UDC: 004.89:005.334 https://doi.org/10.15407/jai2025.02.010

M. Rokosh[1], M. Striletskyi[2]
[1]Ternopil Ivan Puluj National Technical University, Ukraine
56 Rus'ka Str., Ternopil, 46001
[2]West Ukrainian National University, Ukraine
11 Lvivska Str., Ternopil, 46009
[1]mike@apiko.com
[2]mykola@apiko.com
[1]https://orcid.org/0009-0009-0323-3735
[2]https://orcid.org/0009-0009-6167-7289

MODELING THE EVOLUTION OF RISK IN AI SYSTEMS THROUGHOUT THEIR LIFECYCLE USING THE S-CURVE

Abstract. Artificial Intelligence is becoming increasingly embedded in various areas of human life, offering new capabilities that go beyond traditional software systems. Unlike conventional programs that follow fixed instructions, AI can generate its own solutions after processing large volumes of data. However, human input remains essential in designing AI architecture and setting its goals. While AI improves efficiency and decision-making across fields, it also introduces new types of risks. These risks often arise not from malicious intent, but from unpredictable system behavior and user errors. This paper analyzes such risks using a systems perspective and logistic S-curve modeling to examine the AI lifecycle. The analysis shows that the first three stages—development, scaling, and stabilization—carry the highest levels of vulnerability. Key issues include design flaws, insufficient debugging, and lack of continuous monitoring. More advanced systems may evolve through multiple S-curve phases, each introducing new challenges. The study emphasizes the need for stronger legal and ethical standards, drawing on regulatory efforts from the EU, USA, UK, Germany, and France. International cooperation is also highlighted as a key factor in ensuring that AI develops safely and responsibly.

Keywords: artificial intelligence, Al-system, risk, vulnerability, system lifecycle, S-curve.

1. Introduction

The era of Artificial Intelligence (AI) is not only upon us—it is accelerating at an extraordinary pace. What was once considered a futuristic concept is now deeply embedded in everyday life, from personalized recommendations and virtual assistants to predictive analytics in healthcare and autonomous systems in logistics. The development of AI technologies has become so rapid and intensive that even researchers, developers, and end-users often struggle to adapt to its evolving capabilities. This pace of change introduces significant challenges: hesitation, lack of preparedness, and underestimation of AI's transformative potential can lead to serious setbacks, not only for individuals and organizations but for entire sectors of society.

AI cannot be paused, stopped, or meaningfully slowed without incurring opportunity costs that outweigh potential benefits. Such efforts may also encourage fragmentation, as global actors move at different speeds, leading to uneven development and regulation. Rather than resisting the momentum of AI, the focus must be on understanding and guiding it. We must overcome uncertainty with urgency—by embracing AI, studying it critically, building its capabilities responsibly, and mastering its real-world application. This means fostering interdisciplinary collaboration, investing in education and governance, and creating technical safeguards that evolve alongside innovation.

At its core, AI offers a singular advantage: time. With the ability to analyze and synthesize vast datasets at speeds far beyond human capacity, AI can dramatically improve decision-making, optimize processes, and uncover insights that would otherwise remain hidden. However, this power comes with a growing responsibility to ensure that such systems are safe, fair, and aligned with human values. As AI continues to reshape industries and institutions, understanding its full lifecycle—including where risks emerge and how they can be mitigated—becomes not just a technical challenge but a societal imperative. If used wisely, AI can serve as a force multiplier for progress, but if

mismanaged, it can become a source of systemic vulnerability.

2. Problem Statement

As AI becomes increasingly integrated into the fabric of everyday life, it introduces a new class of risks that are distinct from traditional cybersecurity threats. These risks often stem not from intentional harm, but from non-malicious sources such as spontaneous and unpredictable software behaviors, opaque outputs generated by complex neural networks, and inadvertent human errors in system design, deployment, or operation. Unlike conventional software, AI systems can evolve through training data and feedback mechanisms, occasionally producing outcomes that are difficult to anticipate or explain. Such unpredictability, combined with the high-stakes environments in which AI is often deployed—such as healthcare, finance, critical infrastructure, and public administration—can result in system malfunctions, flawed decision-making, and unintended consequences. These vulnerabilities pose a significant challenge to existing risk management practices, which are often ill-equipped to account for the dynamic and emergent nature of AI systems.

3. Related Work

The rapid expansion of Artificial Intelligence has prompted extensive scholarly attention across multiple disciplines, particularly concerning the risks and challenges associated with its development, deployment, and societal integration. Existing literature can be broadly categorized into technical, ethical, and socio-political domains, each contributing to a deeper understanding of the systemic vulnerabilities introduced by AI.

On the socio-political front, Crawford [1] offers a critical perspective on the global implications of AI, framing it not merely as a technical achievement but as an infrastructure embedded with power dynamics and environmental costs. She argues that AI development often reinforces existing inequalities, both within and between nations, while simultaneously consuming substantial ecological resources. This view shifts the focus away from narrow discussions of algorithmic design to the broader material and political conditions under which AI is produced and applied.

From a technical standpoint, foundational texts such as Russell and Norvig [2] have provided a detailed examination of AI algorithms, system architectures, and computational logic. Their work has served as a cornerstone for understanding the capabilities and limitations of intelligent systems. They highlight that while AI can perform tasks once thought exclusive to human intelligence—such as natural language processing, strategic planning, and perception—it remains constrained by issues such as data dependency, brittleness, and lack of interpretability. These technical limitations can become significant risk vectors, particularly when AI systems are deployed in high-stakes environments like healthcare, finance, or law enforcement.

Ethical concerns have received growing attention as AI systems become more autonomous and widely deployed. Shahriar et al. focus on algorithmic bias and data governance, identifying how skewed datasets and opaque decision-making processes can lead to systemic discrimination and loss of public trust. Their analysis emphasizes the need for transparent model evaluation, fairness-aware design practices, and robust legal oversight. Similarly, Luxton [3] explores the implications of AI in psychological and clinical practice, noting that while AI tools offer enhanced diagnostic and predictive capabilities, they also raise significant questions around consent, accountability, and the doctor-patient relationship.

Other scholars have examined risk through the lens of system dynamics and software behavior. Yudkowsky [4], for example, discusses the existential risks posed by misaligned general intelligence, warning that even well-intentioned systems can behave in ways that are unpredictable and harmful when their objectives are not fully aligned with human values. While his perspective is often framed in long-term speculative terms, it has sparked important debates on safety mechanisms, goal specification, and the limits of human control in advanced AI systems.

Further research has emphasized the technical risks of operational deployment. Hutson [5] reports on the reproducibility crisis in AI research, citing that a large number of published models lack sufficient documentation and access to data, which hampers validation and real-world applicability. This lack of transparency not only limits scientific progress but also increases the risk of deploying untested or misunderstood systems in sensitive domains.

In the field of system modeling, Bejan and Lorente [6] introduce the constructal law and the logistic S-curve as a way to understand growth dynamics in complex systems. Their work provides a useful conceptual basis for analyzing the development trajectories of AI systems, particularly in understanding how risk exposure may vary across different stages of growth and saturation.

Taken together, the existing body of research illustrates that AI-related risks are not isolated to any one phase of development or type of application. Instead, they emerge from the interplay of technical limitations, design choices, data quality, regulatory gaps, and social context. However, while these studies offer valuable insights, most tend to address risks as static categories or focus on individual incidents rather than tracing their progression across the AI system lifecycle. This highlights a gap in the literature that this paper aims to address: a structured, lifecycle-based approach to risk analysis that integrates technical, ethical, and systemic factors using a model such as the logistic S-curve. By situating risks within specific developmental stages—such as initial design, accelerated deployment, stabilization, and decline—this research seeks to offer a dynamic and context-aware framework that can better inform future governance, design practices, and policy development.

4. Research Objectives

This paper aims to define and categorize risks inherent in the AI development lifecycle, apply system modeling—particularly the logistic S-curve model—to the AI lifecycle to identify high-risk phases, and recommend strategic, legislative, and ethical measures to reduce or prevent risks. In achieving this aim, the research intends to demonstrate that existing paradigms for AI risk management are inadequate unless contextualized within the evolutionary trajectory of AI systems, acknowledging that risks are not uniformly distributed but rather fluctuate depending on development stage, system architecture, and implementation context.

5. AI Capabilities and Complexities

Artificial Intelligence refers to software systems designed to perform cognitive tasks traditionally associated with human intelligence. These include learning from data, recognizing patterns, and adapting to dynamic conditions. The adaptability and complexity of these systems derive primarily from algorithmic models—especially neural networks—which simulate the functioning of the human brain. Importantly, AI can evolve in non-linear ways. Neural networks may generate new software elements either spontaneously or through incomprehensible internal processes. When such behavior is coupled with human errors in operation or oversight (excluding intentional sabotage), AI becomes a substantial risk vector. In medicine, AI supports both traditional and telemedicine by monitoring physiological parameters and modeling individual biological states [7]. This enables a shift from reactive to preventive healthcare, transforming health from a biographical to a biological construct [9]. In science and engineering, AI facilitates data analysis, synthesis, and iterative refinement, particularly in experimental physics, where recursive analytical loops accelerate hypothesis testing and discovery. In the social sphere, from education and culture to tourism and sports, AI is revolutionizing productivity and access, democratizing opportunities while introducing surveillance and bias-related challenges [9]. In the economy, AI integration reshapes manufacturing, agriculture, and SME operations, demanding comprehensive oversight to balance innovation with inclusivity, and economic sustainability with ethical responsibility.

6. Methodology and Theoretical Framework

The methodological foundation of this research integrates a systems-theoretic perspective, comparative analysis, risk categorization, and graphical modeling to develop a comprehensive understanding of the AI lifecycle and the risks embedded at each of its distinct stages. The analytical core is formed around the logistic S-curve—originally developed by Pierre Verhulst—which offers a valuable mathematical model to describe the growth, stabilization, and eventual decline of complex dynamic systems such as Artificial Intelligence architectures. The S-curve not only allows for visualization of AI's development over time, but also provides a diagnostic tool for identifying when and where specific risk categories are most likely to emerge.

In the context of this paper, the S-curve has been subdivided into four sequential yet overlapping stages: Initial Development, Accelerated Growth, Stabilization, and Decline.

Each stage on the S-curve introduces a unique configuration of vulnerabilities, requiring tailored strategies for mitigation. These stages align respectively with four fundamental types of risk: Foundational, Operational, Systemic, and Residual. These categories encapsulate the evolving nature of AI-related threats and form a coherent framework for targeted risk governance.

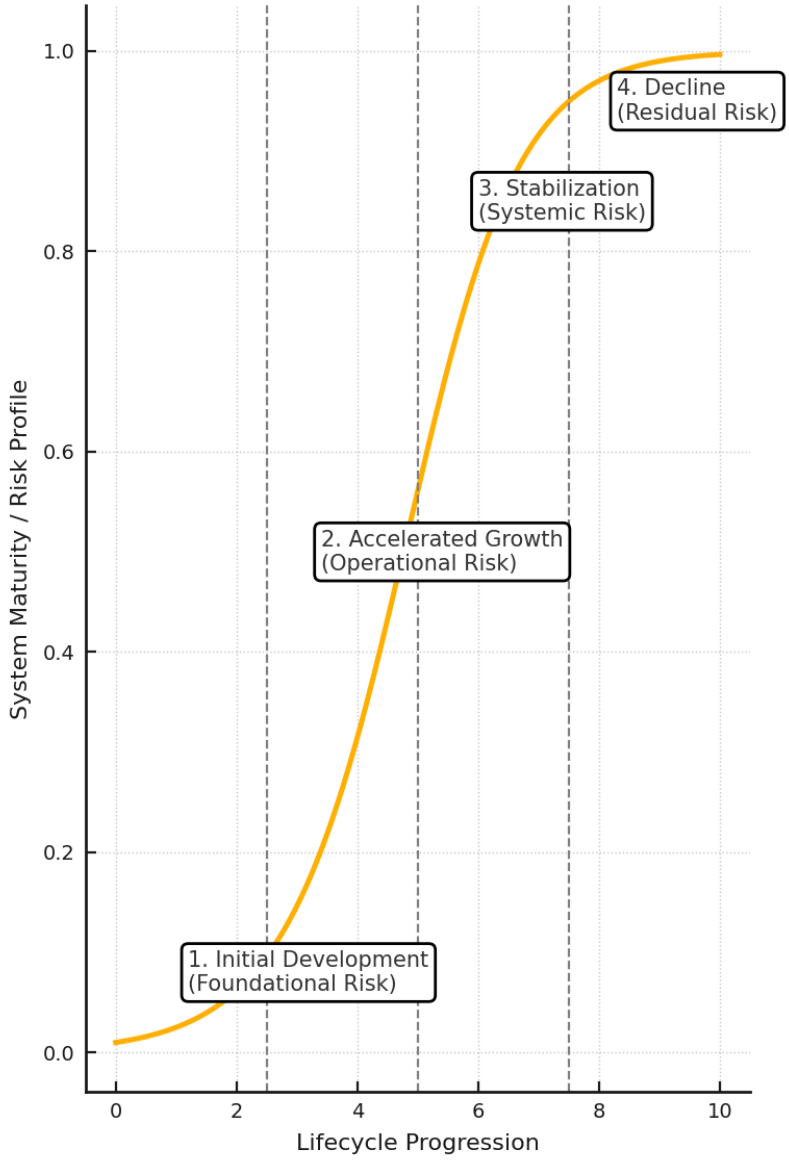

Fig. 1. AI System Lifecycle and Risk Evolution Illustrated by the S-Curve

Foundational Risk characterizes the first stage of the AI lifecycle: the phase of Initial Development. This stage, positioned at the base of the S-curve, is deceptively calm. However, its relative developmental inertia masks an elevated level of latent vulnerability. Foundational risk arises primarily from errors in system architecture, inadequate data quality, biased algorithmic logic, or poorly formulated objectives. Misjudgments at this stage can embed long-lasting flaws that cascade through every subsequent stage of development. A key contributor to foundational risk is the lack of interdisciplinary collaboration at the design level—when engineers, domain experts, ethicists, and legal scholars fail to converge on the intent, scope, and governance of AI functionalities. Data-related issues, such as selection bias, underrepresentation of minority groups, or insufficient volume, also originate at this stage, compromising fairness and reliability. The decisions made in the design and programming of AI systems—such as model type, training regimes, and data input parameters—determine the latent potential for risk exposure. These missteps do not usually reveal their full consequences until the AI is scaled or deployed, making proactive quality assurance and ethical foresight essential.

Operational Risk dominates the second phase: Accelerated Growth. Here, the AI system transitions from concept to real-world implementation, often undergoing rapid iteration, scaling, and exposure to heterogeneous data environments. The steep middle of the S-curve symbolizes exponential development, but this expansion is fraught with implementation challenges. Operational risk encompasses the volatility and unpredictability introduced by rapid deployment: system drift, misaligned performance metrics, or incompatibility with existing infrastructure [10-11]. This stage is often marked by aggressive timelines, market-driven urgency, and inadequate testing. The opacity of neural networks, particularly in black-box models, exacerbates this issue—making it difficult to trace the root cause of malfunctions. Furthermore, operator error becomes a significant factor in this phase. Inadequate training, overreliance on automation, and insufficient human oversight lead to misapplication of AI outputs. The paradox of automation bias emerges here: the more accurate a system appears, the more likely humans are to defer to it uncritically, even when errors occur. Additionally, this phase may involve initial public or client interaction with the AI system, exposing it to legal liabilities and reputational risks. Operational risk is, therefore, both a technical and organizational concern and calls for integrated monitoring systems, stress testing, and scenario planning [12].

Systemic Risk arises in the third phase, corresponding to the Stabilization segment of the S-curve. At this point, the AI system is mature, widely adopted, and deeply embedded in organizational workflows or societal infrastructures. Paradoxically, it is in this phase of relative calm that the most concealed and complex threats emerge. Systemic risk stems from dependencies and interconnections—AI systems influencing or depending on other critical systems such as healthcare, transportation, finance, and government administration. Hidden interdependencies create pathways for cascading failures, whereby an unnoticed error in one AI module may propagate through entire ecosystems. Additionally, adversarial inputs, emergent behavior, and feedback loops introduce new vectors of unpredictability. The stabilization phase is also where governance often becomes complacent, assuming the maturity of the system equates to safety. However, maturity brings scale, and scale multiplies consequences. Algorithms operating at population-level data may inadvertently codify structural inequalities, while long-term use may lead to unanticipated forms of bias, performance decay, or dataset obsolescence. Importantly, this is the stage where AI begins to interact symbiotically with human behavior, shaping decisions, preferences, and institutional policies—blurring the line between automation and authority. Systemic risk is therefore multidimensional and

requires continuous auditing, cross-sectoral policy alignment, and anticipatory regulation.

Residual Risk defines the fourth and final phase: Decline. This stage on the S-curve signifies the waning relevance or functionality of an AI system. However, risk does not disappear at the end of the lifecycle—it transforms. Residual risk emerges from legacy systems that are no longer maintained but still in use, outdated datasets that persist in training environments, or embedded algorithms that continue to influence decisions in invisible ways. In some instances, residual risk is the result of abandonment, where organizations sunset AI tools without thoroughly decommissioning their influence or removing them from operational processes. These abandoned tools may still operate with outdated logic, producing decisions that no longer reflect current ethical standards, legal frameworks, or technical realities. Another form of residual risk involves intellectual inertia—where institutions continue to rely on historically trained AI systems due to sunk costs or regulatory lag, despite knowing better models exist. Residual risks are often overlooked, yet they can be deeply corrosive, undermining trust, safety, and accountability over time. Addressing these requires robust offboarding protocols, long-term data stewardship, and periodic system retirements supported by regulatory incentives.

Taken together, these four categories of risk provide a dynamic and holistic lens for understanding AI system vulnerabilities. The S-curve serves not merely as a visual metaphor but as a structured temporal map that allows researchers, engineers, policymakers, and stakeholders to predict and address risks as they emerge and evolve. It bridges the gap between abstract ethical concerns and practical implementation timelines. This methodological approach grounds the broader research in actionable insights that can inform future standards, interventions, and innovations. In doing so, it advocates for a lifecycle-aware, ethically aligned, and risk-resilient model of AI development.

7. Ethical, Legal, and Strategic Responses

Given the outlined vulnerabilities, proactive legislative and ethical strategies are vital. The study recommends codifying ethical standards and AI codes of conduct, as adopted by global leaders such as the USA, UK, Germany, and France. Risk assessments should be systematically integrated into every AI deployment. Establishing robust national and international regulatory frameworks is critical. However, excessive restrictions that may hinder innovation must be avoided. To ensure balanced development, multidisciplinary legislative collaboration is needed, involving ethicists, technologists, and policymakers. National regulations must be aligned with global standards. A supranational regulatory body for AI—analogous to nuclear non-proliferation agreements—remains elusive but essential. Furthermore, ethical considerations should not remain aspirational; they must be operationalized through compliance protocols, enforceable standards, and algorithmic transparency audits. Without such grounding, ethical declarations risk becoming ceremonial rather than consequential. The ethical implications of AI extend beyond data privacy and discrimination, encompassing issues such as digital personhood, algorithmic autonomy, and long-term societal transformation. Each of these requires deliberate attention, not as peripheral matters but as core components of AI design [13].

8. Geopolitical and Socioeconomic Considerations

Unequal access to AI technologies may widen existing knowledge and development gaps. Nations with robust economies and innovative capacities will continue to lead, increasing global inequality. In response, emerging states must invest in domestic AI development, foster public-private innovation partnerships, and advocate for open knowledge exchange at international forums. The potential for knowledge asymmetries raises concerns not only about economic inequality but also about sovereignty, autonomy, and long-term dependency. If a select group of nations or corporations

monopolize access to transformative AI technologies, others may find themselves locked into subordinate roles within the global knowledge economy. Thus, AI becomes not only a technological but a geopolitical issue, requiring diplomatic coordination, equitable access strategies, and mechanisms to avoid monopolistic dominance that could stifle innovation and exacerbate systemic inequalities across borders.

Conclusions

The swift progression of AI is both a beacon of progress and a source of unprecedented risk. The challenges addressed in this study stem not from malevolence, but from structural and procedural vulnerabilities that manifest across the AI lifecycle. By employing an S-curve model, this research highlights the stages most susceptible to risk, especially the first three phases—design, accelerated growth, and stabilization. To ensure AI's safe integration into global society, legislative foresight, international cooperation, ethical clarity, and technical rigor must converge. The future of AI must be shaped not only by its potential for innovation but also by our collective responsibility to manage its dangers. The stakes are high, and while the benefits are extraordinary, so too are the responsibilities. This is the moment not to fear AI, nor to romanticize it, but to understand it deeply, regulate it wisely, and deploy it judiciously. A well-regulated AI future does not stifle progress; rather, it ensures that the progress achieved serves the broader goals of human well-being, social justice, and planetary sustainability.

References

1. Crawford K. (2021) Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press. 336 p.

2. Russell S.J., Norvig P. (2003) Artificial Intelligence: A Modern Approach. 2nd ed. Prentice Hall.

3. Luxton D.D. (2014) Artificial intelligence in psychological practice: Current and future applications and implications. Professional Psychology: Research and Practice. V. 45, I. 5, P. 332–339. https://doi.org/10.1037/a0034559

4. Yudkowsky E. (2006) Artificial Intelligence as a Positive and Negative Factor in Global Risk. Draft manuscript. [Online]. Available: https://intelligence.org/files/AIPosNegFactor.pdf

5. Hutson M. (2018) Missing data hinder replication of artificial intelligence studies. Science, February 15. https://doi.org/10.1126/science.aat3298

6. Bejan A., Lorente S. (2011) The constructal law origin of the logistics S curve. Journal of Applied Physics. V. 110, 024901. https://doi.org/10.1063/1.3606555

7. Reed T.R., Reed N.E., Fritzson P. (2004) Heart sound analysis for symptom detection and computer-aided diagnosis. Simulation Modelling Practice and Theory. V. 12, I. 2, P. 129–146. https://doi.org/10.1016/j.simpat.2003.11.005

8. The Medical Futurist. (2016) Artificial Intelligence Will Redesign Healthcare. [Online]. Available: https://medicalfuturist.com/artificial-intelligence-will-redesign-healthcare

9. Yorita A., Kubota N. (2011) Cognitive Development in Partner Robots for Information Support to Elderly People. IEEE Transactions on Autonomous Mental Development. V. 3, I. 1, P. 64–73. https://doi.org/10.1109/TAMD.2011.2105868

10. Fawcett T. (2004) ROC Graphs: Notes and Practical Considerations for Researchers. Kluwer Academic Publishers.

11. Davis J., Goadrich M. (2006) The Relationship Between Precision-Recall and ROC Curves. Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA. P. 233–240. https://doi.org/10.1145/1143844.1143874

12. Liang D., Tsai C.-F., Wu H.-T. (2015) The Effect of Feature Selection on Financial Distress Prediction. Knowledge-Based Systems. V. 73, P. 289–297. https://doi.org/10.1016/j.knosys.2014.10.011

13. Luger G., Stubblefield W. (2004) Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 5th ed. The Benjamin/Cummings Publishing Company. 720 p.

The article has been sent to the editors 19.05.25.
After processing 30.05.25.
Submitted for printing 30.06.25.