

DOI: 10.35681/1560-9189.2023.25.2.300436

УДК 004.75

С. Ф. Теленик¹, С. Д. Погорілий², А. А. Крамов², Є. А. Вовк¹

¹Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Проспект Берестейський, 37, 03056 Київ, Україна

²Київський національний університет імені Тараса Шевченка
вул. Володимирська, 60, 01033 Київ, Україна

Виявлення намірів користувача при спілкуванні з ботом

Розглянуто проблему класифікації намірів користувачів у чат-боті, зосереджуючись на використанні моделей машинного та глибокого навчання. Проведено аналіз існуючих методів і наборів даних і визначено задачу виявлення намірів через zero-shot-класифікацію за допомогою моделей Natural Language Inference (NLI). Запропоновано різноманітні підходи до класифікації і використання моделей NLI для виявлення намірів у повідомленнях користувача. Експериментально перевірено їхню ефективність на різних корпусах різних доменів, порівняно з існуючими методами, включаючи попередньо навчені моделі та великі мовні моделі, такі як GPT-3.5. Отримані результати свідчать про перспективність підвищення точності zero-shot-класифікації намірів за допомогою різноманітних моделей NLI та шаблонів гіпотез, хоча важливо враховувати, що найвищі показники точності не досягають значень попередньо навчених моделей і GPT-3.5.

Ключові слова: zero-shot-класифікації, виявлення намірів користувача, чат-бот, моделі NLI.

Вступ

Формування ІТ-середовища і ІТ-суспільства в цілому, про що давно говорять ІТ-фахівці, сприяє зростанню вимог людей до їхнього обслуговування в державних установах, банках, закладах охорони здоров'я, торговельних компаніях тощо. Щоб швидко і ефективно задовольняти потреби людей, ці установи, заклади, компанії реалізують систему шаблонно-алгоритмічних процесів, найчастіше заздалегідь описаних у формі, наприклад бізнес-процесів. Щоб, з одного боку, обслуговування було швидким і зручним для громадян (без очікування в чергах, написання офіційних листів, контактування з працівниками цих установ і закладів) і, з іншого боку, мінімізувало витрати установ і закладів на утримання досвідчених працівників,

здатних швидко допомогти клієнтові реалізувати потребу, сьогодні широко використовуються цифрові помічники. Зазвичай мова йде про чат-боти.

Створення, підтримка чат-ботів, розширення їхньої функціональності, масштабування вимагають відповідних інструментальних засобів. У праці [1] автори запропонували платформу, що інтегрує розроблення, накопичення та використання чат-ботів. Це рішення узагальнює роботу користувача з усіма месенджерами, соціальними мережами та інформаційними системами в аспектах контенту, графіки та функціональних можливостей. Розроблена платформа буде корисною провайдерам і програмним (SaaS), і платформним сервісам (PaaS), якщо допоможе усунути недоліки існуючих децентралізованих рішень і підвищити рівень якості обслуговування користувачів.

Для цього платформа повинна відповідати низці важливих вимог. По-перше, залишаються технічні та технологічні аспекти реалізації платформи, що здатна створювати, підтримувати та розвивати ефективні й корисні чат-боти.

По-друге, не зважаючи на помітний прогрес чат-ботів, клієнти побоюються їхніх можливих помилок і більш схильні взаємодіяти з реальним помічником. Це вимагає перетворення чат-боту в невтомного онлайн-помічника, здатного на ефективну взаємодію з клієнтами, розуміння потреб клієнта, надання відповідей на його запитання.

По-третє, платформа та система чат-ботів повинні належати до розмовних систем, які в процесі спілкування з клієнтом здатні розуміти користувачів і їхні потреби, надати відповідні сервіси для задоволення їхніх потреб. У цьому випадку платформа буде демонструвати природні людські здібності, не обмежуючи, а підсилюючи їх.

По-четверте, платформа збереже та підсилить такі традиційні переваги машинних систем як простоту розгортання, масштабованість, економічність, відсутність затримок в обслуговуванні клієнтів, багатомовність.

Поява великих обсягів природно-мовної інформації в електронному вигляді породжує проблему пошуку потрібної інформації, аналізу тенденцій і формування аналітичних звітів. Тому підсистема оброблення природно-мовних текстів стала важливою складовою інформаційно-аналітичних систем (ІАС), призначеною для витягування з текстів релевантної інформації і формування на її основі аналітичних звітів. Це вимагає дієвих засобів для виявлення елементів знань, інтерпретування, узагальнення, прогнозування, виявлення закономірностей на основі текстової інформації.

Але, щоб імітувати людську поведінку засоби оброблення природної мови треба доповнити засобами ефективною природно-мовної комунікації з користувачами та клієнтами. Тому чат-боти, які часто є складовими ІАС, мають бути інтелегентними, розуміти потреби клієнтів, відповідати на їхні запитання, допомагати вирішувати проблеми, ставлячи в процесі спілкування додаткові запитання. Тут можна використовувати технології розуміння природної мови (NLU — Natural Language Understanding).

При цьому технології розуміння природної мови мають безсумнівно включати *виявлення намірів користувача*, що дозволить ефективно управляти застосуванням системи чат-ботів. Як і системи оброблення природно-мовних текстів (NLP

— Natural Language Processing), вони пов'язані із використанням морфології, синтаксису, семантики, моделей і методів лінгвістики, інших галузей науки, які можна застосовувати для дослідження природної мови. Їх можна використовувати для виявлення нетривіальних намірів користувачів.

Але існує множина запитів користувачів, наміри яких можна назвати тривіальними, оскільки потреба забезпечується реалізацією відповідного бізнес-процесу, тобто запуском і виконанням одного із чат-ботів. Тут можна використовувати шаблони ситуацій, наприклад, описані за допомогою логічних моделей (предикати першого порядку, дескриптивна логіка). Шаблони становлять набір ситуативних правил. Якщо певний набір фраз підпадає під шаблон, то вважається, що ситуація розпізнана, і вона подається користувачеві з контекстом використання. Такий підхід має деякі переваги, насамперед, це відносна простота у реалізації, зменшення часу реакції системи на запит користувача. Основними недоліками підходу є вибірковість, потенційно великий набір шаблонів, вірогідністю тверджень, що витягаються.

Але поєднання технологій розуміння тексту для виявлення тривіальних і нетривіальних намірів глибоко пов'язане з багаторівневою організацією інтелекту людини, надає додаткову інтерпретацію механізмів оброблення природно-мовних текстів, імітує їхнє розуміння, відображення у зрозумілому вигляді і сприяє їхній ефективній реалізації на практиці.

Це є предметом цієї статті, однієї із циклу статей, присвячених створенню платформи розроблення, накопичення та використання чат-ботів, оскільки вона становить складний об'єкт дослідження, проектування та реалізації. З іншого боку, автори сподіваються, що отримані результати будуть важливими для розвитку технологій спілкування природною мовою та розуміння природної мови.

Проблема виявлення намірів користувача

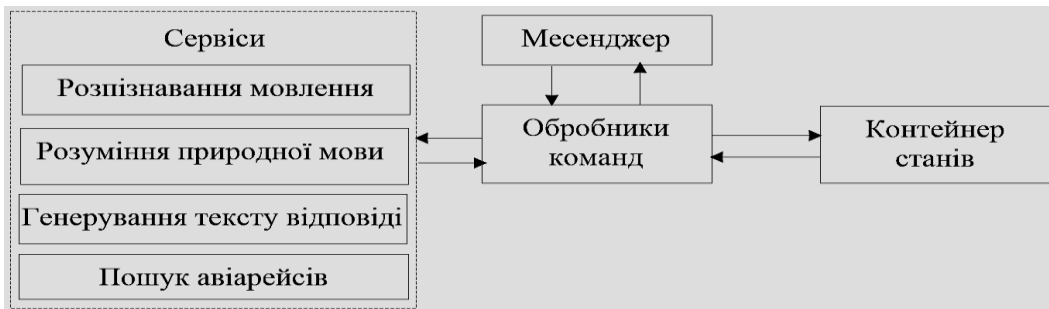
При спілкуванні користувача з чат-ботом, розробленим для підтримки користувача у виконанні певної системи дій, спрямованої на досягнення потрібного результату, надзвичайно важливим є саме зрозуміти наміри користувача. Зазначене завдання є одним із ключових кроків роботи чат-боту, добре визначених у праці [10].

На рисунку наведено архітектуру чат-бота на рівні інтеграції складових. Компонент Месенджер виконує функції обгортки джерела повідомлень. Обробники команд приймають повідомлення від постачальника повідомлень і оброблюють їх, виконуючи такі дії: делегування завдань сервісам; оновлення контейнера стану; визначення відповіді для користувача. Контейнер стану зберігає параметри стану оброблення запитів користувачів. Сервіси реалізують додаткову логіку для оброблення запитів користувачів.

На рисунку добре представлено компоненти, що пов'язані з розпізнаванням, розумінням і генерацією природномовних текстів, їхні взаємозв'язки між собою і з іншими компонентами чат-боту, а також взаємодію, яка спрямована на забезпечення очікуваної поведінки чат-бота. Уточнимо взаємодію складових чат-бота, використовуючи приклад, подібний до прикладу з праці [10]. Усне висловлювання користувача конвертується до текстового представлення сервісом Розпізнавання мовлення. Нехай це буде текст «Прошу знайти якнайраніший рейс з Києва до Львова на четвер в один бік». Сервіс Розуміння природної мови відповідає за виявлення

намірів і слотів (параметрів запити). Отримані за його допомогою дані використовуються для реалізації визначеної чат-ботом бізнес-логіки, визначення наступних станів, звертання до сервісів, визначення структурованих даних відповіді чат-бота. Сервіс генерування тексту створює відповідь природною мовою на основі структурованих даних. Для нашого прикладу намір і слоти могли би мати наступний вигляд у форматі JSON:

```
{
  "intent": "бронювання рейсу",
  "slots": {
    "destination_place": "Львів",
    "departure_place": "Київ",
    "departure_date": "четвер",
    "departure_time": "якнайраніший",
    "ticket": "в один бік"
  }
}
```



Складові чат-бота та їхня взаємодія на прикладі задачі замовлення авіаквитків

Тоді структурований запит на сервіс пошуку авіарейсів мав би включати такі складові:

```
{
  "destination_place": "Львів",
  "departure_place": "Київ",
  "departure_date": "четвер",
  "departure_time": "якнайраніший",
  "ticket": "в один бік"
}
```

Результат пошуку міг би бути представлений у такому вигляді:

```
{
  "destination_place": "Львів Сокільники",
  "departure_place": "Київ Бориспіль",
  "departure_date": "05.10.2023",
  "departure_time": "09:30",
  "ticket": "в один бік",
  "price": "2000 грн"
}
```

Текстове представлення структурованої відповіді могло би мати вигляд «Знайдено квиток в один бік до Львів Сокільниці на четвер 5 жовтня з локальним часом вильоту 09:30 ціною 2000 грн. Бронювати місце на цей рейс?».

У контексті наведеного прикладу зосередимося на виявленні наміру користувача, яке є предметом дослідження. Виявлення наміру користувача разом з пошуком слотів — фраз і ключових слів, що визначають параметри запиту користувача, реалізується сервісом Розуміння природної мови. По суті, сформовані намір і слоти в подальшому *формують запит і критерії роботи бізнес-логіки системи*. Отже, виявлення наміру користувача відповідно до заданого висловлювання є важливим проміжним етапом роботи чат-боту, що водночас впливає на інтерпретацію вхідного запиту та формування відповіді.

Оскільки чат-бот подається для реалізації месенджера у вигляді машини станів, виявлення намірів користувача дозволяє отримати та сформувати наступні елементи роботи системи:

— результат, якого користувач хоче досягнути (перевести вибрану машину станів із поточного чи початкового стану у певний цільовий стан, отримати певний інформаційний об'єкт, наприклад документ, який можна розглядати як ознаку цільового стану, добитися від чат-бота чи інших складових ІАС виконання певних дій, спрямованих на створення нових сутностей, зв'язків, зміну характеристик сутностей, зв'язків тощо);

— перехід який користувач хотів би, щоб машина станів виконала (у станах, з яких є кілька переходів у інші стани, користувач має вибрати потрібний перехід і природною мовою описати певні додаткові умови його здійснення);

— послідовність підказок, яка буде корисною користувачеві в плані вирішення проблеми, яка перед ним постала;

— вибрати потрібну машину станів з наступним формулюванням потрібного результату.

Це комплексна проблема. Природна мова має ряд характеристик, які повинна враховувати будь-яка аналітична система. Це, насамперед, багатозначність, тобто полісемія і омонімія. Ми можемо одну і ту ж саму думку виразити різними засобами. У цьому і полягає складність оброблення тексту для його розуміння технічними засобами.

Таким чином, задача виявлення намірів користувача є однією із проблем у сфері оброблення природної мови. Постановка зазначеної задачі зазвичай формується у наступний спосіб: вхідними даними є текстове представлення висловлювання користувача U ; вихідним результатом є мітка класу $L^{best} \in L$, де L — множина міток класів, які визначають наміри користувача. Отже, виявлення намірів користувача варто інтерпретувати як задачу класифікації тексту відповідно до заздалегідь визначеної множини класів. Розглянемо детальніше існуючі набори даних і методи виявлення намірів користувача, що основані на застосуванні методів машинного та глибинного навчання.

Огляд існуючих рішень

Набори даних

Корпус **ATIS (Airline Travel Information Systems)** [11–13] створений на основі аудіо-запитів користувачів до довідникової служби авіакомпанії щодо допомо-

ги планування подорожей. Транскрибування аудіозаписів і класифікацію запитів користувачів (набір даних містить 17 класів намірів) було здійснено за допомогою команди розмітки даних власноруч. У роботах [14–16] корпус ATIS було перекладено на різні мови з метою створення багатомовних наборів даних з підтримкою іспанської, турецької, французької, китайської та інших мов. Не зважаючи на використання корпусу ATIS у багатьох дослідженнях, недоліком цього набору даних є відносно невисока складність: показники точності виявлення намірів за допомогою сучасних методів складають більше, ніж 98 % [17].

На відміну від інших корпусів, набір даних **TOPv2** [18] передбачає *можливість наявності декількох намірів у висловлюванні користувача*. Наприклад, у випадку запиту користувача «якими маршрутами можна доїхати до футбольного матчу в Києві» необхідно виявити 2 наміри, визначені як наступні мітки: *get_directions* (отримання маршрутів) та *get_location* (отримання місця розташування матчу). Крім того, корпус TOPv2 містить інформацію про домен запиту (погода, годинник, музика тощо) та відповідні дочірні наміри (встановлення годинника, отримання інформації про погоду). Всього набір даних містить 8 доменів і 80 класів намірів. Недоліком корпусу є можливість наявності надлишкових класів для висловлювань з декількома мітками: у роботі [11] зазначено, що більшість міток висловлювань є дочірніми відносно інших міток чи слотів (наприклад, у наведеному вище прикладі набір *get_location* може розглядатися як дочірній відносно набору *get_directions*).

З метою розширення досліджуваної предметної області, в роботі [19] було створено розмічений корпус **Clinc-150**. Насамперед, було додано домени «банківська справа», «кухня», «робота» тощо. Корпус Clinc-150 містить 10 доменів і 150 класів намірів з метою ускладнення класифікації намірів. Однак тренування сучасних моделей (наприклад, RoBERTa) на навчальній вибірці цього корпусу дозволяє отримувати показники точності обробки тестової вибірки, які перевищують 90 % [20]. На відміну від Clinc-150, корпус **Banking-77** [21] містить лише один домен — «банківська справа», однак кількість класів намірів рівна 77. Такий підхід ускладнює процес класифікації висловлювань серед намірів, адже типи багатьох намірів є доволі схожими і належать до одного домену.

Загальними недоліками попередніх корпусів є невеликий розмір і відсутність багатомовної підтримки (крім ATIS). З метою уникнення зазначених недоліків, автори праці [22] запропонували новий корпус **MASSIVE**. Корпус MASSIVE був створений командою Amazon на основі запитів попереднього набору даних SLURP [29], які адресовані персональному домашньому роботу. Набір даних містить 18 доменів, 60 класів намірів і підтримує 51 мову. Загальний розмір корпусу складає близько 1 мільйона записів. Більшість запитів сформована у наказовій формі, що відповідає стилю взаємодії людини та голосового асистента.

Методи виявлення намірів

Як було зазначено раніше, виявлення намірів користувача відповідно до заданого висловлювання розглядається як задача класифікації. Враховуючи ефективність використання попередньо навчених багатошарових моделей на основі архітектури Transformer (наприклад, BERT, RoBERTa, DeBERTa) для вирішення різноманітних задач класифікації тексту природної мови (класифікація емоцій [23] чи

рівня токсичності [24]), зазначені моделі також використовуються для виявлення намірів. Наприклад, додаткове тренування моделі XLM-RoBERTa [25] на навчальній вибірці корпусу MASSIVE дозволило отримати значення метрики F1 (macro), рівного 0,8775 [26], а тренування моделі RoBERTa на корпусі Banking77 — значення F1, рівне 0,9442 [27].

Незважаючи на високі показники метрик, актуальною проблемою залишається можливість створення *універсального класифікатора намірів* відповідно до вимог системи. Розглянуті моделі навчаються на наборах даних, що містять фіксовану множину намірів; крім того, кожен корпус містить запити користувачів відповідно до визначеної множини доменів («годинник», «банківська справа», тощо). У випадку створення нової системи керування чат-ботом виникає проблема використання існуючих корпусів і моделей у зв'язку з наявністю унікального набору намірів — класів класифікації. З метою вирішення зазначеної проблеми може використовуватися *zero-shot-класифікація*, яка передбачає здійснення класифікації вхідних даних за допомогою використання моделі, тренування якої здійснювалося на навчальній вибірці з такими прикладами, які значною мірою відрізняються від вхідних даних. Таким чином, модель, що використовується, обробляє дані такого типу, які були відсутні у навчальній вибірці корпусу тренування моделі. Крім того, модель може бути навчена на корпусі для вирішення принципово іншої проблеми у галузі оброблення природної мови, однак набуті моделлю знання перетворюють її в ефективний інструмент для вирішення проблеми класифікації.

Міркування природною мовою у випадку нефіксованої множини класів класифікації для здійснення zero-shot-класифікації традиційно застосовуються моделі NLI (Natural Language Inference). Задача NLI полягає в наступному: на основі заданої пари речень доказ-гіпотеза (англ. *premise-hypothesis*) необхідно класифікувати цю пару відповідно до заданої категорії класів:

- 1) Entailment — гіпотеза може слідувати з доказу (наприклад, доказ — «У Петра було 6 яблук.», гіпотеза — «У Петра було більше, ніж 3 яблука.»);
- 2) Contradiction — гіпотеза заперечує факти доказу (наприклад, доказ — «У Петра було 6 яблук.», гіпотеза — «У Петра не було 6 яблук.»);
- 3) Neutral — неможливо з'ясувати, чи гіпотеза слідує, чи заперечує доказ (наприклад, доказ — «У Петра було 6 яблук.», гіпотеза — «У Петра був м'яч.»).

Для кожного наміру L_i множини класів L вхідне висловлювання користувача U представляється як доказ $Premise_i$, а гіпотеза Hyp_i — як речення на кшталт «Це речення про {намір}», де {намір} — мітка, відповідна кожному наміру L_i (наприклад, `set_alarm`). Отримані значення ймовірності класу Entailment $P(Entailment | Premise_i, Hyp_i)$ кожного наміру L_i діляться на суму відповідних значень всіх намірів L , утворюючи розподіл імовірностей $P(L|U)$ для кожного наміру $L_i \in L$:

$$P(L_i | U) = \frac{P(Entailment | Premise_i, Hyp_i)}{\sum_j P(Entailment | Premise_j, Hyp_j)}. \quad (1)$$

Відповідно до проаналізованих робіт, застосування моделей NLI для класифікації намірів у розглянутий вище спосіб не завжди є ефективним. У роботі [28] було

здійснено експериментальну перевірку ефективності використання моделі BART-MNLI (модель BART, навчена на корпусі MNLI [30]) для zero-shot-класифікації намірів корпусів Banking77 та SLURP: отримана точність склала 42,2 % і 27,5 % відповідно. Додатково в роботі [31] було розглянуто питання доцільності використання моделей NLI, навчених на корпусі MNLI, для zero-shot-класифікації намірів у корпусі Snips [32] (набір даних, зібраний із платформи Snips Voice Platform, містить 7 намірів); за допомогою зазначеної моделі NLI було отримано показник точності, рівний 77,6 %. Варто зазначити, що у розглянутих роботах експериментальні дослідження проводилися з використанням різних шаблонів речення гіпотези: «I want to {intent}.» [32] і «This example is {intent}.» [31] (в роботі [31] не зазначається шаблон гіпотези, однак, враховуючи використання моделі з ресурсу Hugging Face, можна припустити, що використовувався шаблон за замовчуванням). Отже, доцільним є додаткова перевірка впливу шаблону гіпотези на точність здійснення zero-shot-класифікації.

Окремо варто звернути увагу на навчальний корпус моделей NLI. Всі наведені експериментальні перевірки було здійснено за допомогою моделей NLI, навчених на корпусі MNLI. Однак у роботі [33] було показано наявність так званих «артефактів розмітки» (англ. annotation artifacts) у цьому корпусі: приклади класів Contradiction та Neutral мають високу кореляцію з наявністю певних слів у реченні-гіпотезі (наприклад, слова заперечення «no, never» для класу Contradiction). Наявність таких «артефактів розмітки» ставить під сумнів спроможність навченої моделі обробляти вхідні дані, що не містять відповідних слів; отже, доцільним є здійснення експериментальної перевірки ефективності застосування інших моделей NLI. Крім того, наявність ієрархічної структури «домен-намір» у деяких корпусах (MASSIVE, Clinc-150) передбачає можливість класифікації наміру в 2 послідовних етапи: спочатку визначається домен висловлювання, після чого виконується класифікація наміру серед множини намірів обраного домену. Таким чином, доцільним є здійснення порівняння ефективності застосування одноетапної і двоетапної zero-shot-класифікацій.

Враховуючи результати виконаного аналізу існуючих рішень і доцільність аналізу різних аспектів їхньої роботи, за мету експериментальної частини роботи покладемо визначення показників ефективності застосування різних моделей NLI для виявлення намірів користувача, які можна покласти в основу вибору відповідних компонентів платформи розроблення, накопичення та використання чат-ботів, що пов'язані з опрацюванням природномовного діалогу чат-бота з користувачами. Для досягнення зазначеної мети видається виконати:

- 1) аналіз доцільності застосування одноетапної і двоетапної zero-shot-класифікацій намірів;
- 2) перевірку ефективності використання моделей NLI, навчених на різних корпусах, для виконання zero-shot-класифікації намірів;
- 3) перевірку впливу шаблону гіпотези, що подається на вхід моделі NLI, на точність здійснення zero-shot-класифікації намірів.

Перейдемо до експериментальної частини виконаного дослідження, попередньо проаналізувавши та вибравши існуючі методи класифікації намірів користувача у спілкуванні із чат-ботом на основі застосування різномісних моделей машин-

ного та глибинного навчання (насамперед, на основі zero-shot-класифікації із використанням моделей NLI), наборів даних і шаблонів гіпотез.

Експериментальна перевірка ефективності виконання zero-shot-класифікації намірів

Одноетапна та двоетапна класифікації

Під *одноетапною класифікацією* мається на увазі виявлення наміру $L^{best} \in L$ висловлювання U за допомогою здійснення класифікації вхідного висловлювання на множині всіх намірів незалежно від домену:

$$L^{best} = \arg \max(P(L|U)). \quad (2)$$

Зазначений підхід класифікації висловлювання у множині всіх намірів без попереднього передбачення домену використовується у більшості сучасних існуючих методів, що дозволяє отримувати найвищі показники метрик на відповідних корпусах.

Двоетапна класифікація передбачає виявлення наміру $L^{best} \in L$ висловлювання U за допомогою виконання двох послідовних етапів.

1. Виявлення домену $d \in D$ висловлювання U , де D — множина всіх доменів корпусу.

2. Вибір наміру $L^{best} \in L^d$ серед множини намірів $L^d \subseteq L$, що є дочірніми відносно домену d , виявленого на попередньому етапі.

Наприклад, у випадку вибору домену «годинник» на першому етапі, на наступному кроці класифікація намірів висловлювання здійснюється на множині дочірніх намірів {"встановлення_годинника", "скасування_годинника"}. З метою більш детального дослідження ефективності застосування двоетапної класифікації було вирішено розглянути наступні підходи реалізації розглянутих етапів: *greedy* та *joint*.

Підхід *greedy* відповідає виконанню наведених вище етапів: спочатку здійснюється класифікація домену $d \in D$; далі виконується вибір наміру $L^{best} \in L^d$ серед множини намірів $L^d \subseteq L$, що є дочірніми відносно домену d .

$$\begin{aligned} d &= \operatorname{argmax} (P(D|U)), \\ L^{best} &= \operatorname{argmax} (P(L^d|U)). \end{aligned} \quad (3)$$

Підхід *joint* передбачає одночасне врахування ймовірності класифікації доменів і їхніх дочірніх намірів. Основна мета підходу полягає в оцінці результатів вибору водночас доменів і дочірніх намірів для формування вихідного результату. Розглянемо ймовірності вибору доменів $d \in D$ як ймовірності вибору гіпотез, що формують повну групу подій. Застосуємо формулу повної ймовірності для оцінки ймовірності вибору наміру $P(L_i | U)$ на основі сформованих гіпотез:

$$P(L_i | U) = \sum_{d \in D} P(L_i | U, d) P(d | U), \quad (4)$$

де $P(L_i | U, d)$ — імовірність наміру L_i з урахування заданих висловлювання U та домену d , тобто, ймовірність вибору наміру L_i серед множини дочірніх намірів домену L^d (у випадку $L_i \notin L^d$, $P(L_i | U, d) = 0$); $P(d | U)$ — імовірність вибору домену d серед множини доменів D . Вихідний намір L^{best} обирається відповідно до максимального значення ймовірності $P(L_i | U)$.

Моделі NLI

Як було зазначено раніше, вибір моделі NLI для здійснення zero-shot-класифікації може впливати на точність вирішення поставленої задачі. З метою дослідження впливу моделі NLI на формування вихідного результату було вирішено проаналізувати ефективність застосування моделей, навчених на різних корпусах.

1. Модель *bart-large-mnli* [34] була натренована на корпусі *MNLI* [30] — розміченому наборі даних, що містить приклади письмової та усної мови, тексти різних жанрів (художня література, політика, подорожі, телефонні розмови тощо). Зазначена модель використовувалася у попередніх дослідженнях для оцінки ефективності її застосування для zero-shot-класифікації намірів [28, 31]. Було вирішено дослідити модель *bart-large-mnli* для відтворення результатів розглянутих робіт та їхнього порівняння з результатами роботи інших моделей.

2. Тренування моделі *tals/albert-xlarge-vitaminc* [35] було здійснено на корпусі *VitaminC* [35], який був створений на основі анотації редагувань статей веб-ресурсу Wikipedia. Вибір цієї моделі обумовлений ефективністю її застосування для виявлення помилкових фактів у автоматично згенерованих анотаціях документів [36] (тобто, пошук помилкових тверджень у реченнях анотації, що була сформована зі статті). Потенційним недоліком застосування моделі *tals/albert-xlarge-vitaminc* є відсутність прикладів усного мовлення у навчальному корпусі *VitaminC* та орієнтованість моделі на визначення помилкових тверджень, оснований на редагуванні статей Wikipedia (наприклад, виявлення помилкового дня тижня, імені, дати, що може не підходити принципу zero-shot-класифікації намірів на основі зіставлення доказу та гіпотези як його узагальнення).

3. Модель *MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli* було навчено на комбінації корпусів: *MNLI* [30], *FEVER* [37], *ANLI* [38], *LingNLI* [39] та *WANLI* [40]. Крім навчання моделі на великому наборі даних різного типу, перевагою обраної моделі є високі показники метрик на корпусі *ANLI* (точність ~70%), що є складним корпусом навіть для сучасних великих мовних моделей [41] як *PaLM 2* [42] чи *GPT-3* [43].

Шаблони гіпотез

Шаблони гіпотез можуть бути одним із параметрів, що впливають на точність здійснення zero-shot-класифікації. У ході експериментального дослідження було досліджено ефективність застосування наступних шаблонів гіпотез (текстова частина {intent} означає місце вставлення мітки наміру у рядок шаблону):

1) «This example is {intent}.» (*default*) — стандартний шаблон гіпотези, що використовується у бібліотеці Hugging Face для здійснення zero-shot-класифікації.

Як було зазначено раніше, цей шаблон імовірно використовувався у роботі [28], де було отримано низькі показники точності класифікації;

2) «{intent}' is the topic of the request.» (*topic*) — шаблон гіпотези передбачає узагальнення теми запиту користувача;

3) «I want to {}.» (*want*) — шаблон представлений у наказовій формі, що відповідає взаємодії людини та робота-асистента та описує тему запиту. Цей шаблон застосовувався в роботі [31], що дозволило отримати високий показник точності на наборі даних Snips; однак корпус Snips містить лише 7 міток намірів, що спрощує здійснення zero-shot-класифікації.

Протягом формування кожного шаблону додатково було вирішено замінити всі символи міток «_» на пробіли з метою формування більш когерентного тексту гіпотези (наприклад, для мітки «alarm_set» формується гіпотеза «I want to alarm set.»).

Набори даних для дослідження

Експериментальне дослідження було здійснено на наступних корпусах:

1) *MASSIVE* — лише англійська частина тестового набору даних була обрана для аналізу. Вибір цього корпусу обумовлено наявністю ієрархічної структури «домен-намір», що може використовуватися для дослідження ефективності одноетапної і двоетапної класифікацій. Крім того, більшість запитів користувача у корпусі сформовано у наказовій формі та з урахуванням усного мовлення, що відповідає традиційному спілкуванню користувача та голосового асистента;

2) *Banking77* — набір даних було обрано для дослідження можливості застосування zero-shot-класифікації серед великого набору міток, що належать до одного домену.

Результати

Спочатку проаналізуємо результати застосування одноетапної та двоетапної класифікації для різних моделей NLI для корпусу *MASSIVE* за допомогою використання стандартного шаблону гіпотези *default*. У табл. 1 наведено отримані результати точності для зазначених вище конфігурацій параметрів дослідження. Крім того, наведено точність класифікації намірів у випадку вибору класу наміру як константи — класу, що зустрічається у тестовій вибірці корпусу *MASSIVE* найчастіше (модель *baseline*). Для всіх моделей, крім *tals/albert-xlarge-vitaminc* (що орієнтована на виявлення помилок у фактах), застосування одноетапної класифікації дозволило отримати найвищі показники точності, що може вказувати на доцільність здійснення класифікації *водночас на рівні доменів і намірів* у випадку наявності відповідної ієрархічної структури. Крім того, для заданої конфігурації параметрів (набір даних, шаблон гіпотези) найвищі показники точності очікувано отримано за допомогою моделі *MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli*, що вказує на доцільність навчання моделі NLI на комбінованому наборі даних (а саме, використання корпусу ANLI) для подальшого її використання у zero-shot-класифікації. Найнижчі значення метрики отримано у випадку застосування моделі *tals/albert-xlarge-vitaminc*, що може вказувати на складність використання моделей, що орієнтовані на перевірку фактів, для здійснення zero-shot-класифікації.

Таблиця 1. Результати точності застосування різнотипних підходів класифікації і моделей NLI на корпусі *MASSIVE* із використанням стандартного шаблону гіпотези *default*

Модель NLI	Одноетапна класифікація	Двоетапна класифікація (<i>greedy</i>)	Двоетапна класифікація (<i>joint</i>)
<i>baseline</i>	0,013	0,013	0,013
<i>bart-large-mnli</i>	0,392	0,316	0,316
<i>MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli</i>	<u>0,448</u>	<u>0,316</u>	<u>0,327</u>
<i>tals/albert-xlarge-vitaminc</i>	0,153	0,224	0,158

Багатоаспектність проблеми класифікації намірів користувачів обумовила необхідність здійснення наступного кроку експериментального дослідження. Беручи до уваги важливість ролі зміни шаблону гіпотези для ефективності методів і технологій класифікації намірів користувачів, доцільно дослідити її вплив на точність здійснення класифікації. Враховуючи результати, які наведено в табл. 1, усі розрахунки було проведено за допомогою здійснення *одноетапної класифікації*, адже найвищі значення точності було отримано за допомогою застосування саме зазначеного підходу. У табл. 2 та 3 наведено результати оцінювання точності здійснення *zero-shot*-класифікації за допомогою використання різних шаблонів гіпотези моделей NLI на тестових вибірках корпусів *MASSIVE* і *Banking77* відповідно.

Таблиця 2. Результати оцінювання точності здійснення одноетапної *zero-shot*-класифікації за допомогою використання різних шаблонів гіпотези моделей NLI на тестовій вибірці корпусів *MASSIVE*

Модель NLI	Шаблон <i>default</i>	Шаблон <i>topic</i>	Шаблон <i>want</i>
<i>bart-large-mnli</i>	0,392	0,536	0,539
<i>MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli</i>	0,448	0,546	<u>0,606</u>
<i>tals/albert-xlarge-vitaminc</i>	0,153	0,170	0,214

Таблиця 3. Результати оцінювання точності здійснення одноетапної *zero-shot* класифікації за допомогою використання різних шаблонів гіпотези моделей NLI на тестовій вибірці корпусів *Banking77*

Модель NLI	Шаблон <i>default</i>	Шаблон <i>topic</i>	Шаблон <i>want</i>
<i>bart-large-mnli</i>	0,422	0,457	0,475
<i>MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli</i>	0,449	0,511	<u>0,514</u>
<i>tals/albert-xlarge-vitaminc</i>	0,349	0,400	0,384

Отримані результати свідчать про доцільність використання різнотипних шаблонів гіпотез (крім шаблону *default*) для підвищення показників *zero-shot*-класифікації: за допомогою шаблону гіпотези *want* було підвищено точність моделі *bart-large-mnli* на 14,7 % та 5,3 % (порівняно з шаблоном *default*) на корпусах *MASSIVE* і *Banking77* відповідно; для моделі *MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli* відповідні показники підвищення точності склали 15,8 % та 6,5 %.

Варто зазначити, що текст найбільш ефективного шаблону *want* було сформовано самотужки із загальних міркувань представлення гіпотези як опису запиту до чат-бота відповідно до заданої теми (класу наміру). Однак доцільним є здійснення пошуку оптимального тексту шаблону гіпотези для отримання найвищої точності zero-shot-класифікації на заданому корпусі як окремої задачі. Задача пошуку оптимального тексту шаблону гіпотези виходить за межі поточної роботи та потребує подальших досліджень.

Узагальнюючи отримані результати, можемо стверджувати, що використання моделі NLI *MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli* та шаблону гіпотези *want* дозволило підвищити показники zero-shot-класифікації на 21,4 % і 9,2 % на корпусах *MASSIVE* і *Banking77* порівняно з комбінацією моделі *bart-large-mnli* та шаблону гіпотези *default*, що досліджувалася в попередніх роботах [28, 31]. Незважаючи на зазначені показники підвищення точності, доцільно порівняти отримані кінцеві результати з результатами застосування інших моделей класифікації намірів, а саме:

— моделей, що були натреновані безпосередньо на навчальних вибірках розглянутих корпусів: *cartesinus/xlm-r-base-amazon-massive-intent* [26] для корпусу *MASSIVE* і *philschmid/RobERTA-Banking77* [44] для корпусу *Banking77*. Надалі будемо позначати результати цих моделей для кожного корпусу як *SoTA* (state-of-the-art);

— великих мовних моделей (англ. large language models) як *GPT-3.5* [45], що передбачають можливість текстового опису задачі класифікації намірів природною мовою з подальшою генерацією вихідного результату згідно із бажаним форматом, заданим в описі. Наразі великі мовні моделі широко використовуються для вирішення низки задач галузі обробки природної мови у зв'язку з можливістю текстового формулювання задачі будь-якого типу та вирішення різнотипних завдань без попереднього навчання моделей (або ж за допомогою навчання моделей з використанням декількох прикладів — few-shot learning) [46]. Недоліком застосування великих мовних моделей є складність їхнього локального використання, що потребує значних обчислювальних ресурсів.

У табл. 4 наведено порівняльні результати класифікації найбільш ефективною комбінацією моделі NLI *MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli* та шаблону гіпотези *want* і моделей *SoTA* і *GPT-3.5* для корпусів *MASSIVE* та *Banking77* за показником точності. Результати роботи моделей *SoTA* і *GPT-3.5* було отримано з попередніх досліджень [26, 44, 47, 48]. Варто зазначити, що точність класифікації моделі *GPT-3.5* на корпусі *MASSIVE* представлена як показник точності на корпусі *SLURP*, що відповідає англomовній версії корпусу *MASSIVE* [22].

Таблиця 4. Результати точності класифікації намірів на корпусах *MASSIVE* та *Banking77* за допомогою застосування моделей *MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli*, *SoTA* і *GPT-3.5*

Модель	<i>MASSIVE</i>	<i>Banking77</i>
Модель NLI <i>MoritzLaurer/DeBERTa-v3-large-mnli-fever-anli-ling-wanli</i> , шаблон гіпотези <i>want</i>	0,606	0,514
<i>GPT-3.5</i>	0,728	0,740
<i>SoTA</i>	0,877	0,935

Точність застосування моделі *GPT-3.5* перевищує відповідні показники моделі NLI на 12,2 % і 22,6 % на корпусах *MASSIVE* та *Banking77* відповідно, що свідчить про доцільність подальшого дослідження можливості вдосконалення здійснення zero-shot-класифікації намірів за допомогою моделей NLI чи інших типів моделей [49, 50]. Враховуючи отримані результати та поточну складність локального використання великих мовних моделей, пошук альтернативних шляхів здійснення zero-shot-класифікації є актуальною задачею та потребує подальших досліджень.

Висновки

Проаналізовано існуючі набори даних і методи класифікації намірів користувача у спілкуванні з чат-ботом на основі застосування різнотипних моделей машинного та глибинного навчання. Насамперед, розглянуто проблему виявлення намірів користувача на основі zero-shot-класифікації із використанням моделей NLI. Здійснено описання та дослідження проблеми виявлення намірів користувача, виконано аналіз існуючих підходів, методів і технологій виявлення намірів користувача, моделей класифікації NLI, шаблонів гіпотез, отримано результати експериментального дослідження ефективності застосування різних моделей NLI для виявлення намірів користувача на різних наборах даних дозволило зробити такі висновки:

1) найвищі показники точності було отримано при здійсненні одноетапної zero-shot-класифікації пари «домен-намір» на різних корпусах, що може свідчити про доцільність виконання zero-shot-класифікації без її розділення на послідовні окремі етапи виявлення домену та наміру;

2) доцільним є застосування моделей NLI, тренуваних на різнотипних корпусах (насамперед ANLI), адже використання відповідної моделі дозволило отримати найвищі показники точності порівняно з іншими моделями, що застосовувались у попередніх дослідженнях виявлення намірів;

3) крім вибору моделі NLI, важливим параметром є формування шаблону гіпотези, що значною мірою може впливати на вихідний результат: після заміни тексту гіпотези було зафіксовано підвищення точності на 15,8 %; крім того, відповідне підвищення показників було отримано на всіх корпусах майже для всіх розглянутих моделей NLI;

4) незважаючи на загальний приріст точності на 21,4 % та 9,2 % на проаналізованих наборах даних, що отриманий за допомогою використання різнотипних моделей NLI та шаблону гіпотези, застосування GPT-3.5 чи моделей, попередньо навчених на навчальних вибірках корпусів, дозволяє отримати підвищення точності від 12,2 % порівняно із запропонованим підходом. Таким чином, процес здійснення zero-shot-класифікації за допомогою моделей NLI чи інших моделей NLP є актуальною задачею та потребує подальших досліджень.

1. Сергій Теленик, Гжегож Новаковський, Євгеній Вовк, Ігор Аносов. Розвиток і реалізація технології створення широкого класу застосувань типу чат-ботів на основі формальних моделей. *Наукові записки НаУКМА*. Комп'ютерні науки. 2022. Том 5. С. 97–107. ISSN 2617-3808. DOI: 10.18523/2617-3808.2022.5.97-107.

2. A suite of Text Analytics and Identity Analytics products to analyze Big Data using AI and MachineLearning-based technologies. URL: <http://www.sra.com/netowl/>
3. Leading provider of text analytics solutions: Attensity. URL: <http://www.attensity.com>
4. Stumme G., Hotho A., Berendt B. Semantic Web Mining. State of the art and future directions. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2006. No. 4. P. 124–143.
5. OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax Boris Motik, Peter F. Patel-Schneider, Bijan Parsia, eds. W3C Working Draft, 2008. URL: <http://www.w3.org/TR/2008/WD-owl2-syntax-20081202/>
6. Alexander Koller, Alex Lascarides A Logic of Semantic Representations for Shallow Parsing. Proceedings of the 12th Conference of the European Chapter of the ACL. 200. P. 30–39.
7. Павлов А.А., Теленик С.Ф. Информационные технологии и алгоритмизация в управлении. Киев: Техника, 2002. 344 с.
8. Ann Copestake, Dan Flickinger, Ivan Sag and Carl Pollard. Minimal Recursion Semantics: An introduction. *Journal of Research on Language and Computation*. 2005. No. 3. P. 281–332
9. Tore Amble. The Understanding Computer. Natural Language Understanding in Practice. Norweg, 2008. 201 p.
10. Larson S., & Leach K. (2022). A Survey of Intent Classification and Slot-Filling Datasets for Task-Oriented Dialog. arXiv preprint arXiv:2207.13211.
11. Hemphill C.T., Godfrey J.J., & Doddington G.R. The ATIS spoken language systems pilot corpus. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley*. Pennsylvania. June 24-27, 1990.
12. Hirschman L. Multi-site data collection for a spoken language corpus. In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman*. New York. February 23–26, 1992.
13. Hirschman L., Bates M., Dahl D.A., Fisher W.M., Garofolo J.S., Pallett D.S., ... & Tzoukermann E. Multi-site data collection and evaluation in spoken language understanding. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro*. New Jersey, March 21–24, 1993.
14. Upadhyay S., Faruqui M., Tür G., Dilek H.T., & Heck L. (Almost) zero-shot cross-lingual spoken language understanding. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). 2018, April. P. 6034–6038.
15. Xu W., Haider B., & Mansour S. (2020). End-to-end slot alignment and recognition for cross-lingual NLU arXiv preprint arXiv:2004.14353.
16. Dao M.H., Truong T.H., & Nguyen D.Q. (2021). Intent detection and slot filling for Vietnamese. arXiv preprint arXiv:2104.02021.
17. Wang Y., Shen Y., & Jin H. (2018). A bi-model based rnn semantic frame parsing model for intent detection and slot filling. arXiv preprint arXiv:1812.10235.
18. Chen X., Ghoshal A., Mehdad, Y., Zettlemoyer L., & Gupta S. (2020). Low-resource domain adaptation for compositional task-oriented semantic parsing. arXiv preprint arXiv:2010.03546.
19. Larson S., Mahendran A., Peper J.J., Clarke C., Lee A., Hill P., ... & Mars J. (2019). An evaluation dataset for intent classification and out-of-scope prediction. arXiv preprint arXiv:1909.02027.
20. Papers with Code: Text Classification on clinc_oos. URL: <https://paperswithcode.com/sota/text-classification-on-clinc-oos>
21. Casanueva I., Temčinas T., Gerz D., Henderson M., & Vulić I. (2020). Efficient intent detection with dual sentence encoders. arXiv preprint arXiv:2003.04807.
22. FitzGerald J., Hench C., Peris C., Mackie S., Rottmann K., Sanchez A., ... & Natarajan P. (2022). Massive: A 1m-example multilingual natural language understanding dataset with 51 typologically-diverse languages. arXiv preprint arXiv:2204.08582.
23. Loureiro D., Barbieri F., Neves L., Anke L.E., & Camacho-Collados J. (2022). Timelms: Diachronic language models from twitter. arXiv preprint arXiv:2202.03829.
24. The platform where the machine learning community collaborates on models, datasets, and applications: Toxicity Classification Model. URL: https://huggingface.co/s-nlp/roberta_toxicity_classifier
25. Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., ... & Stoyanov V. (2019). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
26. The platform where the machine learning community collaborates on models, datasets, and applications Sowanski Marcin, 2023: xml-r-base-amazon-massive-intent. URL: <https://huggingface.co/cartesinus/xlm-r-base-amazon-massive-intent>

27. Lin Y.T., Papangelis A., Kim S., Lee S., Hazarika D., Namazifar M., ... & Hakkani-Tur D. (2023). Selective in-context data augmentation for intent detection using pointwise v-information. arXiv preprint arXiv:2302.05096.
28. Clarke C., Heng Y., Kang Y., Flautner K., Tang L., & Mars J. (2023). Label Agnostic Pre-training for Zero-shot Text Classification. arXiv preprint arXiv:2305.16521.
29. Bastianelli E., Vanzo A., Swietojanski P., & Rieser V. (2020). SLURP: A spoken language understanding resource package. arXiv preprint arXiv:2011.13205.
30. Adina Williams, Nikita Nangia, and Samuel Bowman. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018. Volume 1 (Long Papers), P. 1112–1122, New Orleans, Louisiana. Association for Computational Linguistics.
31. Ma T., Yao J. G., Lin C. Y., & Zhao T. (2021, August). Issues with entailment-based zero-shot text classification. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (P. 786–796).
32. Coucke A., Saade A., Ball A., Bluche T., Caulier A., Leroy D., ... & Dureau J. (2018). Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. arXiv preprint arXiv:1805.10190.
33. Gururangan S., Swayamdipta S., Levy O., Schwartz R., Bowman S.R., & Smith N.A. (2018). Annotation artifacts in natural language inference data. arXiv preprint arXiv:1803.02324.
34. The platform where the machine learning community collaborates on models, datasets, and applications: bard-large-mnli. URL: <https://huggingface.co/facebook/bart-large-mnli>
35. Schuster T., Fisch A., & Barzilay R. Get Your Vitamin C! Robust Fact Verification with Contrastive Evidence. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021, June. P. 624–643.
36. Laban P., Schnabel T., Bennett P. N., & Hearst M.A. SummaC: Re-visiting NLI-based models for inconsistency detection in summarization. *Transactions of the Association for Computational Linguistics*. 2022. **10**. P. 163–177.
37. Nie Y., Chen H., & Bansal M. Combining fact extraction and verification with neural semantic matching networks. In Proceedings of the AAAI conference on artificial intelligence. 2019, July. Vol. 33, No. 01. P. 6859–6866.
38. Nie Y., Williams A., Dinan E., Bansal M., Weston J., & Kiela D. Adversarial NLI: A New Benchmark for Natural Language Understanding. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020, July. P. 4885–4901.
39. Parrish A., Huang W., Agha O., Lee S.H., Nangia N., Warstadt A., ... & Bowman S. (2021, November). Does Putting a Linguist in the Loop Improve NLU Data Collection?. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. P. 4886–4901.
40. Liu A., Swayamdipta S., Smith N.A., & Choi. (2022, December). YWANLI: Worker and AI Collaboration for Natural Language Inference Dataset Creation. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. P. 6826–6847.
41. Papers with Code: Natural Language Inference on ANLI test. URL: <https://paperswithcode.com/sota/natural-language-inference-on-anli-test>
42. Anil R., Dai A.M., Firat O., Johnson M., Lepikhin D., Passos A., ... & Wu Y. (2023). Palm 2 technical report. arXiv preprint arXiv:2305.10403.
43. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J.D., Dhariwal P., ... & Amodei D. Language models are few-shot learners. *Advances in neural information processing system*. 2020. **33**. P. 1877–1901.
44. The platform where the machine learning community collaborates on models, datasets, and applications: RoBERTa-Banking77. URL: <https://huggingface.co/philschmid/RoBERTa-Banking77>
45. OpenAI developer platform: Models. URL: <https://platform.openai.com/docs/models>
46. Kocoń J., Cichecki I., Kaszyca O., Kochanek M., Szydło D., Baran J., ... & Kaziemko P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*. 101861.
47. He M., & Garner P N. (2023). Can ChatGPT Detect Intent? Evaluating Large Language Models for Spoken Language Understanding. arXiv preprint arXiv:2305.13512.

48. Loukas L., Stogiannidis I., Malakasiotis P., & Vassos S. (2023). Breaking the Bank with ChatGPT: Few-Shot Text Classification for Finance. arXiv preprint arXiv:2308.14634.

49. Pogorilyy S.D., Kramov A.A. Assessment of Text Coherence by Constructing the Graph of Semantic, Lexical, and Grammatical Consistency of Phrases of Sentences. *Cybern Syst Anal.* 2020. **56**. P. 893–899. <https://doi.org/10.1007/s10559-020-00309-7>.

50. Kramov A., Pogorilyy S. Estimation of the Local and Global Coherence of Ukrainian Texts Using Transformer-Based, LSTM, and Graph Neural Networks. In: Ignatenko O. et al ICTERI 2021 Workshops. ICTERI 2021. *Communications in Computer and Information Science.* 2022. Vol 1635. Springer, Cham. https://doi.org/10.1007/978-3-031-14841-5_33.

Надійшла до редакції 01.10.2023