UDC 004.8

**Асланова Віталіна, Олійник Володимир**

АСИСТЕНТ ПСИХОЛОГІЧНОЇ ПІДТРИМКИ НА ОСНОВІ ДОНАВЧЕНОЇ МОДЕЛІ LLAMA 3

**Aslanova Vitalina, Oliinyk Volodymyr**

PSYCHOLOGICAL SUPPORT ASSISTANT BASED ON FINE-TUNED LLAMA 3 MODEL

У статті пропонується інтелектуальний діалоговий помічник для психологічної підтримки. Система базується на сучасних технологіях обробки природної мови та машинного навчання, використовуючи бібліотеки від Hugging Face для роботи з мовними моделями. Було впроваджено алгоритми для оптимізації файнтюнінгу.

**Ключові слова:** інтелектуальний діалоговий асистент, психологічна допомога, тонке настроювання, Llama-3.

Рис.: 6. Бібл.: 5.

The paper presents an intelligent dialogue assistant for psychological support. The system is based on modern natural language processing and machine learning technologies, using libraries from Hugging Face to work with language models. Algorithms have been implemented to optimize fine-tuning.

**Keywords:** intelligent dialogue assistant, psychological assistance, fine-tuning, Llama-3.

Fig.: 6. Bibliography: 5.

**Relevance of the research topic.** The development of an intelligent dialogue assistant for psychological support has become increasingly pertinent due to the rising demand for accessible mental health resources. In today’s high-stress environment, especially in war zones, high levels of stress have manifested in increased cases of panic attacks, anxiety and post-traumatic stress disorders. Consequently, there is a critical need for systems that can offer reliable and empathetic psychological assistance. This project aims to contribute to the development of an effective and sensitive dialogue system based on fine-tuned large language model – state of the art AI and deep learning solution.

**Problem statement.** The challenge in creating a psychological support chatbot lies in its ability to understand and respond appropriately to the emotional states of users. Developing such a system requires high-quality training datasets that encompass a wide range of psychological scenarios. The primary focus is on ensuring that the model can provide relevant and empathetic responses for open questions in natural language, which is essential for good user experience, real-time interaction and high quality support.

**Analysis of recent research and publications.** Recent advancements in natural language processing and machine learning have led to the development of sophisticated dialogue systems. Techniques such as transfer learning and reinforcement learning have been applied to improve the accuracy and responsiveness of these systems. However, there remains a gap in addressing the specific needs of psychological support, where the model must be particularly adept at emotional recognition and response generation.

**Identification of unexplored aspects of the general problem.** This project focuses on leveraging ORPO (Optimised Reinforcement-based Pre-training Objective) [1], QLoRA (Quantised Low-Rank Adaptation) [2] to enhance the model’s performance in providing psychological support and PEFT (Parameter-Efficient Fine-Tuning) methods that can reduce the computational cost and memory usage while maintaining good performance [3]. By fine-tuning the large language model Llama 3 with these advanced techniques, the research aims to achieve a higher level of sensitivity and relevance in the responses generated by the chatbot.

**Target setting.** The objective is to train a model that can recognise and respond to a variety of psychological states in users, providing timely and appropriate support. The system should be capable of delivering empathetic and contextually relevant responses in real-time, integrated into a Telegram bot for ease of access.

**The statement of basic materials.** The project is divided into three main stages. The first stage involves fine-tuning the pre-trained Llama 3 model on a custom dataset tailored to psychological support scenarios. In the second stage, the model’s responses are evaluated and adjusted to ensure accuracy and empathy. The final stage includes integrating the model into a Telegram bot and conducting extensive testing to refine its performance based on user feedback.

**Model training.** The Llama 3 model was fine-tuned using a custom dataset tailored for various psychological support scenarios. The dataset was formatted to meet the expectations of the ORPO trainer, which requires three entries: 'prompt', 'chosen', and 'rejected'. The 'prompt' represents the patient's message, while the 'chosen' entry corresponds to the model's expected response to the patient's message [1]. To ensure the answers were created in a manner consistent with how a real psychotherapist would respond, an in-depth analysis of actual conversations between patients and therapists was conducted using open-source datasets from Kaggle and Hugging Face. Lastly, the 'rejected' entry was included to provide examples of inappropriate responses, guiding the model on what it should avoid. Techniques such as ORPO, QLoRA and PEFT were employed to optimise the model’s performance and reduce the computational load. PEFT selectively adjusts a small proportion of their parameters, while keeping the rest unaltered. [3] The training process involved using the Adam optimiser, sparse categorical cross-entropy loss, and accuracy metrics, with the model trained over multiple epochs to ensure robust learning.

Subsequently, the model architecture for training was defined, consisting of the following layers:

1. Base Model:

 - Input Layer: This layer accepts the tokenized text input, which represents the user's message or prompt.

2. Transformers from Hugging Face:

 - The core of the architecture is built using the Transformers library from Hugging Face, leveraging pre-trained models for natural language understanding.

The model was trained with the following hyperparameters:

* Optimizer: paged adamw 8bit
* Number of Epochs: 10
* Learning rate: 1e-5
* Batch size: 4

**Model architecture.** The model architecture for the intelligent dialogue assistant is built on the Llama 3 model, therefore consists of several key components: an input tokenizer that processes the user's text into a format suitable for the model, a transformer-based neural network that forms the core of Llama 3, and output layers that generate the response. The tokenizer breaks down the input text into tokens and ensures they are encoded correctly, maintaining the context and meaning of the user's message. At the core, the transformer model uses self-attention mechanisms to process the tokens, understanding the relationships and dependencies between words in the context of the conversation. The final output layers decode the processed tokens into coherent and contextually appropriate responses. The overall model architecture is depicted in Fig. 1.



Fig. 1. Model architecture

**Results.** The obtained results of the psychological support assistant model based on fine-tuned LLaMA 3 are presented in Graph. 2-5. These four graphs display the dynamics of the loss function, samples per second, evaluation of chosen and rejected responses, and accuracy. The graphs illustrate the model's training progress, highlighting improvements in response quality and efficiency over time. Additionally, the trends in accuracy and loss functions provide insight into the model's learning capabilities and overall performance.

Model training results



Graph. 2. Loss function



Fig. 3. Samples per second



Fig. 4. Evaluation of chosen and rejected responses



Fig. 5. Accuracy

**Experiments.** To test the developed intelligent dialogue assistant for psychological support, a bot was created using Telegram. A series of experiments were conducted to analyze the model's performance. For instance, various user queries with different emotional tones were tested. In the first experiment, text containing clear emotional elements was used, and the model responded with relevant and empathetic replies. In the second experiment, text with less clear emotional signals was used. In this case, the model still provided adequate responses but with slightly lower perceived empathy. The test results are displayed in Fig. 6.



Fig. 6. Test Results

**Conclusions.** The paper proposes an approach for creating a reliable and efficient system for psychological support. The use of advanced machine learning algorithms combined with in-depth analysis of user needs has helped to create a tool that can play a significant role in supporting the psychological health of the population. Integration with modern communication platforms has made this assistant accessible and easy to use for a wide range of users. Future research will be aimed at adding support for Ukrainian language taking into account previous achievements with low-resource languages [1] including data augmentation methods [2]. Further improvement of our assistant may include expanding the system's functionality and model performance, integration with other services and platforms to achieve even greater efficiency and accessibility of assistance to users with psychological needs.

**References**

1. Hong J., Lee N., Thorne J. Monolithic Preference Optimization without Reference Model, // arXiv preprint – 2024. – vol. arXiv:2403.07691v2 [cs.CL], 14 Mar.
2. Dettmers T., Pagnoni A., Holtzman A., Zettlemoyer L. QLoRA: Efficient Finetuning of Quantized LLMs, // arXiv preprint – 2023. – vol. arXiv:2305.14314 [cs], 23 May.
3. Han Z., Gao C., Liu J., Zhang J., Zhang S. Q. Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey, // arXiv preprint – 2024. – vol. arXiv:2403.14608 [cs], 21 Mar (v1), last revised 29 Apr 2024 (v5).
4. Oliinyk V. Low-resource text classification using cross-lingual models for bullying detection in the Ukrainian language / Oliinyk V., Matviichuk І. // Adaptive systems of automatic control, 2023. Vol. 1, №42. – P. 87-100.
5. Oliinyk V. Data augmentation with foreign language content in text classification using machine learning / Oliinyk V., Osadcha K. // Adaptive systems of automatic control, 2020. Vol. 1, №36. – P. 51-59.

**ДОВІДКА ПРО АВТОРІВ**

Олійник Володимир Валентинович – к.т.н., доцент, доцент кафедри інформаційних систем та технологій, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Oliinyk Volodymyr – Ph.D., associate professor, Department of Information Systems and Technologies, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: oliinyk.volodymyr@gmail.com

Асланова Віталіна Сергіївна – бакалавр, кафедра інформаційних систем та технологій, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Aslanova Vitalina – BS, Department of Information Systems and Technologies, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: aslanovavitalina@gmail.com

**РОЗШИРЕНА АНОТАЦІЯ**

**В.С.Асланова, В.В.Олійник**

АСИСТЕНТ ПСИХОЛОГІЧНОЇ ПІДТРИМКИ НА ОСНОВІ ДОНАВЧЕНОЇ МОДЕЛІ LLAMA 3

**Актуальність теми дослідження.** Значимість розробки чат-боту для психологічної підтримки є надзвичайно високою у сучасних умовах, де зростання психологічних навантажень та стресів вимагає нових підходів до допомоги людям. Значення такого чат-боту посилюється у світлі останніх тенденцій, які свідчать про збільшення кількості проблем із психічним здоров’ям у всіх вікових категоріях та соціальних групах. Чат-бот відповідатиме критичній потребі у доступності якісної психологічної допомоги, особливо в регіонах з обмеженим доступом до професійних психологічних послуг. Суспільне сприйняття психологічної допомоги також зазнало змін: якщо раніше звернення за психологічною допомогою часто вважалося чимось, за що варто було соромитися, то тепер все більше людей розглядають це як нормальну та необхідну частину догляду за своїм здоров'ям.

**Постановка проблеми.** Проблема створення моделі для психологічної підтримки пов’язана з відсутністю спеціалізованих датасетів для навчання у відкритих джерелах. Отже, для створення даного асистента необхідно мати набір даних, завдяки якому відбувається навчання моделі. Для того аби моделі могли надавати емпатійну, зрозумілу та логічну відповідь користувачу, набір даних повинен містити в собі велику кількість різних варіантів діалогів, які б охоплювали великий вибір тем, які може підняти користувач.

**Аналіз останніх досліджень і публікацій.** Протягом останніх років з’являється більше і більше прикладів діалогових систем, що надають психологічну підтримку. Ці реалізації підкреслюють широкі можливості застосування інтелектуальних діалогових систем у психологічній допомозі, демонструючи їх потенціал у підвищенні доступності та ефективності психотерапевтичних послуг. Тим не менш, більшість з них вузько направленні, як наприклад, для особливої групу людей для яких спеціалізована діалогова система.

**Виділення недосліджених частин загальної проблеми.** Дана стаття присвячена вивченню та аналізу підходів для створення моделі для психологічної підтримки в формі чату. Дослідження сфокусовано на вивченні оптимальних варіантах fine-tining моделі Llama 3.

**Постановка завдання.** Завданням є навчити базову модель визначати емоційний контекст та інтегрувати адаптивні реакції, що буде давати кінцевий результат інтелектуальної діалогової системи.

**Викладення основного матеріалу.** Вирішення поставленої задачі можна розділити на три основні етапи. Спочатку відбувається завантаження моделі, потім відбувається підбір отпимальних гіперпараметрів для навчання моделі і після цього - навчання на спеціально створеному наборі даних. Після цього відбувається аналіз навчання і створення графіків. Кінцевим етапом є створення чат боту в телеграмі для полегшення процесу взаємодії користувача з інтелектуальним діалоговим асистентом.

**Висновки.** В роботі запропоновано підхід до створення діалогової підсистеми для психологічної допомоги користувачам. Завдяки ретельному аналізу існуючих підходів і технологій, вдалося створити чат-бота, здатного до емпатичної взаємодії з користувачами, що сприяє зниженню емоційного напруження і покращенню їх психологічного стану. Обрані методики та технології дозволяють створити надійну та ефективну систему для психологічної підтримки. Використання передових машинних алгоритмів у поєднанні з глибоким аналізом потреб користувачів сприяло створенню інструменту, який може відігравати значну роль у підтримці психологічного здоров'я населення.

**Ключові слова:** інтелектуальний діалоговий асистент, психологічна допомога, тонке настроювання, Llama-3.