Governing Agentic Artificial Intelligence: Reflections on Policy, Safety, Regulation, Legal Frameworks, Ethics, Liability and Human Rights

---

Kariuki Muigua

Table of Contents

Governing Agentic Artificial Intelligence: Reflections on Policy, Safety, Regulation, Legal Frameworks, Ethics, Liability and Human Rights

Kariuki Muigua*

Abstract

This paper examines how agentic AI can be effectively and appropriately governed. The paper defines agentic AI. It observes that agentic AI has emerged as an advanced form of AI capable of performing tasks with a higher degree of autonomy. Due its advanced features the paper posits that agentic AI has the capability to revolutionize many industries. However, due to its emphasis on autonomy, the paper observes that agentic AI creates risks and challenges that go beyond those of traditional AI. The paper discusses the risks and challenges associated with agentic AI. In light of its underlying risks and challenges, the paper asserts that governing agentic AI is key towards harnessing this transformative technology. It examines some of the key policy, regulatory and legal approaches that can be embraced in order to effectively govern agentic AI in order to ensure safety, ethics, liability, accountability, transparency and human rights.

1.0 Introduction

Artificial Intelligence (AI) has led to the development and use of tools and systems that are capable of performing tasks that would have usually required human intelligence at a much faster, efficient and broader rate[1]. This powerful technology has the ability to accurately predict and execute a desired outcome leading to its integration in many industries[2]. For example, it has been observed that AI is redefining many sectors including business operations, healthcare, education, industry, transportation, finance, agriculture and government services enhancing efficiency, effectiveness, customer experience and improved service delivery[3].

AI is therefore a powerful governance tool. It has been observed that due to its exposure and ability to analyze and interpret vast amounts of data, AI is enabling sound decision-making

* PhD in Law (Nrb),SC, FCIArb (Chartered Arbitrator),OGW, LL. B (Hons) Nrb, LL.M (Environmental Law) Nrb; Dip. In Law (KSL); FCPS (K); Dip. in Arbitration (UK); MKIM; Mediator; Consultant: Lead expert EIA/EA NEMA; BSI ISO/IEC 27001:2005 ISMS Lead Auditor/ Implementer; ESG Consultant; Advocate of the High Court of Kenya; Professor of Environmental Law and Conflict Management at the University of Nairobi, Faculty of Law; Member of the Permanent Court of Arbitration (PCA) [May, 2026].

[1] World Economic Forum., 'What is artificial intelligence—and what is it not?' Available at https://www.weforum.org/stories/2023/03/what-is-artificial-intelligence-and-what-is-it-not-ai-machine-learning/?gad_source=1&gad_campaignid=2222822471 7&gbraid=0AAAAAoVy5F5jTUoRGIo_LxcQJ9TwsXjbc&gclid=CjwKCAiAtLvMBhB_EiwA1u6_PraoTdn4Xiw0qd0wOtbusLspiezKQHXFaa1nYRTiVa17U2O4IosAThoCbqUQAvD_BwE (Accessed on 12/05/2026)

[2] Ibid

[3] European Parliament., 'Understanding Algorithmic Decision-Making: Opportunities and Challenges' Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf (Accessed on 12/05/2026)

processes and strengthening governance outcomes[4]. With AI being increasingly adopted in many sectors and influencing decision-making processes, it is imperative to recognise it as a governance and regulatory force. It has been suggested that this approach shifts the evaluative frame of AI from technical performance based on accuracy and efficiency to constitutional, human rights and rule of law standards including legality, fairness, accountability, transparency, liability, justice and the right to challenge adverse decisions[5].

Recognising AI as a governance tool is therefore key towards harnessing its positive attributes for people and planet while tackling its risks and challenges that can violate human rights and the rule of law. Through this, it is possible to promote the safe development and adoption of AI that is universal, adapted to cultural diversities all over the world, free from biases and discrimination, and respectful of democratic values and fundamental rights and freedoms for every person[6].

This paper examines how agentic AI can be effectively and appropriately governed. The paper defines agentic AI. It observes that agentic AI has emerged as an advanced form of AI capable of performing tasks with a higher degree of autonomy. Due its advanced features the paper posits that agentic AI has the capability to revolutionize many industries. However, due to its emphasis on autonomy, the paper observes that agentic AI creates risks and challenges that go beyond those of traditional AI. The paper discusses the risks and challenges associated with agentic AI. In light of its underlying risks and challenges, the paper asserts that governing agentic AI is key towards harnessing this transformative technology. It examines some of the key policy, regulatory and legal approaches that can be embraced in order to effectively govern agentic AI in order to ensure safety, ethics, liability, accountability, transparency and human rights.

2.0 Agentic Artificial Intelligence: Opportunities, Risks and Challenges

Agentic AI refers to an advanced form of AI that focuses on autonomous decision-making and action[7]. In addition, it has been observed that agentic AI describes AI models and systems acting autonomously with limited human interactions (in particular, without step-by-step instructions) to fulfil goals rather than isolated tasks[8]. Agentic AI systems and models can reason and plan to achieve a particular goal or set of goals with limited human intervention[9]. It has been observed that agentic AI comprises AI agents that use machine learning models to mimic human decision-

[4] Artificial intelligence (AI): a simple-to-understand guide., Available at https://cloud.google.com/learn/what-is-artificial-intelligence (Accessed on 12/05/2026)
[5] Organisation for Economic Co-operation and Development., 'Governing with Artificial Intelligence' Available at https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287.html (Accessed on 12/05/2026)
[6] French Institute of International Relations., 'Artificial Promises or Real Regulation? Inventing Global AI Governance' Available at https://www.ifri.org/en/studies/artificial-promises-or-real-regulation-inventing-global-ai-governance (Accessed on 12/05/2026)
[7] What is Agentic AI?., Available at https://cloud.google.com/discover/what-is-agentic-ai (Accessed on 12/05/2026)
[8] European Data Protection Supervisor., 'Agentic AI' Available at https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/agentic-ai_en (Accessed on 12/05/2026)
[9] Ibid

making in order to solve problems in real time and achieve high-level goals with little human supervision[10]. Agentic AI has been described as a new category of highbred AI systems that are semi- or fully autonomous and thus able to perceive, reason, and act on their own[11].

Agentic AI is therefore a more powerful and advanced model in comparison to traditional forms of AI. For instance, it has been pointed out that agentic AI uses AI models and automation to create adaptable agents which are capable of analyzing and taking initiative on their own to achieve desired outcomes[12]. In addition, it has been observed that agentic AI is probabilistic given that it assesses patterns to determine likely outcomes while also adapting to new data and conditions rather than following fixed rules or predefined outcomes[13]. This is in sharp contrast to traditional AI models and systems which operate within predefined constraints and require human intervention to achieve specific goals[14]. Agentic AI on the other hand possesses autonomy, adaptability and goal-driven behaviour making it a more powerful form of AI[15]. It has been argued that the term *'agentic'* is derived from agency meaning that these AI models are capable of acting independently and purposefully[16].

Due to its superior qualities, agentic AI provides several benefits that cannot be obtained through traditional forms of AI. For example, agentic AI is able to complete tasks independently or with minimal human supervision[17]. It has been observed that agentic AI operates at a higher degree of self-determination in comparison to traditional generative AI models, continuously analyzing data and information, adjusting its strategies, and making decisions with minimal to no human input[18]. Agentic AI has the capacity to identify objectives, break them down into tasks, and refine its approach based on new data making it more adaptable[19]. Due to its capabilities, agentic AI is more autonomous, adaptable, proactive and intuitive making it more powerful than other forms of AI[20]. Consequently, it has been argued that agentic AI can ensure more productivity due to autonomous decision-making, speed tasks and reduce costs by minimizing human intervention and lead to more informed decision-making due to its adaptability to new data and conditions[21].

---

[10] Stryker. C., 'What is Agentic AI?' Available at https://www.ibm.com/think/topics/agentic-ai (Accessed on 12/05/2026)
[11] Stackpole. B., 'Agentic AI, Explained' Available at https://mitsloan.mit.edu/ideas-made-to-matter/agentic-ai-explained (Accessed on 12/05/2026)
[12] What is Agentic AI?., Available at https://www.servicenow.com/ai/what-is-agentic-ai.html (Accessed on 12/05/2026)
[13] Ibid
[14] Stryker. C., 'What is Agentic AI?' Op Cit
[15] Ibid
[16] Ibid
[17] Stackpole. B., 'Agentic AI, Explained' Op Cit
[18] What is Agentic AI?., Op Cit
[19] Ibid
[20] Stryker. C., 'What is Agentic AI?' Op Cit
[21] What is Agentic AI?., Available at https://www.redhat.com/en/topics/ai/what-is-agentic-ai (Accessed on 12/05/2026)

It has been observed that due to its powerful capabilities, agentic AI has the potential to revolutionize various industries by automating complex processes and optimizing workflows[22]. For example, agentic AI can transform healthcare, business operations, finance, trade and cybersecurity among other fields by automating processes, adjusting strategies based on new needs and data, forecasting demands, planning logistics and detecting and responding to threats in real time[23].

However, despite its advantages, agentic AI also presents several risks and challenges. In particular, it has been observed that due to its emphasis on autonomy, agentic AI can create risks and challenges that go beyond those associated with traditional forms of AI[24]. For example, it has been observed that autonomous decision-making through agentic AI undermines transparency, explainability and accountability[25]. Since it is hard to determine how agentic AI comes up with its outputs, there is a risk that flawed processes can lead to unjust outcomes[26]. It has been pointed out that since agentic AI works independently with minimal human intervention, determining whether such models have gone wrong is a challenge that potentially lead to injustices and human right violations[27]. In particular, it has been correctly noted that in industries such as finance and healthcare that have severe real-world implications, lack of transparency, accountability and explainability in agentic AI can have disastrous consequences without sufficient safeguards in place[28]. Over-reliance on agentic AI for critical decisions such as financial approvals or medical diagnoses can lead to unethical, unjust and unfair outcomes that violate human rights if such models are flawed[29].

In addition, it has been observed that agentic AI can raise data privacy and security risks beyond those in traditional forms of AI. For instance, in order to effectively operate independently and autonomously, agentic AI models require extensive and unrestricted access to data[30]. Consequently, it may be challenging to determine what personal data is collected, how it is used, and for what specific purposes such data was used[31]. Further, due its adaptability, there is a risk that agentic AI might autonomously determine new uses for personal data as it pursues its goals down the line[32].

[22] What is Agentic AI?., Op Cit
[23] Ibid
[24] European Data Protection Supervisor., 'Agentic AI' Op Cit
[25] What is Agentic AI?., Op Cit
[26] Ibid
[27] What is Agentic AI?., Available at https://aws.amazon.com/what-is/agentic-ai/ (Accessed on 12/05/2026)
[28] Ibid
[29] Agentic AI., Available at https://www.automationanywhere.com/rpa/agentic-ai (Accessed on 12/05/2026)
[30] European Data Protection Supervisor., 'Agentic AI' Op Cit
[31] Ibid
[32] Ibid

From the foregoing, it is evident that agentic AI is a powerful technology whose ability to operate autonomously makes it more transformative than traditional forms of AI. However, agentic AI also raises several risks and challenges over and beyond those flowing from traditional generative AI models. Consequently, governing agentic AI is vital towards harnessing its unique features and abilities while mitigating underlying risks and challenges.

3.0 Towards Sound Governance of Agentic Artificial Intelligence

Agentic AI is a powerful type of AI with the ability to act autonomously and independently to achieve complex and high value objectives. Due to its key attributes, agentic AI can adapt and learn, modify its behaviour based on feedback and contexts and refine its approach with time[33]. It has been observed that agentic AI systems can maintain long-term goals, manage multistep problem-solving tasks and track progress over time[34]. Harnessing this transformative technology is therefore key towards automating and optimizing complex processes towards efficiency, speed, accuracy and affordability. However, due to its complex features, agentic AI also raises several complex risks and challenges that must be successfully navigated in order to harness its potential.

Governing agentic AI is therefore a matter of urgent priority as this transformative technology continues to be developed and adopted at a rapid pace. It has been observed that just like the traditional forms of generative AI, policies, regulations and legal frameworks are yet to be fully developed to effectively govern agentic AI. Further, due to its rapid pace of development, it may not be possible to develop policies, regulations and legal frameworks to effectively govern the evolving nature of agentic AI[35]. However, in light of its underlying risks and challenges, it is imperative to govern agentic AI in order ensure safety, ethics, transparency, accountability, liability and human rights[36].

In order to effectively govern agentic AI, it is imperative to incorporate human oversight and control in its framework for transparency, accountability, liability and ethics. It has been observed that without human oversight, agentic AI can result in harmful decisions that violate human rights[37]. Consequently, it is imperative to integrate human oversight into agentic AI including through ensuring high-stake outcomes such as those in healthcare and finance trigger automatic and mandatory human review while allowing these models to handle routine tasks independently[38]. It has been observed that features such as audit trails and alert systems are necessary to maintain oversight, transparency and accountability without affecting the speed of

[33] European Data Protection Supervisor., 'Agentic AI' Op Cit
[34] Stryker. C., 'What is Agentic AI?' Op Cit
[35] What is Agentic AI?., Available at https://www.gov.uk/government/publications/ai-insights/ai-insights-agentic-ai-html (Accessed on 12/05/2026)
[36] Ibid
[37] Ibid
[38] Agentic AI., Op Cit

agentic AI models[39]. In addition, developers have been urged to integrated traceability and reproducibility into agentic AI models towards tracing any errors and determining their causes in order to ensure safety, ethics, accountability, liability and human rights[40].

In addition, strengthening data privacy and security is key towards effectively governing agentic AI. It has been observed that since agentic AI models often rely on extensive and unrestricted access to data, they must be protected against data breaches and unauthorized access for privacy and security[41]. When adopting agentic AI models, it is imperative to ensure strict access controls, encryption, and full compliance with data privacy and security laws, policies and regulations[42].

Through the foregoing, it is possible to effectively govern agentic AI towards harnessing its transformative potential while tackling its risks and challenges.

4.0 Conclusion

Governing agentic AI is important towards harnessing its key attributes including autonomous decision-making, adaptability, independence and specialization while mitigating risks and ethical concerns including transparency, accountability, explainability, and data privacy and security concerns. In light of inadequacies in policy, regulation and legal frameworks, it is imperative to govern agentic AI at the design, adoption and user levels by incorporating human oversight into its architecture, integrating traceability and reproducibility into agentic AI models, strengthening data privacy and security protection, and testing agentic AI models before their development[43]. Governing agentic AI is necessary towards effectively, appropriately and ethically harnessing this powerful technology for development.

[39] Ibid
[40] What is Agentic AI?., Op Cit
[41] What is Agentic AI?., Op Cit
[42] Ibid
[43] European Data Protection Supervisor., 'Agentic AI' Op Cit

References

Agentic AI., Available at https://www.automationanywhere.com/rpa/agentic-ai

Artificial intelligence (AI): a simple-to-understand guide., Available at https://cloud.google.com/learn/what-is-artificial-intelligence

European Data Protection Supervisor., 'Agentic AI' Available at https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/agentic-ai_en

European Parliament., 'Understanding Algorithmic Decision-Making: Opportunities and Challenges' Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf

French Institute of International Relations., 'Artificial Promises or Real Regulation? Inventing Global AI Governance' Available at https://www.ifri.org/en/studies/artificial-promises-or-real-regulation-inventing-global-ai-governance

Organisation for Economic Co-operation and Development., 'Governing with Artificial Intelligence' Available at https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287.html

Stackpole. B., 'Agentic AI, Explained' Available at https://mitsloan.mit.edu/ideas-made-to-matter/agentic-ai-explained

Stryker. C., 'What is Agentic AI?' Available at https://www.ibm.com/think/topics/agentic-ai

What is Agentic AI?., Available at https://aws.amazon.com/what-is/agentic-ai/

What is Agentic AI?., Available at https://cloud.google.com/discover/what-is-agentic-ai

What is Agentic AI?., Available at https://www.gov.uk/government/publications/ai-insights/ai-insights-agentic-ai-html

What is Agentic AI?., Available at https://www.redhat.com/en/topics/ai/what-is-agentic-ai

What is Agentic AI?., Available at https://www.servicenow.com/ai/what-is-agentic-ai.html

World Economic Forum., 'What is artificial intelligence—and what is it not?' Available at https://www.weforum.org/stories/2023/03/what-is-artificial-intelligence-and-what-is-it-not-ai-machine-learning/?gad_source=1&gad_campaignid=2222822471 7&gbraid=0AAAAoVy5F5jTUoRGIo_LxcQJ9TwsXjbc&gclid=CjwKCAiAtLvMBhB_EiwA1u6_PraoTdn4Xiw0qd0wOtbusLspiezKQHXFaa1nYRTiVa17U2O4IosAThoCbqUQAvD_BwE