BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 28 No. 4. 2023

UDC 004.8
DOI: 10.62660/2306-4412.4.2023.82-90

Integration of artificial intelligence technologies in data engineering: Challenges and prospects in the modern information environment

Vasyl Nesterov*

Data Analyst
Itel Inc.
6676, 107 Corporate Center Pkwy., Jacksonville, United States of America
https://orcid.org/0009-0000-3204-1382

Abstract. The integration of artificial intelligence technologies into data engineering gained significant relevancy in the context of constantly growing volumes and complexity of data, which requires innovative approaches to processing and analysis. The goal of the present study is to conduct a deep analysis of the implementation of artificial intelligence into data engineering with a focus on the challenges occurring and perspectives of this process. Research methods, such as analysis methods, comparison, systematisation, and systemic approach, were used for an objective study of this phenomenon and revealing key aspects of this topic. Analysis revealed key challenges, that include variety and instability of data, the importance of standardisation as well as ensuring security of big data amounts. The importance of ethical aspects is underlined, and perspectives on automation of analytical processes and improving prognostic analysis were also determined. According to the results, employment of common standards improves the consistency of approaches, whereas improved algorithms accelerate the processing of big data amounts. Employment of such technology as Apache Hadoop and Spark for processing big data amounts and step-by-step introduction of artificial intelligence is also useful. Increased decision explication also improves their understanding, simplifying interaction between experts and interested parties, and simultaneously creating conditions for effective implementation and employment of integrated artificial intelligence systems in data engineering. The compilation of ethical standards and legal mechanisms creates an opportunity for responsible and balanced employment of these technologies, ensuring trust and ethical compliance in the process of their implementation into various spheres of human activity. These results determine perspectives for the development of this sphere and highlight its importance in a modern information-based society. Integration of artificial intelligence into data engineering expands capabilities of automating analytical processes, ensuring accurate predictions, and reducing manual labour expenses, creating opportunities for effective management and reasoned decision-making in the data processing sphere

Keywords: intelligent systems; optimisation; security of use; automation of analytical processes; machine learning

Article's History: Received: 28.08.2023; Revised: 30.10.2023; Accepted: 18.12.2023.

INTRODUCTION

In the modern information age, where data volumes are growing exponentially, data engineering is proving to be a key area to ensure effective management and analysis of information. In this context, the integration of artificial intelligence (AI) technologies is becoming a necessity to cope with the complexity and volume of data, which is turning into a real resource, a raw material for strategic decision-making. The use o f machine learning algorithms can greatly facilitate data analysis processes, providing more accurate and faster results. Given the fact that data is becoming an important resource for strategic decision-making, the integration of

Suggested Citation:
Nesterov, V. (2023). Integration of artificial intelligence technologies in data engineering: Challenges and prospects in the modern information environment. *Bulletin of Cherkasy State Technological University,* 28(4), 82-90. doi: 10.62660/2306-4412.4.2023.82-90.

*Corresponding author

AI into data engineering opens up new opportunities for the development of different approaches to data analysis, allowing businesses to adapt to the rapidly changing business environment.

Data engineering and AI, whose interaction is becoming a key factor in development, provide a new level of data analysis and optimisation. High-tech AI algorithms not only accelerate the processing of large amounts of information, but also provide an opportunity to automate and improve processes that require high computing power and analytical skills. Integrating AI technologies into data engineering can improve efficiency and cost-effectiveness in many areas. For example, T. Shmatkovska *et al.* (2023) note in their paper that technologies contribute to improved forecasting accuracy and more informed management decisions based on data analysis. In law enforcement, AI allows automating the processes of verifying individuals, analysing large amounts of data from various sources, and identifying potential risks, which helps to increase the efficiency and speed of response of law enforcement agencies, as pointed out in the article by O. Zachek *et al.* (2023). However, other issues arise here related to the need for changes in legislation on the introduction of AI. M. Ayala-Pazmiño (2023) and U.M. Danchenko (2023) note that in the field of education and science, the use of AI can improve the quality of education and research, contributing to the development of new teaching methods and increasing the productivity of research. A.I. Shevchenko (2023), having conducted an extensive study of this topic, emphasises that the use of this technology allows automating the analysis of large amounts of data, reducing the time and effort required to process them, which contributes to increased efficiency and cost-effectiveness in many industries.

However, it is important to highlight the challenges and prospects of integrating AI into data engineering. From security and ethical issues to optimising data processing and developing new technological solutions, the integration of AI into data engineering is changing the paradigm of information management. Z. Hbur (2022), studying the issues of AI application in Ukraine's information security, emphasises its importance. In particular, it was proved that AI technologies provide an opportunity to develop solutions with higher efficiency, which contributes to the rapid detection of cyberattacks, the selection of optimal strategies for responding to security incidents, automatic assessment of the relevance and consequences of incidents, and timely response. AI's ability to automate complex analytical and decision-making tasks in real time is making significant changes to the way people typically work with large amounts of data. Delving deeper into these aspects allows better understanding the power of AI integration and identifying promising ways to use IT technologies in various industries, from business to science, as studied by V. Khaustova *et al.* (2022). The increased availability of large amounts of data, their complexity, and diversity necessitate the development of new strategies for processing and analysis. The integration of AI allows solving these problems by speeding up decision-making processes and enabling complex analyses that were previously beyond the scope of capabilities.

The research aimed to study in detail the implementation of AI in data engineering, with a focus on identifying the challenges and prospects of this process. The objectives were to highlight key challenges and opportunities, as well as to analyse important aspects that should be taken into account when implementing these technologies in the modern information environment.

MATERIALS AND METHODS

In order to thoroughly study the integration of AI technologies in data engineering and reveal its challenges and prospects, this article used a variety of research methods aimed at an in-depth analysis and objective assessment of this phenomenon. During the research, methods of analysis, comparison, system approach and systematisation were used.

The comparison method included a thorough review of the literature, where different approaches to integrating AI into data engineering were studied. Special emphasis was placed on identifying differences and similarities in methodologies used in known scientific works and projects. An analysis of concepts, theories, and definitions used in relevant scientific sources was carried out to gain a deep understanding of the approaches that define the role of AI in data engineering. This made it possible to establish the common and distinctive features of various concepts and clarify the terminological aspects that were important for the precise definition of the research context. In addition, in the process of comparison, the infrastructure, and tools that contribute to the optimisation of the processing of large volumes of data in the context of AI integration were compared. This included an ecosystem of analytical tools, machine learning platforms and other technical solutions used to ensure the efficiency and speed of data processing in accordance with defined strategies and methods.

During the work, an analytical method was used to examine in detail the key concepts and technical aspects of the integration of AI in data engineering. The analytical approach made it possible to critically evaluate the available data, identify trends and identify gaps in the current understanding of this topic. Reviews of experts actively implementing AI in the practise of their activities were considered in detail. Their experiences and conclusions are analysed, looking at the technical aspects of AI integration, such as the use of programming languages, algorithms, and architectural solutions. This approach provided an opportunity to understand the specific challenges and benefits faced by practitioners, allowing for a detailed examination of their practical experiences. An important stage was the systematisation of the results obtained during the development of technical aspects of AI integration, which included the use of different programming languages, the deployment of algorithms and the choice of architectural solutions. The systematisation of technical details made it possible to structure information according to the effectiveness and potential limitations of AI integration.

The use of a system approach in the work made it possible to study the object of research holistically, in particular, to highlight a holistic approach to understanding the integration of AI in modern science and practise. This, in turn, made it possible to highlight in more detail the scale of the impact of the AI phenomenon itself, not only on the field of data engineering, but also on society as a whole. During the research, information resources, participation in conferences and interaction with professional communities were used. This allowed following the latest trends and innovations in the field of data engineering and AI, providing the research with relevance and contemporary context.

RESULTS

By analysing various aspects of integrating AI technologies into data engineering of this process, successful strategies and solutions have been identified that affect data security, quality and integrity, standardisation, optimisation of big data processing, and other key aspects (Talib *et al.*, 2021; Gröger, 2021; Ebid, 2021). Table 1 summarises the main challenges of AI integration and their solutions.

Table 1. Challenges and solutions in the context of integrating AI into data engineering

Challenges	Decision
Low data quality and reliability	Implementation of technologies for data cleaning and standardisation, use of quality control algorithms
Ethical and legal issues	Development of ethical standards and legal framework for the use of AI, definition of transparent decision-making processes
High complexity of integration with existing systems	Development of specialised interfaces and protocols for interaction between AI and existing systems
High costs and implementation issues	Study and implement effective cost reduction strategies, careful planning, and phased implementation
Low interpretability of models	Use of methods and algorithms that provide greater comprehensibility and interpretability of AI models
Data security and privacy risks	Applying advanced encryption tools, defining mechanisms to protect privacy, and taking into account ethical aspects of security

Source: developed by the author

One of the challenges is to ensure a high level of data security, as the integration of AI makes systems more susceptible to various cyber threats and attacks. Insufficiently protected infrastructure leads to the leakage of confidential information. While adaptive AI algorithms can detect and respond to new types of threats, they are also subject to targeted attacks that are becoming more sophisticated and difficult to detect. In addition, the increasing level of automation in data engineering, thanks to AI technologies, creates potentially vulnerable access points for unauthorised intrusion. AI technologies used to analyse data can lead to confidential information leaks, especially if the algorithms are not adequately protected against attacks and misuse. Integrating AI can push data processing and decision-making forward. However, this can lead to a large amount of load on the system, which can be used for denial-of-service (DDoS) attacks, and create unpredictable situations in the system. Using incorrect or distorted data to train AI models can lead to inaccurate or dangerous results. Employees and other internal users may attempt to exploit the system by using intelligent algorithms to uncover weaknesses and penetrate the system.

The study found that organisations that successfully integrate AI technologies into data engineering actively use a combination of approaches to ensure the highest level of protection. Elements of successful solutions include encryption and decryption. Successful integrations use modern encryption methods to protect sensitive information in real time. Additionally, they consider effective decryption mechanisms for secure access. The use of intelligent threat detection systems allows automatically detecting and responding to potential cyber threats. Machine learning algorithms are used to continuously improve the system in line with new threats. High-quality integrations include access security mechanisms and two-factor authentication to increase the level of aauthorisation and reduce the risk of unauthorised access. Event monitoring and logging systems allow detecting suspicious activity in real time and monitoring system changes. They use analytical tools to identify normal and abnormal patterns of behaviour. In the context of cloud computing, successful solutions provide strong data privacy protection using advanced encryption and key management techniques.

Another challenge to integrating AI into data engineering is data quality and purity. The use of poor quality data in integration processes leads to incorrect analytical conclusions and decision-making errors. Poor data quality can be caused by input errors, lack of standards, or outdated information. Duplication, instability, and a variety of data formats and structures arise when integrating data from different sources. This can create contradictions in information, complicating both analysis and decision-making. The issues of data simplification and standardisation are important aspects of solving these problems. The need to use data cleaning and normalisation algorithms arises to ensure that data is homogeneous and complies with established standards. This process is key to identifying and correcting errors. The use of old or outdated data in AI models can lead to incorrect predictions and incorrect analytical results. Updating and regularly reviewing data is essential to ensure that it remains relevant. For effective AI integration, it is necessary to ensure that the data source is reliable and trustworthy. Unreliable sources can introduce errors and distort analysis results, and identifying and correcting errors in large amounts of data can be time-consuming, especially when using machine learning algorithms.

Addressing data quality and integrity issues in the context of integrating AI technologies into data engineering requires a comprehensive approach and a variety of strategies. For example, the introduction of standards for data formats and structures simplifies data integration. Simplifying data, reducing duplication, and using common standards can make processing and analysis easier. Using automated algorithms to clean and normalise data can help identify and correct errors and inconsistencies. Ensuring that data is regularly updated, and monitoring systems are in place, can help identify and correct incorrect and outdated information in a timely manner. Use machine learning algorithms to detect and correct errors and to predict future data quality. Techniques such as classification and clustering can be used for automated data analysis. Using open sources and publicly available databases allows obtaining high-quality and up-to-date data to compare and update one's own data. Error detection and prevention systems can automatically detect and correct data errors. These systems can use business logic rules and machine learning algorithms. Involve staff and specialists to review and correct data, especially in cases where context and details that are difficult to automate need to be considered.

The diversity of data sources and their heterogeneity in formats and structures is a significant barrier to integration. Data can come from a variety of sources, such as databases, sensors, web services, etc. The diversity of data formats and structures leads to heterogeneity in the presentation of information. This may include differences in field names, data types, or even in the definition of key attributes. In some cases, there may be a lack of standardised formats for representing data in a particular industry or context. The lack of commonly accepted standards makes it difficult to unify them. Solving this problem requires standardisation and the development of unified approaches to data processing. Develop a canonical data model that serves as a common and standardised format for all data sources. This unifies the presentation of information and simplifies integration. Defining a data description language (DDL) to standardise the structures and relationships between data sources promotes schema consistency and facilitates integration. A variety of data integration tools and platforms are available that have built-in tools to address heterogeneity, including Extract, Transform, Load (ETL) processes and the use of middleware. Involve industry participants in the development and adoption of standards for data representation in a particular area. Creating common standards helps to unite the industry and facilitates the exchange of data between different systems. Use of Semantic Web technologies, such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL), to create semantic data models. This allows not only coordinating schemas, but also providing meanings and relationships between data.

Processing and analysing large amounts of data can be a time-consuming task, especially when using machine learning algorithms. Challenges stem from computational complexity and the need for powerful computing and storage resources. This leads to increased costs and difficulties in conducting effective maintenance. The transfer of large amounts of data between different systems often encounters network bandwidth limitations and, as a result, delays in the integration process. Ensuring the security of large volumes of data and reducing the risk of leakage or unauthorised access becomes a challenge. The use of technologies such as Apache Hadoop or Apache Spark for distributed processing of large amounts of data helps to divide computing and speed up analytical operations. Using big data storage systems, such as Apache HBase or Amazon S3, provides scalability and efficient management of large amounts of information. Data compression methods to reduce data size without losing information help reduce storage requirements and improve transmission efficiency. The use of cloud computing platforms allows for scalability of resources depending on the needs and high availability. Data preprocessing, such as filtering, weeding out uninformative or duplicate data, helps to reduce the volume of data and improve its quality. Increase network bandwidth and optimise data transfer mechanisms for efficient exchange of large amounts of information.

One important challenge is to understand and explicate the decisions made by an AI system. Failure to define how algorithms make decisions can make them difficult to adopt and implement. The lack of explication of how models work leads to a lack of trust in them, especially in areas where ethical considerations need to be taken into account and decisions need to be explained. In many cases, such as in medicine or finance, it is important not only to obtain a decision, but also to understand how this decision was made to justify its logic and objectivity. Using models that are easily interpreted allows explicating decision-making processes. This may include the use of simple models that are easy to explain, algorithms that generate explanations or justifications for each decision made. Creating decision support systems that allow users to request and receive detailed explanations for each stage of decision-making.

The use of intelligent systems in the field of data engineering raises ethical issues, in particular, in the areas of privacy, discrimination and responsible use of technology. Bias resulting from improperly trained algorithms can cause discrimination in employment, finance, or justice. The use of AI may violate privacy and personal data protection. In the process of addressing this issue, it is proposed that the creation of independent ethics committees and regulatory bodies that define standards and rules for the use of AI from an ethical and legal perspective will help resolve this issue. It is also worth paying attention to the introduction of algorithms aimed at identifying and reducing bias in decisions, ensuring fairness and equality. Develop and implement mechanisms to ensure confidentiality and protection of personal data, such as encryption and security protocols for information processing. Another solution is to introduce ethical training in the educational programmes of specialists working with artificial intelligence to raise awareness and understanding of ethical requirements. Continuous monitoring and evaluation of the impact of AI on society and individuals to identify ethical issues and address them in a timely manner.

Integrating AI can be costly and require significant investments in hardware and staff retraining. It is important to balance the costs and expected benefits, given that integrating new technologies into existing systems can be complex and time-consuming, especially in companies with large amounts of data and extensive infrastructure. The use of AI requires the availability of qualified specialists who can understand, configure, and manage these technologies, which can be problematic due to the shortage of specialists in the labour market. It is suggested to use a strategy of gradual introduction and scaling of AI technologies, which allows for their gradual integration into existing processes without suddenly affecting the entire business. Consideration may be given to the use of off-the-shelf AI platforms and solutions that can significantly simplify the integration process and reduce costs. Study the experience of other companies and use the best practises in AI implementation, which helps to avoid common mistakes and optimise the process.

High-quality machine learning models can be poorly interpreted, making it difficult to understand their decisions and influence analytical conclusions. Many modern AI models, such as neural networks, have a complex structure, which can make their decisions difficult to interpret and understand for experts and users. Some models may be black boxes, meaning that users cannot access the internal mechanisms and parameters, which affects their openness and verifiability. Hence, when choosing models, give preference to those that are more interpretable and generate explanations for their decisions. Use visualisation techniques to present how the models work and their impacts, which helps to understand the decision-making process. Adding special explanatory modules to the models that create or provide explanations for specific decisions, which helps to make them more interpretable. Use algorithms that not only provide accurate results but also explain how they were arrived at, such as SHapley Additive exPlanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME). Provide training for professionals and users to understand the principles of the models used. Establish ethical standards regarding model interpretability to ensure trust and transparency in areas where this is important.

DISCUSSION

The introduction of AI into data engineering creates several challenges and opportunities that require careful consideration and a rational approach. Data engineering is a key element in ensuring the success of AI adoption by defining a sustainable framework for clean and organised data. Among the key issues related to the implementation of AI in data engineering are data reliability and quality. Data scientists need to ensure that data is reliable and of high quality to ensure that AI systems operate efficiently. As AI relies heavily on input data, the reliability, and quality of this data becomes key to achieving optimal results. Compared to this study, Y. Mohan and A.P. Singh (2022) take a different approach in their article and consider the key aspects of data management in the modern environment. The article emphasises the need to create reliable data for the effective use of AI, which is also evidenced by the results obtained in this study. A high standard of data quality is a prerequisite for the accuracy and reliability of AI analytical models. Data reliability is determined by its validity and resistance to internal or external influences. If the data is not reliable, there is a risk of incorrect analysis and inaccurate decision-making by AI systems. Also, high data quality means that the data is clean, accurate, and complies with specific standards. In this context, data preparation for machine learning models becomes an important issue, requiring a lot of effort from data scientists and engineers. In addition, as noted by S. Martínez-Fernández *et al*. (2022), research related to software testing and software quality is very common, while areas such as software maintenance are ignored. Comparatively, there is a similarity in the findings, which is that data-related issues are the most persistent problems.

Data development turns out to be key in the process of collecting, cleaning, and transforming data to train AI models. Sources and methods of data collection are identified, including sensors, databases, text sources, and more. Data cleaning includes the process of eliminating errors, identifying and correcting incomplete or incorrect records. This stage is important because the quality of the input data determines the accuracy and reliability of AI models. Data transformation ensures that the data is suitable for use in AI training algorithms. This may include standardising formats, extracting important features, or even creating new characteristics to improve the predictive ability of the models. Compared to the author's results J. Serey *et al.* (2021) also offer an in-depth review of data development strategies and techniques for effective AI use. The paper details data sources, including sensors, databases, text sources, and more, and offers practical tips on how to collect and process them. What is different from this work is that J. Serey *et al.* (2021) focus more specifically on the methods of collecting and processing information.

The prospects for implementing AI in data engineering are significant. For example, improved data discovery, where AI technologies can optimise the data search process, saving time and effort for data consumers and providing access to a wider range of data. However, it is also necessary to keep in mind the security and ethical aspects of using AI in data engineering, as outlined in these results and in the work by S. Bharati and P. Podder (2022). It is important to keep in mind that the growing power of AI technologies should be used responsibly. Referring to the article by A. Baird and B. Schuller (2020), ethical standards should become an integral part of the development and use of integrated AI systems, as noted in the case study. The results have much in common. Ensuring transparency and accountability in the choice of algorithms, data processing, and decision-making is an important condition for the adoption of these technologies. Organisations that integrate AI into data engineering must adhere to ethical principles and implement technical means to ensure confidentiality and protect personal data. When integrating AI into data engineering, the possible implications for data privacy and security should be considered. Data engineers and developers should actively work to develop and refine an ethical framework for the use of AI in this context. The overall goal is to create integrated systems that not only effectively harness the power of AI in data engineering, but also ensure a high level of ethics and security for all users and stakeholders.

Simplified data integration and interoperability is a key benefit of implementing AI in data engineering. AI allows easily and efficiently combining different systems and data formats, removing barriers to interaction between different sources of information. AI systems can use algorithms and methods that automatically adapt to different data formats and understand their structure. One example is the introduction of natural language processing algorithms that allow recognition and adaptation to textual information from various sources, as described by J.M. John-Mathews *et al.* (2022) and F. Fui-Hoon Nah *et al.* (2023). This is interesting to consider in further research, as such approaches create a convenient and flexible mechanism for simplified data integration and interaction. Comparing the results of these authors, it is possible to note a difference in the specifics of the research topic and a deeper approach to the study of a particular object. The integration of AI into data processing services is expected to revolutionise data management, making it more scalable and cost-effective, as indicated by I.H. Sarker (2022). When comparing the results, it is worth noting the similarity in the definition of the revolution in information management.

Automation of data pipelines, enhanced by AI tools, is a strategic direction of data engineering development and means the use of algorithms and intelligent systems to optimise and automate various stages of data processing, which greatly facilitates the work of data developers. AI tools can implement technologies for automatic pattern detection, algorithms, and self-adaptive automated monitoring systems. According to A.R. Patel *et al.* (2021), this allows optimising workflows, reducing labour intensity and risks of errors during data development, which is also indicated in this study. In their work, experts note that the introduction of automation into data pipelines allows developers to focus on more valuable tasks, such as strategic planning, analysis, and improving data integration.

Machine learning and big data analytics play a key role in identifying risks as well as building accurate forecasts, which are necessary for strategic planning in various fields, as noted in the work by C. Zhang and Y. Lu (2021) and the study by J. Jöhnk *et al.* (2021). The difference between the results obtained lies in a clearer focus on the application of AI in the field of data analytics, and at the same time, more accurate results. An important aspect of integrating AI into data engineering is addressing the problem of data heterogeneity and standardisation, which requires the development of universal approaches to processing various formats and sources of information, as confirmed by the results of S.S. Gill *et al.* (2022) and this study. All of this indicates that the integration of AI into data engineering is becoming not only a technical challenge, but also requires a comprehensive approach that takes into account the economic, ethical, and social dimensions of this innovation process.

Thus, the prospects for implementing AI in data engineering include enhanced data discovery, facilitated data integration, improved data management efficiency, and automation of data engineering processes. A successful transition to the use of AI in data engineering involves addressing issues related to data reliability, quality, preparation, and automation. Harnessing the potential of AI can optimise data integration and processing, facilitating the effective use of these technologies in data engineering. The development of AI technologies in data engineering opens up new horizons for working with large volumes and diversity of data. Important aspects such as ensuring data security, improving the quality and purity of information, and addressing ethical and legal issues are becoming key focuses on modern information architecture. In the context of dynamic monitoring and forecasting of trends in data analytics, the role of AI is critical.

CONCLUSIONS

The integration of AI technologies into data engineering is a relevant and promising area of development in the modern information environment. Despite challenges such as financial costs, complexity of integration, and ethical issues, the benefits of increased efficiency, speed, and the ability to work with large amounts of data create positive prospects for the development of this area. The study results indicate that the integration of AI technologies into data engineering significantly improves various aspects of data processing, analysis, and security, making it an important area for the development of organisations in the modern information environment. First and foremost, the use of AI technologies allows increasing the efficiency of data processing and analysis, simplifying routine tasks and bringing them to a new level of automation.

Integration of machine learning algorithms helps to improve data security by allowing timely detection and prevention of possible threats. It is important to consider the interpretability and openness of models as a key element of trust in the AI systems used. The introduction of AI stimulates the development of new tools and technological solutions that expand the ability to process and analyse large amounts of data. AI-enhanced data engineering systems are becoming more consistent and interoperable, which contributes to more efficient data processing and the creation of information resources. Integration of AI into data engineering can be used to optimise business processes, which means improving efficiency, reducing costs, and improving the quality of decision-making. For example, implementing AI analytical algorithms can improve forecasting of market trends, and automating certain tasks can free up time to tackle more complex tasks, while contributing to increased productivity and competitiveness.

AI can be used to improve medical diagnostics, predict environmental changes, and optimise transport systems. AI technologies can help in a variety of industries, and improved ethical standards will help create safer and fairer use of AI in various fields, setting the stage for sustainable and successful implementation of these technologies. Prospects for further research include considering the impact of dynamic changes in the technological environment on the integration of AI technologies into data engineering, and developing more effective strategies for taking ethical and security aspects into account.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Ayala-Pazmiño, M. (2023). Artificial intelligence in education: Exploring the potential benefits and risks. *593 Digital Publisher CEIT*, 8(3), 892-899. doi: 10.33386/593dp.2023.3.1827.

[2] Baird, A., & Schuller, B. (2020). Considerations for a more ethical approach to data in AI: On data representation and infrastructure. *Frontiers in Big Data*, 3, article number 25. doi: 10.3389/fdata.2020.00025.

[3] Bharati, S., & Podder, P. (2022). Machine and deep learning for IoT security and privacy: Applications, challenges, and future directions. *Security and Communication Networks*, 2022, article number 8951961. doi: 10.1155/2022/8951961.

[4] Danchenko, U.M. (2023). Relevance of the study of artificial intelligence recipients of higher military education. In *Materials of the all-Ukrainian science-pedagogical professional development "Technologies for the benevolent use of artificial intelligence in the fields of education and science"* (pp. 83-85). Odesa: Helvetica.

[5] Ebid, A.M. (2021). 35 Years of (AI) in geotechnical engineering: State of the art. *Geotechnical and Geological Engineering*, 39, 637-690. doi: 10.1007/s10706-020-01536-7.

[6] Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277-304. doi: 10.1080/15228053.2023.2233814.

[7] Gill, S.S., Xu, M., Ottaviani, C., Patros, P., Bahsoon, R., Shaghaghi, A., Golec, M.,... Uhlig, S. (2022). AI for next generation computing: Emerging trends and future directions. *Internet of Things*, 19, article number 100514. doi: 10.1016/j.iot.2022.100514.

[8] Gröger, C. (2021). There is no AI without data. *Communications of the ACM*, 64(11), 98-108. doi: 10.1145/3448247.

[9] Hbur, Z. (2022). Use of artificial intelligence in information security in Ukraine. *Public Administration: Improvement and Development*, 1. doi: 10.32702/2307-2156-2022.1.2.

[10] Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes – an interview study of organizational AI readiness factors. *Business & Information Systems Engineering*, 63, 5-20. doi: 10.1007/s12599-020-00676-7.

[11] John-Mathews, J.M., Cardon, D. & Balagué, C. (2022). From reality to world. A critical perspective on AI fairness. *Journal of Business Ethics*, 178, 945-959. doi: 10.1007/s10551-022-05055-8.

[12] Khaustova, V., Reshetnyak, O., & Khaustov, M. (2022). Promising areas of IT development in the world. *The Problems of Economy*, 51(1), 3-19. doi: 10.32983/2222-0712-2022-1-3-19.
[13] Martínez-Fernández, S., Bogner, J., Franch, X., Oriol, M., Siebert, J., Trendowicz, A., Vollmer, A.M., & Wagner, S. (2022). Software engineering for AI-based systems: A survey. *ACM Transactions on Software Engineering and Methodology*, 31(2), article number 37e. doi: 10.1145/3487043.
[14] Mohan, Y., & Singh, A.P. (2022). Challenges and opportunities in Big data: A review. *International Journal of Research in Electronics and Computer Engineering*, 10(4), 36-46.
[15] Patel, A.R., Ramaiya, K.K., Bhatia, C.V., Shah, H.N., & Bhavsar, S.N. (2021). Artificial intelligence: Prospect in mechanical engineering field – A review. In K. Kotecha, V. Piuri, H. Shah & R. Patel (Eds.), *Data science and intelligent applications. Lecture notes on data engineering and communications technologies* (pp. 267-282). Singapore: Springer. doi: 10.1007/978-981-15-4474-3_31.
[16] Sarker, I.H. (2022). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science*, 3, article number 158. doi: 10.1007/s42979-022-01043-x.
[17] Serey, J., Quezada, L., Alfaro, M., Fuertes, G., Vargas, E., Ternero, R., Sabattin, J., Durán, C., & Lillo, S. (2021). Artificial intelligence methodologies for data management. *Symmetry*, 13(11), article number 2040. doi: 10.3390/sym13112040.
[18] Shevchenko, A.I. (Ed.). (2023). *Strategy for artificial intelligence development in Ukraine*. Kyiv: Nauka i Osvita. doi: 10.15407/development_strategy_2023.
[19] Shmatkovska, T., Korobchuk, T., & Borysiuk, O. (2023). Modern information and communication technologies in the system of accounting and analytics for the modeling of business processes. *Economy and Society*, 53. doi: 10.32782/2524-0072/2023-53-68.
[20] Talib, M.A., Majzoub, S., Nasir, Q., & Jamal, D. (2021). A systematic literature review on hardware implementation of artificial intelligence algorithms. *The Journal of Supercomputing*, 77, 1897-1938. doi: 10.1007/s11227-020-03325-8.
[21] Zachek, O., Dmytryk, Y., & Senyk, V. (2023). The role of artificial intelligence in increasing efficiency law enforcement activities. *Scientific Bulletin of the Lviv State University of Internal Affairs*, 3, 148-156. doi: 10.32782/2311-8040/2023-3-19.
[22] Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23, article number 100224. doi: 10.1016/j.jii.2021.100224.

Інтеграція технологій штучного інтелекту в інженерію даних: виклики та перспективи у сучасному інформаційному середовищі

Василь Федорович Нестеров

Дата-аналітик
Itel Inc.
6676, бульв. Корпоративний Центр, 107, м. Джексонвіль, Сполучені Штати Америки
https://orcid.org/0009-0000-3204-1382

Анотація. Інтеграція технологій штучного інтелекту в інженерію даних набула значущої актуальності у контексті постійно зростаючого обсягу та складності даних, що вимагає новаторських підходів до їх обробки та аналізу. Метою даної статті є проведення глибокого аналізу процесу впровадження штучного інтелекту в інженерію даних з акцентом на виклики, які виникають, та перспективи цього процесу. Методи дослідження, такі як методи аналізу, порівняння, систематизації, синтезу були використані для об'єктивного вивчення цього явища та для розкриття ключових аспектів даної теми. Аналіз виявив ключові виклики, що включають різноманітність та нестабільність даних, важливість стандартизації, а також забезпечення безпеки великих обсягів інформації. Підкреслюється також важливість етичних аспектів, а також визначені перспективи в автоматизації аналітичних процесів та покращенні прогностичного аналізу. Згідно з результатами, застосування спільних стандартів покращує узгодженість підходів, а вдосконалені алгоритми прискорюють обробку великих обсягів даних. Корисним є використання технологій, таких як Apache Hadoop та Spark, для обробки великих обсягів даних, а також стратегії поступового впровадження штучного інтелекту. Збільшена експлікація рішень сприяє їхньому розумінню, полегшуючи взаємодію між фахівцями та зацікавленими сторонами, та водночас створює умови для ефективного впровадження та використання інтегрованих систем штучного інтелекту в інженерії даних. Вироблення етичних стандартів та правових механізмів відкриває шлях до відповідального та збалансованого використання цих технологій, забезпечуючи довіру та етичність у процесі їх впровадження в різноманітні сфери діяльності. Ці результати визначають перспективи розвитку цієї області та підкреслюють її важливість у сучасному інформаційному середовищі. Інтеграція штучного інтелекту в інженерію даних розширює можливості автоматизації аналітичних процесів, забезпечуючи точні прогнози та зменшуючи витрати на ручну роботу, що відкриває перспективи для ефективного управління та прийняття обґрунтованих стратегічних рішень у галузі обробки інформації

Ключові слова: інтелектуальні системи; оптимізація; безпека користування; автоматизація аналітичних процесів; машинне навчання