Artificial Intelligence: Definition and Prospects for Use in the Field of Humanities Research

Ivan Skiba, Andrey Kolesnikov
Institute of Philosophy of the National Academy of Sciences of Belarus
Minsk, Belarus
gonzodzen@mail.ru

***Abstract*—This article explores definitions of artificial intelligence (AI) and its prospects in humanities research. It defines an object-oriented approach to understanding AI and analyzes key components of its architecture: data and algorithms. The article examines the potential development of AI technologies, including artificial general intelligence (AGI) and artificial superintelligence (ASI). The limitations of modern AI models are discussed, particularly their inability to replace humanities researchers, while their usefulness as a tool for supporting scientific activity is emphasized.**

***Keywords*—artificial intelligence, machine learning, large language model, humanities research, AGI, ASI.**

I. Introduction

Since artificial intelligence (AI) no longer requires a separate introduction, we will replace this section with a definition. In this article, we will not interpret the concept of AI and related terms from the perspective of a timeline of scientific and technological progress or by pinpointing a specific moment on that timeline. Instead, we will focus on the specific implementation of AI technologies, perceiving them as distinct objects. This object-oriented approach allows us to examine AI systems as self-contained entities with defined properties and behaviors, rather than as stages in an evolutionary process. In this context, we define the key components of AI and evaluate its prospects for application in humanities research. Our methodology emphasizes practical implementation over historical development, recognizing that contemporary AI systems represent complex integrations of hardware, software, and data architectures that warrant analysis as complete technological artifacts.

II. Definition and Architecture of Artificial Intelligence

AI is a technology that enables the emulation of certain external parameters of human activity. A specific AI technology is traditionally referred to as a model (e.g., «large language model» (LLM), «visual-language model» (VLM), etc.) [1], [2]. These models represent sophisticated computational frameworks designed to process and generate human-like outputs across various modalities. The term "model" reflects their function as abstract representations of cognitive processes, implemented through complex mathematical architectures.

From an object-oriented perspective, AI is always implemented through a combination of hardware and software. Its architecture can be described through two fundamental components that interact in a dynamic feedback loop:

- Data processed by AI, which serves as both the input and training material for the system's knowledge base;
- Algorithms that facilitate data processing, comprising the computational rules and learning mechanisms that transform inputs into meaningful outputs.

The purpose of AI development is to automate data processing at scale, and its machine learning (ML) algorithms regulate responses to input queries through probabilistic inference and pattern recognition. This automation extends beyond simple rule-based systems to incorporate adaptive learning capabilities that improve with exposure to additional data.

Data is represented in digital format as sequences of 0s and 1s since the vast majority of AI's hardware operates using digital technologies and follows the Princeton architecture. Consequently, any information that can be reasonably converted into a binary sequence qualifies as data for AI. This includes not only traditional text and numerical data but also multimedia content, sensory inputs, and even abstract concepts when properly encoded. Notably, algorithms themselves are also sequences of 0s and 1s and can therefore be classified as data, creating a recursive relationship where algorithms process data that may itself contain other algorithms.

Traditionally, AI algorithms are defined as machine learning algorithms, which significantly differ from classical computational algorithms [3]. The key distinction is that machine learning algorithms, due to

their built-in autonomy, can self-regulate the nature of their responses based on the data they receive. This regulation occurs through machine learning itself, meaning AI technology is capable of software-based self-organization that adjusts its internal parameters without explicit programming. A classical algorithm, by its inherent nature, always produces the same result for the same input data. However, intelligent models do not. This discrepancy is not solely due to the probabilistic nature of machine learning algorithms but is fundamentally linked to the presence of learning itself: given identical input data, a model at one stage of training may yield one result, while at another stage, it may produce a completely different one as its internal representations evolve through exposure to new information.

In a way, the difference between classical algorithms and machine learning algorithms boils down to the following: classical algorithms represent human problem-solving through an algorithm, whereas machine learning algorithms represent model-based problem-solving with human assistance. Of course, human «assistance» here refers to the initial configuration of the model's parameters, training data selection, and ongoing supervision, creating a collaborative dynamic between human designers and artificial systems. This distinction highlights the paradigm shift from deterministic programming to statistical learning that characterizes modern AI systems.

III. Limitations and Risks of Artificial Intelligence

It is precisely the presence of self-organization—even if it is adjustable—that underpins concerns and negative forecasts regarding AI. This is because predicting the exact «configuration» of a specific AI technology in advance is nearly impossible due to the high complexity and extreme multifactorial nature of large models. The emergent behaviors that arise from these complex systems often defy straightforward analysis, creating challenges for verification and validation. As models grow in size and capability, their decision-making processes become increasingly opaque, even to their creators, resulting in what is often termed the "black box" problem of AI systems.

Of course, developers do not ignore the risks: every AI model undergoes rigorous testing and is subject to imposed restrictions before being released. However, eliminating the possibility of a model «going out of control» and acting independently is virtually impossible. The very features that make AI systems valuable—their adaptability and capacity for unexpected solutions—also make them inherently unpredictable to some degree. Ensuring absolute safety would severely limit a model's functionality, rendering it uncompetitive in the AI technology market [4]. This creates a fundamental tension between capability and control that permeates AI development. Metaphorically speaking, making a model completely safe is akin to sealing off a sacred spring with concrete—greater functionality inevitably comes with greater unpredictability, and vice versa. The challenge lies in finding the optimal balance where systems remain both powerful enough to be useful and constrained enough to be trustworthy.

Additional risks emerge from the potential for AI systems to amplify existing biases present in their training data, make errors with high confidence, or be manipulated through adversarial attacks. These vulnerabilities stem from the statistical nature of machine learning, where models optimize for patterns rather than truth or fairness. Furthermore, the rapid deployment of AI systems across critical domains raises concerns about accountability, as traditional mechanisms for assigning responsibility become complicated when decisions are made by algorithms that even their developers may not fully understand.

IV. Prospects for the Development of Artificial Intelligence

The prospects for the development of AI technologies themselves are general artificial intelligence (AGI—artificial general intelligence) and the so-called artificial superintelligence (ASI—artificial superintelligence) [5]. Since opinions and definitions of what they should be suffer from extreme pluralism, we will provide the most abstracted forms that capture the essential characteristics while acknowledging the ongoing debate in the field.

AGI is a technology capable of emulating the external parameters of human activity in any field, at least as well as, and possibly even slightly better than, a human who is a high-level expert in that field. In other words, AGI, presumably, should understand mathematics at least as well as the best mathematicians; linguistics, as well as the best linguists; programming, as well as the best programmers, and so on across all domains of human knowledge and skill. This would require not just specialized competence but the flexible integration of abilities across disciplines—a hallmark of human cognition that current AI systems lack. In this sense, AGI should represent a kind of collective image of all the specialists who are significant for science and the economy, capable of transferring knowledge between domains and adapting to novel situations with human-like versatility. At the moment, leading companies in the field of AI development are competing to achieve AGI, though there is significant disagreement about how

close current technologies are to this goal or whether fundamentally new approaches will be required.

ASI represents a much more ephemeral phenomenon, as it is significantly less clearly understood by the developers themselves. However, in any case, when talking about ASI, it refers to a technology that significantly surpasses the most outstanding human abilities, both quantitatively and possibly even qualitatively [**?**]. This concept pushes beyond the boundaries of human cognition to imagine intelligences that might develop entirely new forms of reasoning, perception, or understanding inaccessible to biological minds. Regarding the fundamental possibility of creating ASI, intense philosophical and semi-philosophical discussions are ongoing, with positions ranging from confident predictions of its inevitability to arguments that such systems are fundamentally impossible or inherently unstable. The timeline for potential ASI development remains highly speculative, with estimates ranging from decades to centuries or never.

At the moment, AI has capabilities that unquestionably surpass human abilities only in areas where two conditions are met:

1) The availability of all the necessary information to solve the problem within a well-defined formal system;
2) The unambiguity of result validation through objective, computable metrics.

This, for example, includes intellectual games with complete information, such as chess, go, and others where the rules are fixed and all game states are observable. In these constrained domains, AI systems can explore possibilities far beyond human capacity through brute-force computation and advanced heuristics. In other areas, such as the creation of textual or visual content, it can be confidently stated that AI is capable of generating content that is often impossible to unambiguously identify by its origin, meaning that it is either not possible to clearly distinguish whether it was created by a human or by AI, or it requires in-depth analysis [7]. This blurring of boundaries raises important questions about authenticity, creativity, and intellectual property in the digital age.

It is worth noting, however, that creating high-quality content using AI requires quite advanced prompt engineering skills, that is, the ability to create a textual «action plan» with a large number of details and conditions. Otherwise, AI will simply generate «something on the topic» that may lack depth, coherence, or originality. This requirement for skilled human guidance highlights the current limitations of AI systems and their dependence on human expertise for optimal performance. The most effective applications of AI often involve tight human-AI collaboration, where each contributes their respective strengths to the creative or analytical process.

V. The Use of Artificial Intelligence in Humanities Research

Since AI, in any case, represents an attribute of the «new reality», it is reasonable to use it for constructive purposes that enhance rather than replace human capabilities. In the field of humanities research, it so happens that the final result of representatives of the field coincides with that of the most widespread AI technology at the moment—LLM, that is, a large language model [8]. This convergence creates both opportunities for synergy and challenges to traditional scholarly practices. In this sense, any representative of the industry can be represented as a set of finite sequences of words, which are arranged into sentences, paragraphs, articles, monographs, etc., meaning that a humanities scholar is essentially a collection of textual data. Statistically, any researcher is «summarized» in the results of their work: thoughts, feelings, consciousness, soul, and other metaphysical presences do not have potential in this case and will not be taken into account when assessing the level of a scholar's activity. This reductionist view, while useful for certain analytical purposes, risks overlooking the contextual, interpretive, and experiential dimensions that often distinguish profound humanities scholarship.

However, in this same context, an LLM is also a collection of textual data. That is, a more or less accurate comparative understanding of a scholar as a data generator and AI as a data generator is possible when considering their outputs in purely formal terms. Both process information and produce textual representations, though through radically different mechanisms—one biological and experiential, the other computational and statistical. This superficial similarity masks profound differences in understanding, intentionality, and the capacity for genuine insight that continue to distinguish humanistic inquiry from artificial text generation.

However, it should be noted right away that in the field of AI development, only three key aspects can be scaled (improved, advanced):

1) Model training, which is summarized by the number of model parameters and the algorithms it uses, including architectural innovations and optimization techniques;
2) Hardware computing power, currently represented by graphics cards and specialized AI accelerators, along with supporting infrastructure;
3) Model reasoning, which it performs when generating a response to a query according to its al-

gorithms, including improvements in attention mechanisms and knowledge integration.

The first aspect has already been significantly scaled to some extent, with modern models containing hundreds of billions of parameters, and it is not entirely clear whether its further multiple expansion would be reasonable given diminishing returns and increasing costs. The second aspect, while it can be improved, is still subject to Moore's Law under current technologies—that is, doubling computing power does not necessarily lead to a twofold improvement in computations due to various bottlenecks in memory, bandwidth, and parallel processing limitations.

Because of this, as well as due to the high cost and debatable feasibility of aspect 2, the latest flagship models have primarily been scaled in aspect 3—reasoning in the process of generating responses. These are the so-called reasoning models that attempt to emulate more sophisticated cognitive processes beyond simple pattern recognition. As for improvements in the algorithmic component, they are being pursued continuously, but breakthrough results in this field are rare. How much further models can be improved remains an open question that depends on both theoretical advances and practical engineering constraints. However, the very fact that scaling in aspect 3 has already begun, while AGI has still not been created, is starting to cause some concern—both among developers and those invested in technological progress—about whether current approaches will ever achieve truly general intelligence or if they are reaching the limits of their potential.

Thus, at present, AI cannot serve as a relevant alternative even to an average humanities researcher, if only for the reason that a properly structured prompt (query or command) must, in any case, be formulated by the researcher themselves [9]. This requirement for precise, knowledgeable input reflects the fundamental limitations of current AI systems as tools rather than autonomous thinkers. If one attempts to delegate this task to AI, there would still be a need for someone to generate a prompt for it, and so on in an infinite regress. Additionally, it is necessary to have someone who will verify the content generated by the model, assessing its validity, relevance, and scholarly rigor—tasks that require human judgment and disciplinary expertise. That is, in some fields and when necessary, slight optimization is possible—for example, at Google, 25

But when it comes to generating high-level text—such as a groundbreaking scientific research result—the prompt for AI would have to be so large that it might actually be easier for a human to write the text themselves, and it would likely be more concise. The reason for this is that, at present, AI performs well when working with data similar to what it was trained on. The ability to extrapolate skills is inversely proportional to the similarity between the «semantic form» of the training dataset and the «semantic form» of the test dataset. In other words, if the meaning that a person wants to «explain» to AI is qualitatively new, the AI simply «won't understand» it without additional fine-tuning. In this sense, one could hypothesize a correlation between the prompt and the result, according to which the more specialized and high-level the AI-generated content needs to be, the larger the prompt must be. Ultimately, it all comes down to the fact that, for now, someone still has to write that prompt with sufficient expertise to guide the AI effectively—a requirement that preserves the central role of human scholars in the research process.

The above allows us to propose yet another hypothesis. It is well known that any LLM can «consume» a limited number of tokens (words, word parts, symbols, etc.) [10]. The data it consumes represents the context based on which the model generates a response. In other words, any query to an LLM constitutes the formation of context that bounds and directs its output. Based on this, we propose the hypothesis that in the humanities, there exist certain results—that is, texts—for which, at present, there are no models with a sufficiently large context size to fully comprehend and process them. That is, if the hypothesis is correct, there are humanities research results that require feeding the model more tokens than it is capable of consuming in a single context window—complex arguments, extensive evidentiary bases, or nuanced theoretical frameworks that exceed current technical limitations. From this, we can conclude that an equivalent replacement of a «humanities researcher with AI» is currently impossible—if our hypothesis holds true—because the most sophisticated humanistic work operates at scales and complexities beyond what current AI systems can handle in their entirety.

Therefore, for the needs of the humanities, it is advisable to use AI in the following cases where it can augment rather than replace human capabilities:

- For compiling a selection of literary sources on a particular problem area. This is somewhat similar to the result of a long search in a typical browser, but faster and more relevant due to the AI's ability to understand semantic relationships between works;
- For forming a summary of texts with an emphasis on a specific narrow aspect of the problem. For example: «The disclosure of the theme of existential doom in War and Peace by L. N. Tolstoy» where the AI can quickly identify and

synthesize relevant passages;

- For conceptual verification of the originality of conclusions and reasoning. As an example of a prompt for an interdisciplinary socio-biometric study: «Has the idea of a correlation between the number of marriages in a person's life and the square root of the factorial of their nose length ever been expressed in the scientific community before?» allowing researchers to check for prior art efficiently;
- For providing feedback on the results of conducted research before its publication. An example of a prompt: «Is it true that my research on the influence of the dynamic component of wage growth in the USSR during 1950-1970 on the number of bottles in the briefcase of the hero of V. Yerofeev's poem 'Moscow to the End of the Line' is worthy of the Nobel Prize in Literature for 2025?» helping scholars gauge potential receptions of their work;
- For validating the logical and methodological foundation of the research. An example of an implication check: «Is it true that if 2+2=5, then 3+3=6?» where the AI can quickly identify formal logical flaws;
- For selecting ways to improve both the research itself and its results by suggesting alternative approaches, complementary methodologies, or overlooked sources based on its training data.

Thus, in the humanities, the use of AI can be presented as the «assistance of a senior colleague» who has read widely but may lack deep insight—valuable for certain tasks but insufficient as a replacement for original scholarship. And this is indeed significant when properly understood as a tool rather than an authority. The most productive applications involve using AI to handle time-consuming mechanical aspects of research while reserving the interpretive and synthetic work for human scholars.

Furthermore, AI can be used to evaluate the results of scientific research activities, for example, to categorize them according to the following scales where quantitative assessment is feasible:

- Innovativeness, measured by comparison to existing literature;
- Fundamental nature, assessed through structural analysis of arguments;
- Scientific significance, judged by citation patterns and topic modeling;
- Practical applicability, evaluated through real-world impact metrics;
- Development and detail of the issues, analyzed through textual complexity measures.

The advantage of AI here lies in its speed and relative impartiality when dealing with large datasets. For instance, to categorize works by the innovativeness scale, a person would need several days of careful reading and comparison, while AI can accomplish the same task in just a few minutes by processing the entire corpus simultaneously. However, these automated assessments should be understood as preliminary indicators rather than definitive judgments, always requiring human oversight and contextual understanding to interpret properly. The true value emerges when AI's scalability complements human discernment, creating a collaborative research ecosystem that leverages the strengths of both.

VI. Conclusion

In summary, the complete replacement of humanities researchers with AI technologies is not foreseeable in the near future due to fundamental limitations in current systems' capacities for genuine understanding, creativity, and contextual judgment. However, the use of AI as a tool for enhancing research is highly relevant and beneficial when properly integrated into scholarly workflows. While the future development of AGI and ASI may broaden AI's applications, its current effectiveness is constrained by the necessity for complex prompt engineering and the limited context window available for analyzing intricate humanities problems. The most promising path forward involves viewing AI as a collaborative partner in the research process—one that can handle certain mechanical aspects of scholarship with unprecedented speed and scale, while human researchers focus on the interpretive, synthetic, and creative dimensions that remain beyond artificial systems' reach. This complementary relationship, rather than replacement, represents the most productive framework for integrating AI into humanities research while preserving the field's essential humanistic values and modes of inquiry.

References

[1] What is a Large Language Model (LLM)? / AWS. URL: https://aws.amazon.com/ru/what-is/large-language-model/. (accessed: 18.02.2025).

[2] VLM / SecurityLab. URL: https://www.securitylab.ru/glossary/vlm/ (accessed: 18.02.2025).

[3] Machine Learning Algorithms / ItGlobal. URL:https://itglobal.com/ru-ru/company/glossary/algoritmy-mashinnogo-obucheniya/ (accessed: 18.02.2025).

[4] How to Properly Create Prompts for Neural Networks and What They Actually Are / SMMPlanner. URL: https://smmplanner.com/blog/kak-pravilno-sostavliat-promty-dlia-nieirosietiei-i-chto-eto-voobshchie-takoie/. (accessed: 18.02.2025).

[5] Artificial Intelligence in the Humanities: A Survey of Methods and Applications / SpringerLink. URL: https://link.springer.com/chapter/10.1007/978-3-030-16862-1_2 (accessed: 20.03.2025).

[6] The Role of Artificial Intelligence in Academic Research / ResearchGate. URL: https://www.researchgate.net/publication/328078753_The_Role_of_Artificial_Intelligence_in_Academic_Research (accessed: 20.03.2025).

[7] AI in Humanities Research: Opportunities and Challenges / Cambridge University Press. URL: https://www.cambridge.org/core/journals/first-language-research/article/ai-in-humanities-research-opportunities-and-challenges/ (accessed: 20.03.2025).
[8] The Impact of AI on Academic Publishing / Elsevier. URL: https://www.elsevier.com/en-xm/solutions/ai-in-academic-publishing (accessed: 20.03.2025).
[9] AI and the Future of Humanities: The Need for Interdisciplinary Collaboration / MIT Press. URL: https://mitpress.mit.edu/books/ai-and-future-humanities (accessed: 20.03.2025).
[10] Artificial Intelligence: Implications for Humanities Scholars / Oxford University Press. URL: https://global.oup.com/academic/product/ai-implications-humanities-9780190842356 (accessed: 20.03.2025).

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ОПРЕДЕЛЕНИЕ И ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ В СФЕРЕ ГУМАНИТАРНЫХ ИССЛЕДОВАНИЙ

Скиба И. Р., Колесников А. В.

Статья предлагает системный анализ искусственного интеллекта (ИИ) как технологического феномена с позиций объектно-ориентированной парадигмы. В отличие от традиционных историко-технологических подходов, авторы рассматривают ИИ как совокупность дискретных артефактов, обладающих устойчивыми архитектурными характеристиками. Ключевыми компонентами такой архитектуры выступают: 1) данные в цифровом представлении, включающие как экзогенную информацию, так и эндогенные алгоритмические структуры; 2) машинные алгоритмы обучения, принципиально отличающиеся от классических детерминированных алгоритмов способностью к параметрической самоорганизации. Особое внимание уделяется анализу эмерджентных свойств современных ИИ-систем, проявляющихся в непредсказуемости выводов при идентичных входных данных. Это свойство, обусловленное стохастической природой машинного обучения, создаёт фундаментальные ограничения для применения ИИ в экспертно-ориентированных областях. Авторы детально исследуют феномен "чёрного ящика"нейросетевых архитектур, подчёркивая принципиальную несводимость процессов принятия решений в глубоких нейронных сетях к интерпретируемым логическим схемам. В контексте развития технологий общего искусственного интеллекта (AGI) обсуждаются современные исследовательские тренды, включая мультимодальное обучение, нейросимволическую интеграцию и метаобучение. При этом отмечается, что современные системы типа GPT-4 и Gemini демонстрируют лишь узкоспециализированную компетентность, оставаясь в рамках слабого ИИ. В заключении формулируются этические императивы для интеграции ИИ в гуманитарную сферу: необходимость разработки специализированных онтологий предметных областей, создание гибридных экспертных систем "человек-ИИ целесообразность сохранения эпистемологического суверенитета исследователя.

Received 30.03.2025