Agentic AI: Autonomous Intelligence for Complex Goals – A Comprehensive Survey

DEEPAK BHASKAR ACHARYA[1], (Senior Member, IEEE), KARTHIGEYAN KUPPAN[2], (Senior Member, IEEE), and DIVYA B[3]

[1]Department of Computer Science, The University of Alabama in Huntsville, Huntsville, AL 35806 USA
[2]JPMorgan Chase, Houston, TX 77082 USA
[3]Department of Electronics & Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal KA, 576104 INDIA

Corresponding author: Divya B (e-mail: divya.ashwin@manipal.edu).

ABSTRACT Agentic AI, an emerging paradigm in artificial intelligence, refers to autonomous systems designed to pursue complex goals with minimal human intervention. Unlike traditional AI, which depends on structured instructions and close oversight, Agentic AI demonstrates adaptability, advanced decision-making capabilities and self-sufficiency, enabling it to operate dynamically in evolving environments. This survey thoroughly explores the foundational concepts, unique characteristics, and core methodologies driving the development of Agentic AI. We examine its current and potential applications across various fields, including healthcare, finance, and adaptive software systems, emphasizing the advantages of deploying agentic systems in real-world scenarios. The paper also addresses the ethical challenges posed by Agentic AI, proposing solutions for goal alignment, resource constraints, and environmental adaptability. We outline a framework for safely and effectively integrating Agentic AI into society, highlighting the need for further research on ethical considerations to ensure beneficial societal impacts. This survey serves as a comprehensive introduction to Agentic AI, guiding researchers, developers, and policymakers in engaging with its transformative potential responsibly and creatively.

INDEX TERMS Agentic AI, Autonomous systems, Human-AI collaboration, Adaptability, Governance frameworks, Ethical AI.

I. INTRODUCTION

A. MOTIVATION AND BACKGROUND

Agentic AIs constitute a qualitative leap in the development of artificial intelligence, defined by their capability to set complex goals in a changing and uncontrolled situation and to pursue them through autonomously managing their resources. Most AI systems, however, were built and operated as tools under supervision with restrictions and definitions provided. These systems are good at doing well-defined tasks within certain boundaries but prominently fail whenever the attempted tasks have no end-state or specific parameters to manipulate. Whereas Agentic AIs can be low-level operative, i.e., goal-directed, even in situations where there are drastic changes and multiple such goals to toggle between.

One of the factors that motivated the design of Agentic AIs is the necessity for tools designed to be able to operate in better yet complex real-world conditions with significant room for flexibility. For example, in disaster relief, healthcare, and cyber security, where proper decisions are needed and the chaos is considerable, the ability to control a situation independently is critical. Agentic AIs don't just assist in human action; they enhance it while taking on tasks that demand high involvement and multi-tasking without constant human intervention. Such a paradigm shift promises to expand the target area of AI from being passive and reactive to focusing on strategic planning, information processing, and problem-solving, enabling a new era once the rights conditions are met.

The impact of Agentic AI [1]–[3] on society is likely to be considerable. As AI becomes embedded in more and more core systems and industries, agentic AI systems would be able to work side by side with humans and take on tasks that can reallocate human effort, increase productivity, and engage in situations where human presence may be undesirable or

dangerous. This change could transform job structures in sectors, allowing for working in concert where AIs perform operational tasks and people deal with more complex and strategic roles.

B. DEFINITION AND SCOPE

In this sense, Agentic AI includes the class of autonomous AI systems [4] that undertake to finish a set of complex tasks that span over long periods of time without human supervision. It learns context and makes decisions. Such systems are designed to operate with a certain level of autonomy, allowing them to traverse changing settings, deal with unexpected situations, and optimize performance over time. Tying them together are the features of autonomy and adaptiveness to cope with task-oriented processes. Unlike classic AI [5], which is rule-based and requires instructions to be performed, and generative AI, which patterns and generates an image, Agentic AI features the best of both worlds.

It is valuable to place agentic AI alongside current AI paradigms to put its boundaries into perspective. For example, classical AI was never focused on image recognition or language translations [6], [7] but instead aimed at achieving very narrow goals.

The first is how Generative AI [8] works. In reverse operations, it combines pieces of information learned through numbers and creates content, such as words or images. On the other hand, Agentic AI surpasses the former two approaches with goal-oriented, input-formed, and adaptable characteristics that allow it to accomplish intricate and multilayered tasks in a period without needing a set of instructions each time.

At this point of the survey, the focus is on why someone should be interested in Agentic AI. From its structural and operational characteristics, why is it based on goals, and how can that advanced type of AI be applied in various fields where its application is unique? In addition, this article also looks into the various practical and moral issues that arise from using systems based on such AI and discusses approaches that deal with safety, transparency, and accountability. The provided definition of Agentic AI allows the authors of this survey to discuss precisely what features of this type of AI system distinguish it from others and aid in its effective and integrated analysis.

C. OBJECTIVES AND CONTRIBUTIONS

This survey paper aims to provide an extensive collection of Agentic AI knowledge and set its boundaries for easier understanding by a wider audience of researchers, developers, and policy-makers. Key contributions of this survey include:

- A systematic review of the main elements that comprise Agentic AI systems with an emphasis on how these systems differ from other commonsense and generative AI systems.
- A comprehensive study of the techniques and concepts used in constructing and assessing Agentic AI, includ-

ing the architectures, learning approaches, and training methods.
- The current and potential uses in different areas, including practical examples of the effectiveness of Agentic AI applications in practice.
- Highlighting engineering problems but not limited to these: goal design and convergence, context adaptation, and limited resources.
- A considers the ethical, societal, and regulatory issues with adopting Agentic AI, including relevant concerns of responsibility, equity, and transparency.
- Recommendations towards future research, advancing suggestions on how the issues of scale, context, and ethics are best integrated into the implementation of Agentic AI.

This paper's expected contributions are more than simply a literature review; rather, they should offer a useful and well-organized background to develop the issues and particulars of Agentic AI. This paper aims to provide important insights into managing operational practices and technological solutions that promote and sustain the development of ethical agentic AI systems.

D. PAPER ORGANIZATION

The remainder of this paper is organized as follows:

- **Section II** Introduces the basic concepts and definitions that help understand Agentic AI as a specific entity in the broader context of artificial intelligence.
- **Section III** Deals with the typical structure of Agentic AI, including independence, flexibility, and the ability to choose what to do.
- **Section IV** Identifies and describes methods of Agentic AI construction – the structural designs, types of learning, and forms of assessment and efficiency measures.
- **Section V** Discusses different sectors where agentic AI has found applications concentrating on applications in industries and cases of joint work with humans.
- **Section VI** Presents a comparative analysis of various systems that exemplify Agentic AI with a focus on the different performance measures and indicators employed in evaluating these systems.
- **Section VII** Explains the technical aspects and difficulties of Agentic AI about aims setting and interaction with the environment aspects.
- **Section VIII** Raises the social, ethical, and governance issues, as concerns responsibility, equity, and compliance with the law.
- **Section IX** Looks at existing systems that facilitate the constructive and responsible implementation of Agentic AI, including oversight and regulation components.
- **Section X** Describes gaps in knowledge and prospected Research and Development work alongside the consideration of refreshing Agentic AI ideas.
- **Section XI** recommends the paper's final remarks, including a conclusion of the findings and a remark that

cross-disciplinary connections are needed to move the field forward in a responsible manner.

This comprehensive survey will provide valuable insights into Agentic AI's current state, future potential, and the challenges that must be addressed to ensure its safe and effective deployment.

II. FOUNDATIONAL CONCEPTS AND DEFINITIONS

A. AGENTIC AI AND ITS ROLE IN THE AI ECOSYSTEM

Within the AI field, Agentic AI serves as a different form of intelligence that can take more autonomously functioning agentic behaviors that are not limited to performing specific tasks or following content-generating algorithms. When viewed in an ecosystem context, Agentic AI stands out because of its purpose, flexibility, and behavior, which enables such AIs to operate almost independently. Rather than following strict guidelines like other robotic AIs, Agentic AI systems are promoted to have rationalism in them, which enables each system [9]to reason and adapt to different scenarios and such circumstances in a sufficient way to accomplish goals. Due to its tendencies to enhance its functions to prepare for any obstacles, Agentic AI has been seen as a potential point of anchor for tasks and goals that require high levels of interactions, for example, autonomous devices, collaborative robots, and interactive decision support systems in the areas of finance and health care.

The increasing demand for systems capable of autonomously handling intricate and dynamic processes has led to growing interest in Agentic AI, particularly in sectors with a scope for AI automation. Though built on basic AI principles, Agentic AI increases the scope of AI achievement by adding the elements of dependence and adaptive independent action. The ecosystem of AI occupies a space between purely reactive and narrowly rule-based AI technologies and broader ideas about AGI, performing an essential function of enabling autonomous decision-making within defined boundaries or structures. This unique position accentuates the ability of Agentic AI to fit in scenarios where the ability to make rapid decisions, manage over-time objectives, and learn on the go are fundamentals to the problem at hand.

B. COMPARISON WITH TRADITIONAL AI

AI, which can be described as 'Agentic' has its core differences when compared to the other advanced types of AIs in terms of autonomy, function, and scope, among others. Such AI systems are integrated into specific tasks like image analysis [10], translation of languages [11], and recommendation engines [12], making them able to perform designated tasks in a highly focused but characteristic narrow manner and scope. They are mostly based on supervised learning approaches on really big datasets where behavior is determined by the input and instructions given to them by people. Thus, traditional AIs are best applied in controlled environments with limited ability to micromanage situations and rather more significant outcomes.

On the other hand, Agentic AI systems can be described as possessing the characteristic of open-endedness where there is no prescription of how the task should be achieved. They work with and adapt to rapidly changing conditions. On the other hand, while conventional AI systems might be accurate, they do not have the situational awareness and goal-directed dynamics inherent in Agentic AI. A case example may include a factory's AI model that is developed to predict equipment failure. As much as the premise is good, AI will not incorporate adaptive shifts in the way it predicts failures due to factors such as changes in the manufacturing schedule or changes in wear patterns across machines. On the other hand, Agentic AI could change its metrics estimation processes depending on the context and adapt both its short and long-term strategies, which is impossible with traditional models.

As shown in Table 1, Agentic AI systems not only adapt based on real-time context but also maintain the flexibility to optimize for complex, long-term goals. These distinctions illustrate Agentic AI's value in scenarios where conventional, rule-bound AI falls short, emphasizing its transformative role in addressing the demands of unpredictable, high-stakes environments.

C. EXPANDED COMPARISON WITH CLASSICAL AGENTS

While classical AI systems are primarily rule-based or supervised-learning models designed for specific tasks, Agentic AI integrates autonomy and adaptability, enabling broader functionality. The differences between these paradigms are best illustrated with examples.

Classical Agents: These agents excel in controlled environments. For example, rule-based financial trading algorithms function effectively when predefined parameters remain constant. However, they struggle with volatile market changes or unpredictable disruptions.

Agentic AI: In contrast, Agentic AI-powered trading systems dynamically adjust strategies based on real-time data, historical trends, and unexpected market shifts, making them more resilient and adaptive.

Reinforcement Learning vs. Language-Model-Based Agents: While reinforcement learning focuses on optimizing cumulative rewards in a specific task environment, language-model-based agents extend this by interpreting complex natural language inputs and interacting with humans seamlessly. For example, an RL agent excels at optimizing gameplay strategies, whereas a language-model-based agent can generate dialogue, interpret rules, and adapt strategies during a live game. Table 2 highlights the key differences.

D. TECHNICAL FOUNDATIONS

The development of Agentic AI systems relies on core algorithms and frameworks that enable goal-directed behavior, contextual adaptation, and autonomous decision-making. These technical foundations incorporate advances in reinforcement learning, goal-oriented architectures, and adaptive control mechanisms.

IEEE *Access*

TABLE 1: Comparison of Traditional AI and Agentic AI

Feature	Traditional AI	Agentic AI
Primary Purpose	Task-Specific Automation	Goal-Oriented Autonomy
Human Intervention	High (Predefined Parameters)	Low (Autonomous Adaptability)
Adaptability	Limited	High
Environment Interaction	Static or Limited Context	Dynamic and Context-Aware
Learning Type	Primarily Supervised	Reinforcement and Self-Supervised
Decision-Making	Data-Driven, Static Rules	Autonomous, Contextual Reasoning

TABLE 2: Comparison of Classical Agents, Reinforcement Learning Agents, and Agentic AI

Feature	Classical Agents	Reinforcement Learning Agents	Agentic AI
Primary Purpose	Fixed-task automation	Reward-driven optimization	Goal-oriented autonomy
Adaptability	Low	Moderate	High
Learning Type	Supervised	Reinforcement Learning	Hybrid, including RAG and Memory
Applications	Static systems	Dynamic environments	Complex, multi-objective tasks

Reinforcement Learning (RL) [13] is central to many agentic systems, as it equips AI models with the ability to learn through trial and error. In RL, agents are trained to maximize cumulative rewards by interacting with an environment, adapting their actions to achieve specific goals over time. This learning paradigm is particularly useful for Agentic AI because it enables systems to continuously refine their strategies based on feedback. As illustrated in Figure 1, reinforcement learning supports "Learning through interaction" and involves a trial-and-error approach to optimize decisions over time.

Goal-Oriented Architectures [14] provide a structural framework for managing complex objectives within Agentic AI systems. Unlike traditional architectures, which often focus on single tasks, goal-oriented architectures enable agents to prioritize and pursue multiple objectives simultaneously. These architectures support a modular structure, where larger goals are broken into manageable sub-goals. In the context of Figure 1, goal-oriented architectures facilitate "Managing complex objectives," allowing agents to approach tasks in structured steps.

Adaptive Control Mechanisms [15] ensure that Agentic AI systems can adjust to changing environments. By incorporating adaptive control, agents recalibrate their parameters in response to external variations, such as data shifts or unexpected disruptions. Techniques like meta-learning, where agents learn to adapt based on prior experiences, enable greater resilience and flexibility. As shown in the flowchart, adaptive control mechanisms provide "Environmental adaptation," allowing agents to maintain optimal performance even under changing conditions.

Figure 1 provides a visual representation of these core frameworks, illustrating how they interact to enable autonomous, adaptable, and goal-driven behavior in Agentic AI systems. Together, these technical foundations equip Agentic AI with the structural and functional capabilities necessary to manage complex, evolving tasks independently, setting it apart from traditional AI systems that rely on rigidly defined parameters and instructions.

By combining reinforcement learning, goal-oriented architectures, and adaptive control, Agentic AI systems achieve a level of autonomy and resilience that allows them to operate effectively in diverse environments. This section establishes the technical basis for understanding the advanced capabilities of Agentic AI, providing a foundation for the methodologies and applications discussed in the following sections.

III. CORE CHARACTERISTICS OF AGENTIC AI
A. AUTONOMY AND GOAL COMPLEXITY

Autonomy is one of the most sought-after qualities that Agentic AI can possess. This is especially needed where complex multi-goal scenarios are involved. Most traditional AI systems focus on completing one task and are programmed with non-complicated input and output requirements to achieve that single goal. In contrast, systems powered by agentic AI can move through multiple needed tasks and shift from one basic one to multiple complex end goals. Such systems come with a certain amount of self-governance where continual supervision of a human being is not mandatory, though at times preferred, and AI agents function independently upon a predetermined or evolving goal structure. One extra step does, however, require acknowledgment; for Agentic AI, Autonomy is not just limited to the completion of a single aim–instead, lesser goals, as well as individual strategies, are substituted to meet larger goals that are long-term in nature. An example can be drawn from autonomous robotics [16], where such an agentic system would demand the complete traversal from point A to point B, with the additional allowance of making diversions around the route, moving constraints such as task hierarchy, time frames, energy expendable ratios, and safety standards. Such an understanding deepens the comprehension of goal complexity, where high-order decision models are used wherein goals guide analysis, planning, and action. Quite clearly, then, as this enables Agentic AI systems to carry out intricate objectives that had been deconstructed into sub-tasks that could fit autonomously into an operational strategy
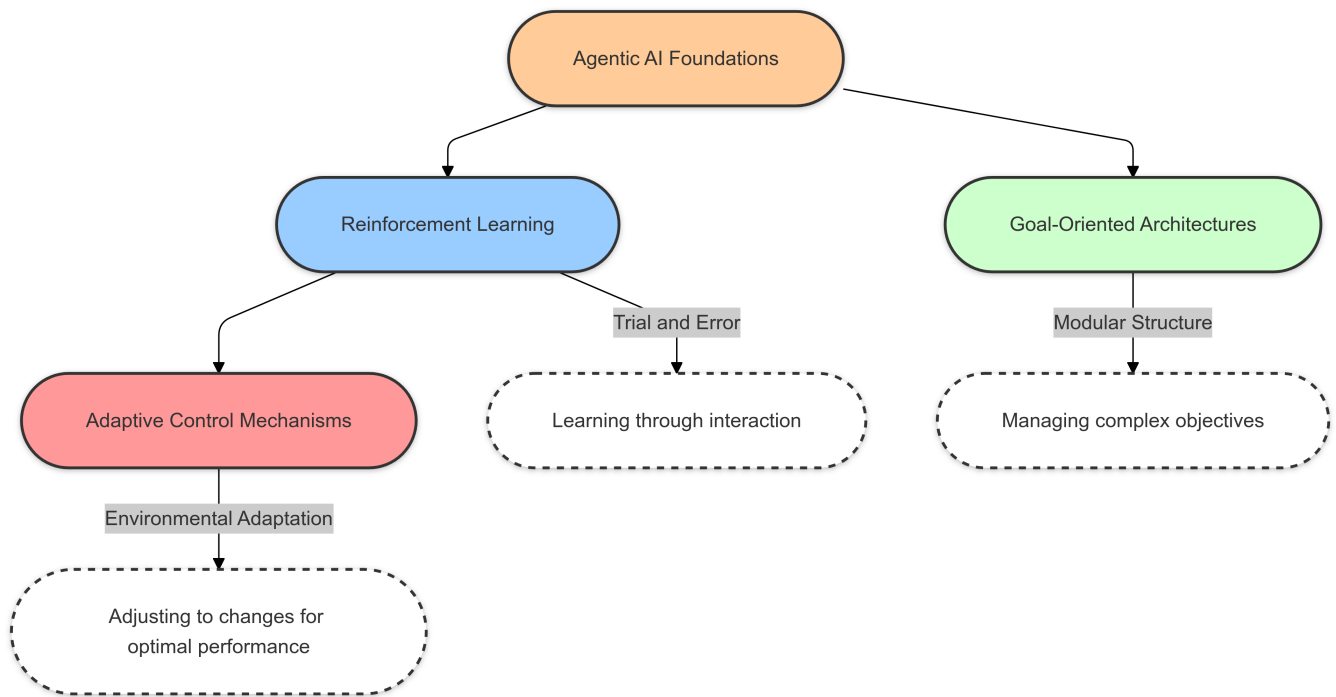
FIGURE 1: Technical Foundations of Agentic AI, illustrating key components: Reinforcement Learning, Goal-Oriented Architectures, and Adaptive Control Mechanisms.

coupled with adjustments.

B. ENVIRONMENTAL AND OPERATIONAL COMPLEXITY

Moreover, Agentic AI's capacity to operate within varied and changing circumstances is yet another characteristic of this AI. Unlike past AI, built for the most optimum functions in a constant and easily predictable environment [17], Agentic AIs integrate all the variability in the real world. This involves adjusting quickly to environmental conditions, data or pattern changes, and even time-worn or newly constituted user demands. In the case of AI agents for self-driving cars, the ideal Agent Did not Photoshop to comply with artificial boundaries contained in traffic laws alone but would learn new road designs and try to understand how the behavior of other drivers would be before deciding how to act in a particular situation when the need arises within a very short period.

To do this, Agentic AI systems, as a general rule, are equipped with means for environmental interaction, on-the-spot data processing, and situational context comprehension. These capabilities allow the system to keep track of and actively participate in the operational parameters that are prone to alterations even at the last minute. For instance, change control [18] in Agentic AI systems is typically embedded within the Agent as reinforcement learning or adaptive algorithms so that the Agent can work optimally regardless of the changing conditions. These characteristic features make Agentic AI fit for use in environments with many dynamics where reactions must be rapid, such as in

Disaster management, healthcare, finance, etc.

C. INDEPENDENT DECISION-MAKING AND ADAPTABILITY

Autonomy and flexibility are core requirements for the Agentic AI to work independently for long hours. In contrast to rule-based systems, which merely do what they are told, Agentic AI has to situate itself in its current context and make decisions as it is working, so it has to learn over time and improve on its behavior. This type of decision-making is commonly done through reinforcement learning or meta-learning when the AI agent gets feedback repeatedly and improves its behavior.

Flexibility allows Agentic AI to act differently in the same scenario and pursue the relevant goals. For instance, in a customer service situation, an agentic AI could change its communication strategies to those that work best with customers' moods to achieve customer satisfaction. This requires prioritizing goals and assessing possible courses of action and their outcomes with respect to the system's goals. Because of flexibility and autonomy in decision-making [19], Agentic AI can reconceptualize its strategies and adapt to new information adopted in the model to operate in a changing environment.

To further illustrate the integration process of Agentic AI into society, Figure 2 outlines the key stages. This includes data collection and preprocessing, the core Agentic AI system's functionality, and its deployment across various industries such as healthcare, finance, manufacturing, and

IEEE *Access*

customer support. The flowchart visually represents the adaptability and independence of Agentic AI in dynamic environments.

D. COMPARATIVE ANALYSIS

Compared to linear systems of regulation, Agentic AI stands out because of its improved ability to remain autonomous, function within an ever-changing context, and cope with multiple goals. In most cases, operative agents are created where the functional area is characterized within rigid boundaries, with unfailing and straightforward preconditions for successful performance. On the other hand, Agentic AI systems are designed to work on complex goals that have yet to be structured and can be defined in a broad context.

As shown in Table 3, traditional agents excel in structured tasks but lack the flexibility required for adaptive, goal-oriented tasks in complex environments. Agentic AI advances these systems by introducing a high degree of autonomy and adaptability, enabling the agent to interact with and respond to its environment in ways that extend beyond simple rule-following. This comparative analysis highlights the distinctive capabilities of Agentic AI, positioning it as a transformative approach in fields where independent, goal-driven, and context-aware behavior is essential.

IV. METHODOLOGIES IN AGENTIC AI DEVELOPMENT
A. ARCHITECTURAL APPROACHES

Architectural approaches in Agentic AI typically involve modular and hierarchical designs that enable the system to manage complex goals and adapt to dynamic environments. Common architectures include multi-agent systems (MAS), hierarchical reinforcement learning (HRL), and goal-oriented modular architectures.

- **Multi-Agent Systems (MAS)**: MAS [20] divides tasks among multiple autonomous agents that collaborate or compete to achieve a common goal. This architecture is particularly useful in scenarios where complex goals can be decomposed into smaller tasks that individual agents can handle.
- **Hierarchical Reinforcement Learning (HRL)**: HRL [21] structures decision-making hierarchically, where high-level agents define sub-goals, and low-level agents execute them. This approach is effective for managing tasks with multiple levels of complexity.
- **Goal-Oriented Modular Architectures**: These architectures [22] organize agent functions into modular components, where each module specializes in specific aspects of the task. Such modularity enables flexibility and scalability, allowing the agent to handle different tasks by reconfiguring modules as needed.

B. LEARNING PARADIGMS

Agentic AI relies on several learning paradigms, each suited to different types of tasks and goals. The main paradigms used in agentic systems are supervised, unsupervised, and reinforcement learning. Table 4 provides a comparison.

C. ADVANCEMENTS IN METHODOLOGIES

Recent advancements in Agentic AI methodologies have focused on key capabilities essential to modern agent design. These include reasoning and planning, tool use, memory mechanisms, Retrieval-Augmented Generation (RAG), and instruction fine-tuning [23].

Reasoning and Planning: These frameworks enable agents to anticipate outcomes, prioritize tasks, and adapt strategies dynamically. They are critical for managing complex, multi-objective tasks in evolving environments like disaster management and autonomous navigation.

Tool Use and Integration: Agents equipped with the ability to interact with external tools and APIs can perform computations, retrieve real-time data, and simulate scenarios, significantly enhancing decision-making processes.

Memory Mechanisms: Episodic and semantic memory models allow Agentic AI systems to retain contextual information, improving their ability to recall past interactions and optimize ongoing tasks.

Retrieval-Augmented Generation (RAG): RAG empowers agents to retrieve external knowledge dynamically, enhancing the relevance and context of their outputs. This capability is particularly significant in conversational agents and real-time decision-making systems.

Instruction Fine-Tuning: This process ensures that agents understand and execute nuanced directives, enabling them to perform multi-step tasks with high precision and adaptability.

D. TRAINING AND EVALUATION TECHNIQUES

Training Agentic AI systems requires techniques that allow agents to learn from interactions with complex environments. Common training techniques include simulation-based training, curriculum learning, and multi-task learning.

- **Simulation-Based Training**: Simulations give students a safe context to investigate numerous situations without real-world consequences [24]. This is very effective in reinforcement learning [25] as it allows agents to design policies that are transferable to the real task.
- **Curriculum Learning**: Structure tasks in increasing order of complexity so that an agent develops basic skills that can be built upon in new, more complex tasks [26]. Such a structure of progressively more complex tasks is fundamental in a multi-goal-oriented environment.
- **Multi-Task Learning**: In multi-task learning [27], agents acquire the capabilities to perform several tasks at once, which expands their generalization abilities over multiple goals, tasks, and scenarios. The problem is of utmost interest in designing agentic AI systems that are supposed to tackle multiple objectives in parallel.

Evaluation techniques for Agentic AI often include metrics such as task success rate, adaptability, resource efficiency, and long-term goal achievement. Table 5 provides an overview of these training and evaluation techniques.
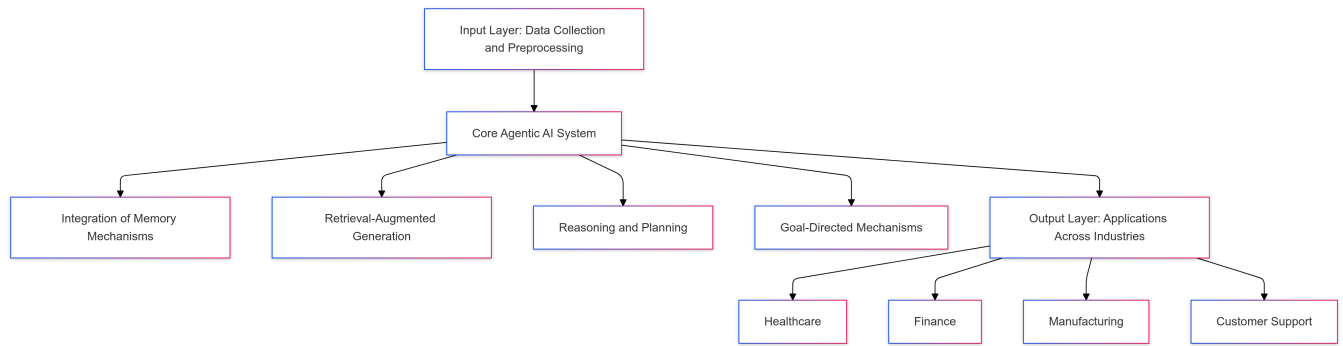
IEEE *Access*

FIGURE 2: Integration Process of Agentic AI into Society.

TABLE 3: Comparison of Traditional Agents and Agentic AI

Feature	Traditional Agents	Agentic AI
Autonomy	Limited, often requires human supervision	High, can operate independently over time
Goal Complexity	Single or limited goals	Multiple, complex, and dynamic goals
Environmental Adaptability	Low, suited for static environments	High, designed for dynamic environments
Decision-Making	Rule-based, predefined actions	Contextual, autonomous reasoning
Learning Approach	Mostly supervised or rule-based learning	Reinforcement and self-supervised learning

TABLE 4: Comparison of Learning Paradigms for Agentic AI

Learning Paradigm	Description	Suitability for Agentic AI
Supervised Learning	Learning from labeled data	Useful for specific tasks but lacks adaptability for complex goals
Unsupervised Learning	Identifying patterns in unlabeled data	Provides insights but requires reinforcement learning for goal-directed behavior
Reinforcement Learning	Learning through trial and error with rewards	Highly suitable due to adaptability and goal-oriented focus

TABLE 5: Training and Evaluation Techniques for Agentic AI

Technique	Description	Usage in Agentic AI
Simulation-Based Training	Training in controlled, simulated environments	Developing strategies in a risk-free setup
Curriculum Learning	Gradual increase in task complexity	Stepwise learning from basic to complex tasks
Multi-Task Learning	Learning multiple tasks simultaneously	Enhances generalization and multi-goal management
Evaluation Metrics	Measures such as success rate, adaptability, and efficiency	Assess agent's performance and goal achievement

E. TOOLS AND FRAMEWORKS

Developing Agentic AI requires specialized tools and frameworks that support reinforcement learning, simulation, and multi-agent system development. Tools such as OpenAI Gym, Unity ML-Agents, TensorFlow Agents, and Rasa offer platforms for developing, training, and evaluating Agentic AI systems across various applications. Each tool provides unique capabilities, from reinforcement learning environments to multi-agent simulations, enabling researchers and developers to experiment with different architectures and training techniques. Figure 3 overviews popular tools and frameworks of the different methodologies of **Agentic AI Development Methodologies**, which are employed in the development and

enhancement of such autonomous AI systems that can work towards achieving some higher-order goals. The approaches are further simplified into four categories:

1) **Architectural Approaches**: This branch deals with the different architectures that autonomous agents use to accomplish their aims. Among key methods are *Multi-Agent Systems*, which consist of agents working in collaboration and competition with others; *Hierarchical Reinforcement Learning*, which focuses on dividing tasks into a hierarchy to help with the learning process; and *Goal-Oriented Modular Architectures* which entail the configuration of the system based on goal or module specification.

IEEE *Access*

2) **Learning Paradigms**: This part emphasizes the related study of autonomous agents derived from different machine learning sources. These consist of *Supervised Learning*, which is based on the use of labeled data; *Unsupervised Learning*, in which the model seeks relations in unlabeled data; and *Reinforcement Learning*, which is trial and error, wherein the agent is being trained to carry out temporal decision-making.

3) **Techniques of Training and Evaluation**: In this case, the focus is on methods that are designed for the training and performance assessment of agentic AI. Techniques include *Simulation-Based Training* which allows for controlled environments to train a number of agents, *Curriculum Learning* in which agents can perform a series of simple tasks and more difficult ones afterward; and *Multi-Task Learning* which allows agents to carry out multiple tasks at a time. Training or evaluating agents in the simulated environment is done using OpenAI Gym [28] and Unity ML-Agents [29].

4) **Computational Tools and Frameworks**: The last subsection discusses more tools or computational frameworks pertinent to the building process of autonomous AI systems. Tools such as Reinforcement Learning (RL) algorithms can be implemented using *TensorFlow Agents*, *PyMARL*, a multi-agent RL library, and *Rasa* [30], a framework for conversational agents.

The structural breakdown outlined in the paragraphs above portrays the range of methodologies and tools that can be used to create agentic AI systems and how they can also enhance the adaptability, efficiency, and functionality of autonomous systems in complex environments.

The methodologies discussed in this section encompass diverse applications and ethical considerations. Table 6 summarizes key methodologies in Agentic AI, their practical applications across domains, and the associated ethical challenges. This summary provides a concise reference for understanding the core techniques driving Agentic AI development.

V. APPLICATIONS OF AGENTIC AI
A. INDUSTRIAL APPLICATIONS
Agentic AI can potentially revolutionize multiple industries, including healthcare, finance, education, and manufacturing. In healthcare, for example, Agentic AI can be used to analyze incoming patient data, identify abnormal patterns, and alert corresponding medical personnel to damage. AI-based devices could, for example, prevent the delay of vital diagnoses by monitoring patients' key health indicators and notifying them when the situation worsens. In finance, for instance, Agentic AI algorithms can assist in making investment transactions, detecting fraudulent activities, and providing tailored investment solutions. They are capable of evaluating the situation on the exchange, independently making decisions regarding the purchase or sale of securities, and adjusting their strategies to changing conditions in real time, which significantly improves the quality of managerial performance and reduces the participation of human beings [1].

In education, intelligent tutoring systems utilizing Agentic AI technology assist learners by tailoring educational content [31] to their needs and addressing their progress and requests. Such an approach leads to enhanced academic performance and decreased pressure on teachers because repetitive tasks such as grading and finding appropriate materials are done automatically. In manufacturing, Agentic AI is applied to predictive maintenance where the state of the machines is assessed, future breakdowns are anticipated and maintenance is carried out when necessary without the involvement of people so that production runs smoothly with minimal delay. It is clear that once implemented, Agentic AI enables industries to operate with greater effectiveness, flexibility, and scalability.

B. HUMAN-AI COLLABORATION
Agentic AI extends the range of human productivity in collaborative and cognitive domains [32]. In such knowledge-intensive ventures as legal or research, it can support professionals by automatically condensing documents, pulling up relevant papers, or performing professional background investigations so the users can concentrate on more complex aspects of the work. For example, in legal practice [33], Agentic AI can examine a corpus of legal text, remember essential documents and help lawyers retrieve cases with similar laws.

In the creative industries, Agentic AI can draft text, develop design concepts, or make creative alterations based on earlier edits or client input. Agentic AI tools are expected to be able to reduce such needless tasks and improve the productivity of content development. Also, in customer services, Agent AI can take care of simple questions, offer assistance, and route complicated problems to human agents, enhancing response time and client satisfaction. This seamless integration of human-AI collaboration enables workers to spend their time in a more strategic and creative workspace. At the same time, Agentic AI takes care of the operational aspects.

C. ADAPTIVE SOFTWARE SYSTEMS
There is a gradual move towards Agentic AI in adaptive or "living" software systems [34], which can change their features without typical limitations and dynamically modify their functionalities depending on environmental evolvement. These variables make the system reconfigure itself rapidly or automatically to ensure that self-learning improves every time around usage. These include real-time updates of suggestions by people as a function of the recommendations and updates of suggestions based on changes in the evolution of the user recommendations dynamically.

Other cases include automated activities in the smart house where, because of the powered agentic AI, boarders can change the lighting, temperatures and security protocols for inbuilt AI to note the behavior of the users. In project development processes, for instance, Agentic AI project management bots can focus on task sequencing, workload distribution, and task timeline complexity to dynamically alter timelines [35] where necessary, which enables the users
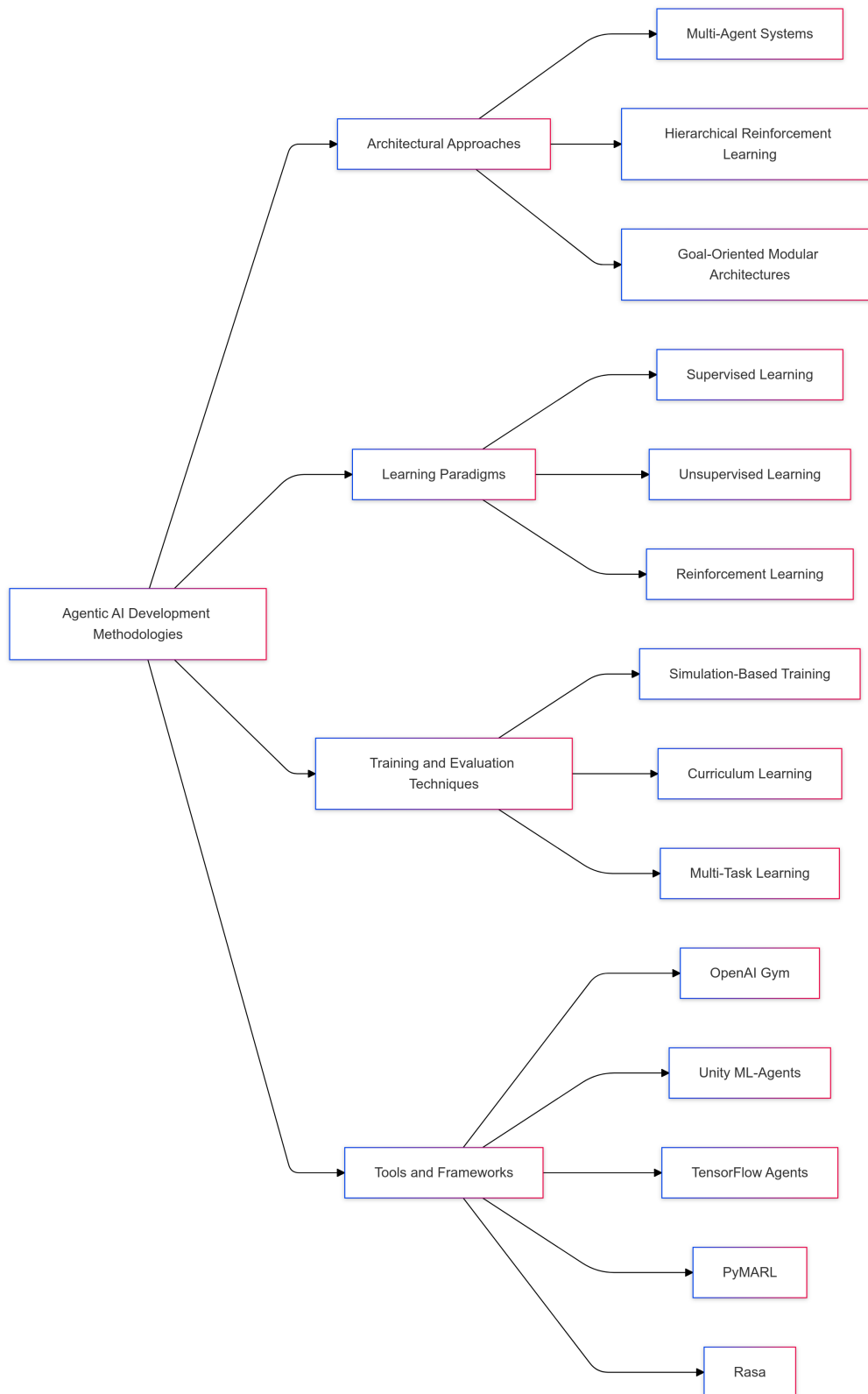
IEEE *Access*

FIGURE 3: Overview of Agentic AI Development Methodologies, including Architectural Approaches, Learning Paradigms, Training Techniques, and Tools.

TABLE 6: Summary of Key Methodologies, Applications, and Ethical Challenges

Methodology	Applications	Ethical Challenges
Memory Mechanisms	Patient monitoring, adaptive healthcare systems	Concerns over data privacy and security
Retrieval-Augmented Generation (RAG)	Real-time knowledge retrieval, adaptive chatbots	Risk of inaccuracies and bias in retrieved data
Reinforcement Learning	Autonomous driving, smart navigation systems	Goal misalignment, safety risks, and overfitting
Instruction Fine-Tuning	Customer service automation, multi-step task execution	Lack of transparency in decision-making processes

to suit the various changes in the project quite effectively. Such adaptive software applications reduce human-in-the-loop requirements and enhance the usability and functionality of the system.

D. EMERGING APPLICATION AREAS

With dynamic patient needs demanding continuous responsiveness, there are new targeted use cases for applying Agentic AI in specific spheres. Such include personalized medicine [36] , where an Agentic AI system could manage chronic patients by overseeing their patient history, sending medication intake reminders [37], and changing treatment recommendations based on other health indicators. Such systems would provide individualistic care-management protocol and even monitor for early indicators of other progressing health conditions, especially for aged people likely to require maximum attention.

For literature creation, Agentic AI is expected to acquire new roles in generating content automatically, targeting wider audiences and meeting precise parameters in content creation. For instance, in marketing, Agentic AI systems would send customized emails and adverts based on user activity and generate content for the adverts in the first place. Furthermore, in the context of self-regulatory research, Scientists using Agentic AI can accomplish the objectives of a literature search, developing new lines of thought, and even creating research designs. Faster research periods could be witnessed in the fields of drug development [38] research or climate change research. Through expansion in these new directions, Agentic AI shows more ability to penetrate high-value markets requiring customized, dynamic and self-servicing applications.

Table 7 provides an overview of Agentic AI applications across various domains, illustrating the diverse range of tasks and contexts where autonomous, goal-directed systems can enhance operations and create value. From healthcare to personalized marketing, the versatility and adaptability of Agentic AI open possibilities for innovative solutions across industries.

E. SCENARIOS DEMONSTRATING ADAPTABILITY

Disaster Management: An Agentic AI system deployed in disaster management autonomously analyzes real-time environmental data during a flood. It reallocates resources, such as rescue teams and medical supplies, to areas most in need, adjusting strategies dynamically based on changing weather conditions and incoming data.

Customer Support: In an e-commerce setting, an Agentic AI chatbot adapts its tone and problem-solving strategies

based on real-time sentiment analysis of customer interactions, improving user satisfaction.

Healthcare Monitoring: A hospital-based Agentic AI system detects patterns in patient vital signs, predicts potential complications, and autonomously notifies healthcare providers, enabling timely interventions without human input.

These scenarios highlight the dynamic and goal-directed capabilities of Agentic AI in real-world applications.

VI. COMPARATIVE ANALYSIS OF AGENTIC AI IMPLEMENTATIONS

A. COMPARISON METRICS

To comprehensively assess and triangulate Agentic AI deployments, it is necessary to identify metrics commensurate with their output, flexibility and leverage. The following metrics are prevailing in the field:

- **Adaptability**: This metric considers the AI system's propensity or capability to react to environmental modifications that are prompt or abrupt actively. A high adaptability [39] score means the AI system can undergo performance losses due to novel conditions like shifts in statistical measures, but not to the extent of incurring significant losses.
- **Goal Achievement Efficiency**: This metric reflects on the AI's inclination or ability to achieve its goals while using minimal resources such as time and person-hours. This metric becomes critical in applications where the provision of the AI model's determinism has strong consequences on the feasibility of the application.
- **Learning Rate and Convergence**: Measures the time the AI takes to learn and attach to a specific task. Quick learning rate and quick convergence are always preferred as they enable the AI to work effectively in dynamic environments requiring continual learning [40], [41].
- **Robustness and Resilience**: This measure determines the system's performance level when its parameters change or during disturbances. Robustness [42], [43]is one attribute that enables the effective design of Agentic AI, which can operate in several vitiating circumstances, such as unexpected scenarios well deepened in the healthcare and autonomous driving sectors.
- **Scalability**: It refers to the system's qualities, which must not change as the range or complexity of tasks changes. This is an important characteristic in industries with significant volumes of data or operational processes, such as the case with finance or manufacturing [44].

TABLE 7: Overview of Agentic AI Applications Across Domains

Domain	Application	Example Use Cases
Healthcare	Autonomous patient monitoring and early warning systems	Chronic disease management, elderly care
Finance	Algorithmic trading and fraud detection	Real-time market analysis, personalized financial advice
Education	Intelligent tutoring and personalized learning paths	Adaptive learning platforms, automated assessments
Manufacturing	Predictive maintenance and quality control	Machinery health monitoring, optimization of production processes
Human-AI Collaboration	Document summarization and creative co-design	Legal document review, collaborative design tools
Adaptive Software Systems	Real-time recommendation engines and adaptive cybersecurity	Personalized recommendations, autonomous threat detection
Emerging Areas	Personalized healthcare and adaptive content generation	Real-time health monitoring, tailored marketing campaigns

- **User Satisfaction and Human-AI Collaboration Efficiency**: User satisfaction with Agentic AI systems dependent on human interaction [45], [46] is a key performance indicator. This metric examines how language, particularly AI's involvement, translates into productivity, usability, and support quality for human doers.

Table 8 summarizes these metrics, providing a basis for evaluating Agentic AI implementations across various applications.

B. CASE STUDIES

The case studies presented in this subsection illustrate the practical uses and performance of agentic AI systems in several scenarios. These examples highlight the capabilities of Agentic AI in accomplishing complex goals in a changing environment.

- **Healthcare Monitoring and Diagnostics**: A healthcare monitoring system based on Agentic AI developed for healthcare monitoring can independently identify a patient's state to be deteriorating through steady tracking of patients' vital signs. This system has been tested in a hospital to enhance the timeliness of patient health intervention [47]. It is becoming critical as it improves the time taken to respond to various health issues. Thanks to the system's versatility, it can operate under different patient conditions and problems, presenting high robustness and reliability.
- **Financial Market Analysis and Algorithmic Trading**: In the finance industry [48], [49], an agentic AI system was used for trading strategies and real-time market optimization with minimum human interaction. To minimize strategies, the AI adjusts them based on past and current data to improve trading outcomes in times of great market volatility. This case study showcases the efficacy and flexibility of agentic AI in fast-paced and high-risk settings as minor enhancements in speed and accuracy in decision-making lead to significant financial returns.
- **Autonomous Customer Support in E-Commerce**: On

an e-commerce platform, a customer service e-agent made possible by Agentic AI technology begs unassisted attention [50], knowing full well the needs of the individual user owing to their past behavior and known preferences. Over time, the AI agent adapts to past interactions and input to the system to improve its responses to questions; this is a classic illustration of a good human-AI relationship [51]. Customers have generally been more satisfied with the individualized and context-responsive adaptive support, and the case demonstrates the merits of incorporating Agentic AI in customer-facing initiatives.

- **Smart Manufacturing and Predictive Maintenance**: On a plant floor, agentic AI computes the predicted time to failure of the machine and the remaining useful life and when to perform maintenance [52] activities so as to maximize operational availability. This system leverages data from a cluster of machines to proactively predict future failures and optimize the distribution of resources, which, in turn, drives the production process. From example deployments, the robustness and scalability of the Agentic AI system have shown its ability to function efficiently in large-scale and data-driven operations.

These case studies emphasize the applicability of Agentic AI technologies in various industries, from healthcare to financial and manufacturing. They all demonstrate various aspects of these technologies in practical use that can be scalable for different levels of complexity issues.

C. BENCHMARKING

Benchmarking is one of the most important procedures for assessing Agentic AI performance and its standard model. This step is important as it allows model researchers and developers to model. Some several standardized benchmarking datasets and environments have been developed and are now used to carry out Agentic AI:

- **Healthcare Datasets (e.g., MIMIC-III, PhysioNet)**: Health care datasets [53], [54] are applied in the training and evaluating systems for AI designs for patient

IEEE *Access*

TABLE 8: Comparison Metrics for Agentic AI Implementations

Metric	Description
Adaptability	Ability to respond effectively to dynamic changes in the environment
Goal Achievement Efficiency	Efficiency in achieving objectives with minimal resources
Learning Rate and Convergence	Speed at which the AI learns and stabilizes in a given domain
Robustness and Resilience	Ability to maintain performance under varying or adverse conditions
Scalability	Performance consistency as task complexity or data scale increases
User Satisfaction and Human-AI Collaboration Efficiency	Impact on productivity and satisfaction in human-AI collaborative tasks

monitoring, clinical diagnostic prediction, and clinical decision support. The claim that an Agentic AI system is doing an excellent job in Health care benchmarks shows that it is possible to make AI systems that can precisely work with such sensitive and life-critical information.

- **Financial Data (e.g., Yahoo Finance, NASDAQ historical data)**: Fundamental datasets of respondents with sales records [55] such as the historical stock prices are very useful for an Agentic AI, capable of predictive analytics and algorithmic trading. These datasets can also be used to rate the performance regarding prediction ability, robustness in fluctuating volatile markets, and profitability of trades.
- **Autonomous Driving Simulators (e.g., CARLA, OpenAI Gym)**: The environments allow for driving testing for autonomous navigation tasks and for AI in vehicles. AI systems' rate of learning, as well as their adaptive capability, are also tested under such environments [56], [57]. Simulators are especially helpful in tests that demand systems to perform similar tasks of safety and decision-making under various conditions.
- **Customer Service Datasets (e.g., MultiWOZ, Amazon Customer Reviews)**: Efficient and personalized AI-augmented customer interactions are evaluated in the dimension of User and Systems or Support collaboration where this set of data assists in the estimation of such measures. The ability to perform well on these benchmarks is a direct indicator of the agent's ability to attend to user concerns as well as hold up to complex inquiries.
- **Manufacturing and IoT Datasets (e.g., NASA Prognostics Data Repository)**: This data source consists of sensor data from industrial equipment that is utilized to train the AI application for predictive maintenance [58], [59]. Evaluation of performance on these data sets seeks to assess the prediction of failures, resource allocation, and how effectively operations respond to different conditions.

Table 9 summarizes these benchmarks, highlighting the diverse application areas and the specific focus of each dataset or environment.

This comprehensive benchmarking approach allows Agentic AI systems to be evaluated and refined based on performance in key application areas, ensuring that they meet industry standards for efficiency, accuracy, and adaptability in real-world scenarios.

D. CRITICAL EVALUATION OF EXISTING IMPLEMENTATIONS

Existing Agentic AI implementations highlight its transformative potential and the challenges associated with its deployment.

Successes

- In healthcare, Agentic AI systems have successfully monitored patients, identified early warning signs, and suggested interventions in real time. These systems enhance healthcare delivery, particularly in high-demand scenarios.
- In finance, algorithmic trading powered by Agentic AI has demonstrated superior performance during volatile market conditions by dynamically adjusting trading strategies.
- In manufacturing, predictive maintenance systems have reduced downtime by proactively anticipating equipment failures and scheduling maintenance.

Limitations

- Healthcare systems often require extensive data preprocessing and struggle with data heterogeneity.
- Financial AI models may overfit historical trends, limiting adaptability to novel events.
- Manufacturing systems face integration issues with legacy equipment and scalability constraints.

Lessons Learned

- Effective Agentic AI implementations require robust data pipelines and high-quality training datasets.
- Incorporating feedback mechanisms and human oversight improves performance and ensures ethical compliance.
- Hybrid models combining classical and agentic paradigms often yield superior results, as seen in complex, multi-stakeholder environments.

Addressing these limitations while leveraging lessons learned will be critical for advancing Agentic AI across diverse domains.

VII. TECHNICAL CHALLENGES AND LIMITATIONS
A. GOAL ALIGNMENT AND COMPLEXITY

A critical consideration in the design of any system is how the autonomously set goals of the AI will agree with the diverse

IEEE *Access*

TABLE 9: Benchmarks and Datasets for Evaluating Agentic AI Implementations

Application Area	Benchmark Dataset/Environment	Evaluation Focus
Healthcare	MIMIC-III, PhysioNet	Patient monitoring, diagnostics, clinical decision support
Finance	Yahoo Finance, NASDAQ historical data	Predictive analytics, algorithmic trading, adaptability to volatility
Autonomous Driving	CARLA, OpenAI Gym	Navigation, safety, decision-making in variable conditions
Customer Service	MultiWOZ, Amazon Customer Reviews	User satisfaction, response efficiency, personalization
Manufacturing and IoT	NASA Prognostics Data Repository	Predictive maintenance, resource optimization, operational adaptability

perspectives of social morals and human users' objectives. In more simplified terms, it is not like how AI generally operates today, where there are set commands to follow. In contrast, agentic AI benefits from possessing a number of highly complex goals, and these goals may evolve over time. There is a straightforward design problem here because people's goals do not always lead to intended outcomes, which could, for example, call for developing inappropriate goals, techniques, or strategies.

However, the question of goal misalignment is sometimes more easily addressed than when the central goals of the projects are multi-dimensional and contingent on the project context. For example, in the case of medical ethics, an AI agent operating for a maximum recovery rate of the patients might be focusing on those strategies that work fast and give results even in the short term but may not be the best thing in the long term. Furthermore, these issues are intertwined with ethical goals and values structures, which are difficult to articulate and integrate within the goal system due to cross-cultural or industry particularities. Addressing this problem is a requirement because other researchers are investigating aligned frameworks like value alignment and inverse reinforcement learning, wherein the etic systems are congruent with the emic reward systems. However, these explanations are still at the very basic understanding and need a lot of work to cope with complex changes in human goals.

B. ENVIRONMENTAL AND SITUATIONAL ADAPTABILITY

As mentioned earlier, agentic AIs tend to be used in dynamic and highly complex environments. This adaptability challenge relates to adapting to real-world conditions, which may change within a narrow time frame without human intervention. In practical terms, it is quite challenging to be highly adaptive, as many dynamics in the real world are highly unpredictable such as market trends in financial investment, epidemiological trends in health care, or even journey events in autonomous cars.

For most agentic AI systems, technological or contextual barriers require action even when there is incomplete information, increasing performance reliability ambiguities. For instance, in the case of autonomous driving, the agentic AI may not be familiar with the surrounding traffic conditions or weather patterns such as snow or heavy rainfall, necessitating

the need to adapt in a safe manner to maintain efficiency. While meta-learning and reinforcement learning can help increase the adaptability of agents by allowing them to learn from their past, the limitations lie within the approaches as well, as adaptation and robustness are difficult goals to achieve. Furthermore, implementing machine learning models based on generalized processes in such complex environments usually requires a high computing time, which is only sometimes practical.

C. RESOURCE CONSTRAINTS

Agentic AIs are complex systems that require lots of computation and energy resources for their training and deployment phases in navigable spaces. One is reinforcement learning, which greatly relies on simulation and data processing and consequently increases the cost and time of training when used. Milling out time for resource demands in real-time decision-making circumstances is a big challenge in applications such as the finance industry and autonomous vehicles.

In addition, hardware is another resource for Agentic AI, as these systems usually require specialized facilities when deployed in real-world environments. For instance, if an autonomous drone or robot is to be used, it would require high-performing GPUs, low-latency sensors, and a vital energy source to enable prompt decision-making. In addition, the proliferation of agentic systems, especially those based on centralized control and monitoring, can put an increasing strain on data storage, processing and network bandwidth resources. Coupled with these challenges are the hardware optimization methods and energy-efficient algorithms that will be required to implement Agentic AI systems largely and cost-effectively efficiently.

D. SCALABILITY

With the complexity of design increasing, scaling improves the system's performance. Some applications of Agent AI, whether in smart cities, healthcare, or financial services, are massive in scale, requiring a higher AI-to-agent ratio, multifaceted tasks, or huge amounts of data to be handled. However, it is nontrivial to ensure that such systems can be scaled up to a higher level without losing some key performance metrics, as it usually involves a multitude of concurrent tasks and data sources.

Multiple agents or components design and their interaction presents a major challenge regarding scalability. For example, in a smart city, there may be an AI that manages traffic, energy, and waste collection. Different architectures must be capable of getting required subsystems to work together and perform complex tasks simultaneously. Besides, as most architects understand, the scaling of Agent AI also magnifies problems like goal dependency, resource usage, and the ability to act in plenty of scenarios. Today, decentralized control, federated learning, and hierarchical structures are being used or researched to improve scalability. However, the problem remains as scaling out seamless techniques is a technological limitation related to AI operating at a larger scale.

Table 10 provides a summary of the primary technical challenges and limitations faced in the development and deployment of Agentic AI. Addressing these challenges is essential to advancing the field and ensuring that Agentic AI systems can function effectively and responsibly in real-world scenarios.

VIII. ETHICAL, SOCIAL AND GOVERNANCE IMPLICATIONS

A. ACCOUNTABILITY AND RESPONSIBILITY

In light of the decision-making independence such systems display, understanding accountability and responsibility in Agentic AI systems is a complex challenge. With the more traditional AI systems, tools fall under the purview of people – the developer, the operator, or the user and the responsibility rests with the person who uses the tool. With Agentic AI, the question of accountability is more contentious due to the nature of an independent-acting AI. In situations where autonomous AI makes a decision that results in a negative outcome, the question of responsibility is often misplaced; is it the developer, the service provider who deploys the AI, or the AI system.

This particular issue comes out clearly in areas such as finance and healthcare, where such systems deal with high-stakes decision-making with high repercussions. There needs to be a more significant gap between the responsibility and liability attribution frameworks and the behavioral intricacies present in Agentic AI systems. Therefore, Regulations and possibly new legal structures must be invoked to delineate responsibility or accountability, especially in environments where multiple stakeholders are participating in the execution of the system, which is most of the time.

B. BIAS, FAIRNESS, AND TRANSPARENCY

It is also known that agentic AI systems may replicate and exacerbate biases in the training data available to them. The notion of bias is especially concerning regarding the use of Agentic AI in hiring, policing, and lending practices. Not only must these biases be acknowledged, but they must also be addressed through appropriate data management, careful algorithm deployment, and active bias reduction programs. Still, it is essential to note that addressing bias and mitigating it in autonomous systems is a much more difficult task than

in traditional AI because of the nature of the tasking such systems are designed to perform.

Another important aspect that also helps enhance fairness and trust in Agentic AI [60] systems is transparency. These systems' operations must be explained through XAI approaches, and there must be an effort to make transparent decision-making processes for end users and relevant stakeholders. This exposure can help effectively check an AI's actions and create a framework for tackling bias or injustice. However, explainability is quite challenging to attain fundamentally, and a more significant deal of it is supposed to be achieved in cases of agentic AI systems, which tend to use deep reinforcement learning models. This area of concern has always been finding the right balance between structural complexity and transparency.

C. PRIVACY AND SECURITY ISSUES

The agents can perform their functions effectively in different environments by relying on relevant information, often private in domains like healthcare and finance, that are already sensitive areas. Such reliance on personal information invites privacy concerns as data can be mishandled and even accessed illegally, leading to a large breach of user privacy. Additionally, in this context, the autonomous ability of the Agentic AI can pose a challenge as the data usage can be difficult to track, thus increasing the likelihood of privacy abuse.

There is also a security challenge in that cyber-attacks can threaten these systems, undermining their intended purpose and exposing weaknesses in the system in place. If agentic AI systems are hacked, they could bring forth a lot of destruction, primarily when the systems govern critical infrastructure or sensitive tasks that need to be decided within that context. There is a great need to maintain a robust and holistic approach towards cybersecurity, as it is required to secure the information that such systems depend on and their functionality. While techniques such as differential privacy and secure multi-party computation may seem promising, privacy and security threats still exist, so implementation should be done carefully to retain the required functional privacy.

D. REGULATORY AND LEGAL PERSPECTIVES

The regulatory and legal perspective on Agentic AI is still a work in progress as regulators try to streamline the unique challenges that autonomous systems present. The AI regulations as they stand today mainly address privacy, transparency, and accountability issues but may not necessarily tackle the challenges posed by Agentic AI's autonomy. For instance, the like of the EU's General Data Protection Regulation (GDPR) [61] imposes stringent obligations with respect to data and user consent, but these requirements may not be adequate when managing AI systems that operate in real-time with autonomous decisions.

Policymakers [62] are also investigating other frameworks, particularly for high-risk AI systems. Some approaches call for an "AI responsibility chain," which can help stakeholders determine the accountability of each party of the AI systems.

TABLE 10: Technical Challenges and Limitations in Agentic AI

Challenge	Description
Goal Alignment and Complexity	Ensuring the AI's goals align with human objectives and ethical standards, particularly in complex or context-dependent tasks
Environmental and Situational Adaptability	Adapting to dynamic environments and making decisions based on incomplete or uncertain data
Resource Constraints	Balancing computational and energy demands with the need for real-time performance in complex applications
Scalability	Maintaining performance as complexity, agent count, and data volumes increase in large-scale applications

In contrast, others recommend a mandatory explainability for some AI applications. Legal standards for Agentic AI are likely to include risk management plans that contain requirements for conducting risk assessments, periodic audits, and certifications for systems operating in susceptible sectors such as health, finance, and security. It is, however, expected that as Agentic AI progresses, developing legal frameworks that enable innovation while enhancing safety and accountability will be crucial.

Table 11 summarizes the primary ethical and governance considerations for Agentic AI, highlighting the complexities that arise as these systems become increasingly autonomous.

IX. CURRENT FRAMEWORKS FOR SAFE AND ACCOUNTABLE AGENTIC AI

A. SAFETY PROTOCOLS

Several safety measures and frameworks have been created as a requirement to prevent the autonomous risks involved with decision-making. One of the most basic methods is **goal safety protocols**, being objective-oriented procedures to suggest to the AI which goals are acceptable to pursue and which actions are harmful. These protocols prevent harm by ensuring that AI endeavors to achieve legal, safe, and ethical goals. Also, to prevent AI from going out of bounds or reaching unanticipated dangerous situations, **fail-safe mechanisms** [63] are deployed, which can adjust or even terminate AI activities altogether.

One such framework primarily employed is the **risk evaluation and management protocol** [64], which leads to the review of the risks posed by the actions of the AI, and subsequently, the control measures are put in place. This is especially important in healthcare and autonomous driving scenarios, where AI takes actions that impact human lives. Additionally, some Agentic AI systems utilize **ethical guardrails** [65], which restrain the AI from making decisions that raise ethical concerns and promote the AI to perform according to social norms and ethics. The illustrated safety protocols work as a system of layers, offering a good level of protection for the abuse of Agentic AI in an ethical manner.

B. MONITORING AND CONTROL MECHANISMS

The autonomous action of agentic AI systems requires monitoring and control mechanisms to govern such actions and allow for human intervention where necessary. **Real-time monitoring systems** [66], [67] oversee the AI's activities and decisions and allow humans to intervene if necessary. For example, in the case of financial trading systems, AI can be employed to identify unusual trade patterns in real-time trading by alerting the relevant authorities and allowing for the conduct of such activities to be at acceptable risk levels.

Such frameworks as **Human-in-the-loop (HITL)** [68] and **human-on-the-loop (HOTL)** [69] have been designed with the idea of maintaining an equilibrium between the extent of autonomy experienced and the necessity of human control. When HITL configurations are deployed, human operators tend to approve some actions and directly interact with the AI while making certain decisions. On the other hand, in HOTL systems, the human supervisor monitors the AI without deciding on every action but can intervene when necessary. These mechanisms are essential for healthcare and other serious decisions involving AI, where the presence of human intelligence is still required.

Override protocols also exist in Agentic AI systems to stop or change actions undertaken by the AI if a situation is classified as unforeseen or falls outside the reasonable range of expected behavior. Last-line systems provide another level of defense by making it possible for the AI to be overpowered by a human if its autonomous conduct becomes damaging. Combined with the aforementioned, these monitoring and control measures are designed to deliver a multitude of defense in depth to ensure that the agentic artificial intelligence does not exceed controlled and supervised settings, with all sorts of human intervention procedures firmly in place.

C. TRANSPARENCY MECHANISMS

Trust and accountability are foundational to the deployment of Agentic AI, and both can be eroded owing to lack of transparency. Explainable AI (XAI) [70], [71] techniques have been proposed to offer a more transparent view of the inner workings of these multi-layered decision-making AI models. As a result, XAI enriches the accountability of the processes involved in reaching decisions as leaders in a particular field, such as that of finance and healthcare, where decisions have profound consequences, are informed about the premises on which specific actions were taken.

Beyond just the transparent nature, **audit trails** can also be added as an extra layer within Agentic AI systems to document every logic and decision so that the actions taken by the AI can be examined afterward. These logs are perfect for compliance

IEEE *Access*

TABLE 11: Ethical and Governance Considerations for Agentic AI

Consideration	Description
Accountability and Responsibility	Determining who is accountable when an AI makes autonomous decisions, especially in complex, multi-stakeholder settings
Bias, Fairness, and Transparency	Ensuring AI actions are fair, free from bias, and explainable to users
Privacy and Security	Protecting sensitive data and securing systems from cyber threats
Regulatory and Legal Perspectives	Adapting legal frameworks to address the unique challenges of autonomous AI systems

because they can assist organizations in ensuring that past choices were compliant with the law. **Self-documenting algorithms** [72] also help promote self-explainable AI by automatically preparing reports and explanations about the system's decisions while it is in operation. Such mechanisms not only foster transparency but also facilitate feedback mechanisms, as they enable developers and stakeholders to improve and understand the decision-making process of the AI over time.

D. CASE STUDIES OF GOVERNANCE APPROACHES

Several industry and academic efforts have created governance frameworks to guarantee the safe, responsible and ethical use of AI. Here are several examples that are worth mentioning [73]:

- **Microsoft's AI Principles and Responsible AI Standard**: These include fairness, reliability, privacy, inclusiveness, and transparency, which govern the shaping of AI systems according to Microsoft's Responsible AI Standard. It also contains the Aether Committee, which fairly addresses the ethical and social issues brought about by AI. There have been other products throughout the business for which this framework has been used to build robust AI risk processes throughout Microsoft's offerings.
- **Google's AI Principles and Model Cards for Transparency**: Google defines AI principles for developing and deploying AI while ensuring privacy, security, and accountability across its platforms. The company also employs **Model Cards**, which contain information on a model's performance, limitations, and intended application, allowing users to appreciate the strengths and weaknesses of AI systems.
- **OpenAI's Charter and Safety Standards**: OpenAI undertook to adopt a charter that contains ethical and safety standards to be observed in developing general-purpose AI. OpenAI takes a human-focused approach and is value-oriented toward alignment with these goals, safety research, and open development of AI. This organization's safety research is directed at developing models meant to interact and operate in unpredictable environments, allowing for the responsible future of Agentic AI.
- **The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems**: IEEE's initiative has for-

mulated standards [74] with corresponding guidelines on ethical AI and addressed areas such as matters of transparency, accountability, as well as data integrity. They are aimed at assisting developers in setting the standard for the UK AI systems that they develop so that these systems will not only work but must be ethical and social, ensuring trust and safety in autonomous AI technologies.

These case studies illustrate diverse governance approaches that balance innovation with safety, accountability, and transparency. Table 12 summarizes these frameworks, comparing key principles and focus areas.

X. OPEN RESEARCH CHALLENGES AND FUTURE DIRECTIONS

The development and deployment of Agentic AI systems present several open research challenges that require attention to ensure safe, effective, and ethically aligned AI. This section explores these challenges and outlines potential future directions for advancing Agentic AI.

A. ENHANCED ADAPTABILITY AND RESILIENCE

In dynamic and uncertain contexts, Agentic AI systems must demonstrate adaptability and resilience as they operate in the environment. Most existing models do not generalize across the environments or change significantly without retraining. Thus, future work might involve solving meta-learning and transfer-learning problems specifically for Agentic AI systems, allowing them to rapidly adapt to new situations based on their previous experiences. Moreover, it would be interesting to research on platforms that enable real-time learning whereby agents learn and modify in response to changes but their primary functions are not affected. This would improve both resilience and adaptability tremendously.

B. IMPROVING GOAL ALIGNMENT WITH HUMAN VALUES

Integrating ethical and value-based human approaches into Agentic AI systems is one of the open branches of research that should be advanced in the future. Given the autonomous nature of operating these systems, there exists the possibility of a goal misalignment problem whereby the AI is driven to do the right thing. Still, the action is contrary to human interests. This work can explore **value alignment** strategies such as inverse reinforcement learning (IRL) and cooperative

IEEE *Access*

TABLE 12: Governance Frameworks for Safe and Accountable Agentic AI

Organization	Framework/Standard	Key Principles and Focus Areas
Microsoft	Responsible AI Standard	Fairness, reliability, privacy, transparency, human oversight
Google	AI Principles and Model Cards	Privacy, security, accountability, performance documentation
OpenAI	Charter and Safety Standards	Alignment with human values, safety research, transparency
IEEE	Ethics of Autonomous and Intelligent Systems	Ethical guidelines, transparency, accountability, data privacy

inverse reinforcement learning (CIRL), in which the AI goes through experience and exposure to human social context to learn human preferences and values. In addition, such frameworks should be created that would permit real-time adjustments to the level of alignment between Agentic AI and human objectives depending on changing social behavior or the targeted audience's cues.

C. INTEGRATION INTO LIVING SOFTWARE AND CYBER-PHYSICAL SYSTEMS

When Agentic AI is widespread, developing robust ethical and global guidelines for this technology application and development will be essential. Autonomous decision-making in the systems complicates ethical issues like bias, privacy, accountability, and transparency. Areas of future research can be directed at how to build a universal ethical framework for Agentic AI, which may include mechanisms of multi-dimensional ethical audit, oversight, and certification. This includes initiatives involving policymakers, ethicists, and AI practitioners to embed supervisory mechanisms that foster advancements without undermining other societal interests [75].

D. ETHICAL FRAMEWORKS AND GLOBAL STANDARDS

The definition of AI agency is still in the making stage, and how advanced decision-making, planning, and reasoning [76] will be embraced within Agentic AI is still being determined. Theoretical advances in multi-agent systems such as coordination, decentralized decision-making, or structured 'goal' management could help increase the autonomy of AI systems without losing control. Also, a better understanding of agents' cognitive functions, such as curiosity, intrinsic motivation, and moral reasoning, could contribute to developing more robust and responsible Agentic AI systems. Formulating scripts for AI agencies will be vital for ensuring that these systems can operate in a deterministic and dependable manner in diverse and complex environments [77].

E. THEORETICAL ADVANCES IN AI AGENCY

The concept of agency in AI is still evolving, and there is much to be explored regarding how Agentic AI can develop advanced decision-making, planning, and reasoning abilities. Theoretical research into topics such as multi-agent coordination, decentralized decision-making, and long-term goal management could provide new insights into enhancing AI autonomy while maintaining control [78]. Additionally,

advancing our understanding of cognitive functions like curiosity, intrinsic motivation, and moral reasoning in artificial agents could help design more robust and ethically aligned Agentic AI systems. Developing formal models of AI agency will be essential for ensuring these systems can act responsibly and predictably in diverse and complex environments.

Table 13 provides a summary of the primary open research challenges and future directions in Agentic AI. Each area highlights critical aspects of advancing the field responsibly and sustainably.

F. SCALABILITY AND EFFICIENCY

The deployment of Agentic AI systems in large-scale and complex environments imposes new constraints, resulting in vast requirements for scalability and efficiency. There are scenarios when such systems need to manage resources and information at large, interact with numerous agents, and make decisions within a short period. For both scalability and efficiency to be achieved, new advancements in algorithms, architectures, and hardware have to be made. This research area is characterized by the use of centerless and hierarchical frameworks and distributed processing or energy-efficient hardware.

Aside from its numerous advantages, a well-known limitation that stems from the emergence of **decentralized architectures** is their integration into large-scale systems. Since the control process can be handled by numerous nodes in a fully distributed way, there is no need for a single centralized controller. Thus, the responsibility and load of this node do not affect the system's scalability. An improvement of specific processes in this-style architecture might include certain structural, behavioral rules that regularize agents' actions regarding network interactions. The more structures decorated with efficient elements of decentralized control, the quicker the optimized directional tasks and networks as a whole system will be, where optimization processes occur in bulks of intelligent agents. As it has become apparent in recent years, these tasks have become possible with systems based on swarm intelligence and decentralized [79] control of structures and processes.

Another area of interest is **the hierarchical models**, which break down decision-making processes into different layers [80], where the system can only focus on general goals and distribute particular actions to subordinate agents. For instance, in the smart city context, a hierarchical Agentic AI could oversee the movement of cars throughout the city. At

TABLE 13: Summary of Open Research Challenges and Future Directions in Agentic AI

Research Area	Future Directions
Scalability and Efficiency	Development of decentralized models, distributed processing, and energy-efficient algorithms
Enhanced Adaptability and Resilience	Meta-learning, transfer learning, and real-time adaptive mechanisms
Goal Alignment with Human Values	Value alignment techniques (e.g., IRL, CIRL), continuous alignment frameworks
Integration into Living Systems	Standards for interoperability, distributed AI architectures
Ethical Frameworks and Global Standards	Establishment of universal ethical standards and global AI governance
Theoretical Advances in AI Agency	Research on multi-agent coordination, moral reasoning, and intrinsic motivation in AI

the same time, local agents manage the traffic at a particular intersection. This configuration diminishes the computation each agent has to perform and enables better system scalability. The hierarchical models have found much application in the coordinated control of multiple, interdependent processes in a system since they allow for agnostic task scheduling as the state of the system changes.

Balanced loading of tasks among processors is another factor that improves scalability since it allows distributing computations over several devices or servers. Within distributed processing frameworks, large scale Agentic AI systems could make the most of cloud or edge computing resources to carry out computation in multiple processes and lower the time taken to respond. Such a situation proves helpful in operating in remote applications or in time-critical situations such as autonomous vehicles or smart grid systems where real-time data processing and decision-making are critical. There is ongoing advancement in deploying distributed AI frameworks, such as federated learning [81] and edge AI [82], that enhance the maneuverability of Agentic AI systems in a wider operational area with dispersed installations [83].

Apart from the AI systems integration, improvement of the Agentic AI systems embodies energy-efficient **algorithms and hardware** optimizations [84]. This is vital when the applications must run endlessly and provide real-time responses. Several Agentic AI models are built using operationally expensive algorithms like deep reinforcement learning, requiring enormous power resources. Further progress is made in energy-efficient machine learning approaches, such as sparse neural networks and quantized models that address the computation requirements at lower levels while retaining the quality of the end products. Furthermore, neuromorphic processors and AI accelerators are some of the emerging hardware technologies that are meant to enable the performance of AI tasks with minimal power requirements. These hardware improvements are critical in expanding the Agentic AI systems to edge devices and IoT applications with limited or no power resources.

Although progress has been achieved, they only partially solve the problem of scalability and efficiency, which continue to be active research issues as real-world environments become increasingly complex. Future directions in this area include the development of **adaptive resource allocation techniques** [85] that dynamically allocate computational resources based on task importance, as well as **self-optimizing algorithms** [86] that adjust their complexity dynamically depending on the requirements. Moreover, thanks to **collaborative intelligence** [87] conception, agents will be able to solve issues without complicated integration processes as they will be able to handle sophisticated operations collectively, thus enhancing system scalability.

G. ENHANCED ADAPTABILITY AND RESILIENCE

Agentic AI systems, particularly those employed by a human worker or soldier, are built to endure persistence under demanding, unpredictable conditions. Therefore, achieving persistent operational success would require these systems to modify their actions autonomously and gain further learning resilience constantly. How can one reach such a high level of learning resilience? One possible answer lies in the use of Agentic AI technologies that can generalize across different environments and concepts, deal with high uncertainty, and cope with unforeseen interruptions.

One exciting way of achieving such resilience is through **meta-learning**, or what children call "learning to learn." Such meta-learning algorithms enable an AI to learn information that may apply to different assignments or environments, allowing the AI to be flexible with virtually no retraining. For example, in robotic tasks [88], a meta-learning agent raised in one environment (let's say, in the house) can move to an alternative environment (the outside) and assume the knowledge of moving within the outside environment. This is because AI can now learn like humans [89] and use its past experiences in the environment to approach new situations [90].

Another significant approach for enhancing adaptability is called **transfer learning** [91]. It has proved that unlike static models trained for a particular purpose, a learning model trained for one task can perform a similar or even related, yet distinct, task. This method is particularly convenient for Agentic AIs acting in many different operational spheres, enabling the system to leverage its previous lessons instead of

IEEE *Access*

acquiring new ones. Transfer learning is known to be effective during these scenarios since it allows the AI to adjust to different patient profiles or drive in different environments, such as autonomous driving.

Resilience in the case of Agentic AI systems is crucial as it allows them to deliver consistent outcomes despite upsets being caused by compromised sensors, intermittent networks or external factors that were not anticipated. To increase resilience, there are ongoing investigations on robust reinforcement learning techniques that enable the reliable functioning of agents in the presence of consistent uncertainty or unknown noise. Robust reinforcement learning methods rely on dual dynamics, allowing the learner to seek performance maximization and impact minimization of stochastic events. This dual objective ensures the agent can effectively implement its functionalities even when the normal conditions vary.

Self-recovery mechanisms are being explored as supplementary to everyday learning strategies. Self-recovery mechanisms allow Agentic AI systems to auto-flag and rectify their duplicates without help. For example, self-recovery strategies in autonomous robotics may include obstacle avoidance strategies where the robot uses an alternative approach or site configuration strategies. System self-recovery mechanisms allow seamless system restoration after a range of disruptions, which is particularly beneficial to mission-critical areas like disaster management and space exploration, where human intervention is not always feasible.

The stand-out areas of research that are helpful in both the modification and robustness of Agentic AI include real-time learning. AI self-recovery mechanisms [92], on the other hand, include real-time learning. This mechanism allows AI systems to receive and integrate new information into relevant models without ceasing operational activities. This is excellent, for example, in the finance or cyber warfare zones models, which require swift AI alterations on market changes or threats. However, with this capability, new challenges also arise related to the real-time processing of data, the ability to conduct computation, model maintenance, and stability, and these need to be addressed in future studies.

Looking to the future, progress in improving adaptability and survivability will likely include the creation of **context-aware learning algorithms** [93], which evaluate the setting and act by it. In other words, some form of contextual AI could switch its unitary decision-making style depending on uncertainty or resource availability. Furthermore, investigation of **multi-modal learning** [94], [95], which fuses information from multiple sources and types to better understand different situations, may assist in improving the versatility of Agentic AI systems in the operative context. With visual, auditory, and other kinds of information, multi-modal learning can be particularly effective for autonomous vehicles trying to function in complex situations.

H. ETHICAL FRAMEWORKS AND GLOBAL STANDARDS

The increasing capabilities of Autonomous Agentic AI systems operating in significant sectors have intensified the call for appropriate ethical considerations and global standards to be put in place. These considerations and standards are pivotal in regulating how Agentic AI interacts with stakeholders and the environment to minimize adverse impacts, biases, and threats to safety and privacy. In addition, appropriate and well-thought-out ethical considerations and global standards also seek to define the scope and areas of transparency, equity, accountability, and user privacy, which need to be emphasized in the design and use of AI.

One of the critical barriers to establishing these frameworks is the existence of ethical and legal variety in different societies. Ethical issues, including data protection, transparency in who took the decision, and reliance on autonomous decisions, may range from one society to another depending on legal order, customs, traditions, or politics. For these differences to be addressed while promoting synergies among regions, a global set of ethical standards for AI governance is fundamental. Establishing universal ethical principles can provide the basis for this alignment by developing common approaches to equity, non-maleficence and beneficence that would override local differences.

Transparency and Explainability are undoubtedly the components of an ethical algorithm since they allow third parties to comprehend the decision-making process of artificial intelligence. In Agentic AI, such attributes are necessary for users' trust, especially in crucial sectors like healthcare and finance, where the rationale behind the decision needs to be accountable. These include XAI techniques that allow stakeholders to explain the interpretation process. Therefore, given the erosion of these fundamental rights, these frameworks should emphasize the need to incorporate stakeholders' responsibilities and authorities' mechanisms to scrutinize AI-generated decisions [96].

More important and unwaveringly related to the development of ethical frameworks is the issue of an **Accountability and Responsibility**. With the growing autonomy of agents in society, there have been significant difficulties in determining who should be held responsible for the results of AI-enhanced acts. Provisions for processes that would enable tracing responsibility in a wide range of stakeholders, including farmers, custodians, and machines, must be in place. Some approaches suggest an **AI responsibility chain**, which separates the roles and responsibilities of each stakeholder during the development and deployment process to prevent over-reliance on one individual [97].

Privacy and Security Standards are especially critical when addressing Agentic AI, which generally requires operating on susceptible data. There should be guidelines prescribing how data is managed, secured or anonymized to meet user privacy. This includes baseline restrictions on processing, such as the use of differential privacy and secure access to databases, which allow for reducing possibilities for breaches or abuse of the data. Furthermore, appropriate measures

should be employed to safeguard the security of Agentic AI, especially in smart cities and autonomous vehicles where such AI systems are exposed to potential cyber threats and can be a danger to the public.

To concretize these principles, several regulating authorities and industry associations have suggested using ethical guidelines and frameworks aimed at the responsible use of artificial intelligence. In particular, some of the core material for the legal issues of privacy and other user rights are contained in the Ethics Guidelines for Trustworthy AI developed by the European Union and the General Data Protection Regulation. Similarly, the IEEE and the Partnership on AI organizations have also formulated some behavioral artificial intelligence ethics addressing responsibility, fairness, and openness issues. Although these guidelines have no universal application, effective international government authority has been advocated to coordinate the interaction between AI policies and standards from different countries.

A further development concerning the issue is establishing an internationally operating **AI coordinating entity** or a consortium like in Europe that could promote the establishment of AI standards and global adherence to ethical principles. This organization would create an objective to standardize AI regulations so that ethical norms and safety requirements will invariably be provided across countries and industries. Moreover, it necessitates research on **adaptive regulatory regimes** [98] that are responsive to technological developments.
Table 14 summarizes the primary components of ethical frameworks and global standards for Agentic AI, outlining essential areas for responsible and safe AI governance.

I. THEORETICAL ADVANCES IN AI AGENCY
AI systems with the potential for intelligent action still need theoretical development regarding number and complexity in understanding agency, self-organization, and decision-making processes. For now, embodiments of Agentic AI exhibit their expected performance in structured tasks and environments, yet transferring such capabilities into higher-order, polycontext environments will necessitate a sensible underlying framework. In a larger perspective, however, some theoretical extensions of the issues, like multi-agent coordination, meta-goal achievement, active persistence, and planner controlling with ethical consideration, assist the AI models in behaving appropriately in different environments.

Multi-agent coordination is an example of such a class of theories, and it is concerned with allowing AI agents to engage, communicate, and collaborate in environments that contain multiple autonomous entities. In particular, practical applications multi-agent coordination is crucial for applications in collaborative robotics, decentralized systems, and smart city architecture. This domain examines how agents learn to pursue their separate aims while achieving mutual goals or seeking to work cooperatively, utilizing materials from game theory, reinforcement learning, and communication theory. Developments in this domain will allow Agentic AI

to carry out interdependent complex tasks where agents must cooperate and compete.

Long-term goal management is decisive and is characterized as having advanced agentic capabilities. There are deeply ingrained limitations concerning timelines with current AI systems. Current AIs tend to optimize for shorter timelines, thus losing the capacity to strategize, forecast trends, and distinguish between sub-objectives. Theoretical work addresses the practical challenges of enabling AIs to set and pursue long-range goals as conditions change and constraints emerge. Hierarchical reinforcement learning and temporal reasoning are two measures that have promise as they allow AI systems to decompose high-level goals into sub-tasks, which take the system toward the overall end state through predetermined objectives.

Agents' ability to wait for a reward and workers' pursuit on their own define two self-rewarding procedures that emerge towards the creation of agentic AI; both procedures in domination with zero hit or standalone provide enough basis to move forward without external incentives or directives. An emerging field of research in self-directed strategy relies on developing agents that can explore, learn, and adapt based solely on their curiosity about their environment. This would be important in cases where the information space is uncertain or where the goal is not defined and other variables are likely to change within the domain of the goal. Further development in the theory of self-rewarding will enhance AI agents' abilities to be goal-oriented and directed by a flexible strategy while dealing with complex problems caused by a lack of structure.

Moral reasoning and ethical decision-making are gaining acceptance in the field of advanced AI agency [99], particularly now that AI is increasingly used for sensitive and high-risk tasks. Research on moral reasoning seeks to offer a basis for ethical deliberations in autonomous systems so that Agentic AI can purposefully choose actions with thought of their effects and values. In this context, integrating ethics, philosophy, and psychology into AI is an overarching aim if ethical AI systems should be created. Codifying moral reasoning in AI will be necessary for health, law and order, and autonomous automobiles, the decision of which can be influential to society.

A few of the areas of theoretical research that need a mention are **self-awareness** and **meta-cognition** [100] in AI which include building systems that understand their actions, abilities, or limitations as self-referential knowledge. All the AI that are self-aware will be able to self-evaluate whether they have performed their tasks optimally, what can be improved, and what actions to take when there are failures or poor performance. Self-agency abilities will also allow AI to assess its strategies and learning processes in the hopes that its agents will be able to enhance their decision-making in the future. Progress in the self-awareness and meta-cognition areas might allow for the emergence of more sophisticated and flexible Agentic AI systems that enhance and improve their performance and robustness in multiple environments.

In this regard, the future possibilities of the works included

IEEE *Access*

TABLE 14: Key Components of Ethical Frameworks and Global Standards for Agentic AI

Component	Description
Transparency and Explainability	Ensure AI decisions are interpretable and can be scrutinized by users and regulators
Accountability and Responsibility	Define and trace responsibility across stakeholders in the AI lifecycle
Privacy and Security Standards	Enforce data privacy and security protocols to protect user information
Global Governance	Develop international regulatory bodies to harmonize AI standards across borders
Dynamic Regulation	Create adaptable frameworks that evolve with technological advancements in AI

the possibility of creating **formal models of AI agency**, which would include the discussed dimensions and form a theoretical framework that could be used for the design and evaluation of agentic behavior. The models could also serve the great purpose of setting metrics and benchmarks for agencies that would, in turn, facilitate ascertaining the efficacy of AI systems in performing complex and nonsupervised tasks. Further, studies on **adaptive moral frameworks** and **contextual decision-making** [101] might enable Agentic AI to take situational ethics and change how it makes decisions in various contexts.

J. ROADMAP FOR FUTURE RESEARCH

The future of Agentic AI depends on addressing current limitations and expanding its applicability. Key areas for research include:

- **Goal Alignment with Human Values:** Developing frameworks for inverse reinforcement learning (IRL) and cooperative IRL to ensure Agentic AI objectives align with societal values.
- **Scalability:** Exploring decentralized architectures and federated learning to enable Agentic AI to handle large-scale, distributed systems effectively.
- **Adaptability and Resilience:** Advancing meta-learning and transfer-learning techniques to allow Agentic AI to adapt to novel situations without retraining.
- **Energy Efficiency:** Innovating energy-efficient hardware and algorithms to reduce the computational cost of deploying Agentic AI in resource-constrained environments.
- **Ethical and Governance Frameworks:** Establishing universal ethical standards, transparency mechanisms, and regulatory guidelines to ensure responsible AI deployment.
- **Real-Time Learning Systems:** Designing systems capable of learning and adapting in real-time without disrupting ongoing operations.

By addressing these areas, future research can enable Agentic AI to achieve its full potential across industries while mitigating risks.

XI. CONCLUSION

A. SUMMARY OF KEY FINDINGS

This survey has explored the foundational characteristics, methodologies, applications, challenges, and future directions of Agentic AI. Key findings highlight that Agentic AI represents a significant advancement in artificial intelligence, characterized by autonomy, goal-oriented behavior, and adaptability across diverse environments. We have identified core applications in industries such as healthcare, finance, and manufacturing, where Agentic AI's ability to make context-aware, autonomous decisions offers transformative benefits. However, deploying these systems in real-world scenarios introduces challenges such as scalability, resource constraints, and ethical concerns, all of which require robust solutions to ensure safe and effective AI deployment. Through a comparative analysis, we examined various implementation frameworks, tools, and methodologies that contribute to the development and evaluation of Agentic AI. We also identified open research challenges, including goal alignment, multi-agent coordination, and regulatory adaptation, which must be addressed to fully realize the potential of Agentic AI.

B. FINAL INSIGHTS ON AGENTIC AI

Agentic AI holds transformative potential across numerous sectors, promising advances in automation, decision-making, and human-AI collaboration. As these systems evolve, they are poised to tackle complex tasks autonomously, significantly expanding the scope of AI applications in both structured and unstructured environments. By combining adaptive learning, robust reinforcement mechanisms, and real-time responsiveness, Agentic AI systems can deliver dynamic solutions that enhance productivity and efficiency. However, with increased autonomy comes the responsibility to address ethical considerations and ensure accountability, transparency, and fairness. As Agentic AI continues to evolve, it is essential that these systems are developed with a clear focus on ethical alignment, resilience, and regulatory compliance to prevent potential misuse or unintended consequences.

REFERENCES

[1] A. Chan, R. Salganik, A. Markelius, C. Pang, N. Rajkumar, D. Krasheninnikov, L. Langosco, Z. He, Y. Duan, M. Carroll et al., "Harms from increasingly agentic algorithmic systems," in Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, pp. 651–666.

IEEE *Access*

[2] A. Paul, C. L. Yu, E. A. Susanto, N. W. L. Lau, and G. I. Meadows, "Agentpeertalk: Empowering students through agentic-ai-driven discernment of bullying and joking in peer interactions in schools," arXiv preprint arXiv:2408.01459, 2024.

[3] Y. Shavit, S. Agarwal, M. Brundage, S. Adler, C. O'Keefe, R. Campbell, T. Lee, P. Mishkin, T. Eloundou, A. Hickey et al., "Practices for governing agentic ai systems," Research Paper, OpenAI, December, 2023.

[4] S. Guenat, P. Purnell, Z. G. Davies, M. Nawrath, L. C. Stringer, G. R. Babu, M. Balasubramanian, E. E. Ballantyne, B. K. Bylappa, B. Chen et al., "Meeting sustainable development goals via robotics and autonomous systems," Nature communications, vol. 13, no. 1, p. 3559, 2022.

[5] E. H. B. Maia, L. C. Assis, T. A. De Oliveira, A. M. Da Silva, and A. G. Taranto, "Structure-based virtual screening: from classical to artificial intelligence," Frontiers in chemistry, vol. 8, p. 343, 2020.

[6] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "React: synergizing reasoning and acting in language models (2022)," arXiv preprint arXiv:2210.03629, 2023.

[7] R. Yang, L. Song, Y. Li, S. Zhao, Y. Ge, X. Li, and Y. Shan, "Gpt4tools: Teaching large language model to use tools via self-instruction," Advances in Neural Information Processing Systems, vol. 36, 2024.

[8] S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, "Generative ai," Business & Information Systems Engineering, vol. 66, no. 1, pp. 111–126, 2024.

[9] T. Chong, T. Yu, D. I. Keeling, and K. de Ruyter, "Ai-chatbots on the services frontline addressing the challenges and opportunities of agency," Journal of Retailing and Consumer Services, vol. 63, p. 102735, 2021.

[10] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of image classification algorithms based on convolutional neural networks," Remote Sensing, vol. 13, no. 22, p. 4712, 2021.

[11] A. Brisset, R. Gill, and R. Gannon, "The search for a native language: Translation and cultural identity," in The translation studies reader. Routledge, 2021, pp. 289–319.

[12] H. Ko, S. Lee, Y. Park, and A. Choi, "A survey of recommendation systems: recommendation models, techniques, and application fields," Electronics, vol. 11, no. 1, p. 141, 2022.

[13] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, and J. Morimoto, "Deep learning, reinforcement learning, and world models," Neural Networks, vol. 152, pp. 267–275, 2022.

[14] F. Binucci, P. Banelli, P. Di Lorenzo, and S. Barbarossa, "Dynamic resource allocation for multi-user goal-oriented communications at the wireless edge," in 2022 30th European Signal Processing Conference (EUSIPCO). IEEE, 2022, pp. 697–701.

[15] L. Zhao, H. Cheng, J. Zhang, and Y. Xia, "Adaptive control for a motion mechanism with pneumatic artificial muscles subject to dead-zones," Mechanical Systems and Signal Processing, vol. 148, p. 107155, 2021.

[16] C. A. Aubin, B. Gorissen, E. Milana, P. R. Buskohl, N. Lazarus, G. A. Slipher, C. Keplinger, J. Bongard, F. Iida, J. A. Lewis et al., "Towards enduring autonomous robots via embodied energy," Nature, vol. 602, no. 7897, pp. 393–402, 2022.

[17] P. Dhamija and S. Bag, "Role of artificial intelligence in operations environment: a review and bibliometric analysis," The TQM Journal, vol. 32, no. 4, pp. 869–896, 2020.

[18] S. Kapoor, B. Stroebl, Z. S. Siegel, N. Nadgir, and A. Narayanan, "Ai agents that matter," arXiv preprint arXiv:2407.01502, 2024.

[19] N. J. Fast and J. Schroeder, "Power and decision making: new directions for research in the age of artificial intelligence," Current opinion in psychology, vol. 33, pp. 172–176, 2020.

[20] R. C. Cardoso and A. Ferrando, "A review of agent-based programming for multi-agent systems," Computers, vol. 10, no. 2, p. 16, 2021.

[21] P. Putta, E. Mills, N. Garg, S. Motwani, C. Finn, D. Garg, and R. Rafailov, "Agent q: Advanced reasoning and learning for autonomous ai agents," arXiv preprint arXiv:2408.07199, 2024.

[22] T. M. Getu, G. Kaddoum, and M. Bennis, "A survey on goal-oriented semantic communication: Techniques, challenges, and future directions," IEEE Access, 2024.

[23] A. Zeng, M. Liu, R. Lu, B. Wang, X. Liu, Y. Dong, and J. Tang, "Agenttuning: Enabling generalized agent abilities for llms," arXiv preprint arXiv:2310.12823, 2023.

[24] C.-P. Dai and F. Ke, "Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review," Computers and Education: Artificial Intelligence, vol. 3, p. 100087, 2022.

[25] N. Shinn, F. Cassano, A. Gopinath, K. Narasimhan, and S. Yao, "Reflexion: Language agents with verbal reinforcement learning," Advances in Neural Information Processing Systems, vol. 36, 2024.

[26] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," IEEE transactions on pattern analysis and machine intelligence, vol. 44, no. 9, pp. 4555–4576, 2021.

[27] R. M. Samant, M. R. Bachute, S. Gite, and K. Kotecha, "Framework for deep learning-based language models using multi-task learning in natural language understanding: A systematic literature review and future directions," IEEE Access, vol. 10, pp. 17 078–17 097, 2022.

[28] J. Arroyo, C. Manna, F. Spiessens, and L. Helsen, "An open-ai gym environment for the building optimization testing (boptest) framework," in Building Simulation 2021, vol. 17. IBPSA, 2021, pp. 175–182.

[29] Y. Song, A. Wojcicki, T. Lukasiewicz, J. Wang, A. Aryan, Z. Xu, M. Xu, Z. Ding, and L. Wu, "Arena: A general evaluation platform and building toolkit for multi-agent intelligence," in Proceedings of the AAAI conference on artificial intelligence, vol. 34, no. 05, 2020, pp. 7253–7260.

[30] X. Kong, G. Wang, and A. Nichol, Conversational AI with Rasa: Build, test, and deploy AI-powered, enterprise-grade virtual assistants and chatbots. Packt Publishing Ltd, 2021.

[31] V. Heilala, P. Jääskelä, M. Saarela, and T. Kärkkäinen, "Adapting teaching and learning in higher education using explainable student agency analytics," in Principles and Applications of Adaptive Artificial Intelligence. IGI Global, 2024, pp. 20–51.

[32] B. Abedin, C. Meske, I. Junglas, F. Rabhi, and H. R. Motahari-Nezhad, "Designing and managing human-ai interactions," Information Systems Frontiers, vol. 24, no. 3, pp. 691–697, 2022.

[33] C. Fang, J. N. Wilkenfeld, N. Navick, and J. L. Gibbs, ""ai am here to represent you": Understanding how institutional logics shape attitudes toward intelligent technologies in legal work," Management Communication Quarterly, vol. 37, no. 4, pp. 941–970, 2023.

[34] S. Hu, C. Lu, and J. Clune, "Automated design of agentic systems," arXiv preprint arXiv:2408.08435, 2024.

[35] A. Baird and L. M. Maruping, "The next generation of research on is use: A theoretical framework of delegation to and from agentic is artifacts." MIS quarterly, vol. 45, no. 1, 2021.

[36] J. J. Woo, A. J. Yang, R. J. Olsen, S. S. Hasan, D. H. Nawabi, B. U. Nwachukwu, R. J. Williams III, and P. N. Ramkumar, "Custom large language models improve accuracy: Comparing retrieval augmented generation and artificial intelligence agents to non-custom models for evidence-based medicine," Arthroscopy: The Journal of Arthroscopic & Related Surgery, 2024.

[37] E. Yang, T. Garcia, H. Williams, B. Kumar, M. Ramé, E. Rivera, Y. Ma, J. Amar, C. Catalani, and Y. Jia, "From barriers to tactics: A behavioral science-informed agentic workflow for personalized nutrition coaching," arXiv preprint arXiv:2410.14041, 2024.

[38] S. Gao, A. Fang, Y. Huang, V. Giunchiglia, A. Noori, J. R. Schwarz, Y. Ektefaie, J. Kondic, and M. Zitnik, "Empowering biomedical discovery with ai agents," Cell, vol. 187, no. 22, pp. 6125–6151, 2024.

[39] M. Murtaza, Y. Ahmed, J. A. Shamsi, F. Sherwani, and M. Usman, "Ai-based personalized e-learning systems: Issues, challenges, and solutions," IEEE access, vol. 10, pp. 81 323–81 342, 2022.

[40] N. Loizou, S. Vaswani, I. H. Laradji, and S. Lacoste-Julien, "Stochastic polyak step-size for sgd: An adaptive learning rate for fast convergence," in International Conference on Artificial Intelligence and Statistics. PMLR, 2021, pp. 1306–1314.

[41] M. Chen, H. V. Poor, W. Saad, and S. Cui, "Convergence time optimization for federated learning over wireless networks," IEEE Transactions on Wireless Communications, vol. 20, no. 4, pp. 2457–2471, 2020.

[42] R. Hamon, H. Junklewitz, I. Sanchez et al., "Robustness and explainability of artificial intelligence," Publications Office of the European Union, vol. 207, p. 2020, 2020.

[43] U. Mittal and D. Panchal, "Ai-based evaluation system for supply chain vulnerabilities and resilience amidst external shocks: An empirical approach," Reports in Mechanical Engineering, vol. 4, no. 1, pp. 276–289, 2023.

[44] S. Tatineni and V. R. Boppana, "Ai-powered devops and mlops frameworks: Enhancing collaboration, automation, and scalability in machine learning pipelines," Journal of Artificial Intelligence Research and Applications, vol. 1, no. 2, pp. 58–88, 2021.

[45] P. Hemmer, M. Westphal, M. Schemmer, S. Vetter, M. Vössing, and G. Satzger, "Human-ai collaboration: the effect of ai delegation on human task performance and task satisfaction," in Proceedings of the 28th International Conference on Intelligent User Interfaces, 2023, pp. 453–463.

[46] M. Fan, X. Yang, T. Yu, Q. V. Liao, and J. Zhao, "Human-ai collaboration for ux evaluation: effects of explanation and synchronization," Proceedings

of the ACM on Human-Computer Interaction, vol. 6, no. CSCW1, pp. 1–32, 2022.

[47] S. P. Pattyam, "Artificial intelligence for healthcare diagnostics: Techniques for disease prediction, personalized treatment, and patient monitoring," Journal of Bioinformatics and Artificial Intelligence, vol. 1, no. 1, pp. 309–343, 2021.

[48] G. Cohen, "Algorithmic trading and financial forecasting using advanced artificial intelligence methodologies," Mathematics, vol. 10, no. 18, p. 3302, 2022.

[49] S. P. Pattyam, "Ai-driven financial market analysis: Advanced techniques for stock price prediction, risk management, and automated trading," African Journal of Artificial Intelligence and Sustainable Development, vol. 1, no. 1, pp. 100–135, 2021.

[50] L. T. Khrais, "Role of artificial intelligence in shaping consumer demand in e-commerce," Future Internet, vol. 12, no. 12, p. 226, 2020.

[51] R. Fedorko, Š. Král', and R. Bačík, "Artificial intelligence in e-commerce: A literature review," in Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 2. Springer, 2022, pp. 677–689.

[52] Z. M. Çınar, A. Abdussalam Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," Sustainability, vol. 12, no. 19, p. 8211, 2020.

[53] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L. A. Celi, and R. Mark, "Mimic-iv," PhysioNet. Available online at: https://physionet.org/content/mimiciv/1.0/(accessed August 23, 2021), pp. 49–55, 2020.

[54] M. Hammad, A. A. Abd El-Latif, A. Hussain, F. E. Abd El-Samie, B. B. Gupta, H. Ugail, and A. Sedik, "Deep learning models for arrhythmia detection in iot healthcare applications," Computers and Electrical Engineering, vol. 100, p. 108011, 2022.

[55] M. Javed Awan, M. S. Mohd Rahim, H. Nobanee, A. Munawar, A. Yasin, and A. M. Zain, "Social media and stock market prediction: a big data approach," MJ Awan, M. Shafry, H. Nobanee, A. Munawar, A. Yasin et al.," Social media and stock market prediction: a big data approach," Computers, Materials & Continua, vol. 67, no. 2, pp. 2569–2583, 2021.

[56] P. Li, A. Kusari, and D. J. LeBlanc, "A novel traffic simulation framework for testing autonomous vehicles using sumo and carla," arXiv preprint arXiv:2110.07111, 2021.

[57] I. Han, D.-H. Park, and K.-J. Kim, "A new open-source off-road environment for benchmark generalization of autonomous driving," IEEE Access, vol. 9, pp. 136 071–136 082, 2021.

[58] S. Hagmeyer, F. Mauthe, and P. Zeiler, "Creation of publicly available data sets for prognostics and diagnostics addressing data scenarios relevant to industrial applications," International Journal of Prognostics and Health Management, vol. 12, no. 2, 2021.

[59] Y. A. Yucesan, A. Dourado, and F. A. Viana, "A survey of modeling for prognosis and health management of industrial equipment," Advanced Engineering Informatics, vol. 50, p. 101404, 2021.

[60] G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan, and A. Anandkumar, "Voyager: An open-ended embodied agent with large language models," arXiv preprint arXiv:2305.16291, 2023.

[61] G. Sartor, F. Lagioia et al., "The impact of the general data protection regulation (gdpr) on artificial intelligence," 2020.

[62] F. Miao, W. Holmes, R. Huang, H. Zhang et al., AI and education: A guidance for policymakers. Unesco Publishing, 2021.

[63] A. V. S. Neto, J. B. Camargo, J. R. Almeida, and P. S. Cugnasca, "Safety assurance of artificial intelligence-based systems: A systematic literature review on the state of the art and guidelines for future work," IEEE Access, vol. 10, pp. 130 733–130 770, 2022.

[64] G. S. Collins, P. Dhiman, C. L. A. Navarro, J. Ma, L. Hooft, J. B. Reitsma, P. Logullo, A. L. Beam, L. Peng, B. Van Calster et al., "Protocol for development of a reporting guideline (tripod-ai) and risk of bias tool (probast-ai) for diagnostic and prognostic prediction model studies based on artificial intelligence," BMJ open, vol. 11, no. 7, p. e048008, 2021.

[65] A. Biswas and W. Talukdar, "Guardrails for trust, safety, and ethical development and deployment of large language models (llm)," Journal of Science & Technology, vol. 4, no. 6, pp. 55–82, 2023.

[66] W. Cai, J. Wang, P. Jiang, L. Cao, G. Mi, and Q. Zhou, "Application of sensing techniques and artificial intelligence-based methods to laser welding real-time monitoring: A critical review of recent literature," Journal of Manufacturing systems, vol. 57, pp. 1–18, 2020.

[67] S. Saponara, A. Elhanashi, and A. Gagliardi, "Implementing a real-time, ai-based, people detection and social distancing measuring system for covid-19," Journal of Real-Time Image Processing, vol. 18, no. 6, pp. 1937–1947, 2021.

[68] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal, "Human-in-the-loop machine learning: a state of the art," Artificial Intelligence Review, vol. 56, no. 4, pp. 3005–3054, 2023.

[69] K. Kuru, "Conceptualisation of human-on-the-loop haptic teleoperation with fully autonomous self-driving vehicles in the urban environment," IEEE Open Journal of Intelligent Transportation Systems, vol. 2, pp. 448–469, 2021.

[70] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "Xai—explainable artificial intelligence," Science robotics, vol. 4, no. 37, p. eaay7120, 2019.

[71] D. B. Acharya, B. Divya, and K. Kuppan, "Explainable and fair ai: Balancing performance in financial and real estate machine learning models," IEEE Access, vol. 12, pp. 154 022–154 034, 2024.

[72] T. Weise and Z. Wu, "Replicable self-documenting experiments with arbitrary search spaces and algorithms," in Proceedings of the Companion Conference on Genetic and Evolutionary Computation, 2023, pp. 1891–1899.

[73] P. B. de Laat, "Companies committed to responsible ai: From principles towards implementation and regulation?" Philosophy & technology, vol. 34, no. 4, pp. 1135–1193, 2021.

[74] K. Michael, R. Abbas, G. Roussos, E. Scornavacca, and S. Fosso-Wamba, "Ethics in ai and autonomous system applications design," IEEE Transactions on Technology and Society, vol. 1, no. 3, pp. 114–127, 2020.

[75] M. Metta, S. Ciliberti, C. Obi, F. Bartolini, L. Klerkx, and G. Brunori, "An integrated socio-cyber-physical system framework to assess responsible digitalisation in agriculture: A first application with living labs in europe," Agricultural Systems, vol. 203, p. 103533, 2022.

[76] E. Zelikman, Y. Wu, J. Mu, and N. Goodman, "Star: Bootstrapping reasoning with reasoning," Advances in Neural Information Processing Systems, vol. 35, pp. 15 476–15 488, 2022.

[77] E. J. Emanuel, G. Persad, A. Kern, A. Buchanan, C. Fabre, D. Halliday, J. Heath, L. Herzog, R. Leland, E. T. Lemango et al., "An ethical framework for global vaccine allocation," Science, vol. 369, no. 6509, pp. 1309–1312, 2020.

[78] S. S. Sundar, "Rise of machine agency: A framework for studying the psychology of human–ai interaction (haii)," Journal of Computer-Mediated Communication, vol. 25, no. 1, pp. 74–88, 2020.

[79] L. Cao, "Decentralized ai: Edge intelligence and smart blockchain, metaverse, web3, and desci," IEEE Intelligent Systems, vol. 37, no. 3, pp. 6–19, 2022.

[80] S. Saha, Z. Gan, L. Cheng, J. Gao, O. L. Kafka, X. Xie, H. Li, M. Tajdari, H. A. Kim, and W. K. Liu, "Hierarchical deep learning neural network (hidenn): an artificial intelligence (ai) framework for computational science and engineering," Computer Methods in Applied Mechanics and Engineering, vol. 373, p. 113452, 2021.

[81] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," IEEE transactions on neural networks and learning systems, vol. 34, no. 12, pp. 9587–9603, 2022.

[82] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, "Communication-efficient edge ai: Algorithms and systems," IEEE Communications Surveys & Tutorials, vol. 22, no. 4, pp. 2167–2191, 2020.

[83] F. Saeik, M. Avgeris, D. Spatharakis, N. Santi, D. Dechouniotis, J. Violos, A. Leivadeas, N. Athanasopoulos, N. Mitton, and S. Papavassiliou, "Task offloading in edge and cloud computing: A survey on mathematical, artificial intelligence and control theory solutions," Computer Networks, vol. 195, p. 108177, 2021.

[84] M. Capra, B. Bussolino, A. Marchisio, G. Masera, M. Martina, and M. Shafique, "Hardware and software optimizations for accelerating deep neural networks: Survey of current trends, challenges, and the road ahead," IEEE Access, vol. 8, pp. 225 134–225 180, 2020.

[85] A. Abid, M. F. Manzoor, M. S. Farooq, U. Farooq, and M. Hussain, "Challenges and issues of resource allocation techniques in cloud computing," KSII Transactions on Internet and Information Systems (TIIS), vol. 14, no. 7, pp. 2815–2839, 2020.

[86] R. Ranjan, N. Vemuri, and K. Venigandla, "Autonomous devops: Integrating rpa, ai, and ml for self-optimizing development pipelines," Asian Journal of Multidisciplinary Research & Review, vol. 3, no. 2, pp. 214–231, 2022.

[87] J. Paschen, M. Wilson, and J. J. Ferreira, "Collaborative intelligence: How human and artificial intelligence create value along the b2b sales funnel," Business Horizons, vol. 63, no. 3, pp. 403–414, 2020.

[88] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn et al., "Rt-2: Vision-language-

IEEE *Access*

action models transfer web knowledge to robotic control," arXiv preprint arXiv:2307.15818, 2023.

[89] S. Hu and J. Clune, "Thought cloning: Learning to think while acting by imitating human thinking," Advances in Neural Information Processing Systems, vol. 36, 2024.

[90] J. X. Wang, "Meta-learning in natural and artificial intelligence," Current Opinion in Behavioral Sciences, vol. 38, pp. 90–95, 2021.

[91] R. Mehrotra, M. Ansari, R. Agrawal, and R. Anand, "A transfer learning approach for ai-based classification of brain tumors," Machine Learning with Applications, vol. 2, p. 100003, 2020.

[92] H. Lee, J. Kim, J. Park, and S. Kang, "Strait: Self-test and self-recovery for ai accelerator," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 42, no. 9, pp. 3092–3104, 2023.

[93] A. K. Y. Yanamala and S. Suryadevara, "Adaptive middleware framework for context-aware pervasive computing environments," International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, vol. 13, no. 1, pp. 35–57, 2022.

[94] D. Driess, F. Xia, M. S. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Vuong, T. Yu et al., "Palm-e: An embodied multimodal language model," arXiv preprint arXiv:2303.03378, 2023.

[95] Y. Huang, C. Du, Z. Xue, X. Chen, H. Zhao, and L. Huang, "What makes multi-modal learning better than single (provably)," Advances in Neural Information Processing Systems, vol. 34, pp. 10 944–10 956, 2021.

[96] N. Balasubramaniam, M. Kauppinen, A. Rannisto, K. Hiekkanen, and S. Kujala, "Transparency and explainability of ai systems: From ethical guidelines to requirements," Information and Software Technology, vol. 159, p. 107197, 2023.

[97] F. Santoni de Sio and G. Mecacci, "Four responsibility gaps with artificial intelligence: Why they matter and how to address them," Philosophy & Technology, vol. 34, no. 4, pp. 1057–1084, 2021.

[98] N. A. Smuha, "From a 'race to ai' to a 'race to ai regulation': regulatory competition for artificial intelligence," Law, Innovation and Technology, vol. 13, no. 1, pp. 57–84, 2021.

[99] C. Small and C. Lew, "Mindfulness, moral reasoning and responsibility: Towards virtue in ethical decision-making," Journal of Business Ethics, vol. 169, no. 1, pp. 103–117, 2021.

[100] A. Langdon, M. Botvinick, H. Nakahara, K. Tanaka, M. Matsumoto, and R. Kanai, "Meta-learning, social cognition and consciousness in brains and machines," Neural Networks, vol. 145, pp. 80–89, 2022.

[101] V. Lai, C. Chen, Q. V. Liao, A. Smith-Renner, and C. Tan, "Towards a science of human-ai decision making: a survey of empirical studies," arXiv preprint arXiv:2112.11471, 2021.

DEEPAK ACHARYA is a scholar and teacher with distinguished research experience in the field of machine learning, deep learning, and computer science applications; he received his PhD and his Master of Science in Computer Science from The University of Alabama in Huntsville (UAH), where he is now a Principal Research Scientist at the Information Technology and Systems Center. He has applied advanced machine learning techniques, especially in NASA-funded projects, to develop super-resolution tools for precipitation data and Earth observation systems to address public health issues in sub-Saharan Africa.

His areas of interest encompass a broad spectrum of Machine Learning (ML) and Deep Learning (DL), including graph neural networks, clustering techniques, and Gumbel-Softmax distribution. He has used his proficiency in ML models to solve many novel domain problems in pattern recognition, predictive analytics, and data-driven decision-making. He understands the theoretical aspects of ML algorithms and the practical challenges of implementing them in real-world problems.

He also teaches part-time at UAH's Computer Science Department. He mentors graduate and undergraduate students and serves on academic review boards for several top-notch journals, helping to further the field of AI and computer science.

With advanced technical skill in engineering languages including Python, C++ and Java, as well as in frameworks such as React JS, he sits on the cusp of theoretical research and practical development. Dr Acharya's multidisciplinary approach makes him uniquely placed to advance industrial applications of AI as well as academic research, with his work at the leading edge of innovation. He works to empower talent, develop AI breakthroughs and foster responsible AI in several sectors.

KARTHEGEYAN KUPPAN is Vice President and Senior Manager of Software Engineering with more than 17 years of experience in designing, developing, and integrating complex software systems in multiple industries. Understanding multiple technologies such as Java, Python, Machine Learning, and Cloud, Karthegeyan Kuppan can produce new systems to meet today's needs while addressing future scalability and efficiency. He is a team leader with multiple years of experience managing cross-functional teams to deliver projects meeting technical requirements and business goals.

He holds a MCA (Master of Computer Applications) degree which he obtained from Anna University while perfecting his computer science and software engineering skills. Throughout his career, he and his team have achieved the highest standards in the delivery of IT services, achieving several large government projects. They retained their leadership position in the development of new IT solutions due to the tendency to follow or predict the development of new technological solutions. The goal is to build these solutions in such a way that they are highly reliable, secure as well as configurable to cater for future increased demand.

He is a life-long learner, a self-starter and a true high-potential professional with a penchant for continuous learning and growth. His passion for new opportunities have led him to participate in multiple open source software projects and further his knowledge in the field of knowledge management through the completion of his master's degree. Moreover, he has served as a mentor with the guidance of persistence and dedication that has influenced many intern students. His visionary approach to problem-solving and passion for innovation have garnered him industry influence.

IEEE *Access*

DIVYA B 's academic milestones include a Bachelor of Engineering (B.E.) in Electronics and Communication Engineering Degree awarded to her by Visvesvaraya Technological University (VTU), Belagavi, with a Master of Technology (M.Tech.) in Signal Processing from Siddaganga Institute of Technology, Tumkur. This specific knowledge of electronics and signal processing and her qualifications made her successful in the academic and research areas.

Currently, she holds the position of assistant professor in the Department of Electronics and Communication Engineering at Manipal Institute of Technology, which is located in Manipal. Her teaching career spans over 14 years, which is one of the factors that has earned her a reputation for dedication and passion towards her profession. She helps students hone their critical and analytical skills and ensures that the latest technology is included in the syllabus as needed.

Her expertise includes applied technology subparts, such as machine learning (ML), deep learning (DL), and signal processing, to which she also actively contributes research and development. Ms. Divya's interest lies in the application of Machine Learning and efforts related to signal processing and biomedical image processing. She has taken part in a number of projects, which implement ML and DL technologies for enhancing image analysis, data analysis, and prediction. All these efforts aim to build more efficient and precise systems that can be used in practice, particularly in the biomedical field. She is very passionate about applying machine learning to solve real-world problems in healthcare. As such, she is currently enrolled in a PhD program at NITK, Surathkal, in biomedical image processing. Ms. Divya is actively participating in the academic community undertaking responsibilities of a mentor, conducting researches, and actively learning the new tools and theories in the area of machine learning and deep learning technologies.

. . .