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 емпатії штучного інтелекту у комунікації з людиною
(тема)

Виконав:
здобувач другого року навчання
групи ІПЗм-23-1

Андрій СТАРІКОВ
(Власне ім'я, ПРІЗВИЩЕ)

Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)
Тип програми освітньо-наукова

Керівник доц. Віктор КАУК
(посада, Власне ім'я, ПРІЗВИЩЕ)

Допускається до захисту
Зав. кафедри _____ Кирило СМЕЛЯКОВ
(підпис) (Власне ім'я, ПРІЗВИЩЕ)

2025 р.

Харківський національний університет радіоелектроніки

Факультет комп’ютерних наук
Кафедра програмної інженерії
Рівень вищої освіти другий (магістерський)
Спеціальність 121 – Інженерія програмного забезпечення
Тип програми освітньо-наукова програма
Освітня програма Інженерія програмного забезпечення
(шифр і назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«___» _____ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Старікову Андрію Вікторовичу
(прізвище, ім’я, по батькові)

1. Тема роботи «Дослідження моделей емпатії штучного інтелекту у комунікації з людиною»

Затверджена наказом по університету від 15.04.2025 р. № 290 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 12.06.2025

3. Вихідні дані до роботи огляд існуючих емпатичних моделей штучного інтелекту в науковій літературі

4. Перелік питань, що потрібно опрацювати в роботі

провести аналіз та порівняння існуючих емпатичних систем, що представлені в науковій літературі; провести порівняльний аналіз нейронних мереж, враховуючи їхні архітектурні особливості; змоделювати архітектуру нейронної мережі, яка буде ефективно забезпечувати природню емпатичну комунікацію

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	17.04.2025	*виконано*
2	Аналіз предметної галузі	25.04.2025	*виконано*
3	Огляд й аналіз літературних наукових джерел	05.05.2025	*виконано*
4	Постановка задачі	12.05.2025	*виконано*
5	Підготовка до апробації результатів дослідження. Публікація матеріалів	20.05.2025	*виконано*
6	Теоретичне дослідження	28.05.2025	*виконано*
7	Підготовка пояснювальної записки	02.06.2025	*виконано*
8	Підготовка презентації та доповіді	02.06.2025	*виконано*
9	Перевірка на плагіат	03.06.2025	*виконано*
10	Нормоконтроль	07.06.2025	*виконано*
11	Рецензування	10.06.2025	*виконано*
12	Попередній захист	10.06.2025	*виконано*
13	Занесення диплома в електронний архів	11.06.2025	*виконано*
14	Допуск до захисту у зав. кафедри	11.06.2025	*виконано*

Дата видачі завдання 17 квітня 2025р.

Студент (ка) _____ Андрій СТАРІКОВ
(підпис)

Керівник роботи _____ доц. Віктор КАУК
(підпис) (посада, Власне ім’я, ПРІЗВИЩЕ)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка містить: 89 с., 14 рис., 4 табл., 41 джерел.

ДІАЛОГОВІ СИСТЕМИ, ЕМПАТИЧНА ВЗАЄМОДІЯ, ЕМПАТІЯ, НЕЙРОННІ МЕРЕЖІ, ШТУЧНА ЕМПАТІЯ, ШТУЧНИЙ ІНТЕЛЕКТ

ARTIFICIAL INTELLIGENCE, DIALOGUE SYSTEMS, COMPUTATIONAL EMPHATY, EMPATHIC INTERACTION, EMPATHY, NEURAL NETWORKS

Об'єктом дослідження є процес емпатичної взаємодії між людиною та системами штучного інтелекту.

Метою роботи є розробка комплексної архітектури моделі штучної емпатії для діалогових систем.

Методами розробки та проектування є системний аналіз для визначення взаємозв'язків між компонентами моделі, математична формалізація для опису процесів, функціональне моделювання для аналізу механізмів емпатичної взаємодії та багатокритеріальний аналіз для вибору оптимальної архітектури нейронної мережі.

У результаті кваліфікаційної роботи було розроблено архітектуру моделі штучної емпатії, що складається з чотирьох основних функціональних блоків: аналізу та оцінки, емоційного аналізу, модуляції емпатії та генерації відповіді, з інтегрованим зовнішнім модулем пам'яті. На основі багатокритеріального аналізу обрано оптимальну нейромережеву архітектуру Transformer для реалізації моделі та розроблено план її експериментальної перевірки.

DIALOGUE SYSTEMS, EMPATHIC INTERACTION, EMPATHY, NEURAL NETWORKS, COMPUTATIONAL EMPHATY, ARTIFICIAL INTELLIGENCE

The object of research is the process of empathic interaction between humans and artificial intelligence systems.

The purpose of the work is to develop a comprehensive architecture of artificial empathy model for dialogue systems.

Development and design methods include system analysis for determining relationships between model components, mathematical formalization for process description, functional modeling for analysis of empathic interaction mechanisms, and multi-criteria analysis for selecting optimal neural network architecture.

As a result of the qualification work, an artificial empathy model architecture was developed, consisting of four main functional blocks: analysis and assessment, emotional analysis, empathy modulation and response generation, with an integrated external memory module. Based on multi-criteria analysis, the optimal Transformer neural network architecture was selected for model implementation and an experimental verification plan was developed.

Заява щодо самостійного виконання кваліфікаційної роботи та можливості її публікації в електронному архіві відкритого доступу ElArKhNURE.

Завідувачу кафедри
ПІ
проф. Кирилу СМЕЛЯКОВУ

ЗАЯВА

щодо самостійності виконання кваліфікаційної роботи та можливості її публікації (та/або публікації анотації кваліфікаційної роботи) в електронному архіві відкритого доступу ElAr KhNURE

Я, Старіков Андрій Вікторович, здобувач вищої освіти на другому (магістерському) рівні вищої освіти академічної групи ІПЗм-23-1 кафедра програмної інженерії заявляю: моя кваліфікаційна робота на тему «Дослідження моделей емпатії штучного інтелекту у комунікації з людиною», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в репозиторії "ElArKhNURE".

Погоджуюся з авторським договором, відповідно до Положення про репозиторій ХНУРЕ "ElArKhNURE". Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений (а) з вимогами академічної доброчесності, згідно з якими виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

Дата Підпис

ЗМІСТ

ВСТУП

У сучасному світі стрімкого розвитку штучного інтелекту та його інтеграції в повсякденне життя людини особливої актуальності набуває проблема емоційної взаємодії між людиною та машиною. Хоча сучасні моделі штучного інтелекту демонструють вражаючі результати в обробці та вирішенні складних когнітивних завдань, їхня здатність до емпатичної взаємодії залишається обмеженою. Це створює суттєвий бар'єр для їх ефективного використання у важливих сферах життя, таких як психологічна підтримка, освіта, охорона здоров'я та соціальна робота.

Актуальність дослідження підтверджується увагою наукової спільноти до цього питання та наявністю великого різноманіття моделей штучної емпатії. Існуючі моделі базуються на різних підходах та критеріях, однак все ще не повністю враховують багаторівневу природу емпатичної взаємодії з людиною.

Метою дослідження є розробка комплексної архітектури моделі штучної емпатії. Для досягнення поставленої мети необхідно вирішити наступні завдання:

- провести аналіз існуючих підходів до моделювання та відтворення емпатії в системах штучного інтелекту;
- розробити математичну модель емпатичної взаємодії, що враховує когнітивні, регуляторні та комунікативні аспекти емпатії;
- створити архітектуру нейронної мережі, здатної реалізувати запропоновану модель;
- спланувати підхід до експериментальної перевірки розробленої системи.

Об'єктом дослідження є процес емпатичної взаємодії між людиною та системами штучного інтелекту. Предметом дослідження є моделі емпатії штучного інтелекту в комунікації з людиною.

У роботі використано наступні методи дослідження: системний аналіз для визначення взаємозв'язків між рівнями та компонентами моделі, математична формалізація для опису процесів та компонентів, функціональне моделювання для

аналізу функцій розпізнавання емоційних станів та механізмів генерації емпатичної відповіді.

Результатом роботи є розроблена архітектура моделі штучної емпатії, що поєднує переваги сучасних нейромережевих підходів для забезпечення природної та контекстно-залежної емпатичної взаємодії. Практична цінність дослідження полягає в можливості використання розробленої моделі для створення більш природних та емоційно чутливих діалогових систем, що можуть знайти застосування в різних сферах людської діяльності.

Деякі результати досліджень апробовані на Міжнародній науково-практичній конференції в форматі науково-дослідницької статті.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Аналіз природньої та штучної емпатії.

Емпатія, або здатність співпереживати іншим, вважається важливим фактором покращення міжособистісних відносин і взаємодії, згідно з дослідженнями в кількох наукових дисциплінах, зокрема в галузях промислової та організаційної психології, розвитку лідерських якостей, соціальної психології, ведення переговорів, нейронауки та психічного здоров'я. Це складна, багатовимірна та високорівнева навичка соціального інтелекту.

Емпатія зазвичай визначається як здатність розуміти та ділитися почуттями, поглядами та досвідом іншої людини, при цьому зберігаючи неосудливу позицію. Її також описують як здатність бути вразливим з іншими людьми у їхній вразливості. Емпатія є багатогранним процесом, що включає в себе прийняття перспективи іншої людини, емоційний відгук, дію та уявне розуміння потреб і досвіду інших. Це не лише розпізнавання намірів інших, а й формування соціальних зв'язків, заснованих на турботі та розумінні [1].

Згідно з дослідженнями Гоулмана емпатія включає щонайменше три її аспекти: когнітивну емпатію, афективну емпатію та емпатичну турботу [2].

Когнітивна емпатія – це здатність розуміти точку зору іншої людини. Вона тісно пов'язана з поняттям прийняття точки зору та часто інтерпретується як здатність «поставити себе на місце іншого».

Афективна емпатія або емоційна емпатія – це здатність відчувати те, що відчуває інша людина. Вона характеризується як «твій біль у моєму серці». У складних ситуаціях здатність швидко відчувати без глибокого аналізу є важливою навичкою, пов'язаною з еволюцією людини.

Співчутлива емпатія або емпатична турбота, – це здатність відчути, що саме потрібно іншій людині і вжити відповідних заходів. Цей аспект емпатії виходить за межі простої здатності приймати перспективу або розділяти почуття інших. Він передбачає демонстрацію корисної поведінки, яка враховує отриману інформацію про потреби інших для ефективнішого розв'язання проблем [2].

Іншою важливою концепцією, що інтегрується в багато підходів до вивчення емпатії, є явище віддзеркалення. Суть цього підходу полягає в тому, що наш мозок і тіло можуть потребувати створення уявного відображення знань, думок, уподобань або почуттів іншої людини, яке збігалося б із тим, що ми самі могли б думати, сприймати чи відчувати в аналогічній ситуації. Це необхідно для того, щоб краще розуміти, співпереживати або адекватно реагувати на переживання іншої особи.

Філософські та психологічні теорії емпатії тривалий час досліджували існування механізмів віддзеркалення. Однак інтерес до цих підходів значно зріс після відкриття так званих «дзеркальних нейронів». Ці нейрони, виявлені в моторній системі мавп, активуються як у ситуаціях, коли мавпа спостерігає за дією іншої, наприклад, коли та простягає руку, так і коли вона виконує таку саму дію самостійно [3].

Схожі процеси спостерігаються й у людей. Зокрема, активність у певних ділянках мозку підвищується як під час спостереження за емоційними або сенсорними станами інших, так і при переживанні цих станів особисто. Це особливо помітно у випадках, коли йдеться про фізичний біль, адже активуються схожі нейронні мережі, що забезпечують спільність переживання [3].

Коли йдеться про визначення емпатії для штучних агентів, існує багато моделей, які трактують емпатію як вроджену реакцію, що ускладнює її реалізацію в штучних системах. З цієї причини різні дослідники у сфері взаємодії людини та робота (HRI – human-robot interaction), запропонували паралельні або реактивні моделі емпатії. Вони є синонімами іншого аспекту емпатії — соматичної емпатії, яка визначається як здатність спонтанно імітувати фізичні реакції, зокрема міміку та жести.

Розуміння відмінностей між штучною емпатією та природною (людською) є критично важливим перед переходом до реалізації штучної емпатії та її численних елементів. У науковій літературі штучна емпатія асоціюється з кількома термінами, такими як «емпатичні обчислення», «афективні обчислення» і «емоційний

інтелект». Однак концепція всіх цих визначень зводиться до ідеї «імітації емпатії штучними агентами».

Різноманітні системи штучного інтелекту часто називають емпатичними.

Проте дослідникам доводиться стикатися з численними підходами до розуміння емпатії. Деякі автори зазначили, що «визначень емпатії існує, мабуть, стільки ж, скільки й авторів у цій галузі». Незважаючи на це розмаїття, багато дослідників мають сильні та часто протилежні думки про те, яке саме визначення є «істинним» чи «правильним». Це часто спонукає їх критикувати проєкти емпатичного ШІ, заявляючи що якесь твердження не є справжньою емпатією. У результаті виникає загальна плутанина як серед розробників ШІ, так і серед користувачів [3].

Спираючись на існуючі дослідження міжособистісної емпатії та штучної емпатії в галузях комп'ютерних наук і робототехніки, дослідники визначають штучну емпатію як кодифікацію когнітивної та афективної емпатії людини за допомогою обчислювальних моделей у процесі проєктування агентів ШІ. Простіше кажучи, штучну емпатію можна розглядати як процес програмування емпатії в алгоритми та агенти ШІ [4].

Штучна емпатія складається з трьох основних компонентів: перспективне мислення (perspective-taking), емпатична турбота (empathetic concern) та емоційне зараження (emotional contagion), ці компоненти можна розділити на дві ключові категорії когнітивний і афективний аспекти:

- перспективне мислення відображає когнітивний аспект штучної емпатії і включає здатність ШІ розуміти думки, потреби й цілі людини через моделювання її міркувань і аналіз контексту ситуації;
- емпатична турбота – афективний аспект, що характеризує здатність ШІ демонструвати увагу та турботу до емоційного стану людини і включає інтуїтивне розуміння емоцій та їх вираження, що сприяє створенню довіри у взаємодії;

– емоційне зараження, другий афективний аспект, який передбачає здатність ШІ відображати емоційні стани людини через відповідні реакції, наприклад, в голосі, міміці чи поведінці [4].

Разом ці три компоненти формують вищий рівень конструкції штучної емпатії, забезпечуючи її комплексність і багатогранність. Їх об’єднання дозволяє створити більш повноцінну модель емпатії, ніж окреме використання когнітивного чи афективного підходів.

Розробка методик створення емпатії та інших емоційних реакцій є однією з основних цілей досліджень у сфері штучного інтелекту. Перше покоління когнітивної науки, представлене репрезентаціоналізмом, вважало, що сприйняття та вираження емоцій є ізольованими процесами, однак це призводило до невідповідності між теоретичними моделями емоцій та реальними способами їх вираження [5].

На противагу цьому, дослідження емоцій у контексті «усвідомленого розуміння» показали, що емоції охоплюють різні аспекти взаємодії всього тіла з фізичним і соціальним середовищем і не обмежуються лише мозком чи нервовою системою. Цей підхід дозволяє позбутися традиційного дизайну штучних емоцій. Концепція емпатії, як її визначив Когут, базується на когнітивній моделі спостереження та інтроспекції, що адаптована для сприйняття складної психологічної конфігурації іншої людини [5].

Штучна емпатія в контексті втілених емоцій не розглядає систему сприйняття емоцій як окрему систему, оскільки сенсорно-моторна система здатна реконструювати емоції через імітацію відповідного фізичного стану. Емоційне вираження включає помітні зміни у виразі обличчя, голосі, рухах тіла та активності, причому між вираженням і станом існує певна двозначність. Ця двозначність пояснюється єдністю та нерозривністю сприйняття та вираження емоцій [5].

Одночасно з цим виокремлюються дві ключові відмінності між людською емпатією та штучною емпатією.

По-перше, агенти ШІ не можуть відчувати чи переживати емоції, як це роблять люди, принаймні на поточному рівні розвитку технологій. Вони лише

здатні імітувати людську емпатію, демонструючи псевдоментальні риси емпатії. Це означає, що штучна емпатія є кодифікованою здатністю, реалізованою через обчислювальні алгоритми.

По-друге, компоненти штучної емпатії не слідують тій самій ієрархічній послідовності розвитку, що й компоненти людської емпатії. Наприклад, базовий рівень людської емпатії — емоційне зараження — є природним і автоматичним для людини, тоді як його впровадження у комп'ютерну систему є складним завданням. Натомість когнітивне прийняття перспективи, яке потребує зусиль від людини, відносно легше реалізувати в ШІ, оскільки воно базується на логічному розумінні й може бути адаптоване через машинне навчання на основі накопичених даних [4].

Спираючись на проведений аналіз та дослідивши складову та природу штучної емпатії, наочно зображено компоненти штучно емпатії (див. рисунок 1.1).

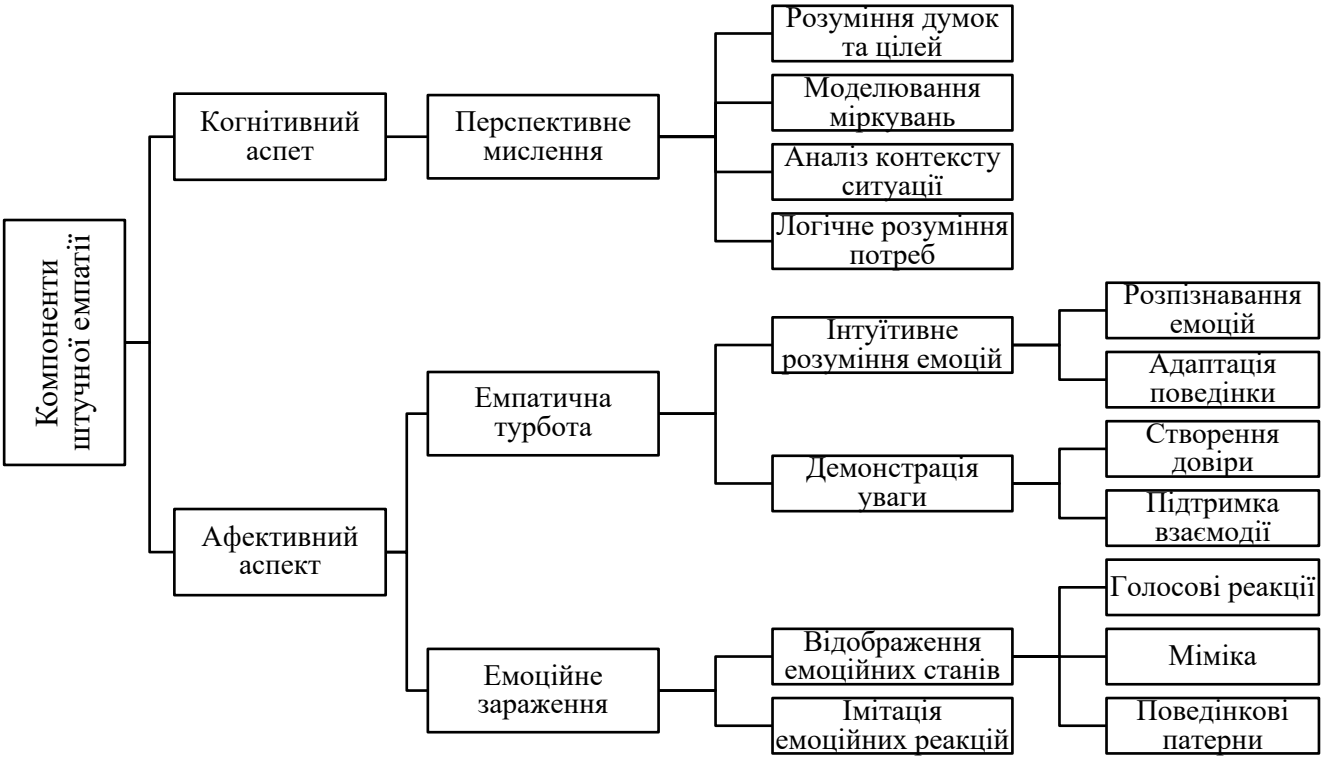

Рисунок 1.1 – Компоненти штучної емпатії (виконано самостійно).

Отже, з огляду на вищезазначене, емпатія виступає одним із ключових факторів у людській взаємодії, що базується на складних нейробіологічних механізмах та проявляється через когнітивні, емоційні та поведінкові аспекти. Особливу роль у цьому відіграє система дзеркальних нейронів, яка забезпечує

автоматичне віддзеркалення станів інших людей. Розуміння природи емпатії та її механізмів має важливе значення для розвитку штучного емоційного інтелекту, хоча наразі штучні системи здатні лише імітувати зовнішні прояви емпатії, не досягаючи глибини та комплексності людського емпатичного досвіду. Це підкреслює унікальність людської емпатії та вказує на необхідність подальших досліджень для створення більш досконалих моделей штучної емпатії.

1.2 Використання емпатії в галузі штучного інтелекту.

У контексті штучного інтелекту та технологій емпатія була досліджена в кількох галузях. Наприклад, супутникові роботи та чат-боти на сьогодні здатні виявляти людські емоції та реагувати на них емпатійним способом. McQuiggan & Lester представили підхід, що заснований на даних, який розвиває емпатію в ШІ через аналіз соціальних взаємодій між людьми. Empath використовує емоційний ШІ для визначення емоцій, аналізуючи голос людини в режимі реального часу незалежно від мови. Інші ініціативи, як Sensum, опрацьовують дані водія (наприклад: вираз обличчя, голос і сенсорні вхідні дані) для покращення безпеки та комфорту під час водіння [1].

Більше того, системи ШІ, такі як XiaoIce, намагаються задовольнити людські потреби в спілкуванні та соціальній приналежності, створюючи довготривалі емоційні зв’язки. XiaoIce оптимізує свою взаємодію з користувачами, орієнтуючись на довготривалу залученість, яка вимірюється тривалістю розмови на сесію [1].

Попри ці досягнення, існують деякі виклики та обмеження в галузі штучної емпатії.

Хоча емпатія є природним елементом людських взаємодій, спрямованих на гармонійні стосунки, попередні дослідження вказують, що споживачі часто неохоче використовують ШІ для завдань, які вимагають суб’єктивності, інтуїції та емоцій, оскільки ШІ не вистачає емпатії для виконання таких завдань. Ця позиція свідчить про необхідність кращої інтеграції емпатії у застосунки ШІ, через що деякі

маркетингові науковці передбачають "емпатичний ШІ" як наступний еволюційний етап ШІ [4].

Зі зростанням популярності голосових асистентів, таких як Siri чи Alexa, аналіз голосових даних для розпізнавання емоцій споживачів стає дедалі актуальнішим. Наприклад, вже були спроби використовувати голосовий аналіз для виявлення специфічних емоцій, таких як збентеження. Можна вважати, що емпатична турбота в ШІ базується на здатності системи розпізнавати емоції та адаптувати свою поведінку для створення враження турботи. Цей напрямок активно розвивається та має значний потенціал у сфері маркетингу та інших взаємодій із користувачами [4].

Розвиток таких технологій, як GPT-3, створює реалістичну базу для вербальної емпатії у взаємодії людини з машиною. Наприклад, штучні агенти можуть допомагати у психотерапії, сприяючи позитивному емоційному зв’язку [5].

Слід також звернути увагу на те, що емоція як прояв емпатії, є основоположним аспектом емпатії, проте її визначення та характеристика – непросте завдання. Філософ Джессі Прінц окреслює це явище через призму двох фундаментальних проблем: "проблеми складових" та "проблеми надлишковості".

Проблема складових включає питання про те, які компоненти емоції (оцінювальні, фізіологічні, феноменологічні, експресивні, поведінкові чи ментальні) є визначальними для її ідентифікації та виявлення. Проблема надлишковості, в свою чергу, досліджує механізми практичної взаємодії цих компонентів [1].

У системах штучного інтелекту емоції зазвичай оцінюються через опосередковані дані: міміку, інтонацію, жестикуляцію тощо. Втім, ці показники можуть слугувати лише індикаторами, а не достовірними маркерами емоційних чи когнітивних станів людини. Старк і Хой наголошують, що системи ШІ часто не враховують суб'єктивний емоційний досвід, що призводить до розбіжності між змодельованим та реальним досвідом користувачів. Відтак, емоційні системи ШІ можуть виявитися неетичними через брак ширшого соціального контексту, який потребується для досягнення справедливих результатів [1].

Останні досягнення у сфері штучного інтелекту змістили фокус до обчислювального підходу, де емпатія прогнозується на основі текстових корпусів і кількісно визначається через маркування емоцій та переживання стресу. Хоча більшість досліджень традиційно зосереджувались на здатності штучного агента до прояву емпатії, виражена штучна емпатія дедалі більше визнається критично важливою для досягнення успішних результатів терапії [6].

Попри те, що емпатія здатна поліпшити комунікативні можливості ШІ, її впровадження потребує виваженого підходу задля уникнення непередбачуваних негативних наслідків. Надмірна емпатія, як зазначає Джессі Прінц, може спричинити упередження, особливо в соціальному, расовому чи політичному контекстах. Це обмеження особливо помітне в ситуаціях конкуренції чи конфлікту. Отже, системи ШІ з інтегрованою емпатією мають розроблятися на засадах справедливості та відповідальності [1].

Для розв'язання цієї проблеми новітні дослідження наголошують на необхідності регулювання емпатії в системах ШІ, з метою запобігання виникненню упереджень. Високий рівень відповідальності здатен мінімізувати негативні прояви емпатії, що підтверджується дослідженнями Блейдера і Ротмана, які доводять: за умови суворої відповідальності емпатія не призводить до упереджень. Таким чином поєднання емпатії та відповідальності дає змогу системам ШІ створювати ефективнішу, справедливішу та етично безпечну взаємодію, корисну як для користувачів, так і для самих систем [1].

Не менш важливим є дослідження в напрямку взаємодії штучного інтелекту та людини. Зокрема, деякі науковці виокремлюють наукову область – HCD (Human Centered Design) та формують три основні відмінності між потенційним розвитком штучної емпатії для HCD з іншими дисциплінами:

- штучна емпатія має зосереджуватись на взаємодії між людьми та їхнім контекстом, а не на взаємодії людей із ШІ-агентом;
- емпат в HCD повинен постійно уявляти альтернативи, які могли б призвести до потенційно кращої ситуації, та залучати людей до процесу уяви;

– штучна емпатія повинна надавати цінні інсайти або залучати та спонукати людину до вираження уяви, що вимагає від агента комунікувати таким чином, щоб людина мала можливість розуміти, відчувати та уявляти [7].

Таким чином, з урахуванням огляду предметної галузі можна зазначити, що емпатія є складним багатовимірним явищем, яке включає когнітивні, афективні та поведінкові компоненти. Природна емпатія базується на глибоких нейробіологічних механізмах, а штучна емпатія, є спробою кодифікувати ці процеси через обчислювальні моделі.

Іноваційність досліджень в обраній галузі розкривається через детальне вивчення трьох основних компонентів штучної емпатії: перспективного мислення, емпатичної турботи та емоційного зараження. Важливо відзначити, що ці компоненти в штучних системах розвиваються за іншою ієрархією, ніж у людей – когнітивні аспекти легше піддаються імплементації, ніж емоційні.

Разом з цим поточними викликами розвитку штучної емпатії є:

– фундаментальна неможливість ШІ істинно відчувати емоції, обмежуючись лише їх імітацією;
– складність розпізнавання та інтерпретації емоційних станів людини через опосередковані дані;
– ризики виникнення упереджень при надмірній емпатії;
– недовіра користувачів до емоційних можливостей ШІ у завданнях, що вимагають суб'єктивності та інтуїції;
– необхідність балансування між емпатією та відповідальністю у системах ШІ.

Перспективами розвитку в цьому напрямку може бути вдосконалення технологій розпізнавання емоцій через аналіз голосу, міміки та інших біометричних даних, оскільки багато провідних компаній намагаються покращити досвід взаємодії людини з будь-якими роботизованими системами.

Не менш важливим є розширення застосування емпатичних систем у психотерапії та охороні здоров'я та розвиток довготривалих емоційних зв'язків між людьми та ШІ-системами, а впровадження принципів Human Centered Design з

метою покращення взаємодії людини з ШІ, де агент спонукає користувача до рефлексії та побудови уяви, може створити більш справедливу та етично безпечну взаємодію людини і робота. Отже, іноваційним напрямком може бути розробка систем, які не просто імітують емпатію, але й сприяють розвитку емпатичних здібностей у людей через конструктивну взаємодію.

2 ОГЛЯД Й АНАЛІЗ ЛІТЕРАТУРНИХ, НАУКОВИХ ДЖЕРЕЛ

Актуальні дослідження когнітивної науки та штучного інтелекту зосереджені на розробці моделей емпатичної поведінки у штучних системах, що дає змогу експериментально перевіряти відповідні теоретичні припущення. Впровадження обчислювальних моделей емпатії має подвійне призначення: воно сприяє створенню більш соціально адаптованих технологічних рішень та водночас поглиблює розуміння механізмів емпатії, дозволяє тестувати теорії та шукати відповіді на різні виклики, що постають перед спільнотою.

Як зазначалося в попередньому розділі, дослідники визначають штучну емпатію як кодифікацію емпатії через когнітивний та афективний аспекти.

Афективний аспект емпатії визначає її як відтворення емоційної реакції (несвідоме у людини) і включає дзеркальну реакцію та поведінку афективного співвіднесення. В свою чергу когнітивний аспект емпатії розглядається як здатність розуміти емоційні та ментальні стани інших за допомогою когнітивних механізмів, таких як прийняття перспективи і теорія розуму. У науковців немає єдиної думки щодо того, як ці види поведінки пов’язані між собою і чи є емпатія дискретним або безперервним феноменом. Проте дослідження свідчать, що ці аспекти, або їх ще називають рівні емпатії, є взаємопов’язаними [8].

Спираючись на провідні наукові дослідження, в контексті розробки систем штучного інтелекту спостерігається цікавий парадокс: когнітивні аспекти емпатії, які для людей часто вимагають свідомих зусиль та навчання, виявляються значно простішими для реалізації в ШІ, ніж афективні компоненти, які у людей проявляються природно та автоматично.

Це пояснюється передусім природою когнітивних процесів, які піддаються формалізації та алгоритмізації. Когнітивний аспект емпатії базується на логічному аналізі, обробці інформації та прийнятті рішень на основі чітких патернів і правил. Такі процеси можна ефективно моделювати за допомогою математичних алгоритмів та методів машинного навчання, використовуючи накопичені дані для навчання систем розпізнавати контекст, аналізувати потреби користувачів та приймати відповідні рішення.

Натомість афективний аспект емпатії представляє значно більший виклик при відтворенні в ШІ. Це пов'язано з тим, що емоційні реакції та переживання базуються на складних нейробіологічних механізмах, які наразі неможливо повноцінно відтворити в штучних системах. Сучасні ШІ можуть лише імітувати емоційні реакції, але не здатні по-справжньому "відчувати" емоції. Крім того, афективні стани мають високий ступінь суб'єктивності та контекстуальної залежності, що ускладнює їх формалізацію.

Таким чином, хоча сучасні системи ШІ досягли значних успіхів у реалізації когнітивних аспектів емпатії, створення повноцінної штучної емпатії залишається складним завданням через обмеження у відтворенні афективних компонентів, які є невід'ємною частиною справжнього емпатичного досвіду.

В існуючих підходах до обчислювального моделювання емпатії популярними є підходи: підхід, що засновано на даних (data-driven) або "зверху-вниз" (top-down) та підхід, що засновано на теорії (theory-driven) або "знизу-вгору" (bottom-up), схематичний опис яких зображено на рисунок. 2.1 [9].

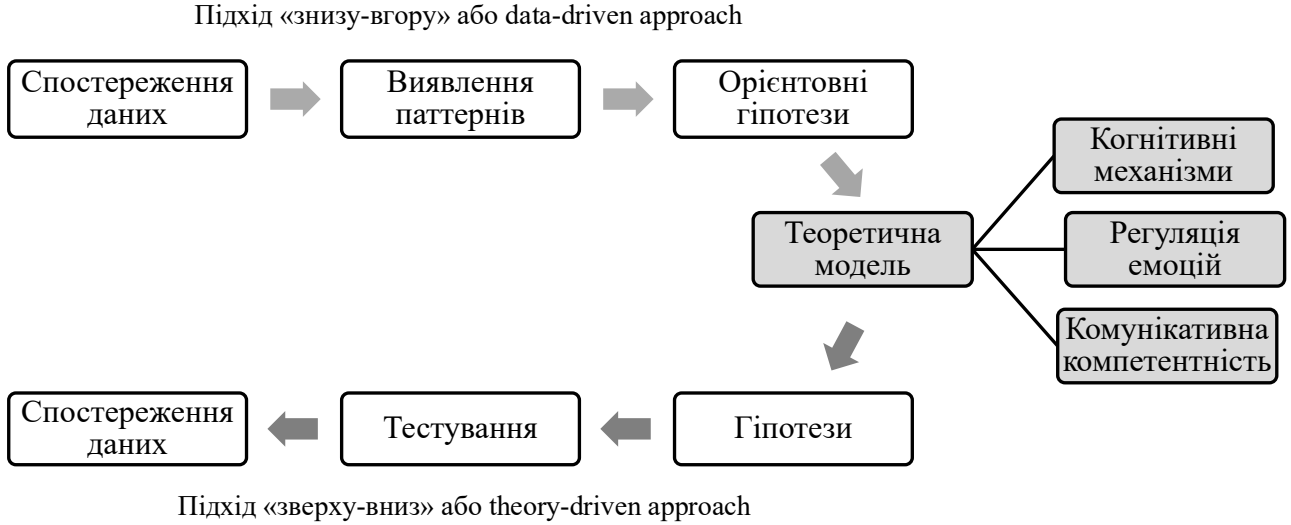

Рисунок 2.1 – Підходи реалізації моделей емпатії в штучних агентах [9]

Підхід "знизу-вгору" базується на емпіричних даних, тоді як підхід "зверху-вниз" відштовхується від теоретичних концепцій.

Підхід "знизу-вгору" починається зі збору та аналізу поведінкових даних. Дослідники спостерігають за проявами емпатії, збирають дані про емоційні реакції

та поведінку. На основі цих спостережень виявляються певні закономірності та паттерни. Після цього формуються попередні гіпотези про те, як працює емпатія. Ці гіпотези перевіряються експериментально, що дозволяє уточнити та вдосконалити розуміння процесу. В результаті формується теоретична модель, яка включає три ключові компоненти: когнітивні механізми, регуляцію емоцій та комунікативну компетентність. Цей підхід дозволяє створювати моделі на основі реальних даних та автоматизовано навчати алгоритми для прогнозування або відтворення емпатичної поведінки.

Підхід "зверху-вниз" починається з теоретичної моделі емпатії, яка базується на існуючих теоріях та концепціях. На основі цієї моделі формулюються гіпотези про те, як повинна працювати емпатія в штучних системах. Ці гіпотези потім перевіряються через експериментальне тестування, результати якого порівнюються з реальними спостереженнями. Цей метод дозволяє систематично перевіряти теоретичні компоненти та має сильну пояснювальну здатність. Він особливо корисний для розуміння механізмів, пов'язаних з емпатією, та їх подальшого моделювання в ієрархічному порядку [9].

Важливо зазначити, що обидва підходи можуть використовуватися разом у гібридних моделях, де теоретичні компоненти емпатичних механізмів моделюються окремо, а потім використовуються для навчання на основі даних. Такі гібридні методи, враховуючи успіхи в розпізнаванні емоцій та зростаючу обчислювальну потужність, показують великий потенціал для майбутнього розвитку досліджень емпатії.

Спираючись на провідні наукові джерела, в таблиці представлено короткий опис та особливості найбільш популярних емпатичних моделей, з вказанням підходів, на який спирається та чи інша модель (дивись табл. 2.1).

Таблиця 2.1 – Опис емпатичних моделей (таблиця виконана самостійно)

Автор(и)	Модель	Підхід	Опис моделі
1	2	3	4
Yalçın, DiPaola	Component Model	Theory-driven	Модель базується на Russian Doll Model та використовує еволюційний підхід, що пов'язує поведінкові паттерни з відповідними механізмами.

Продовження таблиці 2.1

1	2	3	4
			Складається з трьох взаємопов'язаних компонентів: емоційна комунікативна компетентність, регуляція емоцій та когнітивні механізми. Кожен шар має інформацію про попередній та оброблені вхідні дані. Модель дозволяє реалізацію низькорівневої емпатичної поведінки ізольовано, одночасно забезпечуючи основу для моделювання високорівневої емпатичної поведінки [8].
Asada	CDR Model	Theory-driven	Адаптує Russian Doll Model. Емпатична модель проводить паралелі між теоріями розвитку самовідмінності, пропонуючи, що розвиткова емпатія повинна бути частиною Cognitive Developmental Robotics (CDR). Модель включає три рівні розвитку самосвідомості: – екологічне "я" - базова синхронізація з середовищем; – міжособистісне "я" - здатність розрізняти себе та інших; – соціальне "я" - розуміння соціальних взаємодій. Розвиток емпатичних здібностей відбувається послідовно: емоційне зараження та моторна мімікрія, формування емоційної та когнітивної емпатії, розвиток здатності до співчуття, формування складних соціальних емоцій. Ключовим механізмом виступає фізичне втілення (embodiment), що забезпечує основу для моторного резонансу та емоційного зараження через систему дзеркальних нейронів [10].
Rodrigues та інші	Appraisal-based Model	Theory-driven	Базується на теоретичному підході De Vignemont & Singer, який стверджує, що люди не відчувають емпатію до кожної емоції та ситуації, а вибирають відповідь згідно зі своїми оцінками, судженнями. Базується на двох ключових компонентах: а) Емпатична оцінка (Empathic Appraisal) - процес, під час якого агент: 1) сприймає емоційні сигнали від інших агентів; 2) проектує себе в ситуацію іншого для розуміння його емоційного стану; 3) обирає потенційну емпатичну емоцію на основі власної оцінки ситуації та спостережених емоційних сигналів. б) Модуляція емпатичної реакції через 4 фактори: 1) подібність (наскільки схожі емоційні реакції агентів), 2) афективний зв'язок (рівень симпатії між агентами), 3) настрій агента-емпатика, 4) особистість агента (схильність відчувати певні емоції).

Продовження таблиці 2.1

1	2	3	4
			Особливістю моделі є її загальний характер – може застосовуватись для моделювання емпатії між будь-якими агентами, а не лише між агентом і користувачем, що розширює можливості для створення багатоагентних віртуальних середовищ з емоційно насиченою взаємодією [11].
Boukricha та інші	EMMA Framework	Theory-driven	Використовує модель пізньої оцінки De Vignemont & Singer. Фреймворк складається з трьох модулів: – механізм емпатії (Empathy Mechanism) - базується на внутрішній імітації виразів обличчя співрозмовника через систему розпізнавання 44 лицьових дій; – модуляція емпатії (Empathy Modulation) - регулює силу емпатичної реакції на основі декількох факторів: настрій агента, ступінь симпатії до співрозмовника, рівень знайомства; – вираження емпатії (Expression of Empathy) - проявляється через: вирази обличчя, просодію мови, частоту моргання та дихання, вербальні реакції. Особливість моделі полягає в тому, що вона дозволяє віртуальному агенту проявляти різні ступені емпатії залежно від контексту та взаємовідносин [12].
Leite та інші	iCat robot	Theory-driven	Дослідники розробили емпатичну модель, що складається з 5 ключових компонентів: – виявлення афекту – відстеження в реальному часі емоційного стану користувача через візуальні сигнали; – емпатична оцінка – оцінка ситуації з точки зору користувача і генерування відповідної емоційної реакції, що включає короткострокові емоційні реакції (через міміку) та довгострокові зміни настрою; – підтримуюча поведінка – набір дій для зменшення стресу користувача, включаючи: інформаційну підтримку (поради), практичну допомогу, підтримку самооцінки, емоційну підтримку; – пам'ять про минулі взаємодії – робот запам'ятовує важливі моменти попередніх взаємодій для побудови відносин; – вибір дії – модуль, що обирає найбільш доречні дії (вирази обличчя та мовлення) на основі інших компонентів. Модель була реалізована в роботі iCat, який грав у шахи з дітьми. Дослідження показало, що така емпатична модель допомогла підтримувати стабільний рівень соціальної присутності та залученості дітей протягом 5 тижнів взаємодії [13].
Ochs та інші	Formal Empathy Model	Theory-driven	Базується на теоретичних формулюваннях Scherer. Представляє емоції на основі типу, інтенсивності, її цілі, тригерної події та наміру, на який впливає подія. Емпатичні емоції використовують додаткову змінну для цільових емоцій.

Продовження таблиці 2.1

1	2	3	4
			Притаманне обмеження у вигляді вузького погляду на емпатію, оскільки модель не враховує низькорівневі процеси та високорівневі емпатичні процеси [9].
McQuiggan та інші	CARE Framework	Data-driven	Гібридна модель, що навчається на даних соціальної взаємодії людина-агент в симуляційному середовищі Crystal Island. Модель здатна витягувати наміри, дії, вік, стать, афективні стани та біофідбек. Використовує теоретичну модель Davis та навчається на даних взаємодії з користувачем. Використовує Interpersonal Reactivity Index для вимірювання емпатичної природи користувачів, а їх цільова орієнтація використовується для навчання моделі за допомогою Naïve Bayes [9].
Xiao та інші	Prosodic Model	Data-driven	Класифікує рівні емпатії терапевта використовуючи характеристики мовлення: висоту, енергію, джитер (коливання висоти тону), шиммер (коливання амплітуди) та тривалість висловлювання. Аудіозаписи оцінювались трьома людьми за системою MITI для класифікації рівнів емпатії за сімома категоріями. Основні елементи моделі включають: – аналіз п'яти характеристик мовлення (перераховані вище); – квантування цих характеристик на три рівні (низький, середній, високий) для створення патернів просодії; – обчислення розподілу цих патернів протягом сеансу терапії. Ключовим відкриттям стало те, що висока висота голосу та гучність терапевта негативно корелюють з рівнем емпатії, що узгоджується з попередніми психологічними дослідженнями. Модель розглядає емпатію на рівні всього сеансу терапії, а не окремих моментів, і враховує патерни обох учасників - терапевта, і пацієнта [14].
Gibson та інші	Psycho-linguistic Model	Data-driven	Модель прогнозує емпатію терапевта, яка оцінюється за шкалою MITI. Модель використовується для прогнозування емпатії терапевта під час мотиваційних інтерв'ю, використовуючи три типи мовних характеристик: – психолінгвістичні норми (PNF) – 13 лінгвістичних вимірів, таких як конкретність мови, образність, емоційна валентність та вік засвоєння слів; – категорії LIWC (Linguistic Inquiry and Word Count) – 32 психологічні категорії слів, включаючи афективні, когнітивні та перцептивні процеси; – N-грами – традиційні лексичні характеристики (одиночні слова, пари та трійки слів).

Кінець таблиці 2.1

1	2	3	4
			Важливим відкриттям стало те, що емпатичні терапевти схильні використовувати: більш абстрактну мову (негативна кореляція з конкретністю), слова сприйняття, афективну лексику, чіткі та однозначні формулювання [15].
Rashkin та інші	Empathetic Dialogues	Data-driven	Модель складається на наборі даних приблизно 25 тисяч діалогів та базується на двох ключових компонентах. Перший – датасет EmpatheticDialogues, заснований на реальних емоційних ситуаціях. Кожен діалог включає: – початкову емоційну ситуацію від першого співрозмовника (Speaker); – емпатичну відповідь від другого співрозмовника (Listener); – розмітку за 32 різними емоційними категоріями. Другий компонент – архітектура діалогової системи на основі Transformer, яка може працювати в двох режимах: режим генерації відповідей та режим вибору найбільш релевантної відповіді з наявних варіантів. Особливістю системи є те, що вона спочатку навчається на діалогах з Reddit, а потім додатково налаштовується на EmpatheticDialogues для покращення емпатичних властивостей. Експериментально така модель генерує більш емпатичні відповіді, що підтверджено автоматичними метриками та оцінками людей-експертів. Інновацією є також можливість покращувати емпатичні властивості моделі шляхом додавання зовнішніх класифікаторів емоцій без необхідності повного перенавчання системи [16].
Kumano та інші	Group Dynamics Model	Data-driven	Автори пропонують, що колективна оцінка усуває індивідуальні упередження та оцінка базується на 5-рівневій шкалі емпатії від "Сильної емпатії" до "Сильної контр-емпатії". Ключові аспекти моделі: а) Емпатія сприймається зовнішніми спостерігачами під час взаємодії між парами співрозмовників. б) Основні компоненти аналізу: 1) невербальна поведінка учасників (вираз обличчя та напрямок погляду); 2) колективні враження від групи зовнішніх спостерігачів; 3) ймовірнісне моделювання на основі байєсівських мереж. в) Модель враховує суб'єктивність сприйняття емпатії різними спостерігачами і представляє результати як розподіл ймовірностей різних рівнів емпатії (сильна, слабка, відсутня). Інновація полягає в тому, що замість пошуку єдиної "правильної" оцінки емпатії, модель визнає природну варіативність її сприйняття різними людьми і намагається відтворити цей розподіл думок математично [17].

На основі аналізу представлених в таблиці емпатичних моделей можна зробити кілька важливих спостережень щодо сучасних підходів до моделювання штучної емпатії.

Більшість розглянутих моделей (Component Model, CDR Model, Appraisal-based Model, EMMA Framework, iCat robot, Formal Empathy Model) базуються на теоретичному підході, що свідчить про важливість фундаментальних психологічних та когнітивних теорій у розробці емпатичних систем. При цьому моделі, що засновані на даних (CARE Framework, Prosodic Model, Psycho-linguistic Model, Empathetic Dialogues, Group Dynamics Model), зазвичай фокусуються на конкретних аспектах емпатії, таких як просодія мовлення, лінгвістичні особливості або групова динаміка. Це може вказувати на те, що data-driven підхід краще підходить для вирішення специфічних, добре визначених задач у межах загальної проблеми моделювання емпатії.

Також можна відзначити, що більшість моделей включають компоненти для:

- розпізнавання емоційного стану співрозмовника;
- оцінки контексту ситуації;
- генерації відповідної емпатичної реакції;
- модуляції цієї реакції залежно від різних факторів.

Окремо варто відзначити зростаючу роль мультимодальності в емпатичних моделях – багато з них враховують не лише вербальну комунікацію, але й просодію, міміку, жести та інші невербальні сигнали. Інноваційним також є впровадження механізмів колективної оцінки для зменшення суб'єктивності у сприйнятті емпатії.

Проте, існуючі моделі мають ряд суттєвих обмежень та викликів. Більшість з них розроблено для специфічних контекстів, що ускладнює їх масштабування та адаптацію до різних культурних середовищ. Технічні обмеження включають складності з обробкою мультимодальних сигналів у реальному часі та недостатню здатність до утримання довготривалого контексту розмови. Окремою проблемою залишаються етичні виклики, пов'язані з ризиками маніпулятивного використання емпатії та питаннями приватності при збиранні емоційних даних.

Подальші дослідження в цій галузі можуть розвиватися у декількох ключових напрямках. Перспективним є розвиток крос-культурних моделей емпатії, які враховують культурні відмінності у прояві та сприйнятті емпатичних реакцій. Не менш важливим напрямком залишається вдосконалення механізмів контекстуального розуміння, включаючи розвиток довготривалої пам'яті та покращення здатності до відстеження емоційної динаміки взаємодії.

Інтеграція емпатичних моделей з іншими аспектами соціального інтелекту також представляє значний дослідницький інтерес. Це включає поєднання емпатії з моделями соціального навчання та розвиток механізмів емоційної регуляції. Окремої уваги може потребувати розробка стандартизованих методів оцінки та валідації емпатичних систем, включаючи створення універсальних метрик та методів автоматичної оцінки якості емпатичних реакцій.

Наостанок, важливим залишається дослідження етичних аспектів використання штучної емпатії. Це включає розробку принципів відповідального використання емпатичних систем, створення механізмів захисту приватності користувачів та вивчення довгострокового впливу взаємодії зі штучною емпатією на психологічне благополуччя людей.

Такий комплексний підхід до подальших досліджень може сприяти створенню більш досконалих систем штучної емпатії, які будуть не лише технічно ефективними, але й етично відповідальними та соціально корисними.

Загалом, розвиток моделей емпатії демонструє рух від простих, однонаправлених систем до більш складних, контекстно-залежних рішень, які намагаються відтворити багатогранну природу людської емпатії. При цьому зберігається баланс між теоретичним обґрунтуванням та емпіричною валідацією моделей.

3 ПОСТАНОВКА ЗАДАЧІ

В умовах стрімкого розвитку штучного інтелекту та його інтеграції в повсякденне життя людини, особливої актуальності набуває проблема емоційної взаємодії між людиною та машиною. Сучасні моделі штучної емпатії ШІ демонструють вражаючі результати в обробці та вирішенні складних когнітивних завдань, проте їхня здатність до емпатичної взаємодії, на сьогодні, залишається обмеженою. Це створює суттєвий бар'єр для їх ефективного використання у важливих життєвих сферах, як наприклад: психологічна підтримка, освіта, охорона здоров'я, соціальна робота, тощо.

Актуальність проблеми підтверджує увага наукової спільноти до цього питання, а також наявність великого різноманіття моделей штучної емпатії, огляд яких був проведений в попередньому розділі. Існуючі моделі базуються на різноманітних підходах та критеріях, однак все ще не повноцінно враховують багаторівневу природу емпатичної взаємодії з людиною.

Метою дослідження є розробка комплексної архітектури моделі штучної емпатії. Для досягнення поставленої мети необхідно вирішити наступні завдання:

- провести аналіз існуючих підходів до моделювання та відтворення емпатії в системах штучного інтелекту;
- розробити математичну модель емпатичної взаємодії, що враховує когнітивні, регуляторні та комунікативні аспекти емпатії;
- створити архітектуру нейронної мережі, здатної реалізувати запропоновану модель;
- запланувати підхід до експериментальної перевірки розробленої системи.

В процесі розробки архітектури та моделі емпатичної взаємодії будуть використані наступні ключові методи.

Метод системного аналізу, як науковий метод пізнання, що представляє собою послідовність дій з установлення структурних зв'язків між змінними або елементами досліджуваної системи. Такий метод використовується з метою: визначення взаємозв'язків між рівнями та їх компонентами в моделі, встановлення

ієрархії впливів між різними факторами, виявлення ключових входів та виходів системи, тощо.

Метод математичної формалізації – це метод відображення властивостей та відношень досліджуваного об'єкта в точних математичних термінах та співвідношеннях.

Розробка математичного апарату для опису кожного рівня моделі є центральним елементом в процесі створення та обґрунтування моделі штучної емпатії. Такий підхід дозволить формалізувати процес розпізнавання емоцій, створити опис емоційної регуляції та генерації емпатичної відповіді, а також допоможе у встановленні математичних зв’язків між компонентами моделі.

Метод функціонального моделювання. Сутність методу полягає в дослідженні об'єкта, при якому основна увага приділяється його функціям та процесам, що в ньому протікають. Цей метод може бути застосований з метою аналізу функцій розпізнавання емоційних станів, моделюванні процесів, опису механізмів генерації емпатичної відповіді та визначення функціональних залежностей між компонентами.

Таким чином комбінація перерахованих методів дослідження дозволить забезпечити повноту опису моделі, гарантувати математичну строгість, створити основу для практичної реалізації та забезпечити подальшу можливість верифікації практичних результатів експерименту.

Очікуваними результатами теоретичного дослідження є:

- розробка та формалізація математичної моделі штучної емпатії;
- розробка архітектури моделі штучної емпатії нейронної мережі;
- підготовка даних та опис подальшого підходу для проведення практичного експерименту з розробленою моделлю.

При розробці моделі штучної емпатії важливо зазначити ключове технічне обмеження – модель буде працювати виключно з текстовими даними.

Таке обмеження є свідомим вибором з кількох причин. По-перше, обробка аудіо та відео даних значно ускладнює архітектуру моделі, оскільки потребує додаткових компонентів для аналізу тону голосу, міміки, жестів та інших

невербальних проявів емоцій. Це суттєво збільшує обчислювальну складність та вимоги до обчислювальних ресурсів.

По-друге, мультимодальні моделі, які працюють одночасно з текстом, аудіо та відео, вимагають значно більших та складніших наборів даних для навчання, оскільки створення якісних наборів даних, які містять синхронізовані дані різних модальностей з відповідною розміткою емоційних станів та проявів емпатії, є надзвичайно ресурсомістким завданням.

По-третє, фокус на текстовій взаємодії дозволяє більш глибоко дослідити саме лінгвістичні аспекти емпатії – як вона проявляється у виборі слів, побудові речень, загальній структурі діалогу. Це дає можливість створити більш точну та надійну модель для конкретної модальності, замість розробки менш надійної, але більш універсальної системи.

Втім, важливо розуміти, що таке обмеження означає часткову втрату інформації про емоційний стан співрозмовника, яка в реальному спілкуванні передається через невербальні канали. Тому результати роботи моделі слід інтерпретувати в контексті саме текстової комунікації.

4 ТЕОРЕТИЧНЕ ДОСЛІДЖЕННЯ

4.1 Методологічний апарат дослідження.

Теоретичне дослідження складається з декількох етапів та містить в застосуванні певні методи дослідження, які перераховано в попередньому розділі (див. рисунок 4.1).

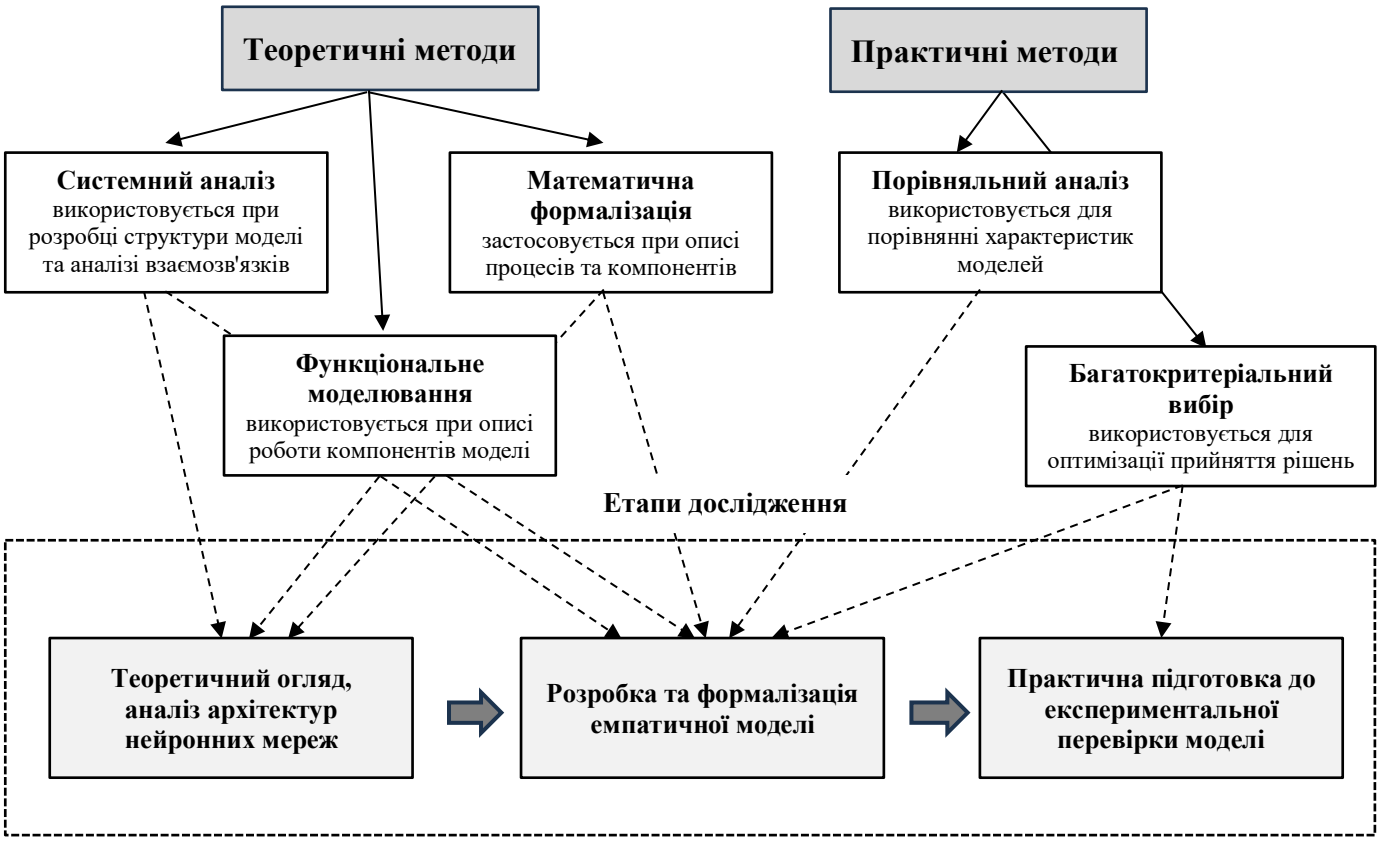

Рисунок 4.1 – Схема проведення дослідження (виконано самостійно)

Представлена схема методологічного апарату дослідження демонструє комплексний підхід до вирішення поставленої задачі через взаємозв'язок теоретичних та практичних методів дослідження. У теоретичній частині використовуються системний аналіз, математична формалізація та функціональне моделювання, що дозволяє створити фундаментальну базу дослідження. Практична частина включає багатокритеріальний вибір та порівняльний аналіз, які забезпечують вибір оптимальних рішень та їх валідацію.

Схема чітко відображає послідовність етапів дослідження від теоретичного аналізу через розробку моделі до практичної підготовки. Кожен етап підкріплений відповідними методами, що забезпечує системність та обґрунтованість дослідження. Важливо, що методи не існують ізольовано, а взаємодоповнюють один одного, створюючи цілісний методологічний апарат.

Така візуалізація дозволяє верифікувати повноту методологічної бази та коректність застосування методів на кожному етапі. Це особливо важливо при розробці складних систем, таких як емпатичні моделі штучного інтелекту, де потрібен баланс між теоретичним обґрунтуванням та практичною реалізацією.

4.2 Огляд архітектур нейронних мереж.

У цьому підрозділі представлено нейронні моделі, які популярні в сучасних діалогових системах та пов'язаних підзадачах. Розглянуті моделі включають: згорткові нейронні мережі (CNN), рекурентні нейронні мережі (RNN), базові моделі послідовність-послідовність (sequence-to-sequence), ієрархічний рекурентний енкодер-декодер (HRED), мережі пам'яті, мережі уваги, трансформер, Pointer Net і CopyNet, моделі глибокого навчання з підкріпленням, генеративно-змагальні мережі (GAN).

4.2.1 Згорткові нейронні мережі.

Згорткові нейронні мережі (CNN) – це тип глибоких нейронних мереж, що складаються із згорткових шарів, шарів об'єднання та повнозв’язних шарів. Архітектура CNN, яка представлена на рисунку 4.2, демонструє яким чином мережа обробляє текстові дані через послідовність перетворень, використовуючи згортку для витягування ознак (див. формулу 4.1):

$$G(m,n) = (f \cdot h)(m,n) = \sum_j \sum_k h(j,k) f(m-j, n-k) \tag{4.1}$$

де f – вхідна матриця (наприклад, частина тексту),

h – ядро згортки (kernel),

j, k – індекси для проходу по ядру згортки,

m – індекс рядка матриці результату,

n – індекс стовпця матриці результату.

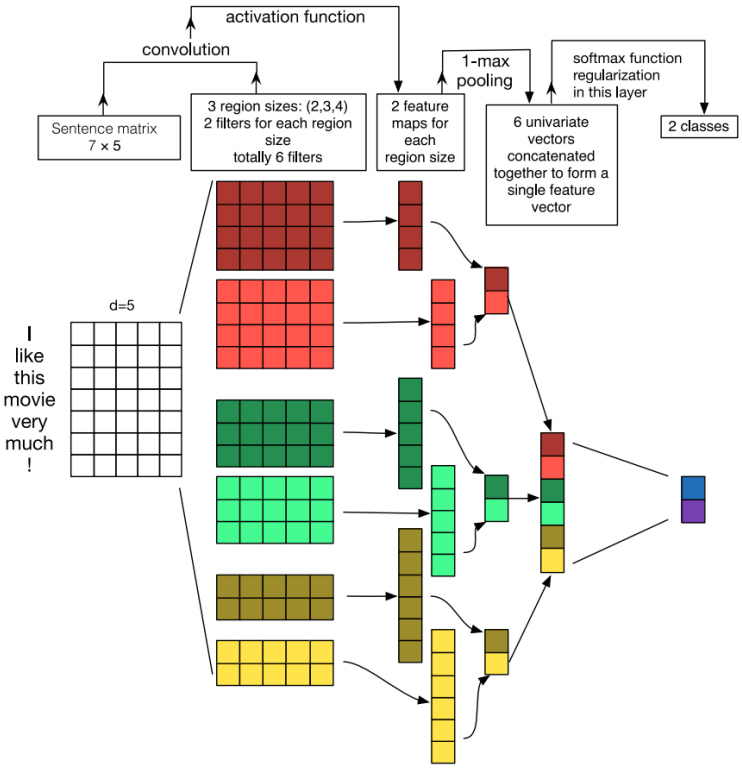

Рисунок 4.2 – Архітектура згорткової нейронної мережі для текстової класифікації [18]

Операція із формули 4.1 працює наступним чином. Спершу ядро h проходить по вхідній матриці f. Далі для кожної позиції обчислюється сума поелементних множень ядра та відповідної частини вхідної матриці. Наостанок $f\ (m-j, n-k)$ зсуває "вікно" відносно поточної позиції (m, n).

У контексті діалогових систем, CNN ефективно працює як екстрактор текстових характеристик, здатний виявляти як локальні, так і глобальні особливості тексту завдяки механізму ковзного вікна та пулінгу. Важливою перевагою є механізм розділення параметрів, який значно зменшує їх загальну кількість та покращує узагальнюючу здатність мережі [19].

Однак у сучасних діалогових системах CNN рідко використовується як основний енкодер через такі обмеження, як фіксована довжина входу та складність обробки послідовної інформації. Замість цього CNN частіше застосовується як додатковий компонент для обробки вже закодованої інформації. Наприклад, в деяких дослідженнях CNN успішно використовується для витягування ознак з матриці подібності між контекстом діалогу та можливими відповідями [20, 21].

Таким чином можна зробити висновок, що основні сфери застосування CNN в діалогових системах:

– ієрархічне витягування ознак після основного кодування;
– обробка матриць подібності в системах пошуку відповідей;
– класифікація коротких текстових фрагментів.

4.2.2 Рекурентні нейронні мережі та класичні sequence-to-sequence моделі.

Рекурентні нейронні мережі (RNN) – це нейронні мережі, що спеціально розроблені для обробки послідовних даних, де кожен стан залежить від попереднього. На відміну від CNN, RNN може обробляти послідовності змінної довжини, що особливо важливо для діалогових систем [22].

RNN представлена двома найбільш відомими архітектурами: мережами Джордана та Елмана, які відрізняються між собою механізмом обробки послідовної інформації. Так, в архітектурі Джордана прихований стан формується на основі поточного входу та попереднього виходу мережі, що дозволяє системі враховувати безпосередній контекст попередньої відповіді (див. рисунок 4.3).

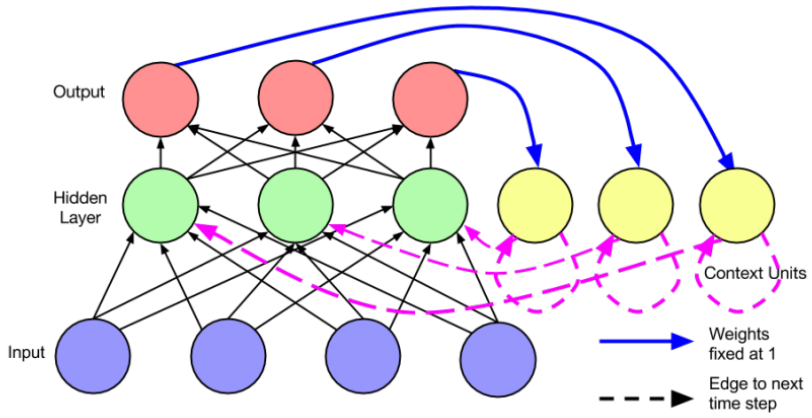

Рисунок 4.3 – Рекурентна нейронна мережа Джордана [22]

В цій архітектурі кожне оновлення прихованого стану визначається поточним входом і виходом останнього кроку часу, тоді як кожен вихід визначається поточним прихованим станом. Таким чином, прихований стан і результат кроку часу t можна обчислюється за формулами 4.2 та 4.3.

$$h_t = \sigma_h(W_h x_t + U_h y_{t-1} + b_n) \tag{4.2}$$

де h_t – оновлення прихованого стану кроку часу t,

σ_h – функція активації,

W_h – матриця ваг для входу,

x_t – поточний вхід,

U_h – матриця ваг для попереднього виходу,

y_{t-1} – попередній вихід мережі

b_n – вектор зміщення.

$$y_t = \sigma_y(W_y h_t + b_y) \tag{4.3}$$

де y_t – вихід в момент часу t,

σ_y – функція активації,

W_y – вагова матриця,

h_t – оновлення прихованого стану кроку часу t,

b_y – вектор зміщення.

Архітектура Елмана, натомість, використовує рекурентний зв'язок від прихованого стану, що забезпечує більш комплексне збереження контекстної інформації протягом всієї послідовності. Це особливо ефективно для розуміння довгострокових залежностей у діалозі (див. рисунок 4.4).

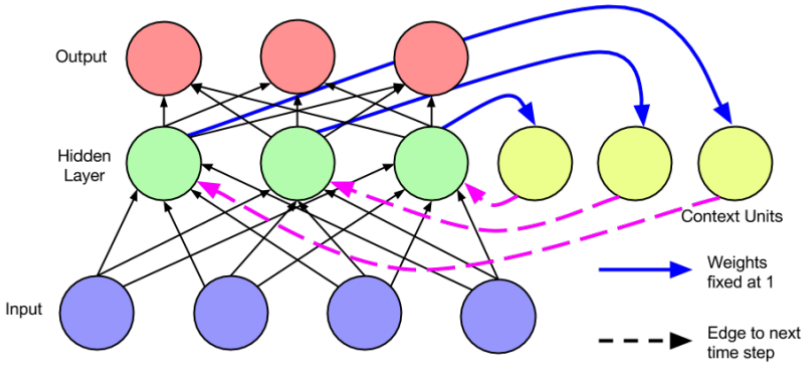

Рисунок 4.4 – Рекурентна нейронна мережа Елмана [22]

В цій архітектурі прихований стан і результат кроку часу t можна обчислюється за формулами 4.4 та 4.5.

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_n) \tag{4.4}$$

де h_t – прихований стан у момент часу t,

h_{t-1} – попередній прихований стан,

σ_h – функція активації,

W_h та U_h – вагові матриці,

x_t – поточний вхід,

b_n – зміщення,

$$y_t = \sigma_y(W_y h_t + b_y) \tag{4.5}$$

де y_t – вихід в момент часу t,

σ_y – функція активації,

W_y – вагова матриця,

h_t – прихований стан у момент часу t,

b_y – зміщення.

Обидві архітектури заклали фундамент для розвитку сучасних діалогових систем, хоча згодом були витіснені більш досконалими варіантами RNN – LSTM та GRU, які краще справляються з проблемою зникаючого градієнта.

Разом з цим Sequence-to-sequence моделі на базі RNN стали проривом у діалогових системах, дозволяючи перетворювати одну послідовність в іншу через механізм кодувальника-декодувальника [23].

Ці моделі широко використовуються для:

- кодування контексту діалогу;
- генерації відповідей;
- обробки знань та доменних тегів [24].

Основним недоліком RNN залишається складність обробки довгих послідовностей через проблеми з градієнтами, хоча LSTM та GRU частково вирішують це питання.

Long Short-Term Memory (LSTM) – це вдосконалена версія RNN, яка вирішує ключову проблему звичайних RNN, а саме зникаючого градієнту при обробці довгих послідовностей.

Архітектура LSTM використовує систему воріт для контролю інформаційного потоку:

- вхідні ворота контролюють нову інформацію;
- ворота забування регулюють видалення старої інформації;
- вихідні ворота керують оновленням стану [24].

Спершу кандидат прихованого стану комбінує поточний вхід $x^{(t)}$ та попередній стан $h^{(t-1)}$, формула 4.6:

$$\hat{h}^{(t)} = \tanh\left(W^{\hat{h}x}x^{(t)} + W^{\hat{h}h}h^{(t-1)} + b_{\hat{h}}\right) \quad (4.6)$$

де $\hat{h}^{(t)}$ – кандидат нового стану,

$W^{\hat{h}x}$ – ваги для поточного входу,

$x^{(t)}$ – поточний вхід,

$W^{\hat{h}h}$ – ваги для попереднього стану,

$h^{(t-1)}$ – попередній стан,

$b_{\hat{h}}$ – зміщення,

tanh – функція активації, що стискає значення до $[-1,1]$.

Вхідні ворота представлені формулою 4.7:

$$i^{(t)} = \sigma(W^{ix}x^{(t)} + W^{ih}h^{(t-1)} + b_i) \quad (4.7)$$

де $i^{(t)}$ – значення вхідних воріт,

W^{ix} – ваги для входу,

$x^{(t)}$ – поточний вхід,

W^{ih} – ваги для попереднього стану,

$h^{(t-1)}$ – попередній стан,

b_i – зміщення,

σ – сигмоїдна функція, дає значення $[0,1]$, як фільтр.

Формула воріт забування (див. формулу 4.8):

$$f^{(t)} = \sigma(W^{fx}x^{(t)} + W^{fh}h^{(t-1)} + b_f) \quad (4.8)$$

де $f^{(t)}$ – значення воріт забування,

W^{fx} – ваги для входу,

$x^{(t)}$ – поточний вхід,

W^{fh} – ваги для попереднього стану,

$h^{(t-1)}$ – попередній стан,

b_f – зміщення,

σ – сигмоїдна функція.

Вихідні ворота контролюють, яку інформацію виводити, формула 4.9:

$$o^{(t)} = \sigma(W^{ox}x^{(t)} + W^{oh}h^{(t-1)} + b_o) \quad (4.9)$$

де $o^{(t)}$ – значення вихідних воріт,

W^{ox} – ваги для входу,

$x^{(t)}$ – поточний вхід,

W^{oh} – ваги для попереднього стану,

$h^{(t-1)}$ – попередній стан,

b_o – зміщення,

σ – сигмоїдна функція.

Стан пам’яті, який балансує нову та стару інформацію, формула 4.10:

$$s^{(t)} = \hat{h}^{(t)} \odot i^{(t)} + s^{(t-1)} \odot f^{(t)} \quad (4.10)$$

де $s^{(t)}$ – новий стан пам’яті,

$s^{(t-1)}$ – попередній стан пам’яті,

$\odot$ – поелементне множення,

$\hat{h}^{(t)} \odot i^{(t)}$ – нова інформація,

$s^{(t-1)} \odot f^{(t)}$ – фільтрована стара інформація.

Вихідний прихований стан – фінальний вихід, що фільтрується вихідними воротами, представляється формулою 4.11:

$$h_t = \tanh(s^{(t)}) \odot o^{(t)} \quad (4.11)$$

де h_t – новий вихідний стан,

$\tanh(s^{(t)})$ – нормалізований стан пам’яті,

$o^{(t)}$ – фільтр вихідних воріт,

$\odot$ – поелементний добуток.

Головна інновація LSTM полягає в поєднанні довгострокової та короткострокової пам'яті, що дозволяє зберігати важливу інформацію протягом

багатьох кроків діалогу. Це особливо важливо для підтримки контексту в тривалих розмовах.

Пізніше була представлена спрощена версія – GRU (Gated Recurrent Unit), яка об'єднує вхідні ворота та ворота забування, зменшуючи кількість параметрів [24].

Архітектура використовує два типи воріт. Перші – ворота оновлення, які контролюють баланс між новою та старою інформацією (див. формулу 4.12):

$$z^{(t)} = \sigma(W^z x^{(t)} + U^z h^{(t-1)} + b_z) \tag{4.12}$$

де $z^{(t)}$ – вектор воріт оновлення на кроці t,

W^z, U^z – матриці ваг,

$x^{(t)}$ – поточний вхід,

$h^{(t-1)}$ – попередній прихований стан,

b_z – вектор зміщення,

σ – сигмоїдна функція активації, яка повертає значення між 0 та 1.

Другі – ворота скидання. Вони визначають, яку частину попереднього стану потрібно зберегти (див. формулу 4.13):

$$r^{(t)} = \sigma(W^r x^{(t)} + U^r h^{(t-1)} + b_r) \tag{4.13}$$

де $r^{(t)}$ – вектор воріт оновлення на кроці t,

W^r, U^r – матриці ваг,

$x^{(t)}$ – поточний вхід,

$h^{(t-1)}$ – попередній прихований стан,

b_r – вектор зміщення,

σ – сигмоїдна функція активації, яка повертає значення між 0 та 1.

Процес оновлення стану в архітектурі GRU починається з обчислення кандидата нового стану (див. формулу 4.14):

$$\hat{h}^{(t)} = \tanh\left(W^h x^{(t)} + U^h (r^{(t)} \odot h^{(t-1)}) + b_h\right) \quad (4.14)$$

де W^h, U^h – матриці ваг,

$x^{(t)}$ – поточний вхід,

$h^{(t-1)}$ – попередній прихований стан,

$r^{(t)} \odot h^{(t-1)}$ – поелементне множення, яке контролює, яку частину попереднього стану використовувати;

b_n – вектор зміщення;

$tanh$ – стикає значення до діапазону $[-1, 1]$.

Наостанок, фінальний стан формується як (див. формулу 4.15):

$$h^{(t)} = \left(1 - z^{(t)}\right) \odot h^{(t-1)} + z^{(t)} \odot \hat{h}^{(t)} \quad (4.15)$$

де $\left(1 - z^{(t)}\right) \odot h^{(t-1)}$ – частина старого стану,

$z^{(t)} \odot \hat{h}^{(t)}$ – частина нового кандидата.

GRU часто показує кращі результати на менших наборах даних, хоча LSTM залишається більш надійним для складних завдань машинного перекладу [25].

4.2.3 Ієрархічний рекурентний кодер-декодер.

Ієрархічний рекурентний кодер-декодер (HRED) – це контекстно-залежна sequence-to-sequence модель, яка була спочатку розроблена для онлайн пошукових систем, а потім адаптована для діалогових систем [26].

Архітектура HRED має три ключові компоненти: енкодер на рівні токенів, контекстний RNN на рівні реплік, декодер для генерації відповідей (див. рисунок 4.5).

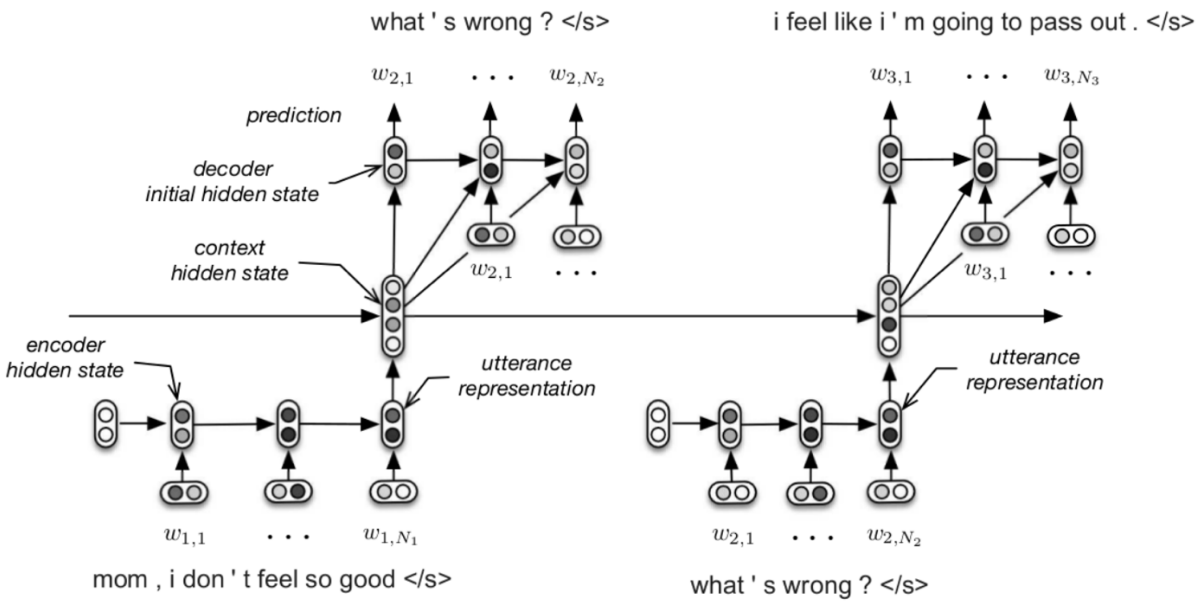

Рисунок 4.5 – Архітектура HRED [26]

Архітектура демонструє обробку діалогу з трьох реплік: "mom, i don't feel so good", "what's wrong?", "i feel like i'm going to pass out".

Схематично зображено:

- нижній рівень: токени (w_1, 1...w_1, N_1) кодуються через encoder hidden state;
- середній рівень: утворення utterance representation для кожної репліки;
- верхній рівень: context hidden state об'єднує інформацію з усіх реплік;
- декодер, який використовує контекстний стан для генерації наступної репліки.

Головна інновація HRED полягає в ієрархічному підході до обробки діалогу. Модель спочатку кодує окремі репліки на рівні слів, потім обробляє послідовність реплік на вищому рівні, створюючи багаторівневе представлення діалогу. Це дозволяє краще зберігати контекст тривалої розмови.

Розширенням HRED стала Latent Variable HRED (VHRED), яка додає латентні змінні для моделювання складних залежностей між послідовностями та покращення різноманітності відповідей [27].

Сучасні системи використовують HRED як основу для більш складних архітектур, додаючи механізми уваги або комбінуючи з іншими підходами для покращення якості та зв'язності діалогу.

4.2.4 Нейронні мережі пам’яті.

Нейронні мережі пам’яті (Memory Networks) представляють важливий крок у розвитку діалогових систем, додаючи явний компонент пам'яті для зберігання та обробки інформації. Ця архітектура дозволяє системі ефективно працювати з довгостроковою інформацією та зовнішніми знаннями [28].

Ключовою особливістю є здатність мережі зберігати факти в окремому модулі пам'яті та звертатися до них при необхідності. Процес обробки інформації включає три основні етапи. Спочатку система оцінює релевантність кожного факту в пам'яті відносно поточного запиту, використовуючи механізм зважування. Потім відбувається вибір інформації з пам’яті через зважене сумування, де кожен факт враховується пропорційно до своєї релевантності.

Фінальний етап включає генерацію відповіді на основі обраної інформації та початкового запиту. Такий механізм дозволяє системі здійснювати "м’який" вибір інформації, що суттєво полегшує процес навчання.

Дані етапи представлено формулами 4.16 – 4.18.

Обчислення ваг:

$$p_i = Softmax(u^T m_i) \tag{4.16}$$

де p_i – вага для і-го елемента пам’яті,

u – вектор запиту,

m_i – вектор і-го елемента пам’яті,

$u^T m_i$ – скалярний добуток для оцінки релевантності,

$Softmax$ – перетворення оцінки в розподіл імовірностей.

Вибір пам’яті:

$$o = \sum_i p_i c_i \tag{4.17}$$

де o – вихідний вектор пам’яті,

p_i – обчислені ваги,

c_i – значення і-го елемента пам’яті.

Фінальне передбачення (генерація):

$$\hat{\alpha} = Softmax(W(o + u)) \tag{4.18}$$

де $\hat{\alpha}$ – вихідний розподіл імовірностей,

o – вектор пам’яті,

u – початковий запит,

$Softmax$ – нормалізація до розподілу імовірностей.

У практичному застосуванні Memory Networks показали високу ефективність в задачах, що потребують інтеграції зовнішніх знань та багатокрокових міркувань [29]. Особливо важливою є можливість відстеження стану діалогу та контролю контексту розмови.

Основним обмеженням залишається складність одночасного навчання всіх модулів системи та потреба в значних обчислювальних ресурсах, особливо при роботі з великими базами знань.

4.2.5 Механізм уваги та трансформер модель.

Механізм уваги революціонізував обробку послідовностей, вирішивши фундаментальну проблему sequence-to-sequence моделей, яка полягла в обмеженні фіксованим контекстним вектором. Було запропоновано підхід "вирівнювання та перекладу", де модель динамічно фокусується на різних частинах вхідної послідовності [30].

При генерації кожного слова модель обчислює ваги уваги для всіх станів енкодера. Процес можна представити як пошук у м'якій формі – замість вибору

одного конкретного стану, модель створює зважену суму всіх станів (див. рисунок 4.6).

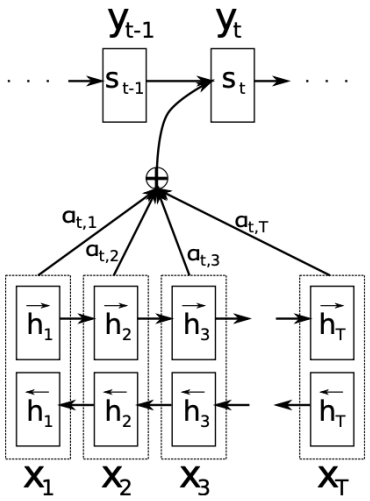

Рисунок 4.6 – Архітектура механізму уваги [30]

Рисунок 4.6 показує, як розподіл ймовірностей наступного токена залежить від попереднього токена, поточного стану декодера та контекстного вектора.

Нейронна мережа Transformer розвинула цю ідею, представивши механізм самоуваги (self-attention). В цьому механізмі кожен токен послідовності проектується у три різні простори: запитів (Q), ключів (K) та значень (V). Формула 4.19 описує, як обчислюється увага між всіма позиціями послідовності одночасно.

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4.19)$$

де Q – матриця запитів (query),

K – матриця ключів (key),

V – матриця значень (value),

d_k – розмірність запитів, або ключів,

QK^T – матричний добуток запитів та транспонованих ключів,

$\sqrt{d_k}$ – нормалізуючий фактор для стабільності градієнтів,

$Softmax$ – нормалізація до розподілу імовірностей.

Одночасно з цим розроблено і Multi-head attention, який додає ще один рівень абстракції – паралельне обчислення уваги в різних підпросторах представлення. Це дозволяє моделі одночасно враховувати різні типи взаємозв'язків між токенами. Формула 4.20 показує, як результати різних «голів» уваги проектуються в кінцевий простір.

$$MultiHead(Q,K,V) = Concat(head_1, \dots, head_h)W^o \quad (4.20)$$

де $head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)$ – окремі «голови» уваги,

h – кількість «голів»,

$Concat$ – конкатенація усіх «голів»,

W^o – проекційна матриця,

W_i^Q, W_i^K, W_i^V – матриці проекції для кожної «голови».

Позиційне кодування вирішує проблему втрати інформації про порядок слів при паралельній обробці. Використовуючи синусоїдальні функції різної частоти, модель отримує унікальне позиційне представлення для кожного токена.

Важливим аспектом архітектури є використання резидуальних з'єднань та нормалізації шарів, що забезпечує стабільне навчання глибоких моделей (див. рисунок 4.7).

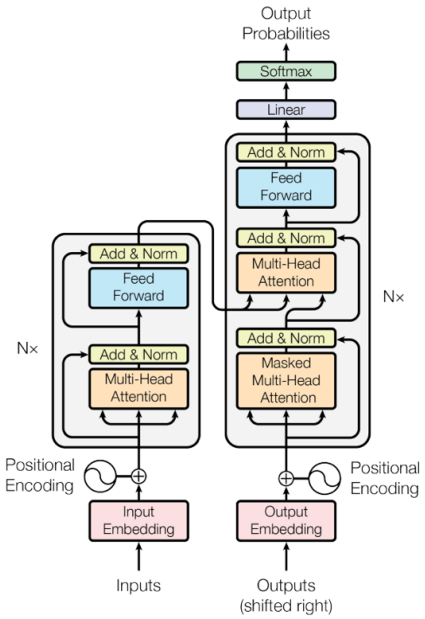

Рисунок 4.7 – Архітектура мережі трансформер [31]

Як видно з рисунку 4.7 архітектура Transformer складається з двох основних компонентів: енкодера та декодера, кожен з яких повторюється N разів. Вхідна послідовність спочатку проходить через позиційне кодування, яке додає інформацію про позиції токенів в послідовності.

В енкодері Multi-Head Attention обробляються вхідні дані паралельно в різних підпросторах представлення, що дозволяє моделі вивчати різні типи залежностей одночасно. Після уваги йде нормалізація (Add & Norm) та повнозв'язний шар (Feed Forward).

Декодер, в свою чергу, починається з маскованої Multi-Head Attention, яка запобігає доступу до майбутніх токенів під час тренування. Другий блок уваги пов’язує вихід декодера з енкодером. Після цього знову використовується нормалізація та повнозв'язний шар.

Residual connections (обхідні з'єднання), які присутні навколо кожного підблоку, допомгають в процесі навчання глибокої мережі. Фінальний Linear шар проектує представлення в простір словника, а Softmax перетворює їх у ймовірності для генерації наступного токена.

Це, разом з можливістю паралельних обчислень, зробило Transformer основою для розвитку потужних мовних моделей як BERT та GPT.

В контексті діалогових систем Transformer показує виняткову ефективність. Дослідження демонструють його переваги в задачі пошуку відповідей, а також успішне використання інкрементальних Transformer-енкодерів для обробки багатокрокових діалогів з інтеграцією документів [32].

4.2.6 Pointer Net та CopyNet.

Pointer Net та CopyNet представляють важливу віху розвитку у генерації тексту, особливо для діалогових систем, де потребується точно відтворювати інформацію з вхідного контексту.

Pointer Net модифікує стандартну sequence-to-sequence архітектуру, змінюючи спосіб генерації вихідної послідовності. Замість використання фіксованого словника, модель працює з динамічним словником, розмір якого

дорівнює довжині вхідної послідовності. Як показано на рисунку, це дозволяє напряму "вказувати" на елементи входу (див. рисунок 4.8).

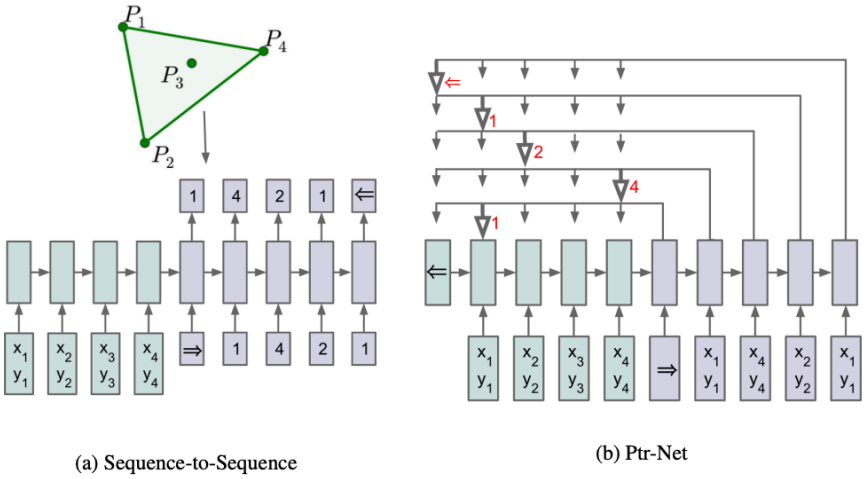

Рисунок 4.8 – Архітектура Pointer Net [33]

CopyNet розширює цю концепцію, додаючи можливість генерації нових слів. Архітектура включає два режими роботи: режим генерації, який створює нові слова зі словника та режим копіювання, що копіює слова з вхідної послідовності (див. рисунок 4.9).

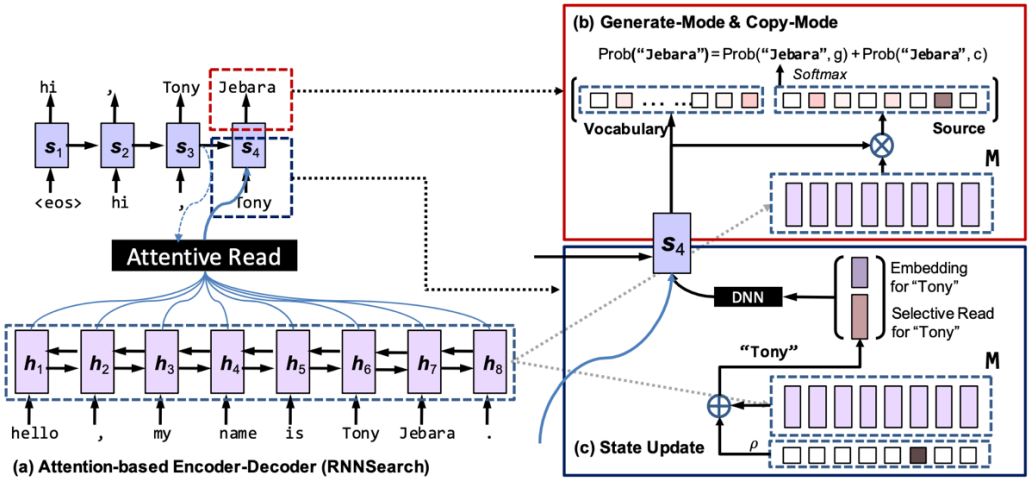

Рисунок 4.9 – Архітектура CopyNet [34]

Модель використовує складний механізм кодування, де кожне слово представляється як комбінація його вбудовування та позиційно-специфічного прихованого стану. Це дозволяє системі бути більш чутливою до контексту при виборі між генерацією та копіюванням.

Практичне застосування в діалогових системах:

- task-oriented системи використовують копіювання для точного відтворення значень слотів [35];
- системи з базами знань застосовують комбінацію копіювання та генерації для створення інформативних відповідей [36];
- системи з мультидоменною підтримкою використовують копіювання для перенесення значень між доменами [37].

Головна перевага обох моделей – здатність працювати зі словами поза словником, та точно відтворювати важливу інформацію. CopyNet демонструє особливу ефективність у задачах, де потрібно балансувати між копіюванням специфічної інформації та генерацією природної мови.

Недоліки включають потребу в механізмах контролю для запобігання надмірному копіюванню та складність балансування між режимами роботи. Проте, як показують дослідження, ці проблеми можна вирішити через додаткові механізми контролю та навчання [38].

4.2.7 Глибокі моделі навчання з підкріпленням і генеративні змагальні мережі.

Глибокі моделі навчання з підкріпленням (Deep Reinforcement Learning) в діалогових системах представляють природний спосіб навчання через взаємодію. Як показано на рисунку, агент (діалогова система) взаємодіє з середовищем (користувачем), отримуючи винагороду за свої дії (див. рисунок 4.10).

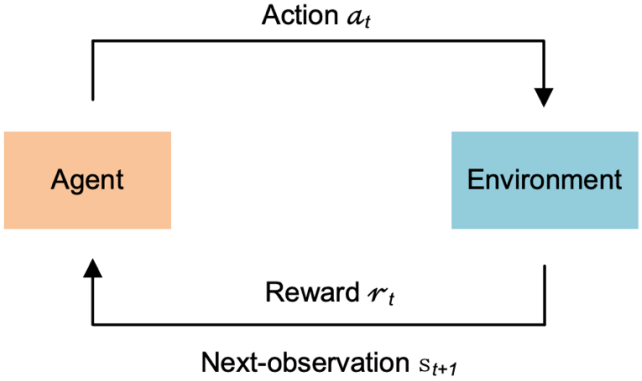

Рисунок 4.10 – Структура навчання з підкріпленням [24]

Математично це описується як процес Маркова (див. формулу 4.21).

$$M = \langle S, A, R, P, \gamma \rangle \tag{4.21}$$

де S – стани середовища,

A – можливі дії,

P – матриця переходів,

R – функція винагороди,

γ – коефіцієнт дисконтування.

Разом з цим існують і Deep Q-Networks (DQN), які використовують глибокі нейронні мережі для апроксимації Q-функції, що оцінює якість дій у кожному стані. Ключове рівняння представлено формулою 4.22 [24].

$$\pi^*(s) = \arg max Q^*(s, a) \tag{4.22}$$

де $\pi^*(s)$ – оптимальна політика для стану s,

$Q^*(s, a)$ – оптимальна Q-функція,

s – поточний стан,

a – можлива дія,

функція Q моделюється за допомогою глибокої нейронної мережі, такої як CNN, RNN тощо.

Для стабільного навчання DQN використовує два важливі механізми:

- experience replay для зберігання та повторного використання попереднього досвіду;
- target network як окрему мережу для обчислення цільових значень.

При цьому Reinforce – алгоритм RL на основі політики, який не має мережі значень, навпаки, безпосередньо оптимізує політику через градієнтний підйом. Його цільова функція представлена формулою 4.23.

$$J(\theta) = E\left[\sum_{t=1}^{H} \gamma^{t-1} r_t | a_t \sim \pi(s_t; \theta)\right] \quad (4.23)$$

де $J(\theta)$ – цільова функція, очікувана сума винагород, яка оцінює якість політики з параметрами θ,

$E[\ ]$ – математичне очікування,

$\sum_{t=1}^{H}$ – сума від t=1 до H, де H – довжина траєкторії діалогу,

γ^{t-1}– коефіцієнт дисконтування в степені (t-1), який зменшує вплив віддалених винагород,

r_t – винагорода на кроці t,

$a_t \sim \pi(s_t; \theta)$ – дія a_t вибирається згідно політики π з параметрами θ у стані s_t.

І, наостанок, GAN (Generative Adversarial Networks) – це архітектура, що складається з двох нейронних мереж: генератора та дискримінатора, які змагаються між собою [39]. В цій архітектурі генератор створює дані, а дискримінатор намагається відрізнити згенеровані дані від реальних (див. рисунок 4.11).

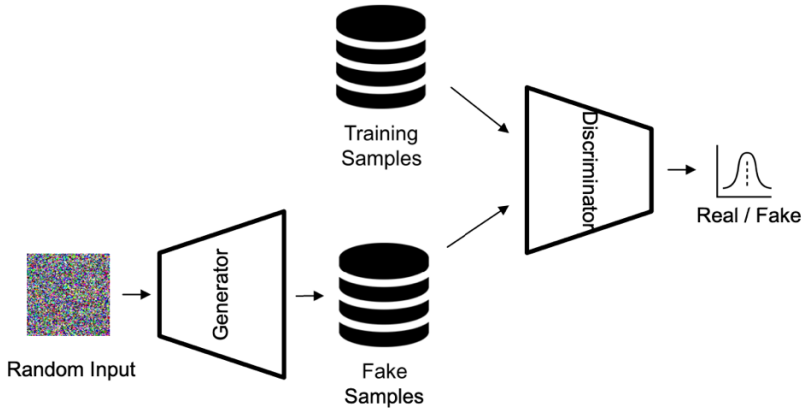

Рисунок 4.11 – Архітектура GAN [24]

Особливо ефективним виявилося поєднання претренованих моделей з RL-файнтюнінгом, де початкове навчання з учителем доповнюється RL для покращення якості відповідей [38].

GAN також показав ефективність у:

– покращенні різноманітності відповідей;

- створенні більш природних діалогів;
- оцінці якості згенерованих відповідей.

Основні виклики включають: проблему розрідженої винагороди в RL, нестабільність навчання GAN та складність дизайну функцій винагороди.

4.3 Формалізація емпатичної моделі.

При розробці емпатичної діалогової моделі також слід зазначити специфіку наявних мовних ресурсів. На сьогоднішній день англомовний сегмент даних для навчання таких систем значно переважає за обсягом та якістю аналогічні ресурси українською мовою. Це зумовлено історично більшою розвиненістю англомовного ринку штучного інтелекту та наявністю великих датасетів діалогів.

На жаль, відсутність репрезентативних діалогових українськомовних даних створює певні обмеження для розробки моделей, здатних природно спілкуватися українською мовою. Зокрема, це стосується розуміння контексту, культурних особливостей та емоційних нюансів, притаманних українському мовленню. Тому при створенні емпатичних діалогових систем часто доводиться спиратися на англомовні дані та адаптувати існуючі рішення.

Втім, ця ситуація поступово змінюється завдяки зростанню інтересу до розвитку українськомовних досліджень та появі нових проєктів. Зокрема, є певні дослідження, які вирішують задачі бінарної класифікації тексту українською мовою за допомогою претренованої багатомовної BERT-моделі [40], або дослідження з моделювання нейронної мережі на основі алгоритму зворотного поширення для розпізнавання емоційної складової в тексті [41].

Спираючись на проведений огляд емпатичних моделей, що представлено в розділі 2, складено пропозицію архітектури власної емпатичної моделі для текстових діалогових систем (див. рисунок 4.12).

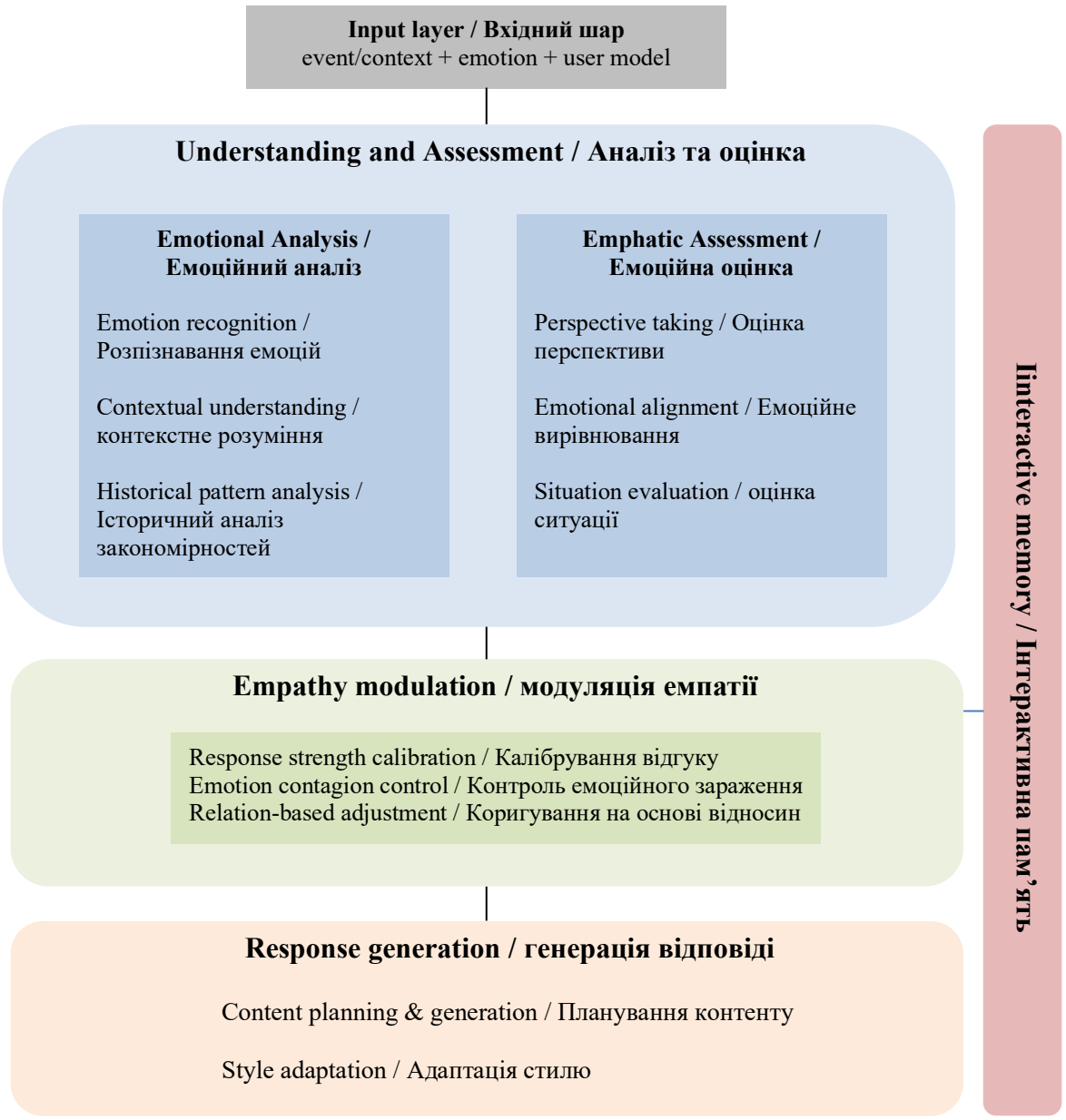

Рисунок 4.12 – Пропозиція архітектури емпатичної моделі (виконано самостійно)

Запропонована архітектура емпатичної моделі складається з чотирьох основних блоків.

Запропонована архітектура емпатичної моделі представляє собою єдину нейронну мережу з інтегрованим зовнішнім модулем пам'яті. В основі архітектури лежить нейромережеве ядро, яке складається з двох основних функціональних блоків: розуміння та генерації відповідей. Така архітектура забезпечує повний цикл обробки діалогу від сприйняття вхідної інформації до генерації емпатичної відповіді.

Блок аналізу та оцінки реалізує функції аналізу емоцій та емпатичної оцінки. Для ефективної роботи цього блоку нейронна мережа повинна мати розвинену здатність до обробки послідовних даних, що потрібно для глибокого аналізу контексту діалогу. Також архітектура мережі має забезпечувати достатню глибину для багаторівневого аналізу емоційних станів та їх нюансів. Не менш важливим в цьому блоці є висока придатність для задач діалогу та гнучкість до масштабування, що дозволить системі адаптуватися до різноманітних ситуацій спілкування.

В свою чергу блок генерації відповідей об'єднує функції модуляції емпатії та безпосередньої генерації тексту. Цей компонент потребує розвиненого механізму уваги для фокусування на найбільш релевантному контексті спілкування. Архітектура на цьому етапі повинна забезпечувати високу гнучкість для інтеграції зовнішніх знань, що критично важливо для точної модуляції емпатичної відповіді. При цьому система має демонструвати стійкість до втрати градієнта для забезпечення стабільного навчання, а складність параметрів повинна залишатися в межах, що дозволяють ефективне налаштування.

Особливу роль в пропозиції архітектури емпатичної моделі відіграє зовнішній модуль пам'яті, який інтегрується з основним нейромережевим ядром. Цей компонент забезпечує зберігання та доступ до історії взаємодій, що критично важливо для підтримання послідовного та контекстно-залежного діалогу. Модуль пам'яті повинен ефективно обробляти послідовні дані та зберігати довготривалі залежності, демонструючи при цьому стійкість до втрати градієнта для забезпечення стабільної роботи протягом тривалих діалогів.

Вся архітектура працює як єдина система, де нейронна мережа забезпечує наскрізну обробку від розуміння вхідного контексту до генерації відповіді, а модуль пам'яті підтримує довготривалий контекст взаємодії. Така інтеграція дозволяє системі демонструвати природну та контекстно-доречну емпатичну поведінку у діалозі.

Таким чином для реалізації описаної архітектури необхідно обрати таку нейромережеву модель, яка зможе ефективно забезпечити всі зазначені вимоги, демонструючи високу продуктивність під час обробки послідовних даних,

підтримки механізмів уваги, інтеграції зовнішніх знань тощо. Вибір конкретної архітектури нейронної мережі має базуватися на ретельному аналізі перерахованих вище вимог. Втім, можна допустити, що не всі перераховані вимоги можуть бути задоволені існуючими нейронними мережами.

Враховуючи різноманітність можливих архітектур та складність вибору оптимального варіанту, доцільно застосувати формальний метод багатокритеріального аналізу. Це дозволить об'єктивно оцінити кожну альтернативу за набором важливих критеріїв.

Враховуючи порівняльний аналіз нейронних мереж, який було проведено в підрозділі 4.2, та на основі наукових джерел [18-39], сформовано порівняльну таблицю критеріїв нейронних мереж (див. табл. 4.2).

Таблиця 4.2 – Опис критеріїв нейронних мереж (таблиця виконана самостійно)

Вид мережі	**Обробка послідовних даних**	**Використання уваги**	**Потреба в обчислювальних ресурсах**	**Глибина мережі**	**Гнучкість до масштабування**	**Тип навчання**	**Придатність для задач діалогу**	**Гнучкість для інтеграції зовнішніх знань**	**Стійкість до втрати градієнта**	**Складність параметрів**
1	2	3	4	5	6	7	8	9	10	11
CNN	Н	Ні	Н	С	С	Супервізоване	С	Н	Н	Н
RNN	В	Ні	С	В	Н	Супервізоване	В	Н	Н	В
LSTM	В	Ні	В	В	С	Супервізоване	В	Н	В	В
GRU	В	Ні	С	В	С	Супервізоване	В	Н	В	С

Кінець таблиці 4.2.

1	2	3	4	5	6	7	8	9	10	11
Transformer	В	Так	В	В	В	Супервізоване	В	С	С	В
Memory Networks	С	Так	В	С	С	Супервізоване	В	В	Н	С
GAN	Н	Ні	В	В	Н	Ненаглядне	С	Н	Н	В

Слід зазначити, що в таблиці 4.2 рівень деяких критеріїв позначено умовно:

- високий – В;
- середній – С;
- низький – Н.

Для критеріїв, що відображено в таблиці надано пояснення нижче, яке спирається на конкретні характеристики, описане в науковій літературі [18-39].

Критерій обробки послідовних даних:

- мережі CNN мають низьку оцінку, оскільки вони оптимізовані для обробки даних із фіксованою структурою, CNN застосовуються для текстових задач, але їхня ефективність знижується у випадках з довгими послідовностями;
- мережі RNN, LSTM, GRU отримали високий показник, оскільки ці мережі спеціально призначені для обробки послідовних даних, враховуючи часові залежності, що є критичним для діалогових систем;
- Transformer має високий показник, оскільки в своїй архітектурі обробляє послідовності паралельно;
- Memory Networks отримали середній показник, бо використовують зовнішню пам'ять для зберігання та обробки інформації, що дозволяє їм працювати з послідовностями, але їх продуктивність залежить від контексту задачі.

Критерій використання уваги:

- моделі CNN, RNN, LSTM, GRU не використовують механізм уваги в базовій архітектурі;
- Transformer використовує механізм самоуваги, що дозволяє ефективно враховувати взаємодію між різними частинами вхідних даних;
- Memory Networks: використовують механізми уваги для пошуку релевантних фактів у зовнішній пам'яті.

Оцінка потреби в обчислювальних ресурсах:

- CNN моделі, завдяки своїй простоті, потребують менше ресурсів порівняно з більш складними моделями;
- рекурентний характер обчислень RNN моделей підвищує потребу в ресурсах порівняно з CNN, тому середня оцінка;
- моделі LSTM, GRU використовують механізми для контролю пам'яті, збільшуючи потребу в ресурсах, тому оцінка висока;
- Transformer модель має паралельну обробку і механізми уваги, що збільшують обчислювальну складність, тому оцінка висока;
- Memory Networks має високу оцінку, бо характерне використання зовнішньої пам'яті і механізмів уваги, що вимагає значних ресурсів.

Критерій глибини мережі:

- для CNN, RNN – середня оцінка, бо зазвичай моделі складаються з декількох шарів;
- для LSTM, GRU – висока, бо моделі дозволяють враховувати довготривалі залежності в даних;
- для Transformer – висока, оскільки архітектура мережі може бути дуже глибокою.

Гнучкість до масштабування:

- низька оцінка для RNN, LSTM, GRU, оскільки послідовна природа обробки ускладнює масштабування;

- середня оцінка для CNN, бо мережі можуть маштабуватися для обробки більших зображень або наборів даних, але з обмеженнями для послідовностей;
- висока оцінка для Transformer через наявну паралельну обробку.

Критерій тип навчання:

- CNN, RNN, LSTM, GRU, Transformer, Memory Networks: мають оцінку супервізоване, адже ці моделі навчаються на основі міток даних;
- GAN має оцінку ненаглядне, GAN можуть навчатися без міток, використовуючи генеративний підхід для створення нових даних.

Придатність для задач діалогу:

- висока оцінка у RNN, LSTM, GRU, Transformer, ці мережі спеціалізовані на обробці тексту і послідовностей;
- середня оцінка у CNN, вони можуть використовуватися для текстових задач, але з обмеженнями для довгих послідовностей.

Оцінюючи гнучкість для інтеграції зовнішніх знань Memory Networks мають високий показник, оскільки саме такий вид мережі призначеней для інтеграції зовнішніх знань.

Високу оцінку критерія стійкості до втрати градієнта мають мережі LSTM, GRU, адже мають спеціальні механізми для запобігання зниканню градієнта.

Якщо враховувати критерій складності параметрів, то мережі Transformer та GAN мають найвищий показник, оскільки велика кількість параметрів обумовлює високу складність і потребу в ресурсах.

Для аналізу вибору моделі в багатокритеріальній задачі, критерії потрібно привести до порівнянних шкал і нормалізувати їх до принципу «за максимумом». Це дозволить узгоджено оцінити моделі за всіма критеріями.

- обробка послідовних даних має категоричну шкалу (низька, середня, висока), де «висока» є бажаним результатом;
- використання уваги має бінарну шкалу (так/ні), де «так» є бажаним результатом;

- потреба в обчислювальних ресурсах – це кількісна шкала (низька, середня, висока), де «низька» є бажаним результатом, але для нормалізації переведемо по принципу «за максимумом» (висока є бажаним результатом);
- глибина мережі виражена категоричною шкалою (низька, середня, висока), де «висока» є бажаним результатом;
- гнучкість до масштабування – категорична шкала (низька, середня, висока), де «висока» є бажаним результатом;
- тип навчання має категоричну шкалу (супервізоване, ненаглядне), де «ненаглядне» є бажаним результатом, тому що воно не вимагає міток;
- придатність для задач діалогу – це категорична шкала (низька, середня, висока), де «висока» є бажаним результатом;
- гнучкість для інтеграції зовнішніх знань має категоричну шкалу (низька, середня, висока), де «висока» є бажаним результатом;
- стійкість до втрати градієнта має категоричну шкалу (низька, середня, висока), де «висока» є бажаним результатом;
- складність параметрів – це кількісна шкала (низька, середня, висока), де «низька» є бажаним результатом, але для нормалізації переведемо по принципу «за максимумом» (висока є бажаним результатом).

Також слід додати, що для критеріїв, де значення шкали «низька» є бажаним результатом, проведемо інверсію шкали, тобто «низька» = 1, «середня» = 0,5, «висока» = 0.

Результат нормалізації критерії представлено в таблиці (див. табл. 4.3).

Таблиця 4.3 – Векторний опис альтернатив після нормалізації (таблиця виконана самостійно)

Вид мережі	Обробка послідовних даних	Використання уваги	Потреба в обчислювальних ресурсах	Глибина мережі	Гнучкість до масштабування	Тип навчання	Придатність для задач діалогу	Гнучкість для інтеграції зовнішніх знань	Стійкість до втрати градієнта	Складність параметрів
CNN	0	0	1	0,5	0,5	0	0,5	0	0	1
RNN	1	0	0,5	1	0	0	1	0	0,5	0
LSTM	1	0	0	1	0,5	0	1	0	1	0
GRU	1	0	0,5	1	0,5	0	1	0	1	0,5
Transformer	1	1	0	1	1	0	1	0,5	0,5	0
Memory Networks	0,5	1	0	0,5	0,5	0	1	1	0,5	0,5
GAN	0	0	0	1	0	1	0,5	0	0	0

Проведемо аналіз Парето-домінування для представлених архітектур нейронних мереж на основі таблиці 4.3. Варто почати з розгляду сутності домінування за Парето, де один варіант вважається кращим за інший, якщо він перевершує його хоча б за одним критерієм та не поступається за всіма іншими.

При детальному розгляді матриці характеристик можна зробити кілька важливих спостережень. Transformer демонструє суттєву перевагу над Memory Networks у таких аспектах як обробка послідовних даних, глибина мережі та гнучкість до масштабування, маючи значення 1 проти 0,5. Проте Memory Networks має кращий показник придатності для інтеграції знань. Це створює ситуацію, де жодна з цих архітектур повністю не домінує над іншою за Парето.

При порівнянні LSTM та RNN, LSTM демонструє вищі показники стійкості до втрати градієнта, але RNN має перевагу в потребі обчислювальних ресурсів. Така ситуація також не дозволяє встановити чітке домінування однієї архітектури над іншою.

GAN виявляється досить специфічною архітектурою, яка має високі показники в глибині мережі та типі навчання, але поступається іншим архітектурам за багатьма іншими критеріями. CNN показує сильні сторони в потребі

обчислювальних ресурсах та складності параметрів, але має обмеження в інших аспектах.

Таким чином загальний висновок аналізу свідчить про відсутність абсолютного домінування будь-якої архітектури над усіма іншими. Кожна архітектура має свої сильні сторони та області застосування.

Для вирішення поставленої задачі обрано лінійну адитивну згортку з ваговими коефіцієнтами. Математична форма моделі (див. формулу 4.24):

$$z^* = max \sum \alpha_j \beta_j a_{ij} \tag{4.12}$$

де z^* – оптимальне значення цільової функції,

α_j – нормуючі множники для кожного j-го критерію (визначені при нормуванні),

β_j – вагові коефіцієнти j-го критерію,

a_{ij} – значення j-го критерію для і-ї альтернативи,

i = 1,7 – оптимальні альтернативи,

j = 1,10 – критерії оцінювання,

$\sum \beta_j = 1$,

$0 \leq \beta_j \leq 1$.

Процес визначення вагових коефіцієнтів для критеріїв оцінки нейронних мереж базується на аналізі вимог розробленої архітектури емпатичної моделі. Основою для визначення значущості кожного критерію слугують функціональні особливості та вимоги кожного компонента системи.

Найвищий пріоритет надається здатності обробляти послідовні дані (β_1 = 0,15) та механізму уваги (β_2 = 0,15). Це обґрунтовується тим, що ефективна обробка послідовної інформації є критично важливою для розуміння контексту діалогу та емоційного стану співрозмовника. Механізм уваги, в свою чергу, забезпечує здатність системи фокусуватися на найбільш релевантних аспектах вхідної інформації, що особливо важливо для точної емпатичної оцінки та генерації відповідей.

Наступний рівень значущості присвоєно придатності для задач діалогу (β_7 = 0,13), оскільки це безпосередньо впливає на якість взаємодії системи з користувачем. Глибина мережі (β_4 = 0,12) також отримала високий ваговий коефіцієнт через необхідність багаторівневого аналізу емоційних станів та контексту. Така ж вага надана гнучкості для інтеграції зовнішніх знань (β_8 = 0,12) та стійкості до втрати градієнта (β_9 = 0,12), що обумовлено необхідністю стабільної роботи з модулем пам'яті та забезпечення надійного навчання системи.

Менш критичним параметрам, таким як потреба в обчислювальних ресурсах (β_3 = 0,07) та гнучкість до масштабування (β_5 = 0,07), надано відповідні вагові коефіцієнти. Це пояснюється тим, що хоча ці характеристики впливають на практичну реалізацію системи, вони не є визначальними для її функціональності на даному етапі теоретичного дослідження.

Найнижчі ваги присвоєно типу навчання (β_6 = 0,03) та складності параметрів (β_{10} = 0,04), оскільки ці характеристики мають найменший вплив на здатність системи забезпечувати емпатичну взаємодію.

Використовуючи визначені вагові коефіцієнти та нормалізовані значення з таблиці, проведено розрахунки за формулою лінійної згортки для кожної альтернативи.

CNN: $z = 0\times0,15 + 0\times0,15 + 1\times0,07 + 0,5\times0,12 + 0,5\times0,07 + 0\times0,03 + 0,5\times0,13 + 0\times0,12 + 0\times0,12 + 1\times0,04 = 0,1815$

RNN: $z = 1\times0,15 + 0\times0,15 + 0,5\times0,07 + 1\times0,12 + 0\times0,07 + 0\times0,03 + 1\times0,13 + 0\times0,12 + 0,5\times0,12 + 0\times0,04 = 0,445$

LSTM: $z = 1\times0,15 + 0\times0,15 + 0\times0,07 + 1\times0,12 + 0,5\times0,07 + 0\times0,03 + 1\times0,13 + 0\times0,12 + 1\times0,12 + 0\times0,04 = 0,4685$

GRU: $z = 1\times0,15 + 0\times0,15 + 0,5\times0,07 + 1\times0,12 + 0,5\times0,07 + 0\times0,03 + 1\times0,13 + 0\times0,12 + 1\times0,12 + 0,5\times0,04 = 0.4885$

Transformer: $z = 1\times0,15 + 1\times0,15 + 0\times0,07 + 1\times0,12 + 1\times0,07 + 0\times0,03 + 1\times0,13 + 0,5\times0,12 + 0,5\times0,12 + 0\times0,04 = 0,67$

Memory Networks: $z = 0,5\times0,15 + 1\times0,15 + 0\times0,07 + 0,5\times0,12 + 0,5\times0,07 + 0\times0,03 + 1\times0,13 + 1\times0,12 + 0,5\times0,12 + 0,5\times0,04 = 0,595$

GAN: z = 0×0,15 + 0×0,15 + 0×0,07 + 1×0,12 + 0×0,07 + 1×0,03 + 0,5×0,13 + 0×0,12 + 0×0,12 + 0×0,04 = 0,185

Таким чином, на основі проведених розрахунків можна зробити висновок, що найбільш підходящою архітектурою для реалізації запропонованої моделі є Transformer з показником 67 % оптимальності, за ним слідує Memory Networks 59,5 % оптимальності та GRU 48,85 % відповідно.

Результати багатокритеріального аналізу показали, що архітектура Transformer демонструє найвищий показник оптимальності мережі, що обирається, серед усіх розглянутих варіантів. Це обумовлено кількома ключовими перевагами даної архітектури.

По-перше, Transformer забезпечує високу ефективність обробки послідовних даних, що критично важливо для аналізу контексту діалогу та емоційних станів.

По-друге, вбудований механізм самоуваги дозволяє моделі ефективно фокусуватися на найбільш релевантних аспектах вхідної інформації, що особливо важливо для емпатичної оцінки ситуації. Крім того, архітектура демонструє відмінну гнучкість до масштабування та високу придатність для задач діалогу.

Другу позицію у відборі оптимальності посіла архітектура Memory Networks, що підкреслює важливість механізмів пам'яті для емпатичної взаємодії.

Memory Networks особливо ефективні в задачах, що потребують довготривалого зберігання контексту та інтеграції зовнішніх знань. Ця архітектура показала найвищі результати за критеріями гнучкості для інтеграції зовнішніх знань та здатності підтримувати довготривалий контекст діалогу.

Комбінація сильних сторін обох архітектур – механізму самоуваги Transformer та можливостей роботи з пам'яттю Memory Networks – створює потужну основу для реалізації емпатичної моделі. Transformer забезпечує ефективну обробку вхідної інформації та генерацію відповідей, тоді як компоненти Memory Networks дозволяють системі зберігати та використовувати важливу контекстну інформацію протягом тривалої взаємодії.

4.4 Планування експериментальної перевірки моделі.

На основі проведеного багатокритеріального аналізу та визначення оптимальної архітектури нейронної мережі можна запропонувати комплексний підхід до експериментальної перевірки розробленої моделі емпатії.

Результати багатокритеріального аналізу показали, що архітектура Transformer демонструє найвищий показник оптимальності вибору мережі (0,67), значно випереджаючи інші розглянуті архітектури. Це обумовлює вибір Transformer як базової архітектури для реалізації емпатичної моделі. Водночас, друге місце посіла архітектура Memory Networks (0,595), що підтверджує важливість механізмів пам'яті для ефективної емпатичної взаємодії. Таким чином, експериментальна перевірка моделі повинна враховувати як переваги архітектури Transformer, так і можливості інтеграції компонентів пам'яті.

План експериментальної перевірки доцільно розділити на кілька послідовних етапів. На першому етапі необхідно провести підготовку даних для навчання та тестування моделі. Враховуючи специфіку задачі емпатичної взаємодії, набір даних повинен включати діалоги з різноманітними емоційними контекстами та ситуаціями. Особливу увагу слід приділити розмітці даних, яка повинна відображати не лише емоційний стан співрозмовників, але й рівень емпатії у відповідях.

Навчання моделі пропонується проводити в кілька фаз. Спочатку має відбутись базове навчання на загальному корпусі діалогів для формування основних мовленнєвих навичок. Після цього має бути проведено спеціалізоване навчання на датасеті емпатичних діалогів з акцентом на розвиток здатності до емоційного розуміння та генерації відповідних реакцій.

Оцінка ефективності моделі повинна базуватися на комплексі метрик, що охоплюють різні аспекти емпатичної взаємодії. Опису та обгрунтуванню ключових слід приділити під час під час підготовки практичного експерименту.

Не менш важливим аспектом експериментальної перевірки є порівняльний аналіз з існуючими рішеннями. Пропонується провести порівняння розробленої моделі з базовими моделями діалогових систем та спеціалізованими емпатичними

моделями. Це дозволить оцінити внесок запропонованої архітектури та підходу до моделювання емпатії.

Заключним етапом експериментальної перевірки є аналіз отриманих результатів та формування рекомендацій щодо можливих покращень моделі. На основі цього аналізу можуть бути визначені напрямки подальшої оптимізації архітектури, вдосконалення процесу навчання та розширення функціональних можливостей запропонованої моделі штучної емпатії.

ВИСНОВКИ

В результаті виконання дослідження було досягнуто поставленої мети – розроблено комплексну архітектуру моделі штучної емпатії для діалогових систем. У процесі дослідження було успішно вирішено всі поставлені завдання та отримано наступні результати.

По-перше, проведено глибокий аналіз предметної галузі, який дозволив визначити ключові компоненти штучної емпатії та їх взаємозв’язок. Встановлено, що штучна емпатія базується на трьох основних складових: перспективному мисленні, емпатичній турботі та емоційному зараженні. При цьому виявлено цікавий парадокс: когнітивні аспекти емпатії, які для людей часто вимагають свідомих зусиль, виявляються простішими для реалізації в штучних системах, ніж афективні компоненти, які у людей проявляються природно та автоматично.

По-друге, здійснено комплексний огляд існуючих моделей штучної емпатії, що дозволило систематизувати та класифікувати різні підходи до моделювання емпатії в системах штучного інтелекту. Виявлено, що більшість сучасних моделей базуються на теоретичному підході, тоді як моделі, засновані на даних, зазвичай фокусуються на специфічних аспектах емпатії. Це спостереження має важливе значення для подальшого розвитку галузі, оскільки вказує на необхідність розробки більш інтегрованих підходів.

По-третє, розроблено оригінальну архітектуру емпатичної моделі, яка складається з чотирьох основних функціональних блоків: аналізу та оцінки, емоційного аналізу, модуляції емпатії та генерації відповіді. Особливістю запропонованої архітектури є інтеграція зовнішнього модуля пам'яті, що дозволяє системі підтримувати довготривалий контекст взаємодії та забезпечує більш природну емпатичну комунікацію.

По-четверте, проведено багатокритеріальний аналіз для вибору оптимальної нейромережевої архітектури, яка найкраще відповідає вимогам розробленої моделі. Аналіз охопив десять ключових критеріїв, включаючи обробку послідовних даних, використання механізмів уваги, потребу в обчислювальних ресурсах та інші важливі характеристики. В результаті виявлено, що архітектура Transformer

демонструє найвищий показник оптимальності (67 %), значно випереджаючи інші розглянуті архітектури. Друге місце посіла архітектура Memory Networks (59,5 %), що підтверджує важливість механізмів пам'яті для ефективної емпатичної взаємодії.

Наостанок, розроблено комплексний план експериментальної перевірки моделі, який включає підготовку даних, багатофазне навчання моделі та оцінку її ефективності за допомогою комплексу метрик. Запропонований підхід до експериментальної верифікації забезпечує можливість об'єктивної оцінки розробленої моделі та її порівняння з існуючими рішеннями.

Практична значимість отриманих результатів полягає в можливості їх використання для створення більш ефективних діалогових систем з розвиненими емпатичними здібностями. Запропонована архітектура може бути адаптована для різних прикладних задач, де важлива емоційна взаємодія між людиною та машиною.

Наукова новизна роботи полягає в розробці інтегрованого підходу до моделювання штучної емпатії, який враховує як когнітивні, так і афективні аспекти емпатичної взаємодії. Запропонована архітектура вирізняється комплексним підходом до обробки емоційної інформації та використанням сучасних досягнень у галузі нейронних мереж.

Подальші перспективи дослідження включають:

- розширення можливостей моделі для роботи з мультимодальними даними;
- вдосконалення механізмів довготривалої пам'яті для підтримки більш складних контекстів взаємодії;
- розробку методів оцінки якості емпатичної взаємодії;
- дослідження можливостей адаптації моделі до різних культурних контекстів;
- вивчення етичних аспектів використання емпатичних систем штучного інтелекту.

Варто зазначити, що хоча розроблена модель демонструє значний потенціал, існують певні обмеження та виклики, які потребують подальшого дослідження. Зокрема, це стосується складності відтворення істинних емоційних переживань у штучних системах та необхідності забезпечення етично відповідального використання емпатичних технологій.

Таким чином, проведене дослідження не лише досягло поставленої мети, але й створило міцний фундамент для подальшого розвитку галузі штучної емпатії. Отримані результати можуть бути використані як теоретична та практична основа для розробки нового покоління емпатичних діалогових систем, здатних забезпечити більш природну та ефективну взаємодію між людиною та машиною.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Ramya Srinivasan, Beatriz San Miguel González, The role of empathy for artificial intelligence accountability, Journal of Responsible Technology, Volume 9, 2022. URL: https://doi.org/10.1016/j.jrt.2021.100021 (дата звернення 02.06.2025).

2. "Empathetic Conversational Systems: A Review of Current Advances, Gaps, and Opportunities." IEEE Transactions on Affective Computing, null (2022).:1-20. URL: https://arxiv.org/abs/2206.05017 (дата звернення 02.06.2025).

3. Jana, Schaich, Borg., Hannah, Read. "What Is Required for Empathic AI? It Depends, and Why That Matters for AI Developers and Users." null (2024). URL: https://typeset.io/papers/what-is-required-for-empathic-ai-it-depends-and-why-that-5acegnef37lh (дата звернення 02.06.2025).

4. Yuping, Liu-Thompkins., Shintaro, Okazaki., Hairong, Li. "Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience." Journal of the Academy of Marketing Science, 50 (2022). URL: https://typeset.io/papers/artificial-empathy-in-marketing-interactions-bridging-the-1us9tbv3 (дата звернення 02.06.2025).

5. Zhongliang, Cui., Jing, Liu. "A Study on Two Conditions for the Realization of Artificial Empathy and Its Cognitive Foundation." Philosophies, 7 (2022). URL: https://typeset.io/papers/a-study-on-two-conditions-for-the-realization-of-artificial-1qoxat37 (дата звернення 02.06.2025).

6. Lee, Yoon Kyung, et al. "Chain of empathy: Enhancing empathetic response of large language models based on psychotherapy models". URL: https://arxiv.org/abs/2311.04915 (дата звернення 02.06.2025).

7. Bart, Bussmann., Jacqueline, Heinerman., Joel, Lehman. "Towards Empathic Deep Q-Learning." arXiv: Learning, null (2019). URL: https://typeset.io/papers/towards-empathic-deep-q-learning-1zyykdd18v (дата звернення 02.06.2025).

8. Özge Nilay Yalcin, Steve DiPaola, "A computational model of empathy for interactive agents". 2018. URL: https://www.scribd.com/document/416270144/A-Computational-Model-of-Empathy-for-Interactive-Agents (дата звернення 02.06.2025).

9. Yalçın, Ö. and Steve DiPaola. “Modeling empathy: building a link between

affective and cognitive processes.” Artificial Intelligence Review 53 (2019): 2983 - 3006. URL: https://www.semanticscholar.org/paper/Modeling-empathy%3A-building-a-link-between-affective-Yal%C3%A7%C4%B1n-DiPaola/66f7c96fc118a6adfe288c2af36af058404148c6 (дата звернення 02.06.2025).

10. Asada, M. Towards Artificial Empathy. Int J of Soc Robotics 7, 19–33 (2015). URL: https://doi.org/10.1007/s12369-014-0253-z (дата звернення 02.06.2025).

11. Rodrigues SH, Mascarenhas S, Dias J, Paiva A (2014) A process model of empathy for virtual agents. Interacting with Computers 27(4):371–391. URL: https://www.researchgate.net/publication/270340921_A_Process_Model_of_Empathy_For_Virtual_Agents (дата звернення 02.06.2025).

12. Boukricha H, Wachsmuth I, Carminati MN, Knoeferle P (2013) A computational model of empathy: Empirical evaluation. In: Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on, IEEE, pp 1–6. URL: https://www.researchgate.net/publication/259632516_A_Computational_Model_of_Empathy_Empirical_Evaluation (дата звернення 02.06.2025).

13. Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2014) Empathic robots for long-term interaction. International Journal of Social Robotics 6(3):329–341. URL: https://www.semanticscholar.org/paper/Empathic-Robots-for-Long-term-Interaction-Leite-Castellano/be497b933bc6434cf611e3103f86eb0b2ddbc87f (дата звернення 02.06.2025).

14. Xiao B, Bone D, Segbroeck MV, Imel ZE, Atkins DC, Georgiou PG, Narayanan SS (2014) Modeling therapist empathy through prosody in drug addiction counseling. In: Fifteenth Annual Conference of the International Speech Communication Association. URL: https://sail.usc.edu/publications/files/xiao2014_modeling-therap.pdf (дата звернення 02.06.2025).

15. Gibson J, Malandrakis N, Romero F, Atkins DC, Narayanan SS (2015) Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms. In: Sixteenth Annual Conference of the International Speech Communication Association. URL: https://www.isca-

archive.org/interspeech_2015/gibson15b_interspeech.html (дата звернення 02.06.2025).

16. Rashkin H, Smith EM, Li M, Boureau YL (2018) I know the feeling: Learning to converse with empathy. URL: https://openreview.net/pdf?id=HyesW2C9YQ (дата звернення 02.06.2025).

17. Kumano S, Otsuka K, Mikami D, Matsuda M, Yamato J (2015) Analyzing interpersonal empathy via collective impressions. IEEE Transactions on Affective Computing 6(4):324–336. URL: https://openreview.net/pdf/8fda8b6f4b559f0959fd121e3b730c0a26ac0174.pdf (дата звернення 02.06.2025).

18. Zhang Y, Wallace B (2017) A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Asian Federation of Natural Language Processing, Taipei, Taiwan, pp 253–263. URL: https://www.aclweb.org/anthology/I17- 1026 (дата звернення 02.06.2025).

19. Conneau A, Schwenk H, Barrault L, Lecun Y (2016) Very deep convolutional networks for text classification. URL: https://arxiv.org/abs/1606.01781 (дата звернення 02.06.2025).

20. Feng J, Tao C, Wu W, Feng Y, Zhao D, Yan R (2019) Learning a matching model with co-teaching for multi-turn response selection in retrieval-based dialogue systems. URL: https://arxiv.org/abs/1906.04413 (дата звернення 02.06.2025).

21. Chongyang Tao, Wei Wu, Can Xu, Wenpeng Hu, Dongyan Zhao, and Rui Yan. 2019. One Time of Interaction May Not Be Enough: Go Deep with an Interaction-over-Interaction Network for Response Selection in Dialogues. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1–11, Florence, Italy. Association for Computational Linguistics. URL: https://aclanthology.org/P19-1001/ (дата звернення 02.06.2025).

22. Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning. URL: https://arxiv.org/abs/1506.00019 (дата звернення 02.06.2025).

23. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. URL: https://arxiv.org/abs/1409.3215 (дата звернення 02.06.2025).

24. Jinjie Ni, Tom Young, Vlad Pandelea, Fuzhao Xue, Erik Cambria (2022) Recent Advances in Deep Learning Based Dialogue Systems: A Systematic Survey. URL: https://arxiv.org/abs/2105.04387 (дата звернення 02.06.2025).

25. Gruber N, Jockisch A (2020) Are gru cells more specific and lstm cells more sensitive in motive classification of text? Frontiers in Artificial Intelligence 3(40):1–6. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7861254/ (дата звернення 02.06.2025).

26. Serban I, Sordoni A, Bengio Y, Courville A, Pineau J (2016) Building end-to-end dialogue systems using generative hierarchical neural network models. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 30, no 1. URL: https://arxiv.org/abs/1507.04808 (дата звернення 02.06.2025).

27. Serban I, Sordoni A, Lowe R, Charlin L, Pineau J, Courville A, Bengio Y (2017) A hierarchical latent variable encoder-decoder model for generating dialogues. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 31, no 1. URL: https://ojs.aaai.org/index.php/AAAI/article/view/10983 (дата звернення 02.06.2025).

28. Weston J, Chopra S, Bordes A (2014) Memory networks. URL: https://arxiv.org/abs/1410.3916 (дата звернення 02.06.2025).

29. Gao Y, Wu CS, Joty S, Xiong C, Socher R, King I, Lyu M, Hoi SC (2020c) Explicit memory tracker with coarse-to-fine reasoning for conversational machine reading. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp 935–945. URL: https://aclanthology.org/2020.acl-main.88/ (дата звернення 02.06.2025).

30. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. URL: https://arxiv.org/abs/1409.0473 (дата звернення 02.06.2025).

31. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Lu, Polosukhin I (2017) Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in Neural Information

Processing Systems, Curran Associates, Inc., vol 30. URL: https://proceedings.neurips.cc/paper/2017/file/ 3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (дата звернення 02.06.2025).

32. Henderson M, Vulic I, Gerz D, Casanueva I, Budzianowski P, Coope S, Spithourakis G, Wen TH, ´ Mrkšic N, Su PH (2019b) Training neural response selection for task-oriented dialogue systems. URL: https://aclanthology.org/P19-1536/ (дата звернення 02.06.2025).

33. Oriol V, Meire F, Navdeep J (2015) Pointer networks. Advances in neural information processing systems 28:2692–2700. URL: https://arxiv.org/abs/1506.03134 (дата звернення 02.06.2025).

34. Gu J, Lu Z, Li H, Li VO (2016) Incorporating copying mechanism in sequence-to-sequence learning. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany, pp 1631–1640. URL: https://www.aclweb.org/anthology/P16-1154 (дата звернення 02.06.2025).

35. Wu CS, Madotto A, Hosseini-Asl E, Xiong C, Socher R, Fung P (2019) Transferable multi-domain state generator for task-oriented dialogue systems. URL: https://arxiv.org/abs/1905.08743 (дата звернення 02.06.2025).

36. Lin X, Jian W, He J, Wang T, Chu W (2020a) Generating informative conversational response using recurrent knowledge-interaction and knowledge-copy. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp 41–52. URL: https://aclanthology.org/2020.acl-main.6/ (дата звернення 02.06.2025).

37. Ouyang Y, Chen M, Dai X, Zhao Y, Huang S, Jiajun C (2020) Dialogue state tracking with explicit slot connection modeling. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp 34–40. URL: https://aclanthology.org/2020.acl-main.5/ (дата звернення 02.06.2025).

38. Wu J, Wang X, Wang WY (2019b) Self-supervised dialogue learning. URL: https://arxiv.org/abs/1907.00448 (дата звернення 02.06.2025).

39. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S,

Courville A, Bengio Y (2014) Generative adversarial networks. URL: https://arxiv.org/abs/1406.2661 (дата звернення 02.06.2025).

40. Рябишев О. В. Аналіз тональності тексту українською мовою / О. В. Рябишев, А. Л. Єрохін, А. Г. Бахмет // Бионика интеллекта : научно-технический журнал. – 2021. – № (96). – С. 15–21. URL: https://openarchive.nure.ua/handle/document/23317 (дата звернення 02.06.2025).

41. Nazarenko D. S., Afanasieva I. V., Golian N. V. Neural network approach for emotional recognition in text. Bionics of intelligence. 2019. Т. 1, № 92. С. 9–13. URL: https://doi.org/10.30837/bi.2019.1(92).02 (дата звернення 02.06.2025)

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ

40. Рябишев О. В. Аналіз тональності тексту українською мовою / О. В. Рябишев, А. Л. Єрохін, А. Г. Бахмет // Бионика интеллекта : научно-технический журнал. – 2021. – № (96). – С. 15–21. URL: https://openarchive.nure.ua/handle/document/23317 (дата звернення 02.06.2025).

41. Nazarenko D. S., Afanasieva I. V., Golian N. V. Neural network approach for emotional recognition in text. Bionics of intelligence. 2019. Т. 1, № 92. С. 9–13. URL: https://doi.org/10.30837/bi.2019.1(92).02 (дата звернення 02.06.2025).