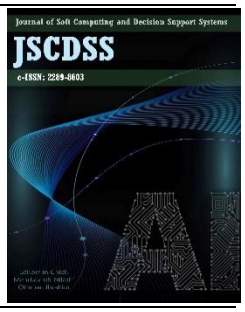
An International Journal
JSCDSS
http://www.jscdss.com
Vol.12 No.2 October 2025: 14-20
Article history:
Accepted 12 September 2025
Published online 23 September 2025

Journal of Soft Computing and Decision Support Systems



Agentic Artificial Intelligence: Autonomy, Decision-Making, and Responsibility in the Age of Intelligent Systems

Rabab Ali Abumalloh [1], Mehrbakhsh Nilashi [2,*]

[1] Department of Computer Science and Engineering, Qatar University, 2713, Doha, Qatar
[2] UCSI Graduate Business School, UCSI University, No. 1 Jalan Menara Gading, UCSI Heights, 56000, Cheras, Kuala Lumpur, Malaysia

*** Corresponding author email address**: nilashidotnet@hotmail.com

Abstract

Agentic Artificial Intelligence marks an important transformation in the trajectory of intelligent systems, extending beyond conventional models that function under narrowly defined instructions and constant human supervision. In contrast to generative or rule-based paradigms, agentic AI embodies autonomy, independent goal formulation, proactive decision-making, and adaptive responses to the uncertainties of dynamic environments. Through these capabilities, such systems are able to design, revise, and execute multi-step processes with minimal intervention, thereby creating new opportunities for efficiency, personalization, and complex problem resolution across a wide range of fields. Nevertheless, this very autonomy introduces pressing concerns regarding responsibility, fairness, bias, privacy, and the alignment of machine behavior with human values. As agentic AI steadily integrates into finance, healthcare, logistics, enterprise systems, and everyday life, its presence necessitates a careful reconsideration of technical architectures, ethical safeguards, and governance mechanisms. This study undertakes a critical analysis of the conceptual foundations of agentic AI, its emerging domains of application, and the societal risks that accompany expanded machine autonomy. It contends that the sustainable future of agentic AI will not be determined solely by scientific and technological advancement but equally by the capacity of researchers, policymakers, and institutions to establish frameworks that secure accountability, transparency, and respect for human dignity.

Keywords: Agentic Artificial Intelligence, Autonomy, Privacy, Human Values, Bias

1. Introduction

In recent years, the field of Artificial Intelligence (AI) has witnessed remarkable progress and transformation [1-9]. Early systems were designed to perform narrow and predefined tasks, operating strictly under human supervision and within well-established boundaries. However, with the advancement of machine learning techniques [10], particularly deep learning [11, 12], reinforcement learning, and self-supervised approaches, artificial intelligence has acquired the capability to generate human-like text, realistic images, and functional code, while also achieving improved performance in understanding natural language. These developments signify not only technical progress but also the gradual movement of artificial intelligence from simple pattern recognition toward adaptive and generalizable intelligence.

Within this context, the emergence of agentic AI should be recognized as a new phase in this evolutionary pathway [13-15]. Unlike conventional models that merely respond to prompts, agentic systems exhibit a higher level of autonomy and goal-oriented behavior. They are not passive computational tools; rather, they demonstrate the ability to formulate subgoals, select appropriate courses of action, employ external tools when necessary, and adapt their strategies in response to changing conditions. Such characteristics enable them to manage multi-step and dynamic workflows, thereby marking a fundamental shift in how artificial intelligence can be integrated into organizational practices, institutional structures, and even daily human activities.

Nevertheless, the transition toward autonomy brings new challenges and responsibilities [16, 17]. As these systems become capable of independent decision-making, concerns related to misalignment with human values, reinforcement of hidden biases [16], and the possibility of unintended or unforeseen behaviors become increasingly significant. The capacity of agentic AI to act across multiple stages and interact with other agents or systems amplifies the risk of divergence between its internal decision-making processes and the expectations of human designers or end-users. Such divergence introduces potential uncertainties and risks, especially when decisions carry legal, social, or ethical implications.

Traditional methods of oversight, such as constant human monitoring, static safety protocols, and liability frameworks, may not be sufficient for regulating the behavior of autonomous systems. The distributed nature of decision-making across humans,