

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерних наук
(повна назва)

Кафедра _____ програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти _____ другий (магістерський)

Дослідження моделей та методів проєктивних методик тестування на основі
природної мови
(тема)

Виконав

студент 2 курсу, групи ІПЗм-22-1

Світенко Г.М.

(прізвище, ініціали)

Спеціальність 121 – Інженерія програмного

забезпечення

(код і повна назва спеціальності)

Тип програми освітньо-наукова

(повна назва освітньої програми)

Керівник _____ доцент каф. ІІІ, к.т.н.,

доцент Турута О.П.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ Дудар З.В.
(підпис) (прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Програмної інженерії
(повна назва)

Рівень вищої освіти другий (магістерський)
Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми освітньо-наукова програма
(освітньо-професійна або освітньо-наукова)

Освітня програма Інженерія програмного забезпечення
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ

Зав. кафедри _____
(підпис)

«29» березня 2024 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Світенку Георгію Миколайовичу
(прізвище, ім'я, по батькові)

1. Тема роботи: Дослідження моделей та методів проєктивних методик тестування
на основі природної мови

затверджена наказом університету від «29» березня 2024 р. № 250Ст

2. Термін подання студентом роботи до екзаменаційної комісії 20 червня 2024 р.

3. Вихідні дані до роботи: електронні ресурси за обраною тематикою, алгоритми обробки природної мови, великі мовні моделі, методичні вказівки до виконання кваліфікаційної роботи магістра, пояснювальна записка.

4. Перелік питань, що потрібно опрацювати в роботі: аналіз предметної області, постановка задачі, огляд існуючих методів, опис розробленої системи.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Огляд наукової та патентної літератури	29.03.2024	Виконано
2	Постановка задачі	01.04.2024	Виконано
3	Розробка моделі	05.04.2024	Виконано
4	Збір тестових даних	10.04.2024	Виконано
5	Проектування та реалізація програмної системи	23.04.2024	Виконано
6	Розробка плану проведення експериментальних досліджень	28.04.2024	Виконано
7	Проведення експериментальних досліджень	05.05.2024	Виконано
8	Аналіз отриманих результатів	10.05.2024	Виконано
9	Підготовка пояснювальної записки	25.05.2024	Виконано
10	Підготовка презентації та доповіді	08.06.2024	Виконано
11	Перевірка на плагіат та нормоконтроль	10.06.2024	Виконано
12	Рецензування	15.06.2024	Виконано
13	Архівування	17.06.2024	Виконано
14	Попередній захист	18.06.2024	Виконано
15	Допуск до захисту	18.06.2024	

Дата видачі завдання 29 березня 2024 р.

Студент _____ Світенко Г.М.
(підпис)

Керівник роботи _____ доцент каф. ІІІ, к.т.н., доцент Турута О.П.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи бакалавра: 69 с., 20 рис., 48 джерел.

ПСИХОАНАЛІЗ, ПРОЕКТИВНІ МЕТОДИКИ ТЕСТУВАННЯ, ОБРОБКА ПРИРОДНИХ МОВ, ВЕЛИКІ МОВНІ МОДЕЛІ, ІНТЕРПРЕТАЦІЯ ВІДПОВІДЕЙ.

Об'єктом дослідження є застосування підходів обробки природної мови для проведення тестування за проективними методиками.

Метою роботи є дослідження та аналіз методів обробки природної мови при проведенні тестування за проективними методиками та їх подальша реалізація.

Методи дослідження, що були використані: аналіз та синтез, методи порівняння, математичні методи.

Результатом роботи є реалізація методу автоматичної інтерпретації та оцінки результатів опитування за проективною методикою тестування з використанням засобів та підходів обробки природних мов.

PSYCHOANALYSIS, PROJECTIVE TESTING METHODS, NATURAL LANGUAGE PROCESSING, LARGE LANGUAGE MODELS, ANSWER INTERPRETATION.

The object of the study is the application of natural language processing approaches for projective testing.

The purpose of the study is to investigate and analyze natural language processing methods for projective testing and their further implementation.

Research methods used: analysis and synthesis, comparison methods, mathematical methods.

The result of the work is the implementation of a method for automatic interpretation and evaluation of survey results based on projective testing methods using natural language processing tools and approaches.

Умови публікації звіту: заява щодо самостійного виконання кваліфікаційної роботи та можливості її публікації в електронному архіві відкритого доступу E1ArKhNURE.

Я, Світенко Георгій Миколайович
(прізвище, ім'я, по батькові)

студент групи ІПЗм-22-1, здобувач вищої освіти на другому (магістерському) рівні
кафедра програмної інженерії
(повна назва кафедри)

заявляю: моя кваліфікаційна робота на тему

Дослідження моделей та методів проєктивних методик тестування на основі
природної мови,
(назва роботи)

що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу E1ArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

Дата

Підпис

ЗМІСТ

Перелік умовних скорочень	8
Вступ.....	9
1 Аналіз предметної галузі.....	10
1.1 Аналіз предметної галузі дослідження	10
1.2 Огляд патентної та наукової літератури	16
1.3 Виявлення проблем та актуалізація рішень	21
2 Науково-технічна задача	25
3 Опис теоретичних та експериментальних досліджень	26
3.1 Опис теоретичних досліджень.....	26
3.2 План проведення експериментів	42
3.3 Метрики ефективності.....	43
3.4 Збір даних та підготовка експерименту.....	44
4 Аналіз результатів дослідження	45
4.1 Результати проведених експериментальних досліджень.....	45
4.2 Аналіз результатів проведених експериментальних досліджень	45
5 Опис програмної системи.....	48
5.1 Підсистема для проведення проєктивного тестування	48
5.2 Підсистема інтерпретації та оцінювання відповідей ТАТ.....	49
6 Можливість впровадження у науковій та практичній діяльності.....	50
Висновки	51
Перелік джерел посилань	53
Перелік джерел посилань за науковими напрямками керівника та науковців кафедри програмної інженерії.....	57
Додаток А Тестові та експериментальні данні	58
Додаток Б Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ.....	59

Додаток В Слайди презентації.....	60
Додаток Г Апробація результатів кваліфікаційної роботи.....	67
Додаток Д Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008:2015	69

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

BERT – Bidirectional Encoder Representations from Transformers.

GPT – Generative Pre-trained Transformer.

LLM – Large language model, велика мовна модель.

Llama – Large Language Model Meta AI.

ML – Machine learning, машинне навчання.

NLP – Natural language processing, обробка природної мови.

TAT – Thematic apperception test, тематичний апперцептивний тест.

ВСТУП

Особистість є об'єктом дослідження багатьох наук – філософії, соціології, педагогіки, психології тощо. Психологія, як наука, розглядає особистість як складну систему, де взаємодіють різноманітні фактори – від генетичних до соціокультурних. В цьому контексті розуміння природи особистості стає ключовим завданням для вивчення психології [1]. Щоб удосконалити це розуміння, психологи вдаються до використання проєктивних методик та теорій, спрямованих на вивчення внутрішніх процесів, які лежать в основі формування і виявлення індивідуальних особливостей.

Засади сучасних методів та підходів до вивчення особистості, оцінювання її рис та якостей, були закладені наприкінці 19 та початку 20 століть, і вони донині залишаються основоположними для сучасної психології.

Застосування сучасних методів та новітніх підходів дозволяє поглиблено досліджувати когнітивні, емоційні та соціальні аспекти індивідуальності. Їхня актуальність у сьогоденні підкреслюється потребою в постійному вдосконаленні стратегій аналізу та виражається в розвитку новітніх підходів, що базуються на передових технологіях та здобутках науки таких як останні досягнення у сферах аналізу даних та обробки природної мови.

Ці засади становлять основу для розкриття найглибших та найскладніших аспектів людської індивідуальності, що є важливим кроком у розумінні та підтримці психічного здоров'я та особистісного розвитку [2]. А також створюють підґрунтя для впровадження на практиці більш ефективних та комплексних підходів та методів психотерапії, що має позитивний вплив на психосоціальний стан сучасної людини.

Отже, метою дослідження є синтез методу автоматичної інтерпретації та оцінки результатів опитування за проєктивною методикою тестування з використанням засобів та підходів обробки природних мов.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗИ

1.1 Аналіз предметної галузі дослідження

Одним із важливих напрямків сучасної психології є психологія особистості – галузь психологічних знань, що спрямована на поглиблене розуміння особистості [3]. В рамках цієї галузі, дослідники вивчають внутрішні процеси та реакції, що формують унікальний характер кожної людини та аналізують структурні та функціональні аспекти особистості, включаючи емоції, мотивацію, свідомість, поведінку та когнітивні процеси. Психологія особистості вивчає рушійні сили і умови розвитку особистості, періодизацію розвитку людини і її індивідуальності, індивідуальні властивості особистості та їх роль у розвитку особистості [4], структуру особистості і різні методичні підходи до її вивчення, теорії особистості тощо.

1.1.1 Структура особистості у психології

Структура особистості складається з рис, що є відносно стабільними характеристиками, вони визначають поведінкові тенденції людини. До неї належать самооцінка, що є суб'єктивним сприйняттям і оцінкою людиною своєї цінності та здібностей та самоідентичність, що визначає унікальність та відмінність людини від інших. Функції особистості включають мотивацію, що спонукає людину до дії та визначає її поведінку; емоції, які є афективними станами та впливають на думки та поведінку; і когніції, що включають: мислення, пам'ять, увагу та інші психічні процеси, що визначають свідоме і несвідоме пізнання світу.

Виділяють кілька основних теорій особистості. Психоаналітична теорія, запропонована З. Фрейдом [5], розглядає особистість як результат конфліктів між трьома основними компонентами: Ід, Его і Суперего. Ід представляє несвідомі інстинкти та прагнення, що шукають негайного задоволення. Его виконує роль раціонального посередника, що балансує між бажаннями Ід і вимогами реальності. Суперего являє собою внутрішні моральні стандарти і соціальні норми, які регулюють поведінку через відчуття провини та сорому. Основна ідея цієї теорії

полягає в тому, що внутрішні конфлікти між цими компонентами формують особистість та її поведінку.

Гуманістична теорія, представлена К. Роджерсом і А. Маслоу, акцентує увагу на прагненні людини до самореалізації та особистісного зростання [6]. Роджерс вважав, що люди мають вроджену тенденцію до розвитку та самовдосконалення, а важливою умовою для цього є наявність безумовної позитивної оцінки. Маслоу розробив ієрархію потреб людини, в якій самореалізація знаходиться на вершині, припускаючи, що людина повинна задовольнити базові потреби (фізіологічні потреби, потреби в безпеці, потреби в любові й прихильності, потреби в визнанні та оцінці) перед тим, як досягне самореалізації. Гуманістична теорія підкреслює важливість суб'єктивного досвіду і вільної волі в розвитку особистості.

Теорія рис, розроблена Г. Олпортом і Р. Кеттелом, стверджує, що особистість складається з певних рис, що є відносно стабільними характеристиками. Олпорт запропонував концепцію, що включає три рівні рис: кардинальні риси (основні риси, які домінують в особистості), центральні риси (основні риси, що описують людину), і вторинні риси (рис, що проявляються в певних ситуаціях) [7]. Кеттел розробив 16-факторну модель особистості, яка включає основні риси, що визначають поведінкові тенденції людини [8]. Ця теорія дозволяє кількісно вимірювати особистісні характеристики і використовувати їх у діагностиці та прогнозуванні поведінки.

Теорія соціального навчання, представлена А. Бандурою, підкреслює роль соціального середовища та спостереження у формуванні особистості [9]. Бандура вважає, що люди вчаться нових поведінкових моделей через спостереження за іншими (моделювання) і через взаємодію з навколишнім середовищем. Він також розробив концепцію самоефективності, що описує віру людини у власну здатність досягати поставлених цілей. Ця теорія наголошує на важливості взаємодії між особистістю, поведінкою і середовищем (тріадичний взаємодійний детермінізм).

Когнітивна теорія, розроблена А. Беком і А. Еллісом, акцентує увагу на ролі когнітивних процесів у формуванні особистості. Бек створив когнітивну модель депресії, в якій зазначено, що негативні автоматичні думки, когнітивні спотворення

та дисфункціональні переконання, впливають на емоції і поведінку людини [10]. Елліс розробив раціонально-емоційно-поведінкову терапію, яка наголошує на зміні ірраціональних думок і переконань для поліпшення емоційного стану і поведінки. Когнітивна теорія фокусується на тому, як мислення і переконання впливають на психологічне здоров'я і розвиток особистості.

Кожна із зазначених теорій пропонує унікальний підхід до розуміння особистості, її структури і функцій, а також методів її дослідження та розвитку. Кожна з теорій має свої сильні та слабкі сторони, і часто використовується в поєднанні з іншими підходами для досягнення більш всебічного розуміння особистості.

1.1.2 Методи дослідження психології особистості

Психологія особистості використовує широкий спектр методів, щоб дослідити та зрозуміти індивідуальні відмінності. Їх можна узагальнити за наступними категоріями: оцінювальні методи, методи спостереження, дослідницькі методи та експериментальні методи.

Серед оцінювальних методів одним з найпоширеніших підходів є тестування, що включає стандартизовані опитувальники та проєктивні тести. Особистісні опитувальники, де респонденти самотійно оцінюють свої риси, поведінку та емоційні стани, такі як MMPI (Minnesota Multiphasic Personality Inventory) [11] або NEO-PI-R (Revised NEO Personality Inventory) [12], використовуються для вимірювання широкого спектра особистісних рис і поведінкових тенденцій. Недоліком опитувальників є те, що вони можуть бути піддані суб'єктивності та соціальній бажаності респондента тесту. На противагу особистісним опитувальникам, де непередбачена пряма взаємодія дослідника з респондентом, інтерв'ю є методом збору інформації, що передбачає безпосередню взаємодію дослідника з респондентом. Інтерв'ю можуть бути структурованими, напівструктурованими або неструктурованими. Вони дозволяють дослідникам отримати глибоке розуміння особистих переживань, думок і почуттів респондента. Одним із форматів проведення дослідження у вигляді інтерв'ю є проєктивні тести,

такі як тест Роршаха або ТАТ, які пропонують респондентам інтерпретувати неоднозначні стимули, що дозволяє отримати інформацію про їхні несвідомі аспекти особистості.

Метод спостереження дає змогу дослідникам отримувати інформацію про поведінку особистості в природних або лабораторних умовах. Цей метод часто застосовується для вивчення соціальної поведінки та взаємодії з іншими людьми. Це може бути пряме спостереження, коли дослідник безпосередньо спостерігає за поведінкою респондента, або непряме, коли використовуються відеозаписи або звіти інших осіб.

До дослідницьких методів відносять біографічний метод, що включає вивчення життєвого шляху та особистих документів, таких як автобіографії, щоденники або листи респондента та лонгітюдні дослідження, які передбачають тривале спостереження за однією групою респондентів протягом значного періоду часу. Такі методи допомагають зрозуміти, як життєві події, досвід і соціальні умови впливають на формування особистості, а також дозволяють відстежувати зміни в особистості та поведінці респондентів, а також виявляти фактори, що впливають на їх розвиток [13].

Експериментальні методи передбачають контрольовані умови дослідження, де дослідник може маніпулювати змінними і спостерігати їх вплив на поведінку респондента. Цей метод дозволяє встановлювати причинно-наслідкові зв'язки між змінними.

Кожен з представлених методів володіє унікальними перевагами та обмеженнями. Їх комплексне використання дозволяє отримати всебічне та глибоке розуміння особистості та її розвитку. Завдяки різноманіттю теорій та підходів, психологія особистості постійно збагачує розуміння людської природи та сприяє розвитку психологічної практики.

Однак, в рамках клінічної психології та психотерапії, найбільш поширеними та популярними інструментами залишаються особистісні опитувальники та проєктивні тести [14]. Ці методи дають можливість психологу швидко та

ефективно оцінити різні аспекти особистості, виявити потенційні проблеми та розробити індивідуальний план діагностики та лікування.

1.1.3 Особистісні опитувальники та проєктивні методики тестування

Особистісні опитувальники є одним з основних інструментів, що використовуються в психології для вимірювання та оцінки різних аспектів особистості. Вони являють собою стандартизовані набори питань або тверджень, на які респонденти відповідають, щоб розкрити свої особистісні характеристики, поведінкові тенденції та емоційні реакції.

Серед основних видів опитувальників виділяють опитувальники на основі теорії рис, клінічні опитувальники, проєктивні тести та самооцінювальні опитувальники.

Представником опитувальників на основі теорії рис є тест NEO-PI-R, який дозволяє оцінювати п'ять основних факторів особистості, відомих як модель «Big Five»: екстраверсія, доброзичливість, сумлінність, емоційна стабільність та відкритість до досвіду. Кожен з цих факторів розділений на шість підшкал, що дозволяє детально оцінити різні аспекти кожного фактора. NEO-PI-R забезпечує комплексний підхід до оцінки особистісних характеристик і дозволяє отримати глибоке розуміння індивідуальних відмінностей.

Тест MMPI є одним з найбільш широко використовуваних і досліджених клінічних опитувальників для оцінки психопатології та особистісних характеристик. Він включає кілька шкал, спрямованих на виявлення різних психічних розладів і особистісних рис, таких як депресія, шизофренія, паранойя та інші. MMPI складається з понад 500 тверджень, на які респонденти відповідають «правда» або «неправда». Його результати допомагають психіатрам в діагностиці та розробці планів лікування, а також можуть бути використані в юридичних і дослідницьких контекстах. MMPI забезпечує надійні та валідні результати, що робить його важливим інструментом у психологічній практиці.

Результати опитувальників є надійними та зручними в порівнянні, що дозволяє проводити об'єктивну оцінку особистості. Вони є ефективними, оскільки

дозволяють швидко збирати дані від великої кількості респондентів, роблячи їх зручним інструментом для досліджень та практичного застосування. Також, різноманітні види опитувальників дозволяють оцінити широкий спектр особистісних рис, емоційних реакцій та поведінкових тенденцій, надаючи комплексне уявлення про індивідуальність.

До недоліків особистісних опитувальників можна віднести такі фактори, як вплив соціальної бажаності, коли респонденти можуть свідомо або несвідомо спотворювати свої відповіді, щоб відповідати очікуванням суспільства або дослідника, що може призвести до неточної оцінки. Також існує суб'єктивність інтерпретації результатів, яка може залежати від досвіду та упереджень дослідника, що робить інтерпретацію дещо суб'єктивною. Культурні відмінності також можуть впливати на валідність інтерпретації результатів, оскільки норми та очікування, що стосуються особистісних рис, можуть відрізнятися в різних культурах. Деякі опитувальники потребують ретельної підготовки та знань для правильного тлумачення результатів, що може обмежувати їх доступність для ширшого кола користувачів.

Проективні методики тестування, зокрема тести Роршаха, ТАТ, метод фотографій природних об'єктів та інші, визначаються як спеціальний вид психологічних тестів, спрямованих на дослідження непрямо виражених аспектів особистості, психологічних процесів та ставлення особистості до різноманітних ситуацій [15]. Вони ґрунтуються на припущенні, що проекція (перенесення власних почуттів, думок та почуттів на інших) може допомогти розкрити ті психологічні характеристики, які залишаються скритими або непрямими.

Ці методики використовуються в різних сферах, таких як психологічна діагностика, клінічна практика, психологічне дослідження, надаючи можливість розкрити різноманітні аспекти психологічної та особистісної інформації про особу. Проте, незважаючи на значні можливості в дослідженні індивідуальних проявів особистості, проективні методи мають свої обмеження.

Суттєвим недоліком є невідповідність вимогам психометричних інструментів, оскільки можливий вплив на відповіді людини визначається багатьма

ситуаційними факторами – інструкціями, особистістю експериментатора, стимулами, тимчасовими станами суб'єкта тощо. Саме тому використання проєктивних методів вимагає широкої теоретичної підготовки психолога та практичного досвіду в реалізації конкретних методик.

Існує проблема надійності та валідності проєктивних тестів, оскільки у процесі їх застосування спостерігається нестабільність результатів та суперечливість інтерпретацій даних [16]. Проте, незважаючи на ці обмеження, проєктивні методи залишаються цінним інструментом, оскільки сприяють виявленню внутрішніх аспектів суб'єкта, які важко виразити безпосередньо. Вони допомагають зрозуміти складні властивості особистості, які не завжди піддаються точному вимірюванню або оцінці.

1.2 Огляд патентної та наукової літератури

Стаття «Let's Pretend: Projective Methods Reconsidered» Д. В. Рука розглядає історію та розвиток проєктивних методів у дослідженнях поведінки та їхнє сучасне переосмислення [17]. Проєктивні методи, які з'явилися наприкінці XIX століття, набули популярності в період між 1940 і 1960 роками, особливо в клінічній і дитячій психології, а також у соціологічних та антропологічних дослідженнях.

Автор підкреслює, що після Другої світової війни проєктивні техніки поширилися з поведінкових наук до рекламних агентств і фірм з маркетингових досліджень, де вони застосовувалися у розробці нових продуктів і вивченні споживчих потреб. Проєктивні методи дозволяють дослідникам отримати доступ до підсвідомих думок і почуттів респондентів через інтерпретацію їхніх реакцій на неоднозначні стимули, такі як малюнки або оповідання.

Д. В. Рук наголошує на необхідності переосмислення ролі проєктивних методів у сучасних дослідженнях. Хоча ці методи мають свої обмеження, вони залишаються цінними інструментами для глибокого розуміння психологічних аспектів поведінки споживачів. Автор обговорює сучасні підходи до застосування проєктивних методів, включаючи інтеграцію нових технологій, що можуть підвищити їхню ефективність та точність.

У статті «Psychological assessment: Future challenges and progresses», автор Р. Фернандес-Бальєстерос, досліджує досягнення у сфері психологічної оцінки завдяки інтеграції комп'ютерів та інформаційних технологій [18]. Основна увага приділяється тому, як ці технології можуть вдосконалити як об'єктивні, так і проєктивні техніки, що використовуються в психологічному тестуванні.

Автор виділяє покращення у зборі, аналізі та інтерпретації даних, які стали можливими завдяки технологічному прогресу, а також потенційні майбутні розробки в цій галузі. Незважаючи на переваги, також обговорюються етичні аспекти та виклики, пов'язані зі збільшеним використанням технологій у психологічних оцінках, включаючи питання конфіденційності, безпеки даних та необхідність збереження людського контролю при інтерпретації результатів. Інтеграція комп'ютерних наук у особистісні опитувальники та тести висвітлюється як значна область прогресу, пропонуючи більш складні та точні інструменти для психологів для оцінки та розуміння складних психологічних рис та станів.

У статті «Psychological test usage with adolescent clients: Survey update» розглядаються сучасні практики використання психологічних тестів серед психологів, які працюють з підлітками [19]. Основна увага приділяється тому, як фахівці використовують різні види тестів, включаючи як об'єктивні, так і проєктивні методи, та яку роль у цьому процесі відіграють комп'ютерні технології. У статті підкреслюється що саме проєктивні методи залишаються популярними серед фахівців, хоча їх використання часто вимагає високого рівня підготовки та знань. Зазначається, що використання комп'ютерів дозволяє психологам не тільки збирати та аналізувати дані більш ефективно, але й підвищувати точність інтерпретацій відповідей респондентів. Завдяки цьому психологи можуть краще зрозуміти поведінкові та емоційні стани підлітків.

Автори «The role of information and communication technologies in using projective techniques as survey tools to meet the challenges of bounded rationality» підкреслюють, що традиційні підходи до досліджень, які використовують стандартні кількісні опитування та підтримуються якісними методами, такими як напівструктуровані інтерв'ю, часто не здатні виміряти питання, що виходять за

межі звичайного розуміння [20]. Вони пропонують підхід, який полягає в тому, що виклики обмеженої раціональності можна частково вирішити шляхом включення проєктивних технік у підхід «змішаних методів».

Стаття акцентує увагу на тому, як інформаційні та комунікаційні технології можуть підсилити застосування проєктивних технік у дослідженнях. Використання інформаційних та комунікаційних технологій дозволяє розширити можливості збору та аналізу даних, роблячи процес більш ефективним і точним. Це включає застосування комп'ютерних програм та онлайн-інструментів для проведення опитувань та аналізу результатів, що дозволяє отримати більш глибоке розуміння психологічних аспектів поведінки респондентів. Використання подібного підходу має на меті подолання обмежень традиційних методів одночасно підвищуючи валідність і надійність отриманих даних.

У статті «Using the projective thematic apperception test for automatic personality recognition in texts» досліджується можливість застосування проєктивних технік, і зокрема ТАТ, разом із методами та підходами комп'ютерних наук для автоматичного розпізнавання особистісних рис у текстах [21]. ТАТ традиційно використовується для оцінки психологічного стану та особистісних характеристик на основі історій, які респонденти складають у відповідь на неоднозначні зображення. У цьому дослідженні розглядається можливість застосування алгоритмів машинного навчання для аналізу текстів, створених за допомогою ТАТ, з метою автоматичного визначення особистісних рис.

Метою дослідження є автоматизація процесу інтерпретації відповідей респондентів ТАТ, що зазвичай вимагає значних зусиль з боку досвідчених психологів та довготривалого навчання і підготовки спеціалістів. Автоматизація цього процесу може зробити оцінку більш ефективною та доступною. Тому в рамках дослідження автори розглядають застосування різних методів ML та NLP для аналізу текстових відповідей респондентів.

Результати дослідження показують, що використання алгоритмів машинного навчання для аналізу текстів, створених за допомогою ТАТ, може бути ефективним інструментом для автоматичного розпізнавання особистісних рис. Це відкриває

нові можливості для застосування комп'ютерних технологій у психологічній діагностиці та оцінці, роблячи їх більш доступними та точними. Водночас автори наголошують на необхідності продовження досліджень за цим напрямком з метою валідації отриманих результатів.

У роботі «The new technologies in personality assessment: a review» автори розглядають сучасні підходи до оцінки особистості, що включають нові технології [22]. Стаття ділиться на п'ять основних блоків: big data, носимі технології, гейміфікацію (gamification), відео-резюме та автоматизоване тестування особистості. Автори детально описують кожен з цих підходів, зокрема їх використання для оцінки особистості та які переваги зазначені методики пропонують у порівнянні з традиційними методами психологічних наук.

Так наприклад, застосування підходів роботи з big data та алгоритмів їх подальшої обробки дозволяє зібрати значні обсяги інформації про поведінку індивідуумів, що може бути корисним для точнішої оцінки особистісних рис.

Носимі технології, такі як смарт-годинники та фітнес-трекери, дозволяють збирати дані про фізичну активність і фізіологічні показники, що також можуть бути використані для оцінки особистості.

Гейміфікація процесу тестування включає інтеграцію ігрових елементів у звичні сценарії психологічних тестів, що може зробити процес більш привабливим для респондентів та підвищити їх мотивацію.

Відео-резюме дозволяють отримати більше інформації про невербальну комунікацію та поведінку респондентів, що також доповнює традиційні підходи до оцінки. Автоматизоване тестування особистості використовує алгоритми машинного навчання для аналізу даних та інтерпретації результатів, що може зробити процес більш ефективним і менш затратним.

Автори статті вказують, що на даний момент немає достатньо доказів, які б підтверджували психометричні властивості (тобто надійність та валідність) нових підходів до оцінки особистості. Однак, також зазначається, що кількість досліджень у цій галузі зростає, що свідчить про можливість встановлення ефективності та надійності цих методів у майбутньому. Підсумовуючи, дослідники

висловлюють думка, що хоча поки що доказів недостатньо, є підстави вважати, що нові методи мають потенціал стати важливими інструментами для оцінки особистості, коли буде проведено більше досліджень.

Стаття «Application of Imagery Communication Psychotherapy in Projective Test» досліджує використання психотерапії образної комунікації (Imagery Communication Psychotherapy, ICP) у проєктивних тестах [23]. ICP, розроблений як ідейний послідовник ТАТ, використовує образи як засіб для глибокого спілкування, що дозволяє досліджувати підсвідомі аспекти особистості. Метод передбачає, що респонденти створюють історії на основі неоднозначних зображень, які потім аналізуються для виявлення глибоких психологічних процесів і конфліктів.

Автори статті описують, як сучасні інформаційні та комунікаційні технології можуть підсилити цей підхід. Використання комп'ютерних алгоритмів для аналізу текстів, створених під час тестування, дозволяє автоматизувати процес інтерпретації та робити його більш ефективним і точним. Це допомагає забезпечити більш глибоке розуміння особистісних рис і психологічного стану респондентів, що в свою чергу підвищує точність і доступність подібних психологічних оцінок.

У статті «Computerized clinical psychological test interpretations: Unvalidated plus all mean and no sigma» автор Дж. Матараццо аналізує потенціал та обмеження комп'ютеризованих інтерпретацій клінічних психологічних тестів [24]. Автор зазначає, що хоча ці технології мають значний потенціал для покращення роботи фахівців у сфері охорони здоров'я, наразі вони не відповідають навіть базовим науковим критеріям валідації.

Матараццо підкреслює, що комп'ютеризовані інтерпретації можуть стати цінним інструментом для клініцистів, але лише за умови, що їх використовують фахівці, які пройшли відповідну підготовку. Автор критикує існуючі системи за те, що вони надають середні значення без врахування розсіювання даних, що може призвести до помилкових висновків і некоректних інтерпретацій.

Основною проблемою, на думку автора, є відсутність належної валідації подібних технологій, що ставить під сумнів їхню надійність та точність. Матараццо наголошує на необхідності проведення додаткових досліджень для встановлення наукової обґрунтованості цих методів перед тим, як вони будуть широко впроваджені у клінічну практику.

Підсумовуючи, дослідник робить висновок, що хоча комп'ютеризовані інтерпретації психологічних тестів мають значний потенціал для майбутнього, їх слід використовувати обережно і тільки як допоміжний інструмент під наглядом кваліфікованих фахівців. До тих пір, поки не буде доведено їх наукову валідність, ці технології не можуть замінити традиційні методи оцінки.

1.3 Виявлення проблем та актуалізація рішень

Психологія особистості, вивчаючи індивідуальні особливості та їх вплив на поведінку, є динамічною та багатогранною наукою, що використовує різноманітні теоретичні та методологічні підходи для дослідження особистості. Отриманні знання дають змогу людині не лише глибше зрозуміти себе та оточуючих, але й дозволяють психіатрам розробити ефективні методи діагностики та терапії психологічних розладів, а також персоналізувати підходи в освіті, менеджменті та інших сферах.

Технічний прогрес значно розширив можливості психологів, надавши їм нові інструменти для дослідження та роботи з людьми. Використання комп'ютерів та програмного забезпечення дозволяє вирішувати велику кількість прикладних задач серед яких:

- автоматизація рутинних завдань, що звільняє час психологів для більш творчої та продуктивної роботи;
- збір та аналіз великих обсягів даних, що допомагає виявляти приховані закономірності та проводити складні статистичні дослідження;
- розробка та застосування нових методів дослідження і терапії. Прикладом може бути використання комп'ютерних симуляцій та віртуальна реальність для вивчення поведінки людей у різних ситуаціях;

– персоналізація терапії. Так, застосування комп'ютерних технологій на кшталт штучного інтелекту та машинного навчання дозволяє розробляти індивідуальні плани терапії, що враховують унікальні особливості кожного пацієнта;

– зробити психотерапію доступнішою. За рахунок використання онлайн-платформ та чат-ботів своєчасна та якісна психологічна допомога стає доступнішою для людей, які живуть у віддалених місцевостях або мають обмежені можливості.

Серед усіх зазначених застосувань комп'ютерних технологій у галузі психології найбільшу поширеність здобули чат-боти психологічної підтримки [25]. Вони використовують спеціально підготовлену базу питань та відповідей, а також методику NLP – виокремлення сутностей (Entity recognition), що дозволяє автоматично збирати та аналізувати інформацію про респондента, його потреби і проблеми. Залежно від складності питання, чат-бот може відповісти самостійно або передати сигнал людині психологу для відповіді на більш складні питання і ситуації.

Використання чат-ботів дозволяє одному спеціалісту надавати підтримку багатьом клієнтам майже одночасно, знижуючи вартість послуг з психологічної підтримки. Це робить якісну психологічну допомогу доступною для людей з менш захищених верств населення і розширює клієнтську базу психологів-спеціалістів.

Так прикладом може бути платформа Wysa [26], яка використовує метод когнітивно-поведінкової терапії (КПТ) та пропонує щоденник, відстеження настрою, техніки релаксації, ігри та історії. Це робить його корисним інструментом для людей, які хочуть самостійно впоратися з емоціями та стресом.

На противагу, чат-бот Woebot [27] поєднує метод КПТ з використанням штучного інтелекту, який дозволяє персоналізувати користувацький досвід, пропонуючи інтерактивні вправи, ігри, історії, відстеження настрою та техніки релаксації покликані для боротьби з стресом, тривогою та депресією.

Чат-бот Yourer [28], окрім зазначеного раніше функціоналу застосунків конкурентів, також надає доступ до терапевтів 24/7 та пропонує самостійні

програми з КПТ, відстеження настрою та техніки релаксації такі як глибоке дихання та медитація.

Іншим поширеним використанням комп'ютерних технологій у психології є застосунки, що полегшують працю психотерапевта шляхом автоматизації рутинних задач та процесів таких як ведення картки пацієнта, аналіз історії хвороби чи спрощення проведення терапії. Прикладом таких застосунків можуть бути EMDR Helper [29] та Beck Depression Inventory (BDI-II) [30].

EMDR Helper – це програмне забезпечення, яке допомагає терапевтам використовувати метод десенсибілізації та переробки очних рухів (EMDR) для лікування травм пов'язаними із травматичними спогадами та емоціями. Програма пропонує різні візуальні та звукові стимули, які терапевти можуть використовувати під час сеансів EMDR, а також дозволяє відстежувати прогрес своїх пацієнтів, що допомагає у плануванні терапії. Крім того, EMDR Helper може генерувати звіти про сеанси, що корисно як для терапевтів, так і для пацієнтів.

Beck Depression Inventory (BDI-II) – це самозвітний опитувальник, який використовується для оцінки тяжкості симптомів депресії, моніторингу прогресу лікування та допомоги у дослідженнях депресії. Він складається з 21 питання, які оцінюють різні симптоми депресії, такі як сум, втрата інтересу, втома, порушення сну та апетиту, труднощі з концентрацією уваги та суїцидальні думки. Кожному питанню присвоюється бал від 0 до 3, а сумарний бал використовується для оцінки тяжкості депресії. Важливо зазначити, що BDI-II не є діагностичним інструментом, і діагноз депресії повинен бути поставлений кваліфікованим психіатром або психологом.

В той же час застосування більш складних інструментів обробки природних мов, таких як великі мовні моделі, раніше не давало суттєвих переваг через помилки та неточності у відповідях. Проте, з появою мовних моделей сімейства BERT та сімейства GPT, результати генерації відповідей на запитання та аналіз текстових даних значно покращилися, відкриваючи нові можливості для застосування у всіх галузях, включаючи психологію.

Питання можливості застосування цих інструментів для проведення повноцінного психологічного тестування майже не розглядалося через складність інтерпретації відповідей респондентів, яка не може бути описана набором чітких правил і інструкцій через загальну нечіткість психологічних наук.

Незважаючи на значні успіхи, психологія особистості все ще стикається з певними викликами. Одним із ключових є етичний аспект використання комп'ютерних технологій. Важливо, щоб дані про особистість були захищені, а комп'ютеризовані системи використовувалися відповідально та не завдавали шкоди людям.

Ще одним викликом є необхідність інтеграції комп'ютерних методів з традиційними методами психологічної роботи. Важливо, щоб психологи не покладались лише на комп'ютери, а використовували їх як доповнення до своїх навичок та досвіду.

2 НАУКОВО-ТЕХНІЧНА ЗАДАЧА

В рамках кваліфікаційної роботи магістра необхідно вирішити ряд задач. Першим етапом дослідження стане проведення аналізу теоретичних відомостей та огляд тематичних публікацій, що характеризують розвиток та поточний стан проєктивних методик тестування, а також застосування сучасних комп'ютерних технологій в галузі психології та психотерапії. Наступною стадією дослідження є пошук та порівняльний аналіз сучасних підходів NLP для вирішення задач аналізу довільних текстів з метою відповідей на питання по тексту. Подальшими етапами роботи стане підготовка та планування експериментальної частини, збір даних, проведення експериментів, валідація та оцінка отриманих результатів.

Метою даного дослідження є синтез методу автоматичної інтерпретації та оцінки результатів опитування за проєктивною методикою тестування з використанням засобів та підходів обробки природніх мов.

Задачами дослідження є:

- вивчення існуючих наукових робіт на тему використання проєктивних методик тестування та підходів на їх основі у психології особистості;
- ознайомлення з дослідженнями та методами NLP для вирішення проблем question answering (відповіді на питання) та summarization (підбиття підсумків) для довільних текстів;
- порівняльний аналіз наявних технологій, проєктування архітектури та імплементація програмного забезпечення для проведення експериментальної частини дослідження;
- визначення критеріїв та метрик оцінювання ефективності запропонованого методу інтерпретації та оцінки результатів опитування за проєктивною методикою тестування;
- збір експериментальних даних, планування та опис схеми проведення експерименту, аналіз отриманих результатів.

Висновки за роботою повинні передбачати рекомендації до впровадження запропонованих підходів на підґрунті проведеного аналізу.

3 ОПИС ТЕОРЕТИЧНИХ ТА ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ

3.1 Опис теоретичних досліджень

Аналіз предметної галузі та огляд наукової літератури виявили низку проблем, з якими стикається сучасна психологія особистості. Однією з ключових проблем є складність та трудомісткість проєктивних методик тестування, які є фундаментом психологічних досліджень. Їхнє використання потребує кваліфікованих спеціалістів, що значно обмежує сферу застосування цих методів та робить їх недоступними для широкого загалу. Дослідники вбачають перспективне рішення цієї проблеми в автоматизації тестування за проєктивними методиками за допомогою комп'ютерних технологій. Це дозволить не лише спростити та прискорити процес дослідження, але й зробити його доступнішим для більш широкого кола людей, відкриваючи нові можливості для розуміння людської особистості та надання ефективної психологічної допомоги.

Як було зазначено раніше, проєктивні методики тестування побудовані на принципі пропозиції респондентам інтерпретувати неоднозначні стимули, що дозволяє отримати інформацію про їхні несвідомі аспекти особистості. Через цю особливість відповіді на запитання проєктивних тестів зазвичай є неструктурованими. Вивчаючи їх, дослідник виокремлює ключові слова та речення, аналізує відповіді на наявність прихованих підтекстів та патернів, порівнює відповіді з певними шаблонами поведінки.

Для автоматизації подібних задач пропонується використовувати підходи та методи природної обробки мов. Окреслені задачі можуть бути зведені до двох загальних задач NLP – question answering (відповіді на питання) та summarization (підбиття підсумків) для довільних текстів [31]. Одним з інструментів, який покликаний вирішувати подібні задачі та який за останній час отримав значне поширення є великі мовні моделі.

В рамках дослідження пропонується розробити систему для проведення тестування за проєктивними методиками із подальшою інтерпретацією та аналізом відповідей респондента тестування з використанням великих мовних моделей.

3.1.1 Вибір проєктивної методики тестування

Проєктивні методики тестування включають багато підходів, таких як тест Роршаха, тематичний апперцептивний тест, проєктивний малюнок людини, метод асоціацій тощо. Кожен з цих тестів має свої переваги та недоліки, що дозволяють їм не втрачати актуальність та мати свою нішу в проєктивній психодіагностиці.

Однак, при проведенні дослідження в якості основної методики проєктивного тестування пропонується використати тематичний апперцептивний тест [32].

ТАТ вирізняється своєю універсальністю та високою застосовністю в різних областях психологічного дослідження. Цей метод вивчає особистісні мотивації через інтерпретацію зображень персонажів, викликаючи спонтанні реакції та дозволяючи глибше зануритися в психічний світ респондента.

ТАТ являє набір з декількох (від 6 до 30) чорно-білих картинок ситуаційного типу, які по черзі пред'являються респонденту. Для тесту підібрані спеціальні картинки, де зображені ситуації відповідають двом основним вимогам. При достатньо чітких і яскравих зображеннях предметів і дійових осіб самі по собі композиції цих ситуацій є в той же час об'єктивно невизначеними, що не допускають однозначного тлумачення.

При сприйнятті таких картин у випробуваного створюється своєрідна ілюзія досконалої ясності картини: чіткі зображення одразу визначають сферу дій респондента, його середовище і напрямки реакції, але в той же час невизначеність композиції (що виключає однозначне і однакове рішення) змушує його тлумачити картини по-своєму. Випробуваний в кожне тлумачення вкладає свої знання, свій індивідуальний досвід, свій спосіб дії і свої оцінки.

При проходженні тесту, респондент має за відведений час у довільній формі розкрити певні аспекти кожного зображення в наборі та виконати наступні завдання:

- описати те, що, на його думку, відбувається на картині (тобто в даний момент),
- розповісти, що, на його думку, сприяло виникненню даній ситуації (що її породило);
- передбачити, що станеться надалі;
- розповісти про почуття та переживання зображених осіб;
- описати думки людей.

До переваг ТАТ над іншими проєктивними методиками тестування можна віднести те, що ТАТ забезпечує контекст для відповідей респондента, що допомагає зрозуміти, як людина реагує на соціальні та міжособистісні ситуації. Це робить результати більш релевантними для аналізу поведінки в реальному житті порівняно наприклад із тестом Роршаха. ТАТ дає можливість варіативної інтерпретації історій, дозволяючи психологу застосовувати різні теоретичні підходи залежно від контексту дослідження та специфіки респондента.

Окрім цього, ТАТ дозволяє виявляти не лише статичні особливості особистості, але й динамічні аспекти, такі як зміни у часі, розвиток внутрішніх конфліктів та способи їх розв'язання [33]. Цей тест може бути адаптований для використання з різними віковими групами, включаючи дітей, підлітків та дорослих, що робить його універсальним інструментом для психологічної діагностики у широкому віковому спектрі.

Таким чином, вибір ТАТ в якості основного підходу проєктивного тестування в даному дослідженні обґрунтований його універсальністю, широким застосуванням та можливістю глибокого вивчення особистісних мотивацій. Цей метод володіє потужним потенціалом для отримання інсайтів у внутрішній світ особистості та її ставлення до різних сценаріїв життя.

3.1.2 Обґрунтування схеми оцінювання результатів проєктивного тестування

У попередньому розділі було відзначено, що серед основних недоліків проєктивних методик тестування спостерігається обмежена стандартизація та суб'єктивність інтерпретацій результатів. Проте варто зазначити, що для деяких з цих методик існують і застосовуються схеми оцінювання, які надають можливість аналізу відповідей респондентів за конкретним шаблоном. Такий підхід дозволяє класифікувати респондентів за їхніми особливостями та рисами поведінки використовуючи кількісні оцінки, визначені на певному наборі критеріїв.

Прикладом такої схеми оцінювання є система, розроблена Річардом Даном для тематичного апперцептивного тесту [34]. Його система була створена з метою надання простого та економічного підходу до проєктивного тестування, особливо підходячи для клінічних умов. Використовуючи три категорії оцінок – організація сприйняття, діапазон сприйняття і персоналізація сприйняття – система Дана дозволяє визначити різні ступені психічних захворювань у респондентів.

До основних аспектів системи оцінювання Дана можна віднести:

- структурованість оцінки – система Дана надає чіткі критерії та категорії для аналізу розповідей респондентів. Це дозволяє зменшити суб'єктивність інтерпретацій і забезпечити послідовність у оцінюванні;

- фокус на тематичному змісті – відповіді респондентів оцінюються на основі тематичного змісту історій, включаючи головні персонажі, їхні мотивації, емоційні стани, конфлікти та розв'язання. Це дозволяє глибше зрозуміти внутрішній світ респондента;

- категоризація відповідей – система Дана передбачає категоризацію відповідей за різними параметрами, такими як головні теми, типи конфліктів, міжособистісні стосунки та емоційні реакції. Це допомагає створити структурований профіль особистості респондента;

- кількісна оцінка – на відміну від класичного підходу до оцінювання проєктивних методик з використанням якісних оцінок, система Дана використовує

кількісні оцінки, що дозволяє провести статистичний аналіз даних і порівняти результати між різними респондентами або групами.

Недоліком системи оцінювання Дана можна вважати те, що результат оцінювання не розкриває деталей про стан психічного здоров'я респондента, лише вказуючи на загальну приналежність опитуваної особи до певної клінічної групи. Але зважаючи на задачі даного дослідження, такий формат оцінки стану респондентів є цілком прийнятним і дозволяє суттєво спростити роботу спеціалістів-психологів шляхом попередньої класифікації відповідей учасників ТАТ тестування.

Система Дана відрізняється високою надійністю, навіть при використанні непідготовленими користувачами. За допомогою п'яти різних карток ТАТ для чоловіків і жінок, система Дана показала не лише високу критеріальну валідність, але й правильно розподіляла людей по клінічних групах (нормальні, невротичні, психотичні) у значній кількості випадків. Її актуальність і в сучасності підтверджуються як для клінічних, так і для дослідницьких цілей. Результатом оригінального дослідження став показник точності визначення приналежності респондента до певної клінічної групи на рівні 83 %.

В оригінальному дослідженні було проведено тестування на валідаційних та крос-валідаційних вибірках. Валідаційна вибірка складалася з 300 осіб - 150 чоловіків і 150 жінок. З них 100 були представниками «нормальної» клінічної групи, 100 – «невротичної» і 100 – «психотичної». Крос-валідаційна вибірка включала 90 чоловіків, серед яких було по 30 представників кожної групи. Учасників «нормальної» клінічної групи було відібрано серед студентів коледжів, без психіатричних захворювань, з показниками тесту ММРІ нижче 70. Респондентів з «невротичної» та «психотичної» груп були відібрані за попередньо встановленими критеріями, такими як наявність діагнозу неврозу або шизофренії, або були відібрані з амбулаторних пацієнтів. Усі учасники дослідження були у віці від 20 до 40 років з показниками IQ не нижче середнього рівня.

Враховуючи викладені відомості, в рамках даного дослідження пропонується використати систему оцінювання Ричарда Дана для проведення тестування за методикою тематичного апперцептивного тесту та інтерпретації його результатів.

3.1.3 Аналіз схеми оцінювання Ричарда Дана

Як було зазначено в попередньому підрозділі, система оцінювання запропонована Даном, використовує три категорії оцінок – організація сприйняття, діапазон сприйняття і персоналізація сприйняття.

Категорія «організація сприйняття» відображає здатність суб'єкта дотримуватися стандартних інструкцій для розповіді історії. До неї входять сім можливих компонентів:

- опис картки – хто та які предмети зображені на картці, опис деталей оточення;
- поточні події – опис того, що відбувається на картці зараз;
- минулі події – опис подій та передумов ситуації, що зображена на картці;
- майбутні події – опис подій, які можливо стануться надалі, яким буде результат поточної ситуації в майбутньому;
- почуття – які почуття викликає у респондента ситуація та дійові особи зображені на картці;
- думка – яке ставлення в респондента до ситуації що зображена на картці;
- результат – висновок або завершення історії, яку розповів респондент по запропонованій картці.

Оцінка цієї категорії відбувається за допомогою спеціального бланку (див. рис. 3.1). Бланк заповнюється дослідником шляхом оцінки кожного окремого компоненту історії – якщо компонент присутній у розповіді респондента, то у відповідній клітинці бланку проставляється відмітка «+», якщо компонент відсутній у розповіді, то ставиться позначка «-».

Загальна оцінка респондента за цією категорією складається з суми балів (позначок «+») за усі картки використанні в тестуванні.

Категорія «діапазон сприйняття» була розроблена спираючись на статистичні данні отриманні при проведенні проєктивного тестування за допомогою ТАТ. Аналізуючи їх Р. Дана виявив закономірність, що респонденти з «нормальної» клінічної групи у 90 % випадків звертають увагу та описують одні й ті самі деталі зображені на картці. І навпаки, представники з «невротичної» та «психотичної» груп схильні пропускати об'єкти зазначені представниками «нормальної» клінічної групи, акцентуючи свою увагу на інших аспектах зображеної ситуації. В рамках цієї категорії, дослідник оцінює наскільки історія респондента збігається у деталях розповіді зі середньостатистичною історією.

Процес оцінювання та розрахунок загальної оцінки виконується у той же спосіб, що і у попередній категорії.

TAT Score Sheet: Male
Perceptual Organization (PO), Perceptual Range (PR), Perceptual Personalization (PP)

Name _____ Age _____

Component	FO							Card	PR Criteria	PP (List by card #)
	2	3	4	6	7	Tot. F				
CD								2	<input type="checkbox"/> a. Family: young girl, woman, activity specified; adult male.	_____
PB									<input type="checkbox"/> b. Fields or farm.	_____
PE								3	<input type="checkbox"/> d. Figure, sex and age specified.	_____
FE									<input type="checkbox"/> e. Emotions noted.	_____
f									<input type="checkbox"/> f. Activity specified.	_____
T								4	<input type="checkbox"/> g. Male; emotions noted, activity specified.	_____
O									<input type="checkbox"/> h. Woman; activity specified.	_____
Total N								6	<input type="checkbox"/> i. Conflict or cooperation.	_____
									<input type="checkbox"/> j. Man; emotions noted, activity specified.	_____
Total Add.									<input type="checkbox"/> k. Woman; emotions noted.	_____
									<input type="checkbox"/> l. Personality referrent.	_____
								7	<input type="checkbox"/> m. Older male, activity specified, relationship specified.	_____
									<input type="checkbox"/> n. Male, emotions noted.	_____
									<input type="checkbox"/> o. Personality referrent.	_____

Sum T T T Scorer _____

Рисунок 3.1 – Приклад бланку оцінювання за системою Дана (за даними джерела [34])

Категорія «персоналізація сприйняття» розглядає та оцінює вирази, слова та фрази, використані в оповіданні, які є недоречними і не мають очевидного зв'язку з історією, яку респондент намагається розповісти. Ці включення мають бути ні

відтворенням стимулу, ні доповненням до стимулу, і є відхиленнями від відносно послідовного, організованого, зв'язної розповіді за процедурою ТАТ.

Умовою оцінки таких відхилень є їх повторюваність та «екстремальність». Вони можуть стосуватися таких речей, як адекватність виконання процедури ТАТ, коментарі, що не стосуються зображення, критика малюнків, додаткові описи, розпливчастість, ухиляння або прямі особисті посилання.

Дослідник оцінює кожне повторюване слово, фразу або коментар як один бал. Загальна оцінка за цією категорією виставляється як сума оцінок цієї категорії за кожну історію.

Для обчислення загальної оцінки відповіді респондента, оцінка за кожен критерій переводиться за допомогою таблиці наведеної на рисунку 3.2, після чого обчислюється сума отриманих значень.

Table 16
T Scores for Categories PO, PR, PP

T Score	Category		
	PO	PR	PP
78	35-37		
72	32-34		
67		15	
65	29-31		
62		14	
59	26-28		
58		13	0-2
56		12	
55	23-25		
53		11	
52	20-22		
51		10	
50			3-5
49	17-19	9	
47		8	
46	14-16		
45		7	6-8
43	11-13	6	
42			9-11
40		5	
39			12-14
38	8-10		
37		4	15-17
35			18-20
33		3	
32	5-7		21-23
30		2	24-26
28			27-29
26		1	30-35
25			36-38
24			39-41
19		0	42-44

Рисунок 3.2 – Таблиця переводів оцінок за кожну категорію до загальної шкали
(за даними джерела [34])

Порогові значення фінальної оцінки за тестування для «нормальної» клінічної групи становлять від 166 балів до 203 балів. Оцінки в межах 134 балів та 165 балів визначають приналежність респондента до «невротичної» клінічної групи. Значення оцінки нижче 133 балів вказують на те, що респондент є представником «психотичної» клінічної групи.

3.1.4 Вибір інструментів NLP

Для проведення дослідження та побудови програмної системи здатної до інтерпретації відповідей респондентів ТАТ та їх подальшої оцінки пропонується використати комплекс інструментів та методів NLP.

Головним компонентом системи стане сучасна LLM, як найбільш комплексний та досконалий інструмент NLP на поточний час. В задачі LLM входить рішення завдань з пошуку та виділення іменованих сутностей, аналіз змістовної частини текстового представлення відповіді респондента, відповіді на запитання по тексту відповіді.

Іншими засобами NLP, які пропонується використати в дослідженні є розпізнавання іменованих сутностей (Named Entity Recognition, NER) [35], що буде застосовано для оцінки відповідей респондентів за критерієм «діапазон сприйняття», та алгоритми виявлення текстових аномалій (Text Anomaly Detection, TAD) [36], що буде застосовано для знаходження стилістичних та лексичних аномалій у відповідях.

3.1.5 Огляд LLM

Для реалізації програмної системи експериментальної частини дослідження необхідно обрати одну або декілька LLM, які в подальшому будуть використанні для перевірки можливості інтерпретації та аналізу результатів тестування за проєктивними методиками. Для вибору моделей необхідно провести їх порівняння зважаючи на особливості дослідження, що проводиться. Адже узагальнена інформація про ту, чи іншу модель може не розкривати важливі аспекти про архітектуру і роботу моделі.

На момент проведення магістерського дослідження найбільш просунутими великими мовними моделями є GPT [37] від компанії OpenAI, LLaMA [38] та RoBERTa [39] від компанії Meta AI, PaLM 2 [40] та T5 [41] від компанії Google, які були взяті за досліджуванні альтернативи.

Моделі сімейства GPT — це глибинні мовні моделі від OpenAI, які здатні генерувати зв'язні тексти на задану тему. Остання версія, GPT-4, має 1.76 трильйони параметрів та навчена на великому обсязі текстів з мережі Інтернет. Моделі цього сімейства використовують трансформерну нейронну мережу з авторегресивною архітектурою та самоувагою. Переваги моделей сімейства GPT включають генерацію зв'язного тексту на різні теми та можливість швидко адаптуватися до нових завдань. Недоліками є можливі помилки у розумінні логіки, відображення упереджень та високі обчислювальні витрати.

LLaMA 2 – велика мовна модель від компанії Meta AI. Модель має 65 мільярдів параметрів та навчена на 3 трільйонах токенів текстових даних з ресурсу Вікіпедія, книг та веб-сторінок. Вона використовує трансформерну нейронну мережу з двосторонньою архітектурою та самоувагою. Перевагами моделі є розуміння природної мови та доступ до докладної інформації з багатьох джерел. Недоліки – можливість обмеженої генерації тексту та відображення упереджень.

RoBERTa – це модель побудована на архітектурі BERT від компанії Meta AI для обробки природної мови [42]. Модель навчена на 160 ГБ тексту, що у 10 разів перевищує датасет використаний для навчання оригінальної моделі (BERT). Суттєвими змінами від оригінальної архітектури є вилучене в окреме завдання передбачення наступного речення, використання динамічного маскування, більші розміри партій та оптимізовані гіперпараметри. Переваги RoBERTa включають високу ефективність в завданнях обробки природної мови, здатність до навчання з невеликою кількістю прикладів, і відкритий код. Недоліками моделі є можливість невірної розуміння логіки та відображення упереджень.

PaLM 2 – це велика мовна модель від компаній Google AI для обробки природної мови, навчена на 540 мільярдах параметрів та 780 мільярдах токенів з різних джерел. Вона використовує архітектуру декодера та трансформерну

нейронну мережу з механізмом самоуваги. Її особливості включають систему Pathways для ефективного масштабування та навчання з меншою кількістю параметрів. Плюси PaLM 2 – це висока ефективність у різних завданнях обробки природної мови, здатність до роботи з текстом та кодом на різних мовах та в галузях, високорівневі розуміння завдань коду та математичних задач. Серед мінусів моделі можна назвати можливість відсутності логічного розуміння та відображення упереджень, обчислювальна складність та важкість налаштування, що суттєво обмежує її доступність та масштабованість для менших компаній та окремих користувачів.

T5 – це інша велика мовна модель від Google AI для обробки природної мови, яка використовує трансформерну нейронну мережу. Вона використовує формат text-to-text для виконання різноманітних завдань, таких як переклад, резюмування та відповіді на запитання. T5 може працювати з великою кількістю інформації з різних джерел і є відкритою моделлю. Модель може виявляти відсутність логічного розуміння та відображати упередження.

Незважаючи на те, що моделі T5 та PaLM 2 побудовані з використанням різних підходів та архітектур, їх спільними рисами є сфера застосування та основні недоліки, такі як вплив упереджень та висока обчислювальна складність.

3.1.6 Формування критеріїв та шкал оцінювання

З метою об'єктивно порівняти обрані альтернативи було відібрано п'ять критеріїв, що дозволять проаналізувати кожен модель на відповідність вимогам дослідження. Так критеріями порівняння є швидкодія, максимальний розмір вхідних даних, підтримка семантичного аналізу, придатність до додаткового тренування (fine tuning) та простота використання та розробки.

Швидкодія – це показник того, наскільки модель здатна швидко аналізувати вхідний текст та генерувати текст-відповідь. Вимірюється в токенах на секунду. Порівняння моделей на основі часу, який вони зазвичай витрачають на генерацію тексту, не є простим завданням, оскільки час генерації може залежати від різних факторів, таких як довжина і складність вхідних і вихідних послідовностей,

наявність і конфігурація обчислювальних ресурсів, а також оптимізація і реалізація моделей. Однак, спираючись на відкриті джерела було отримано інформацію про приблизний час генерації відповідей моделями, які були запущені на однакових обчислювальних потужностях (один пристрій TPU v4) та мали одне і те саме завдання.

Максимальний розмір вхідних даних – це показник того, текст якого максимального розміру (у токенах) модель може отримати на вхід.

Підтримка семантичного аналізу текстів. Семантичний аналіз тексту - це процес вилучення та аналізу значення і структури тексту, таких як теми, сутності, зв'язки, настрої та наміри. Деякі з моделей мають більшу підтримку семантичного аналізу тексту, ніж інші, залежно від їхніх архітектури, цілей і можливостей. Через неможливість виміряти значення цієї характеристики в абсолютних величинах, було введено рядкову шкалу:

Повна підтримка (10 балів) – модель може вирішувати завдання розуміння природної мови, що включає передбачення відсутніх лексем у частково замаскованій вхідній послідовності, а також виконання розпізнавання спільних посилань та розпізнавання іменованих сутностей;

Часткова підтримка (5 балів) – модель може вирішувати частину завдання розуміння природної мови, але не повністю чи/або потребує додаткового налаштування для виконання цього завдання;

Не підтримується (1 бал) – модель не може вирішувати завдання розуміння природної мови.

Придатність до тренування (fine tuning) – показник того наскільки модель піддається процесу коригування параметрів попередньо навченої моделі для покращення її роботи на конкретному завданні або предметній області. Деякі з моделей краще підходять для налаштування, ніж інші, залежно від їхньої архітектури, навчальних наборів даних і цілей.

Повна придатність (10 балів) – модель призначена для подальшого доопрацювання та покращення продуктивності, вона має модульну та гнучку архітектуру, яка може бути легко адаптована до різних завдань та застосувань;

Часткова придатність (5 балів) – модель менш піддається додатковому налаштуванню і оптимізована для виконання завдань за своїм основним призначенням;

Не підтримується (1 бал) – модель не підтримує додаткове налаштування.

Простота використання та розробки. Простота використання та розробки стосується того, наскільки легко отримати доступ, використовувати та модифікувати ці моделі для різних цілей та потреб. Цей показник залежить наявності моделей, їх доступності та масштабованості.

Відмінно (10 балів) – мовна модель поставляється за open-source моделлю, має добру документацію та зручний програмний інтерфейс. Модель легко масштабується.

Добре (8 балів) – мовна модель поставляється за open-source моделлю, має документацію та зрозумілий програмний інтерфейс.

Задовільно (5 балів) – мовна модель закрита для вільної зміни та налаштування. Модель має детальну документацію розробника та зовнішній програмний інтерфейс;

Погано (3 бали) – мовна модель закрита для вільної зміни та налаштування. Модель має погано описану документацію розробника та зовнішній програмний інтерфейс.

Використання наведених вище шкал оцінювання дозволить порівняти відібрані альтернативи формальними методами такими як принцип Парето.

3.1.7 Порівняльний аналіз LLM

Після визначення альтернатив, критеріїв вибору та виміру відповідних даних була побудована таблиця порівняння великих мовних моделей (див. табл. 3.1).

Таблиця 3.2 демонструє порівняння з показниками критеріїв переведеними до числових значень.

Через те, що всі критерії відповідають принципу – «більше значення - краще», додаткові перетворення значень критеріїв непотрібні.

Таблиця 3.1 – Порівняння великих мовних моделей за визначеними критеріями (таблиця виконана самостійно)

Мовна модель	Швидкодія (токен/сек)	Розмір вхідних даних (токен)	Підтримка семантичного аналізу	Придатність до fine tuning	Простота використання та розробки
GPT-4	4.5	32768	Повна підтримка	Часткова придатність	Задовільно
LLaMA 2	190	1024	Повна підтримка	Повна придатність	Відмінно
RoBERTa	100	512	Повна підтримка	Часткова придатність	Задовільно
PaLM 2	10	4096	Часткова підтримка	Часткова придатність	Погано
T5	40	1024	Часткова підтримка	Повна придатність	Добре

Використовуючи принцип Парето, з таблиці альтернатив (див. табл. 3.2) було прибрано модель PaLM 2 через її неконкурентноспроможність у порівнянні з іншими розглянутими моделями.

Таблиця 3.2 – Порівняння великих мовних моделей за визначеними критеріями (таблиця виконана самостійно)

Мовна модель	Швидкодія (токен/сек)	Розмір вхідних даних (токен)	Підтримка семантичного аналізу	Придатність до fine tuning	Простота використання та розробки
GPT-4	4.5	32768	10	5	5
LLaMA 2	190	1024	10	10	8
RoBERTa	100	512	10	5	5
PaLM 2	10	4096	5	5	3
T5	40	1024	5	10	8

Модель GPT-4, не зважаючи на загально слабкі показники критеріїв порівняння, виявилась кращою за показником «Максимальний розмір вхідних даних», і тому була залишена для подальшого порівняння (відповідно до принципу «альтернатива має бути не гірше за інші хоча б за одним критерієм»).

Таблиця 3.3 демонструє список альтернатив після застосування принципу Парето.

Таблиця 3.3 – Список альтернатив після застосування принципу Парето (таблиця виконана самостійно)

Мовна модель	Швидкодія (токен/сек)	Розмір вхідних даних (токен)	Підтримка семантичного аналізу	Придатність до fine tuning	Простота використання та розробки
GPT-4	4.5	32768	10	5	5
LlaMA 2	190	1024	10	10	8
RoBERTa	100	512	10	5	5
T5	40	1024	5	10	8

Наступним кроком аналізу стало нормування показників критеріїв за принципом «min-max» (3.1).

$$f = \frac{f_{\text{вимір}} - f_{\text{min}}}{f_{\text{max}} - f_{\text{min}}} \quad (3.1)$$

Результатом (див. табл. 3.4) цього стала нормалізація показників критеріїв в рамках від 0 до 1 (включно з кінцями).

Через те, що при проведенні магістерського дослідження деякі з характеристик є важливішими за інші, для вирішення багатокритеріальної задачі вибору доцільно використовувати адитивну згортку із ваговими коефіцієнтами.

Так, ваговими коефіцієнтами в порядку зменшення будуть:

- підтримка семантичного аналізу – 0.4, цей показник є на порядок важливішим за інші, адже визначає придатність моделі виконувати завдання дослідження без додаткових налаштувань;
- максимальний розмір вхідних даних – 0.25, визначає наскільки деталізовану текстову відповідь респондента, модель може опрацювати;
- підтримка додаткових налаштувань – 0.15, адже визначає наскільки модель гнучка і може бути змінена для виконання задач дослідження;
- швидкодія та простота використання по 0.1, адже ці показники є найменш критичними при проведенні дослідження.

Таблиця 3.4 – Нормовані значення критеріїв (таблиця виконана самостійно)

Мовна модель	Швидкодія (токен/сек)	Розмір вхідних даних (токен)	Підтримка семантичного аналізу	Придатність до fine tuning	Простота використання та розробки
GPT-4	0	1	1	0	0
LlaMA 2	1	0.015	1	1	1
RoBERTa	0.51	0	1	0	0
T5	0.19	0.015	0	1	1

Розрахунок згортки відбувається за формулою 3.2:

$$Z^* = \max_{i=1,m} \sum_{j=1}^n \alpha_j \beta_j a_{ij} \quad (3.2)$$

де α_j – нормуючі множники;

β_j – вагові коефіцієнти;

a_{ij} – значення критерію із таблиці.

Таблиця 3.5 демонструє результат після проведення розрахунків згортки для альтернатив відбору.

Таблиця 3.5 – Результат вирішення багатокритеріального задачі вибору (таблиця виконана самостійно)

Мовна модель	Швидкодія (токен/сек)	Розмір вхідних даних (токен)	Підтримка семантичного аналізу	Придатність до fine tuning	Простота використання та розробки	Z^*
GPT-4	0	1	1	0	0	0.65
LlaMA 2	1	0.015	1	1	1	0.75375
RoBERTa	0.51	0	1	0	0	0.451
T5	0.19	0.015	0	1	1	0.44375
Вагові коефіцієнти	0.1	0.25	0.4	0.15	0.1	

Таким чином, кращою мовною моделлю для проведення дослідження є LlaMa 2 з результатом 0.75375. На другому місці модель GPT-4 (0.65). На третьому місці RoBERTa (0.451) з відривом від четвертого місця у 0.00725. На останньому

місці, через посередні загальні характеристики та відсутність підтримки семантичного аналізу, модель T5 з результатом 0.44375.

Результатом проведеного аналізу є те, що при проведенні даного дослідження основною моделлю буде LLaMa 2, яка має найкращі показники по всіх обраних критеріях, окрім максимального розміру вхідних даних.

В якості альтернативної моделі може бути використана мовна модель GPT-4, яка є другою моделлю, що має найкращий показник розміру вхідних даних серед усіх розглянутих альтернатив та гарний показник підтримки семантичного аналізу.

3.2 План проведення експериментів

В рамках експериментальної частини дослідження має бути перевірена ефективність роботи системи в різних комбінаціях використаних компонентів:

- інтерпретація та оцінка відповідей респондентів з використанням моделі Llama 2;
- інтерпретація та оцінка відповідей респондентів з використанням моделі Llama 2 та алгоритмів NER і TAD для попередньої обробки тексту відповіді учасника опитування;
- інтерпретація та оцінка відповідей респондентів з використанням моделі GPT-4;
- інтерпретація та оцінка відповідей респондентів з використанням моделі GPT-4 та алгоритмів NER і TAD для попередньої обробки тексту відповіді учасника опитування;
- інтерпретація та оцінка відповідей респондентів з використанням моделей Llama 2 та GPT-4;
- інтерпретація та оцінка відповідей респондентів з використанням моделей Llama 2 і GPT-4 та алгоритмів NER і TAD для попередньої обробки тексту відповіді учасника опитування.

В експериментах із застосуванням обох LLM агрегація результатів оцінювання передбачається за вагове середнє. Такий підхід дозволить має на меті компенсувати недоліки однієї моделі перевагами іншої. Значення вагових

коefficientів має бути встановлено після проведення експериментів, де великі мовні моделі застосовуються окремо.

Передумовами проведення експериментальної частини дослідження є:

- імплементація програмної системи, що реалізує окреслені сценарії проведення експериментів;
- збір та попередня обробка тестових даних;
- визначення метрик оцінювання результатів роботи програмної системи в рамках експериментів.

Виконання зазначених кроків є необхідною умовою для проведення експериментальної частини даного дослідження.

3.3 Метрики ефективності

В якості метрики визначення ефективності класифікації відповідей респондентів ТАТ тестування за однією з трьох раніше зазначених груп пропонується використати F-міру. Цей показник характеризує точність тесту і широко застосовується у галузях ML та NLP [43].

Значення F-міри (3.3) обчислюється як середнє гармонійне значень влучності (Precision) та повноти (Recall).

$$F = 2 \frac{Precision * Recall}{Precision + Recall} \quad (3.3)$$

Влучність (3.4) обчислюється як відношення кількості правильно визначених позитивних результатів до числа всіх позитивних результатів разом із тими, які визначено неправильно.

$$Precision = \frac{TP}{TP + FP} \quad (3.4)$$

де TP – кількість істинних позитивних рішень;

FP – кількість помилково позитивних рішень.

У формулі 3.5 наведено обчислення значення повноти, що є числом правильно визначених позитивних результатів, поділеним на число всіх зразків, які повинно було бути визначено як позитивні.

$$Recall = \frac{TP}{TP + FN} \quad (3.5)$$

де TP – кількість істинних позитивних рішень;

FN – кількість помилково негативних рішень.

Для визначення рівня подібності отриманих кількісних оцінок відповідей респондентів відносно еталонних значень буде використано коефіцієнт кореляції Пірсона. Цей показник дозволить виявити та оцінити лінійну залежність між наборами даних.

3.4 Збір даних та підготовка експерименту

В рамках підготовки дослідження з метою збору тестових даних було проведено тестування за методикою ТАТ для групи з 18 осіб.

Віковий склад групи розподілено наступним чином: 10 осіб віком від 18 до 25 років, 5 осіб віком від 26 до 40 років та 3 особи віком від 40 до 65 років. Група має рівний розподіл за статтю: однакова кількість чоловіків та жінок.

Усі респонденти не мають історії психологічних захворювань, причому лише троє з них мали родичів із психічними захворюваннями або розладами. Учасники дослідження самостійно оцінили свій психічний стан як нормальний. Однак, семеро з них згадали в своїх анкетах про підвищену тривожність та втомленість протягом останнього року.

Для отриманих відповідей було проведено процедуру ручного оцінювання за системою Р. Дана. З метою зменшення впливу суб'єктивних чинників на оцінки відповідей респондентів, у процедурі оцінювання взяли участь три особи, які не брали участь у ТАТ тестуванні та не мають психологічної освіти. Отримані оцінки для кожної відповіді були узгоджені між особами, що виконували перевірку.

Значення оцінок, отриманих під час ручної перевірки (див. табл. А.1), було прийнято за еталонні. Під час аналізу даних, отриманих в ході експерименту, слід виконувати порівняння із зазначеними еталонними даними.

4 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

4.1 Результати проведених експериментальних досліджень

Згідно із планом проведення експериментів розроблену систему було перевірено в зазначених конфігураціях на тестових даних. В таблиці А.2 наведено результати оцінок відповідей респондентів в кожному проведеному експерименті.

На результатах експериментів були розраховані показник коефіцієнту кореляції Пірсона, значення точності (precision), повноти (recall) та F-міри відносно еталонного набору даних. Отриманні результати наведено у таблиці 4.1.

Таблиця 4.1 – Значення оцінок для кожного експерименту (таблиця виконана самостійно)

Експеримент	Кореляція	Precision	Recall	F-міра
Експеримент 1	0.44	0.38	0.38	0.33
Експеримент 2	0.81	0.72	0.62	0.63
Експеримент 3	0.68	0.62	0.62	0.62
Експеримент 4	0.84	0.62	0.45	0.51
Експеримент 5	0.95	0.62	0.5	0.53
Експеримент 6	0.97	0.81	0.67	0.71

Розраховані показники є достатніми для аналізу ефективності застосування засобів NLP для інтерпретації та оцінювання результатів проєктивного тестування за ТАТ методикою.

4.2 Аналіз результатів проведених експериментальних досліджень

Результати першого експерименту, коли для оцінки тестування було використано велику мовну модель Llama 2, демонструють, що зазначена модель погано справляється із поставленим завданням без попередньої обробки тексту. Модель показала найнижчі результати кореляції та F-міри серед усіх проведених проб. Аналізуючи отриманні результати для кожної оцінюваної категорії за системою Дана, можна побачити закономірність, що модель показує кращий

результат при виконанні задач пов'язаних із роботою з контекстом тексту-відповіді (категорія «організація сприйняття») і в той же час робить багато помилок в завданнях на виокремлення сутностей та пошук аномалій у тексті.

Ця ситуація частково покращується у третьому експерименті, коли тексти відповідей респондентів додатково було опрацьовано спеціалізованими алгоритмами, які вирішують зазначені завдання. Незважаючи на це, результатом третього експерименту є посереднє значення кореляції (0.68) отриманих оцінок з еталонними оцінками. При цьому, значення F-міри при порівнянні отриманої класифікації відповідей є одним із найбільших серед всіх експериментів.

На противагу моделі Llama-2, модель GPT-4 показала гарний результат в експерименті без використання додаткових засобів NLP: показник кореляції – 0.81 та значення F-міри – 0.63. Хоча GPT-4 показала значно кращий результат за модель Llama-2, закономірність виявлена в першому експерименті прослідковується і в результатах цієї спроби – LLM краще порадиться із завданнями на розуміння контексту тексту, аніж із завданнями на виявлення сутностей та пошук аномалій.

Результати тесту, з використанням моделі GPT-4 в комбінації із іншими алгоритмами NLP продемонструвало погіршення показників precision та recall, що вплинуло на фінальну оцінку F-міри, яка склала 0.51.

Використання обох LLM у форматі «зміцнення результатів» оцінки в експериментах 5 та 6 показало суттєвий приріст в точності отриманих оцінок відповідей респондентів TAT тесту. Спираючись на результати отриманні в попередніх експериментах, в якості основної моделі було обрано GPT-4 і її оцінка отримала ваговий коефіцієнт 0.65. Відповідно ваговий коефіцієнт оцінки моделі Llama-2 склав 0.35.

Показник кореляції є високим в обох випадках – 0.95 та 0.97, в той час як значення F-міри для експерименту 5 є нижчим на 18 пунктів за показник для експерименту 6. Ця аномалія чітко прослідковується на відповідному графіку на рисунку 4.1, де значення перших чотирьох оцінок суттєво відрізняються від еталонних значень. Ця різниця зумовила отримання для перших чотирьох

респондентів класів відмінних від еталонних, що в результаті вплинуло на значення F-міри.

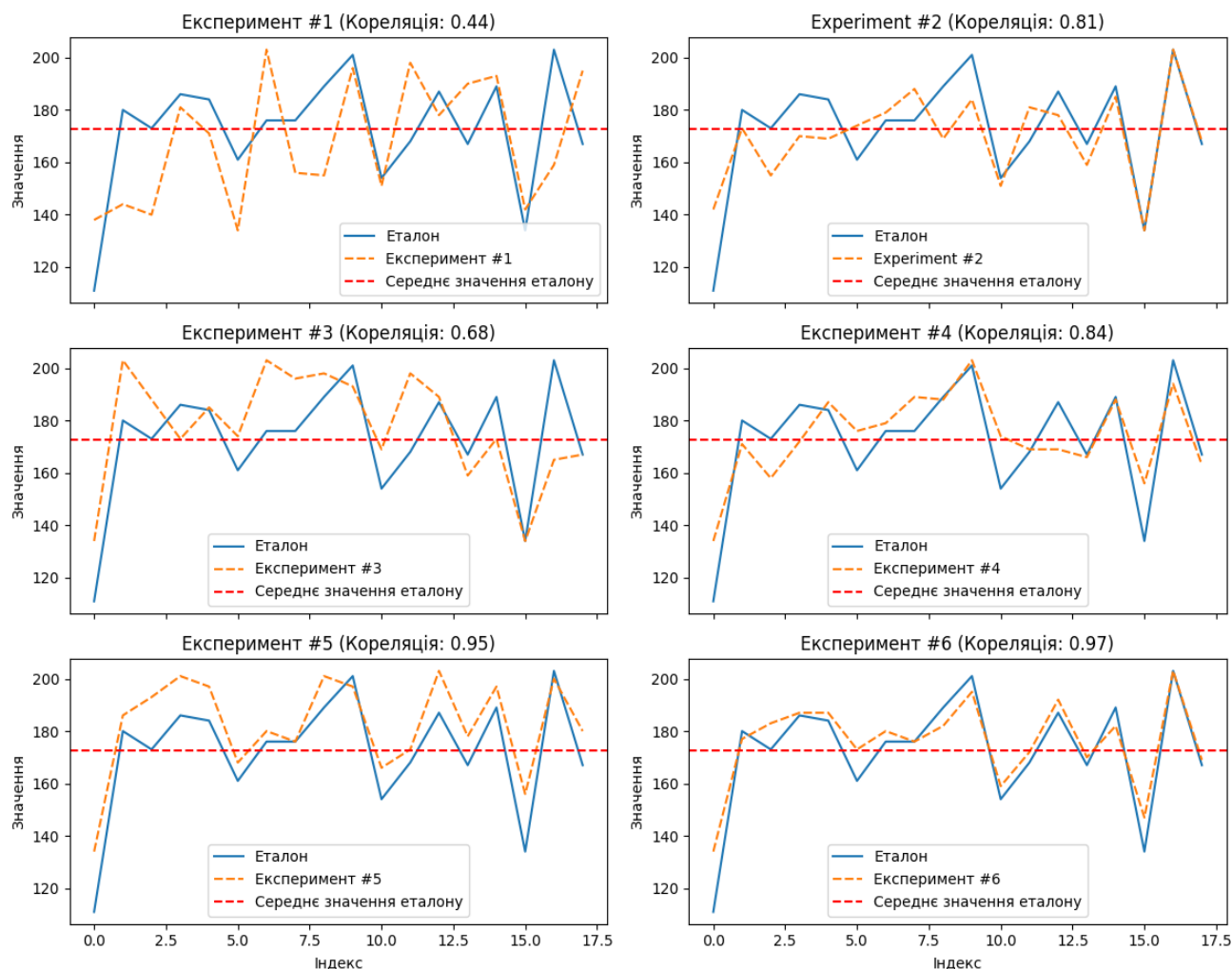


Рисунок 4.1 – Графіки порівняння експериментальних значень з еталонними значеннями

Підсумовуючи можна зробити висновок, що найкращий результат продемонстрував підхід із застосування комплексу різних засобів NLP. Такий метод дозволив компенсувати недоліки кожного окремого інструменту та алгоритму, підвищивши загальну точність системи. Недоліками такої конфігурації програмної системи можна вважати погіршену швидкодію, порівняно з використанням кожного засобу окремо, та загальна складність імплементації і використання подібних комплексних систем.

5 ОПИС ПРОГРАМНОЇ СИСТЕМИ

Розроблена програмна система складається з двох підсистем – підсистеми проведення проєктивного тестування за ТАТ методикою із записом відповідей респондента, та підсистеми інтерпретації та оцінювання відповідей ТАТ тестування.

5.1 Підсистема для проведення проєктивного тестування

Задачею даної підсистеми є анкетування респондентів, проведення ТАТ тесту та зберігання отриманих відповідей для їх подальшої обробки та інтерпретації іншою підсистемою програмної системи.

Підсистему було побудовано як веб-додаток з метою підвищення зручності її використання та полегшення доставки ПЗ до кінцевого користувача. Клієнтську частину додатку було побудовано з використанням мови програмування TypeScript та веб-фреймворку Vue 3 [44]. Серверна частина додатку реалізована з використанням мови програмування C# та фреймворку ASP.NET [45]. В якості бази даних було обрано NOSQL систему керування базами даних MongoDB [46].

Основний функціонал додатку включає в себе анонімне анкетування респондента під час якого система записує інформацію про вік, стать, наявність підтвердженої історії психологічних хвороб в респондента, або в його родичів, власну оцінку психологічного стану та додаткові коментарі. Після етапу збору анкетних даних, користувачу послідовно демонструється п'ять ТАТ карток, про кожну з яких користувач має дати розгорнуту відповідь. Користувачу додатку доступні два способи вводу інформації – ручне введення тексту, або введення голосом. Для обробки голосового вводу було використано Speech to text API від компанії Google.

Отриманні під час анкетування та тестування відповіді передаються на серверну частину додатку, де потім зберігаються у форматі JSON у базу даних системи.

5.2 Підсистема інтерпретації та оцінювання відповідей ТАТ

Задачею підсистеми є обробка зібраних тестових даних за визначеною планом експерименту конфігурацією.

Так основний функціонал ПЗ включає в себе оцінку відповіді респондента з використанням інструментів передбачених в експерименті, збереження отриманих результатів до бази даних та подальший розрахунок метрик ефективності роботи системи оцінювання таких як коефіцієнт кореляції Пірсона, значення точності (precision), повноти (recall) та F-міри відносно еталонного набору даних.

Додаток було реалізовано у вигляді набору самостійних скриптів, кожен з яких описує умови одного експерименту, без використання користувацького інтерфейсу.

При реалізації було використано API компанії OpenAI для доступу до великої мовної моделі GPT-4. Зазначений API дозволяє отримати доступ до різних варіантів LLM від компанії OpenAI з можливістю fine tuning налаштувань моделі під завдання дослідження. Модель Llama 2 було розгорнуто у хмарному сервісі Azure AI Studio [47], з доступом до моделі через відповідний API. Такий підхід дозволив підвищити точність відповідей моделі за рахунок використання більших версій Llama-2 та скоротити час генерування відповіді у порівнянні із локальним розгортанням зазначеної моделі.

Для реалізації було обрано мову програмування Python, що пов'язано з наявністю великої кількості реалізації NLP інструментів та засобів обробки даних у складі наукових пакетів таких як: pandas, numpy, Scikit-learn та spacy.

Для візуалізації отриманих результатів було використано пакети matplotlib та seaborn.

6 МОЖЛИВІСТЬ ВПРОВАДЖЕННЯ У НАУКОВІЙ ТА ПРАКТИЧНІЙ ДІЯЛЬНОСТІ

За тематикою даного дослідження було підготовлено та опубліковано тези доповіді, де було розглянуто актуальність питання впровадження комп'ютеризованих систем з використанням сучасних інструментів NLP у галузі психології особистості при проведенні тестування за проєктивними методиками [48].

Перспективним напрямком для практичного застосування розробленої програмної системи є психологічна діагностика та терапія. Впровадження програмної системи може значно покращити точність та об'єктивність оцінок, знижуючи ризик суб'єктивних помилок, що можуть виникати через людський фактор. Перевагою запропонованого методу є швидкість обробки великих обсягів даних, що сприяє прискоренню процесу діагностики. Крім того, автоматизована система може забезпечити доступність якісної психологічної оцінки в регіонах з обмеженим доступом до кваліфікованих спеціалістів.

Логічним продовженням даного дослідження може стати верифікація отриманих результатів на більшій вибірці даних. Це дозволить підтвердити надійність і репрезентативність попередніх результатів, а також виявити можливі закономірності та аномалії.

Окрім цього, одним із перспективних напрямків є використання LLM для аналізу проєктивних методик, відмінних від ТАТ. Здатність LLM розпізнавати складні мовні конструкції та контексти дозволяє глибше аналізувати відповіді респондентів і розширювати можливості автоматизованого оцінювання.

Також важливо дослідити можливості використання альтернативних систем оцінювання для ТАТ. Порівняння різних систем може допомогти визначити найефективніші методи, забезпечуючи більш комплексний підхід до аналізу результатів тестування.

ВИСНОВКИ

Темою кваліфікаційної роботи є дослідження моделей та методів проєктивних методик тестування на основі природної мови.

В ході виконання роботи було проведено аналіз предметної області в межах якого, було розглянуто наукову та патентну літературу, що присвячена проблематиці створення та впровадження комп'ютеризованих систем для проведення тестування за проєктивними методиками. Це дало підґрунтя для формулювання постановки науково-технічної задачі.

Так в рамках дослідження було перевірено гіпотезу можливості застосування інструментів обробки природної мови, таких як великі мовні моделі, алгоритми розпізнавання іменованих сутностей тощо, для проведення тестування за проєктивними методиками з подальшою інтерпретацією та оцінкою відповідей респондентів.

Під час проведення теоретичних досліджень в якості досліджуваної методики проєктивного тестування було обрано TAT тест через поширеність його застосування та переваги над іншими проєктивними тестами. Відповідно до обраної методики було відібрано схему її оцінювання, що дозволяє перейти до кількісних характеристик та визначило критерії відбору засобів NLP таких як швидкодія, розмір вхідних даних у токенах, підтримка семантичного аналізу, придатність до fine tuning та простота використання. Аналіз інструментів NLP за допомогою принципу Парето та принципу «min-max» за визначеними критеріями дозволив обрати LLM Llama-2 і GPT-4 та алгоритми пошуку сутностей і виявлення аномалій в тексті для подальшої програмної реалізації схеми оцінювання.

Розроблена програмна система складається з двох підсистем. Підсистему проведення проєктивного тестування за TAT методикою із записом відповідей респондента було реалізовано у вигляді веб-додатку з використанням фреймворків Vue 3 та ASP.NET. В свою чергу, підсистема інтерпретації та оцінювання відповідей TAT тестування складається з набору Python скриптів, кожен з яких реалізує умови проведення окремого експерименту дослідження.

Для підготовки експерименту було обрано тестові конфігурації програмної реалізації, підготовлено тестові данні шляхом проведення проєктивного тестування за ТАТ методикою групи з 18 осіб та розраховано еталонні оцінки відповідей респондентів.

Оцінка експериментальних даних відбувалась за показниками коефіцієнту кореляції Пірсона, значення точності (precision), повноти (recall) та F-міри відносно еталонного набору даних.

Аналіз отриманих даних свідчить, що на непідготовлених даних модель GPT-4 показує кращі результати ніж модель Llama-2 за показниками кореляції та F-міри. Використання попередньо підготовлених за допомогою алгоритмів NER та TAD текстових даних, дозволяє покращити результати великих мовних моделей. Застосування комплексу з двох LLM з агрегацією їх оцінок як вагове середнє, продемонструвало найкраще приближення до еталонних даних.

В ході проведення експерименту було виявлено аномальне зменшення показника F-міри для конфігурації з використанням GPT-4 на попередньо підготовлених даних та для конфігурації з використанням обох моделей на непідготовлених даних. Це може бути пов'язано з недостатнім об'ємом статистичної вибірки, через що невірне визначення навіть одного класу (клінічної групи) респондента призводить до значної зміни значення метрики.

Логічним продовженням даного дослідження може стати верифікація отриманих результатів на більшій вибірці даних. Це дозволить підтвердити надійність і репрезентативність попередніх результатів.

Подальшим напрямком розвитку досліджень також може стати удосконалення етапу підготовки даних та способу комплексування результатів різних великих мовних моделей.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Столяренко О. Б. Психологія особистості / О. Б. Столяренко. – Київ: Центр учбової літератури, 2012. – 280 с.
2. Галян І. М. Психодіагностика. Навчальний посібник / І. М. Галян . - К.: Академвидав. - 2009. - 464 с.
3. Поліщук С.А. Методичний довідник з психодіагностики: навч.-метод. Посіб / С.А. Поліщук. – Суми : Університетська книга - 2009. – 226 с.
4. Мак-Вільямс Н. Психоаналітична діагностика / Н. Мак-Вільямс. — М.: Клас, 1998. – 240 с
5. Фрейд З. Вступ до психоаналізу. Переклад з німецької: Петро Тарашук / З. Фрейд. – Київ: Основи, 1998. – 709 с.
6. Maslow A. H. A Theory of Human Motivation / Maslow. // Psychological Review. – 1943. – №50. – С. 370–396.
7. Allport G. W. Personality: a psychological interpretation. / Allport. – Нью-Йорк: Holt, Rinehart, & Winston, 1937.
8. Cattell R. B. The Description and Measurement of Personality / Raymond Cattell. – Нью-Йорк: Yonkers-on-Hudson, 1946.
9. Бандура А. Moral Disengagement: How People Do Harm and Live with Themselves / Альберт Бандура. – Нью-Йорк: NY: Worth, 2015.
10. Ellis A. Recent research with personality inventories / Albert Ellis. // Journal of Consulting Psychology. – 1953. – №17. – С. 45–49.
11. Schiele B. C. The Minnesota multiphasic personality inventory / B. C. Schiele, A. B. Baker, S. R. Hathaway. // Journal-Lancet. – 1943. – №63. – С. 292–297.
12. McCrae R. R. The NEO-PI-3: A More Readable Revised NEO Personality Inventory / R. R. McCrae, P. T. Costa, T. A. Martin. // Journal of Personality Assessment. – 2005. – №84. – С. 261–270.
13. Орбан-Лембрик Л. Е. Соціальна психологія: Посібник / Л. Е. Орбан-Лембрик. – Київ: Академвидав, 2003. – 448 с.

14. Горноста́й П. П. Психологія особистості: Словник-довідник / П. П. Горноста́й. – Київ: Рута, 2001. – 320 с.

15. Tuerlinckx F. Measuring needs with the Thematic Apperception Test: A psychometric study. / F. Tuerlinckx, P. De Boeck, W. Lens. // *Journal of Personality and Social Psychology*. – 2002. – №82. – С. 448–461.

16. Aronow E. A practical guide to the Thematic Apperception Test: The TAT in clinical practice / E. Aronow, K. Weiss, M. Reznikoff., 2001. – 121 с.

17. Rook D. W. Let's pretend: Projective methods reconsidered / Rook // *Handbook of qualitative research methods in marketing*. / Rook., 2006.

18. Fernández-Ballesteros R. Psychological assessment: Future challenges and progresses / Fernández-Ballesteros. // *European psychologist*. – 1999. – №4. – С. 248.

19. Archer R. P. Psychological test usage with adolescent clients: Survey update / R. P. Archer, C. R. Newsom. // *Assessment*. – 2000. – №3. – С. 227–235.

20. Bond R. The role of information and communication technologies in using projective techniques as survey tools to meet the challenges of bounded rationality / R. Bond, E. Ramsey. // *Qualitative Market Research: an international journal*. – 2010. – №4. – С. 430–440.

21. Camati R. S. Using the projective thematic apperception test for automatic personality recognition in texts / R. S. Camati, A. A. Scaduto, F. Enembreck. // *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. – 2021. – С. 78–85.

22. Ihsan Z. The new technologies in personality assessment: A review / Z. Ihsan, A. Furnham. // *Consulting Psychology Journal: Practice and Research*. – 2018. – №2. – С. 147.

23. Yuan Y. Application of Imagery Communication Psychotherapy in Projective Test / Y. Yuan, M. Y. Cao, Y. Liu. // *Psychology and Behavioral Sciences*. – 2018. – №7. – С. 38–44.

24. Matarazzo J. D. Computerized clinical psychological test interpretations: Unvalidated plus all mean and no sigma / Matarazzo. // *American Psychologist*. – 1986. – №1. – С. 14.

25. Abrams Z. AI is changing every aspect of psychology. Here's what to watch for / Zara Abrams. // *Monitor on Psychology*. – 2023. – №54. – С. 46.
26. Wysa [Електронний ресурс] – Режим доступу до ресурсу: <https://www.wysa.com/>.
27. Woebot [Електронний ресурс] – Режим доступу до ресурсу: <https://woebothealth.com/>.
28. Youper [Електронний ресурс] – Режим доступу до ресурсу: <https://www.youper.ai/>.
29. EMDR Helper [Електронний ресурс] – Режим доступу до ресурсу: <https://play.google.com/store/apps/details?id=fr.mrgaroby.emdrtherapy&hl=uk&pli=1>.
30. Beck Depression Inventory
31. Dashenkov D. Methods of Multilanguage Question Answering / D. Dashenkov, K. Smelyakov, O. Turuta. // 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T). – 2021. – №1.
32. Thematic Apperception Test (TAT) interpretation: Practice recommendations from a survey of clinical psychology doctoral programs accredited by the American Psychological Association.
33. An inventory for measuring depression / [A. T. Beck, C. H. Ward, Mendelson та ін.]. // *Archives of General Psychiatry*. – 1961. – №4. – С. 561–571.
34. Dana R. H. A Manual for Objective TAT Scoring / Richard Dana. // REGIONAL RESEARCH INSTITUTE FOR HUMAN SERVICES. – 1956. – №34.
35. Люліна К. П., Турута О. П. Дослідження методів розпізнавання іменованих сутностей у неструктурованому тексті // Сучасні напрями розвитку інформаційнокомунікаційних технологій та засобів управління. Тези доповідей дванадцятої міжнародної науково-технічної конференції 27 – 28 квітня 2022 року. – Баку – Харків – Жиліна, 2022, т. 2. с. 143.
36. Neural Natural Language Generation: A Survey on Multilinguality, Multimodality, Controllability and Learning / Erdem, Kuyu, Babii, Turuta. // *Journal of Artificial Intelligence Research*. – 2022. – №73. – С. 1131–1207.

37. GPT-4 LLM [Електронний ресурс] – Режим доступу до ресурсу: <https://openai.com/index/gpt-4/>.
38. Llama 2 LLM [Електронний ресурс] – Режим доступу до ресурсу: <https://llama.meta.com/>.
39. Liu Y. RoBERTa: A Robustly Optimized BERT Pretraining Approach / Y. Liu, M. Ott, N. Goyal. – 2019.
40. PaLM 2 LLM [Електронний ресурс] – Режим доступу до ресурсу: <https://ai.google/discover/palm2/>.
41. Raffel C. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer / C. Raffel, N. Shazeer, K. Lee. // Journal of Machine Learning Research. – 2020. – №21. – С. 1–67.
42. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J.Devlin, M. Chang, K. Lee, K. Toutanova. // North American Chapter of the Association for Computational Linguistics. – 2019. – С. 4171–4186.
43. Ukrainian News Corpus As Text Classification Benchmark / D.Panchenko, D. Maksymenko, O. Turuta, M. Luzan. // Proceedings of the 17th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer.. – 2021. – С. 717–726.
44. Vue.js framework documentation [Електронний ресурс] – Режим доступу до ресурсу: <https://vuejs.org/>.
45. ASP.NET Web framework documentation [Електронний ресурс] – Режим доступу до ресурсу: <https://learn.microsoft.com/en-us/aspnet/overview>.
46. MongoDB documentation [Електронний ресурс] – Режим доступу до ресурсу: <https://www.mongodb.com/>.
47. Azure AI Studio documentation [Електронний ресурс] – Режим доступу до ресурсу: <https://azure.microsoft.com/en-us/products/ai-studio>.
48. Світенко Г.М., Турута О.П. Дослідження моделей та методів проєктивних методик тестування на основі природної мови. // 28-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті». Зб. матеріалів форуму. Т. 6. – 2024. – С. 423–424.

**ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ ЗА НАУКОВИМИ НАПРЯМАМИ
КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ**

31. Dashenkov D. Methods of Multilanguage Question Answering / D. Dashenkov, K. Smelyakov, O. Turuta. // 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T). – 2021. – №1.

35. Люліна К. П., Турута О. П. Дослідження методів розпізнавання іменованих сутностей у неструктурованому тексті // Сучасні напрями розвитку інформаційнокомунікаційних технологій та засобів управління. Тези доповідей дванадцятої міжнародної науково-технічної конференції 27 – 28 квітня 2022 року. – Баку – Харків – Жиліна, 2022, т. 2. с. 143.

36. Neural Natural Language Generation: A Survey on Multilinguality, Multimodality, Controllability and Learning / Erdem, Kuyu, Babii, Turuta. // Journal of Artificial Intelligence Research. – 2022. – №73. – С. 1131–1207.

43. Ukrainian News Corpus As Text Classification Benchmark / D.Panchenko, D. Maksymenko, O. Turuta, M. Luzan. // Proceedings of the 17th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer.. – 2021. – С. 717–726.