The Use of Distilled Large Language Models to Determine the Sentiment of a Text

Ksenia Andrenko
Brest State
A.S. Pushkin University
Brest, Belarus
ksenia.andrenko2002@gmail.com

Aliaksandr Kroshchanka
Brest State
Technical University
Brest, Belarus
kroschenko@gmail.com

Olga Golovko
Brest State
Technical University
Brest, Belarus
olechkagolovko@gmail.com

***Abstract*—This paper describes a distilled T5-large Transformer-based model for text sentiment analysis optimised for educational applications. The model, created by transferring knowledge from Large Language Models, demonstrates an accuracy of 97.22% on a test sample with a significant reduction in computational requirements. The distillation process, training methodology and experimental results are described. Application scenarios in educational contexts including adaptive learning and personalisation of feedback are proposed. The study shows promise for the use of distilled Large Language Models in resource-limited educational technologies.**

***Keywords*—Large Language Model, reduction, sentiment detection, knowledge distillation, linguodidactics.**

I. Introduction

Large Language Models (LLMs), for instance Grok, ChatGPT, DeepSeek, Llama, Qwen, Gemini, Claude and Mistral, have demonstrated high performance in many linguistic tasks (dialogue communication, syntactically and semantically correct text generation, summarisation, interlingual translation, text stylisation and attribution). This has resulted in a notable surge of interest in LLMs from educators and learners, including language teachers. LLMs have the capacity to automatically generate learning tasks, select examples, provide age-appropriate explanations of material, create a quiz, develop a didactic crossword puzzle, illustrate learning material, and numerous other functions. LLMs offer several advantages, including their perpetual availability and the provision of instantaneous, refined feedback. They also provide a diverse array of materials and the capability for personalised learning [1]. The personalisation of learning is inextricably linked to the development of the learner's personality [2]. The personalisation of learning with LLMs is a distinctive feature, with the capacity to adapt to the needs of the interlocutor (gender, age, level of education), including the emotional background, by determining the sentiment of the texts.

II. Statement of the problem of sentiment determination

Sentiment is a complex linguistic category that reflects the emotional and evaluative components of a communicative act. It is realised both at the lexical level through individual linguistic units (lexical tonality) and at the level of integral communicative fragments [3].

Sentiment analysis involves the identification and classification of those elements of a text or speech that express the emotional attitude of the subject of communication (speaker, author) to the object that is the subject of emotional evaluation. The subject's emotionality can be represented in the form of three main categories: positive, neutral or negative. Thus, sentiment reflects the author's position towards the topic under discussion, which, as a rule, is described within a one-dimensional emotional space, where the main poles are "positive" and "negative" evaluations.

For the purposes of linguodidactics, where the adaptation of scientific concepts to educational tasks is required, the definition of sentiment can be simplified and reformulated to meet the needs of learners. In this context, sentiment is defined as the emotional colouring of a speech or text that reflects the speaker's (or writer's) attitude to the subject of discussion, interlocutor or the situation being described.

III. Limitations of LLMs in education

In the context of pedagogical applications, one of the most salient features of LLMs is the partial unpredictability and unreliability of the results they produce. All LLMs are prone to errors, and they have been observed to 'hallucinate' (i.e. construct imaginary facts and operate on them as if they were real). This phenomenon remains to be the subject of rigorous scientific explanation, and its mitigation through technical means is challenging without significant disruption to communication. Furthermore, LLMs are trained on an unrestricted corpus of internet texts, including blogs and forums, which contain profanity, factually inaccurate information, and other such content. This renders their use in the educational process both risky and uncontrolled.

Large-scale neural network architectures, in particular transformers with several billion tunable parameters, demonstrate high accuracy, but face significant operational constraints during training: they require powerful

GPU/TPU accelerators, occupy a significant amount of memory, have high power consumption and significant infrastructure costs, making them inaccessible to a wide range of users and necessitating the development of size optimisation techniques.

IV. Distilled models

In the field of large neural network model optimisation, a number of approaches have been developed to create compact architectures, comprising several major categories. Distilled models, such as DistilBert and TinyLlama, exploit the process of knowledge transfer from a large teacher model to a smaller student model, achieving preservation of up to 97% efficiency while significantly reducing the number of parameters. Efficiently designed models, represented by the MobileNet, ShuffleNet, SqueezeNet, EfficientNet and ALBert families, are initially developed with computational resource optimisation as the primary consideration, using special architectural solutions such as depthwise separable convolutions and channel shuffle. Quantisation allows reducing the model dimensionality by reducing the parameter type to 8-bit integers, while the pruning technique helps to achieve a similar effect by thinning the model, in the process of which low-value weights, neurons and even whole layers of neural elements are removed. A separate category is tensor compact models, which use special tensor representations to reduce the number of parameters while maintaining the generalisability of the network. Each of these approaches offers a different balance between performance and resource efficiency, allowing to choose the optimal solution for specific tasks and application conditions.

All of the above techniques, except for the use of special architectural solutions, can be labelled as methods of compression (reduction) of neural network models, allowing to reduce the size of the model while maintaining its efficiency.

This paper focuses on the technique of knowledge distillation, which is particularly important for language models because it preserves the underlying patterns that large models extract from large amounts of data in a more compact form. For example, Hinton describes how distillation can transfer knowledge from complex neural networks to smaller ones, preserving their efficiency [4].

In the context of educational technology, knowledge distillation becomes critical for text sentiment detection, providing four key benefits. Firstly, the efficiency and increased speed of compact models (e.g., DistilBert demonstrates a 60% speedup while maintaining 97% efficiency compared to the BERT model) allows for rapid, real-time text sentiment detection, which is especially important for fast-paced classroom work [5]. Secondly, the significant reduction in equipment and maintenance costs makes AI technologies affordable even for educational institutions with limited budgets [6]. Thirdly, the ability to adapt distilled models to specific educational tasks, from creating grammar exercises to mathematical problems of varying levels of difficulty, helps to personalise learning and increase its effectiveness [7]. Finally, fourthly, the use of distilled models that solve specific problems allow us to achieve manageability, through the possibility of direct pre-training, architectural changes, local deployment, etc. This eliminates dependence on large models deployed in the cloud and available only as a service.

V. Computational experiment

To determine the sentiment of a text, we trained a distilled Transformer t5 model. The experiments were conducted with a dataset of texts totalling 1080 sentences in English with different emotional colouring. The dataset was balanced into three sentiment classes: negative, neutral and positive (360 sentences of each class). To increase the representativeness of the data, texts from different thematic areas were included in the dataset: education, interpersonal relationships, parenting, travelling, etc. Subsequently, the data were divided into training (900 sentences, 83.3%) and test (180 sentences, 16.7%) datasets while maintaining a proportional distribution of classes. The said datasets were generated using the Claude 3.7 Sonnet, Grok 3 and Gemini 2.0 Pro models, followed by manual checking and correction by linguists to ensure naturalness and stylistic diversity.

As a basic model for distillation we used the seq2seq model of T5-large transformer, with the number of adjustable parameters 770 million [8]. The input of the model during training was JSON-formatted data, while the output was JSON-formatted reference values containing the type of sentiment, the confidence level of the model, and an explanation of why such a decision was made when classifying a sentence.

Example of input and output sequences:

```
{
"instruction": "Analyse the sentiment
    of the following text.",
 "input": "Please open your books to
    page 10."
}
```

```
{
 "sentiment": "neutral",
 "confidence": 0.98,
 "explanation": "A polite request/
    instruction, expressing no
    explicit emotions."
}
```

Training was performed for 100 epochs. Main parameters: mini-batch size – 16 instances, learning rate – 5e-5, and Adam method with weight decay was used for minimisation [9], input sequence size – 64 tokens, output sequence size – 128 tokens.

We used the A100 video accelerator in the training process [10].

In Fig. 1 the evolution of the error on the training dataset during the training process is depicted.

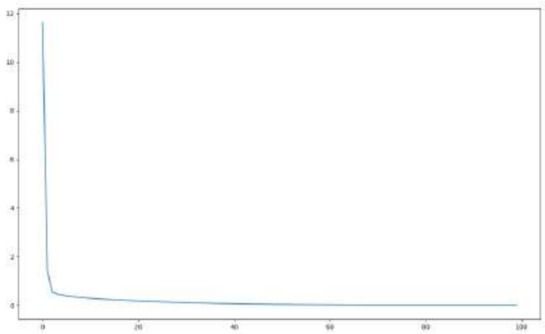

Figure 1. Evolution of the error on the training dataset.

The confusion matrix calculated on the test dataset is given in Table I.

Table I
Confusion matrix

		Actual type		
		NEG	NEU	POS
Type, received using the model	NEG	59	0	0
	NEU	1	59	3
	POS	0	1	57

The final accuracy of the model on the training dataset was 99.89% (one error per 900 instances), on the test dataset it was 97.22%.

Examples of errors made by the model for sentences from the test dataset, taking into account the explanatory part, are given in the table II (T_a – actual type of sentence, T_a – type obtained by the model, Exp_a – actual explanation, Exp_m – explanation obtained by the model).

Thus, during the computational experiment, a distilled t5 transformer model was trained with high accuracy for text sentiment detection.

The obtained high accuracy rates (97.22% on the test dataset) are conditioned by several key factors: the specificity of the source data (sentences generated with LLMs contain more pronounced lexical sentiment markers compared to natural texts, which considerably simplifies the classification task), the advantages of the T5 architecture, which was originally developed as a universal model for various NLP tasks, providing high efficiency in tasks that require deep understanding of context and semantics, the efficiency of the T5 architecture, and the efficiency of the T5 model in tasks that require deep understanding of context and semantics.

Table II
Examples of errors

Sentence	T_a	T_m	Exp_a	Exp_m
International trade has brilliantly lifted millions of families out of poverty	POS	NEU	Uses positive language ('brilliantly') to highlight poverty reduction benefits of global trade.	The text uses strong positive evaluation ('brilliantly') to praise the success of international trade, implying admiration, but not strong emotion
The human toll of globalisation can be seen in devastated factory towns and abandoned communities	NEG	NEU	Uses negative language ('human toll', 'devastated', 'abandoned') to describe social impacts	The text presents a factual observation about the impact of a common technological approach without expressing an opinion about the moral or environmental value of this impact.
Tourism has expanded dramatically as a result of globalisation	NEU	POS	A factual observation about tourism trends without clear emotional judgement	The text uses a strong positive descriptor ('dramatically') to describe the growth of tourism as a result of modern technological development

VI. Personalisation of education at the systemic level

In the long term, sentiment analysis can become an integral part of the educational process, transforming approaches to learning, interaction and assessment.

With the emergence of sentiment analysis technologies, there is a concomitant shift in the role of the teacher from the traditional "knowledge transmitter" to that of a facilitator and mentor. This new role is characterised by the utilisation of data to provide students with support and guidance. It is predicted that teachers will come to rely on automated systems for routine tasks such as the analysis of texts or monitoring of engagement, thus allowing them to focus more strategically on issues such as the development of individual learning trajectories or the nurturing of interpersonal relationships in the classroom.

Sentiment analysis has the potential to underpin the creation of personalised educational systems, whereby each student receives learning materials, assignments and feedback that are tailored to their cognitive, emotional and social characteristics. This is of particular importance in mass education settings, where traditional approaches often fail to account for individual differences.

For instance, the education platform employs sentiment analysis in conjunction with other data (e.g., test scores, task completion rates) to generate an individual profile for each student. Utilising this comprehensive profile, the system then proposes an optimal pace for learning, the most suitable types of assignments, and the instructor's preferred communication style.

Sentiment analysis also can contribute to the integration of emotional intelligence into educational programmes, making it an integral part of learning. This will prepare students not only for professional activity, but also for successful interaction in society, where the ability to understand and manage emotions is becoming increasingly important.

Sentiment analysis can be a tool to promote intercultural education, helping students to understand how emotions are expressed in different languages and cultures. This is particularly relevant in the context of globalisation, where students increasingly interact with people from other cultures, both in educational and professional contexts.

VII. Prospects for the application of the model in linguodidactics

Sentiment analysis has considerable potential for application in an educational context, covering a wide range of tasks.

For example, sentiment analysis plays a significant role in creating adaptive learning materials. It is possible to track preferred exercises, identify the most difficult and uninteresting topics, and find ways to increase student engagement. For example, a sentiment detection system can analyse students‘ comments and, detecting negative sentiment ("This topic is boring.", "Upbringing of children is tiresome, difficult and time-consuming.", "I don't like make up, so I'm not interesting in these words and the words are complicated to memorise."), suggest additional exercises to memorise vocabulary and factual information in game form or video lessons with native speakers. When integrated with artificial intelligence systems, it is possible to create more complex adaptive systems that will take into account not only the emotional state, but also the cognitive characteristics of learners.

The monitoring of student engagement is another important task that can be addressed by analysing the sentiment of the text. In the context of traditional teaching, the teacher is able to read non-verbal signals (gestures, facial expressions). However, it is simply impossible to track all students during a lesson, and the temporary manifestations of any emotions in the lesson may be caused by external factors unrelated to the educational process itself. Unlike traditional teaching, where the teacher focuses on non-verbal cues, automatic analysis helps to take more factors into account. The system can automatically analyse students‘ written answers, identifying markers of emotional state, such as the use of emotionally coloured vocabulary ("awful", "good", "clear", "curious") or patterns of behaviour (reducing the amount of text in answers). Thus, the teacher can promptly adjust the educational process by adding motivating elements, changing the focus and format of tasks, teaching material.

If attitudes towards the assignment or topic have been identified, there is still the problem of feedback. Sometimes it is difficult for the teacher to formulate a response that is not clichéd. At the same time, an inappropriate tone (e.g., overly critical or formal) can demotivate the student and reinforce his or her negative attitude towards the subject. For example, in an essay, a student writes "I'm not sure, whether my essay has a convincing argument for my point of view". The sentiment analysis system classifies this as negative or uncertain sentiment and offers the teacher the following comment: "You've already laid a good foundation for your argument, let's add some examples to strengthen your position". This approach helps to increase the student's confidence and motivate them to work further.

Sentiment analysis is not limited to written work. Modern automatic speech recognition systems are able to record the speech of teachers and students in the classroom, after which the sentiment analysis system can process this data.

This can be useful for teachers who wish to improve their speech in the classroom from the position of emotional colouring. For example, it is extremely important for student teachers to control their speech during the formation of professional speech (tracking negative vocabulary, imperatives, etc.). For example, a teacher uses a large number of imperatives: "Go to the blackboard. Open your books.", modal verbs with the meaning of owing "You have to do it. You must write down these words."

It is important for students of linguistic specialities to understand the emotional colouring of a text. Thus, the sentiment analysis system can check essays, oral recordings, comments, simple sentences, determining the level of emotional colouring of the text. For example, "The book left me with a deep sense of sorrow and contemplation. The protagonist's fate was truly heartbreaking." The text contains emotionally coloured words ("sorrow", "heartbreaking"), which indicates a negative sentiment, namely sadness and sorrowfulness. "This explanation is so confusing. I have no idea what the teacher meant." The phrases "so confusing", "no idea" indicate frustration and misunderstanding. Perhaps a further explanation should be offered.

Thus, sentiment analysis can help students better understand the emotional colouring of a text and improve their language skills. However, despite its considerable potential, it is important to consider a number of limitations associated with its application in an educational setting. Consider the main ones.

VIII. Limitations in the application of the model

Modern sentiment analysis algorithms, although demonstrating high accuracy in standard situations, may make mistakes when analysing complex texts containing irony, sarcasm, cultural allusions or non-standard language constructions. This is especially true in educational contexts where students' texts may be informal, contain errors, or reflect individual differences in style.

The system categorises the student's message "Oh great, another five-page essay due tomorrow. I just love staying up all night writing!". The expected sentiment of this comment is negative (sarcasm, dissatisfaction).

A possible algorithm error is evaluating sentiment as positive (because of the words "great", "love") without recognising sarcasm.

The use of sentiment analysis to monitor students' emotional state raises serious issues of confidentiality and ethics. One of the key issues is the collection of data without the explicit consent of students. This can lead to them holding back in expressing their thoughts for fear that their comments will be used against them.

Another important aspect is the use of the analysis results for academic decision making. If the algorithm determines that a student often uses negatively coloured expressions such as "I don't understand anything" or "This topic is frustrating", and on this basis the system marks him as under-motivated, this may affect his grade or level of learning. However, the emotional colouring of speech does not always correlate with actual performance.

In addition, there is a risk of stigmatisation. If a teacher receives a report in which one of the students is labelled as having a tendency to make negative comments, this may influence his/her attitude towards the student. Human perception is influenced by such labelling and the teacher may start to see the student as a problem student, even if there is no real reason for this.

The possibility of manipulation with the system should also be considered. If students realise that sentiment analysis affects their learning process, they may deliberately use positive expressions such as "This lesson was amazing!" or "I feel super confident!" even if they are actually experiencing difficulties. As a result, the teacher receives unreliable data, which reduces the effectiveness of feedback and hinders the adaptation of the learning process to the real needs of students.

Consequently, sentiment analysis holds considerable potential for implementation within the domain of linguodidactics. This is due to its capacity to personalise learning experiences, monitor student engagement, and consider their emotional state. Nevertheless, its effective implementation necessitates a comprehensive approach encompassing not only technological but also ethical components. It is imperative to consider the implications of language and cultural idiosyncrasies, as well as the potential inaccuracies inherent in algorithmic processes. The introduction of such technologies must be carried out in compliance with the principles of transparency, voluntariness and protection of personal data.

Furthermore, it is imperative to acknowledge that sentiment analysis should not be regarded as a standalone tool. Indeed, it is the combination of sentiment analysis with other methods of learning assessment and adaptation, including artificial intelligence, personalised learning strategies and pedagogical analysis, that will ensure optimal efficacy. It is only by adopting an integrated approach that a truly flexible and supportive learning environment can be created, where technology can take into account not only students' knowledge and skills, but also their emotional state.

IX. Distilled language models in OSTIS systems

To integrate a trained distilled language model, it is initially necessary to decide on the role that the model will perform in the OSTIS Ecosystem.

On the basis of the model, a problem solver can be developed – for example, to analyse the sentiment of a text. In this case, integration can be performed by embedding the model in a broader knowledge processing context. For example, the model analyses the text "Product is great, but delivery is slow". It determines the mixed sentiment, and then this result is passed to the OSTIS knowledge base, where it can be linked to the "Product" and "Delivery" entities for further analysis or decision making.

In the process of integration it is necessary to perform transformation of input and output data into SC-code. For this purpose, it is necessary to formalise the analysed text and encode the resulting sentiment as a semantic node associated with the analysed text. Such transformations allow the result to be used for more complex inferences.

Integration can also be done by implementing an agent that interacts with the model. In this case, the most proven approach is the use of microservice architecture with the implementation of an appropriate API for interaction with the OSTIS-agent.

Integration of the distilled model with the OSTIS system will allow it to be used in more complex scenarios, for example:

- **feedback analysis**: the model determines the sentiment of the reviews, and OSTIS makes recommendations or identifies problem areas (e.g., "slow delivery" is found in 70% of negative reviews);
- **training systems**: the sentiment of a student's text can be analysed to assess their emotional state and OSTIS will offer adapted learning content;
- **social media monitoring**: integration with data from different sources will allow real-time tracking

of public opinion, and OSTIS will link sentiment to specific topics or events.

X. Conclusion

Thus, we have developed a distilled sentiment analysis model based on the T5 architecture that achieves 97.22% accuracy on a test dataset while reducing the number of parameters by 92.2% compared to the original model, demonstrated the effectiveness of the knowledge distillation technique for creating resource-efficient natural language analysis models, and proposed scenarios for applying the developed model in an educational context, including adaptive learning, monitoring student engagement, and personalising feedback. The results obtained indicate the promising application of distilled language models in educational technologies, especially in the context of limited computational resources. The compact size and high speed of operation make the developed model suitable for integration into mobile applications and online learning platforms.

References

[1] K. Andrenko, "Linguodidactic potential of large language models in english language teaching," in *BIG DATA and Advanced Analytics : a collection of scientific articles. X International scientific and practical conference*. Minsk : BSUIR, 2024, pp. 393–397.

[2] A. Asmolov, "Over-specification. you can't turn a standard into a corset," in *Educational policy*, vol. № 1–2 (77–78), 2019, pp. 18–23.

[3] B. Gasparov, *Language, memory, image. Linguistics of language existence*. Moscow: New Literary Review, 1996.

[4] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," 2015. [Online]. Available: https://arxiv.org/abs/1503.02531

[5] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," 2020. [Online]. Available: https://arxiv.org/abs/1910.01108

[6] H. Xu, W. Gan, Z. Qi, J. Wu, and P. S. Yu, "Large language models for education: A survey," 2024. [Online]. Available: https://arxiv.org/abs/2405.13001

[7] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, and G. Kasneci, "ChatGPT for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1041608023000195

[8] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," 2023. [Online]. Available: https://arxiv.org/abs/1910.10683

[9] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019. [Online]. Available: https://arxiv.org/abs/1711.05101

[10] "NVIDIA. NVIDIA A100 Tensor Core GPU Architecture. Technical white paper," 2020. [Online]. Available: https://images.nvidia.com/aem-dam/en-zz/Solutions/data-center/nvidia-ampere-architecture-whitepaper.pdf

ИСПОЛЬЗОВАНИЕ ДИСТИЛЛИРОВАННЫХ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ ТЕКСТА

Андренко К. В., Крощенко А. А.,
Головко О. В.

В статье описана дистиллированная модель на основе трансформера T5-large для анализа тональности текста, оптимизированная для образовательных приложений. Модель, созданная путем передачи знаний от больших языковых моделей, демонстрирует точность 97.22% на тестовой выборке при значительном снижении вычислительных требований. Описан процесс дистилляции, методология обучения и результаты экспериментов. Предложены сценарии применения в образовательном контексте, включая адаптивное обучение и персонализацию обратной связи. Исследование показывает перспективность использования дистиллированных больших языковых моделей в образовательных технологиях с ограниченными ресурсами.

Received 05.04.2025