Research article
UDC 34:004:340.1:004.8
EDN: https://elibrary.ru/pltfwo
DOI: https://doi.org/10.21202/jdtl.2025.17

Agentic Artificial Intelligence: Legal and Ethical Challenges of Autonomous Systems

Gordon Bowen

Anglia Ruskin University, Cambridge, United Kingdom

Keywords

agentic artificial intelligence,
artificial intelligence,
autonomy,
digital technologies,
ethics,
law,
legal regulation,
liability,
programming,
risk

Abstract

Objective: to identify specific legal and ethical problems of agentic artificial intelligence and develop recommendations for the creation of protective mechanisms to ensure the responsible functioning of autonomous AI systems.

Methods: the research is conceptual in nature and is based on a systematic analysis of scientific literature on the ethics of artificial intelligence, legal regulation of autonomous systems and social interaction of AI agents. The work uses a comparative analysis of various types of AI systems, a study of the potential risks and benefits of agentic artificial intelligence, as well as an interdisciplinary approach that integrates advances in law, ethics, and computer science to form a comprehensive understanding of the issue.

Results: the research has established that agentic artificial intelligence, possessing the decision-making autonomy and ability to social interaction, creates qualitatively new legal and ethical challenges compared to traditional AI assistants. The main categories of potential harm were identified: direct impact on users through overt and covert actions, manipulative influence on behavior, and cumulative harm from prolonged interaction. The author stipulates the need for distributing responsibility between three key actors: the user, the developer and the owner of the agentic artificial intelligence system.

Scientific novelty: for the first time, the research presents a systematic analysis of the ethical aspects of agentic artificial intelligence as a qualitatively new class of autonomous systems that differ from traditional AI assistants in the degree of independence and social interactivity. The author developed a typology of potential risks of social interaction with agent-based intelligent systems and proposes a conceptual model for the distribution of legal and ethical responsibilities in the user-developer-owner triad.

Practical significance: the research forms the theoretical basis for the development of ethical principles and legal norms governing agentic based artificial intelligence in a growing market for autonomous intelligent systems. The findings will be useful for legislators creating a regulatory framework, developers designing protective mechanisms, as well as organizations implementing agentic artificial intelligence systems in various economic fields.

For citation

Bowen, G. (2025). Agentic Artificial Intelligence: Legal and Ethical Challenges of Autonomous Systems. *Journal of Digital Technologies and Law, 3*(3), 431–445. https://doi.org/10.21202/jdtl.2025.17

Contents

Introduction

The capabilities of AI agents, such as communications skills and sophisticated reasoning that does not require human intervention, are increasing. AI assistants are anchored to users, but agentic AI agents have degrees of freedom[1]. The global market value of agentic AI in 2024 was US$5.1 billion, which is expected to increase to US$47 billion by 2030, with a compound annual growth rate of 44 % [2]. The degree of freedom now experienced by agentic AI agents re- quires legal and ethical frameworks to regulate agentic AI agents' behaviour. The owner/developer of an agentic AI agent and an agentic AI agent's software behaviour both need monitoring from a legal and ethical perspective. However, a balance needs to be struck to not lose the advantages of agentic AI agents. The ethics and legal frameworks for "static" AI assistants, which are controlled or tethered to ethical behaviour, require redefining for agentic AI agents. How should these frameworks be changed for agentic AI agents? Do agentic AI agents require social consciousness

1 Morris, B. (2024). Beyond Intelligence: The Impact of Advanced AI Agents. https://clck.ru/3NedB2

2 Vailshery, L. S. (2025). Global market value of agentic AI 2030. https://clck.ru/3NedDe

to navigate the new ethics and legal environments? The paper is structured as follows: introduction, literature review, implications and conclusion. The overarching aim of the paper is to understand how the debate on ethics and the legal landscape for agentic AI agents will need to evolve.

1. Literature review

1.1. Three types of AI agents

AI agents are also known as compound AI systems and are a growing area of research (Kapoor et al., 2024). Compound AI systems are the best way to leverage and maximise AI models and may have been one of the most important trends in 2024[3]. Compound AI systems differ from AI systems (large language models) in many ways, for example, they tackle and complete harder tasks, they have more real-world use and can solve problems that do not have a single answer; compound AI systems might require custom-built agent–computer interfaces (Yang, Jimenez et al., 2024). A compound AI system can manage several agents, which has a price implication (Kapoor et al., 2024).

In traditional AI, agents were considered to be able to perceive and act upon the environment (Russell & Norvig, 1995); from a traditional perspective, a thermostat could be classified as an agent (Kapoor et al., 2024). Agentic AI systems are often viewed as a spectrum of AI systems with more or less agentic capability[4]. There are three types of AI agents (Alberts, Keeling et al., 2024): Artifacts (an agentic agent interprets data in a social environment), Interactive Systems (behave as a social actor) and Conversational Agents (the agents have social roles). AI agents as social actors need to be conversational and thus interact with the user, but this goes beyond being agreeable, friendly, truthful and not using harmful language. Interactions are expected to be contextual, which requires awareness of the individual user, the social environment and the situational context. This will translate into AI agents giving information unprompted, which leads to them making suggestions (Alberts, Keeling et al., 2024).

1.2. The social interaction of AI agents

Technologies can be considered social because they are situated and embedded in social environments. This assumption is accepted by researchers who research the ideological perspective of technology, such as culture biases and values that are embedded in the technologies we use (Bender et al., 2021; Shelby et al., 2023). Systems that are not socially interactive can be seen to be harmful (Alberts, Keeling et al., 2024). An example

[3] Zaharia, M., Khattab, O., Chen, L., Davis, J. Q., Miller, H., Potts, C., Zou, J., Carbin, M., Frankle, J., Rao, N., & Ghodsi, A. (2024). The Shift from Models to Compound AI Systems. https://clck.ru/3NedKd

[4] Ng, A. (2024). Welcoming Diverse Approaches Keeps Machine Learning Strong. https://clck.ru/3NedNZ

of harm in a "passive" technological system includes training data that misrepresent demographics (Bender et al., 2021).

Looking beyond passive technologies, people treat interactive technology as having a purpose/intention and social meaning (Grimes et al., 2021). Furthermore, the situated actions of a system are interpreted in human social ways, which is a central tenet of the Computers Are Social Actors research philosophy, where humans interacting with computer systems apply human social norms and expectations to their interactions with technology (Nass et al., 1994).

Interactive systems mimic human behaviour or qualities. This is achieved through social cues (speaking in the first person and expressing emotion) (Grimes et al., 2021). Conversational agents can respond in familiar social situations. An AI system could be situated in a social role in which the AI agent is a friend or therapist; however, the AI agent could utter something offensive that could upset the user, or familiarity in a relationship with the user could breed contempt and insensitivity (Alberts, Keeling et al., 2024). Social interactions that could cause harm are categorised as follows (Alberts, Keeling et al., 2024):

– Direct harm to the user – giving rise to overt action such as offensiveness due to the language used or behaviour.

– Direct harm to the user – giving rise to covert action such as opinions that appear positive or neutral.

– Interactions that exert a harmful influence on behaviour – misleading or giving false information.

– Interactions that exert a harmful influence on behaviour – manipulating and persuading users to do things they would not normally do.

– Interactions that collectively harm users – harm emerging over time in relationships.

Harm from interactions means that harm resides inside language (Shelby et al., 2023). Direct harm to the user is dependent on the contextual language and relational language. Language can be conceived as positive (endearment using derogatory language) or less positive (women or elderly being patronised) depending on the situation (Coghlan et al., 2021).

Interactions from an influencer cause secondary harm effects by influencing the thinking or doing of an individual. Thus, agentic AI can have undue influence on an individual by producing false or misleading behaviour in the individual (Alberts, Keeling et al., 2024). Engaging in social cues makes systems more intuitive and engaging (Kocielnik et al., 2021); humans react to social cues emotionally and not rationally, and this can be used to manipulate individuals (Alberts, Lyngs et al., 2024; Shamsudhin & Jotterand, 2021).

Interactions that collectively cause harm include dismissive actions that are tactless and controlling. The collective effect of harm is cumulative, for example, a single instance of tactlessness can be ignored, but if it is repeated then it could affect an individual adversely over time (Alberts, Keeling et al., 2024).

1.3. Agentic AI decision making

Agentic AI systems require unprecedented autonomy and contextual awareness (Martinez & Kifle, 2024; Mohanarangan et al., 2024). The decision-making process in the agentic AI algorithm needs to be revolutionary to fulfil the requirements to work independently and make logical and coherent decisions in the environment it operates within. The algorithm will be making real-time decisions and synthesising complex data and datasets (Abuelsaad et al., 2024). Agentic AI has two capabilities that go beyond the capabilities of AI assistants. The first is decision making that operates at different levels, from low-level responses to high-level strategic responses, which requires long-term thoughtfulness (Abuelsaad et al., 2024). A second capability is the move from reactive to proactive goal-oriented behaviour, which requires the system to identify complex tasks and determine the necessary subtasks. Thus, to pursue its objectives, agentic AI will require a flexible architecture in its goal management software (Martinez & Kifle, 2024). Agentic AI requires an adaptive learning style that can harness different learning styles, and the ability to reinforce learning so it can proactively apply the learning style that is appropriate to situations. The adaptive learning system must be able to learn from experience (Abuelsaad et al., 2024).

Agentic AI acts as an outsourcer for an organisation. Early adopters will have a first mover advantage (market position, innovation, customer relations, operational efficiencies, learning curve, market share), and last movers will potentially lose their competitive advantage (incur a loss of market share and increased costs, experience slowness in business and process innovation, lag in personalisation of customer services, experience higher opportunity costs leading to higher operational costs, miss early learning opportunities, and potentially experience a higher barrier to entry and pay a lower entry fee to enter the market but with fewer tests) (Beulen et al., 2022). Agentic AI offers many benefits: positive impacts on operating costs; higher efficiency because AI can perform tasks automatically and with greater accuracy; scalability without the need for additional resources and investment; and core goal focus, because organisations can focus on their core business activities and leave the minor or less important activities to AI (Hosseini & Seilani, 2025). However, the use of agentic AI has drawbacks: dependence on technology (overdependence

on technology could lead to operational disruption in AI service); limited range of personalisation because many of the tasks require extensive customisation; privacy and security issues, for example, the outsourcing of data to third parties raises concerns on privacy and security; and hidden costs relating to training, deployment and implementation.

Applications of agentic AI will span many industries, including robotics and manufacturing[5], healthcare systems[6], transport and logistics[7], traffic management systems[8] and financial services[9]. One emerging application of agentic AI is dynamic patient needs, which leads to personalised medicines (Hasan et al., 2025). In this application, agentic AI manages patients with chronic illnesses by overseeing patient history and sending reminders to patients (Yang, Garcia et al., 2024); this leads to recommendations on treatment using health indicators. This type of agentic AI system could manage individualistic patient care management and monitor for early warning signs of the health progression of patients, especially for older patients (Acharya et al., 2025). Another application of agentic AI is the automatic generation of new content that targets wider audiences and meets content requirements based on set criteria. This application would be helpful to marketing activities, such as sending customised emails to customers and potential customers. Literature searches by businesses, scientists and academics would be faster with agentic AI, and lead to new thoughts and ideas. Agentic AI could empower drug discovery, development and delivery (Gao et al., 2024).

Agentic AI research is gaining traction in moral reasoning and ethical decision making (Small & Lew, 2021). The importance of privacy and security in the handling of sensitive information has pushed this type of research to the fore. Research on moral reasoning seeks to establish an ethical basis for autonomous systems so that agentic AI systems can select actions with thought of their effects and values. Thus, in this context, the integration of psychology, ethics and philosophy create an overarching goal for AI systems that is ethical. All agentic systems need to be ethical in their decision making, especially health, autonomous, and law and order systems, because these decisions influence society (Acharya et al., 2025).

---

5 Randieri, C. (2025, January 3). Agentic AI: A New Paradigm In Autonomous Artificial Intelligence. Forbes. https://clck.ru/3NedZf

6 Automation Anywhere. (n.d.). What is agentic AI? Key benefits & features. https://clck.ru/3Nedsx

7 Ibid.

8 Randieri, C. (2025, January 3). Agentic AI: A New Paradigm In Autonomous Artificial Intelligence. Forbes. https://clck.ru/3NedZf

9 Ibid; Automation Anywhere. (n.d.). What is agentic AI? Key benefits & features. https://clck.ru/3Nedsx

Agentic AI systems require self-awareness and meta-cognition (Langdon et al., 2022), which can be achieved by building systems that understand their actions, abilities and limitations as self-referential knowledge. Self-awareness in AI systems can be done through self-evaluation on whether they have carried out tasks optimally, what can be improved, and what actions should be taken when failures occur or performance is poor. Self-agency skills (carrying out of the task) and ability (to detect the need to carry out the task) will enable agentic AI to assess its strategies and learning processes to improve the effectiveness of its decision making. Progress in research on self-awareness and meta-cognition might lead to more flexible and sophisticated agentic AI systems, thereby leading to enhanced and improved performance with the robustness to operate in multi-environments (Acharya et al., 2025). This will require further research on creating AI agency models, adaptive moral frameworks and contextual decision making (Lai et al., 2021).

2. Implications

Compound AI agents are more powerful than AI systems; thus, the ethical and legal issues are more complex. This is exacerbated by the social independence of agentic AI agents. A user/owner/developer has a degree of control over passive AI systems (AI assistants) because passive AI systems are tethered to a position and are singular in problem solving or tasks.

The owner/developer of an AI agent, the user of agentic AI, and agentic AI agent algorithm all need to behave ethically. An ethical and legal perspective of an agentic agent's interactions with a user needs to be taken so that the agentic agent behaves responsibly and does not cause harm. The technical developer and the owner of the algorithm need to ensure that an agentic AI system is applied ethically and legally. Why? Agentic software becomes independent once released; thus, guardrails need to be in place to monitor its behaviour and personality. Where does the legal responsibility lie if an agentic AI system runs amok or goes rogue? What if it causes harm, for example in the retrieving of data and information from a third party because it has a degree of decision-making capability, and one could argue consciousness[10] (Lim et al., 2025). However, the user of the AI agentic algorithm might have some ethical and legal responsibilities. What if a user asked an agentic AI system to do something that is unethical and illegal, such as transferring information without due process? In this situation, who is liable? Is it the user or the developer/owner of the AI algorithm? What if the relationship between an agentic AI and a user becomes toxic, and the agentic AI goes rogue and causes harm

[10] Al-Sibai, N. (2022). OpenAI Chief Scientist Says Advanced AI May Already Be Conscious. https://clck.ru/3Nee2Z

(Alberts, Keeling et al., 2024)? Agentic AI can contextualise the environmental landscape, so they have a sense of awareness, but does this let the user off the hook? There are similarities between agentic AI and autonomous automobiles. Parties will try to absolve themselves of blame and liability.

The problems and issues identified are not that prevalent with AI assistants. Moving from passive technological systems to interactive systems raises additional concerns not only about the legal and ethical implications but also about the scope and ability of agentic AI systems. Agentic AI is the future direction of AI, and it is becoming unstoppable, given the many benefits; however, there are challenges which need to be acknowledged and acted upon by the AI community for the protection of society and humankind. Treating individuals with respect is the starting point to making agentic AI ethically and legally responsible. Basic Psychological Needs Theory and the field of human–robot interaction could assist in the development of suitable frameworks (Li et al., 2025; Hosseini & Seilani, 2025; Korzynski et al., 2025; Kshteri, 2025).

New applications of agentic AI have been established in healthcare, logistics and transportation, and financial services. However, security and privacy concerns are rife because of the level of data and the independence of actions in the decision making of agentic AI. It makes decisions by breaking down complex tasks and compartmentalising them into subunits. The question is: How robust is the decision-making architecture and the understanding of the environmental ecosystem in which the decision-making process operates? Reliability and accuracy of the decisions is underpinned by, and dependent on, these areas. The decision-making process and the environmental ecosystem are starting points and foundational for the decision outcomes. An unreliability of the foundational aspects of agentic AI could contribute to the algorithm running amok and suffering from hallucinations. The range of applications and emerging applications of agentic AI makes it necessary for guardrails to be implemented at all levels of the architecture, which requires a hierarchical architecture of the agentic AI system. However, feedback will be needed to test subsystems of the various architectures of the algorithm so parts that are underperforming or are exhibiting worrying behaviour can be isolated or corrected. Will this require redundancy in the agentic AI architecture? If this is the case, then costs to own and implement agentic systems will rise. The application of the reasoning and consciousness that are speculated to exist in AI systems could be a pointer in the right direction. The emailing of customers and potential customers by agentic AI has business-related risks; thus, guardrails are necessary throughout the agentic AI system. Business-affecting problems risk reputational damage, brand credibility and relationship damage. The benefits of applying agentic AI in new and emerging applications, such as drug development, could drive the application without the necessary guardrails and reliability in the architecture in place. Does the societal value of agentic AI outweigh the benefits of ensuring rigorousness in safeguarding the regulatory and legal frameworks? Should agentic AI safeguarding, legal and regulatory, be more trial and error or experiential learning?

Conclusions

Compound AI agents (agentic AI) have many benefits that range from being able to work independently to the ability to reason; hence, they have a level of awareness and consciousness. Nevertheless, there are dangers, which requires a balanced approached to the deployment of agentic AI. Autonomous automobile scenarios are applicable to agentic AI, and lessons learnt from the autonomous automobile industry is a good starting point to understand the ethical and legal situations that are applicable to agentic AI. The risks of agentic AI need to be balanced with suitable guardrails that do not reduce or hinder innovation in the application of agentic AI. This requires an evolving legal and ethical framework that continues to protect society but also delivers the benefits of agentic AI to businesses and industries. Agentic AI takes decision making from human–machine interface to machine–machine interaction without the need for human intervention in decisions, but what if something goes wrong. Guardrails need to be implemented that are rigorous, resilient and robust to sustain ethical and legal frameworks.

References

Abuelsaad, T., Akkil, D., Dey, P., Jagmohan, A., & Vempaty, A. (2024). Agent-E: From Autonomous Web Navigation to Foundational Design Principles in Agentic Systems. *arXiv preprint arXiv:2407.13032*. https://doi.org/10.48550/arXiv.2407.13032

Acharya, D. B., Kuppan, K., & Ashwin, D. B. (2025). Agentic AI: Autonomous intelligence for complex goals – a comprehensive survey. In *IEEE Access (*vol. 13, pp. 18912–18936). https://doi.org/10.1109/ACCESS.2025.3532853

Alberts, L., Keeling, G., & McCroskery, A. (2024). Should agentic conversational AI change how we think about ethics? Characterising an interactional ethics centred on respect. *arXiv:2401.09082v2*. https://doi.org/10.48550/arXiv.2401.09082

Alberts, L., Lyngs, U., & Van Kleek, M. (2024). Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces that Act Socially. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–25. https://doi.org/10.1145/3653693

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Virtual Event Canada: ACM. https://doi.org/10.1145/3442188.3445922

Beulen, E., Plugge, A., & van Hillegersberg. J. (2022). Formal and relational governance of artificial intelligence outsourcing. *Information System E Business Management*, *20*(4), 719–748. https://doi.org/10.1007/s10257-022-00562-7

Coghlan, S., Waycott, J., Lazar, A., & Neves, B. (2021). Dignity, Autonomy, and Style of Company: Dimensions Older Adults Consider for Robot Companions. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–25. https://doi.org/10.1145/3449178

Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J. R., Ektefaie, Y., Kondic, J., & Zitnik, M. (2024). Empowering biomedical discovery with AI agents. *Cell*, *187*(22), 6125–6151. https://doi.org/10.1016/j.cell.2024.09.022

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*, 113515.

Hasan, S. S., Fury, M. S., Woo, J. J., Kunze, K. N., & Ramkumar, P. N. (2025). Ethical Application of Generative Artificial Intelligence in Medicine. *Arthroscopy: Journal of Arthroscopic Related Surgery*, *41*(4), 874–885. https://doi.org/10.1016/j.arthro.2024.12.011

Hosseini, S., & Seilani, H. (2025). The Role of Agentic AI in Shaping a Smart Future: A Systematic review. *Array*, *26*, 100399. https://doi.org/10.1016/j.array.2025.100399

Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., & Narayanan, A. (2024). AI Agents That Matter. *arXiv:2407.01502v1*.

Kocielnik, R., Langevin, R., George, J. S., Akenaga, S., Wang, A., Jones, D. P., Argyle, A., Fockele, C., Anderson, L., Hsieh, D. T., Kabir, Y., Duber, H., Hsieh, G., & Hartzler, A. L. (2021). Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *3rd Conference on Conversational User Interfaces (CUI '21), July 27–29, 2021, Bilbao (online), Spain*. ACM, New York, NY, USA. https://doi.org/10.1145/3469595.3469599

Korzynski, P., Edwards, A., Gupta, M. C., Mazurek, G., & Wirtz, J. (2025). Humanoid robotics and agentic AI: reframing management theories and future research directions. *European Management Journa*l, *43*(4), 548–560. https://doi.org/10.1016/j.emj.2025.06.002

Kshetri, N. (2025). Transforming cybersecurity with agentic AI to combat emerging cyber threats. *Telecommunications Policy*, *49*(6), 102976. https://doi.org/10.1016/j.telpol.2025.102976

Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). Towards a science of human-AI decision making: A survey of empirical studies. *arXiv:2112.11471*. https://doi.org/10.48550/arXiv.2112.11471

Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M., & Kanai, R. (2022). Meta-learning, social cognition and consciousness in brains and machines. *Neural Network*, *145*, 80–89. https://doi.org/10.1016/j.neunet.2021.10.004

Li, X., Shi, W., Zhang, H., Peng, C., Wu, S., & Tong, W. (2025). The Agentic-AI Core: an AI-Empowered, Mission-Oriented core network for Next-Generation mobile telecommunications. *Engineering*. https://doi.org/10.1016/j.eng.2025.06.027

Lim, S., Schmälzle, R., & Bente, G. (2025). Artificial Social Influence via Human-Embodied AI Agent Interaction in Immersive Virtual Reality (VR): Effects of Similarity-Matching during health conversations. *Computers in Human Behavior Artificial Humans*, *5*, 100172. https://doi.org/10.1016/j.chbah.2025.100172

Martinez, D. R., & Kifle, B. M. (2024). *Artificial Intelligence: A Systems Approach from Architecture Principles to Deployment*. MIT Press eBooks, IEEE Xplore2. https://doi.org/10.7551/mitpress/14806.001.0001

Mohanarangan, S., Karthika, D., Moohambigai, B., & Sangeetha, R. (2024). Unleashing the Power of AI and Machine Learning: Integration Strategies for IoT Systems. *International Journal of Scientific Research in Computer Science and Engineering*, *12*(2), 25–32.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). https://doi.org/10.1145/259963.260288

Russell, S. J., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall. Google-Books-ID: CUVeMwAACAAJ.

Shamsudhin, N., & Jotterand, F. (2021). Social Robots and Dark Patterns: Where Does Persuasion End and Deception Begin? In F. Jotterand, & M. Ienca (Eds.), *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues* (pp. 89–110). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-74188-4_7

Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., Yilla, N., Gallegos, J., Smart, A., Garcia, E., & Virk, G. (2023). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. *arXiv:2210.05791*. https://doi.org/10.48550/arXiv.2210.05791

Small., C., & Lew, C. (2021). Mindfulness, moral reasoning and responsibility: Towards virtue in ethical decision-making. *Journal of Business Ethics*, *169*(1), 103–117. https://doi.org/10.1007/s10551-019-04272-y

Yang, J., Jimenez, C. E., Wettig, A., Lieret, K., Yao, S., Narasimhan, K., & Press, O. (2024). SWE-AGENT: Agent-Computer Interfaces Enable Automated Software Engineering. *arXiv:2405.15793*. https://doi.org/10.48550/arXiv.2405.15793

Yang, E., Garcia, T., Williams, H., Kumar, B., Ramé, M., Rivera, E., Ma, Y., Amar, J., Catalani, C., & Jia, Y. (2024). From barriers to tactics: A behavioural science-informed agentic workflow for personalized nutrition coaching. *arXiv:2410.14041*. https://doi.org/10.48550/arXiv.2410.14041

Author information

Gordon Bowen – DBA, Associate Professor, School of Management, Anglia Ruskin University
Address: East Road, CB1 1PT, Cambridge, United Kingdom
E-mail: gordon.bowen@aru.ac.uk
ORCID ID: https://orcid.org/0009-0007-4082-0336
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=56943078600
WoS Researcher ID: https://www.webofscience.com/wos/author/record/65121803
Google Scholar ID: https://scholar.google.com/citations?user=zm_Qgw4AAAAJ

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – June 10, 2025
Date of approval – June 26, 2025
Date of acceptance – September 25, 2025
Date of online placement – September 30, 2025

Научная статья
УДК 34:004:340.1:004.8
EDN: https://elibrary.ru/pltfwo
DOI: https://doi.org/10.21202/jdtl.2025.17

Агентный искусственный интеллект: правовые и этические вызовы автономных систем

Гордон Боуэн

Университет Англия Рёскин, Кембридж, Великобритания

Ключевые слова

автономность,
агентный искусственный интеллект,
искусственный интеллект,
ответственность,
право,
правовое регулирование,
программирование,
риск,
цифровые технологии,
этика

Аннотация

Цель: определить специфические правовые и этические проблемы агентного искусственного интеллекта и выработать рекомендации по созданию защитных механизмов для обеспечения ответственного функционирования автономных ИИ-систем.

Методы: исследование носит концептуальный характер и основано на системном анализе научной литературы по вопросам этики искусственного интеллекта, правового регулирования автономных систем и социального взаимодействия ИИ-агентов. В работе применяются сравнительный анализ различных типов ИИ-систем, исследование потенциальных рисков и преимуществ агентного искусственного интеллекта, а также междисциплинарный подход, интегрирующий достижения в сфере права, этики и компьютерных наук для формирования комплексного понимания проблематики.

Результаты: установлено, что агентный искусственный интеллект, обладая автономностью принятия решений и способностью к социальному взаимодействию, создает качественно новые правовые и этические вызовы по сравнению с традиционными ИИ-ассистентами. Выявлены основные категории потенциального вреда: прямое воздействие на пользователей через открытые и скрытые действия, манипулятивное влияние на поведение и кумулятивный вред от длительного взаимодействия. Определена необходимость распределения ответственности между тремя ключевыми субъектами: пользователем, разработчиком и владельцем системы агентного искусственного интеллекта.

Научная новизна: впервые проведен системный анализ этических аспектов агентного искусственного интеллекта как качественно нового класса автономных систем, отличающихся от традиционных ИИ-ассистентов степенью независимости и социальной интерактивности. Разработана типология потенциальных рисков социального взаимодействия с агентными интеллектуальными системами, и предложена концептуальная модель распределения правовой и этической ответственности в триаде «пользователь – разработчик – владелец».

Практическая значимость: результаты исследования формируют теоретическую основу для разработки этических принципов и правовых норм регулирования агентного искусственного интеллекта в условиях растущего рынка автономных интеллектуальных систем. Полученные выводы могут быть использованы законодателями при создании нормативной базы, разработчиками при проектировании защитных механизмов, а также организациями при внедрении агентных систем искусственного интеллекта в различных сферах экономической деятельности.

Для цитирования

Боуэн, Г. (2025). Агентный искусственный интеллект: правовые и этические вызовы автономных систем. *Journal of Digital Technologies and Law*, *3*(3), 431–445. https://doi.org/10.21202/jdtl.2025.17

Список литературы

Abuelsaad, T., Akkil, D., Dey, P., Jagmohan, A., & Vempaty, A. (2024). Agent-E: From Autonomous Web Navigation to Foundational Design Principles in Agentic Systems. *arXiv preprint arXiv:2407.13032*. https://doi.org/10.48550/arXiv.2407.13032

Acharya, D. B., Kuppan, K., & Ashwin, D. B. (2025). Agentic AI: Autonomous intelligence for complex goals – a comprehensive survey. In *IEEE Access (*vol. 13, pp. 18912–18936). https://doi.org/10.1109/ACCESS.2025.3532853

Alberts, L., Keeling, G., & McCroskery, A. (2024). Should agentic conversational AI change how we think about ethics? Characterising an interactional ethics centred on respect. *arXiv:2401.09082v2*. https://doi.org/10.48550/arXiv.2401.09082

Alberts, L., Lyngs, U., & Van Kleek, M. (2024). Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces that Act Socially. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–25. https://doi.org/10.1145/3653693

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Virtual Event Canada: ACM. https://doi.org/10.1145/3442188.3445922

Beulen, E., Plugge, A., & van Hillegersberg. J. (2022). Formal and relational governance of artificial intelligence outsourcing. *Information System E Business Management*, *20*(4), 719–748. https://doi.org/10.1007/s10257-022-00562-7

Coghlan, S., Waycott, J., Lazar, A., & Neves, B. (2021). Dignity, Autonomy, and Style of Company: Dimensions Older Adults Consider for Robot Companions. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–25. https://doi.org/10.1145/3449178

Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J. R., Ektefaie, Y., Kondic, J., & Zitnik, M. (2024). Empowering biomedical discovery with AI agents. *Cell*, *187*(22), 6125–6151. https://doi.org/10.1016/j.cell.2024.09.022

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*, 113515.

Hasan, S. S., Fury, M. S., Woo, J. J., Kunze, K. N., & Ramkumar, P. N. (2025). Ethical Application of Generative Artificial Intelligence in Medicine. *Arthroscopy: Journal of Arthroscopic Related Surgery*, *41*(4), 874–885. https://doi.org/10.1016/j.arthro.2024.12.011

Hosseini, S., & Seilani, H. (2025). The Role of Agentic AI in Shaping a Smart Future: A Systematic review. *Array*, *26*, 100399. https://doi.org/10.1016/j.array.2025.100399

Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., & Narayanan, A. (2024). AI Agents That Matter. *arXiv:2407.01502v1*.

Kocielnik, R., Langevin, R., George, J. S., Akenaga, S., Wang, A., Jones, D. P., Argyle, A., Fockele, C., Anderson, L., Hsieh, D. T., Kabir, Y., Duber, H., Hsieh, G., & Hartzler, A. L. (2021). Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *3rd Conference on Conversational User Interfaces (CUI '21), July 27–29, 2021, Bilbao (online), Spain*. ACM, New York, NY, USA. https://doi.org/10.1145/3469595.3469599

Korzynski, P., Edwards, A., Gupta, M. C., Mazurek, G., & Wirtz, J. (2025). Humanoid robotics and agentic AI: reframing management theories and future research directions. *European Management Journa*l, *43*(4), 548–560. https://doi.org/10.1016/j.emj.2025.06.002

Kshetri, N. (2025). Transforming cybersecurity with agentic AI to combat emerging cyber threats. *Telecommunications Policy*, *49*(6), 102976. https://doi.org/10.1016/j.telpol.2025.102976

Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). Towards a science of human-AI decision making: A survey of empirical studies. *arXiv:2112.11471*. https://doi.org/10.48550/arXiv.2112.11471

Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M., & Kanai, R. (2022). Meta-learning, social cognition and consciousness in brains and machines. *Neural Network*, *145*, 80–89. https://doi.org/10.1016/j.neunet.2021.10.004

Li, X., Shi, W., Zhang, H., Peng, C., Wu, S., & Tong, W. (2025). The Agentic-AI Core: an AI-Empowered, Mission-Oriented core network for Next-Generation mobile telecommunications. *Engineering*. https://doi.org/10.1016/j.eng.2025.06.027

Lim, S., Schmälzle, R., & Bente, G. (2025). Artificial Social Influence via Human-Embodied AI Agent Interaction in Immersive Virtual Reality (VR): Effects of Similarity-Matching during health conversations. *Computers in Human Behavior Artificial Humans*, *5*, 100172. https://doi.org/10.1016/j.chbah.2025.100172

Martinez, D. R., & Kifle, B. M. (2024). *Artificial Intelligence: A Systems Approach from Architecture Principles to Deployment*. MIT Press eBooks, IEEE Xplore2. https://doi.org/10.7551/mitpress/14806.001.0001

Mohanarangan, S., Karthika, D., Moohambigai, B., & Sangeetha, R. (2024). Unleashing the Power of AI and Machine Learning: Integration Strategies for IoT Systems. *International Journal of Scientific Research in Computer Science and Engineering*, *12*(2), 25–32.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). https://doi.org/10.1145/259963.260288

Russell, S. J., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall. Google-Books-ID: CUVeMwAACAAJ.

Shamsudhin, N., & Jotterand, F. (2021). Social Robots and Dark Patterns: Where Does Persuasion End and Deception Begin? In F. Jotterand, & M. Ienca (Eds.), *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues* (pp. 89–110). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-74188-4_7

Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., Yilla, N., Gallegos, J., Smart, A., Garcia, E., & Virk, G. (2023). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. *arXiv:2210.05791*. https://doi.org/10.48550/arXiv.2210.05791

Small., C., & Lew, C. (2021). Mindfulness, moral reasoning and responsibility: Towards virtue in ethical decision-making. *Journal of Business Ethics*, *169*(1), 103–117. https://doi.org/10.1007/s10551-019-04272-y

Yang, J., Jimenez, C. E., Wettig, A., Lieret, K., Yao, S., Narasimhan, K., & Press, O. (2024). SWE-AGENT: Agent-Computer Interfaces Enable Automated Software Engineering. *arXiv:2405.15793*. https://doi.org/10.48550/arXiv.2405.15793

Yang, E., Garcia, T., Williams, H., Kumar, B., Ramé, M., Rivera, E., Ma, Y., Amar, J., Catalani, C., & Jia, Y. (2024). From barriers to tactics: A behavioural science-informed agentic workflow for personalized nutrition coaching. *arXiv:2410.14041*. https://doi.org/10.48550/arXiv.2410.14041

Сведения об авторе

Боуэн Гордон– доктор делового администрирования, доцент, школа менеджмента, Университет Англия Рёскин
Адрес: Великобритания, г. Кембридж, CB1 1PT, Ист Роуд
E-mail: gordon.bowen@aru.ac.uk
ORCID ID: https://orcid.org/0009-0007-4082-0336
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=56943078600
WoS Researcher ID: https://www.webofscience.com/wos/author/record/65121803
Google Scholar ID: https://scholar.google.com/citations?user=zm_Qgw4AAAAJ

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.07.45 / Право и научно-технический прогресс
Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 10 июня 2025 г.
Дата одобрения после рецензирования – 26 июня 2025 г.
Дата принятия к опубликованию – 25 сентября 2025 г.
Дата онлайн-размещения – 30 сентября 2025 г.