ISSN 2078-5119. Теорія і практика викладання української мови як іноземної. 2024. Випуск 18. С. 180–191
Theory and Practice of Teaching Ukrainian as a Foreign Language. 2024. Issue 18. P. 180–191

УДК 811.161'32]:004.032.26:004.055:004.423.46

REVIEW OF THE SOURCE BASE OF THE STUDY OF NEURAL NETWORK MODELING OF LINGUISTIC UNITS

Oleksii Dovhan

Mykhailo Drahomanov Ukrainian State University
Department of Slavic, Romance, and Oriental Languages
8-14 Oleksandra Konyskoho Str., 14th floor, room 14-12, 01601, Kyiv, Ukraine
phone: 044 486 44 65
e-mail: oleksiivdovhan@gmail.com
http://orcid.org/0000-0002-6728-818X

The article highlights the peculiarities of the source base of neural network modeling research, which consists of the differentiation of the modern concept of 'source', as well as transformational changes in the functional role of the latter in modern (in particular, humanitarian and linguistic) research. The author emphasizes the productivity of dividing the above-mentioned base into: a) legal documents (the Constitution of Ukraine, laws, orders, regulations, concepts), which allow to identify the core principles of regulation of the analyzed issues and its specific features; b) analytical and reporting documents (reports, plans, etc.), which are representative of the formal indicators of achievement of certain tasks, deadlines, compliance with the work algorithm, etc; c) mass media (Internet platforms, portals, social networks, scientific blogs, websites, telegram channels, etc.) that illustrate socio-cultural changes in the linguistic landscape and can be productive for studying various regional changes (languages of national minorities, studies of certain age groups, etc.) and comprehending changes in the linguistic polysystem, contributing to a deeper understanding of the correlation between the components of the problem under study; d) artificial neural networks (ChatGPT (3.5 and 4), DALL E2, Gemini, LeiaPix, Narakeet, Pika, Runway, Sora, etc.) – a controversial group of sources, which is produced by the dual nature of the latter: they are both a tool and a source of research on neural network modeling of linguistic units; by their nature, they are one of the most representative types, since their actualization produces the practical implementation of certain hypotheses of the researcher with the comprehension of the proper parameterization of their functioning. Thus, the actualization of a diverse range of sources produces objectivity and completeness of modern humanitarian (in particular, linguistic) research, allowing a comprehensive analysis of the studied issues and ensuring the reproducibility, knowledge intensity, modernity, etc. of the scientific work carried out.

Key words: neural network modeling, neural network modeling research, applied linguistics, computational linguistics, artificial neural networks, neural network models, sources of modern research.

DOI:

Statement of the problem in general terms and its connection with important scientific or practical tasks. Today, the dynamics of digitalization processes have become exponential, the essence of which is the widespread use of the latest technologies (from recommendations to YouTube videos to predicting the success of surgical interventions), which naturally has affected scientific activity. Thus, the approach to conducting certain humanities research has changed, more and more integrated, interdisciplinary, etc. works have appeared, and the integration of the mathematical paradigm into humanities research has become unprecedented. In particular, modern scientific research (including linguistic research) is increasingly moving away from professional limitations, acquiring the characteristics of digital humanities.

The above has undoubtedly influenced the dynamics, topics, structure, methodology (including tools), etc. of such humanitarian (including linguistic) studies, while emphasizing the immutability of the basis for their actualization. We are talking about the historiography and sources of a particular study, which have become increasingly impor-

tant in the context of digitalization processes. Thus, based on the specifics of this study, we will take a closer look at the sources, since they represent the analyzed issues and become its cornerstone.

We are referring to the transformational changes that have taken place in the specifics and role of the above-mentioned concepts in the context of constructing a particular study. Undoubtedly, they (historiography and sources) have remained important, but in line with current trends in the analysis, processing, representation, etc. of scientific data (let us recall data science, scientometrics, etc.), the date of publication of a particular study has become the pivotal one. Accordingly, if it was published recently and analyzes the relevant trends in the development of digitalization tools for linguistic research, its relevance will be higher than that of the same study published five or ten years ago. This is because the latter will no longer be representative of the current state of the above-mentioned problem but of the originality of its genesis and changes in the understanding of the role of the latter.

At the same time, the above-mentioned digitalization processes are unstable, which naturally affects the actualization of certain innovative practices. First of all, we are talking about artificial neural networks and neural network modeling of linguistic units, which is made possible by their work. This instability produces a parallelism in the functioning of traditional sources of data acquisition, processing, analysis, and generation with a rather rare actualization of their innovative counterparts. In the above context, neural network modeling should be positioned as a manifestation of such processes, which allows optimizing modern linguistic research. It is noteworthy that they (these studies) are mostly focused on natural language processing (NLP): the latter, in turn, correlates with linguistic analysis in the broadest sense, which is based on the source base mentioned above. We are talking about the actualization of the range from syntactic localization to machine translation and text generation by neural network models [30], which is why it is important to study the source base of neural network modeling of linguistic units.

Analysis of recent research and publications that have initiated the solution of this problem and on which the author relies. Thus, the problem of the originality of the role of the source base for the study of neural network modeling of linguistic units is quite complex formally (the variety of types, types, etc. of sources and the specifics of the stages of actualization in modern linguistic research) and functionally (we are talking about the role, selection, ranking by importance, etc. related to predictive power: architecture, type, features of the layers of the neural network model, etc.). The above naturally leads to the multiplicity of studies of the issues under consideration, producing a multi-interpretation of the presentation of such a source base. Therefore, the main feature of the source base of the study of neural network modeling of linguistic units is its discursive nature, which is the specificity of the analyzed problem. The latter is related to the need for a specific actualization of the studied issue of linguistic analysis, taking into account machine and deep learning, data science, etc.

The specifics of the evaluation of recurrent neural networks (RNNs) in the context of the functioning of edge intelligence applications and human activity recognition are studied by V. Lalapura et al. [25]. In the analyzed research, the authors study the architecture features and efficiency of edge-intelligent RNNs (EI-RNNs) to achieve higher performance. Scientists note that they have evaluated and optimized six different recurrent units, viz: Vanilla RNNs, Long Short-Term Memory (LSTM) units, Gated Recurrent Units (GRUs), Fast Gated Recurrent Neural Networks (FGRNNs), Fast Recurrent Neural Networks (FRNNs), and Unitary Gated Recurrent Neural Networks (UGRNNs) on eight publicly available time series of recognition

data. The researchers found the relevance of efficient training (dataset processing and preprocessing procedures, hyperparameter tuning, etc.) and appropriate compression methods (e.g., low-rank parameterization and interactive clipping) that correlate with optimizing the performance of an artificial neural network model in the context of RNN memory usage.

The study of text sentiment analysis as a component of NLP is highlighted in the research of W. Yu et al. [26], in which the authors analyze quantum machine learning models (in particular, the originality of using the property of qubit superposition to increase the expressiveness of the model and achieve faster computation compared to classical systems) of South Asian languages (for example, Bengali), which are characterized by flexible word order and high inflectional capacity. The scientists see the solution to the analyzed problem in the use of a quantum-classical neural network model with a hybrid structure developed by them. It is about updating a pre-trained multilingual Bidirectional Encoder Representations from Transformers (BERT) model for word vectorization with Batched Upload Quantum Recurrent Neural Network (BUQRNN) and Parameter Non-shared Batched Upload Quantum Recurrent Neural Network (PN-BUQRNN) as models for feature extraction and further sentiment analysis in Bengali.

The problem of human speech recognition by artificial neural networks was studied by C. Brodbeck, T. Hannagan, J. Magnuson [27], where the authors assert the performance of RNNs for such activities. The latter, according to the scientists, receives input data block by block and can imitate several core behavioral features of human speech, but they doubt the ability of RNNs to take into account the dynamics of such speech (actually, neural processing). The researchers emphasize that the internal dynamics of the long short-term memory (LSTM) of RNNs trained to recognize speech from auditory spectrograms predict the responses of human neurons to the same incentives, even though they do not depend on auditory features. The authors note that variations in the RNN architecture are motivated by cognitive principles that improve this predictive ability more. Moreover, the different components of hierarchical RNNs predict individual components of the brain's responses to speech in an anatomically structured way, suggesting that such neural networks reproduce the hierarchy of speech recognition in the brain.

The dissertation of Bulgarian researcher T. Mihaylov [28] is devoted to the peculiarities of machine reading comprehension, a task of understanding a language poly system, in which an artificial neural network has to read a certain passage of text and, as a rule, answer control questions about its content. In the analyzed study, the author explores ways to solve the problem of machine reading comprehension based on an anthropocentric approach, which is characterized by a human assessment of reading comprehension tasks. This is about updating background knowledge, common sense, language skills, and, in fact, understanding, sequence, and logic of data presentation in the text, etc. The scientist notes that, in contrast to the above, existing updated neural network approaches to solving the problem are focused on training end-to-end systems based on annotated data that correspond to a local task and are not suitable for generalization, concluding certain trends, etc.

The researcher emphasizes that his study updates the above-mentioned anthropological approach using knowledge from various linguistic tasks, which, in turn, is incorporated into the artificial neural network model under development. The latter, according to the author, should integrate external common sense knowledge encoded in the form of key-value memory. The scientist argues that this approach allows us to go beyond the document-query system or discrete features, emphasizing several external

knowledge in the work of the artificial neural network model and combining them with a representation of the context before generating the result. This approach, according to the researcher, is productive in attracting knowledge from an external data source that is explicit in the textual array, but which is relevant to the formation of the result.

In addition, T. Mihaylov studied the dynamics of transferring linguistic knowledge from low-level tasks in the reading comprehension system using neural representations. According to the author, the data transferred in this way, subject to prior training on linguistic tasks, is productively adapted and combined to improve the machine reading process at an early stage of training and during training with small portions of data. In addition, the researcher proposes to update the structured linguistic annotations as the basis for a discourse-aware semantic self-attention coder, which the researcher used for narrative texts. This approach has also been productive in isolating the connections between discourse units, events, and their arguments, and references that relate to each other, updating the existing annotation tools.

The research on the problem of automated essay evaluation is devoted to the study of J. Lohmann et al. [29] in which the authors note the productivity of using hybrid neural network models. The scientists explain the representativeness of the latter by the fact that their structure combines functional engineering and powerful tools of deep neural networks (DNNs), which ensures speed and quality of work. The researchers conducted a series of experiments using three variants of neural network models that update different input data: a) models based on linguistic features, b) models that use contextual inserts at the essay level from the distilled version of the pre-trained BERT transformer (DistilBERT), and c) a hybrid model where both types of features are updated. The results show that model A outperforms model B, but both are outperformed by model C.

Thus, the analysis of the historiography on the originality of the role of the source base for the study of neural network modeling of linguistic units has shown the fragmentary nature of information presentation. Thus, despite the existence of several studies on certain aspects of the problem, no study would cover the specifics of the source base not from the standpoint of updated materials (in fact, the basis for neural network modeling), but rather in terms of organization, methodology, certain aspects of updating, etc. That is why this study aims to fill the existing gap, namely, to highlight the peculiarities of updating the source base of the study of neural network modeling of linguistic units.

Identification of previously unresolved parts of the general problem to which this article is devoted. The aforementioned formulation of the problem and the analysis of recent studies have shown the relevance of the issue under consideration, which is also determined by a certain blurring of the category and conceptual apparatus. Thus, most source studies position the term "source" as a certain construct reflecting the genesis of human society, i.e., everything created in the process of human activity and representing information about the analyzed issue. That is why the lack of specificity in understanding the concept of source produces several unresolved and unexplained parts of the general problem of the source base of modern humanities research (in particular, linguistic research).

The point is that for classical linguistics research, such a problem simply does not exist, since everything is regulated, while for research in the context of the digital humanities, the problem becomes significant. For example, the study of neural network modeling of language unit recognition is based on the digital humanities approach. Thus, the structure, methodology, and experiments of such a study are built not only within the framework of the conventional areas (applied, computer, etc. linguistics) but also

within, for example, data science, computational linguistics, etc. In turn, the above causes several fluctuations in the conceptual system, within which we believe it is necessary to reconsider the basis of modern scientific humanitarian research (in particular, linguistic research), which is, in fact, the unresolved and unexplained part of the analyzed problem.

Formation of the article's objectives (statement of the task). *The purpose* of the article is to consider the originality of the source base of modern scientific humanitarian research. *The subject* is the specificity of such a base in the context of neural network modeling as an innovative tool of linguistic science.

Presentation of the main research material with full justification of the scientific results obtained. The above naturally allows us to position the source base of neural network modeling research as a basis for highlighting and analyzing the transformational changes in modern linguistic research. The latter carried out within the framework of the digital humanities (applied, computer, computational, etc. linguistics), are characterized by the understanding of words, phrases, sentences, paragraphs, texts, etc. as, first of all, data sets, the analysis, processing, representation, etc. of the results of which is carried out within the framework of data science or similar sciences. This approach is productive because it is representative of the pragmatic foundations of linguistic analysis in terms of information processing as its basis. At the same time, within the framework of this approach, the specified source base plays a special role as a cornerstone of neural network modeling of linguistic units. Thus, to study its features in different discourses (in particular, political Internet discourse), we have actualized a wide array of sources, which are divided into the following groups:

1. *Legal documents (LDs)* are a group of documents (the Constitution of Ukraine, laws, orders, regulations, concepts, etc.) that are representative of the originality of the legal framework for the existence of digitalization processes in general (in particular, the innovative tools of modern scientific (including humanities and linguistic) research). This group is characterized by the following features: a) *hierarchy:* produced by the different status or importance of a particular type of source, consistency in terms of levels of influence, within which one document has a higher status or area of influence than another: thus, there may be documents that are valid only in a certain territory (for example, in the Autonomous Republic of Crimea or Europe), etc; b) *openness or accessibility:* represented by the nature of the legal framework, which is freely available on several official resources (for example, the regulatory bulletin "Vidomosti Verkhovna Rada of Ukraine", the portal of the Verkhovna Rada of Ukraine, etc.); c) *temporal dynamics of changes:* expressed in the systematic updating of the legal framework (amendments to laws, their expiration, etc.).

Therefore, the legal basis for the functioning and actualization of digitalization processes in our country is naturally determined by the provisions of the Constitution of Ukraine [3], several laws regulating certain aspects of related (the Law of Ukraine "On Ensuring the Functioning of the Ukrainian Language as the State Language" [7]) and information (the Laws of Ukraine "On Information" [11], "On Electronic Communications" [6], "On Protection of Information in Information and Telecommunication Systems" [9], "On Protection of Personal Data" [10] and "On Media" [12]) activities. Several regulatory documents relating to the functioning, work plan, and reporting of the Ministry of Digital Transformation of Ukraine can be separately distinguished (for example, "The Regulation on the Ministry of Digital Transformation of Ukraine" [5], "The Report on the Implementation of the Work Plan of the Ministry of Digital Transformation of Ukraine for 2023" [2]).

Naturally related to the activities of the relevant Ministry are several legal acts that regulate general issues of digitalization in Ukraine (some orders of the Cabinet of Ministers of Ukraine: "Some Issues of Digital Transformation" [1], "On Approval of the Action Plan for the Implementation of the Concept of Artificial Intelligence Development in Ukraine for 2021-2024" [8], "On Approval of the Concept of Artificial Intelligence Development in Ukraine" [15], "On Approval of the Concept for the Development of Digital Competencies and Approval of the Action Plan for its Implementation" [13], "On Approval of the Concept for the Development of the Digital Economy and Society of Ukraine for 2018-2020 and Approval of the Action Plan for its Implementation" [14], etc.) and Europe ("The Digital Europe Program" [24]).

We emphasize that the above-mentioned legal framework contains the main documents and is representative of the implementation of institutional reforms and ensuring the development of digitalization processes, the Ukrainian language as the state language, and scientific activity in general. Thus, the analysis of the above laws, regulations, concepts, orders, etc. has shown the relevant trends in the legal development of the Ukrainian state and has outlined the legal context for the use of artificial neural networks (in particular, neural network modeling of linguistic units). In addition, the actualization of this type of source is representative of the degree of compliance of the aforementioned context with international standards and norms, which made it possible to speak about the prospects for the development of this area.

2. *Analytical and reporting documents (ARDs)* are the core for mastering the specifics, transformational changes, etc. of modern linguistic research (in particular, neural network modeling of linguistic units). This group includes: *a) statistical data:* indicators of the implementation of a particular study, research topic, grant, etc., illustrations of the above, which are produced by the authors in the process of testing, practical implementation of certain hypotheses; *b) analytical reviews, plans, reports, etc.* – a cyclically created group of documents that represents formal indicators of the effectiveness of work (in particular, scientific) within the forms mentioned in paragraph A (for example, the report on the implementation of the plan [4] of the Ministry of Digital Transformation of Ukraine for 2023 [2], scientometric indicators of a researcher's activity, success in meeting the deadlines for the planned stage of research, etc.), are differentiated by their nature (can be in the form of tables, texts, graphs, diagrams, etc;) c) *the results several scientific studies*, represented by an integrated historiography of neural network modeling of linguistic units, which highlights certain aspects of the problem under study.

3. *Mass media* is a group representative of the realities of modern discourse (in particular, Internet discourse): its actualization allows us to track new trends in the information space, record several sociocultural phenomena (for example, neologization), etc. with the identification of possible correlations with the functioning of scientific innovation trends. In the context of studying neural network modeling of linguistic units, it is advisable to use Internet platforms (for example, the project of the Ministry of Education and Science of Ukraine – "Science and Business" (https://goo.su/wBSzKnR), etc.), portals (for example, "Verkhovna Rada of Ukraine" (https://goo.su/NIaiM), "Government Portal" (https://goo.su/iHycV)), social networks (for example, Facebook, X, BlueSky, etc.), scientific blogs (from some blogs on Google's Blogger to: IFL SCIENCE (https://goo.su/vAZSMUO), Nautilus (https://goo.su/Kw9o6i), PLOS (https://goo.su/58mc2UB), and others), websites (for example, KUNST (https://goo.su/nvPr), "My Science" (https://goo.

su/00cHu), etc.), telegram channels (for example, "Laboratory Mouse" (https://goo.su/d111), "Neural Networks UA: IT, Technology, AI" (https://goo.su/dx26dU7), etc.) and others that publish (or broadcast online) news, analytical studies, interviews, etc.

The specificity of the above group of documents is the relevance of the data presented in them (promptness of information provision), the variety of its types (text, audio, video, photo, or integration of different formats in one product or on one resource), quality and reliability of information (first of all, we are talking about media that are parametrically consistent with the definition given in the relevant Law of Ukraine [12]) and influence on public opinion (popularization of certain ideas, trends or, conversely, devaluation of prevailing trends and tendencies).

4. *Artificial neural networks (ANNs)* are a controversial group of sources because, on the one hand, they are tools for work, and, on the other hand, they are a source in their own right. This is because, in the case of neural network modeling of linguistic units, it is the peculiarities of their operation and work with them that become the cornerstone of modern linguistic research. In a certain sense, this type is the most informative of all, since it is thanks to their actualization that the researcher can study the specifics of the latter's actualization (data generation: ChatGPT (3.5 and 4) [16]; Gemini [18]), specific learning principles (machine learning: ChatGPT (3.5 and 4) [16], DALL E2 [17], Gemini [18], LeiaPix [19], Narakeet [20], Pika [21], Runway [22], Sora [23], etc.), correlation of the above results and their algorithm [ChatGPT (3.5 and 4) [16], DALL E2 [17], Gemini [18], LeiaPix [19], Narakeet [20], Pika [21], Runway [22], Sora [23], etc.), which are indicative in the context of the problem under study. The latter is because it is when using artificial neural networks, choosing a specific combination of layers within them, selecting an original training sample for their training, etc. that a researcher comes to certain conclusions, testing his or her scientific hypotheses.

It is noteworthy that artificial neural networks are fundamentally and functionally open as a source of research on neural network modeling of linguistic units. Thus, the above-mentioned productivity of using recurrent, graph, etc. neural networks for linguistic research is noted, and large linguistic models (LLMs) are also commonly used, but this does not mean that this approach is productive. The point is to individualize the proper tools of modern humanitarian (in particular, linguistic) research, which is produced by the development of technologies that are the basis of neural network modeling. This results in constant transformational changes that produce the necessary flexibility: if one tool is productive today, tomorrow it may not be very successful for a particular task. The latter is evidenced by the emergence in February 2024 of the Sora neural network model [23], which is ostensibly designed for animating images and creating high-quality videos, but whose algorithm for working with textual input deserves the attention of modern linguists. In turn, this situation shapes the understanding of the tools of modern (primarily humanitarian) research as an unstable set of tools.

Naturally, the above leads to a certain blurring of the boundaries between the proper tools of modern humanitarian (in particular, linguistic) research, its historiography (the specifics of the data corpora involved), and the source base (in fact, we are talking about a certain uncertainty of the latter), which is well illustrated by artificial neural networks as a means and source of research data), as well as subject correlation (the increasing degree of integration of the mathematical paradigm into specifically humanitarian research produces an increasingly thin line between humanitarian and technical research, as evidenced by the emergence of computational linguistics, etc.) Instead, a modern researcher (regardless of

his or her specialty) needs to constantly update and improve existing knowledge, skills, and abilities (for example, media literacy, data analysis, cybersecurity, etc.) by tracking current trends in working with artificial intelligence (from software (from browser extensions to several smartphone applications) to hardware) and actualizing such knowledge in integrated humanities research within the digital humanities.

Conclusions from this research and prospects for further research in this area. Thus, the problem of the source base of the study of neural network modeling of linguistic units lies in the differentiation of the modern concept of 'source', as well as in the transformational changes in the functional role of the latter in modern (in particular, humanitarian and linguistic) research. The sources of such works should be divided into: *a) legal documents* (the Constitution of Ukraine, laws, orders, regulations, concepts), which allow us to identify the core principles of regulation of the analyzed issues and its specific features; *b) analytical and reporting documents* (reports, plans, etc.), which are representative of the formal indicators of achievement of certain tasks, deadlines, compliance with the work algorithm, etc; *c) mass media* (Internet platforms, portals, social networks, scientific blogs, websites, telegram channels, etc.) that illustrate socio-cultural changes in the linguistic landscape and can be productive for studying various regional changes (languages of national minorities, studies of certain age groups, etc.) and comprehending changes in the linguistic poly system, contributing to a deeper understanding of the correlation between the components of the problem under study; *d) artificial neural networks* (ChatGPT (3.5 and 4), DALL E2, Gemini, LeiaPix, Narakeet, Pika, Runway, Sora, etc.) – a controversial group of sources, which is produced by the dual nature of the latter: they are both a tool and a source of research on neural network modeling of linguistic units; by their nature, they are one of the most representative types, since their actualization produces the practical implementation of certain hypotheses of the researcher with the comprehension of the proper parameterization of their functioning. Thus, the actualization of a diverse range of sources produces objectivity and completeness of modern humanitarian (in particular, linguistic) research, allowing for a comprehensive analysis of the studied issues and ensuring reproducibility, knowledge intensity, modernity, etc. of the scientific work.

Список використаної літератури

Джерела:

1. Деякі питання цифрової трансформації : розпорядження Кабінету Міністрів України від 17 лютого 2021 р. № 365-р. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/P69Z (дата звернення: 31.03.24).

2. Звіт про виконання плану роботи Міністерства цифрової трансформації України на 2023 рік : наказ Міністерства цифрової трансформації України від 15 лютого 2024 р. № 26. *Міністерство цифрової трансформації України* : вебсайт. URL : https://goo.su/pEcR6l (дата звернення: 31.03.24).

3. Конституція України. *Президент України* : офіційне інтернет-представництво. URL : https://goo.su/KMoTyl (дата звернення: 31.03.24).

4. План роботи Міністерства цифрової трансформації України на 2023 рік : наказ Міністерства цифрової трансформації України від 16 травня 2023 р. № 56. *Міністерство цифрової трансформації України* : вебсайт. URL : https://goo.su/hJ3G (дата звернення: 31.03.24).

5. Положення про Міністерство цифрової трансформації України : постанова від 18 вересня 2019 р. № 856. *Урядовий портал* : єдиний вебпортал органів виконавчої влади України. URL : https://goo.su/Kdct (дата звернення: 31.03.24).

6. Про електронні комунікації : Закон України від 16 грудня 2020 р. № 1089-IX. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/Z7lmdcT (дата звернення: 31.03.24).

7. Про забезпечення функціонування української мови як державної : Закон України від 25 квітня 2019 р. № 2704-VII. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/RKsNw (дата звернення: 31.03.24).

8. Про затвердження плану заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2021-2024 роки : розпорядження Кабінету Міністрів України від 12 травня 2021 р. № 438-р. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/y7zULXy (дата звернення: 31.03.24).

9. Про захист інформації в інформаційно-телекомунікаційних системах : Закон України від 05 липня 1994 р. № 80/94-ВР. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/Sg1Ez5 (дата звернення: 31.03.24).

10. Про захист персональних даних : Закон України від 01 червня 2010 р. № 2297-VI. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/I31on8J (дата звернення: 31.03.24).

11. Про інформацію : Закон України від 2 жовтня 1992 р. № 2657-XII. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/uv6ic1O (дата звернення: 31.03.24).

12. Про медіа : Закон України від 13 грудня 2022 р. № 2849-IX. *Верховна Рада України :* офіційний вебпортал парламенту України. URL : https://goo.su/lSP5y (дата звернення: 31.03.24).

13. Про схвалення Концепції розвитку цифрових компетентностей та затвердження плану заходів з її реалізації : розпорядження Кабінету Міністрів України від 03 березня 2021 р. № 167-р. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/lhmjQVY (дата звернення: 31.03.24).

14. Про схвалення Концепції розвитку цифрової економіки та суспільства України на 2018-2020 роки та затвердження плану заходів щодо її реалізації : розпорядження Кабінету Міністрів України від 17 січня 2018 р. № 67-р. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/c8iI (дата звернення: 31.03.24).

15. Про схвалення Концепції розвитку штучного інтелекту в Україні : розпорядження Кабінету Міністрів України від 02 грудня 2020 р. № 1556-р. *Верховна Рада України* : офіційний вебпортал парламенту України. URL : https://goo.su/ZWClB6 (дата звернення: 31.03.24).

16. ChatGPT. *OpenAI* : website. 2024. URL : https://goo.su/IHf5S (date of application: 31.03.24).

17. DALL E2. *OpenAI* : website. 2024. URL : https://goo.su/EdjJMG2 (date of application: 31.03.24).

18. Gemini. *Gemini* : website. 2024. URL : https://goo.su/pkW1vK (date of application: 31.03.24).

19. LeiaPix. *LeiaPix* : website. 2024. URL : https://goo.su/Ibj6bD (date of application: 31.03.24).

20. Narakeet. *Narakeet* : website. 2024. URL : https://goo.su/WzRB (date of application: 31.03.24).

21. Pika. *Pika* : website. 2024. URL : https://goo.su/eRrORM (date of application: 31.03.24).

22. Runway. *Runway* : website. 2024. URL : https://goo.su/xYz3RPl (date of application: 31.03.24).

23. Sora. *OpenAI* : website. 2024. URL : https://goo.su/3Rm1x (date of application: 31.03.24).

24. The Digital Europe Programme. *European Commission* : website. URL : https://goo.su/Gdfyz (date of application: 31.03.24).

Література:

25. A Systematic Evaluation of Recurrent Neural Network Models for Edge Intelligence

and Human Activity Recognition Applications / V. S. Lalapura et al. *Algorithms*. 2024. Volume 17. Issue 3. https://doi.org/10.3390/a17030104 *MDPI* : website. URL : https://goo.su/zIP0bZ (date of application: 31.03.24).

26. Application of Quantum Recurrent Neural Network in Low Resource Language Text Classification / W. Yu et al. *IEEE Transactions on Quantum Engineering*. 2024. https://doi.org/10.1109/TQE.2024.3373903 *IEEE Xplore* : website. URL : https://goo.su/zgtJjc (date of application: 31.03.24).

27. Brodbeck C., Hannagan T., Magnuson J. S. Recurrent neural networks as neurocomputational models of human speech recognition. https://doi.org/10.1101/2024.02.20.580731 *bioRxiv* : website. 2024. URL : https://goo.su/ziCc (date of application: 31.03.24).

28. Mihaylov T. B. Knowledge-Enhanced Neural Networks for Machine Reading Comprehension. https://doi.org/10.11588/heidok.00034352 *heiDOK* : website. 2024. URL : https://goo.su/E6yXm (date of application: 31.03.24).

29. Neural Networks or Linguistic Features?-Comparing Different Machine-Learning Approaches for Automated Assessment of Text Quality Traits Among L1-and L2-Learners' Argumentative Essays / J. F. Lohmann et al. https://doi.org/10.21203/rs.3.rs-3979182/v1 *Research Square* : website. 2024. URL : https://goo.su/qraxW8 (date of application: 31.03.24).

30. Universal Neurons in GPT2 Language Models / W. Gurnee et al. *arXiv preprint arXiv:2401.12181*. 2024. URL : https://goo.su/hT2WOW (date of application: 31.03.24).

Sourses and references

Sources:

1. Deiaki pytannia tsyfrovoi transformatsii : rozporiadzhennia Kabinetu Ministriv Ukrainy vid 17 liutoho 2021 r. № 365-r. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/P69Z (data zvernennia: 31.03.24).

2. Zvit pro vykonannia planu roboty Ministerstva tsyfrovoi transformatsii Ukrainy na 2023 rik : nakaz Ministerstva tsyfrovoi transformatsii Ukrainy vid 15 liutoho 2024 r. № 26. *Ministerstvo tsyfrovoi transformatsii Ukrainy* : vebsait. URL : https://goo.su/pEcR6l (data zvernennia: 31.03.24).

3. Konstytutsiia Ukrainy. *Prezydent Ukrainy* : ofitsiine internet predstavnytstvo. URL : https://goo.su/KMoTyl (data zvernennia: 31.03.24).

4. Plan roboty Ministerstva tsyfrovoi transformatsii Ukrainy na 2023 rik : nakaz Ministerstva tsyfrovoi transformatsii Ukrainy vid 16 travnia 2023 r. № 56. *Ministerstvo tsyfrovoi transformatsii Ukrainy* : vebsait. URL : https://goo.su/hJ3G (data zvernennia: 31.03.24).

5. Polozhennia pro Ministerstvo tsyfrovoi transformatsii Ukrainy : postanova vid 18 veresnia 2019 r. № 856. *Uriadovyi portal* : yedynyi vebportal orhaniv vykonavchoi vlady Ukrainy. URL : https://goo.su/Kdct (data zvernennia: 31.03.24).

6. Pro elektronni komunikatsii : Zakon Ukrainy vid 16 hrudnia 2020 r. № 1089-IX. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/Z7lmdcT (data zvernennia: 31.03.24).

7. Pro zabezpechennia funktsionuvannia ukrainskoi movy yak derzhavnoi : Zakon Ukrainy vid 25 kvitnia 2019 r. № 2704-VII. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/RKsNw (data zvernennia: 31.03.24).

8. Pro zatverdzhennia planu zakhodiv z realizatsii Kontseptsii rozvytku shtuchnoho intelektu v Ukraini na 2021-2024 roky : rozporiadzhennia Kabinetu Ministriv Ukrainy vid 12 travnia 2021 r. № 438-r. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/y7zULXy (data zvernennia: 31.03.24).

9. Pro zakhyst informatsii v informatsiino-telekomunikatsiinykh systemakh : Zakon Ukrainy vid 05 lypnia 1994 r. № 80/94-VR. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/Sg1Ez5 (data zvernennia: 31.03.24).

10. Pro zakhyst personalnykh danykh : Zakon Ukrainy vid 01 chervnia 2010 r. № 2297-VI. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/I31on8J (data zvernennia: 31.03.24).

11. Pro informatsiiu : Zakon Ukrainy vid 2 zhovtnia 1992 r. № 2657-KhII. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/uv6ic1O (data zvernennia: 31.03.24).

12. Pro media : Zakon Ukrainy vid 13 hrudnia 2022 r. № 2849-IKh. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/lSP5y (data zvernennia: 31.03.24).

13. Pro skhvalennia Kontseptsii rozvytku tsyfrovykh kompetentnostei ta zatverdzhennia planu zakhodiv z yii realizatsii : rozporiadzhennia Kabinetu Ministriv Ukrainy vid 03 bereznia 2021 r. № 167-r. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/lhmjQVY (data zvernennia: 31.03.24).

14. Pro skhvalennia Kontseptsii rozvytku tsyfrovoi ekonomiky ta suspilstva Ukrainy na 2018-2020 roky ta zatverdzhennia planu zakhodiv shchodo yii realizatsii : rozporiadzhennia Kabinetu Ministriv Ukrainy vid 17 sichnia 2018 r. № 67-r. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/c8iI (data zvernennia: 31.03.24).

15. Pro skhvalennia Kontseptsii rozvytku shtuchnoho intelektu v Ukraini : rozporiadzhennia Kabinetu Ministriv Ukrainy vid 02 hrudnia 2020 r. № 1556-r. *Verkhovna Rada Ukrainy* : ofitsiinyi vebportal parlamentu Ukrainy. URL : https://goo.su/ZWClB6 (data zvernennia: 31.03.24).

16. ChatGPT. *OpenAI* : website. 2024. URL : https://goo.su/IHf5S (date of application: 31.03.24).

17. DALL E2. *OpenAI* : website. 2024. URL : https://goo.su/EdjJMG2 (date of application: 31.03.24).

18. Gemini. *Gemini* : website. 2024. URL : https://goo.su/pkW1vK (date of application: 31.03.24).

19. LeiaPix. *LeiaPix* : website. 2024. URL : https://goo.su/Ibj6bD (date of application: 31.03.24).

20. Narakeet. *Narakeet* : website. 2024. URL : https://goo.su/WzRB (date of application: 31.03.24).

21. Pika. *Pika* : website. 2024. URL: https://goo.su/eRrORM (date of application: 31.03.24).

22. Runway. *Runway* : website. 2024. URL : https://goo.su/xYz3RPl (date of application: 31.03.24).

23. Sora. *OpenAI* : website. 2024. URL: https://goo.su/3Rm1x (date of application: 31.03.24).

24. The Digital Europe Programme. *European Commission* : website. URL : https://goo.su/Gdfyz (date of application: 31.03.24).

References:

25. A Systematic Evaluation of Recurrent Neural Network Models for Edge Intelligence and Human Activity Recognition Applications / V. S. Lalapura et al. *Algorithms*. 2024. Volume 17. Issue 3. https://doi.org/10.3390/a17030104 *MDPI* : website. URL : https://goo.su/zIP0bZ (date of application: 31.03.24).

26. Application of Quantum Recurrent Neural Network in Low Resource Language Text Classification / W. Yu et al. *IEEE Transactions on Quantum Engineering*. 2024. https://doi.org/10.1109/TQE.2024.3373903 *IEEE Xplore* : website. URL : https://goo.su/zgtJjc (date of application: 31.03.24).

27. Brodbeck C., Hannagan T., Magnuson J. S. Recurrent neural networks as neuro-computational models of human speech recognition. https://doi.org/10.1101/2024.02.20.580731 *bioRxiv* : website. 2024. URL : https://goo.su/ziCc (date of application: 31.03.24).

28. Mihaylov T. B. Knowledge-Enhanced Neural Networks for Machine Reading Comprehension. https://doi.org/10.11588/heidok.00034352 *heiDOK* : website. 2024. URL : https://goo.su/E6yXm (date of application: 31.03.24).

29. Neural Networks or Linguistic Features?-Comparing Different Machine-Learning Approaches for Automated Assessment of Text Quality Traits Among L1-and L2-Learners' Argumentative Essays / J. F. Lohmann et al. https://doi.org/10.21203/rs.3.rs-3979182/v1 *Research Square* : website. 2024. URL : https://goo.su/qraxW8 (date of application: 31.03.24).

30. Universal Neurons in GPT2 Language Models / W. Gurnee et al. *arXiv preprint arXiv:2401.12181*. 2024. URL : https://goo.su/hT2WOW (date of application: 31.03.24).

ОГЛЯД ДЖЕРЕЛЬНОЇ БАЗИ ДОСЛІДЖЕННЯ НЕЙРОМЕРЕЖЕВОГО МОДЕЛЮВАННЯ МОВНИХ ОДИНИЦЬ

Олексій Довгань

Український державний університет імені Михайла Драгоманова
кафедра слов'янських, романських і східних мов
вул. Олександра Кониського, 8-14, 14 поверх, кімн. 14-12, 01601, Київ, Україна
тел.: 044 486 44 65
ел. пошта: oleksiivdovhan@gmail.com
http://orcid.org/0000-0002-6728-818X

У статті висвітлено особливості джерельної бази дослідження нейромережевого моделювання, яке полягає у диференційованості сучасного поняття 'джерела', а також трансформаційних змінах функційної ролі останнього в сучасних (зокрема, гуманітарних та лінгвістичних) дослідженнях. Підкреслено продуктивність розділення вищезазначеної бази на: а) *нормативно-правові документи* (Конституція України, закони, накази, положення, концепції), які дозволяють виокремити стрижневі принципи регулювання аналізованої проблематики та її специфічні риси; б) *аналітично-звітні документи* (звіти, плани тощо), що репрезентативні щодо формальних показників досягнення тих чи тих завдань, термінів виконання, дотримання алгоритму роботи і такого іншого; в) *масмедіа* (інтернет-платформи, портали, соціальні мережі, наукові блоги, вебсайти, телеграм-канали тощо), які унаочнюють соціокультурні зміни мовного ландшафту й можуть бути продуктивними для дослідження різноманітних регіональних змін (мов національних меншин, дослідження певних вікових груп та інших) та осягнення змін мовної полісистеми, сприяючи глибшому розумінню кореляції між складниками досліджуваної проблематики; г) *штучні нейронні мережі* (ChatGPT (3.5 та 4), DALL E2, Gemini, LeiaPix, Narakeet, Pika, Runway, Sora тощо) – дискусійна група джерел, що продуковано дуальною природою останніх: вони виступають одразу й інструментом, й джерелом дослідження нейромережевого моделювання мовних одиниць; за своєю природою є одним з найрепрезентивніших видів, оскільки їх актуалізація продукує практичне втілення тих чи тих гіпотез дослідника з осягненням належної параметризації їх функціонування при цьому. Таким чином, актуалізація різноманітного спектра джерел продукує об'єктивність та повноту сучасного гуманітарного (зокрема, лінгвістичного) дослідження, дозволяючи комплексно проаналізувати досліджувану проблематику й забезпечити відтворюваність, наукоємність, сучасність тощо проведеної наукової роботи.

Ключові слова: нейромережеве моделювання, дослідження нейромережевого моделювання, прикладна лінгвістика, комп'ютерна лінгвістика, штучні нейронні мережі, нейромережеві моделі, джерела сучасного дослідження.

Стаття надійшла до редакції 27.03.2024
доопрацьована 03.04.2024
прийнята до друку 09.04.2024