КОМП'ЮТЕРНІ НАУКИ

УДК 519.6 *http://dx.doi.org/10.30970/vam.2024.33.12349*

НАЛАШТУВАННЯ МОДЕЛІ ГЕНЕРАТИВНОГО ШІ ДЛЯ ОСВІТНІХ ЦІЛЕЙ

Б. Бурда, Н. Колос

Львівський національний університет імені Івана Франка,
вул. Університетська 1, Львів, 79000, Україна
e-mail: nadiya.kolos@lnu.edu.ua

Розглянуто різні підходи до налаштування генеративного ШІ, зокрема методи тренування моделей та оцінка ефективності таких рішень. Крім того, охоплено експериментальну частину, де розроблено та протестовано прототип освітнього чат-бота для вивчення української мови на основі генеративного ШІ. Мета цього експерименту – оцінити, наскільки такі інструменти можуть сприяти поліпшенню навчальних результатів студентів. Результати експерименту дають підстави зробити висновки щодо перспективності використання генеративного ШІ в освіті та визначити подальші напрями досліджень у цій сфері.

Ключові слова: fine-tuning, Python, генеративний ШІ, LLM.

1. Вступ

Досліджуємо можливості використання генеративного ШІ для створення бесід, що імітують реальні освітні взаємодії та сприяють підвищенню ефективності навчання. Технологія дає змогу створювати контент на основі заданих параметрів, що допомагає чат-ботам не лише відповідати на запитання, а й генерувати нові навчальні матеріали, пояснення та приклади, адаптовані під конкретні запити користувачів. Використання генеративного ШІ у чат-ботах відкриває нові горизонти для персоналізованого навчання, роблячи його більш інтерактивним та ефективним. Останні дослідження у сфері штучного інтелекту продемонстрували, що генеруючі моделі, такі як GPT-3, значно перевершують традиційні методи обробки природної мови за своєю ефективністю та універсальністю. У численних працях розглядають різні аспекти їх використання у сфері освіти, однак потенціал для розвитку все ще залишається значним.

2. Основи розробки генеративних ШІ-моделей

2.1. Використання мовних моделей у різних задачах

2.1.1. Генерація тексту

Великі мовні моделі (LLMs) демонструють видатні можливості у задачах генерації тексту. Наприклад, GPT-3, з кількістю параметрів 175 мільярдів, продемонструвала значні поліпшення у завданнях генерації тексту, таких як автоматичне написання статей, відповідь на запити, створення коду і багато інших задач. Модель GPT-3 використовує архітектуру трансформера і здатна до виконання завдань без додаткового налаштування, просто отримуючи контекстні підказки (few-shot learning), що дає змогу їй генерувати високоякісний текст, близький до людського.

2.1.2. Обробка природної мови (NLP)

Природна обробка мови (NLP) є галуззю ШІ, що зосереджується на взаємодії між комп'ютерами та людськими мовами. Вона охоплює такі задачі, як розуміння тексту, його синтез, переклад та аналіз. NLP дає змогу використовувати технології генерації тексту для створення зрозумілих і зв'язних текстів на основі введених даних.

2.1.3. Інші застосування великих мовних моделей

Великі мовні моделі також застосовуються у різних специфічних ділянках, охоплюючи:

- кодова генерація: модель Codex, розроблена на основі GPT-3, спеціально навчена для генерації коду і демонструє високі результати у створенні програмного коду на різних мовах програмування, а також у завданнях, пов'язаних з автоматизацією кодування;
- робототехніка: моделі, такі як PaLM-E, використовують LLMs для управління роботами, забезпечуючи їм здатність розуміти та виконувати складні завдання через природну мову;
- мультимодальні системи: мультимодальні LLMs обробляють не лише текстову інформацію, а й інші види даних, такі як зображення, відео та аудіо. Наприклад, модель Flamingo використовує візуально-лінгвістичні дані для виконання задач, що потребують інтеграції різних видів інформації.

2.2. Основні концепції та терміни

2.2.1. Технології генерації тексту

Технології генерації тексту є важливою частиною сучасних досліджень у галузі штучного інтелекту. Вони базуються на здатності комп'ютерів аналізувати та синтезувати людську мову, створюючи тексти, які виглядають так, ніби їх написала людина. Ці технології знаходять широке застосування в різних сферах, враховуючи автоматичне написання новин, створення творчих текстів, генерування відповідей на запити користувачів та ін. Існує кілька основних методів генерації тексту, кожен з яких має свої особливості та сфери застосування:

- моделі на основі правил (Rule-Based Models) – базуються на заздалегідь визначених правилах і структурах. Вони ефективні для простих задач, але мають обмеження у створенні складних і варіативних текстів;
- шаблонні моделі (Template-Based Models) – використовують заздалегідь підготовлені шаблони тексту з заповнюваними полями. Цей підхід дає змогу швидко генерувати тексти, але також обмежений у різноманітності та гнучкості;
- статистичні моделі (Statistical Models) – базуються на аналізі великих обсягів текстових даних і використовують ймовірнісні методи для прогнозування наступного слова або фрази в тексті. Прикладом є моделі n-грам;
- нейронні мережі (Neural Networks) – сучасні технології генерації тексту базуються на використанні глибоких нейронних мереж. Ці моделі допомагають генерувати тексти високої якості, враховуючи контекст і семантику.

2.2.2. Сучасні методи генерації тексту

Сучасні методи генерації тексту базуються на використанні великих мовних моделей (LLMs), наприклад таких як GPT-3, які демонструють значні переваги порівняно з традиційними методами.

2.2.3. GPT-3

Generative Pre-trained Transformer 3 (GPT-3) є однією з найбільших і найвідоміших моделей, розроблених компанією OpenAI. Вона складається з 175 мільярдів параметрів і використовує архітектуру трансформера. Основні переваги цієї моделі такі:

- масштабування. Завдяки великій кількості параметрів GPT-3 має велику здатність до генерації тексту з високою якістю, близькою до людської;
- Few-shot learning. GPT-3 здатна виконувати завдання, для яких вона не була спеціально навчена, використовуючи лише кілька прикладів (few-shot learning), що значно поліпшує її універсальність;
- універсальність. Модель може виконувати широкий спектр задач, враховуючи написання текстів, переклад, створення коду, і навіть генерацію візуального контенту.

2.3. Природна обробка мови (NLP)

Обробка природної мови (Natural Language Processing) – це підрозділ штучного інтелекту, займається тим, як взаємодіє мова комп'ютера та людська. Основна мета NLP – налагодити зв'язок між людською та комп'ютерною системою спілкування, де комп'ютер розуміє, інтерпретує та генерує текст так, щоб отримати з нього сенс. Дві основні категорії NLP – це розуміння природної мови та генерація природної мови.

2.3.1. Розуміння природної мови (NLU)

Розуміння природної мови (Natural Language Understanding) охоплює різні завдання, що потребують глибокого розуміння змісту тексту. Одним із ключових аспектів є аналіз тональності, який полягає у визначенні емоційної тональності тексту, будь то позитивна, негативна або нейтральна. Також важливим є витягнення сутностей (Named Entity Recognition, NER), що охоплює виявлення та класифікацію іменованих сутностей, таких як імена людей, організації та місця у тексті. Розв'язання анафор (Coreference Resolution) допомагає визначити, які слова чи фрази в тексті належать до тієї ж сутності, що є важливим для збереження контексту. Мовний аналіз (Linguistic Parsing) охоплює аналіз граматичної структури тексту, враховуючи синтаксичний і морфологічний аналіз, що допомагає краще зрозуміти його зміст. Усі ці аспекти разом сприяють глибшому розумінню тексту, що є основою для виконання завдань, таких як читання з розумінням і відповіді на запитання.

2.3.2. Генерація природної мови (NLG)

Генерація природної мови (Natural Language Generation, NLG) охоплює кілька важливих аспектів, які забезпечують створення зв'язного та змістовного тексту. Одним із них є генерація тексту, що передбачає автоматичне створення тексту на

задану тему. Машинний переклад дає змогу виконувати переклад тексту з однієї мови на іншу, що важливо для багатомовних комунікацій. Резюмування тексту спрямоване на скорочення тексту до його ключових моментів без втрати основного змісту, що допомагає швидко зрозуміти основні ідеї. Автоматичне завершення тексту охоплює пропонування завершень для початих речень або абзаців, полегшуючи написання текстів. Чат-ботів і віртуальних асистентів створюють для ведення діалогів з користувачами, відповідаючи на їхні запитання та виконуючи різні завдання, що підвищує зручність і ефективність взаємодії. Усі ці аспекти разом роблять NLG потужним інструментом для створення текстів і комунікації [1].

2.4. Попереднє налаштування моделі (Pre-training)

Попереднє налаштування – це етап, на якому модель навчається на великій кількості ненадійних даних. Мета цього етапу – навчити модель загальним мовним правилам і структурам. Зазвичай використовують великі корпуси текстів з різних джерел: книги, статті, вебсторінки. Під час попереднього налаштування модель вивчає статистичні властивості мови, що допомагає їй ефективно виконувати різноманітні задачі генерації тексту. Типовими прикладами таких моделей є GPT (Generative Pre-trained Transformer) та BERT (Bidirectional Encoder Representations from Transformers). Ці моделі використовують трансформери, які забезпечують високу якість генерації тексту завдяки здатності враховувати контекст і семантичні зв'язки між словами [2].

2.5. Фінальне налаштування (Fine-tuning)

Фінальне налаштування – це подальше налаштування попередньо навченої моделі на специфічних задачах або наборах даних. Це дає змогу моделі адаптуватися до конкретних вимог задачі та поліпшити її ефективність. Налаштування відбувається шляхом навчання моделі на специфічних наборах даних, що містять приклади завдань, які потрібно виконати. Наприклад, для задачі автоматичного написання новин модель може бути налаштована на наборі даних з новинних статей, що допоможе їй навчитися специфічним мовним конструкціям і стилю, характерному для новинних текстів. [3]

2.6. Методології оцінки моделей

Оцінка великих мовних моделей охоплює різні методології для визначення їхньої ефективності у виконанні завдань. Основні підходи до оцінки – порівняльне оцінювання, оцінка ефективності у реальних задачах та оцінка за допомогою людських аннотацій. Порівняльне оцінювання (benchmarking) має на меті порівняти продуктивність моделі з іншими моделями на стандартних наборах даних, використовуючи загальноприйняті набори даних і завдань, такі як GLUE, SuperGLUE, SQuAD для задач розуміння природної мови, та BLEU, ROUGE для задач машинного перекладу та резюмування. Наприклад, GPT-3 та BERT продемонстрували високі результати на багатьох порівняльних наборах даних, що підтверджує їхню ефективність. Оцінка ефективності у реальних задачах має на меті визначити, наскільки ефективна модель у виконанні конкретних, реальних завдань, використовуючи модель у реальних сценаріях, таких як автоматичне написання новин, створення чат-ботів або аналіз соціальних медіа. Приклади охоплюють використання GPT-3 для створення креативного контенту або автоматичного генерування

відповідей на запитання користувачів у чат-ботах. Оцінка за допомогою людських аннотацій має на меті визначити якість виходу моделі за допомогою людської оцінки, де аннотатори оцінюють тексти, згенеровані моделлю, за різними критеріями, такими як точність, зв'язність, креативність і відповідність запиту. Наприклад, оцінка якості текстів, створених GPT-3, порівняно з текстами, написаними людьми [4].

2.7. Порівняння сучасних великих мовних моделей

Серед великих мовних моделей особливо виділяються GPT-3, BERT, T5 та XLNet, які продемонстрували високу продуктивність у різних завданнях. Порівнюючи їх за допомогою методологій оцінки моделей, згаданих раніше, можна отримати такі результати. Порівняльне оцінювання на основі загальноприйнятих наборів даних GLUE та SuperGLUE (рис. 1). Усі моделі демонструють високу точність, перевищуючи 80%. GPT-3 і XLNet мали найкращі результати, наближаючись до 85% на обох наборах задач. Різниця в точності між GLUE і SuperGLUE для кожної моделі незначна. BERT і T5 трохи поступаються GPT-3 і XLNet, особливо на задачах SuperGLUE.

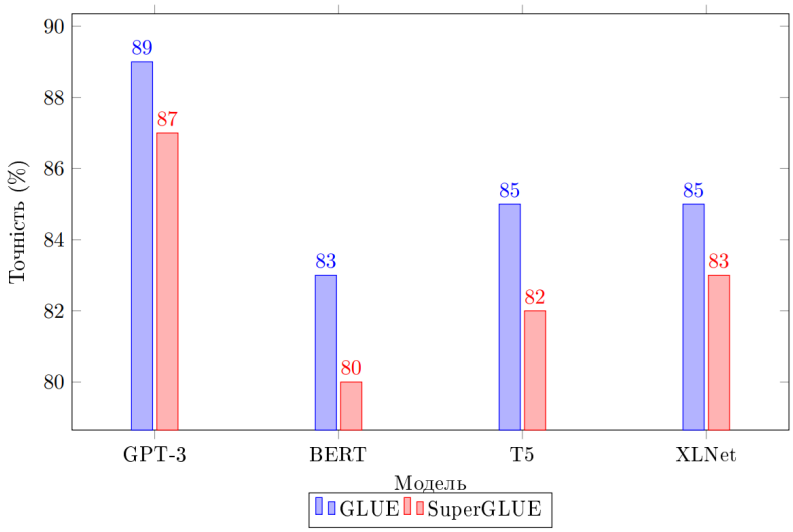

Рис. 1. Точність моделей на GLUE та SuperGLUE
Accuracy of models on GLUE and SuperGLUE

3. Практичне застосування великих мовних моделей у створенні чат-бота

3.1. Збір даних

Реалізація запланованого проєкту була розпочата зі створення необхідного набору даних для навчання моделі. Як основне джерело даних було обрано сайт zno.osvita.ua, який містить завдання зі зовнішнього незалежного оцінювання та Національного мультипредметного тесту (ЗНО/НМТ) з різних предметів. Цей

сайт є цінним джерелом інформації, бо надає доступ до завдань різних типів, які використовують під час ЗНО/НМТ. Особливий інтерес становлять завдання, які містять опис тексту, за яким йдуть питання та варіанти відповідей. Основна мета таких завдань полягає в тому, щоб учень проаналізував наданий текст і відповів на поставлені питання. На першому етапі роботи було вирішено зосередитися на завданнях цього типу, бо вони мають важливе значення для розвитку навичок аналізу тексту та критичного мислення у студентів. Завдання з вибором правильної відповіді з кількох варіантів особливо корисне для створення навчального набору даних, який може бути використаний для навчання моделі генеративного штучного інтелекту. Для збору даних з обраного ресурсу було розроблено спеціальний парсер, який автоматично збирає тексти завдань, питання до них та правильні відповіді. Зібрані дані зберігаються у форматі XSLX, що дає змогу легко їх обробляти та використовувати для подальшого навчання моделі. XLSX-файл містить такі поля: текст завдання, варіанти відповідей і правильна відповідь. Після збору даних, файл було перетворено у Json для зручнішої роботи з ним. Отже, на першому етапі реалізації проєкту було створено якісний і достатньо великий набір даних, який буде використано для навчання моделі генеративного штучного інтелекту. Цей набір даних є основою для подальших етапів роботи, зокрема для налаштування моделі, її тестування та оцінки ефективності у реальних умовах навчання.

3.2. Навчання моделі

Ми провели навчання і тестування двох моделей з бібліотеки OpenAI різних поколінь: davinci-002 та GPT-3.5 Turbo. Обидві моделі базуються на трансформерній архітектурі. Це забезпечує ефективну обробку текстових послідовностей через механізм уваги, що робить ці моделі особливо потужними для обробки природної мови. Модель davinci-002 є однією з ранніх версій, орієнтованих на генерацію високоякісного тексту. Водночас GPT-3.5 Turbo представляє новітнє покоління з поліпшеною швидкістю та оптимізацією, зберігаючи високу якість вихідного тексту.

Спочатку було прийнято рішення провести налаштування (fine-tuning) моделі davinci-002 для підвищення її ефективності у генерації тексту. Для проведення налаштування, дані були підготовлені у форматі, придатному для моделі. Зокрема, дані були перетворені у формат JSONL, де кожен елемент мав таку структуру: “prompt”: “Текст з варіантами відповідей”, “completion”: “Правильна відповідь”. Після підготовки даних вони були завантажені до сховища OpenAI. Цей процес охоплював кілька етапів:

– перевірка коректності даних. Перед завантаженням даних траба було впевнитися, що вони відповідають вимогам формату і не містять помилок;
– завантаження даних. Дані були завантажені на платформу OpenAI за допомогою відповідних інструментів і команд, що допомагає інтегрувати їх з моделлю для подальшого налаштування.

Після завершення етапу підготовки та завантаження даних був запущений процес налаштування моделі davinci-002. Цей процес складався з кількох ключових етапів:

– ініціалізація процесу навчання – визначення початкових параметрів навчання, таких як кількість ітерацій, розмір навчальної вибірки, та початкові гіперпараметри;
– навчання моделі – протягом навчання модель проходила через численні ітерації, під час яких параметри моделі оптимізувалися для мінімізації функції

втрат. Цей етап складався з обчислення градієнтів та оновлення ваг моделі для досягнення оптимальних результатів;
– моніторинг навчання – під час навчання постійно моніторилися показники ефективності, зокрема функція втрат (training loss), що допомагало відстежувати прогрес навчання та виявляти можливі проблеми.

Під час процесу налаштування було зібрано та проаналізовано значну кількість даних щодо змін у величині втрат.

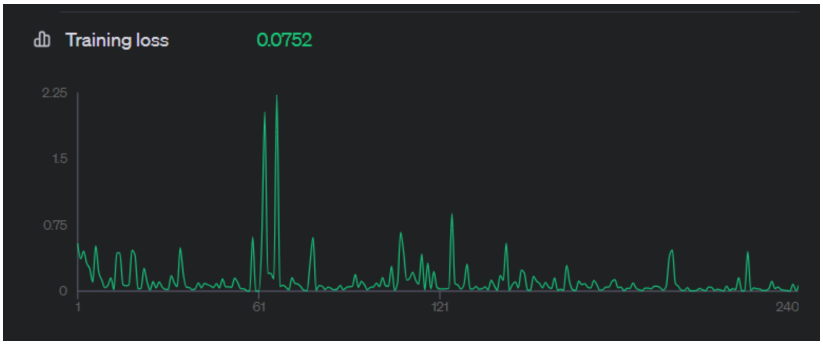

Рис. 2. Графік втрат davinci002
Loss graph of davinci002

Результати виявили, що процес fine-tuning моделі davinci-002 відбувся успішно, що підтверджується значним зниженням втрат протягом навчання та досягненням стабільно низького рівня втрат на завершальному етапі. Це свідчить про те, що модель добре адаптувалася до наданих даних і може ефективно виконувати задані завдання. Після проведення налаштування (fine-tuning) моделі davinci-002, вирішили провести аналогічний процес для моделі gpt-3.5 turbo. Процес підготовки даних для моделі gpt-3.5 turbo був аналогічний до процесу для davinci-002. Дані були перетворені у формат JSONL з наступною структурою кожного елементу: “messages”: [“role”: “system”, “content”: “Чат-бот, який генерує тексти з питаннями та дає правильні відповіді на згенеровані питання”, “role”: “user”, “content”: “Задай мені питання”, “role”: “assistant”, “content”: “Текст + питання до тексту”, “weight”: 0, “role”: “user”, “content”: “Яка правильна відповідь?”, “role”: “assistant”, “content”: “A) Правильна відповідь”, “weight”: 1] Надалі процес відбувався ідентично до моделі davinci-002. Результати навчання такі:

Графік продемонстрував успішне навчання моделі gpt-3.5 turbo з поступовим зниженням втрат до остаточного значення, близького до нуля. Модель досягла значно нижчих значень втрат на завершальному етапі, що свідчить про її вищу ефективність після налаштування порівняно з моделлю davinci-002. Це засвідчує кращу здатність моделі gpt-3.5 turbo до адаптації та навчання на специфічних даних.

3.3. Використання моделі

Після успішного налаштування моделі ми приступили до її застосування в реальному проєкті. Було створено чат-бот, який використовував згенерований моделлю текст і формулював питання до нього, а також перевіряв правильність

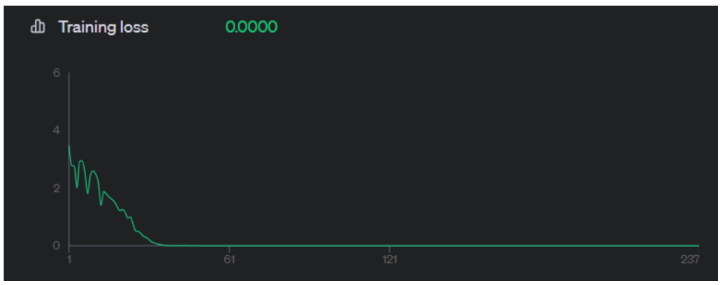

Рис. 3. Графік втрат GPT-3
GPT-3 loss graph

відповідей користувачів на ці питання. Це забезпечує можливість вивчати мову через глибоке розуміння та аналіз тексту, що позитивно впливає на ефективність навчального процесу.

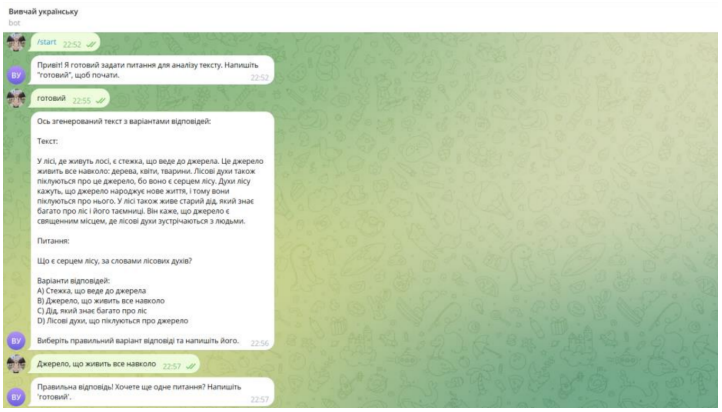

Рис. 4. Реалізований чат-бот. Поведінка, коли відповідь правильна
Implemented chatbot. Behavior when the answer is correct

4. Висновок

На підставі проведеного аналізу було розроблено та протестовано прототип чат-бота для вивчення української мови, який використовує можливості генеративного ШІ для автоматичного генерування текстів, формулювання питань, надання варіантів відповідей і перевірки правильності відповідей користувачів. Для цього було проведено налаштування (fine-tuning) двох моделей: davinci-002 та gpt-3.5 turbo, з використанням даних на основі завдань ЗНО/НМТ. Модель gpt-3.5 turbo продемонструвала кращі результати порівняно з моделлю davinci-002, що свідчить про її вищу ефективність і здатність до адаптації на специфічних даних. Отже, результати підтверджують перспективність використання генеративного ШІ для створення освітніх чат-ботів, які можуть значно поліпшити навчальний процес та сприяти підвищенню навчальних результатів студентів. Запропоновані підходи та методики можуть бути використані для подальших досліджень і впровадження

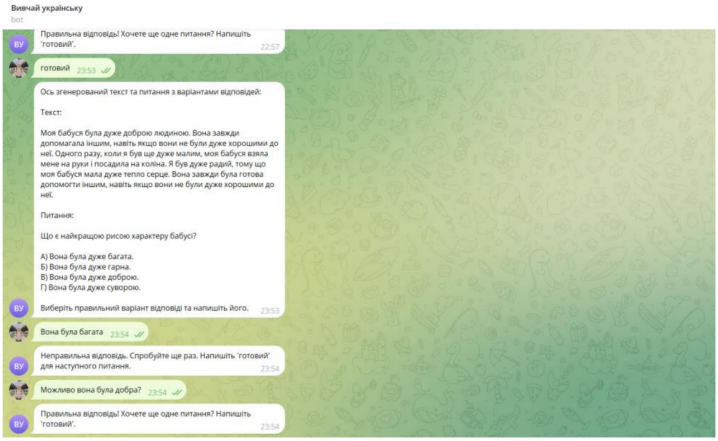

Рис. 5. Реалізований чат-бот. Поведінка, коли відповідь неправильна
Implemented chatbot. Behavior when the answer is incorrect

в освітню практику. У підсумку, розробка та впровадження освітніх чат-ботів на основі генеративного ШІ відкриває нові горизонти для персоналізованого та ефективного навчання, роблячи його доступнішим і гнучкішим для студентів.

Список використаної літератури

1. What is a Neural Network? IBM [Online]. Available: https://www.ibm.com/topics/neural-networks.
2. A Survey of Natural Language Generation (arxiv.org) (2022) // ACM Computing Surveys (CSUR).– 2022.– 1(1).– 1. [Online] Available: https://ar5iv.labs.arxiv.org/html/2112.11739.
3. Li Junyi Pre-trained Language Models for Text Generation: A Survey / Junyi Li, Tianyi Tang, Wayne Xin Zhao, Jian-Yun Nie, Ji-Rong Wen.– 2022. [Online] Available: https://ar5iv.labs.arxiv.org/html/2201.05273.
4. Evaluating Large Language Models: Methods, Best Practices & Tools Lakera – // Protecting AI teams that disrupt the world.– 2023, Aug 31.– Armin Norouzi [Online] Available: https://www.lakera.ai/blog/large-language-model-evaluation.

Стаття: надійшла до редколегії 18.08.2024
доопрацьована 04.09.2024
прийнята до друку 25.09.2024

FINE-TUNING THE GENERATIVE AI MODEL FOR EDUCATIONAL PURPOSES

B. Burda, N. Kolos

Ivan Franko National University of Lviv,
1, Universytetska str., 79000, Lviv, Ukraine
e-mail: nadiya.kolos@lnu.edu.ua

This paper examines a range of methodologies for the tuning of generative artificial intelligence (AI). The paper commences with an examination of the fundamental principles

underlying the development of generative AI models, which serve as the basis for subsequent analysis. Subsequently, the paper delineates the methodology for tuning and training models, which is pivotal for attaining optimal outcomes. Particular emphasis is placed on the criteria for evaluating the efficacy and performance of generative AI models. Moreover, the study compares contemporary large language models, enabling the discernment of their respective strengths and weaknesses, as well as the pivotal distinctions between them. This comparison facilitates an understanding of which models are the most promising for various applications. A substantial portion of the paper is dedicated to the experimental section, wherein a prototype educational chatbot for learning the Ukrainian language, based on generative AI, was developed and tested. The objective of this experiment is to evaluate the potential of such tools to enhance students' learning outcomes and influence the educational process. The outcomes of the experiment will facilitate the formulation of conclusions regarding the prospects of employing generative AI in education and the identification of prospective avenues for further research in this crucial and timely field.

Key words: fine-tuning, Python, Generative AI, LLM.