УПРАВЛІННЯ РОЗВИТКОМ ОСВІТИ

UDC 378.09
DOI https://doi.org/10.52726/as.pedagogy/2024.2.14

M. V. NADUTENKO
Candidate of Philological Sciences, Senior Researcher,
Ukrainian Lingua-Information Fund of the National Academy
of Sciences of Ukraine, Kyiv, Ukraine
E-mail: margo.nadutenko@gmail.com
https://orcid.org/0000-0003-3215-0970

M. V. NADUTENKO
Candidate of Technical Sciences in Structural, Applied, and Mathematical Linguistics, Head of Informatics Department of Ukrainian Lingua-Information Foundation of National Academy of Sciences (NAS) of Ukraine,
Secretary of Scientific Council "Information. Language. Intellect" of NAS of Ukraine, Kyiv, Ukraine
E-mail: hostmaster@ulif.org.ua
https://orcid.org/0000-0001-6732-8455

O. L. FAST
Candidate of Sciences (Pedagogics), Associate Professor,
Vice-Rector for Research, Teaching and International Relations,
Municipal Higher Educational Institution "Lutsk Pedagogical College"
of the Volyn Regional Council, Lutsk, Ukraine
E-mail: ofast@lpc.ukr.education
https://orcid.org/0000-0001-7216-0044

THE TRIAD "INTELLECTUALIZATION – VIRTUALIZATION – BIG DATA": INTERDISCIPLINARY ISSUES ON NATURAL LANGUAGE STUDIES AND DOCTORAL EDUCATION PERSPECTIVES

This article explores the interdisciplinary issues of natural language studies and the perspectives of doctoral education in the context of the triad «Intellectualization – Virtualization – Big Data.» Recent advancements in technology have significantly propelled progress in the field of natural language processing (NLP). Modern NLP relies heavily on machine learning and deep learning techniques, enabling computers to learn from extensive datasets and execute tasks with high precision.

Key aspects highlighted in the article include the use of neural networks, attention mechanisms, and transfer learning. The primary challenges facing contemporary NLP, such as ambiguity, data scarcity, lack of context, and ethical concerns, are thoroughly discussed. Important research directions include multimodality, explainable neural networks, and the development of NLP systems for low-resource languages.

The «Results and Discussion» section emphasizes the technological status of NLP and its significance in the realms of intellectualization, virtualization, and big data analytics. Potential research areas are identified, such as enhancing the intellectualization of NLP systems, creating virtual environments for NLP applications, and leveraging big data analytics to improve the accuracy and efficiency of NLP systems.

Additionally, the article examines the training of doctoral students in the context of NLP and big data analytics, highlighting the development of competencies in research data management and the methodological capital of educational data science. The conclusion underscores the importance of integrating ethical considerations into the use of NLP technologies and their relevance to contemporary doctoral education in Ukraine.

The study highlights the critical need for an integrated approach to language analysis, virtual environments, and big data for the development of NLP systems and the modernization of doctoral education.

Key words: natural language processing (NLP), big data, doctoral education, doctoral students training, intellectualization, virtualization.

Introduction. In recent years, the development of technology has enabled significant progress in the field of natural language processing (NLP).

The basis of modern NLP is machine and deep learning. These techniques allow computers to learn from large data sets and perform tasks such as language modeling, text classification, and sentiment analysis with high accuracy. Machine learning algorithms, particularly deep neural networks, have shown extraordinary success in NLP tasks, outperforming traditional rule-based approaches.

Neural networks are a type of machine learning algorithm that has shown extraordinary success in NLP tasks. They are particularly effective in tasks such as language modeling, where they can learn to predict the next word in a sentence based on the context provided by previous words.

Word embedding is a technique used to represent words as vectors in a high-dimensional space. This allows computers to capture the semantic meaning of words and phrases, allowing them to perform tasks such as text classification and tonality analysis with greater accuracy.

Attention mechanisms are a technique used to focus the computer's attention on specific parts of the input data when performing NLP tasks. This can be particularly useful in tasks such as machine translation, where the computer must pay attention to both the source language and the target language.

Transfer learning is a technique that allows computers to use knowledge gained from one task to perform another related task. In practice, when using large language models, you don't have to deal with creating new models and training them from scratch on the client side. Most often, it will be necessary to create models on the basis of already existing ones. This technique is called Transfer Learning. This is a technology that has greatly advanced the field of NLP. This approach involves using pre-trained models and fine-tuning them for specific tasks, resulting in more accurate and efficient work. Transfer learning has enabled NLP systems to achieve state-of-the-art results in a variety of applications, including translation, text tonality analysis, and text generation.

Challenges of modern NLP:

– Ambiguity. One of the biggest challenges in NLP is dealing with ambiguity. Natural language is inherently ambiguous, and words and phrases often have multiple meanings. This can make it difficult for computers to accurately interpret and understand human language.

– Limited data. Another problem in NLP is the limited availability of data. Many languages do not have a large amount of data available for training and testing, which can limit the accuracy of NLP systems.

– Lack of context. NLP systems often struggle to understand the context in which language is used. This can lead to errors and misunderstandings, especially in tasks such as machine translation.

– Ethical problems. NLP raises a number of ethical concerns, such as the potential for bias in algorithms and the impact of automation on employment. These issues need to be addressed to ensure that NLP systems are used responsibly and ethically.

One of the most important areas of research is multimodality, which involves the analysis and understanding of language in combination with other modalities, such as computer vision, speech recognition and generation. Another direction of research is explanatory (self-explanatory) neural networks, the purpose of which is to provide transparency and interpretation of NLP systems. In addition, there is growing interest in developing NLP systems that can work with low-resource languages, which can have a significant impact on language and cultural heritage preservation.

Results and Discussion. The technological status of natural language has made significant progress in recent years thanks to advances in machine learning and deep learning. However, to ensure responsible and ethical use of NLP systems, issues such as ambiguity, data limitations, and ethical issues must be addressed.

The triad "Intellectualization – Virtualization – Big Data" had a significant impact on the technological status of natural language. Research directions that determine the interaction between these three factors are:

Intellectualization of NLP. With the increasing availability of large data sets and advances in machine learning, NLP has become an intellectually demanding field that requires a deep understanding of linguistics, computer science, and statistics. Scientists are exploring new methods of intellectualizing NLP, such as incorporating subject-specific knowledge, using cognitive computing techniques, and developing more complex models that can capture the nuances of human language.

Virtualization of NLP. The advent of virtual assistants, chatbots, and other virtual interfaces has led to the virtualization of NLP. Researchers are looking for ways to create more interactive virtual environments that can mimic human conversations and improve user interaction. This includes the development of more advanced dialogue systems, the integration of NLP with computer vision and other means of sensory input, and the creation of personalized avatars that can be adapted to the user's preferences and needs.

Big Data Analytics for NLP. The abundance of digital data has created new opportunities for research and application of NLP. Researchers use big data analytics to improve the accuracy and efficiency of NLP systems, develop more specialized models for specific domains, and explore new applications of NLP in areas such as social network analysis, sentiment analysis, and information retrieval.

In our opinion, possible areas of research in this field are:

Research in the field of intellectualization of NLP systems to better understand the complexity and nuances of human language.

Exploring opportunities for creating virtual environments for NLP programs.

Investigating the potential application of big data analytics in NLP to improve the accuracy and efficiency of NLP systems.

Exploring the interrelationships in the Intelligence-Virtualization-Big Data triad in NLP systems, the potential trade-offs and limitations of each approach.

Overall, the interplay between intelligence, virtualization, and big data analytics can significantly impact the state of natural language technology, enabling more sophisticated and accurate NLP systems that can mimic all the nuances of human speech and improve user interaction.

Research in the field of intellectualization of NLP systems.

Improving the intellectualization of NLP systems to better capture the complexity and nuances of human language is an ongoing challenge in the field. Possible directions of research in this field are as follows:

Multimodal NLP. Incorporating multiple modalities, such as computer vision, speech, and gesture recognition, into NLP systems can provide a more comprehensive understanding of human language and behavior. It is necessary to study the possibilities of using methods of combining information from different modalities and increasing the accuracy of NLP systems.

Cognitive NLP. Integrating cognitive science and NLP can help create more intelligent and human-like NLP systems. Research is needed on the use of cognitive architectures, such as the SOAR cognitive architecture, the core cognitive pathway, and SCC (Structure, Substance, Subject) to model human cognition and improve the performance of NLP systems.

The SOAR (State-Object-Action-Result) cognitive architecture is a theoretical framework for understanding human cognition and decision-making. It was developed by psychologists and cognitive scientists, notably John R. Anderson and his colleagues, as an alternative to more traditional cognitive architectures such as information processing theory.

According to the SOAR system, cognition is based on the interaction of four main components:

State: This refers to a person's current mental state, including their beliefs, desires, and goals. The state component represents the overall cognitive context in which the individual operates.

Object: This refers to external stimuli or objects that a person perceives and processes. An object component represents specific aspects of the environment to which an individual pays attention.

Action: This refers to actions or responses that a person is considering or planning. The action component represents a person's intentions and goals and the strategies they use to achieve those goals.

Result: This refers to the results or consequences of a person's actions. The result component is the feedback that the individual receives about the effectiveness of his actions.

The SOAR concept emphasizes the dynamic and iterative nature of cognition, suggesting that cognitive processes constantly interact and influence each other. For example, the constituent component can influence the object component, shaping the attention and perception of the personality of the environment. Similarly, an action component can influence an outcome component, influencing the outcome of a person's actions.

One of the key strengths of the SOAR framework is its ability to explain a wide range of cognitive phenomena, including decision making, problem solving, and learning. It also provides a useful framework for understanding the role of emotion and motivation in cognition, as well as the influence of cognitive biases and heuristics on decision making.

Overall, the SOAR cognitive architecture provides a comprehensive and flexible framework for understanding the complex and dynamic nature of human cognition. Its emphasis on the interconnectedness of cognitive processes and its ability to explain a wide range of cognitive phenomena make it a valuable tool for researchers and practitioners in fields such as psychology, neuroscience, and education.

Competitive NLP. Competitive training techniques can be used to improve the reliability and generalization abilities of NLP systems. Research is needed on the use of adversarial methods to assess the strengths and weaknesses of NLP systems and develop more resilient and adaptive systems.

In 1991, Jürgen Schmidhuber published adversarial neural networks, which compete with each other in the form of an antagonistic game, where the win of one network is the loss of the other. The first network is a generative model that simulates a probability distribution over the output images. A second network is trained by gradient descent to predict the environment's response to these images. This was called "artificial curiosity".

Metalearning NLP. Metalearning is a machine learning technique that involves learning how to learn. Research is needed on the use of metalearning in NLP to improve the adaptability and rapid learning ability of NLP systems.

NLP "Human in the loop". Collaborative human-computer interaction can improve the intelligence of NLP systems by incorporating human feedback and judgment into the learning process. Research and development of techniques for using human-in-the-loop techniques such as active learning to improve the performance of NLP systems is needed.

A typical learning process for an ML model is a cascading process in which the next stage begins after the previous one is completed: the task is set and the project structure is defined; data is collected and manually marked, often by a team of employees; data-scientists try several model architectures and learning techniques, from the simplest to the most powerful, to see which ones work best; if the chosen architecture and technique provide good performance, it is deployed on the server.

A considerate disadvantage of this approach is that in the first stages it requires a lot of time to receive feedback, that is, the iteration cycles are very slow, and the stages must be repeated. This approach has several significant disadvantages: at the beginning of the task, the task is often given vaguely; it is difficult to understand in advance how the data should be annotated; test data are not real data, and this will be revealed only at the end; new data must be repeatedly annotated to track current model performance and improve the system.

Due to the "human-in-the-loop" approach, instead of hoping for smooth and linear progress, an iterative approach is used to build the learning model. The working model is trained from the first data elements. It is updated frequently as new data is added. The model and subject matter experts work together to build, adapt, and improve the model by annotating data or modifying the task to refine requirements and performance. Active learning is a type of human-in-the-loop learning in which the data for annotation is selected by the model. By focusing on the most informative data, you can significantly reduce the amount of data required. Thanks to the presence in the feedback loop of the model, it is possible to perform preliminary marking of data. This allows you to organize a workflow where annotation turns into a confirmation/rejection task of recommended AI predictions.

Attributable to all of the above, this method is much better suited for training models. A serious obstacle to implementation is the complexity of preparation. Most annotation interfaces do not allow working with a learning model.

The disadvantage of human-in-the-loop systems is that humans make mistakes, so any human-in-the-loop system is prone to error. Another disadvantage of such systems is their expensive maintenance, as it should involve significant costs for human intellectual work.

Explainable (self-explanatory) NLP systems. Developing NLP systems that can provide transparent and interpretable results is essential to building trust and confidence in these systems. Research into the use of techniques such as attention mechanisms, feature importance, and visualization to improve the explainability of NLP systems is needed.

Multilingual NLP systems. NLP systems that can handle multiple languages and dialects can provide a more comprehensive understanding of human language and culture. Research is needed on the use of multilingual models and transfer learning to improve the performance of multilingual NLP systems.

Emotional NLP systems. NLP systems that can recognize and respond to emotions can provide a more human-like and empathetic interaction experience. Research is needed to explore the use of affective technologies, such as emotion detection and sentiment analysis, to improve the emotional intelligence of NLP systems.

Social NLP. NLP systems that can understand and respond to social cues such as social norms and expectations can provide a more human-like and socially acceptable interaction experience. Research is needed on the use of social learning and social signal processing methods to improve the social intelligence of NLP systems.

Ethical NLP. Ensuring that NLP systems are fair and respect privacy and security is essential to their widespread adoption. Research is needed on the development and use of techniques such as de-biasing, maintaining confidentiality and secure communication to enhance the ethicality of NLP systems.

Exploring opportunities for creating virtual environments for NLP programs.

Creating interactive virtual environments for NLP applications can be achieved through combining the following technologies and techniques.

Virtual Reality (VR) and Augmented Reality (AR). VR and AR technologies can be used to create immersive, interactive virtual environments that mimic real-world settings. For example, a VR environment can be designed to simulate customer interactions, allowing users to practice their language skills in a realistic environment.

3D avatars and characters. 3D avatars and characters can be used to create more engaging and interactive virtual environments. These avatars can be customized to represent different genders, ages, and cultures, and animated to express a variety of emotions and reactions.

Dialogue systems. Dialogue systems such as chatbots and voice assistants can be used to create more interactive and engaging virtual environments. These interfaces can be integrated with NLP systems to mimic human conversation and provide a more personalized and responsive interaction.

Gamification. Gamification techniques such as points, badges, and leaderboards can be used to make NLP programs more engaging and interactive. These techniques can be used to encourage users to practice their language skills and create a sense of accomplishment and achievement.

Social learning. Social learning techniques can be used to create more immersive and interactive virtual environments. For example, users can interact with other students in a virtual classroom or collaborative project, providing a sense of community and social support.

Personalization and feedback. Personalization methods can be used to adapt the virtual environment to the individual user. For example, the system can adjust the level of difficulty, content and feedback based on the user's progress and preferences.

In a virtual environment, users can be given real-time feedback, allowing them to improve their language skills and receive instant feedback on their performance. This can be achieved using NLP algorithms and machine learning techniques.

Doctoral education contexts

NLP and Big data analytics are among the core issues in navigating doctoral education according to the recent research articles analysis:

– Sound research data management (RDM) competencies are seen as elementary tools used by researchers to ensure integrated, reliable, and re-usable data, and to produce high-quality research results (Jukka Rantasaari, 2021).

– Education data science might be understood as assembling a new form of 'methodological capital' as educational institutions generate increasing quantities of digital data (Ben Williamson, 2017).

– Training the next generation of doctoral researchers in data science: challenges and recommendations (Papagiannidis, Meadows & Panagiotopoulos, 2023), etc.

In line with many studies, data science techniques have expanded research opportunities by creating novel methodological landscapes and strategies for tackling research topics (Papagiannidis, Meadows & Panagiotopoulos, 2023). "The single biggest stimulus of new tools and theories of data science is the analysis of data to solve problems posed in terms of the subject matter under investigation. Creative researchers, faced with problems posed by data, will respond with a wealth of new ideas that often apply much more widely than the particular data sets that gave rise to the ideas." (Cleveland, p. 22).

Extolling the potential of data science methods, researchers caution that research should be driven by "big questions" and not by "big data" (McKenna, Myers, and Newman, 2017). Moreover, according to Boyd and Crawford (2017), certain mythology also takes place: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy".

B. Williamson (2017) states that, "the methodological capital of educational data science consists of competence in big data analyses, the ability to secure funding and strategic partnerships and the capacity to produce knowledge and theory that may be effective in the competition for control over contemporary understandings of e-learning, digital media, and education" (p. 120).

Based on the needs of training "cross-functional, cross-discipline, omni-knowledgeable researchers who can tackle any research objective that can deliver valuable analytical insights from day one", Papagiannidis, Meadows & Panagiotopoulos, (2023) have developed the framework that identifies three areas of tension related to big data applications: organisational learning (Learning), organisational leadership (Leading) and societal tensions (Linking), putting forward a set of recommended actions (Table 1).

Table 1

Learning, Leading, Linking Framework Recommendations

Challenges and areas of attention		**Key recommendations**
Learning	*Hard skills*	Provide regular training and upskilling in response to changes in data science practice and publication expectations. Involve external partners to fill gaps in advanced skills and support doctoral supervision and training. These could include interdisciplinary collaborations within universities or engagement with national data centres, institutes and government statistical units.
	Soft skills	Provide training that builds awareness of the individual characteristics (such as personality traits and competencies, motivation, social skills, etc.) that support effective use of data skills. Introduce training on soft skills that are important for complex data projects, which are often interdisciplinary in nature.
	Learning processes	Create awareness, not just of how to operate in a diverse environment in relation to data, but also how to contribute to such an environment, for example, how to actively address data management and data quality issues in complex projects. Ensure that researchers have experience of engaging with data science experts and other domain specialists in at least one interdisciplinary project. This could be organised in the form of a data study group, consulting project or placement.
Leading	*Strategic vision*	Establish academic development that incorporates data skills as a key dimension for all research activities. For example, create awareness of how academic projects can develop comprehensive strategies that reflect data management challenges. Create mentoring schemes that encourage not just sharing of experiences but also create opportunities for long-term collaborations.
	Academic impact	Explain how researchers can use data-driven approaches to create value for stakeholders beyond academic publications, for example, in providing training related to the sustainable data management of outputs from research projects.
	Data knowledge practice	Facilitate knowledge sharing and participation in mentorship networks and communities with expertise in data science practice.
Linking	*Data ecosystems*	Introduce training that creates awareness of diverse data ecosystems and their associated characteristics, opportunities and challenges (e.g. smart cities, health data, data utilities, marketplace platforms). Develop an appreciation of the differences between academic-to-academic, academic-to-business and academic-to-government collaborations. Provide training that integrates the perceived legitimacy and trustworthiness of data in the eyes of key stakeholders involved in projects.
	Responsible data science practice	Incorporate guiding principles, ethical considerations and practical examples of responsible use of data science methods.

Conclusions. The use of big data analytics in NLP can improve the accuracy and efficiency of NLP systems in several ways.

Big data analytics allows you to collect huge amounts of data from various sources, such as social networks, news or publications. This data can be used to train NLP models, improving their accuracy and generalization ability.

With big data analytics, NLP models can be trained on larger and more diverse data sets, resulting in increased accuracy and reliability. This can be achieved using such methods as transfer learning, ensemble learning, and multi-task learning.

Big data analytics allows processing of large data sets in real time, enabling faster and more efficient NLP system development. This can be particularly useful in applications such as tonal analysis of texts, where timely understanding is critical.

Big data analytics can help personalize NLP systems for specific users or domains, increasing their effectiveness and relevance. For example, by analyzing user behavior and preferences, NLP systems can be adapted to provide more relevant recommendations or responses.

Big data analytics can help detect and correct errors in NLP systems, improving their overall accuracy and reliability. This can be done by analyzing large volumes of data to identify patterns and trends in errors and adjust NLP models accordingly.

Big data analytics can help in understanding decision-making processes in NLP systems, improving their explainability and reliability. This can be achieved through techniques such as feature importance analysis and visualization, which can help identify the most influential factors in NLP solutions.

With big data analytics, it is possible to combine multiple modalities such as text, speech and computer vision to improve the accuracy and reliability of NLP systems. This can be achieved by analyzing large amounts of data from multiple sources and combining the information obtained from each method.

Thus, the use of big data analytics in NLP can improve the accuracy and efficiency of NLP systems by enabling large-scale data collection, improved model learning, real-time processing, personalization, adaptation, error analysis, explainability, continuous improvement, and multi-modality fusion. However, it is important to consider the ethical considerations associated with big data analytics in NLP to ensure that these systems are used responsibly and for the benefit of society.

The issues raised in this discussion are extremely relevant when it comes to the doctoral students training. It is undeniable that Big data analytics, NLP technologies will complement the landscape of doctoral education research with new contexts, challenges, and form new perspectives. A cursory analysis of foreign pedagogical discourse evidently confirms this, which at the same time actualizes the deepening in researching this problem and examining the effective strategies for the doctoral education modernization in Ukraine.

REFERENCES

1. B. McKenna, M. D. Myers, and M. Newman, (2017). Social media in qualitative research: Challenges and recommendations. *Information and Organization,* vol. 27, no. 2. Pp. 87–99.
2. Cleveland, W. S., (2001). Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review,* 69, no. 1. Pp. 21–26.
3. D. Boyd and K. Crawford, 92012). Critical Questions for Big Data. *Information, Communication & Society*, vol. 15, no. 5. Pp. 662–679.
4. G. T. L. Brown, (2017). Doctoral Education in Quantitative Research Methods: Some Thoughts about Preparing Future Scholars. *Frontiers in Applied Mathematics and Statistics,* vol. 3. P. 25.
5. Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua, (2014). Generative Adversarial Networks Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014) (англ.). С. 2672–2680. https://web.archive.org/web/20191122034612/http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf
6. Papagiannidis, S., Meadows, M. & Panagiotopoulos, P., (2023). Training the next generation of doctoral researchers in data science: The impact on publications and beyond. *IEEE Transactions on Technology and Society,* vol. (In-Press), pp. (InPress). https://doi.org/10.1109/TTS.2023.3243395
7. Schmidhuber, Jürgen (1991). A possibility for implementing curiosity and boredom in model-building neural controllers. Proc. SAB'1991 (англ.). MIT Press/Bradford Books. С. 222–227.
8. Schmidhuber, Jürgen (2010). Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990-2010). *IEEE Transactions on Autonomous Mental Development* (англ.). 2 (3): 230–247. doi:10.1109/TAMD.2010.2056368.

9. Schmidhuber, Jürgen (2020). Generative Adversarial Networks are Special Cases of Artificial Curiosity (1990) and also Closely Related to Predictability Minimization (1991). *Neural Networks* (англ.). 127: 58–66. PMID 32334341. arXiv:1906.04493. doi:10.1016/j.neunet.2020.04.008.

10. Sioux McKenna & Susan van Schalkwyk (2023): A scoping review of the changing landscape of doctoral education, Compare: *A Journal of Comparative and International Education*. DOI: 10.1080/03057925.2023.2168121

11. Williamson, B. (2017). Who owns educational theory? Big data, algorithms and the expert power of education data science. *E-Learning and Digital Media.* 2017, Vol. 14(3) 105–122.

12. Nadutenko M., Prykhodniuk V., Shyrokov V., Stryzhak O. Ontology-Driven Lexicographic Systems. Advances in Information and Communication. FICC 2022. *Lecture Notes in Networks and Systems. Cham : Springer.* 2022. С. 204–215. DOI: 10.1007/978-3-030-98012-2_16

13. V.A. Shyrokov, M.V. Nadutenko, et al. Linguistic-Informational Studies: Works of the Ukrainian Lingua-Information Fund of the National Academy of Sciences of Ukraine: in 5 volumes. Volume 5. Virtualization of Linguistic Technologies. Kyiv, Ukrainian Lingua-Information Fund of the National Academy of Sciences of Ukraine. 2018. 286 pages. ISBN 978-966-02-8683-2. ISBN 978-966-02-8690-0 (Vol. 5). DOI 10.33190/978-966-02-8683-2/8690-0

М. В. НАДУТЕНКО
кандидат філологічних наук, старший науковий співробітник,
Український мовно-інформаційний фонд Національної академії наук України, м. Київ, Україна
Електронна пошта: margo.nadutenko@gmail.com
http://orcid.org/0000-0003-3215-0970

М. В. НАДУТЕНКО
кандидат технічних наук зі структурної, прикладної та математичної лінгвістики,
завідувач відділу інформатики,
Український мовно-інформаційний фонд Національної академії наук України,
секретар наукової ради «Інформація. Мова. Інтелект» НАН України, м. Київ, Україна
Електронна пошта: hostmaster@ulif.org.ua
https://orcid.org/0000-0001-6732-8455

О. Л. ФАСТ
кандидат педагогічних наук, доцент,
проректор з науково-педагогічної роботи та міжнародної співпраці,
Комунальний заклад вищої освіти «Луцький педагогічний коледж»
Волинської обласної ради, м. Луцьк, Україна
Електронна пошта: olhafast.ua@gmail.com
http://orcid.org/0000-0001-7216-0044

ТРІАДА «ІНТЕЛЕКТУАЛІЗАЦІЯ – ВІРТУАЛІЗАЦІЯ – ВЕЛИКІ ДАНІ»: МІЖДИСЦИПЛІНАРНІ АСПЕКТИ ПРИРОДОМОВНИХ ДОСЛІДЖЕНЬ ТА ПЕРСПЕКТИВ ДОКТОРСЬКОЇ ОСВІТИ

У цій статті розглядаються міждисциплінарні питання природномовних досліджень та перспективи докторської освіти в контексті тріади «Інтелектуалізація – Віртуалізація – Великі дані». Останні технологічні досягнення значно сприяли прогресу у сфері обробки природної мови (NLP). Сучасне NLP значною мірою покладається на методи машинного та глибинного навчання, що дозволяє комп’ютерам навчатися на великих наборах даних і виконувати завдання з високою точністю.

Основні аспекти, які висвітлені в статті, включають використання нейронних мереж, механізми уваги та трансферного навчання. Обговорюються головні виклики, з якими стикається сучасне NLP, такі як неоднозначність, обмеженість даних, відсутність контексту та етичні проблеми. Важливими напрямками досліджень є мультимодальність, пояснювальні нейронні мережі та розвиток систем NLP для мов з обмеженими ресурсами.

Стаття акцентує увагу на технологічному статусі NLP та його значущості в умовах інтелектуалізації, віртуалізації та аналізу великих даних. Визначені можливі напрямки досліджень, такі як покращення інтелектуалізації систем NLP, створення віртуальних середовищ для NLP-додатків та використання аналітики великих даних для підвищення точності та ефективності систем NLP.

Крім того, у статті розглядається навчання докторантів в умовах використання NLP та аналізу великих даних, акцентуючи увагу на розробці компетенцій в управлінні дослідницькими даними та методологічному капіталі освітньої науки даних. У підсумках підкреслюється важливість інтеграції етичних аспектів у використання технологій NLP та їх значення для сучасної докторської освіти в Україні.

У дослідженні підкреслено критичну необхідність комплексного підходу до аналізу мови, віртуальних середовищ та великих даних для розвитку систем NLP та модернізації докторської освіти.

Ключові слова: обробка природної мови, великі дані, докторська освіта, підготовка аспірантів, інтелектуалізація, віртуалізація.