Вісник Національного технічного університету "ХПІ", 2026, № 1 (15)
ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

UDC 004.8 **DOI: 10.20998/2411-0558.2026.01.05**

D. SALNIKOV, PhD, associate professor, National technical University "Kharkiv polytechnic institute",
V. KRYLOVA, PhD, associate professor, National technical University "Kharkiv polytechnic institute",
R. KOTKO, PhD, postgraduate student, National technical University "Kharkiv polytechnic institute"

SEMANTIC PREPROCESSING AND SELECTIVE TRANSLATION OF POWERPOINT SLIDES USING EMBEDDING'S AND GPT-4o

This article presents an artificial intelligence approach for automatic translation of presentation slides using semantic embedding's and a lightweight neural network for intelligent preprocessing. The system aims to selectively translate only relevant slide content – excluding references, metadata, and technical terms – by semantically classifying text chunks before passing them to the GPT-4o model. The approach enhances translation quality, reduces token consumption, and maintains slide formatting integrity. The proposed method demonstrates significant gains in translation efficiency and contextual accuracy, especially for academic educational presentations. The obtained results confirm the potential of integrating semantic models and generative language systems in the field of automated translation of educational materials. Future research plans include extending the system for multilingual support and adaptation to various presentation formats. Figs.: 5. Refs.: 12 items.

Ключові слова: слова: slides translation, semantic embedding's, semantic preprocessing, GPT-4o, NLP pipeline, document translation.

Introduction. Recent advancements in large language models have demonstrated their capacity to produce high-quality translations across a wide range of languages. In addition to general-purpose models, specialized architectures have been developed to address machine translation tasks with increased accuracy and efficiency. Concurrently, text classification has been extensively explored in the natural language processing literature, with numerous methods proposed to categorize textual data based on semantic or contextual features. A central objective of these efforts is to develop algorithms that offer both high performance and computational efficiency.

A multi-label text classification approach that enhances performance by integrating semantic relationships among labels through label vector fusion is proposed in [1]. Another approach is to use fastText for efficient and scalable

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

text classification, as demonstrated in [2], achieving strong performance on short-text data with low computational cost. In [3] word embedding's are combined with a multi-grained cascade forest, offering a non-deep learning approach that balances accuracy and efficiency in multi-class classification tasks.

In contrast, non-neural network approaches have also been explored. In [4], authors utilize a multi-level fuzzy neural network that incorporates fuzzy logic to handle linguistic uncertainty, improving classification accuracy. Similarly, [5] proposes an integrated algorithm that combines multiple techniques to effectively classify short texts, addressing challenges related to sparse features and limited context.

Semantic preprocessing and filtering represent crucial techniques in current machine translation research. Filtering and pruning low - quality data, improves translation accuracy and quality [6]

The following studies highlight the importance of integrating semantic and contextual information to improve translation accuracy and coherence.
The approach demonstrated in [7] introduces graph convolutional networks to predicate – argument structures into neural machine translation, boosting BLEU scores on English to German translation tasks. Authors demonstrate semantic-aware encoding strategies that allow detection and classification of text chunks (e.g. notes, references).

LLM-based (like GPT-3.5, GPT-4) preprocessing on a document-level translation method, as described in [8], outperforms commercial translation systems on discourse coherence. Such applications are directly applicable to improve the translation quality of slide content isolated in chunks.

Some studies suggest including the whole document context in translation [9]. Such a straightforward method allows for enhancing English to German, English to French and French to English translations.

On the other hand, [10] introduces hierarchical attention mechanisms to integrate broader document context into translation, achieving BLEU score improvements over sentence-level models. This approach is particularly relevant for managing slide sequences and ensuring contextual coherence, thereby supporting continuity across presentations.

Another pivotal consideration in deploying LLMs for translation tasks is the computational expense, which scales markedly with model capacity. For instance, large-capacity models offer enhanced fluency and better handling of multilingual context but incur substantially higher costs. Empirical data from OpenAI's API pricing illustrates this disparity. The GPT-4o model charges approximately $2.50 per million input tokens and $10.00 per million output tokens. In contrast, its smaller counterpart, GPT-4o Mini, dramatically reduces costs to $0.15 per million input tokens and $0.60 per million output tokens. At the upper end of the performance spectrum, GPT- 4.5 – a more powerful and resource-intensive model – entails much steeper pricing: $75 per million input tokens and $150 per million output tokens Wikipedia. Therefore, there is a salient trade-off: while higher-capacity models may yield superior translation quality, they also result in exponentially greater operating costs. This dynamic necessitates careful model selection to strike a balance between translation fidelity and economic feasibility, particularly when designing systems meant for high-volume or real-time translation workflows.

The escalating cost of high-parameter language models underscores the need for more efficient translation algorithms. To achieve high translation quality at lower computational expense, future work should focus on methods that selectively utilize LLMs through intelligent preprocessing, semantic filtering, and lightweight neural components. Such approaches can improve the quality-per-dollar ratio, enabling scalable and cost-effective translation solutions.

Additionally, it is important to note that, since the focus of this article is on translating PowerPoint slides, the available context depth is limited by the presence of graphical elements, images, captions, and other non-textual content.

Quality evaluation method. The Bilingual Evaluation Understudy (BLEU) metric proposed in [11] is one of the most widely used automatic measures for assessing the quality of machine translation. It operates by comparing machine-generated translations with one or more human reference translations, relying on n-gram overlap to quantify lexical similarity

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log\, p_n\right). \quad (1)$$

Another approach to evaluate translation quality is BERTS crore[12] that uses contextual embedding's from BERT model to compute semantic similarity between candidate and reference translations. The BERTS crore is calculated by aligning each token in x with a token in $\hat{x}$ to compute recall and precision. A greedy matching strategy is employed, ensuring that each token is paired with

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

the most similar token in the opposite sentence to maximize the similarity score. Precision and recall are then combined into an F1 measure. For a given reference x and candidate $\hat{x}$, the recall, precision, and F1 scores are defined as follows.

$$R_{BERT} = \frac{1}{|x|} \sum_{x_i \in x} \max_{\hat{x}_j \in x} x_i^T \hat{x}_j, \tag{2}$$

$$P_{BERT} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_i \in \hat{x}} \max_{x_i \in x} x_i^T \hat{x}_j, \tag{3}$$

$$F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}}. \tag{4}$$

While BLEU, BERTScrore and others, provides a standardized and reproducible way to evaluate textual output, they have important limitations. Specifically, such metrics does not account for semantic meaning beyond surface-level word matching and is entirely text-centric. As a result, they are not applicable for evaluating translation quality in contexts where graphical elements, images, or other non-textual content play a central role, such as PowerPoint slides. In such cases, additional evaluation methods that consider layout, multimodal context, and the interaction between text and graphics are required.

The following figure illustrates an important limitation of purely text-based evaluation metrics. While the translation in this example receives a satisfactory BLEU score, it nevertheless includes segments that should not be translated, such as licenses and reference entries. These elements are typically standardized, legally binding, or context-independent, and their translation may introduce inconsistencies or even inaccuracies.

Fig. 2 presents original (*a*) and translated (*b*) slide images. Although translation achieve acceptable BLEU score, our preference is to retain technical terms in English and produce translations that align with the target style and formatting width.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

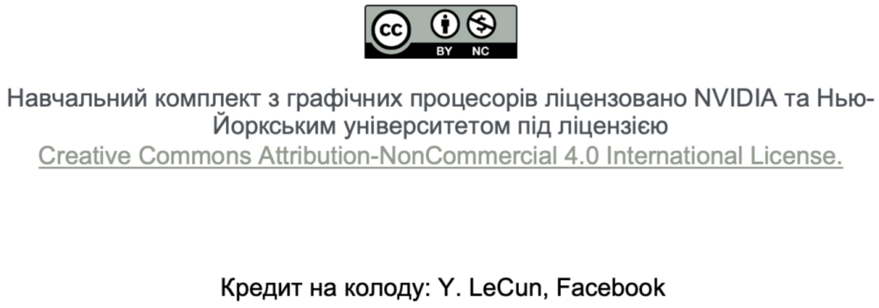
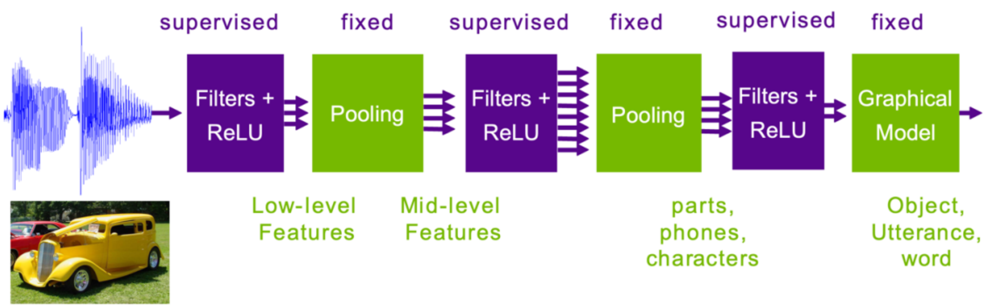
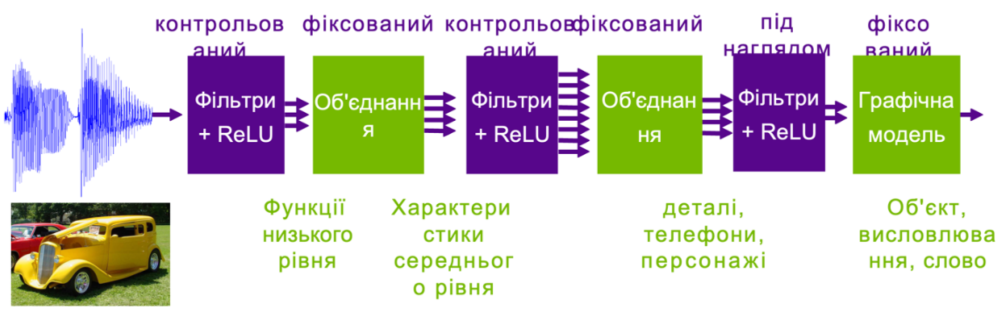


Fig. 1. Example of the slide with untranslatable data



(*a*)

Fig. 2. Example of the slide with technical terms and styling – original slide



(*b*)

Fig. 2. Example of the slide with technical terms and styling – translation

A GPT-4o – based evaluation agent was designed to evaluate translation quality of visual represented data. In this setup, the agent receives as input the

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

original slide image and a candidate translated version of the same slide. The block diagram of such approach is presented on Fig. 3.

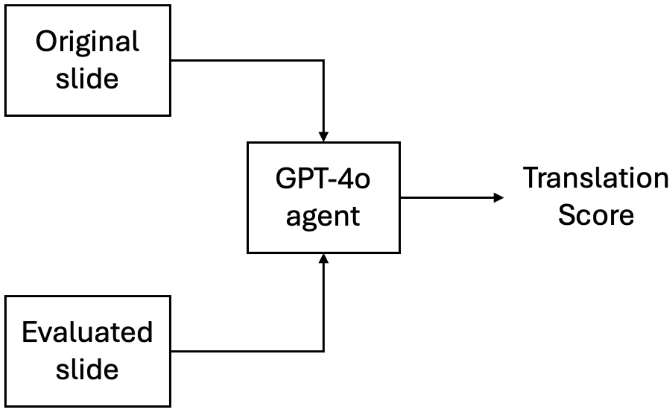


Fig. 3. Block diagram of the translation quality evaluation algorithm

By analyzing textual content, layout, and visual elements, the model generates a translation score that reflects not only lexical accuracy but also contextual fidelity and alignment with graphical components. Such an approach extends beyond purely text-based metrics like BLEU by incorporating multimodal reasoning, enabling more reliable evaluation of translation quality in slides where text, images, and formatting are tightly interdependent.

Filtering neural network architecture. Text embedding's are high-dimensional numerical representations of language that capture semantic relationships between textual inputs. These representations enable quantitative assessment of similarity or relatedness between text segments and are widely applied in scientific and engineering domains, including information retrieval, clustering, recommendation systems, anomaly detection, and text classification In the proposed method we use OpenAI's text-embedding-3-small model, an enhanced and more efficient successor to the Ada embedding model. Thus, input of the networks is an embedding vector of the tokenized text chunk.

The neural network is intentionally designed to be lightweight, to allow efficient execution on relatively low-performance CPUs (Fig. 4).. This reduces preprocessing overhead and ensures that filtering can be performed without the need for specialized hardware. By minimizing computational requirements, the

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

approach lowers overall system costs by limiting the amount of unnecessary text passed to LLM for translation.

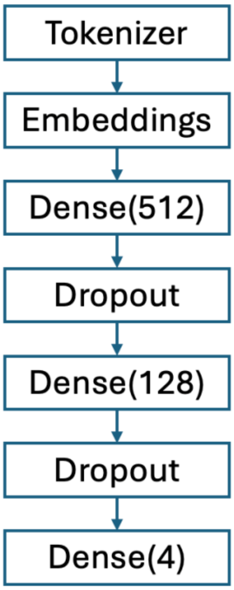


Fig. 4. Architecture of the proposed neural network

The proposed filtering model is implemented as a feed forward neural network operating on embedding vectors of text chunks extracted from slides. The architecture begins with a dense layer of 512 units with ReLU activation, followed by a dropout layer with a rate of 0.3 to reduce over fitting. This is succeeded by a second dense layer of 128 units with ReLU activation and an additional dropout layer with a rate of 0.2. The final layer is a four-unit dense layer with softmax activation, producing class probabilities corresponding to “translatable text”, “reference”, “term” and “note”.

The network is trained using categorical cross-entropy loss and optimized with Adam, employing a learning rate scheduler to ensure stable convergence. Training data consists of approximately 5000 manually labeled text chunks, providing the model with sufficient examples to learn effective discrimination between content that should and should not be translated.

Results. The proposed approach demonstrates improved translation performance, as confirmed by experiments on English-to-Ukrainian slide

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

translation tasks. An example of the usage of the proposed preprocessing system is illustrated in the Fig. 5.

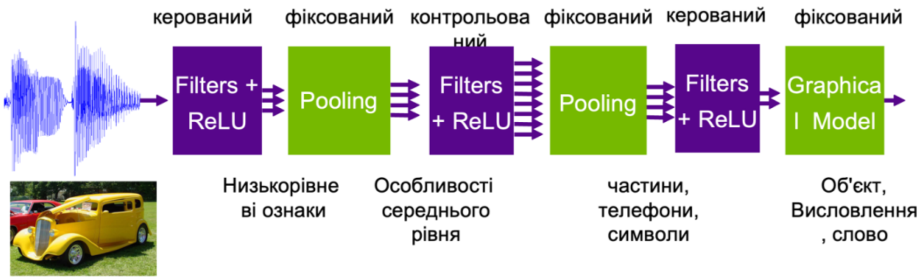


Fig. 5. Example of the slide with not translated technical terms and preserved styling

Using the proposed method, we conducted an evaluation on 163 translated slides, comparing the results against translations produced by DeepL. The analysis employed the proposed scoring approach, which considers both textual accuracy and contextual integration with graphical elements.

Our findings indicate that slides translated with the proposed method consistently outperformed those generated by DeepL, achieving improvements of approximately 5 to 15% in translation quality scores. This demonstrates the effectiveness of semantic preprocessing and selective translation in enhancing the overall fidelity of slide translations.

References:

1. Labels Vector Fusion," in 2018 International Conference on Promising Electronic Technologies (ICPET), 2018, pp. 80–85. https://doi:10.1109/ICPET.2018.00021

2. T. Yao, Z. Zhai, and B. Gao, "Text Classification Model Based on fastText" in 2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS), 2020, pp. 154–157. https://doi:10.1109/ICAIIS49377.2020.9194939

3. W. Wang, G. He, and X. Liu, "Text Multi-classification Based on Word Embedding and Multi-Grained Cascade Forest," in 2019 IEEE 5th International Conference on Computer and Communications (ICCC), 2019, pp. 13–17. https://doi:10.1109/ICCC47050.2019.9064153

4. S. Zobeidi, M. Naderan, and S. E. Alavi, "Effective text classification using multi-level fuzzy neuralnetwork," in 2017 5th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), 2017, pp. 91–96. https://doi:10.1109/CFIS.2017.8003664

5. D. Zhao, N. Du, and L. Qin, "Study on Short Text Classification with Integrated Algorithm," in 2016 13th Web Information Systems and Applications Conference (WISA), 2016, pp. 121–124. https://doi:10.1109/WISA.2016.33

6. G. Xu, Y. Ko, and J. Seo, "Improving Neural Machine Translation by Filtering Synthetic Parallel Data," Entropy, vol. 21, no. 12, 2019, https://doi:10.3390/e21121213

7. D. Marcheggiani, J. Bastings, and I. Titov, "Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks." [Online]. Available: https://arxiv.org/abs/1804.08313

8. L. Wang et al., "Document-Level Machine Translation with Large Language Models." [Online]. Available: https://arxiv.org/abs/2304.02210

9. V. Macé and C. Servan, "Using Whole Document Context in Neural Machine Translation." [On-line]. Available: https://arxiv.org/abs/1910.07481.

10. L. Miculicich, D. Ram, N. Pappas, and J. Henderson, "Document-Level Neural Machine Translation with Hierarchical Attention Networks." [Online]. Available: https://arxiv.org/abs/1809.01576

11. K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a Method for Automatic Evaluation of Machine Translation," in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, P. Isabelle, E. Charniak, and D. Lin, Eds., Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, Jul. 2002, pp. 311–318. https://doi:10.3115/1073083.1073135

12. Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger,Yoav Artzi, "BERTScore: Evaluating Text Generation with BERT". 2020. doi: https://doi.org/10.48550/arXiv.1904.09675

Статтю представив д-р техн. наук, професор НТУ "ХПІ" Качанов П.О.

Поступила (received) 01.10.2025
Стаття прийнята до друку 25.01.2026
Опублікована 27.02.2026

Salnikov Dmytro
PhD, associate professor,
National Technical University «Kharkiv Polytechnic Institute»
Kyrpychova str., 2, Kharkov, Ukraine, 61002
E-mail: dmytro.salnikov@khpi.edu.ua
ORCID: 0009-0007-6201-5370

Krylova Viktoriia
Ph.D, associate professor
National Technical University "Kharkiv Polytechnic Institute"
Kyrpychova str., 2, Kharkiv, Ukraine, 61002
E-mail: viktoriia.krylova@khpi.edu.ua

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

ORCID ID: 0000-0002-4540-8670

Kotko Roman
postgraduate student
National Technical University “Kharkiv Polytechnic Institute”
Kyrpychova str., 2, Kharkiv, Ukraine, 61002
E-mail: roman.kotko@cit.khpi.edu.ua
ORCID ID: 0009-0006-2847-4349

УДК 004.8

Семантична попередня обробка та вибірковий переклад слайдів PowerPoint з використанням вбудовування та GPT-4o / Сальніков Д.В., Крилова В.А., Котко Р.О. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2026. – № 1 (15). – С. 83 – 93.

У цій статті представлено підхід штучного інтелекту до автоматичного перекладу слайдів презентацій з використанням семантичного вбудовування та легкої нейронної мережі для інтелектуальної попередньої обробки. Система спрямована на вибірковий переклад лише релевантного вмісту слайдів, виключаючи посилання, метадані та технічні терміни, шляхом семантичної класифікації фрагментів тексту перед їх передачею до моделі GPT-4o. Цей підхід підвищує якість перекладу, зменшує споживання токенів та зберігає цілісність форматування слайдів. Запропонований метод демонструє значне підвищення ефективності перекладу та контекстної точності, особливо для академічних освітніх презентацій. Отримані результати підтверджують потенціал інтеграції семантичних моделей та генеративних мовних систем у сфері автоматизованого перекладу навчальних матеріалів. Майбутні плани досліджень включають розширення системи для багатомовної підтримки та адаптацію до різних форматів презентацій. Іл.: 5. Бібліогр.: 12 назв.

Ключові слова: переклад слайдів, семантичне вбудовування, семантична попередня обробка, GPT-4o, конвеєр NLP, переклад документів.

UDC 004.8

Semantic Preprocessing and Selective Translation of PowerPoint Slides using Embeddings and GPT-4o / Salnikov D., Krylova V., Kotko R. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkiv: NTU "KhPI". – 2026. – № 1 (15). – P. 83 – 93.

This article presents an artificial intelligence approach for automatic translation of presentation slides using semantic embedding’s and a lightweight neural network for intelligent preprocessing. The system aims to selectively translate only relevant slide content – excluding references, metadata, and technical terms – by semantically classifying text chunks before passing them to the GPT-4o model. The approach enhances translation quality, reduces token consumption, and maintains slide formatting integrity. The proposed method demonstrates significant gains in translation efficiency and contextual accuracy, especially for academic educational presentations. The obtained results confirm the potential of integrating semantic models and generative language systems in the field of automated translation of educational materials. Future research plans include extending the system for multilingual support and adaptation to various presentation formats. Figs.: 5. Ref.: 12 items.

Keywords: slides translation, semantic embedding’s, semantic preprocessing, GPT-4o, NLP pipeline, document translation.