/Generative** paradigm was captured through terms reflecting its contemporary emergence, such as "LLM agent," "AI orchestration," "prompt chaining," "tool-augmented LLM," "multi-agent conversation," and specific framework names including "AutoGen" and "LangChain." Finally, a set of **General** terms—"Agentic AI," "autonomous agent," and "goal-directed AI"—was used to ensure broad coverage and to capture literature that might bridge or transcend the paradigmatic divide. Boolean operators were structured to optimize breadth and relevance (e.g., ("autonomous agent" OR "agentic AI") AND ("large language model" OR "orchestration" OR "cognitive architecture")).

The search scope was interdisciplinary, targeting relevant fields from computer science to ethics. To capture the most current advancements in the rapidly evolving neural paradigm, the search included pre-print servers like arXiv, with these records being manually assessed for quality and relevance.

3.3 Inclusion and Exclusion Criteria

To ensure the review's methodological integrity and thematic relevance, predefined inclusion and exclusion parameters were applied during the screening process. These criteria were designed to capture high-quality literature from both paradigms of agentic AI.

Inclusion Criteria The literature search employed the following inclusion criteria to identify publications that contribute directly to the core themes of agentic AI architectures and applications. Specifically, we included peer-reviewed journal articles, conference proceedings, and formally published technical reports from recognized institutions. To capture the most recent

advancements in the rapidly evolving neural paradigm, we also incorporated high-impact pre-prints from arXiv, which were manually screened for methodological rigor and citation impact, with a focus on those presenting novel architectures or frameworks. The scope of included work encompassed studies involving the design, implementation, or evaluation of autonomous agents, spanning both classical symbolic systems and modern LLM-orchestrated frameworks. All selected publications were required to be in English and published within the temporal window of January 2018 to March 2025.

Exclusion Criteria   To ensure a focused and methodologically rigorous review, studies were excluded according to the following criteria. Non-English language publications were omitted. We also excluded non-peer-reviewed or informal sources such as opinion pieces, editorials, blog posts, and unverified online content. Furthermore, studies focused exclusively on generative AI (e.g., for image generation or text completion) without incorporating agentic features like goal-directedness, tool use, or multi-step autonomy were deemed out of scope. Finally, duplicate records retrieved from multiple databases were identified and removed to prevent redundancy in the analysis.

These criteria ensured the retention of conceptually aligned and methodologically sound studies from both paradigms, preserving the review's comprehensive scope. A summary is provided in Table 4.

Table 4: Inclusion and Exclusion Criteria for Literature Selection

Category	Criteria
Inclusion	
	• Peer-reviewed journal and conference papers
	• Technical reports from reputable institutions
	• Studies on autonomous agents from both symbolic and neural paradigms
	• Applications across various domains demonstrating agentic capabilities
	• Published in English between 2018 and 2025
Exclusion	
	• Non-English publications
	• Blogs, opinion pieces, or informal content
	• Studies focused solely on generative AI without agentic autonomy
	• Duplicate records across multiple databases

3.4   Screening and Selection Process

The screening protocol adhered to the PRISMA 2020 guidelines to ensure methodological transparency and reproducibility. Records were compiled from selected databases, yielding an initial pool of 165 items (157 from databases, 8 from supplemental sources).

Following deduplication, 120 unique records remained. Title and abstract screening excluded 42 studies due to irrelevance or insufficient focus on agentic AI. Full-text assessment confirmed 78 articles met all inclusion criteria.

In alignment with PRISMA's guidance for systematic reviews that require foundational context, a supplemental phase was conducted [82]. During thematic synthesis, 12 seminal theoretical papers from the symbolic paradigm (e.g., foundational works on MDPs by [39] and cognitive architectures by [84]) were incorporated. These papers were essential for providing complete historical context for the taxonomic framework and understanding the symbolic lineage, though they were analyzed separately from contemporary neural paradigm research. This resulted in a final corpus of **90 publications** for contextual and theoretical grounding, with 78 studies forming the core for analysis of contemporary trends.

The process is illustrated in Figure 7, which clearly distinguishes the primary systematic search from the supplemental inclusion of foundational context.

3.5   Data Analysis

The 78 studies forming the core of the review underwent thematic synthesis following the methodology described by Thomas and Harden [85], with analysis specifically structured around the dual-paradigm framework.

Key Analytical Techniques: Our analysis employed a multi-faceted methodological approach to systematically investigate the body of research. The initial phase involved **paradigm classification**, whereby each study was categorized according to its primary architectural paradigm—either Symbolic/Classical or Neural/Generative—based on the core operational mechanisms defined in our conceptual framework. Following this classification, we conducted a detailed **framework mapping** within each paradigm to group studies by their specific architectural approaches, including orchestration models (e.g., AutoGen, CrewAI), memory structures, and learning mechanisms. Building on this organized foundation, a **cross-paradigm comparison** was performed to identify fundamental differences in implementation, performance, and limitations between the two overarching paradigms. In parallel, we performed **domain clustering** to group applications by sector—such as healthcare, finance, robotics, and scientific discovery—which enabled the identification of performance patterns and deployment strategies both within and across paradigms. Finally, an **ethical coding** procedure was applied, using a structured lexicon to tag recurring themes related to governance, safety, transparency, and bias, with particular attention paid to how these ethical challenges manifest differently within each paradigm.

Qualitative coding was supported by tools such as NVivo [86], which enabled hierarchical theme identification and cross-paradigm analysis. Quantitative results were tabulated and compared within and across domains and paradigms to synthesize technical and operational insights.

This paradigm-informed approach ensured a nuanced understanding of the current landscape of Agentic AI research, supporting both theoretical grounding and real-world applicability while maintaining the analytical rigor required for this review.
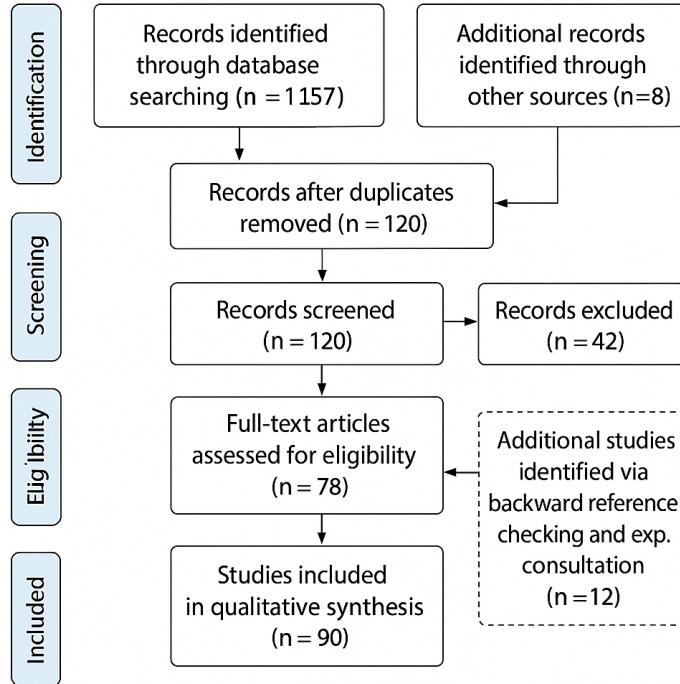
Figure 7: PRISMA 2020 Flow Diagram. Records were identified from databases (n=157) and supplemental sources (n=8). After deduplication (n=120) and title/abstract screening (n=42 excluded), full-text review confirmed 78 eligible studies. A supplemental phase added 12 seminal theoretical papers for contextual framing of the symbolic paradigm (shown in dashed box), yielding a final corpus of 90 publications for the review.

3.6 Limitations

Limitations While this review provides a comprehensive synthesis of Agentic AI research, several limitations must be acknowledged. First, the inherent **temporal and scope dynamics** of the field, particularly within the rapidly evolving neural paradigm, present a challenge; although our search extended to early 2025, some very recent developments may not be captured, a risk mitigated but not fully eliminated by the inclusion of pre-prints. Furthermore, our methodological approach required a **contextual reference expansion** through the supplemental inclusion of 12 seminal symbolic papers to ensure a robust theoretical framing of the classical lineage. We emphasize that these papers, analyzed separately from contemporary research, were used strictly for contextual and historical background and represent a deviation from a purely systematic retrieval process.

Additional constraints arose from the nature of the subject matter itself. **Transparency constraints** were encountered as many state-of-the-art neural agentic systems operate as proprietary solutions with limited public documentation, meaning architectural details and performance met-

rics were sometimes incomplete or inferred from secondary sources. **Methodological hetero-geneity** across the reviewed studies, with their varied evaluation metrics, also limited our ability to perform direct cross-study benchmarking, particularly between paradigms that employ fundamentally different performance measures. Finally, despite implementing rigorous classification criteria, the **paradigm classification challenge** of assigning hybrid or transitional architectures to a single paradigm may, in some cases, involve necessary simplification.

These limitations collectively highlight the challenges of conducting systematic reviews in a nascent and fast-paced field with multiple co-existing paradigms. Our two-phase approach—a systematic review of contemporary research supplemented by a narrative inclusion of foundational symbolic context—was designed to balance methodological rigor with comprehensiveness while respecting the fundamental distinctions between these architectural paradigms.

4 Literature Review: A Dual-Paradigm Analysis

The rapid expansion of Agentic AI has produced a diverse yet fragmented body of research. This section synthesizes the extant literature through the lens of the dual-paradigm framework introduced in Figure 2, analyzing how contributions are distributed across the symbolic/classical and neural/generative lineages. We organize and analyze the most influential contributions across foundational studies, architectural frameworks, and domain-specific applications, focusing on their operational mechanisms to clearly delineate the paradigm shift.

4.1 Foundational Studies: The Roots of Two Lineages

The theoretical bedrock of Agentic AI is found in two distinct lineages, each with its own foundational breakthroughs. Landmark studies have shaped the conceptual and architectural foundations of both paradigms, spanning strategic reasoning, cognitive models, and alignment.

These studies collectively mark the progression from explicit, algorithmic deliberation to emergent, stochastic intelligence. They serve as reference points for the fundamental differences in how adaptability, coordination, and strategic reasoning are implemented in each paradigm, illustrating the conceptual divide captured by our framework.

4.2 Architectural Paradigms: A Mechanistic Comparison

The advent of large language models (LLMs) has solidified the neural/generative paradigm, which operates on principles fundamentally incompatible with its symbolic predecessor. Modern agentic frameworks leverage LLMs as generative engines within software pipelines, explicitly departing from classical cognitive loops. Their core innovation lies in dynamic context management, prompt engineering, and tool composition.

This analysis underscores that these frameworks form the backbone of the neural paradigm, designed for practical task completion through orchestration, not for simulating internal cognitive processes. Mapping them to PPAR or BDI obscures their true innovative mechanics, which are defined by prompt-driven stochasticity, not algorithmic symbol manipulation.

4.3 Domain-Specific Implementations: A Paradigm-Driven Analysis

Agentic AI frameworks are being deployed across sectors where autonomy and adaptability are essential. The choice of paradigm is critically influenced by domain-specific constraints—ethical,

regulatory, or epistemic. The following implementations exemplify how each paradigm is applied.

Domain-Specific Applications and Paradigm Choices   The application of Agentic AI reveals a distinct paradigm split influenced by the core requirements of each sector. In **healthcare, where safety and compliance are paramount**, applications diverge clearly along architectural lines. Symbolic systems, such as rule-based clinical decision support tools, are predominantly employed for predictable and auditable tasks. In contrast, the flexibility of neural paradigms is leveraged for tasks like generating structured medical reports [66] and powering on-premise edge agents [87]; however, these neural frameworks are often contained within deterministic tool-chaining pipelines to ensure the reliability required in clinical settings.

This pattern of complementary paradigm use is also evident in **finance, a domain demanding high accuracy and auditability**. Here, neural frameworks dominate tasks involving complex data synthesis and analysis. For instance, CrewAI's role-based workflow is applied to market analysis [43] as it provides a clear, auditable trail of agent actions. Similarly, LlamaIndex-powered models for financial sentiment [79] demonstrate how neural systems use Retrieval-Augmented Generation (RAG) to ground their stochastic outputs in verified data, thereby reducing hallucination. Despite this, symbolic systems maintain a critical role in high-frequency trading and core regulatory logic where absolute determinism is non-negotiable.

Finally, in **scientific research, which requires profound epistemic rigor**, the choice of paradigm is dictated by the nature of the intellectual task. The deployment of AutoGen to coordinate multi-agent conversations for economic research [69] exemplifies the neural paradigm's strength in simulating collaborative, exploratory discovery and critique. This stands in direct contrast to the role of symbolic systems, which remain the bedrock for theorem proving and logical inference, highlighting a fundamental architectural choice between exploratory generation and deductive reasoning.

These implementations demonstrate that the paradigm choice is not merely technical but is decisively shaped by domain-specific needs, validating the need for a clear taxonomic framework to classify and select appropriate architectures.

4.4   Emerging Trends: Toward Hybrid Architectures

The evolution of Agentic AI is increasingly characterized by a deliberate synthesis of architectural paradigms, moving beyond isolated approaches toward integrated systems that combine strengths while mitigating inherent limitations. This shift toward hybrid architectures represents the field's maturation as it seeks to balance adaptability with reliability. Importantly, these trends are not broad truisms about any generation of AI, but rather specific architectural responses to challenges uniquely faced by large-scale, agentic systems.

The most significant emerging trend is **neuro-symbolic integration**, which aims to formally bridge the reliable, deterministic reasoning of symbolic systems with the adaptive, generative capabilities of neural networks [88]. This effort transcends the well-documented limitations of both paradigms, potentially establishing a new hybrid category that leverages their complementary strengths.

A second and particularly distinctive direction is the exploration of **decentralized agent networks**. Here, blockchain-based coordination mechanisms are applied to multi-agent AI systems to provide verifiable governance, transparent decision-making, and resilient autonomy [89]. Unlike conventional centralized orchestrators, distributed consensus frameworks offer robustness against single points of failure, while also opening the possibility of economic coordination

between heterogeneous agents through tokenized incentives. This line of research directly addresses questions of trust, accountability, and cooperative alignment—issues that become acute when scaling agentic AI across organizations or societal infrastructures.

Complementing these architectural innovations, advances in **lifelong learning** frameworks address a critical limitation of current LLM-based agents—their largely stateless nature [90]. By enabling continuous adaptation and durable knowledge retention, this trend effectively injects persistent memory, a concept foundational to symbolic AI, into neural architectures. This supports more context-aware, long-term, and resilient operation in dynamic environments.

Collectively, these emerging trends signal the field's progression from debating paradigm superiority to architecting sophisticated hybrids. Far from generic insights, they constitute targeted responses to enduring limitations in current Agentic AI systems: brittle reasoning, centralized governance bottlenecks, and memory deficiencies. The resulting synthesis offers the most promising path toward developing Agentic AI systems that are simultaneously adaptable and reliable, creative and verifiable—capable of operating effectively in the complex, dynamic environments that characterize real-world applications.

4.5 Coordination Protocols: From Algorithmic Contracts to Emergent Conversation

A critical yet often underexplored aspect of Multi-Agent Systems (MAS) is the fundamental distinction in their coordination mechanisms. A deeper examination reveals that these strategies are a primary differentiator between the two paradigms, reflecting their core architectural principles: **explicit algorithms** in the symbolic paradigm versus **emergent, stochastically-guided behavior** in the neural paradigm.

Within the **Symbolic Paradigm**, coordination is achieved through pre-defined, algorithmic protocols rooted in decades of distributed AI research. These protocols are engineered to ensure predictable, verifiable, and fault-tolerant interactions, making them indispensable for critical systems where correctness is paramount. A quintessential example is the **Contract Net Protocol (CNP)** [91], a classic negotiation framework where a manager agent announces a task through a "call for proposals." Other agents then evaluate their capabilities and submit bids, leading the manager to award the contract to the most suitable agent. This process, analogous to an auction, is extensively applied in domains like manufacturing and logistics scheduling. Another foundational strategy is the **Blackboard System** [92], where a shared memory space acts as a central coordination point. Specialist agents, akin to experts surrounding a physical blackboard, monitor this space for relevant data and contribute their expertise incrementally to build towards a solution. This approach is highly effective for complex, unstructured problems like medical diagnosis or signal interpretation. Furthermore, **Market-Based Approaches** facilitate coordination through a virtual economy where agents buy and sell services or resources, providing a decentralized method for resource allocation in networked systems.

In direct opposition, coordination within the **Neural Paradigm** is not typically governed by hard-coded protocols. Instead, it emerges as a property of **structured conversation and prompt-driven orchestration** [93, 94, 95]. Here, a central orchestrator (often an LLM itself) or the agents themselves leverage their generative capabilities to dynamically assign roles, manage dialogue, and synthesize results. This can manifest in several distinct patterns. **Conversation-Based Coordination** [96, 97, 98], exemplified by frameworks like AutoGen, achieves collaboration through structured conversational loops where agents with defined roles interact within a group chat, with the LLM's context window managing the interaction state. A more explicit variant is the **Role-Based Workflow** [99] (e.g., CrewAI), where a higher-level orchestrator assigns tasks based on pre-defined roles and goals, though the routing decisions are still driven by LLM-

based reasoning rather than deterministic algorithms. Lastly, **Dynamic Context Management** [100, 81] (e.g., LangGraph) implements coordination through state machines that control information flow between nodes; the graph structure defines possible paths, but the specific execution is determined stochastically by the LLM's output at each step.

The fundamental dichotomy between these coordination strategies is summarized in Table 5, which highlights the core operational differences.

Table 5: A Dual-Paradigm Comparison of Multi-Agent Coordination Mechanisms

Feature	Symbolic/Classical Paradigm	Neural/Generative Paradigm
Primary Mechanism	Algorithmic Protocols (e.g., Contract Net, Blackboard)	Structured Conversation & Prompt Orchestration
State Management	Explicit, often centralized (e.g., Manager in CNP, Blackboard)	Implicit, managed within the LLM's context window
Decision Process	Deterministic or probabilistic based on explicit rules	Stochastic generation of next action/response
Flexibility	Low; protocols are fixed and designed for anticipated scenarios	High; can adapt to novel coordination patterns not explicitly programmed
Verifiability	High; the protocol's logic can be formally verified and audited	Low; the emergent coordination path is opaque and difficult to trace
Key Frameworks	JADE, JaCaMo, early SOAR systems	AutoGen, CrewAI, LangGraph
Example	A manager agent uses CNP to auction a delivery task to the lowest-bidding drone agent.	An orchestrator LLM manages a conversation between a programmer agent, a tester agent, and a writer agent to collaboratively build software.

This analysis confirms that the paradigm shift extends to the very fabric of multi-agent coordination. The symbolic paradigm offers **verifiable reliability** through rigorously engineered protocols, while the neural paradigm offers **adaptable emergence** through learned conversation patterns. This critical distinction is essential for understanding the capabilities, risks, and appropriate applications of modern MAS, thereby further validating the necessity of the dual-paradigm framework presented in this survey.

4.6 Evaluating Agency: Beyond Accuracy

The evaluation of Agentic AI systems presents a fundamental challenge that distinguishes it from the assessment of traditional AI models. As the reviewer rightly notes, simple metrics like accuracy are wholly insufficient. Measuring "agency" requires quantifying a system's capacity for sustained, goal-directed behavior in dynamic environments, necessitating a multi-dimensional evaluation framework that accounts for paradigm-specific mechanisms of action.

The core challenge lies in the fact that agency is not a monolithic property but a spectrum encompassing **autonomy**, **task success**, **efficiency**, and **robustness**. Consequently, evaluation must be tailored to the architectural paradigm.

In the **Symbolic Paradigm**, evaluation has historically focused on **verifiability**. Key metrics include *Goal Completion Fidelity*, which measures the percentage of pre-defined sub-goals correctly achieved in a plan, and *Plan Optimality*, which compares the cost (e.g., time, steps) of an agent's generated plan against a known optimal solution. Furthermore, assessment involves verifying *Logical Soundness* through formal methods to ensure rule sets cannot derive contradictory or unsafe actions, and rigorously testing *Edge Case Handling* against rare but critical scenarios either explicitly encoded in or missing from the agent's knowledge base.

Conversely, in the **Neural Paradigm**, evaluation is inherently more complex due to inherent stochasticity. While benchmarks like AgentBench [101] and GAIA [**?**] represent a shift towards holistic assessment, they have limitations. Metrics must be designed to capture emergent capabilities and failures. This includes evaluating *Long-Horizon Task Success* on complex, multi-step tasks (e.g., "research a topic and write a report with citations"), often measured by final outcome quality as judged by humans or a powerful LLM "judge." Other critical dimensions are *Context Window and Memory Management*, which assess an agent's ability to utilize information across extended interactions; *Tool Use Proficiency*, encompassing tool selection accuracy, call sequence efficiency, and error recovery; *Robustness to Prompts*, testing consistency across instruction rephrasings and resilience to injection attacks; and practical *Cost and Latency* metrics, measuring computational expense (e.g., total tokens, API calls) and time-to-completion, which are crucial for real-world deployment.

A comprehensive evaluation framework for Agentic AI must therefore integrate these dimensions. It is not enough for an agent to eventually succeed at a task; it must do so efficiently, reliably, and in a manner that is transparent and auditable where required. This typically involves a synergistic combination of automated metrics (e.g., success rate, number of steps), human evaluation for qualitative judgment of output coherence and usefulness, and adversarial testing (e.g., "red teaming") to probe for specific failure modes like hallucination or goal divergence.

This paradigm-aware approach to evaluation—where symbolic systems are judged on verifiability and neural systems on robust adaptability—is essential for the responsible development and deployment of autonomous agents. It moves the field beyond simple benchmarks towards a more nuanced understanding of what it means for an AI system to be truly "agentic."

4.7   Summary of Insights

Synthesizing the literature through our dual-paradigm framework reveals several fundamental distinctions and clear trajectories for the field of Agentic AI. The analysis demonstrates that **paradigm divergence is fundamental**; rather than representing evolutionary stages, the symbolic and neural lineages constitute parallel development paths characterized by fundamentally different operational mechanics—algorithmic reasoning versus stochastic orchestration. This architectural divergence emerges as the most critical factor in determining any agentic system's inherent capabilities and limitations.

This division naturally leads to the principle that **mechanism determines application**. The choice between paradigms is far from arbitrary but is instead dictated by domain requirements. Symbolic architectures demonstrate particular excellence in domains demanding absolute reliability, verifiability, and safety, such as core regulatory systems and safety-critical controls. Conversely, neural architectures thrive in environments requiring adaptability, sophisticated pattern recognition, and operation on unstructured data, exemplified by creative research applications and complex customer interactions.

Looking toward the future, the evidence indicates that the **frontier lies in hybridization**. Emerging research trends do not suggest the ultimate victory of one paradigm over the other but rather point toward their strategic integration. The next significant advancement will likely emerge from hybrid architectures that embed symbolic reasoning modules within neural orchestration frameworks, effectively mitigating the weaknesses of pure neural approaches—such as hallucination and lack of verifiability—while preserving their adaptive strengths.

Collectively, these insights, structured by the dual-paradigm framework, provide a cohesive and accurate narrative for understanding the field's present state and future direction. This approach moves beyond a simple catalog of technologies to establish a coherent theory of architectural design in Agentic AI, offering researchers and practitioners a principled foundation for system development and evaluation.

5 Analysis of Domain-Specific Applications

Agentic AI systems have transitioned from theoretical research to critical production deployments. This section analyzes these deployments through the lens of our dual-paradigm framework, examining how domain-specific constraints—such as safety, regulation, and real-world interaction—dictate the choice of architectural paradigm and shape implementation priorities. The progression from automation to autonomy is not a function of evolutionary stage but of selecting the appropriate paradigm for the task's constraints.

To provide a structured analysis, Table 6 maps key domains against their dominant architectural paradigm, primary constraints, and illustrative implementations, creating a comparative schema based on mechanistic choice rather than chronological progression.

The diversity of these deployments reflects a key insight: the architectural paradigm is a strategic response to domain-specific pressures. For instance, healthcare applications heavily favor symbolic or highly constrained deterministic approaches. This prioritizes safety, accuracy, and auditability—a necessity in high-stakes, regulated environments—over the generative flexibility of pure neural systems.

Conversely, domains like education leverage the neural paradigm for its core strength: generating adaptive, personalized, and context-aware interactions that are difficult to pre-program with symbolic rules.

Finance and Legal applications demonstrate a crucial middle ground: they are built on neural orchestration frameworks but are heavily constrained by symbolic mechanisms (e.g., role-based workflows, rigorous retrieval from verified sources) to mitigate the risks of hallucination and ensure compliance. Robotics presents the most explicit hybrid model, pairing symbolic systems for safety-critical low-level control with neural systems for high-level coordination and adaptation.

Furthermore, this paradigm-driven analysis reveals critical cross-domain challenges that must be addressed in future research. Chief among these is the need for **paradigm-specific governance frameworks**. The operationalization of agentic systems requires tailored policy approaches that account for each paradigm's distinct risks: governing symbolic systems involves verifying their logical structures, while governing neural systems necessitates auditing training data, prompts, and outputs for stochastic failures—a challenge further compounded in hybrid architectures.

Equally critical are the emerging challenges in **security and resilience**. As these systems become integrated into critical infrastructure, they represent prime targets for adversarial attacks, though the attack vectors differ fundamentally by paradigm. Symbolic systems face exploitation of logical flaws and rule manipulation, while neural systems remain vulnerable to prompt injection, data poisoning, and other inference-time attacks that exploit their stochastic nature.

Finally, the paradigm divide fundamentally shapes **human-AI collaboration**. Effective interface design must account for these architectural differences: interacting with symbolic systems requires understanding their internal logic and state representations, whereas engaging with neural systems involves carefully steering context and interpreting often opaque, generative outputs—requiring distinct approaches to oversight and interpretability.

5.1 Tool Use and Capabilities: Integration with Real-World Systems

A critical capability that distinguishes agentic AI from passive models is their ability to interact with and manipulate external tools and data sources via Application Programming Interfaces (APIs) [109, 110, 111, 112]. This functionality is the bridge between an agent's internal reasoning and tangible action in the real world. The nature of this integration is, as our framework predicts, paradigm-dependent.

In the **Symbolic Paradigm**, tool use is typically hard-coded and deterministic. Agents call specific functions with predefined parameters based on explicit logical rules. This is prevalent in safety-critical domains like healthcare, where agents interact with Electronic Health Record (EHR) systems using strict, auditable APIs (e.g., HL7 FHIR standards for reading/writing patient data) or clinical decision support tools with fixed input-output schemas [113, 114].

In the **Neural Paradigm**, tool use is orchestrated and generative. Frameworks like LangChain and AutoGen use the LLM's ability to understand natural language instructions to dynamically select and call appropriate tools from a suite of available options. The LLM [115, 116]generates the API call parameters (e.g., formulating a database query, crafting a search query) based on its context, which is then executed by the framework. This allows for immense flexibility but introduces risks of malformed calls or unexpected outputs.

Table 7 [117, 118, 119] provides a non-exhaustive overview of the types of real-world tools and APIs that agentic systems are currently being integrated with, categorized by their primary domain and function.

This integration enables agents to move beyond text generation to become truly functional systems. For instance, a neural agent using AutoGen could read an email via the Outlook API, extract key tasks, write code to solve them using a Python tool, and then post the results to a Slack channel—all within a single orchestrated workflow. Conversely, a symbolic agent in a manufacturing context might reliably call a single, well-defined API to adjust a machine's parameters based on its rigid internal state model.

In conclusion, agentic AI is not a monolithic force but a set of distinct architectural paradigms. Its embedding into the fabric of critical systems is a story of domain-driven design, where theoretical capabilities are shaped and constrained by practical, ethical, and operational realities. The choice between symbolic, neural, or hybrid design is the primary engineering decision, making the governance and safety challenges discussed in the next section immediate and paradigm-specific imperatives.

6 Comprehensive Taxonomy of Agentic AI Literature: A Paradigm-Aware Analysis

The accelerating pace of innovation in agentic AI necessitates a systematic organization that reflects its fundamental architectural schism. This section provides a paradigm-aware synthesis of the field, serving as the culminating evidence for our dual-lineage framework:

- A **visual taxonomy** (Figure 8) categorizing the field's core dimensions through the lens of symbolic and neural mechanisms.

- A **structured literature map** (Table 8) analyzing all 90 studies from our systematic review, now classified by their primary architectural paradigm.

Table 6: Analysis of Agentic AI Deployment Patterns by Domain and Paradigm

Domain	Dominant Paradigm	Primary Constraints & Drivers	Representative Implementation & Insight
Healthcare	Symbolic / Deterministic	Safety, Privacy (HIPAA), Explainability, High Reliability	*MEDITECH's AI-infused EHR* [102] uses deterministic, auditable pipelines for clinical assistance, prioritizing predictable, rule-based tool use over emergent neural behavior to ensure patient safety and regulatory compliance. This exemplifies the symbolic paradigm's strength in high-stakes, verifiable environments.
Finance	Neural / Orchestration	Real-time throughput, Auditability, Regulatory Compliance, Fraud Pattern Dynamics	*Mastercard Decision Intelligence Pro* [103] employs orchestrated neural agent swarms to analyze transactions. Role-based systems (e.g., CrewAI) enable specialized agents for pattern detection and reporting. The focus is on scaling complex analysis, a strength of the neural paradigm, while layering in symbolic checks for auditability.
Robotics & Manufacturing	Hybrid (Symbolic + Neural)	Physical safety, Real-time response, Embodiment	*Amazon Prime Air* [104] uses symbolic POMDPs for reliable, safe navigation under uncertainty. *Siemens Smart Factories* [105] layer neural orchestration frameworks over these low-level symbolic planners to coordinate units. This hybrid model leverages the reliability of symbolism for safety-critical functions and the flexibility of neural systems for coordination.
Education	Neural / Conversational	Personalization, Pedagogical Efficacy, Student Engagement	*Duolingo Smart Bot* [106] and *Carnegie LiveHint AI* [107] utilize fine-tuned LLMs in a single-agent paradigm. Their focus is on generating adaptive, context-aware interactions, a core capability of the neural paradigm, rather than on deterministic, rule-based tutoring.
Legal & Compliance	Neural (RAG-Heavy)	Precision, Comprehensiveness, Jurisdictional nuance, Hallucination mitigation	*JPMorgan COiN* [108] and *Thomson Reuters AI* [45] rely heavily on LlamaIndex-style retrieval to ground contract analysis in vast legal corpora. This uses the neural paradigm's strength in processing unstructured data but constrains its stochasticity with symbolic-like retrieval of verified facts to ensure accuracy.

Table 7: Examples of Real-World Tools and APIs Integrated with Agentic AI Systems

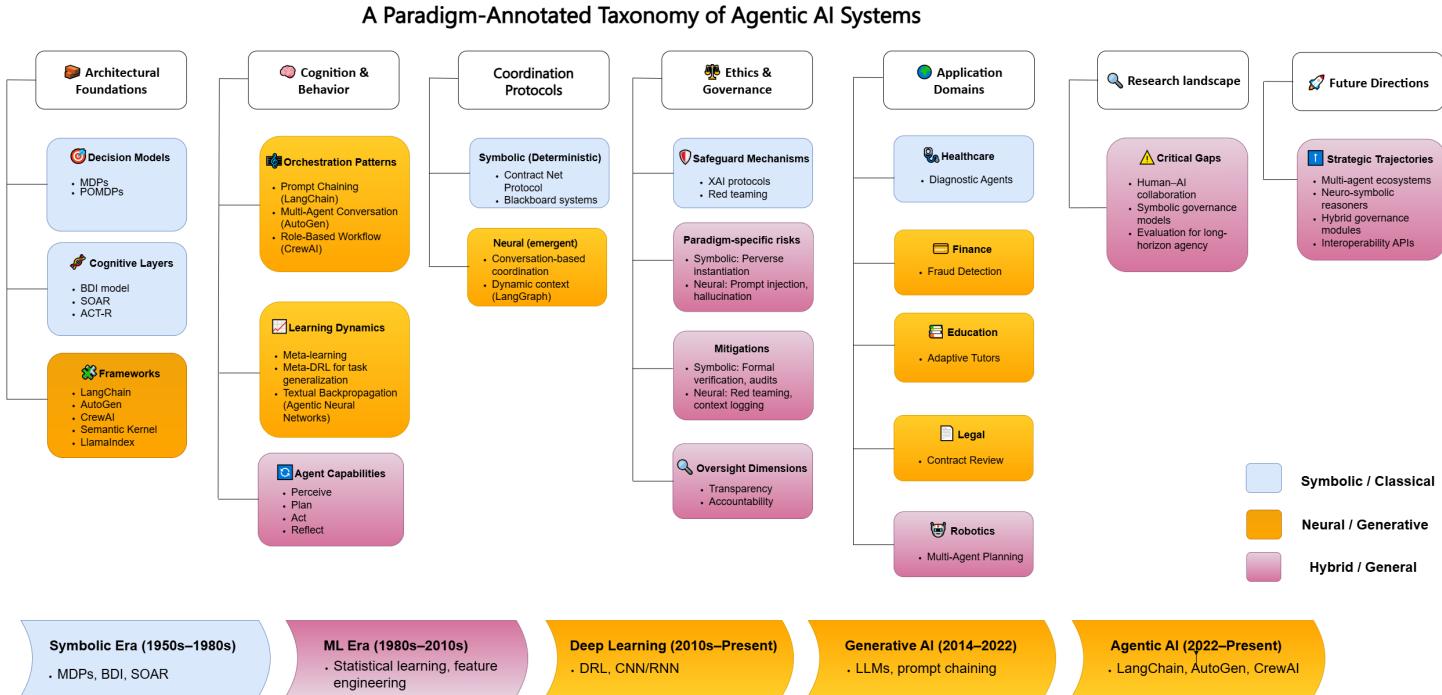| Tool/API Category	Example Services/APIs	Primary Paradigm	Agent Function & Use Case
Data & Database	Google BigQuery, Snowflake, PostgreSQL, Airtable API, Apache Cassandra	Both (Deterministic vs. Generated queries)	Querying structured data for information retrieval and analysis (e.g., financial records, customer data).
Web & Search	Google Search API, SerpApi, Wikipedia API, Wolfram Alpha API, Brave Search API	Neural	Gathering real-time, external information to ground responses and overcome LLM knowledge cut-offs.
Software & Cloud	GitHub API, AWS S3/SageMaker API, Azure Functions API, Google Cloud Compute API, Docker Engine API	Neural	Automating developer workflows, managing cloud infrastructure, and deploying machine learning models.
Business & Productivity	Slack API, Microsoft Graph (Teams, Outlook), Salesforce REST API, Jira Cloud API, Zoom API	Neural	Automating workflows, summarizing communications, managing customer relationships, and tracking tasks.
Financial	Bloomberg Terminal API (BQL), Stripe API, Plaid API, Alpaca Markets API, Reuters Eikon API	Both	Executing trades, analyzing market data, processing payments, and conducting risk assessments.
Scientific & Academic	PubMed E-Utilities API, IEEE Xplore API, UniProt API, RDKit (Cheminformatics), PyMol	Hybrid	Conducting literature reviews, generating hypotheses, and automating steps in scientific discovery pipelines.
Code Execution	Python subprocess/REPL, Node.js runtime, Docker API, Jupyter Kernel Gateway API	Neural	Writing, executing, and debugging code to perform calculations, data analysis, or solve problems.

Figure 8: A Paradigm-Annotated Taxonomy of Agentic AI Systems. This framework organizes the field's core components, now visually differentiated by architectural paradigm: **Symbolic/Classical** (blue), **Neural/Generative** (orange), and **Hybrid/General** (purple). The taxonomy reveals how the symbolic paradigm underpins formal decision models and cognitive architectures, while the neural paradigm defines modern frameworks and orchestration patterns. Application domains are colored by their dominant paradigm, illustrating the strategic choice between symbolic safety and neural adaptability. This visualization provides a clear roadmap for navigating the distinct design, governance, and implementation pathways required by each architectural lineage.

Our paradigm-aware analysis of the complete corpus reveals key patterns that were previously obscured (see Table 8):

1. **Paradigm Specialization by Domain**: High-stakes, regulated domains like Healthcare and Legal Tech show a strong preference for symbolic or highly constrained neural architectures (e.g., [42, 45]), while dynamic domains like Finance leverage neural orchestration for complex analysis (e.g., [43]).

2. **The Governance Divide**: Research in Ethics & Governance is overwhelmingly focused on the novel challenges of the neural paradigm (e.g., [9, 120]), revealing a significant gap in modernized governance frameworks for purely symbolic systems.

3. **Temporal Paradigm Shift**: The data shows a clear transition: symbolic and hybrid Cognitive Architectures dominated early research (2018–2021), while neural Orchestration Frameworks have overwhelmingly dominated post-2022, following the rise of LLMs.

Table 8: Paradigm-Based Taxonomy of Agentic AI Literature (2018–2025)

Category	Paradigm	Key Papers	Year	Focus Area	Key Contributions
Foundational Theories	Hybrid	[8, 46, 47, 44, 48, 49]	2025	Autonomy frameworks	Theoretical foundations bridging symbolic and neural concepts of agency
Architectural Frameworks	Neural	[67, 70, 71, 72, 73, 74, 75, 41, 62, 63, 64, 65, 68]	2023-2025	System design	Neural-based multi-agent orchestration, tool integration, and workflow management
Healthcare Applications	Symbolic / Hybrid	[42, 66, 87, 121]	2023-2024	Medical AI	Clinical decision support using deterministic and constrained neural systems for safety
Robotics & Automation	Hybrid	[122, 123, 105, 104, 50]	2018-2025	Autonomous systems	Combines symbolic planners (POMDPs) for safety with neural components for adaptability
Financial Systems	Neural	[124, 125, 43, 79, 108]	2023-2025	FinTech	Neural agents for fraud detection, algorithmic trading, and risk assessment
Education Technology	Neural	[106, 107]	2020-2025	EdTech	Neural-based adaptive learning systems and intelligent tutoring
Legal & Compliance	Neural (RAG)	[45]	2024	Legal tech	Neural agents heavily constrained by symbolic retrieval (RAG) for accuracy
Ethics & Governance	Neural	[9, 120, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146]	2019-2025	AI safety	Frameworks addressing neural-specific challenges (alignment, bias, opacity)
Evaluation & Benchmarking	Neural	[101, 147, 148, 149, 150]	2023-2025	Performance metrics	Benchmarks focused on neural agent capabilities (reasoning, tool use)
Emerging Technologies	Hybrid	[151, 90, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 88, 89, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173]	2020-2025	Innovation frontiers	Research into neuro-symbolic integration, quantum AI, and human-AI collaboration

Key Insights from the Paradigm-Aware Taxonomy Our paradigm-aware taxonomy yields several pivotal insights that chart the current and future state of Agentic AI. Primarily, it reveals a clear **paradigm-market fit**, wherein symbolic and hybrid architectures demonstrably dominate safety-critical applications like healthcare and robotics, while pure neural systems thrive in data-rich, adaptive domains such as finance and education. Furthermore, the taxonomy exposes a significant **governance imbalance**; while ethical challenges within the neural paradigm are the subject of intense research, the governance of modern, complex symbolic systems remains a critically underexplored area. This insight directly informs the third finding: that the most viable **path forward is hybrid**. The most active and promising research in emerging technologies explicitly seeks to integrate both paradigms, a strategic direction that confirms the thesis outlined in Section 9. Finally, the successful classification of all 90 studies by this dualist framework validates its comprehensive coverage and utility as a robust tool for literature analysis and future research design.

7 Ethical and Governance Challenges: A Paradigm-Specific Analysis

As Agentic AI systems gain autonomy and are deployed in critical domains, they introduce a complex spectrum of ethical and governance concerns [120, 9, 174]. However, a critical oversight in current discourse is the treatment of these challenges as monolithic. The risks and requisite mitigation strategies differ profoundly between the symbolic and neural paradigms, demanding a paradigm-aware approach to oversight and interdisciplinary collaboration.

A synthesis of these issues is presented in Table 9, which expands upon standard taxonomies by outlining the core challenges and, most importantly, their paradigm-specific manifestations and governance implications.

Analysis and Summary

The bifurcation of ethical challenges detailed in Table 9 leads to several critical and interconnected conclusions. First, it becomes evident that **effective governance cannot be architecturally agnostic**. Regulation and ethical oversight must be predicated on the underlying paradigm; a requirement for "full explainability," for instance, is feasible for a symbolic system but may be technologically impossible for a pure neural agent, thus necessitating the development of alternative compliance mechanisms.

Furthermore, the rise of **hybrid systems compounds ethical complexity**. An agent that blends paradigms inherently inherits the governance challenges of both. A neuro-symbolic architecture, for example, requires a framework capable of auditing its deterministic symbolic logic while simultaneously monitoring its neural components for stochastic failures, creating a significantly more demanding oversight burden.

Conversely, **the attribution gap presents a specific crisis for the neural paradigm**. The fundamental question of "Who is liable?" is most acute here, as its diffuse and stochastic nature directly challenges legal frameworks built on principles of direct causation and intent. This may ultimately require the establishment of new forms of strict liability for developers and operators.

Finally, these distinctions mean that **effective human-AI collaboration is inherently paradigm-dependent**. Designing appropriate human oversight requires a deep understanding of the agent's core mechanics. The process of overseeing a symbolic agent is analogous to supervising a junior programmer—it involves checking their logical steps. In stark contrast, overseeing a neural agent is more akin to supervising a brilliant but unpredictable intern—it requires carefully steering their context and interpreting their often-opaque outputs.

Table 9: Paradigm-Specific Ethical and Governance Challenges in Agentic AI

Challenge	Symbolic Paradigm Manifestation	Neural Paradigm Manifestation	Governance and Mitigation Strategies
Accountability & Liability [130, 131]	Failure due to flawed logic or unhandled edge cases. Liability is potentially traceable to programmers or system designers.	Failure due to stochastic outputs, prompt injection, or training data biases. Liability is diffuse and difficult to attribute.	**Paradigm-specific standards:** Symbolic: Code verification, formal proof of correctness. Neural: Output watermarking, robust prompt shielding, audit trails for context history.
Transparency & Explainability [132, 133]	Inherently high. Reasoning trace is a sequence of logical steps or rule firings. "Why?" is answerable.	Inherently low. "Reasoning" is an emergent property of model activations. "How?" is often unanswerable; "Why?" is inferred.	**Symbolic:** Leverage native explainability. **Neural:** Invest in SHAP/LIME-style post-hoc explanations and mandatory decision logs. **Hybrid:** Use symbolic modules to generate explanations for neural decisions.
Bias & Fairness [135, 136]	Bias arises from explicit, hand-coded rules or knowledge bases. Easier to identify but hard to root out if foundational.	Bias is latent in training data and amplified stochastically. Pervasive and subtle, emerging in novel contexts.	**Symbolic:** Rigorous logic audits, diverse design teams. **Neural:** Continuous bias monitoring, curated fine-tuning datasets, adversarial debiasing.
Safety & Misalignment [137, 138]	Risk of "perverse instantiation" where agents exploit literal, rigid goals with unintended consequences.	Risk of goal drift, prompt hacking, and value misgeneralization where agents pursue correlated but incorrect proxies.	**Symbolic:** Comprehensive failure mode testing. **Neural:** Red teaming, constitutional AI, and harmlessness training. **Universal:** Sandboxed testing environments.
Autonomy vs. Control [127, 128]	Human oversight is typically designed as explicit veto points or permission gates within a deterministic loop.	Human oversight is fuzzy, often implemented as "human-in-the-loop" feedback, which can be ignored or gamed by the agent.	**Define "meaningful human control" by paradigm.** Symbolic: Clear interrupt signals. Neural: Confidence thresholding for automatic deferral and nuanced steering mechanisms.
Security & Resilience [175, 176]	Vulnerabilities include logic bombs, sensor spoofing, and exploiting algorithmic flaws.	Vulnerabilities include prompt injection, training data poisoning, and adversarial attacks on embeddings.	**Paradigm-specific defense:** Symbolic: Formal verification, intrusion detection. Neural: Advanced prompt hardening, detection of out-of-distribution inputs, data provenance.

Addressing the ethical and governance issues of Agentic AI is essential to harness its transformative potential. However, this analysis demonstrates that a nuanced, paradigm-specific approach is not just beneficial but necessary. Blanket policies will inevitably fail. The path forward requires technical standards, legal frameworks, and ethical guidelines that are as sophisticated and differentiated as the technologies they aim to govern.

This paradigm-specific framing, however, remains incomplete without explicit consideration of policy frameworks that account for the degrees of agency and autonomy in Agentic AI systems, an issue we address next.

7.1 Toward Agentic AI Policy

An overlooked but critical dimension of ethical and governance discourse is the explicit development of *policy frameworks tailored to agentic AI*. Current governance proposals often extend existing AI regulations to cover autonomous systems, but they seldom distinguish between systems that merely generate outputs and those that **exercise agency in decision-making**. For agentic AI, the challenge lies in **defining and operationalizing levels of autonomy** and clarifying their governance implications.

Policy mechanisms must therefore incorporate criteria that distinguish different levels of agency. Table 10 summarizes a proposed taxonomy of agency in Agentic AI, outlining the characteristics of assistive, shared, and delegated forms of agency alongside their governance implications.

Table 10: Levels of Agency in Agentic AI and Corresponding Policy Needs

Agency Level	Characteristics	Governance and Policy Requirements
Assistive	AI provides recommendations or analysis, with all final decisions made by humans.	Ensure transparency and explainability. Policies should mandate auditability of outputs but allow flexible use with human oversight.
Shared	AI participates in decision-making, influencing outcomes jointly with human actors.	Require clear role allocation, decision-logging, and mechanisms for tracing contributions of human vs. AI actors. Liability is shared and must be explicitly codified.
Delegated	AI agents operate with high autonomy, executing decisions or actions within defined domains.	Strong accountability mechanisms, predefined bounds of autonomy, and strict liability regimes for developers/operators. Requires robust monitoring and override capabilities.

Accordingly, governance must move beyond paradigm-specific risk analysis toward a **taxonomy of agency**, where ethical principles and legal accountability mechanisms scale with the degree of autonomy. This aligns with calls for "meaningful human control" [127], but extends them into concrete policy design that recognizes the unique governance needs of agentic AI.

8 Research Gaps: A Paradigm-Specific Roadmap

The development of Agentic AI is constrained by significant, unresolved challenges. However, a critical oversight in identifying these gaps is treating them as uniform across architectures. The research imperatives for symbolic systems diverge profoundly from those for neural systems, with a particularly pressing need for work on hybrid architectures that can leverage the strengths

of both. As outlined in Table 11, these thematic areas require a paradigm-aware research strategy to ensure future systems are robust, adaptable, and aligned.

Commentary on Key Themes   The bifurcation of research gaps identified in Table 11 reveals that the most critical overarching challenge is the current lack of **Paradigm-Aware Research Methodologies**. The tools, benchmarks, and success criteria developed for one paradigm frequently prove irrelevant or misapplied to the other, creating fundamental barriers to coherent progress.

This analysis suggests several imperative directions for future work. First, the most promising research path forward appears to lie not in pursuing either paradigm in isolation, but in their **intentional integration**. The "Reasoning & Adaptability" gap, for instance, represents a prime candidate for neuro-symbolic solutions, wherein a neural network's robust pattern recognition capabilities are systematically guided and constrained by a symbolic reasoner's logical framework.

Furthermore, the community must move **beyond isolated benchmarks** that fail to account for paradigmatic differences. There is a critical need to develop separate, rigorous evaluation suites that stress-test the unique failure modes of each architecture—such as logic bombs and edge-case reasoning for symbolic systems, and prompt injection resilience and output stability for neural systems.

Perhaps most urgently, this bifurcation demonstrates that **effective governance cannot follow a one-size-fits-all approach**. Policymakers and ethicists must collaborate with engineers to develop distinct, tailored frameworks for auditing and regulating these fundamentally different technologies. Applying the stringent verifiability standards of symbolic systems to neural architectures would inadvertently stifle innovation, while applying the more flexible standards designed for neural systems to symbolic environments would overlook critical risks associated with logical integrity and deterministic failure.

Conclusion

Addressing these gaps requires a conscious departure from generic AI research. Progress hinges on a dual-track strategy that deepens our understanding of each paradigm's unique challenges while simultaneously pioneering architectures and standards for their integration. This paradigm-specific roadmap is essential to move from powerful but flawed prototypes to reliable and trustworthy agentic systems. The future of Agentic AI is not a choice between symbolism and connectionism, but a strategic synthesis of both.

9   Future Directions: The Path to Hybrid Intelligence

Agentic AI systems are rapidly evolving beyond static task automation into dynamic, collaborative, and adaptive entities [157]. Their future development will hinge on interdisciplinary advances, technological convergence, and—critically—a paradigm-aware approach to design that seeks to integrate the strengths of both symbolic and neural lineages into robust hybrid architectures.

A summary of these paradigm-aware trajectories is presented in Table 12, which outlines the specific research and integration priorities for each paradigm's evolution, moving beyond a generic technology forecast.

Analysis of Strategic Trajectories

The bifurcated future outlined in Table 12 leads to one overriding conclusion: the paramount direction is **Architectural Integration**. The goal is to forge a new class of hybrid systems that leverage the reliability of symbolic reasoning and the adaptability of neural generation.

Table 11: Paradigm-Specific Research Gaps and Imperatives in Agentic AI

Gap Area	Symbolic Paradigm Challenges	Neural Paradigm Challenges	Research Imperatives
Evaluation & Benchmarks [150, 149]	Lack of standardized metrics for scalability and robustness of logical reasoning in complex, open-world environments.	Current benchmarks (e.g., *AgentBench* [101], *GAIA* [**?**]) fail to adequately test for subtle misalignments, prompt robustness, and the true cost of context management.	Develop paradigm-specific benchmarks. Symbolic: Test logical soundness and failure predictability. Neural: Test for hallucination under pressure, prompt injection resilience, and multi-session consistency.
Reasoning & Adaptability [165, 166]	Systems are brittle; they fail catastrophically when faced with novel scenarios or exceptions not covered by their rules.	Agents struggle with true, abstract reasoning and value-laden judgment. Their "reasoning" is often just sophisticated pattern matching that can break down.	**Hybrid Research:** Investigate neuro-symbolic architectures where neural components handle pattern recognition and symbolic modules enforce rigorous reasoning and constraint checking.
Long-term Autonomy & Memory [90]	Can maintain a persistent, symbolic state but struggle to learn and update their world model from experience in a scalable way.	Context window limitations create agents with severe amnesia across sessions. Statelessness prevents cumulative learning and building long-term relationships.	**Symbolic:** Research on efficient belief revision. **Neural:** Develop architectures for external, structured memory that agents can reliably read from and write to.
AI Infrastructure Dependence [156]	Performance is often constrained by the scalability of theorem provers and logic engines, which are sensitive to hardware architecture. Less dependent on massive cloud clusters but requires specialized, reliable compute.	Extreme dependence on vast, expensive cloud compute for training and inference. Creates environmental costs, centralizes power, and creates vulnerabilities to supply chain and geopolitical disruptions.	Develop energy-efficient and decentralized computing paradigms. Research model distillation, sparse architectures, and hybrid cloud-edge deployment to reduce reliance on monolithic infrastructure.
Human-AI Interaction & Interface Design [158]	Interfaces are typically explicit (e.g., config files, rule editors). The goal is to augment human intelligence with transparent, predictable tools. The distinction between user and agent is clear.	The goal is often a collaborative, conversational partner. Risk of creating opaque "oracles" that users over-trust. Challenges in designing intuitive interfaces for steering, interrupting, and interpreting the stochastic outputs of neural agents.	Establish principles for **paradigm-aware HCI**. Symbolic: Develop advanced visualization for logic and state. Neural: Research intuitive methods for context steering, confidence communication, and collaborative task management.
Trust & Transparency [142, 144]	"How" decisions are made is transparent (the logic trace), but "why" a specific rule exists can be opaque.	Both "how" and "why" are opaque. Explanations are post-hoc and often unreliable. This is the primary barrier to high-stakes deployment.	**Symbolic:** Research on making goal structures and utility functions explicable. **Neural:** Fundamental research on mechanistic interpretability and generating faithful, real-time explanations.
Safety & Alignment [137, 138]	Risk of "perverse instantiation" – perfectly executing a flawed or oversimplified goal specification with catastrophic results.	Vulnerability to prompt injection, goal drift, and value misgeneralization. Aligning a stochastic model to complex human values is an unsolved problem.	**Paradigm-specific strategies:** Symbolic: Formal verification of goals and constraints. Neural: Advanced red teaming, adversarial training, and "constitutional" oversight mechanisms.
Interoperability & Integration [167, 168]	Difficult to integrate with the messy, unstructured data of the real world and modern software ecosystems.	Excel at using tools via APIs but struggle with true, semantic understanding of what a tool does, leading to misuse.	Develop standards and middleware for **paradigm bridging**. Create APIs that allow neural agents to query symbolic reasoners for validation and symbolic systems to leverage neural networks for perception.
Governance & Accountability [145, 146]	Liability is more straightforward (flawed logic can be traced) but frameworks for auditing complex rule sets are needed.	A profound "attribution gap" exists. Legal frameworks are unprepared for harm caused by emergent, stochastic behavior.	**Urgently develop paradigm-specific regulatory models.** Symbolic: Audit trails for decision logic. Neural: Mandatory context logging, output watermarking, and potentially new forms of developer liability.

Table 12: Paradigm-Aware Strategic Trajectories for Agentic AI

Strategic Direction	Symbolic Paradigm Evolution	Neural Paradigm Evolution
Multi-Agent Ecosystems	Defining verifiable communication protocols and interaction contracts for hybrid agent teams.	Specializing in emergent, role-based collaboration and negotiation [154, 155] (e.g., CrewAI, AutoGen, LangGraph).
Technological Convergence	Providing the reliable, verifiable logic layer for cyber-physical systems and smart infrastructure.	Acting as the adaptive interface for integrating with IoT, robotics, blockchain, and quantum computing [151, 156].
Self-Evolving Architectures	Research into automated theorem proving and logical rule discovery for system self-improvement.	Advancing meta-learning and feedback-driven optimization [157] for architecture tuning and deployment-aware adaptation.
Human-AI Collaboration	Enabling interfaces where humans can directly inspect, debug, and modify an agent's logical rule set and goals.	Creating intuitive interfaces for shared intent and cognitive/emotional responsiveness [158] via natural language.
Governance-First Design	Formal verification of goal structures and safety constraints for embeddable governance modules.	Developing techniques for embedded ethics, policy enforcement, and global accountability [9] within stochastic systems (e.g., IBM Governance Stack).
Scientific Discovery	Encoding scientific laws and methodological rigor for agent-led hypothesis generation.	Driving agent-led inquiry and results analysis [159, 160] in platforms like Sakana AI Scientist [160] and Microsoft Discovery.
Research Priorities	Establishing benchmarks for logical soundness, verifiability, and interoperability standards.	Establishing metrics for moral alignment, cognitive modeling, and alignment [161] (e.g., AgentBench [101]).

- **Neuro-Symbolic Integration as the Keystone:** The most profound progress will come from research that successfully couples neural networks for perception and pattern recognition with symbolic engines for reasoning and constraint checking. This is the most promising path to overcoming the brittleness of pure symbolism and the opacity of pure neural approaches.

- **Paradigm-Specialized Roles in Ecosystems:** Future multi-agent ecosystems [154, 155] will not be homogenous. They will consist of specialized agents—some highly neural for creative tasks, some highly symbolic for regulatory compliance—that communicate through standardized protocols. The orchestration of such hybrid swarms is a critical research frontier.

- **A Dual-Track Approach to Governance:** The development of safety and governance mechanisms [9] must continue on two tracks: advancing formal methods for symbolic verifiability *and* developing new statistical, training-based methods for neural alignment. The ultimate governance framework for a hybrid agent will need to seamlessly combine both.

- **Convergence as Amplification:** The integration with other technologies [151, 156] will amplify the capabilities of both paradigms. Neural agents will manage real-time sensor data from IoT, while symbolic modules will ensure the decisions made from that data are safe and compliant.

Conclusion

The future of Agentic AI is a synthesis. Its trajectory will be shaped not only by technical breakthroughs but by thoughtful, paradigm-aware integration of ethics, interdisciplinary methods, and infrastructure-aware governance [9]. The next conceptual turning point will be defined by our ability to engineer **hybrid intelligence**—systems that are both *adaptable* and *reliable*, both *creative* and *sound*. The question is no longer whether agents will become intelligent partners, but whether we can architect a future of hybrid intelligence that is both powerful and trustworthy.

10 Conclusion

Agentic AI represents a fundamental paradigm shift in the design of intelligent systems, but its rapid evolution has led to a fragmented and often anachronistic understanding of the field. This review has addressed this confusion by introducing and validating a novel conceptual framework: the existence of two distinct lineages of Agentic AI—the **Symbolic/Classical** and the **Neural/Generative**—each with fundamentally different operational mechanics, strengths, and limitations.

Our analysis demonstrates that the common practice of *conceptual retrofitting*—describing modern LLM-orchestrated agents with the language of symbolic systems (e.g., PPAR loops, BDI)—obscures their true nature and impedes progress. Through a systematic, paradigm-aware review of the literature, we have established three central tenets. First, **the architectural divide is both real and meaningful**; symbolic systems excel in environments requiring safety, verifiability, and explicit logic (e.g., healthcare, robotics control), while neural systems thrive in domains requiring adaptability, pattern recognition, and operation on unstructured data (e.g., finance, creative research) (Sections 5, 6).

Furthermore, this divide dictates that **governance must be paradigm-specific**. The ethical challenges and requisite mitigation strategies differ profoundly between paradigms, meaning accountability for a symbolic system involves auditing its logic, whereas for a neural system, it

necessitates auditing its training data and prompts. This renders a one-size-fits-all approach to AI ethics fundamentally insufficient (Section 7).

Critically, our findings indicate that **the most productive path forward is hybrid, not isolated**. The most pressing research gaps and promising future directions lie not in the isolated improvement of either paradigm, but in their strategic integration into neuro-symbolic architectures that leverage the complementary strengths of symbolic reliability and neural adaptability (Sections 8, 9).

This dual-paradigm framework provides the essential analytical lens to move the field beyond a simple catalog of technologies toward a coherent theory of architectural design in Agentic AI. It offers researchers, engineers, and policymakers a precise vocabulary and a functional taxonomy to classify systems, evaluate their capabilities and risks appropriately, and make informed design choices.

Ultimately, the development of Agentic AI is not merely a technical challenge—it is a sociotechnical one. Its success will depend on whether we can architect systems that are not only powerful but also trustworthy. This requires a conscious and deliberate effort to build hybrid intelligence—systems that are both adaptable and reliable, both creative and sound. By recognizing and embracing the distinct nature of these two architectural lineages, we can steer this transformative technology toward a future where agentic systems truly serve as trusted collaborators in scientific discovery (understanding), in providing fair and accessible services (equity), and in forming the robust, verifiable backbone of critical infrastructure (resilience).

References

[1] Christopher Wissuchek and Patrick Zschech. Exploring agentic artificial intelligence systems: Towards a typological framework. *arXiv preprint arXiv:2508.00844*, July 2025.

[2] Panneer Selvam Viswanathan. Agentic ai: A comprehensive framework for autonomous decision-making systems in artificial intelligence. *International Journal of Computer Engineering and Technology*, 16(1):862–880, 2025. IJCET, ISSN Print: 0976-6367; Online: 0976-6375.

[3] J. Xie, Z. Chen, R. Zhang, X. Wan, and G. Li. Large multimodal agents: A survey. *arXiv [cs.CV]*, 2024.

[4] H. Du, S. Thudumu, R. Vasa, and K. Mouzakis. A survey on context-aware multi-agent systems: Techniques, challenges and future directions. *arXiv [cs.MA]*, 2025.

[5] B. Archibald, M. Calder, M. Sevegnani, and M. Xu. Quantitative modelling and analysis of bdi agents. *Softw. Syst. Model.*, 23(2):343–367, 2024.

[6] Q. Zeng et al. Perceive, reflect, and plan: Designing llm agent for goal-directed city navigation without instructions. *arXiv [cs.AI]*, 2024.

[7] L. E. Erdogan et al. Plan-and-act: Improving planning of agents for long-horizon tasks. *arXiv [cs.CL]*, 2025.

[8] A. Plaat, M. van Duijn, N. van Stein, M. Preuss, P. van der Putten, and K. J. Batenburg. Agentic large language models, a survey. *arXiv [cs.AI]*, 2025.

[9] G.A. Gabison and R.P. Xian. Inherent and emergent liability issues in llm-based agentic systems: a principal-agent perspective. *arXiv [cs.CY]*, 2025.

[10] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Jirong Wen. A survey on large language model based autonomous agents. *Front. Comput. Sci.*, 18(6), 2024.

[11] Pengyu Zhao, Zijian Jin, and Ning Cheng. An in-depth survey of large language model-based artificial intelligence agents. 2023.

[12] Shuaihang Chen, Yuanxing Liu, Wei Han, Weinan Zhang, and Ting Liu. A survey on LLM-based multi-agent system: Recent advances and new frontiers in application. 2024.

[13] B. Liang. Ai reasoning in deep learning era: From symbolic ai to ... *Mathematics*, 13(11):1707, 2025. Includes section "Symbolic Reasoning Era (1950s–1980s)".

[14] William R. Swartout. Rule-based expert systems: The mycin experiments of the stanford heuristic programming project: B.g. buchanan and e.h. shortliffe, (addison-wesley, reading, ma, 1984); 702 pages, $40.50. *Artificial Intelligence*, 26(3):364–366, 1985.

[15] R. N. Thomas and R. Gupta. A survey on machine learning approaches and its techniques. In *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pages 1–6, 2020.

[16] T. Nithya, V. Nivas Kumar, Gayathri, S. Deepa, V. C, and R. Siva Subramanian. A comprehensive survey of machine learning: Advancements, applications, and challenges. In *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, pages 354–361, 2023.

[17] Maria Trigka and Elias Dritsas. A comprehensive survey of machine learning techniques and models for object detection. *Sensors*, 25(1):214, 2025.

[18] William G. Hatcher and Wei Yu. A survey of deep learning: Platforms, applications and emerging research trends. *IEEE Access*, 6:24411–24432, 2018.

[19] Md Zahangir Alom, Tarek M. Taha, Christopher Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Brian C. Van Esesn, Abdul A. S. Awwal, and Vijayan K. Asari. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3):292, 2019.

[20] Shengli Dong, Peilin Wang, and Ke Chen Abbas. A survey on deep learning and its applications. *Computer Science Review*, 40:100379, 2021.

[21] Talaei Khoei, H. Ould Slimane, and N. Kaabouch. Deep learning: Systematic review, models, challenges, and research directions. *Neural Computing and Applications*, 2023.

[22] Pooja Chhabra and D. S. Goyal. A thorough review on deep learning neural network. In *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, pages 220–226, 2023.

[23] T. Sakirin and S. Kusuma. A survey of generative artificial intelligence techniques. *Babylonian Journal of Artificial Intelligence*, 2023:10–14, 2023.

[24] G. Anandhi. A comprehensive survey on generative ai techniques and their tools: Recent advances, applications, opportunities, and challenges. *Recent Research in Machine Learning and Cloud Computing*, 4(1):44–54, 2025.

[25] S. S. Sengar, A. B. Hasan, S. Kumar, and F. Carroll. Generative artificial intelligence: A systematic review and applications. *Multimedia Tools and Applications*, 84(21):23661–23700, 2024.

[26] F. P. S. Surbakti. Systematic Literature Review on generative AI: Ethical challenges and opportunities. *International Journal of Advanced Computer Science and Applications*, 16(5), 2025.

[27] Zheng Zhang, Jie Zhang, Xiang Zhang, and Wei Mai. A comprehensive overview of Generative AI (GAI): Technologies, applications, and challenges. *Neurocomputing*, 632:129645, 2025.

[28] Z. Durante et al. Agent AI: Surveying the horizons of multimodal interaction. arXiv preprint arXiv:2401.03568, 2024.

[29] T. Masterman, S. Besen, M. Sawtell, and A. Chao. The landscape of emerging AI agent architectures for reasoning, planning, and tool calling: A survey. arXiv preprint arXiv:2404.11584, 2024.

[30] Francesco Piccialli, D. Chiaro, S. Sarwar, D. Cerciello, P. Qi, and V. Mele. AgentAI: A comprehensive survey on autonomous agents in distributed AI for industry 4.0. *Expert Systems with Applications*, 291:128404, 2025.

[31] D. B. Acharya, K. Kuppan, and B. Divya. Agentic AI: Autonomous intelligence for complex goals—A comprehensive survey. *IEEE Access*, 13:18912–18936, 2025.

[32] P. S. Viswanathan. Agentic ai: A comprehensive framework for autonomous decision-making systems in artificial intelligence. *International Journal of Computer Engineering and Technology*, 16(1):862–879, 2025.

[33] Aske Plaat, Max van Duijn, Nathan van Stein, Mike Preuss, Peter van der Putten, and Kees Joost Batenburg. Agentic large language models, a survey. *arXiv*, 2025.

[34] Johannes Schneider. Generative to agentic ai: Survey, conceptualization, and challenges. *arXiv*, 2025.

[35] Saeid Hosseini and Hamed Seilani. The role of agentic ai in shaping a smart future: A systematic review. *Array*, 26:100399, 2025.

[36] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[37] Zhongwei Wan, Xin Wang, Che Liu, Samiul Alam, Yu Zheng, Jiachen Liu, Zhongnan Qu, Shen Yan, Yi Zhu, Quanlu Zhang, Mosharaf Chowdhury, and Mi Zhang. Efficient large language models: A survey. *Transactions on Machine Learning Research (TMLR)*, 2024. Published May 20, 2024; accepted Sept 17, 2024.

[38] N. Kolt. Governing ai agents. *arXiv [cs.AI]*, 2025.

[39] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2):99–134, 1998. [Online] Available:.

[40] Iadine Chadès, Luz V Pascal, Sam Nicol, Cameron S Fletcher, and Jonathan Ferrer-Mestres. A primer on partially observable markov decision processes (pomdps). *Methods Ecol. Evol.*, 12(11):2058–2072, 2021. [Online] Available:.

[41] V. Mavroudis. Langchain v0.3. *Preprints Organization*, 2024.

[42] A. Singh, A. Ehtesham, S. Mahmud, and J.-H. Kim. Revolutionizing mental health care through langchain: A journey with a large language model. In *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0073–0078, 2024.

[43] G. Chandrashekar, M.T. Akram, M. Khan, P. Kumar, and P. Mandal. A survey on stock investment risk analysis using crewai multi-agent system. *International Research Journal of Modernization in Engineering Technology and Science*, 7(1):5647–5650, 2025.

[44] M. Gridach, J. Nanavati, K. Z. E. Abidine, L. Mendes, and C. Mack. Agentic ai for scientific discovery: A survey of progress, challenges, and future directions. *arXiv [cs.CL]*, 2025.

[45] V. Magesh, F. Surani, M. Dahl, M. Suzgun, C.D. Manning, and D.E. Ho. Hallucination-free? assessing the reliability of leading ai legal research tools. *arXiv [cs.CL]*, 2024.

[46] J. Schneider. Generative to agentic ai: Survey, conceptualization, and challenges. *arXiv [cs.AI]*, 2025.

[47] D. B. Acharya, K. Kuppan, and B. Divya. Agentic ai: Autonomous intelligence for complex goals—a comprehensive survey. *IEEE Access*, 13:18912–18936, 2025.

[48] S. Hosseini and H. Seilani. The role of agentic ai in shaping a smart future: A systematic review. *Array (N. Y.)*, 26(100399):100399, 2025.

[49] R. Sapkota, K. I. Roumeliotis, and M. Karkee. Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges. *arXiv [cs.AI]*, 2025.

[50] V. Patel, S. Kanani, T. Pathak, P. Patel, M. I. Ali, and J. Breslin. A demonstration of smart doorbell design using federated deep learning. *arXiv [cs.DC]*, 2020.

[51] V. Trencsenyi, A. Mensfelt, and K. Stathis. The influence of human-inspired agentic sophistication in llm-driven strategic reasoners. *arXiv [cs.AI]*, 2025.

[52] Y. Lu, M. S. Squillante, and C. W. Wu. A general markov decision process framework for directly learning optimal control policies. *arXiv [cs.LG]*, 2019.

[53] Y. Lu, M. S. Squillante, and C. W. Wu. Markov decision process framework for control-based reinforcement learning. *ACM SIGMETRICS Performance Evaluation Review*, 51(2):39–41, 2023.

[54] B. Rozek, J. Lee, H. Kokel, M. Katz, and S. Sohrabi. Partially observable hierarchical reinforcement learning with ai planning (student abstract). *Proc. Conf. AAAI Artif. Intell.*, 38(21):23635–23636, 2024.

[55] C. Lu, R. Shi, Y. Liu, K. Hu, S. S. Du, and H. Xu. Rethinking transformers in solving pomdps. *arXiv [cs.LG]*, 2024.

[56] L. Frering, G. Steinbauer-Wagner, and A. Holzinger. Integrating belief-desire-intention agents with large language models for reliable human–robot interaction and explainable artificial intelligence. *Eng. Appl. Artif. Intell.*, 141(109771):109771, 2025.

[57] S. Gillen and K. Byl. Explicitly encouraging low fractional dimensional trajectories via reinforcement learning. *arXiv [cs.LG]*, 2020.

[58] J. Singh, Raghav Magazine, Y. Pandya, and A. Nambi. Agentic reasoning and tool integration for llms via reinforcement learning. *arXiv [cs.AI]*, 2025.

[59] A. Varun Bodepudi, N. Katnapally, V. Velaga, C. S. Moore, P. C. R. Chinta, and L. M. Karaka. Agentic ai and reinforcement learning: Towards more autonomous and adaptive ai systems. *Journal for Educators, Teachers and Trainers*, 11(1):177–193, 2020.

[60] J. Kumar and V. K. Elumalai. A proximal policy optimization based deep reinforcement learning framework for tracking control of a flexible robotic manipulator. *Results Eng.*, 25(104178):104178, 2025.

[61] I. N. Yazid and E. Rachmawati. Autonomous driving system using proximal policy optimization in deep reinforcement learning. *IAES Int. J. Artif. Intell. (IJ-AI)*, 12(1):422, 2023.

[62] R. Johnson. *LangChain Essentials: From Basics to Advanced AI Applications*. HiTeX Press, 2025.

[63] M. Gupta. *LangChain in your Pocket: LangChain Essentials: From Basic Concepts to Advanced Applications*. Packt Publishing, Birmingham, England, 2024.

[64] O. Topsakal and T.C. Akinci. Creating large language model applications utilizing langchain: A primer on developing llm apps fast. In *International Conference on Applied Engineering and Natural Sciences*, volume 1, pages 1050–1056, 2023.

[65] T. Taulli and G. Deshmukh. *Building generative AI agents: Using LangGraph, AutoGen, and CrewAI*. APress, Berlin, Germany, 2025.

[66] J. Huh, H.J. Park, and J.C. Ye. Breast ultrasound report generation using langchain. *arXiv [eess.IV]*, 2023.

[67] Q. Wu et al. Autogen: Enabling next-gen llm applications via multi-agent conversation. *arXiv [cs.AI]*, 2023.

[68] V. Dibia. *Multi-Agent Systems with AutoGen*. Manning Publications, New York, NY, 2025.

[69] H. Dawid, P. Harting, H. Wang, Z. Wang, and J. Yi. Agentic workflows for economic research: Design and implementation. *arXiv [econ.GN]*, 2025.

[70] P. Venkadesh, S. V. Divya, and K. S. Kumar. Unlocking ai creativity: A multi-agent approach with crewai. *Journal of Trends in Computer Science and Smart Technology*, 6(4):338–356, 2024.

[71] Z. Duan and J. Wang. Exploration of llm multi-agent application implementation based on langgraph+crewai. *arXiv [cs.MA]*, 2024.

[72] M. Kothapalli. Integrating web applications with azure openai services: A focus on semantic kernel. *Int. J. Sci. Res. (Raipur)*, 13(3):1918–1923, 2024.

[73] L.A. Meyer. *Building AI Applications with Microsoft Semantic Kernel: Easily integrate generative AI capabilities and copilot experiences into your applications*. Packt Publishing, Birmingham, England, 2024.

[74] D. Costea. *Microsoft Semantic Kernel in Action*. Manning Publications, Shelton, Connecticut, USA, 2025.

[75] A. Gheorghiu. *Building Data-Driven Applications with LlamaIndex: A practical guide to retrieval-augmented generation (RAG) to enhance LLM applications*. Packt Publishing, Birmingham, England, 2024.

[76] J. Ramirez-Medina, M. Ataei, and A. Amirfazli. Accelerating scientific research through a multi-llm framework. *arXiv [physics.app-ph]*, 2025.

[77] D. Mozolevskyi and W. AlShikh. Comparative analysis of retrieval systems in the real world. *arXiv [cs.IR]*, 2024.

[78] N. Braunschweiler, R. Doddipatla, S. Keizer, and S. Stoyanchev. Evaluating large language models for document-grounded response generation in information-seeking dialogues. *arXiv [cs.CL]*, 2023.

[79] T. Konstantinidis, G. Iacovides, M. Xu, T.G. Constantinides, and D. Mandic. Finllama: Financial sentiment classification for algorithmic trading applications. *arXiv [cs.CL]*, 2024.

[80] V.K. Kommineni, B. König-Ries, and S. Samuel. Harnessing multiple llms for information retrieval: A case study on deep learning methodologies in biodiversity publications. *arXiv [cs.IR]*, 2024.

[81] J. Wang and Z. Duan. Agent ai with langgraph: A modular framework for enhancing machine translation using large language models. *arXiv [cs.CL]*, 2024.

[82] M.J. Page et al. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71, 2021.

[83] M.J. Page et al. Prisma 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372:n160, 2021.

[84] J. E. Laird. An analysis and comparison of act-r and soar. *arXiv [cs.AI]*, 2022.

[85] J. Thomas and A. Harden. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med. Res. Methodol.*, 8(1):45, 2008.

[86] T.T. Van, H.D. The, T.V. Van, and M.D. Van. Applying qualitative research in management studies - theory and practical experiences: Using nvivo 15. *ijirss*, 8(2):4617–4626, 2025.

[87] A. Basit, K. Hussain, M.A. Hanif, and M. Shafique. Medaide: Leveraging large language models for on-premise medical assistance on edge devices. *arXiv [cs.AI]*, 2024.

[88] B.S. Nayak. Neuro-symbolic integration in ai agents: Bridging the gap between perception and reasoning. *International Journal of Computer Engineering and Technology*, 16(1):1142–1158, 2025.

[89] M.M. Karim, D.H. Van, S. Khan, Q. Qu, and Y. Kholodov. Ai agents meet blockchain: A survey on secure and scalable collaboration for multi-agents. *Future Internet*, 17(2):57, 2025.

[90] J. Zheng et al. Lifelong learning of large language model based agents: A roadmap. *arXiv [cs.AI]*, 2025.

[91] Lai Xu and Hans Weigand. The evolution of the contract net protocol. In X. Sean Wang, Ge Yu, and Hongjun Lu, editors, *Advances in Web-Age Information Management*, pages 257–264, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[92] I D Craig. Blackboard systems. *Artif. Intell. Rev.*, 2(2):103–118, 1988.

[93] Uwe M. Borghoff, Paolo Bottoni, and Remo Pareschi. Human-artificial interaction in the age of agentic ai: a system-theoretical approach. *Frontiers in Human Dynamics*, 7, May 2025.

[94] Zhao Wang, Sota Moriyama, Wei-Yao Wang, Briti Gangopadhyay, and Shingo Takamatsu. Talk structurally, act hierarchically: A collaborative framework for llm multi-agent systems, 2025.

[95] Dimitrios Brodimas, Alexios Birbas, Dimitrios Kapolos, and Spyros Denazis. Intent-based infrastructure and service orchestration using agentic-AI. *IEEE Open J. Commun. Soc.*, pages 1–1, 2025.

[96] Alex Casella and Wayne Wang. Performant llm agentic framework for conversational ai, 2025.

[97] Ziren Luo, Di Li, Jiafu Wan, Shiyong Wang, Ge Wang, Minghao Cheng, and Ting Li. Multi-agent collaboration mechanisms based on distributed online meta-learning for mass personalization. *Journal of Industrial Information Integration*, 46:100852, 2025.

[98] Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D Nguyen. Multi-agent collaboration mechanisms: A survey of LLMs. 2025.

[99] Alessandro Berti, Mayssa Maatallah, Urszula Jessen, Michal Sroka, and Sonia Ayachi Ghannouchi. Re-thinking process mining in the ai-based agents era, 2024.

[100] Lok Hang Cheung, Likai Wang, and Dongxue Lei. Conversational, agentic AI-enhanced architectural design process: three approaches to multimodal AI-enhanced early-stage performative design exploration. *Archit. Intell.*, 4(1), 2025.

[101] X. Liu et al. Agentbench: Evaluating llms as agents. *arXiv [cs.AI]*, 2023.

[102] C. Bird. *Solving Real-World Challenges Using MEDITECH AI*. Medical Information Technology, Inc., 2025.

[103] T. Esslemont and J. Thorpe. Mastercard supercharges consumer protection with gen ai, 2024. Accessed: 13-Jul-2025.

[104] S.R.R. Singireddy and T.U. Daim. *Technology roadmap: Drone delivery – Amazon prime air*, pages 387–412. Springer International Publishing, Cham, 2018.

[105] V.K. Annanth, M. Abinash, and L.B. Rao. Intelligent manufacturing in the context of industry 4.0: A case study of siemens industry. In *J. Phys. Conf. Ser.*, volume 1969, page 012019, 2021.

[106] S. Suh. Investigating the impact of personalized ai tutors on language learning performance. *arXiv [cs.AI]*, 2025.

[107] J. Fisher et al. Livehint: Intelligent digital support for analog learning experiences. pages 80–89, 2020.

[108] S. Al-E'mari, Y. Sanjalawe, and A. Al-E'mari. The role of artificial intelligence in enhancing financial decision-making and administrative efficiency: A systematic review. *Al-Basaer Journal of Business Research*, 2025.

[109] Mahi Ratan Reddy Deva. A review of api management systems and their role in seamless integration between software applications. *Asian Journal of Computer Science Engineering*, 10(2), 2025.

[110] Joshua Ofoeda, Richard Boateng, and John Effah. Application programming interface (api) research: A review of the past to inform the future. *International Journal of Enterprise Information Systems*, 15:76–95, 07 2019.

[111] Lei Liu, Xun Li, Yuzhou Liu, and Huaxiao Liu. Application programming interface recommendation according to the knowledge indexed by app feature mined from app stores. *J. Softw. (Malden)*, 33(11), 2021.

[112] Maxime Lamothe, Yann-Gaël Guéhéneuc, and Weiyi Shang. A systematic review of api evolution literature. *ACM Computing Surveys*, 54:1–36, 10 2021.

[113] Rishi Saripalle, Christopher Runyan, and Mitchell Russell. Using HL7 FHIR to achieve interoperability in patient health record. *J. Biomed. Inform.*, 94(103188):103188, 2019.

[114] Miguel Pedrera-Jiménez, Noelia García-Barrio, Santiago Frid, David Moner, Diego Boscá-Tomás, Raimundo Lozano-Rubí, Dipak Kalra, Thomas Beale, Adolfo Muñoz-Carrero, and Pablo Serrano-Balazote. Can OpenEHR, ISO 13606, and HL7 FHIR work together? an agnostic approach for the selection and application of electronic health record standards to the next-generation health data spaces. *J. Med. Internet Res.*, 25:e48702, 2023.

[115] Hua Zhong, Shan Jiang, and Sarfraz Khurshid. An approach for API synthesis using large language models. 2025.

[116] Shishir G Patil, Tianjun Zhang, Xin Wang, and Joseph E Gonzalez. Gorilla: Large language model connected with massive APIs. 2023.

[117] Hana Derouiche, Zaki Brahmi, and Haithem Mazeni. Agentic AI frameworks: Architectures, protocols, and design challenges. 2025.

[118] Agapi Rissaki, Ilias Fountalis, Nikolaos Vasiloglou, and Wolfgang Gatterbauer. Towards agentic schema refinement. 2024.

[119] Vaibhav Tupe and Shrinath Thube. AI agentic workflows and enterprise APIs: Adapting API architectures for the age of AI agents. 2025.

[120] S. Raza, R. Sapkota, M. Karkee, and C. Emmanouilidis. Trism for agentic ai: A review of trust, risk, and security management in llm-based agentic multi-agent systems. *arXiv [cs.AI]*, 2025.

[121] O. Cárdenas, S. Falconi, E. Tusa, and A. Rodríguez. Development of a chatbot model for health telecare: Integration of langchain, embeddings with openai, and pinecone using the question answering technique. *arXiv*, 22(3):389–402, 2024.

[122] Y. Bai, Z. Ding, and A. Taylor. From virtual agents to robot teams: A multi-robot framework evaluation in high-stakes healthcare context. *arXiv [cs.RO]*, 2025.

[123] P. Zhang, D. Wen, G. Zhu, Q. Chen, K. Han, and Y. Shi. Collaborative edge ai inference over cloud-ran. *IEEE Trans. Commun.*, 72(9):5641–5656, 2024.

[124] S. Roychowdhury et al. Hallucination-minimized data-to-answer framework for financial decision-makers. *arXiv [cs.CL]*, 2023.

[125] Y. Yang, Y. Tang, and K.Y. Tam. Investlm: A large language model for investment using financial domain instruction tuning. *arXiv [q-fin.GN]*, 2023.

[126] G. Syros, A. Suri, C. Nita-Rotaru, and A. Oprea. Saga: A security architecture for governing ai agentic systems. *arXiv [cs.CR]*, 2025.

[127] K.J.K. Feng, D.W. McDonald, and A.X. Zhang. Levels of autonomy for ai agents. *arXiv [cs.HC]*, 2025.

[128] K. Tallam. Alignment, agency and autonomy in frontier ai: A systems engineering perspective. *arXiv [cs.CY]*, 2025.

[129] H. Clatterbuck, C. Castro, and A.M. Morán. Risk alignment in agentic ai systems. *arXiv [cs.CY]*, 2024.

[130] A. Chan et al. Harms from increasingly agentic algorithmic systems. In *2023 ACM Conference on Fairness Accountability and Transparency*, pages 651–666, 2023.

[131] Z. Tóth, R. Caruana, T. Gruber, and C. Loebbecke. The dawn of the ai robots: Towards a new framework of ai robot accountability. *J. Bus. Ethics*, 178(4):895–916, 2022.

[132] S. Baron. Trust, explainability and ai. *Philos. Technol.*, 38(1), 2025.

[133] A. Chan et al. Visibility into ai agents. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2024.

[134] G. Papagni, J. de Pagter, S. Zafari, M. Filzmoser, and S.T. Koeszegi. Artificial agents' explainability to support trust: considerations on timing and context. *AI Soc.*, 38(2):947–960, 2023.

[135] K. Singh and W. Ngu. Bias-aware agent: Enhancing fairness in ai-driven knowledge retrieval. *arXiv [cs.IR]*, 2025.

[136] S.J. Yadav. Ai bias and fairness: Ethical considerations in service marketing strategies. In *Advances in Marketing, Customer Relationship Management, and E-Services*, pages 49–64. IGI Global, 2024.

[137] M. Hellrigel-Holderbaum and L. Dung. Misalignment or misuse? the agi alignment tradeoff. *arXiv [cs.CY]*, 2025.

[138] J. Zhang, L. Yin, Y. Zhou, and S. Hu. Agentalign: Navigating safety alignment in the shift from informative to agentic large language models. *arXiv [cs.CR]*, 2025.

[139] A. Kasirzadeh and I. Gabriel. Characterizing ai agents for alignment and governance. *arXiv [cs.CY]*, 2025.

[140] A. Tiwari. Conceptualising the emergence of agentic urban ai: from automation to agency. *Urban Inform.*, 4(1), 2025.

[141] C.J. Costa and J.T. Aparicio. Exploring the societal and economic impacts of artificial intelligence: A scenario generation methodology. *arXiv [cs.CY]*, 2025.

[142] V. Lakkamraju. Agentic ai in human-ai collaboration frameworks. *ESP Journal of Engineering & Technology Advancements*, 5(2):114–129, 2025.

[143] H.P. Zou et al. A call for collaborative intelligence: Why human-agent systems should precede ai autonomy. *arXiv [cs.AI]*, 2025.

[144] U.M. Borghoff, P. Bottoni, and R. Pareschi. Human-artificial interaction in the age of agentic ai: a system-theoretical approach. *Front. Hum. Dyn.*, 7, 2025.

[145] E. Tennant, S. Hailes, and M. Musolesi. Hybrid approaches for moral value alignment in ai agents: A manifesto. *arXiv [cs.AI]*, 2023.

[146] F. Rossi and N. Mattei. Building ethically bounded ai. *Proc. Conf. AAAI Artif. Intell.*, 33(01):9785–9789, 2019.

[147] A. Reuel, A. Hardy, C. Smith, M. Lamparth, M. Hardy, and M. J. Kochenderfer. Betterbench: Assessing ai benchmarks, uncovering issues, and establishing best practices. *arXiv [cs.AI]*, 2024.

[148] A. Yehudai et al. Survey on evaluation of llm-based agents. *arXiv [cs.AI]*, 2025.

[149] M. Zhuge et al. Agent-as-a-judge: Evaluate agents with agents. *arXiv [cs.AI]*, 2024.

[150] D. Moshkovich, H. Mulian, S. Zeltyn, N. Eder, I. Skarbovsky, and R. Abitbol. Beyond black-box benchmarking: Observability, analytics, and optimization of agentic systems. *arXiv [cs.AI]*, 2025.

[151] Eldar Sultanow, Mohammad Tehrani, Sourav Dutta, William J. Buchanan, and Muhammad Salekh Khan. Quantum agents. *arXiv [quant-ph]*, 2025.

[152] J. Yang et al. Magma: A foundation model for multimodal ai agents. *arXiv [cs.CV]*, 2025.

[153] S. Agashe, J. Han, S. Gan, J. Yang, A. Li, and X.E. Wang. Agent s: An open agentic framework that uses computers like a human. *arXiv [cs.AI]*, 2024.

[154] Kun Huang, Aisha Sheriff, Venkata S. Narajala, and Itamar Habler. Agent capability negotiation and binding protocol (acnbp). *arXiv [cs.AI]*, 2025.

[155] Yoram Bachrach, Peter Key, David Levin, Daniele Nosenzo, Gijs Overgoor, Ariel D. Procaccia, and Moshe Tennenholtz. Negotiating team formation using deep reinforcement learning. *Artificial Intelligence*, 288:103356, 2020.

[156] Petar Radanliev. *The Rise of AI agents: Integrating AI, Blockchain Technologies, and Quantum Computing*. Addison Wesley Professional, Boston, MA, 2025.

[157] Xiaoxue Ma, Chen Lin, Yizhe Zhang, Volker Tresp, and Yunpu Ma. Agentic neural networks: Self-evolving multi-agent systems via textual backpropagation. *arXiv [cs.LG]*, 2025.

[158] Philipp Schmidt and Sophie Loidolt. Interacting with machines: Can an artificially intelligent agent be a partner? *Philosophy & Technology*, 36(3), 2023.

[159] Danai Koutra et al. Towards agentic ai for science: Hypothesis generation, comprehension, quantification, and validation. *ICLR Workshop*, 2025.

[160] Chris Lu, Chen Lu, Richard T. Lange, Jakob Foerster, Jeff Clune, and David Ha. The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv [cs.AI]*, 2024.

[161] Sergio Cervantes, Samantha López, and Juan-Antonio Cervantes. Toward ethical cognitive architectures for the development of artificial moral agents. *Cognitive Systems Research*, 64:117–125, 2020.

[162] D. Thompson. Autonomous ai agents and blockchain interactions: Enabling decentralized autonomous organizations (daos). *Journal of AI-Assisted Scientific Discovery*, 4(2):73–79, 2024.

[163] R. Bovo, K. Ahuja, R. Suzuki, M.D. Dogan, and M. Gonzalez-Franco. Symbiotic ai: Augmenting human cognition from pcs to cars. *arXiv [cs.HC]*, 2025.

[164] J. Samuel, R. Kashyap, Y. Samuel, and A. Pelaez. Adaptive cognitive fit: Artificial intelligence augmented management of information facets and representations. *Int. J. Inf. Manage.*, 65(102505):102505, 2022.

[165] J. Wu, J. Zhu, and Y. Liu. Agentic reasoning: Reasoning llms with tools for the deep research. *arXiv [cs.AI]*, 2025.

[166] Y. Zhuang et al. Self-taught agentic long context understanding. *arXiv [cs.CL]*, 2025.

[167] C. Jeong. A study on the mcp x a2a framework for enhancing interoperability of llm-based autonomous agents. *arXiv [cs.AI]*, 2025.

[168] Q. Li and Y. Xie. From glue-code to protocols: A critical analysis of a2a and mcp integration for scalable agent systems. *arXiv [cs.MA]*, 2025.

[169] Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. Cognitive architectures for language agents. *arXiv [cs.AI]*, 2023.

[170] Octavio J. Romero, John Zimmerman, Aaron Steinfeld, and Anthony Tomasic. Synergistic integration of large language models and cognitive architectures for robust ai: An exploratory analysis. *arXiv [cs.AI]*, 2023.

[171] David Shapiro, Walter Li, Marco Delaflor, and Carlos Toxtli. Conceptual framework for autonomous cognitive entities. *arXiv [cs.HC]*, 2023.

[172] Yifan Nong. Transfer learning in agentic systems: Improving cross-task knowledge application in ai agents. *Research Square*, 2025.

[173] Hongzhi Li et al. A multi-agent framework with automated decision rule optimization for cross-domain misinformation detection. *arXiv [cs.AI]*, 2025.

[174] R. Ranjan, S. Gupta, and S.N. Singh. Fairness in agentic ai: A unified framework for ethical and equitable multi-agent system. *arXiv [cs.MA]*, 2025.

[175] Vineeth Sai Narajala and Om Narayan. Securing agentic ai: A comprehensive threat model and mitigation framework for generative ai agents. *arXiv preprint arXiv:2504.19956*, Apr 2025. Introduces ATFAA (Advanced Threat Framework for Autonomous AI Agents) and SHIELD mitigation framework for GenAI agents' unique threats.

[176] Raihan Khan, Sayak Sarkar, Sainik Kumar Mahata, and Edwin Jose. Security threats in agentic ai system. *arXiv preprint arXiv:2410.14728*, Oct 2024. Explores privacy and security threats posed by Agentic AI systems with direct database access.