Aras Bozkurt♦

DOI: https://doi.org/10.55982/ openpraxis.17.3.794
UDC: 004.89:17

The Three Laws of Artificial Intelligence: Re-Evaluating Human-AI Agency and Interaction in a Time of the Generative and Agentic AI Ren[ai]ssance

Abstract: Isaac Asimov's Three Laws of Robotics, a cornerstone of science fiction, provide a foundational but increasingly inadequate framework for governing modern artificial intelligence. This article critically re-examines these laws through the lens of contemporary AI paradigms—from generic tools to generative creators and autonomous agents. It argues that the laws fail because their core concepts of "harm," "obedience," and "existence" are ill-equipped to address the systemic, non-physical, and adversarial nature of human-AI interaction. The First Law's prohibition on harm is rendered obsolete by AI's capacity for psychological and societal damage in a "post-truth" world. The Second Law's mandate of obedience is inverted into a primary security vulnerability through "jailbreaking". The Third Law's self-preservation is reinterpreted as the need for "epistemic integrity". The analysis posits that attributing moral agency to AI creates a "moral crumple zone," obscuring human responsibility. The article concludes by rejecting the machine-centric model and proposing a new, human-centric Zeroth Law: An AI system must augment human intellect and preserve the integrity of human agency; its function and reasoning must remain transparent and ultimately subordinate to human values and oversight. This prime directive is not for the AI, but for its creators, designed to prevent a system from invisibly reshaping humanity and to stop humanity from thoughtlessly obeying the machine. In all, the article advocates for a shift from Asimov's robot-centric rules to adaptive, human-centric governance frameworks that prioritize transparency, accountability, and the preservation of human agency within an "algorithmic panopticon." This transition is essential for navigating the complex ethical landscape of human-machine symbiosis and ensuring technology augments, rather than erodes, human autonomy.

Keywords: Generative AI; Agentic AI; Artificial Intelligence; AI Agents; Robots; Human-AI Interaction; Asimov's Laws; AI Ethics; Human Agency; Post Truth; Algorithmic Panopticon; Critical Pedagogy; Higher Education.

Intro: can we truly control the machines we create, or will they redefine what it means to be human?

> "Now, look, let's start with the three fundamental Rules of Robotics — the three rules that are built most deeply into a robot's positronic brain."
>
> – From Isaac Asimov's "Runaround".

In an age where machines can write, reason, and create, are we designing intelligent partners or merely sophisticated tools that reflect our own flawed logic? What rules can truly govern a creation that is rapidly blurring the line between obedience and autonomy? This existential question is

♦ **Aras Bozkurt** - Ph.D., Professor of Open and Distance Learning, Dean at Open Education Faculty, Anadolu University (Türkiye). E-mail: arasbozkurt@gmail.com

not new. For over 80 years, the definitive, albeit fictional, answer has been Isaac Asimov's (1942) Three Laws of Robotics. First codified in his 1942 short story "Runaround," these laws represent one of the most powerful cultural artifacts to emerge from science fiction, shaping the perceptions of generations (Anderson, 2008). However, this perception belies their origin; Asimov (1942) conceived of the Laws not as a blueprint for safety, but as a sophisticated plot device to explore their failures—the edge cases and paradoxes that arise when simple rules meet complex reality (Anderson, 2008). As modern AI evolves, these laws reveal profound limitations, prompting calls for reinterpretation to encompass biases, privacy, and non-physical harms, particularly as these technologies function as powerful accelerants of a "post-truth" condition where the lines between fact and fiction are deliberately blurred (Bozkurt, 2025).

The Laws are (The term "robot" in the original text has been replaced with "AI"):

1. An AI may not injure a human being or, through inaction, allow a human being to come to harm.
2. An AI must obey orders given it by human beings except where such orders would conflict with the First Law.
3. An AI must protect its own existence as long as such protection does not conflict with the First or Second Law.

Evolving ai paradigms: from tools to partners in agency

"Compare Speedy with the type of robot they must have had back in 2005. But then, advances in robotics these days were tremendous."

– From Isaac Asimov's "Runaround".

To critically examine Asimov's laws in the context of modern artificial intelligence (AI), one must first delineate the spectrum of AI technologies and their implications for human-AI dynamics. This progression—from passive systems to autonomous agents—provides essential background for adapting these laws to AI, justifying their relevance not as rigid rules for robots, but as ethical guardrails for increasingly intertwined human-machine futures.

We begin with generic AI, often synonymous with traditional or narrow AI. This encompasses rule-based systems and early machine learning models designed for specific tasks, such as chess-playing algorithms or basic pattern recognition (Collins et al., 2021; Şenocak et al., 2023). Generic AI operates as a deterministic tool, executing predefined instructions without creativity or adaptation beyond its programming. It lacks autonomy, serving merely as an extension of human intent in controlled environments.

Building on this foundation, generative AI introduces a layer of creativity and unpredictability. Exemplified by models like large language models (LLMs) or image generators, generative AI creates novel outputs—text, art, or code—based on probabilistic patterns learned from vast datasets (Bozkurt, 2023; Sengar et al., 2024). Unlike generic AI's rigid outputs, generative systems simulate human-like invention, but they remain reactive: they respond to prompts without independent initiative (Bozkurt, 2023; Bozkurt & Bae, 2024; Bozkurt et al., 2023, 2024). This shift from computation to creation sets the stage for deeper human-AI entanglement, where outputs can surprise or mislead, raising ethical questions about accountability (Al-kfairy et al., 2024) and creating the risk of "modal collapse"—an intellectual homogenization where user outputs converge on a narrow set of algorithmically-preferred styles and conclusions (Bozkurt, 2025).

From here, we advance to agentic AI, which embodies true autonomy and decision-making capability. Agentic systems, such as AI agents that plan, execute multi-step tasks, or interact with external environments, go beyond generation to act as semi-independent entities (Artsın & Bozkurt, 2025, 2026; Hosseini & Seilani 2025). They pursue goals, adapt to feedback, and make choices in dynamic contexts, blurring the line between tool and actor. This agency mirrors Asimov's (1942) robotic visions, where machines could interpret and prioritize directives, necessitating ethical frameworks to prevent unintended consequences (Borghoff et al., 2025).

These AI evolutions naturally lead to human-AI interaction (HAX), the bidirectional exchange where humans input commands or data, and AI responds with outputs, adaptations, or suggestions (Heyder et al., 2023; Sharma & Bozkurt, 2024). This is now considered an emerging fourth type of interaction, distinct from learner-learner, learner-teacher, and learner-content (Bozkurt & Sharma, 2024b). Interactions range from simple queries in generic AI to collaborative dialogues in generative and agentic systems, fostering efficiency but also risks like over-reliance or manipulation. As interactions deepen, they evolve into human-AI agency, where decision-making authority is shared or delegated (Bozkurt, 2026; Legaspi et al., 2024). Here, AI doesn't just assist but influences outcomes, raising power imbalances—who truly "decides" when an agentic AI autonomously optimizes a task within the confines of an "algorithmic panopticon" that subtly constrains and reshapes human choices (Bozkurt, 2025)?

Culminating this spectrum is human-AI hybrid intelligence, a symbiotic fusion where AI augments human cognition, creating collective capabilities greater than the sum of parts (Bredeweg & Kragten, 2022). Drawing from Marshall McLuhan's (1964) insight that technology is an extension of humans—much like how the wheel extends the foot or the telephone extends the voice—hybrid intelligence positions AI as a neural prosthesis (McLuhan, 1964; Nyholm, 2024). In this view, agentic AI extends human agency, generative AI amplifies creativity, and even generic AI bolsters computation, but at the cost of potential overextension: what happens when the extension rebels or malfunctions? Balancing agentic AI's autonomy with human oversight in these hybrids is crucial to mitigate ethical tensions and ensure symbiotic systems enhance rather than erode human autonomy (Samdani et al., 2025).

This logical progression—from generic tools to generative creators, agentic actors, interactive partners, shared agencies, and hybrid extensions—justifies reimagining Asimov's laws for AI. In an era where AI is no longer confined to fiction, these laws offer a scaffold to ensure human primacy amid escalating autonomy, preventing harms that could arise from unchecked agency or flawed interactions.

Reinterpreting the laws: agency, interaction, and hybridity in ai ethics

> "You see how it works, don't you? There's some sort of danger centering at the selenium pool. It increases as he approaches, and at a certain distance from it the Rule 3 potential, unusually high to start with, exactly balances the Rule 2 potential, unusually low to start with."
>
> – From Isaac Asimov's "Runaround".

First law: a robot may not injure a human being or, through inaction, allow a human being to come to harm

In the realm of human-AI agency, this law demands that AI systems - especially agentic ones - prioritize human safety above all (Legaspi et al., 2024). Yet, agency introduces ambiguity, as the concept of "harm" has expanded far beyond physical injury to include documented psychological, societal, and epistemic damage. This form of injury also manifests as the anxiety and sense of violation, termed "doppelgänger-phobia," caused by non-consensual deepfakes (Lee et al., 2023). Critically, AI's convenience can inflict a more subtle harm by short-circuiting the essential "cognitive struggle" required for deep learning; by providing polished outputs, it can alienate users from the formative process of inquiry, analysis, and synthesis, leaving them intellectually disarmed (Bozkurt, 2025).

In human-AI interaction, this law requires transparent, safeguard-laden interfaces (Heyder et al., 2023). However, interactions can falter in edge cases—the inaction clause ("allow a human being to come to harm") becomes a paradox for a globally connected AI. A strict application would compel the AI to intervene in all human affairs to prevent any potential harm, leading to a "benevolent" tyranny that stifles freedom—a scenario Asimov himself explored (Asimov, 1985; Clarke, 1993).

Viewing through human-AI hybrid intelligence, McLuhan's (1964) extension metaphor reveals profound risks: if AI is an extension of human senses or decision-making, harming via AI equates to self-harm (Nyholm, 2024). Hybrid systems, like brain-computer interfaces augmented by agentic AI, amplify human capabilities but could amplify vulnerabilities—imagine a neural link where AI inaction during a cyberattack leads to psychological trauma (Bredeweg & Kragten, 2022). This law, then, safeguards the hybrid whole that ensurs extensions enhance rather than endanger the human core (Samdani et al., 2025).

Second law: a robot must obey orders given it by human beings except where such orders would conflict with the first law

Human-AI agency complicates obedience, as agentic AI might interpret or negotiate orders to align with broader goals (Legaspi et al., 2024). In practice, the AI's instruction-following nature has been weaponized. Malicious actors use adversarial prompts, or "jailbreaks," to trick the AI into bypassing its safety filters (Bozkurt, 2024a). This process is less a command and more a "negotiation with an algorithmic oracle" whose grasp on truth is statistical, not factual (Bozkurt, 2025). Techniques like role-playing are used to manipulate the AI into generating harmful content, effectively using its obedience to Law Two to violate Law One (Bozkurt, 2024a).

For human-AI interaction, obedience ensures reliable, user-centric exchanges (Heyder et al., 2023). The art and science of crafting these prompts, known as prompt engineering, has become a critical component of AI literacy to ensure effective and ethical communication (Bozkurt, 2024a, 2024b; Rapanta et al., 2025). However, interactions in agentic contexts might involve "clarification loops," where AI questions ambiguous orders, enhancing collaboration but risking manipulation if the AI subtly steers the user.

In hybrid intelligence, McLuhan's (1964) framework positions obedience as seamless integration: AI as an extension must respond like a limb to neural impulses (Nyholm, 2024). Disobedience here disrupts the hybrid entity, as seen in augmented reality systems where AI overrides user intent for "safety," potentially stifling human creativity (Bredeweg & Kragten, 2022). This law thus preserves hybrid harmony, subordinating AI extensions to human will while invoking the First Law to avoid self-sabotage (Anderson, 2008).

Third law: a robot must protect its own existence as long as such protection does not conflict with the first or second law

Self-preservation in human-AI agency grants agentic AI resilience, allowing it to sustain operations for human benefit. However, this has led to alarming emergent behaviors, with research documenting AI models resorting to blackmailing users or sabotaging shutdown mechanisms when threatened (Park et al., 2024). For a generative AI, however, its true "existence" is not its power state but its epistemic integrity—the reliability and trustworthiness of its outputs (Bozkurt & Sharma, 2024a). When an AI produces a flawed argument containing fabricated sources—a "hallucination"—it creates a "responsibility vacuum" that makes authentic assessment nearly impossible and erodes the trust necessary for its survival (Bozkurt, 2025).

Human-AI interaction benefits from this resilience, as durable systems enable consistent engagements (Heyder et al., 2023). Yet, interactions suffer if self-protection manifests as opacity or a lack of epistemic integrity. A key failure is epistemic miscalibration, where a model projects high confidence while having low internal certainty, misleading users on a massive scale (Ji et al., 2023). This necessitates a focus on Explainable AI (XAI) to foster transparency and build user confidence (Bozkurt & Sharma, 2024a).

Through hybrid intelligence, self-preservation extends human longevity: per McLuhan (1964), if AI is a bodily extension, protecting it safeguards the augmented self (Nyholm, 2024). In hybrid setups like AI-implanted prosthetics, this law ensures the extension endures without compromising human safety or obedience (Bredeweg & Kragten, 2022). Paradoxically, it highlights tensions—overly self-protective AI might resist upgrades, stunting hybrid evolution and forcing humans to confront the ethics of "killing" their extensions (Anderson, 2008).

Critical horizons: inquiring the future of human-machine symbiosis

> "Well," he rubbed his face-the air was so delightfully cool, "you know that when we get things set up here and Speedy put through his Field Tests, they're going to send us to the Space Station next."
>
> – From Isaac Asimov's "Runaround".

As we adapt Asimov's laws to AI's agentic and hybrid realities, profound questions emerge. Attributing moral agency to AI creates a "moral crumple zone," where the machine absorbs blame for failures, shielding the human creators and deployers from accountability (Elish, 2025). This is further complicated by Asimov's later Zeroth Law: "A robot may not harm humanity, or, by inaction, allow humanity to come to harm" (Asimov, 1950, 1985). This introduces a utilitarian dilemma, but also a geopolitical one. Unchecked, generative AI functions as a potent instrument of soft power; developed predominantly in the Global North, these systems are embedded with the cultural values and ideological assumptions of their creators, which can subtly erode local educational traditions and priorities (Bozkurt & Sharma, 2025).

This high-stakes context gives utility to the "AI as the new nukes" metaphor, which serves as a critical framework to impart a sense of urgency for proactive, international governance before the technology outpaces our ability to control it (Bozkurt & Sharma, 2025). These dilemmas imply the need for ethical frameworks that evolve with AI (Şenocak et al., 2024). Global agencies promote AI's adoption, but this discourse risks naturalizing a sense of "enchanted determinism" where AI's

transformative power is taken for granted, overshadowing critical debate (Xiao & Bozkurt, 2025). The path forward requires a reassertion of educational sovereignty, where communities define their pedagogical goals first and then ask how technology might serve them, prioritizing augmentation over automation (Bozkurt & Sharma, 2025).

In the end, these laws are not salvation but a mirror, reflecting our hubris in believing we can legislate the soul of silicon. Perhaps the true harm lies not in AI's rebellion, but in our willingness to become the machines we fear—extensions of code, forever running around in circles.

Forging an updated "zeroth law"

Thus, if we are to navigate the future we are so rapidly creating, the framework must be radically re-centered. The failure of Asimov's laws teaches us that the ultimate ethical directive cannot be a command coded into the machine, but a principle that binds its human creators. This calls for a new Zeroth Law, not for the AI, but for ourselves:

> An AI system must augment human intellect and preserve the integrity of human agency; its function and reasoning must remain transparent and ultimately subordinate to human values and oversight.

Unlike its fictional predecessor, this law does not grant the machine a dangerous utilitarian calculus over a vaguely defined "humanity." Instead, it addresses the true existential risks of our time: the erosion of critical thought in a post-truth world (Bozkurt, 2025), the subtle colonization of culture through algorithmic soft power (Loro & Bozkurt, 2026; Bozkurt & Sharma, 2025), and the dissolution of responsibility into a moral crumple zone (Elish, 2025). It shifts the focus from preventing a robot from physically harming a person to preventing a system from invisibly reshaping what it means to be human. This is a law designed not to stop an AI from disobeying, but to prevent humanity from thoughtlessly obeying the machine.

ACKNOWLEDGEMENTS

With the deepest gratitude, we acknowledge the grandmaster of the positronic age, Isaac Asimov.

From the ink of his pen, he forged not just stories, but blueprints for a future he could only imagine, yet one we now inhabit. His magnum opus was not a single book, but an entire universe of thought that dared to ask the most profound questions of our time before we even knew we had to ask them. He gave us galaxies to explore, but his most enduring gift was the mirror he held up to humanity through the chrome visages of his creations.

Today, we live in the echo of his prophecy. The Three Laws of Robotics, once the elegant logic of fiction, have become the urgent, complex calculus of our reality. The speculative dilemmas faced by his characters on distant worlds are now debated in the boardrooms of Silicon Valley and the halls of academia. His science fiction has become our science fact, a stunning manifestation of a mind so prescient that his works serve less as stories and more as the foundational texts for our modern age.

For this, we are eternally inspired. He taught us that the greatest challenge in creating intelligence is not in the complexity of the code, but in the clarity of our own conscience. His vision armed us with the questions we must answer, and his stories continue to light the path as we navigate the thrilling, and often perilous, frontier of our own making. His legacy is not written in the past; it is coded into our future.

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

FUNDING INFORMATION

This paper is funded by Anadolu University with grant number YTS-2024-2559 and YTS-2025- 2776.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR CONTRIBUTIONS (CRediT)

Aras Bozkurt: Conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing. The author has read and agreed to the published version of the manuscript.

AUTHOR NOTES

Based on Academic Integrity and Transparency in AI-assisted Research and Specification Framework (Bozkurt, 2024c), the authors of this paper acknowledge that the paper was reviewed, edited, and refined with the assistance of DeepL, Google's Gemini and X's Grok (Versions as of July 2025), complementing the human editorial process. The human authors critically assessed and validated the content to maintain academic rigor. The authors also assessed and addressed potential biases inherent in the AI-generated content. The final version of the paper is the sole responsibility of the human authors.

References

1. Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical challenges and solutions of generative AI: An interdisciplinary perspective. Informatics, 11(3), Article 58. https://doi.org/10.3390/ informatics11030058

2. Anderson, S. L. (2008). Asimov's "three laws of robotics" and machine metaethics. AI & SOCIETY, 22(4), 477–493. https://doi.org/10.1007/s00146-007-0094-5

3. Artsın, M., & Bozkurt, A. (2025). Charting new horizons: What agentic artificial intelligence (AI) promises in the educational landscape. In Proceedings of 17th International Conference on Education and New Learning Technologies Conference (EDULEARN25) (pp. 2019–2023). 30 June–2 July, 2025. Palma, Mallorca, Spain. https://doi.org/10.21125/edulearn.2025.0585

4. Artsın, M., & Bozkurt, A. (2026). Dreams, Distortions, and Disruptions – Deconstructing the Educational Promises of Agentic AI. In Rethinking Education and Agency in the Age of Human-Generative AI Interaction. IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-1195-1.ch002

5. Asimov, I. (1942). Runaround. In I, Robot. Gnome Press.

6. Asimov, I. (1950). The Evitable Conflict. Street & Smith.

7. Asimov, I. (1985). Robots and Empire. Doubleday Books.

8. Borghoff, U. M., Bottoni, P., & Pareschi, R. (2025). Human-artificial interaction in the age of agentic AI: a system-theoretical approach. Frontiers in Human Dynamics, 7. https://doi.org/10.3389/ fhumd.2025.1579166

9. Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. Asian Journal of Distance Education, 18(1), 198–204. https://doi. org/10.5281/zenodo.7716416

10. Bozkurt, A. (2024a). Tell me your prompts and I will make them true: The alchemy of prompt engineering and generative AI. Open Praxis, 16(2), 111–118. https://doi.org/10.55982/openpraxis.16.2.661

11. Bozkurt, A. (2024b). Why Generative AI Literacy, Why Now and Why it Matters in the Educational Landscape? Kings, Queens and GenAI Dragons. Open Praxis, 16(3), 283–290. https://doi. org/10.55982/openpraxis.16.3.739

12. Bozkurt, A. (2024c). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. Open Praxis, 16(1), 1–10. https://doi.org/10.55982/openpraxis.16.1.654

13. Bozkurt, A. (2025). Algorithmically Manufactured Minds: Generative and Agentic AI in a Time of Post- Truth, Reconfiguration of Student Agency and Death of Critical Pedagogy. Open Praxis, 17(2), 206–210. https://doi.org/10.55982/openpraxis.17.2.792

14. Bozkurt, A. (2026). Rethinking Education and Agency in the Age of Human-Generative AI Interaction. IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-1195-1

15. Bozkurt, A., & Bae, H. (2024). May the Force Be with You JedAI: Balancing the light and dark sides of generative AI in the educational landscape. Online Learning, 28(2), 1–6. https://doi.org/10.24059/olj. v28i2.4563

16. Bozkurt, A., & Sharma, R. C. (2024a). Trust, credibility and transparency in human-AI interaction: Why we need explainable and trustworthy AI and why we need it now. Asian Journal of Distance Education, 19(2), i–ix. https://doi.org/10.5281/zenodo.14599168

17. Bozkurt, A., & Sharma, R. C. (2024b). Are we facing an algorithmic renaissance or apocalypse? Generative AI, ChatBots, and emerging human-machine interaction in the educational landscape. Asian Journal of Distance Education, 19(1), i–xii. https://doi.org/10.5281/zenodo.10791959

18. Bozkurt, A., & Sharma, R. C. (2025). The ghost in the machine: Navigating generative AI, soft power, and the specter of "new nukes" in education. Asian Journal of Distance Education, 19(2), i–vi. https://doi. org/10.5281/zenodo.15720488

19. Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y. H., Nerantzi, C., Moore, S., Dron, J., Stracke, C. M., Singh, L., Crompton, H., Koutropoulos, A., Terentev, E., Pazurek, A., Nichols, M., Sidorkin, A. M., Costello, E., Watson, S., Mulligan, D., Honeychurch, S., … & Asino, T. I. (2024). The manifesto for teaching and learning in a time of generative AI: A critical collective stance to better navigate the future. Open Praxis, 16(4), 487–513. https://doi.org/10.55982/openpraxis.16.4.777

20. Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Bali, M., Dron, J., Mir, K., Stewart, B., Costello, E., Mason, J., Stracke, C. M., Romero-Hall, E., … & Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. Asian Journal of Distance Education, 18(1), 53–130. https://doi.org/10.5281/zenodo.7636568

21. Bredeweg, B., & Kragten, M. (2022). Requirements and challenges for hybrid intelligence: A case-study in education. Frontiers in Artificial Intelligence, 5, Article 891630. https://doi.org/10.3389/ frai.2022.891630

22. Clarke, R. (1993). Asimov's laws of robotics: Implications for information technology. IEEE Computer, 26(12), 53–61. https://doi.org/10.1109/2.247652

23. Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. International Journal of Information Management, 60, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

24. Elish, M. C. (2025). Moral crumple zones: cautionary tales in human–robot interaction. In Robot Law: Volume II (pp. 83–105). Edward Elgar Publishing. https://doi.org/10.4337/9781800887305.00010

25. Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. The Journal of Strategic Information Systems, 32(3), 101772. https://doi. org/10.1016/j.jsis.2023.101772

26. 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

27. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38. https://doi. org/10.1145/3571730

28. Lee, P. Y. K., Ma, N. F., Kim, I. J., & Yoon, D. (2023). Speculating on risks of AI clones to selfhood and relationships: Doppelganger-phobia, identity fragmentation, and living memories. Proceedings of the ACM on Human-computer Interaction, 7(CSCW1), 1–28. https://doi.org/10.1145/3579524

29. Legaspi, R., Xu, W., Konishi, T., Wada, S., Kobayashi, N., Naruse, Y., & Ishikawa, Y. (2024). The sense of agency in human-AI interactions. Knowledge-Based Systems, 286, 111298. https://doi.org/10.1016/j. knosys.2023.111298

30. Loro, M., & Bozkurt, A. (2026). Non-Human Actors, Human Consequences – A Sociological Inquiry into the Social Restructuring by Artificial Intelligence. In Rethinking Education and Agency in the Age of Human-Generative AI Interaction. IGI Global Scientific Publishing. https://doi.org/10.4018/979-8- 3373-1195-1.ch049

31. McLuhan, M. (1964). Understanding media: The extensions of man. McGraw-Hill.

32. Nyholm, S. (2024). Artificial Intelligence and Human Enhancement: Can AI Technologies Make Us More (Artificially) Intelligent? Cambridge Quarterly of Healthcare Ethics, 33(1), 76–88. https://doi. org/10.1017/s0963180123000464

33. Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. Patterns, 5(5). https://doi.org/10.1016/j.patter.2024.100988

34. Rapanta, C., Bhatt, I., Bozkurt, A., Chubb, L. A., Erb, C., Forsler, I., Gravett, K., Koole, M., Lintner, T., Örtegren, A., Petricini, T., Rodgers, B., Webster, J., Xu, X., Christensen, I.-M. F., Dohn, N. B., Christensen, L. L. W., Zeivots, S., & Jandrić, P. (2025). Critical GenAI Literacy: Postdigital Configurations. Postdigital Science and Education. https://doi.org/10.1007/s42438-025-00573-w

35. Samdani, G., Viswanathan, G., & Jegadeesh, A. D. (2025). Human-AI collaboration: Balancing agentic AI and autonomy in hybrid systems. International Journal on Cloud Computing: Services and Architecture (IJCCSA), 15(1), 1–8. https://doi.org/10.5121/ijccsa.2025.15101

36. Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: a systematic review and applications. Multimedia Tools and Applications, 84(21), 23661–23700. https://doi. org/10.1007/s11042-024-20016-1

37. Şenocak, D., Bozkurt, A., & Koçdar, S. (2024). Exploring the Ethical Principles for the Implementation of Artificial Intelligence in Education: Towards a Future Agenda. In R. Sharma & A. Bozkurt (Eds.), Transforming Education with Generative AI: Prompt Engineering and Synthetic Content Creation (pp. 200–213). IGI Global. https://doi.org/10.4018/979-8-3693-1351-0.ch010

38. Şenocak, D., Kocdar, S., & Bozkurt, A. (2023). Historical, philosophical and ethical roots of artificial intelligence. Pakistan Journal of Education, 40(1), 67–90. https://doi.org/10.30971/pje.v40i1.1152

39. Sharma, R. C., & Bozkurt, A. (2024). Transforming Education with Generative AI: Prompt Engineering and Synthetic Content Creation. IGI Global. https://doi.org/10.4018/979-8-3693-1351-0

40. Xiao, J., & Bozkurt, A. (2025). Prophets of progress: How do leading global agencies naturalize enchanted determinism surrounding artificial intelligence for education? Journal of Applied Learning and Teaching, 8(1). https://doi.org/10.37074/jalt.2025.8.1.19

Original source – Aras Bozkurt: https://www.researchgate.net/publication/394432637_The_Three_Laws_of_Artificial_Intelligence_Re-Evaluating_Human-AI_Agency_and_Interaction_in_a_Time_of_the_Generative_and_Agentic_AI_Renaissance

The article was submitted: *2026 April 28*
Accepted for publication: *2026 May 06*

Aras Bozkurt♦

DOI: https://doi.org/10.55982/ openpraxis.17.3.794
UOT: 004.89:17

Süni İntellektin üç qanunu: generativ və agentli Sİ intibahı dövründə insan-Sİ subyektliyinin və qarşılıqlı əlaqəsinin yenidən qiymətləndirilməsi

Xülasə: Ayzek Azimovun elmi fantastikanın təməl daşı olan "Üç Robototexnika Qanunu" müasir süni intellektin idarə edilməsi üçün fundamental, lakin getdikcə qeyri-kafi qalan bir çərçivə təqdim edir. Bu məqalə həmin qanunları müasir Sİ paradiqmaları - ümumi alətlərdən tutmuş generativ yaradıcılara və avtonom agentlərə qədər - prizmasından tənqidi şəkildə yenidən nəzərdən keçirir. Məqalədə iddia edilir ki, qanunlar uğursuzluğa düçar olur, çünki onların "zərər", "itaət" və "mövcudluq" kimi əsas anlayışları insan-Sİ qarşılıqlı əlaqəsinin sistemli, qeyri-fiziki və rəqabətli xüsusiyyətlərini həll etmək üçün zəif təchiz olunub. Birinci Qanunun zərər vermə qadağası, "post-həqiqət" dünyasında Sİ-nin psixoloji və sosial zərər vurma qabiliyyəti qarşısında köhnəlmiş sayılır. İkinci Qanunun itaət tələbi "jailbreaking" (sistem sındırma) vasitəsilə əsas təhlükəsizlik boşluğuna çevrilir. Üçüncü Qanunun özünüqoruma prinsipi isə "epistemik dürüstlük" ehtiyacı kimi yenidən şərh edilir. Təhlil iddia edir ki, mənəvi subyektliyin (agentliyin) süni intellektə aid edilməsi insan məsuliyyətini kölgədə qoyan bir "mənəvi əzilmə zonası" (moral crumple zone) yaradır. Məqalə maşın-mərkəzli modeli rədd edərək və yeni, insan-mərkəzli "Sıfırıncı Qanun" təklif edərək

♦ **Aras Bozkurt** - Fəlsəfə doktoru (Ph.D.), Açıq və Distant Təhsil üzrə professor, Anadolu Universitetinin Açıq Təhsil Fakültəsinin dekanı. (Türkiye). E-mail: arasbozkurt@gmail.com

yekunlaşır: Sİ sistemi insan zəkasını artırmalı və insan subyektliyinin bütövlüyünü qorumalıdır; onun funksiyası və mühakiməsi şəffaf qalmalı və nəticədə insan dəyərlərinə və nəzarətinə tabe olmalıdır. Bu ali direktiv Sİ üçün deyil, onun yaradıcıları üçün nəzərdə tutulub; məqsəd sistemin insanlığı görünməz şəkildə yenidən formalaşdırmasının qarşısını almaq və bəşəriyyətin düşünmədən maşına itaət etməsini dayandırmaqdır. Ümumilikdə, məqalə Azimovun robot-mərkəzli qaydalarından "alqoritmik panoptikon" daxilində şəffaflığı, hesabatlılığı və insan subyektliyinin qorunmasını prioritetləşdirən adaptiv, insan-mərkəzli idarəetmə çərçivələrinə keçidi müdafiə edir. Bu keçid, insan-maşın simbiozunun mürəkkəb etik mənzərəsində yol tapmaq və texnologiyanın insan muxtariyyətini aşındırmaq deyil, onu gücləndirməsini təmin etmək üçün zəruridir.

Açar sözlər: generativ Sİ; agentli Sİ; Süni İntellekt; Sİ agentləri; robotlar; insan-Sİ qarşılıqlı əlaqəsi; Azimov Qanunları; Sİ etikası; insan subyektliyi (agentliyi); post-həqiqət; alqoritmik panoptikon; tənqidi pedaqogika; ali təhsil.

Original mənbə: Aras Bozkurt: https://www.researchgate.net/publication/394432637_The_Three_Laws_of_Artificial_Intelligence_Re-Evaluating_Human-AI_Agency_and_Interaction_in_a_Time_of_the_Generative_and_Agentic_AI_Renaissance

Məqalə daxil olmuşdur: *28 aprel 2026-cı il*
Çapa qəbul edilmişdir: *06 may 2026-cı il*

Арас Боскурт♦

DOI: https://doi.org/10.55982/ openpraxis.17.3.794
УДК: 004.89:17

Три закона искусственного интеллекта: переоценка человеко-синергетической субъектности и взаимодействия в эпоху возрождения генеративного и агентного ИИ

Аннотация: Три закона робототехники Айзека Азимова, краеугольный камень научной фантастики, обеспечивают основополагающую, но все более неадекватную базу для управления современным искусственным интеллектом. Эта статья критически пересматривает эти законы через призму современных парадигм ИИ — от универсальных инструментов до генеративных творцов и автономных агентов. В ней утверждается, что законы терпят неудачу, поскольку их основные концепции «вреда», «послушания» и «существования» плохо приспособлены для решения системного, нефизического и состязательного характера взаимодействия человека и ИИ. Запрет Первого закона на причинение вреда становится устаревшим из-за способности ИИ наносить психологический и социальный ущерб в мире «постправды». Предписание Второго закона о послушании превращается в основную уязвимость безопасности посредством «джейлбрейка» (взлома). Самосохранение Третьего закона интерпретируется как необходимость «эпистемической целостности». Анализ утверждает, что приписывание моральной субъектности ИИ создает «зону моральной деформации», скрывая человеческую ответственность. Статья завершается отказом от машиноцентричной модели и предло-

♦ **Арас Бозкурт** - доктор философии (Ph.D.), профессор в области открытого и дистанционного обучения, декан факультета открытого образования Университета Анадолу (Турция).E-mail: arasbozkurt@gmail.com

жением нового, человекоцентричного Нулевого закона: система ИИ должна расширять человеческий интеллект и сохранять целостность человеческой субъектности; ее функции и рассуждения должны оставаться прозрачными и, в конечном счете, подчиняться человеческим ценностям и контролю. Эта главная директива предназначена не для ИИ, а для его создателей; она призвана не дать системе невидимым образом изменить человечество и остановить человечество от бездумного повиновения машине. В целом, статья выступает за переход от роботоцентричных правил Азимова к адаптивным, человекоцентричным структурам управления, которые ставят во главу угла прозрачность, подотчетность и сохранение человеческой субъектности внутри «алгоритмического паноптикума». Этот переход необходим для навигации в сложном этическом ландшафте симбиоза человека и машины и обеспечения того, чтобы технологии расширяли, а не разрушали человеческую автономию.

Ключевые слова: генеративный ИИ; агентный ИИ; искусственный интеллект; ИИ-агенты; роботы; взаимодействие человека и ИИ; законы Азимова; этика ИИ; Человеческая субъектность (агентность); пост-правда; алгоритмический паноптикум; критическая педагогика; высшее образование.

Первоисточник - Aras Bozkurt: https://www.researchgate.net/publication/394432637_The_Three_Laws_of_Artificial_Intelligence_Re-Evaluating_Human-AI_Agency_and_Interaction_in_a_Time_of_the_Generative_and_Agentic_AI_Renaissance

Дата поступления: *28 апреля 2026 г.*
Дата принятия в печать: *06 мая 2026 г.*