УДК 81'324:004.89
DOI https://doi.org/10.24919/2308-4863/89-1-41

Володимир ПАСІЧНИК,
orcid.org/0000-0002-5231-6395
доктор технічних наук, професор,
професор кафедри інформаційних систем та мереж
Національного університету «Львівська політехніка»
(Львів, Україна) vpasichnyk@gmail.com

Максим ЯРОМИЧ,
orcid.org/0009-0005-3299-6695
аспірант кафедри прикладної лінгвістики
Національного університету «Львівська політехніка»
(Львів, Україна) yamax0312@gmail.com

ПОРІВНЯЛЬНИЙ АНАЛІЗ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ В КОНТЕКСТІ ВИРІШЕННЯ ЛІНГВІСТИЧНИХ ЗАДАЧ

У статті здійснено міждисциплінарне дослідження порівняльної ефективності великих мовних моделей у контексті вирішення лінгвістичних задач. У фокусі – моделі нового покоління, такі як GPT-4o, Claude 3, Gemini, DeepSeek R1, Grok, а також відкриті системи (BLOOM, LLaMA 2), які активно використовуються у лінгвістиці, філології, перекладознавстві та цифрових гуманітарних науках. Робота обґрунтовує необхідність концептуального переосмислення LLM не лише як інструментів генерації тексту, а як "помічників-дослідників", здатних виконувати складні завдання: семантичний аналіз, стилістичну інтерпретацію, тематичне моделювання, автоматизоване узагальнення тексту та міжмовне порівняння.

Методологічною основою аналізу виступає модифікований груповий метод аналізу ієрархій Саати (Group AHP), що дозволяє поєднати судження трьох груп експертів – лінгвістів, гуманітаріїв та ІТ-фахівців – в єдиний вектор пріоритетів. Було сформовано 15 критеріальних параметрів, що охоплюють як лінгвістичні аспекти (розуміння контексту, корекція граматики, стилістичне багатство, багатомовність), так і технічні (апаратні вимоги, швидкість генерації, API-інтеграція). Побудова матриць парних порівнянь здійснювалася окремо для кожної групи експертів, після чого результати були агреговані шляхом геометричного середнього. В результаті отримано узгоджений вектор ваг критеріїв та ранжування моделей за їхньою ефективністю в лінгвістичних задачах.

У результаті порівняння за груповим методом AHP найвищі глобальні ваги отримали Claude 3, GPT-4o, GPT-4.5 та Gemini, які вирізняються високим рівнем контекстуального розуміння, точністю генерації, лексичним багатством та стабільною багатомовною продуктивністю. Водночас такі моделі, як Nemotron-4, Llama 3.1, Qwen2.5 і DeepSeek R1, хоча й поступаються флагманам, демонструють високу ефективність з урахуванням відкритого коду та оптимізації на ресурсах.

Описані результати можуть бути використані для практичного вибору LLM у філологічних дослідженнях, створенні цифрових архівів, розробці освітніх платформ і онтологічних систем, адаптованих до мовної специфіки. Стаття також пропонує формалізований підхід до комплексної оцінки LLM у гуманітарному просторі, що може стати підґрунтям для створення прозорих метрик і стандартів якості моделей штучного інтелекту у лінгвістичних задачах.

Ключові слова: *великі мовні моделі (LLM), лінгвістичні задачі, аналіз ієрархій Саати, Claude 3, GPT-4o, Gemini, стилістичний аналіз, онтології, штучний інтелект у філології.*

Volodymyr PASICHNYK,
orcid.org/0000-0002-5231-6395
Doctor of Technical Sciences, Professor,
Professor at the Department of Information Systems and Networks
Lviv Polytechnic National University
(Lviv, Ukraine) vpasichnyk@gmail.com

Maksym YAROMYCH,
orcid.org/0009-0005-3299-6695
PhD student at the Department of Applied Linguistics
Lviv Polytechnic National University
(Lviv, Ukraine) yamax0312@gmail.com

COMPARATIVE ANALYSIS OF LARGE LANGUAGE MODELS IN THE CONTEXT OF SOLVING LINGUISTIC TASKS

This article presents an interdisciplinary study of the comparative effectiveness of large language models in the context of solving linguistic tasks. The focus is on next-generation models such as GPT-4o, Claude 3, Gemini, DeepSeek R1, Grok, as well as open systems like BLOOM and LLaMA 2, which are actively used in linguistics, philology, translation studies, and digital humanities. The study substantiates the need to conceptually reframe LLMs not merely as text generation tools, but as "research assistants" capable of performing complex tasks including semantic analysis, stylistic interpretation, topic modeling, automatic text summarization, and cross-linguistic comparison.

The methodological foundation of the analysis is the modified Group Analytic Hierarchy Process (Group AHP), which enables the integration of assessments from three expert groups–linguists, humanities scholars, and IT professionals–into a unified priority vector. A total of 15 evaluation criteria were defined, encompassing both linguistic aspects (context understanding, grammar correction, stylistic richness, multilinguality) and technical parameters (hardware requirements, generation speed, API integration). Pairwise comparison matrices were constructed separately for each expert group, and the results were then aggregated using the geometric mean. This led to a consolidated weighting vector of the criteria and a ranking of models based on their effectiveness in linguistic tasks.

As a result of the Group AHP evaluation, the models with the highest global weights were Claude 3, GPT-4o, GPT-4.5, and Gemini, distinguished by their high levels of contextual understanding, text generation accuracy, lexical richness, and stable multilingual performance. Meanwhile, models such as Nemotron-4, Llama 3.1, Qwen2.5, and DeepSeek R1, while less dominant, demonstrate strong effectiveness considering their open-source nature and hardware efficiency.

These findings can inform practical choices of LLMs for philological research, digital archive construction, educational platform development, and ontology-based systems tailored to linguistic specificity. The article also introduces a formalized approach to comprehensive LLM assessment in the humanities, which may serve as a foundation for creating transparent metrics and quality standards for AI models used in linguistic contexts.

Key words: *large language models (LLM), linguistic tasks, Analytic Hierarchy Process, Claude 3, GPT-4o, Gemini, stylistic analysis, ontologies, artificial intelligence in philology.*

Постановка проблеми. В умовах стрімкого розвитку штучного інтелекту великі мовні моделі стали не лише технологічним проривом, а й інструментами, що дедалі глибше проникають у сферу гуманітарних знань. Якщо ще донедавна генеративні моделі розглядалися переважно як допоміжні засоби для створення текстів чи пошуку інформації, то сьогодні вони дедалі частіше використовуються у складніших задачах – аналізі дискурсу, стилістичній трансформації, семантичному узагальненні, автоматичній референтації, інтерпретації літературних текстів (Neuhaus, 2024). Усе це свідчить про необхідність серйозного переосмислення їхньої ролі в лінгвістичних дослідженнях і водночас про потребу в ретельному порівняльному оцінюванні їхньої якості.

Великі мовні моделі на кшталт GPT-4, Claude 3, Gemini, DeepSeek або Grok розрізняються за архітектурою, кількістю параметрів, типами навчальних корпусів, підходами до фільтрації даних і подальшого тонкого налаштування. Однак за межами технічних характеристик не менш важливим постає питання практичної ефективності моделей – зокрема, у таких складних і багатовимірних завданнях, як лінгвістичний аналіз і філологічна інтерпретація тексту. Саме тут постає проблема: як порівняти LLM не лише як мовні інтерфейси або генератори загальних відповідей, а як потенційних «помічників-дослідників» у гуманітарній сфері, зокрема мовознавстві?

У цьому контексті актуальність дослідження полягає в кількох важливих аспектах. По-перше, LLM вже активно використовуються у лінгвістичних задачах, таких як автоматичне позначення частин мови, синтаксичне та семантичне розборювання, виявлення стилістичних особливостей,

тематичне моделювання та навіть елементи літературної критики (Tkachenko, 2021). Відповідно, якість таких застосувань залежить не лише від загальної продуктивності моделі, а й від її здатності до глибокого контекстуального розуміння, лексико-граматичної чутливості, точності відтворення стилістичних норм, дотримання фактологічної достовірності (Mukanova et al., 2024). Це вимагає особливого типу оцінювання – більш структурованого, ніж звичайне інтуїтивне тестування.

По-друге, лінгвістичні задачі є міждисциплінарними за своєю суттю. Вони передбачають співпрацю між лінгвістами, філологами, перекладознавцями, культурологами та інженерами штучного інтелекту. У такій ситуації виникає потреба у спільній експертній оцінці, де кожна група фахівців оцінює модель зі своєї перспективи: лінгвісти – за граматичною точністю та семантичними нюансами, гуманітарії – за інтерпретативною глибиною, аналітики – за технічними характеристиками, масштабованістю й стабільністю (Giglou et al., 2023). Саме тому для дослідження було обрано груповий метод аналізу ієрархій (Group AHP), який дозволяє зібрати й об'єднати такі різновекторні оцінки в єдину систему пріоритетів.

По-третє, важливо не лише встановити поточний стан ефективності моделей, а й окреслити напрями для їхнього подальшого вдосконалення. Порівняння моделей за 15 критеріями, релевантними до лінгвістичних задач – такими як розуміння контексту, лексичне багатство, багатомовність, розпізнавання граматичних помилок, стилістична чутливість – дозволяє сформувати чіткий профіль кожної моделі: її сильні й слабкі сторони в контексті практичного використання в лінгвістичних дослідженнях або освітніх завданнях. Це особливо корисно для викладачів, науковців, редакторів і всіх, хто інтегрує LLM у свою фахову діяльність.

Відтак, порівняльне оцінювання великих мовних моделей за допомогою групового методу аналізу ієрархій забезпечує об'єктивний рейтинг моделей, а й створює нову методологічну основу для аналізу якості штучного інтелекту у гуманітарному просторі. Це особливо важливо в контексті зростання інтеграції ШІ в освітній процес, наукове письмо, культурну аналітику та цифрові гуманітарні дослідження (Mu et al., 2024). Така робота дозволяє не просто сказати, яка модель є «кращою», а й пояснити – чому, для кого і в яких умовах вона буде ефективнішою.

Аналіз останніх досліджень. *Порівнювані великі мовні моделі.* **GPT-4o** (від "omni") – це найновіша флагманська мультимодальна модель від OpenAI, яка виводить концепцію інтеграції вхідних модальностей на новий рівень. Вперше в історії OpenAI ця модель має єдину архітектуру, здатну нативно приймати та комбінувати текст, зображення та аудіо в режимі реального часу, що усуває необхідність у допоміжних конвертаційних модулях (Introducing GPT-4o, 2024). GPT-4o працює швидше, ніж GPT-4-turbo, водночас будучи дешевшою для розробників. Її ключова перевага – не просто багатомодальність, а повна синхронізація модальностей, яка дозволяє миттєво реагувати на голос, візуальний контент і текстовий контекст.

Ця модель забезпечує значно більш природну взаємодію з користувачами завдяки наднизькій затримці відповіді (~232 мс для голосу), імітуючи живу розмову. GPT-4o має поліпшене емоційне розпізнавання, підтримку понад 50 мов (включно з українською), здатність описувати зображення, допомагати в математиці та програмуванні. Попри те, що точна кількість параметрів не розголошується, модель демонструє рівень когнітивної продуктивності, порівнянний із GPT-4-turbo, і є кроком до «агентного» ШІ, здатного бачити, слухати й говорити. GPT-4o – це не просто оновлення, а фундаментальний стрибок у бік багатоканальної ШІ-взаємодії.

GPT-4.5 – це перехідна модель між GPT-4 і GPT-5, яка була розгорнута восени 2024 року. Вона значно покращила здатність до довготривалого контекстного аналізу завдяки вікну контексту до 128 тисяч токенів. Модель орієнтована на високу точність, кращу фактологічну узгодженість і стабільність у складних логічних та багатоетапних завданнях (OpenAI Platform, 2025). У порівнянні з GPT-4, GPT-4.5 демонструє кращу поведінкову інерцію (alignment), що дозволяє їй краще дотримуватись інструкцій, уникати упереджень і забезпечувати більш «людиноподібну» мову.

Технічні деталі GPT-4.5 залишаються частково засекреченими, однак відомо, що вона використовує оновлену форму моделі на базі інфраструктури Azure AI Supercomputing. Ймовірна кількість параметрів – близько 12,8 трильйона, хоча ці дані не були офіційно підтверджені. GPT-4.5 – це також модель із частково мультимодальними можливостями (текст + зображення), інтегрована з функцією function calling, memory API та tool use. У практичному сенсі вона використовується в системах, де потрібна висока стабільність, великий контекст і точність – зокрема, в юридичних, медичних, технічних консультаціях та науковій генерації.

DeepSeek R1 – це проривна модель від китайського стартапу DeepSeek, що належить до нової

генерації MoE-моделей (Mixture of Experts). Хоча загальна кількість параметрів становить 671 мільярд, модель активує лише ~37 мільярдів на кожен запит, що дозволяє зберегти баланс між високою продуктивністю та ефективним використанням ресурсів. DeepSeek R1 розроблена як повністю відкрита альтернатива GPT-4, з відкритим кодом і доступною ваговою базою, що робить її привабливою для академічного і корпоративного використання (DeepSeek-AI, 2024).

Особливістю DeepSeek є її навчання на величезному і високоякісному корпусі, що включає код, технічну документацію, енциклопедичні й освітні тексти. Вона демонструє чудові результати в математичних і логічних бенчмарках, програмуванні, а також у багатомовному генеративному письмі. Завдяки своїй архітектурі модель має кращу латентність і масштабованість, ніж звичайні dense-моделі (типу GPT-3/4), а також забезпечує стабільну якість генерації навіть при обмеженій обчислювальній потужності. Це робить її однією з найперспективніших моделей open-weight класу станом на 2025 рік.

Gemini – флагманська мультимодальна модель від Google DeepMind, яка поступово еволюціонувала від Gemini 1.0 до Gemini 2.5 Pro. Модель обробляє текст, зображення, аудіо та відео, що дозволяє їй працювати в широкому спектрі завдань – від генерації коду до аналізу зображень і створення відео. Gemini 2.5 Pro, представлена в березні 2025 року, має контекстне вікно до 1 мільйона токенів і покращені можливості логічного мислення, зокрема завдяки використанню техніки «chain-of-thought prompting» (Georgiev et al., 2024).

Модель інтегрована в екосистему Google, включаючи Gmail, Docs, Android-пристрої та Google Search. Gemini також доступна через API для розробників та в хмарних сервісах Google Cloud. Особливістю Gemini є її здатність до «агентного» мислення, що дозволяє їй самостійно планувати та виконувати складні завдання, що робить її потужним інструментом для бізнесу та розробників

Grok 3 – це третє покоління мовної моделі від xAI, компанії, заснованої Ілоном Маском. Модель була представлена в лютому 2025 року і позиціонується як одна з найрозумніших на ринку. Grok 3 має три версії: базову, Think (з покращеним логічним мисленням) та Big Brain (з розширеними обчислювальними можливостями). Модель демонструє високі результати в тестах з математики, науки та програмування, перевершуючи конкурентів, таких як GPT-4 та Gemini (Grok, 2025).

Особливістю Grok 3 є функція DeepSearch, яка дозволяє моделі отримувати актуальну інформацію з інтернету в реальному часі. Модель також інтегрована в соціальну мережу X (колишній Twitter), де доступна для преміум-користувачів. Grok 3 орієнтована на надання точних, логічно обґрунтованих відповідей, що робить її корисною для професійного використання та досліджень .

Claude 3 – це сімейство моделей від компанії Anthropic, яке включає три версії: Haiku (швидка та легка), Sonnet (баланс між швидкістю та потужністю) та Opus (найпотужніша). Модель була представлена в березні 2024 року і демонструє високі результати в когнітивних завданнях, перевершуючи GPT-4 та Gemini Ultra в деяких бенчмарках (Anthropic Team, 2025).

Claude 3 підтримує обробку тексту та зображень, має контекстне вікно до 200 тисяч токенів, а в деяких випадках – до 1 мільйона. Модель використовує підхід «Constitutional AI», що забезпечує безпечну та етичну взаємодію з користувачами. Claude 3 інтегрована в хмарні сервіси Amazon Bedrock та доступна через API для розробників, що робить її зручною для бізнесу та наукових досліджень .

Mistral 7B – це відкрита мовна модель з 7,3 мільярдами параметрів, розроблена французькою компанією Mistral AI. Вона використовує архітектуру трансформера з інноваційними механізмами, такими як Grouped-Query Attention (GQA) для швидшого виведення та Sliding Window Attention (SWA) для ефективної обробки довгих послідовностей. Ці технічні рішення дозволяють моделі перевершувати Llama 2 13B на всіх бенчмарках та Llama 1 34B у багатьох завданнях, зокрема в генерації коду та обробці англійської мови (Mistral 7B, 2024).

Mistral 7B випущена під ліцензією Apache 2.0, що дозволяє вільне використання та модифікацію. Вона доступна у двох версіях: базовій та інструкційно налаштованій (Instruct), оптимізованій для завдань, що вимагають дотримання інструкцій. Завдяки своїй ефективності та відкритості, Mistral 7B стала популярною серед розробників та дослідників, які шукають потужну модель з помірними вимогами до ресурсів.

Phi-2 – компактна мовна модель з 2,7 мільярдами параметрів, розроблена Microsoft Research. Незважаючи на свій невеликий розмір, вона демонструє видатні результати в завданнях, що вимагають логічного мислення та розуміння мови, перевершуючи моделі, які в 25 разів більші. Такий ефект досягнуто завдяки ретельному відбору навчальних даних, що включають синтетичні тексти та фільтровані веб-джерела з акцентом на безпеку та освітню цінність.

Phi-2 є частиною серії «Phi» від Microsoft, спрямованої на створення високопродуктивних малих мовних моделей. Вона особливо підходить для використання на пристроях з обмеженими ресурсами, таких як мобільні телефони та вбудовані системи, забезпечуючи при цьому високу якість генерації тексту та логічного виведення.

Llama 3.1 – це найновіша версія відкритої мовної моделі від Meta AI, представлена у липні 2024 року. Вона доступна в трьох конфігураціях: 8B, 70B та 405B параметрів. Найбільша версія, Llama 3.1 405B, є найбільшою відкритою мовною моделлю на сьогоднішній день, перевершуючи GPT-4o та Claude 3.5 Sonnet у кількох бенчмарках. Модель була навчена на понад 15 трильйонах токенів за допомогою 16 000 GPU Nvidia H100, що забезпечує її високу продуктивність та універсальність (Phi-2, 2023).

Llama 3.1 підтримує багатомовність, генерацію коду та має розширені можливості для інтеграції в різні продукти Meta, такі як WhatsApp, Instagram та Facebook. Вона також доступна для розробників через API та платформи, такі як Hugging Face. Відкрита ліцензія моделі сприяє її широкому впровадженню в дослідницьких та комерційних проектах.

Nemotron-4 340B – відкрита мовна модель від NVIDIA, розроблена для генерації синтетичних даних високої якості, що дозволяє зменшити залежність від дорогих та обмежених реальних датасетів. Модель має 340 мільярдів параметрів і представлена у трьох варіантах: базова, інструкційно налаштована та модель з підкріпленням. Вона оптимізована для роботи з фреймворком NVIDIA NeMo та бібліотекою TensorRT-LLM, що забезпечує ефективне навчання та інференцію на апаратному забезпеченні NVIDIA (Nemotron-4 340B, 2024).

Nemotron-4 340B демонструє високу продуктивність у різноманітних бенчмарках, часто перевершуючи інші відкриті моделі, такі як Llama-3 70B та Mixtral 8X7B. Вона здатна працювати на одному сервері NVIDIA DGX H100 з 8 GPU, що робить її доступною для широкого кола дослідників та розробників. Модель також використовує методи Supervised Fine-Tuning (SFT) та Reinforcement Learning with Human Feedback (RLHF) для покращення здатності до слідування інструкціям та взаємодії з користувачами.

Qwen2.5 – серія великих мовних моделей від Alibaba Cloud, яка включає як щільні (dense), так і розріджені (MoE) архітектури з кількістю параметрів від 0,5 до 72 мільярдів. Моделі були попередньо навчені на 18 трильйонах токенів та додатково налаштовані за допомогою понад 1 мільйона зразків для Supervised Fine-Tuning (SFT) та багатоступеневого Reinforcement Learning. Qwen2.5 підтримує до 128 тисяч токенів у контекстному вікні та демонструє високі результати в завданнях, що вимагають логічного мислення, генерації коду та багатомовної обробки тексту (Qwen2-0.5B-Instruct, 2023).

Особливістю Qwen2.5 є її відкритість: багато моделей серії доступні з відкритим кодом під ліцензією Apache 2.0. Крім того, Alibaba представила мультимодальні версії, такі як Qwen2.5-Omni-7B, здатні обробляти текст, зображення, аудіо та відео, що робить їх придатними для використання на мобільних пристроях та вбудованих системах. Моделі Qwen2.5 також інтегровані в екосистему Alibaba Cloud, надаючи розробникам гнучкі та потужні інструменти для створення інтелектуальних застосунків.

PaLM 2 – це друга генерація великих мовних моделей від Google, яка відзначається покращеними можливостями багатомовної обробки, логічного мислення та генерації коду. Модель була навчена на великому корпусі текстів понад 100 мовами, що дозволяє їй досягати високих результатів у багатомовних тестах та завданнях перекладу. Крім того, PaLM 2 демонструє покращену ефективність у порівнянні з попередником, що дозволяє її інтегрувати в різноманітні продукти та сервіси Google (Introducing PaLM 2, 2025).

PaLM 2 вже використовується в понад 25 продуктах Google, включаючи Gmail, Google Docs та Google Bard, забезпечуючи функції, такі як автоматичне резюмування, генерація тексту та інтелектуальні відповіді. Модель також доступна для розробників через API, що дозволяє створювати власні застосунки з використанням потужностей PaLM 2 у завданнях генерації тексту, кодування, аналізу даних та інших.

LaMDA (Language Model for Dialogue Applications) – це серія мовних моделей від Google, спрямованих на покращення якості діалогових систем. Перша версія була представлена у 2021 році, з подальшими оновленнями, включаючи LaMDA 2 у 2022 році. Модель розроблена для ведення природних, відкритих розмов на широкий спектр тем, демонструючи глибоке розуміння контексту та нюансів мови (LaMDA, 2025).

LaMDA стала основою для чат-бота Bard, який був запущений у 2023 році. Bard використовує можливості LaMDA для надання інформативних та контекстуально релевантних відповідей, що робить його конкурентом інших передових мовних моделей, таких як ChatGPT.

Megatron-Turing NLG 530B – це спільна розробка Microsoft і NVIDIA, яка на момент випуску була найбільшою монолітною трансформерною мовною моделлю з 530 мільярдами параметрів. Модель побудована на 105-шаровій архітектурі трансформера з 128 головками уваги, що забезпечує її високу продуктивність у завданнях генерації тексту та розуміння мови. T-NLG демонструє передові результати в різноманітних NLP-бенчмарках, включаючи завдання логічного мислення та генерації коду. Її розробка стала можливою завдяки використанню фреймворків DeepSpeed і Megatron, що дозволило ефективно масштабувати навчання моделі на великій кількості GPU (Smith et al., 2022).

BLOOM (BigScience Large Open-science Open-access Multilingual Language Model) – це відкрита багатомовна мовна модель з 176 мільярдами параметрів, розроблена в рамках міжнародного проєкту BigScience. Модель була навчена на даних з 46 природних мов і 13 мов програмування, що робить її однією з наймасштабніших відкритих моделей такого типу. BLOOM призначена для генерації тексту, продовження підказок та інших завдань обробки природної мови. Її відкритий доступ сприяє дослідженням і розвитку в галузі штучного інтелекту, дозволяючи дослідникам і розробникам використовувати модель без обмежень, характерних для комерційних альтернатив (BigScience, 2024).

LLaMA 2 (Large Language Model Meta AI) – це друга генерація відкритих мовних моделей від Meta AI, представлена у липні 2023 року. Модель доступна в трьох конфігураціях: 7B, 13B та 70B параметрів, і була навчена на значно більшому обсязі даних порівняно з попередньою версією. LLaMA 2 включає як базові моделі, так і версії, оптимізовані для діалогових застосувань (LLaMA 2-Chat). Модель демонструє високу продуктивність у завданнях генерації тексту, логічного мислення та програмування, і доступна для дослідників і розробників з відкритим кодом, що сприяє її широкому впровадженню в різних сферах (Touvron et al., 2023).

Метод аналізу ієрархій Саати. Для проведення дослідження використаємо метод аналізу ієрархій (AHP, від англ. *Analytic Hierarchy Process*) був розроблений американським вченим Томасом Сааті у 1970-х роках як універсальний підхід до вирішення складних задач прийняття рішень (Saaty, 1980). Основна ідея методу полягає у структурованому представленні проблеми у вигляді ієрархії, де верхній рівень визначає загальну мету, середній – сукупність критеріїв, а нижній – конкретні альтернативи, між якими здійснюється вибір. Такий підхід дозволяє не лише впорядкувати складні багатофакторні завдання, а й поєднати у процесі прийняття рішень як кількісні, так і якісні показники.

AHP особливо ефективний у ситуаціях, коли необхідно оцінити альтернативи за багатьма критеріями, об'єднати експертні судження різного ступеня суб'єктивності або інтегрувати кількісні та якісні фактори в єдину систему (Сааті, 2008). Типовими прикладами застосування є вибір постачальника, визначення пріоритетів інвестицій, оцінювання ризиків або формування стратегій.

Застосування AHP починається з побудови ієрархічної структури задачі. На її вершині розміщується головна мета – наприклад, вибір оптимального рішення. Наступний рівень ієрархії складається з критеріїв, за якими здійснюється оцінка альтернатив (таких як вартість, якість, надійність, терміни тощо). На нижньому рівні подаються самі альтернативи – об'єкти або варіанти, між якими здійснюється вибір.

Ключовим елементом методу є формування матриць парних порівнянь. У рамках кожного рівня ієрархії всі елементи порівнюються між собою попарно щодо їхнього впливу на вищий рівень. Для цього використовується шкала Сааті – дев'ятибальна шкала, де значення 1 означає рівнозначність двох елементів, а 3, 5, 7 і 9 відповідають поступовому зростанню переваги одного елемента над іншим. Проміжні значення (2, 4, 6, 8) використовуються для відтінків переваг.

Після заповнення матриць здійснюється нормалізація, на основі якої обчислюються вагові коефіцієнти (власні вектори), що відображають відносну важливість кожного елемента. З огляду на те, що метод передбачає використання людських суджень, важливо перевірити їх узгодженість. Для цього розраховується індекс узгодженості. Якщо його значення перевищує 0.1, це свідчить про наявність суперечностей у судженнях, і відповідні оцінки потребують перегляду.

На завершальному етапі здійснюється агрегування результатів: ваги альтернатив підсумовуються відповідно до важливості кожного критерію, і таким чином визначається загальний пріоритет кожної альтернативи. У підсумку приймається рішення на користь тієї альтернативи, яка має найбільшу підсумкову вагу.

Таким чином, метод аналізу ієрархій є надійним, гнучким і прозорим інструментом багатокритеріального аналізу, що дозволяє ухвалювати виважені рішення навіть у складних і нечітко структурованих ситуаціях.

Мета статті – проаналізувати ефективність великих мовних моделей у вирішенні лінгвістичних завдань шляхом застосування групового методу аналізу ієрархій Сааті для їх оцінювання за 15 лінгвістичними та технічними критеріями та розробити формалізований підхід для створення прозорих метрик якості в гуманітарних науках із наданням практичних рекомендацій для використання LLM у філологічних дослідженнях.

Виклад основного матеріалу. *Критерії оцінювання.* Визначимо основні критерії оцінювання ефективності різних великих мовних моделей для розв'язування лігвістичних задач:

1. Обсяг пам'яті, необхідний для роботи. Для багатьох лінгвістичних задач, особливо при роботі з великими корпусами або інтеграції у локальні середовища, важливо, щоб модель могла функціонувати в межах доступних ресурсів. Високі вимоги до пам'яті ускладнюють розгортання на звичайних комп'ютерах або мобільних пристроях, що обмежує її застосування в освітніх чи польових умовах.

2. Масштабованість. Масштабованість дозволяє адаптувати модель до різних обсягів задач – від аналізу окремого документа до обробки великих корпусів. Це критично для наукових досліджень, де обсяг даних часто зростає, а також для реальних додатків (наприклад, чат-ботів або мовних інтерфейсів), що мають витримувати навантаження.

3. Взаємодія з іншими системами (API, інтеграція). Багато лінгвістичних задач вимагають поєднання LLM із іншими сервісами – базами даних, перекладачами, лінгвістичними аналізаторами або освітніми платформами. Потужний і гнучкий API, можливість інтеграції через REST/GraphQL чи Python SDK – усе це критично для практичного використання моделей у лінгвістичному середовищі.

4. Вартість використання. Особливо важлива у контексті досліджень, освіти та малих проєктів. Надто висока вартість (оплата за токени або запити) обмежує частоту використання моделі та її застосування в довготривалих експериментах або викладанні.

5. Дата останнього оновлення знань, наявність автоматичних оновлень. Актуальність знань моделі є вирішальною для лінгвістичних задач, де мова, терміни, соціокультурні контексти та тренди швидко змінюються. Модель, яка «застрягла» в датасетах 2021 року, не здатна коректно обробити сучасні тексти, нові мовні явища чи медіа-дискурс.

6. Якість генерації тексту. Висока якість генерації необхідна для задач парафразування, автоматичного написання текстів, пояснення граматики, лексики, побудови прикладів і тестів. Погана генерація веде до помилок, штучності або стилістичної убогості.

7. Розуміння контексту. Контекст – ключ до інтерпретації лексики, граматики, дискурсу. Для лінгвістичних задач важливо, щоб модель могла розпізнавати співвідношення між репліками у діалозі, тематичну цілісність тексту, співвідношення анафор та референтів. Це критично для синтаксичного та семантичного аналізу.

8. Багатомовність. У сучасній прикладній лінгвістиці багатомовність – це не лише перевага, а необхідність. Моделі мають бути здатні працювати не лише англійською, а й українською, німецькою, іспанською тощо. Це важливо для перекладу, навчання мов, порівняльного мовознавства та мовної інклюзивності.

9. Можливість адаптації до спеціалізованих завдань. Лінгвістичні задачі можуть бути дуже вузькими: аналіз юридичної мови, технічної термінології, поетичного стилю. Важливо, щоб модель могла бути донавчена або адаптована до таких специфічних доменів – інакше її генерації будуть загальними, поверхневими.

10.Лексичне та стилістичне багатство. Багатство лексики і стилю є критично важливим для задач типу: стилістичний аналіз, жанрова класифікація, автоматичне редагування, написання художніх чи академічних текстів. Бідна модель не здатна відтворити стилістичні нюанси мови – вона звучить «машинно»

11.Швидкість генерації тексту. У задачах, пов'язаних із реальним часом – наприклад, мовна підтримка в інтерфейсах, живі переклади або інтерактивні тести – затримка в генерації псує користувацький досвід. Швидкість також критична при обробці великих корпусів.

12.Врахування контексту знань та фактів. Для лінгвістичних задач важливо, щоб модель не вигадувала інформацію (галюцинації), а ґрунтувалася на реальних фактах. Наприклад, при поясненні граматики, етимології або перекладу термінів, неточна інформація шкодить навчанню або дослідженню.

13.Розпізнавання та корекція граматичних помилок. Цей критерій важливий для створення систем автоматичного редагування, перевірки письма, навчальних інструментів для мов. Модель має не лише знаходити помилки, а й пояснювати їх та пропонувати стилістично і граматично правильні варіанти.

14.Модельні упередження та безпека. У лінгвістичних застосунках, пов'язаних із соціально

чутливою лексикою, питанням ідентичності, культур – надзвичайно важливо, щоб модель не продукувала дискримінаційні або упереджені висловлювання. Також важливо, щоб вона не відповідала на «провокаційні» або шкідливі запити (на кшталт написання плагіату).

15. Апаратні вимоги та оптимальність роботи. Для використання в освітніх установах, польових дослідженнях або мобільних додатках важливо, щоб модель не потребувала дорогих обчислювальних ресурсів. Оптимізовані моделі (наприклад, Phi-2 або Mistral 7B) забезпечують високу якість при мінімальних витратах, що робить їх більш придатними для широкого використання.

Лінгвістичні задачі, що можуть бути вирішені засобами великих мовних моделей. У сучасній філології та лінгвістиці великі мовні моделі відкривають нові горизонти для автоматизації, аналізу й генерації текстів. Завдяки глибокому навчанню на масивних текстових корпусах і трансформерній архітектурі, такі моделі, як GPT, BERT, Claude, можуть розв'язувати завдання, які раніше вимагали багато часу і ручної роботи (Пасічник, 2025a). Особливо перспективною є їх інтеграція з онтологіями – формалізованими структурами знань, які забезпечують точність, узгодженість і контроль над термінами й контекстами.

1. Семантичний аналіз. Однією з найважливіших задач, яку можна автоматизувати за допомогою великих мовних моделей, є семантичний аналіз – визначення значення слів, фраз, синтаксичних конструкцій та їх зв'язків у контексті. Ця задача особливо актуальна при роботі з полісемічними словами, синонімами, метафорами та іншими складними мовними явищами (Giglou et al., 2023). LLM, натреновані на багатомовних корпусах, дозволяють ідентифікувати значення з урахуванням контексту, що є особливо важливим для точного лексичного аналізу. У поєднанні з онтологіями, які задають формальні межі понять, моделі здатні структурувати знання і мінімізувати неоднозначності.

2. Стилістичний і стилометричний аналіз. Завдяки здатності великих мовних моделей до розпізнавання стилістичних патернів, стає можливим проведення стилістичного аналізу з високою точністю. Моделі можуть визначати тональність тексту, рівень формальності, вживання фігур мови, а також робити стилометричну класифікацію – наприклад, для визначення авторства твору або стилістичної належності (Opara, 2024). У дослідженнях доведено, що LLM перевершують традиційні підходи у виявленні стилістичних маркерів, таких як метафори, алегорії чи іронія, оскільки здатні враховувати широкий контекст, включно з культурним і жанровим.

3. Лексикографія і побудова словників. Великі мовні моделі значно спрощують процес створення й редагування словників, адже здатні автоматично витягувати лексичні одиниці, наводити їх контекстуальні значення та приклади вживання (Кунанець, 2025; Пасічник, 2025d). Це дозволяє модернізувати лексикографічну практику, уникаючи громіздких ручних процедур. Моделі можуть також порівнювати значення в різних мовах і фіксувати міжмовні відповідності, що особливо цінно для створення перекладних і термінологічних словників у багатомовних контекстах.

4. Жанрова класифікація та тематичне моделювання. Великі мовні моделі здатні класифікувати тексти за жанрами й тематиками, виявляючи внутрішні закономірності, стилістичні особливості та ключові концепти. Це корисно для літературознавства, де важливо визначити жанрову приналежність твору, його інтертекстуальні зв'язки, домінуючі мотиви або семантичні ядра (Schreiber, 2016). LLM використовують тематичне моделювання (topic modeling), кластеризацію й аналіз ключових слів, що дозволяє виявляти структуру тексту навіть без попереднього маркування (Пасічник, 2025c).

5. Інтертекстуальний аналіз. Завдяки здатності працювати з великими обсягами тексту, великі мовні моделі здатні виявляти цитати, алюзії, ремінісценції, пародії та інші інтертекстуальні елементи. Це дозволяє створювати карти інтертекстуальності, виявляти неочевидні зв'язки між творами та аналізувати взаємовплив літературних традицій. Застосування онтологій у цьому процесі підвищує точність, оскільки допомагає моделі структурувати й обмежувати простір пошуку відповідно до заданих концептів і жанрів (Umphrey et al., 2024).

6. Порівняльне мовознавство і багатомовний аналіз. Великі мовні моделі, натреновані на багатомовних корпусах, надають потужні інструменти для порівняльного аналізу мов. Вони здатні одночасно працювати з десятками мов, порівнювати граматичні структури, лексичні поля, стилістичні відмінності, що особливо цінно для лінгвістів, які працюють у сфері контрастивної лінгвістики, перекладознавства або типології. Поєднання LLM з онтологіями дозволяє ще глибше зануритись у міжкультурні паралелі, фіксуючи відмінності не лише на мовному, але й на концептуальному рівні.

7. Робота з історичними та літературними текстами. Філологи можуть використовувати великі мовні моделі для аналізу стилістики історич-

них текстів, реконструкції контексту епохи, імітації мовного стилю минулих періодів. Моделі можуть також генерувати варіанти перекладу з урахуванням стилістичних особливостей джерела або створювати узагальнення про літературні стилі різних епох. Особливо ефективною є їх взаємодія з онтологіями, які враховують культуру, епоху, лексичний склад і стилістичні фігури текстів минулого.

8. Генерація та переписування текстів. Великі мовні моделі здатні автоматично генерувати різні види текстів – від коротких рефератів і анотацій до літературних оглядів і навчальних матеріалів. Це може бути використано як допоміжний інструмент при створенні текстів із чітко заданою структурою або стилем (Пасічник, 2025b). Також моделі здатні переписувати тексти з урахуванням рівня складності, жанру, стилю чи аудиторії, що є цінним у викладацькій та редакторській діяльності. При цьому онтології допомагають уникати фактичних помилок, задаючи базу знань для контролю достовірності.

9. Побудова цифрових архівів та онтологічних баз знань. Великі мовні моделі у поєднанні з онтологіями дозволяють створювати інтерактивні цифрові архіви, які включають автоматично класифіковані, анотовані й тематично згруповані тексти. Це сприяє створенню нових інструментів для навчання, дослідження й популяризації літературної спадщини (Mukanova et al., 2024). Такі архіви можуть інтегрувати знання з різних дисциплін – мовознавства, літературознавства, культурології – і забезпечити інноваційний підхід до вивчення гуманітарних наук

Визначення ваг критеріїв. Після формування повного переліку критеріїв оцінювання великих мовних моделей у лінгвістичних задачах наступним етапом є побудова матриці парних порівнянь критеріїв. Цей процес здійснюється відповідно до методології аналізу ієрархій Сааті й передбачає попарне зіставлення кожного критерію з усіма іншими з точки зору їх відносної важливості. Експерти порівнюють, наскільки один критерій є важливішим за інший у контексті мовної ефективності моделей, використовуючи дев'ятибальну шкалу Сааті, де значення 1 означає рівнозначність, а значення 3, 5, 7 і 9 вказують на перевагу одного критерію над іншим у зростаючому порядку.

Усі ці оцінки заносяться в квадратну матрицю, у якій елементи відображають перевагу одного критерію над іншим. Матриця є взаємно оберненою, тобто якщо критерій А важливіший за критерій В з оцінкою 5, то навпаки – критерій В порівняно з А матиме значення 1/5. Після заповнення всієї матриці на її основі обчислюється вектор ваг критеріїв, який відображає відносну важливість кожного критерію в загальній структурі прийняття рішень.

Для обчислення вектора ваг використовується метод головного власного вектора. Ці ваги критеріїв використовуються на наступному рівні ієрархії для зважування локальних векторів альтернатив – тобто для обчислення фінального вектора глобальних пріоритетів мовних моделей, який узагальнює їхню ефективність у всіх обраних параметрах одночасно. Такий підхід забезпечує логічну послідовність оцінювання, формалізує суб'єктивні судження експертів і дозволяє приймати виважені рішення на основі багатокритеріального аналізу.

Перевірка узгодженості. Наступним важливим етапом є перевірка, чи були порівняння між

Таблиця 1

Матриця попарних порівнянь критеріїв

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1.	1	1/3	1/3	1/3	1/3	1/7	1/9	1/7	1/7	1/7	1/5	1/9	1/9	1/5	1/5
2.	3	1	1	1	1	1/5	1/7	1/5	1/5	1/5	1/3	1/7	1/7	1/3	1/3
3.	3	1	1	1	1	1/5	1/7	1/5	1/5	1/5	1/3	1/7	1/7	1/3	1/3
4.	3	1	1	1	1	1/5	1/7	1/5	1/5	1/5	1/3	1/7	1/7	1/3	1/3
5.	3	1	1	1	1	1/5	1/7	1/5	1/5	1/5	1/3	1/7	1/7	1/3	1/3
6.	7	5	5	5	5	1	1	3	3	3	3	1	1	3	5
7.	9	7	7	7	7	1	1	3	3	3	3	1	1	3	5
8.	7	5	5	5	5	1/3	1/3	1	1	1	1	1/3	1/3	1	3
9.	7	5	5	5	5	1/3	1/3	1	1	1	1	1/3	1/3	1	3
10.	7	5	5	5	5	1/3	1/3	1	1	1	1	1/3	1/3	1	3
11.	5	3	3	3	3	1/3	1/3	1	1	1	1	1/5	1/5	1	3
12.	9	7	7	7	7	1	1	3	3	3	5	1	1	3	5
13.	9	7	7	7	7	1	1	3	3	3	5	1	1	3	5
14.	5	3	3	3	3	1/3	1/3	1	1	1	1	1/3	1/3	1	3
15.	5	3	3	3	3	1/5	1/5	1/3	1/3	1/3	1/3	1/5	1/5	1/3	1

Таблиця 2

Ваги критеріїв

12	Врахування знань та фактів	0.1510226055795294
13	Корекція граматики	0.15102260557952932
7	Розуміння контексту	0.14427477428180232
6	Якість генерації тексту	0.13381626539730092
8	Багатомовність	0.06572900311756194
9	Адаптація до завдань	0.06572900311756194
10	Лексичне багатство	0.06572900311756192
14	Безпека та упередження	0.05526992372165737
11	Швидкість генерації	0.05269089309149482
15	Апаратні вимоги	0.03305520337322098
3	Взаємодія з іншими системами (API)	0.017867924248544556
4	Вартість використання	0.017867924248544556
5	Дата оновлення знань	0.017867924248544556
2	Масштабованість	0.017867924248544553
1	Обсяг пам'яті	0.010187942627500064

критеріями достатньо послідовними (не суперечливими). Для цього використовується індекс узгодженості (Consistency Index, CI) та коефіцієнт узгодженості (Consistency Ratio, CR).

Формула для CI:

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

де: *n* – кількість критеріїв (n = 15), а λmax – максимальне власне значення.

Після цього визначаємо коефіцієнт узгодженості CR:

$$CR = \frac{CI}{RI}$$

де: CI – середнє випадкове узгодження (Random Index), для n=15 дорівнює 1.59 (за стандартною таблицею Сааті).

Інтерпретація CR:

Таблиця 2. Ваги критеріїв Якщо CR < 0.1, то узгодженість хороша, оцінки є коректними.

– Якщо CR > 0.1, то варто переглянути порівняння критеріїв через суперечливість оцінок.

У даному випадку після розрахунків отримуємо:

– Індекс узгодженості (CI): 0.0439
– Коефіцієнт узгодженості (CR): 0.0276

Значення CR = 0.0276 суттєво менше за критичне значення 0.1, що вказує на дуже хорошу узгодженість експертних оцінок.

Формування груп експертів. У процесі формування ваг критеріїв оцінювання великих мовних моделей у лінгвістичних задачах було враховано пріоритети саме з точки зору фахівців-мовознавців. Згідно з результатами парних порівнянь, найбільш вагомими критеріями виявилися ті, що безпосередньо пов'язані з глибинною мовною обробкою та змістовою релевантністю результатів генерації. Зокрема, «розуміння контексту», «якість генерації тексту», «розпізнавання та корекція граматичних помилок», а також «врахування знань та фактів» отримали найвищі пріоритети. Їхня перевага над іншими критеріями була оцінена на рівні 7–9 балів за шкалою Сааті, що свідчить про їх критичне значення для успішного застосування LLM у лінгвістичних дослідженнях.

До групи критеріїв з помірною, але все ж важливою вагою потрапили «лексичне та стилістичне багатство», «багатомовність» і «можливість адаптації до спеціалізованих завдань». Хоча ці характеристики є суттєвими для практичної ефективності мовної моделі, особливо в багатомовному або спеціалізованому контексті (наприклад, в аналізі наукових, художніх або технічних текстів), їхня значущість порівняно з базовими параметрами семантичного аналізу була оцінена дещо нижче – в межах 3–5 балів.

Окрему позицію посів критерій «безпека та модельні упередження». Він також визнаний важливим, зокрема у завданнях, пов'язаних з навчанням, автоматичним оцінюванням і публічним використанням LLM. Однак з погляду суто лінгвістичного функціоналу, цей критерій не визначає ключової ефективності мовної обробки, тому отримав оцінку на рівні 3–5 балів.

Натомість усі технічні параметри, такі як обсяг пам'яті, масштабованість, можливість інтеграції через API, вартість використання, дата оновлення знань, апаратні вимоги та швидкість генерації, були розглянуті як вторинні з точки зору лінгвістики. Їхня роль у забезпеченні загальної продуктивності моделі визнається, однак вони не є визначальними у розв'язанні лінгвістичних задач. Тому ці критерії отримали найнижчі оцінки у системі пріоритетів.

Таким чином, ваги критеріїв, сформовані в результаті експертного оцінювання, відображають чітку домінацію якісних мовних параметрів над технічними характеристиками, що повністю відповідає специфіці філологічних і лінгвістичних досліджень, де на першому місці стоїть глибина мовного розуміння, точність, достовірність і стилістична чутливість.

Оцінювання ефективності великих мовних моделей у контексті розв'язання лінгвістичних задач потребує міждисциплінарного підходу. Через складну природу таких задач – на перетині мовознавства, штучного інтелекту та академічної

практики – доцільним є залучення трьох окремих груп експертів: лінгвістів, наукових працівників та ІТ-фахівців. Кожна з цих груп має унікальну компетенцію, яка дозволяє оцінити LLM з різних кутів зору: мовного, дослідницького та технічного. Нижче подано обґрунтування такого підходу.

Лінгвісти, як фахівці з аналізу структури, значення та функціонування мови, відіграють ключову роль у перевірці здатності моделей до виконання класичних лінгвістичних задач: морфологічного, синтаксичного, семантичного та прагматичного аналізу. Саме лінгвісти здатні оцінити, наскільки точно LLM відтворюють граматичні структури, розпізнають багатозначність, використовують стилістичні фігури та зберігають дискурсивну цілісність.

У завданнях типу: жанрова класифікація, розпізнавання стилістичних маркерів, виявлення семантичної неоднозначності – саме лінгвістична експертиза дозволяє встановити якісні межі коректної роботи моделі. Лінгвісти також здатні виявити випадки «галюцинацій» моделі, коли вона створює граматично правильні, але мовно чи логічно хибні твердження, що особливо важливо у навчальних і дослідницьких контекстах.

Друга група експертів – науковці, які працюють у галузі гуманітарних досліджень, зокрема філології, культурології, історії літератури чи перекладознавства. Вони не обов'язково є лінгвістами за фахом, однак мають практичний досвід роботи з великими текстами, джерелами, літературною критикою чи навчальними матеріалами. Саме ця група фокусується на інтерпретативному та джерелознавчому аспекті використання LLM: чи зберігає модель стилістичні особливості тексту? чи не спрощує вона культурні відтінки? чи коректно посилається на джерела?

Науковці можуть оцінити, наскільки модель здатна допомогти у вирішенні прикладних академічних задач – таких як анотація, тематичне моделювання, узагальнення текстів, аналіз художніх образів. Важливою частиною їхньої оцінки є перевірка наукової достовірності: чи генерує модель фейкові факти, неправильні дати, вигадані цитати тощо. Таким чином, вони фокусуються не лише на лінгвістичній якості, а й на інформаційній надійності й аналітичній цінності моделі для досліджень.

Третій аспект – технічний. ІТ-фахівці (зокрема NLP-інженери, аналітики даних, розробники ШІ-рішень) забезпечують критичну оцінку моделі з погляду інтеграції, продуктивності та обмежень. Вони аналізують швидкість генерації, апаратні вимоги, оптимізацію для локального або хмарного середовища, можливості API, конфігурації безпеки, а також функціональність донавчання.

Крім того, саме ця група здатна оцінити масштабованість, багатомовну підтримку, можливість адаптації моделі до вузької предметної області (domain adaptation), а також рівень модульності: наскільки легко LLM можна вбудувати в існуючу освітню платформу, науковий проєкт або цифровий архів. Вони також перевіряють, наскільки добре модель справляється з підтримкою Unicode, транслітерації, мультимодальності (текст + зображення) та розмітки.

Поєднання трьох експертних перспектив дає змогу забезпечити повноцінне і збалансоване оцінювання мовної моделі. Лінгвісти відповідають за мову, науковці – за зміст і академічну цінність, а ІТ-фахівці – за технічну реалізацію та інфраструктурну ефективність. Такий підхід дозволяє не лише оцінити вже наявну модель, а й сформулювати вимоги до майбутніх – з урахуванням специфіки лінгвістичних і філологічних задач.

Груповий метод. Груповий метод аналізу ієрархій постає як найдоцільніший інструмент для реалізації цього міждисциплінарного підходу (Сааті, 1989). На відміну від класичного варіанту AHP, орієнтованого на одного експерта, груповий метод дозволяє поєднати оцінки та судження кількох фахівців, що працюють у різних професійних доменах. У контексті оцінювання великих мовних моделей для лінгвістичних задач це означає можливість врахування вагомих, але різновекторних критеріїв, які формуються на основі специфіки кожної експертної групи.

Суть групового AHP полягає у створенні матриць парних порівнянь критеріїв або альтернатив кожним із експертів або експертних груп, після чого ці оцінки агрегуються в єдину, узгоджену матрицю.

Це дозволяє уникнути викривлень, властивих арифметичному середньому, і краще зберігає співвідношення переваг, особливо коли між експертами існують суттєві розбіжності. Після формування агрегованої матриці проводиться її нормалізація і обчислення вектора ваг – власного вектора, що представляє пріоритети критеріїв або альтернатив. У класичному варіанті цей вектор наближається через середнє нормалізованих рядків або через знаходження головного власного вектора матриці (методом степеневого наближення або іншим чисельним способом).

Наступним важливим кроком є перевірка узгодженості оцінок. Для цього обчислюється індекс узгодженості, який дозволяє оцінити, наскільки логічно послідовними були парні порівняння.

Якщо $CR < 0.1$, то оцінки вважаються узгодженими. У випадку перевищення цього порогу, необ-

хідно переглянути вихідні оцінки окремих експертів, аби забезпечити більшу послідовність.

У практичному сенсі це означає, що лінгвісти зможуть сфокусуватись на таких критеріях, як семантична точність, стильове багатство, граматична коректність і здатність моделі до адекватного мовного аналізу; наукові працівники – на вмінні моделі працювати з джерелами, створювати інтерпретації, узагальнення, дотримуватись академічної доброчесності; ІТ-фахівці – на технічній стабільності, продуктивності, масштабованості, можливості адаптації та інтеграції. Кожна з цих груп, працюючи в межах власної експертизи, формує власну матрицю оцінювання, яка потім об'єднується у єдину модель, що відображає колективний консенсус.

Таким чином, груповий метод не лише технічно вирішує проблему багатоголосся в оцінюванні, а й формує концептуально прозору методологію для гармонізації міждисциплінарних поглядів. Його перевагою є те, що він забезпечує як кількісну точність, так і якісну обґрунтованість пріоритетів. У результаті оцінювання можна отримати єдиний інтегрований вектор ваг критеріїв або рейтинг альтернатив, що враховує комплексний погляд усіх залучених сторін. У галузі гуманітарних досліджень, де важко застосувати суто технократичні або формальні методи оцінки, Group AHP стає ключовим інструментом, який дозволяє об'єднати мовну, змістову й технологічну логіку у збалансовану систему ухвалення рішень.

Побудова матриць парних порівнянь для критеріїв. Першим ключовим кроком у застосуванні групового методу аналізу ієрархій є побудова матриць парних порівнянь для кожного з обраних критеріїв кожною групою експертів. Саме на цьому етапі фахівці з відповідної галузі – у нашому випадку, лінгвісти – здійснюють попарне зіставлення альтернативних мовних моделей за конкретним критерієм, формуючи основу для подальших обчислень ваг. Продемонструємо цей процес на прикладі критерію «розуміння контексту», який є надзвичайно важливим у лінгвістичних задачах, зокрема при аналізі текстової цілісності, зв'язності, інтерпретації анафоричних елементів або багатозначних конструкцій.

Для оцінювання було залучено групу експертів-лінгвістів, які здійснили порівняння 16 великих мовних моделей, визначаючи для кожної пари, яка модель краще справляється із завданням збереження й розгортання контексту. Оцінки надавалися за дев'ятибальною шкалою Сааті, де 1 означає рівнозначну ефективність, 3 – помірну перевагу, 5 – істотну, 7 – сильну, 9 – абсолютну перевагу однієї моделі над іншою. Усі оцінки фіксувалися в матриці розміром 16×16, де для кожної пари моделей було записано відповідну експертну перевагу. Аналогічно було побудовано матриці і для інших двох груп експертів. Для наглядніщого подання інформації пронумеруємо моделі таким чином:

1. GPT-4o	9. Llama 3.1
2. GPT-4.5	10. Nemotron-4 340B
3. DeepSeek R1	11. Qwen2.5
4. Gemini	12. PALM 2.
5. Grok-3	13. LaMDA 3
6. Claude 3	14. Megatron-Turing NLG 530B
7. Mistral 7B	15. BLOOM
8. Phi-2	16. LLaMA 2

Таблиця 3

Матриця попарних порівнянь за оцінкою експертів-лінгвістів

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	1	1	5	1	1	1	7	7	5	3	5	3	3	3	7	7
2	1	1	5	1	1	1	7	7	5	3	5	3	3	3	7	7
3	1/5	1/5	1	1/5	1/5	1/5	3	5	1	1/3	1	1/3	1/3	1/3	3	5
4	1	1	5	1	1	1	7	7	5	3	5	3	3	3	7	7
5	1	1	5	1	1	1	7	7	5	3	5	3	3	3	7	7
6	1	1	5	1	1	1	7	7	5	3	5	3	3	3	7	7
7	1/7	1/7	1/3	1/7	1/7	1/7	1	3	1/3	1/5	1/3	1/5	1/5	1/5	1	3
8	1/7	1/7	1/5	1/7	1/7	1/7	1/3	1	1/5	1/7	1/5	1/7	1/7	1/7	1/3	1
9	1/5	1/5	1	1/5	1/5	1/5	3	5	1	1/3	1	1/3	1/3	1/3	3	5
10	1/3	1/3	3	1/3	1/3	1/3	5	7	3	1	3	1	1	1	5	7
11	1/5	1/5	1	1/5	1/5	1/5	3	5	1	1/3	1	1/3	1/3	1/3	3	5
12	1/3	1/3	3	1/3	1/3	1/3	5	7	3	1	3	1	1	1	5	7
13	1/3	1/3	3	1/3	1/3	1/3	5	7	3	1	3	1	1	1	5	7
14	1/3	1/3	3	1/3	1/3	1/3	5	7	3	1	3	1	1	1	5	7
15	1/7	1/7	1/3	1/7	1/7	1/7	1	3	1/3	1/5	1/3	1/5	1/5	1/5	1	3
16	1/7	1/7	1/5	1/7	1/7	1/7	1/3	1	1/5	1/7	1/5	1/7	1/7	1/7	1/3	1

Наведемо матриці парних порівнянь для критерію «Розуміння контексту».

Агрегована матриця. Після завершення побудови трьох окремих матриць парних порівнянь для кожного з критеріїв – «розуміння контексту», «лексичне та стилістичне багатство» та «розпізнавання граматичних помилок» – наступним етапом є агрегація оцінок для отримання єдиної узагальненої матриці. Оскільки ми маємо справу з груповим методом AHP і використовуємо оцінки кількох експертів (у даному випадку – групи лінгвістів), усі індивідуальні матриці по кожному критерію об'єднуються шляхом обчислення середнього геометричного значень відповідних елементів.

Ваги моделей. Після побудови агрегованої матриці парних порівнянь для кожного критерію наступним кроком є визначення вектора локальних ваг, який відображає відносну пріоритетність кожної з альтернатив (у нашому випадку – мовних моделей) за цим критерієм. Найбільш точним і теоретично обґрунтованим способом для цього є метод головного власного вектора.

Фінальна матриця ваг моделей за критеріями. Після завершення обчислення всіх локальних векторів ваг за кожним з 15 критеріїв, отриманих у межах окремих матриць парних порівнянь, переходимо до етапу агрегації результатів. На цьому кроці формується фінальна матриця, в якій кожен

Таблиця 4

Матриця попарних порівнянь за оцінкою ІТ-фахівців

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	1	1	5	3	3	1	7	7	5	5	5	5	5	5	7	7
2	1	1	5	3	3	1	7	7	5	5	5	5	5	5	7	7
3	1/5	1/5	1	1	1	1/3	3	5	1	1	1	1	1	1	3	5
4	1/3	1/3	1	1	1	1/3	3	5	3	1	1	1	1	1	3	5
5	1/3	1/3	1	1	1	1/3	3	5	3	1	1	1	1	1	3	5
6	1	1	3	3	3	1	5	5	5	5	5	5	5	5	7	7
7	1/7	1/7	1/3	1/3	1/3	1/5	1	3	1/3	1/3	1/3	1/3	1/3	1/3	1	3
8	1/7	1/7	1/5	1/5	1/5	1/5	1/3	1	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1
9	1/5	1/5	1	1/3	1/3	1/5	3	5	1	1	1	1	1	1	3	5
10	1/5	1/5	1	1	1	1/5	3	5	1	1	1	1	1	1	3	5
11	1/5	1/5	1	1	1	1/5	3	3	1	1	1	1	1	1	3	5
12	1/5	1/5	1	1	1	1/5	3	3	1	1	1	1	1	1	3	5
13	1/5	1/5	1	1	1	1/5	3	3	1	1	1	1	1	1	3	5
14	1/5	1/5	1	1	1	1/5	3	3	1	1	1	1	1	1	3	5
15	1/7	1/7	1/3	1/3	1/3	1/7	1	3	1/3	1/3	1/3	1/3	1/3	1/3	1	3
16	1/7	1/7	1/5	1/5	1/5	1/7	1/3	1	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1

Таблиця 5

Матриця попарних порівнянь за оцінкою експертів-лінгвістів

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	1	1	5	3	3	1	7	7	5	5	5	3	3	3	7	7
2	1	1	5	3	3	1	7	7	5	5	5	3	3	3	7	7
3	1/5	1/5	1	1	1	1/3	3	5	1	3	1	1	1	1	3	5
4	1/3	1/3	1	1	1	1/3	3	5	3	1	1	3	3	1	3	5
5	1/3	1/3	1	1	1	1/3	3	5	3	1	1	3	3	1	3	5
6	1	1	3	3	3	1	5	5	5	5	5	3	3	3	7	7
7	1/7	1/7	1/3	1/3	1/3	1/5	1	3	1/3	1/3	1/3	1/3	1/3	1/3	1	3
8	1/7	1/7	1/5	1/5	1/5	1/5	1/3	1	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1
9	1/5	1/5	1	1/3	1/3	1/5	3	5	1	1	1	1	1	1	3	5
10	1/5	1/5	1/3	1	1	1/5	3	5	1	1	1	1	1	3	3	5
11	1/5	1/5	1	1	1	1/5	3	3	1	1	1	3	3	1	3	5
12	1/3	1/3	1	1/3	1/3	1/3	3	3	1	1	1/3	1	3	1	3	5
13	1/3	1/3	1	1/3	1/3	1/3	3	3	1	1	1/3	1/3	1	1	3	5
14	1/3	1/3	1	1	1	1/3	3	3	1	1/3	1	1	1	1	3	5
15	1/7	1/7	1/3	1/3	1/3	1/7	1	3	1/3	1/3	1/3	1/3	1/3	1/3	1	3
16	1/7	1/7	1/5	1/5	1/5	1/7	1/3	1	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1

Таблиця 6

Агрегована матриця попарних порівнянь

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	1	4.983	1	1	1	6.977	6.977	4.983	2.990	4.983	2.990	2.990	2.990	6.977	6.977
2	1	1	4.983	1	1	1	6.977	6.977	4.983	2.990	4.983	2.990	2.990	2.990	6.977	6.977
3	0.200	0.200	1	0.200	0.200	0.200	2.990	4.983	1	0.333	1	0.333	0.333	0.333	2.990	4.983
4	1	1	4.983	1	1	1	6.977	6.977	4.983	2.990	4.983	2.990	2.990	2.990	6.977	6.977
5	1	1	4.983	1	1	1	6.977	6.977	4.983	2.990	4.983	2.990	2.990	2.990	6.977	6.977
6	1	1	4.983	1	1	1	6.977	6.977	4.983	2.990	4.983	2.990	2.990	2.990	6.977	6.977
7	0.143	0.143	0.334	0.143	0.143	0.143	1	2.990	0.334	0.200	0.334	0.200	0.200	0.200	1	2.990
8	0.143	0.143	0.200	0.143	0.143	0.143	0.334	1	0.200	0.143	0.200	0.143	0.143	0.143	0.334	1
9	0.200	0.200	1	0.200	0.200	0.200	2.990	4.983	1	0.333	1	0.333	0.333	0.333	2.990	4.983
10	0.334	0.334	2.990	0.334	0.334	0.334	4.983	6.977	2.990	1	2.990	1	1	1	4.983	6.977
11	0.200	0.200	1	0.200	0.200	0.200	2.990	4.983	1	0.334	1	0.333	0.333	0.333	2.990	4.983
12	0.334	0.334	2.990	0.334	0.334	0.334	4.983	6.977	2.990	1	2.990	1	1	1	4.983	6.977
13	0.334	0.334	2.990	0.334	0.334	0.334	4.983	6.977	2.990	1	2.990	1	1	1	4.983	6.977
14	0.334	0.334	2.990	0.334	0.334	0.334	4.983	6.977	2.990	1	2.990	1	1	1	4.983	6.977
15	0.143	0.143	0.334	0.143	0.143	0.143	1	2.990	0.334	0.200	0.334	0.200	0.200	0.200	1	2.990
16	0.143	0.143	0.200	0.143	0.143	0.143	0.334	1	0.200	0.143	0.200	0.143	0.143	0.143	0.334	1

Таблиця 7

Вектор ваг моделей за критерієм «Розуміння контексту»

№	Модель	Групова вага
1	GPT-4o	0.125518
2	GPT-4.5	0.125518
3	Gemini	0.125518
4	Grok-3	0.125518
5	Claude 3	0.125518
6	Nemotron-4 340B	0.059010
7	PaLM 2	0.059010
8	LaMDA 3	0.059010
9	Megatron-Turing NLG 530B	0.059010
10	DeepSeek R1	0.028574
11	Llama 3.1	0.028574
12	Qwen2.5	0.028574
13	Mistral 7B	0.015200
14	BLOOM	0.015200
15	Phi-2	0.010124
16	LLaMA 2	0.010124

стовпець відповідає окремому критерію, а кожен рядок – конкретній альтернативі (тобто мовній моделі). Усі значення в цій матриці – це локальні ваги моделей за відповідним критерієм, які були попередньо розраховані методом головного власного вектора.

Наступним кроком є побудова фінального вектора глобальних ваг альтернатив. Для цього до кожного стовпця (тобто до ваг альтернатив за конкретним критерієм) застосовується зважування відповідно до важливості самого критерію, яка була визначена на попередньому рівні ієрархії (тобто у векторі ваг критеріїв). Іншими словами, кожен локальний вектор ваг множиться на вагу відповідного критерію, після чого здійснюється підсумовування по рядках – тобто обчислюється зважена сума для кожної альтернативи.

Результатом цього процесу є фінальний глобальний вектор пріоритетів, який відображає загальну ефективність кожної великої мовної моделі з урахуванням усіх 15 критеріїв одночасно. Саме ці значення використовуються для ранжування альтернатив і прийняття остаточного рішення щодо вибору моделі, яка найкраще відповідає поставленим лінгвістичним цілям. Для зручності використаємо застосовану вище нумерацію великих мовних моделей, а також пронумеруємо критерії:

1. Розуміння контексту	9. Обсяг пам’яті
2. Якість генерації	10.Масштабованість
3. Корекція граматики	11. Інтеграція/API
4. Врахування знань/фактів	12. Вартість
5. Лексичне багатство	13. Дата оновлення
6. Багатомовність	14. Апаратні вимоги
7. Адаптація до завдань	15. Швидкість генерації
8. Безпека	

Згідно з результатами, найвищі позиції у підсумковому рейтингу посіли Claude 3, GPT-4o та GPT-4.5, що свідчить про високу узгодженість оцінок між трьома групами експертів – лінгвістами, науковими працівниками та ІТ-фахівцями – щодо якості цих моделей.

Модель Claude 3 була визнана найефективнішою передусім завдяки своїй відомій здатності генерувати логічно узгоджений, стилістично багатий та контекстуально точний текст. Експерти-

Таблиця 8

Фінальна агрегована матриця парних порівнянь моделей

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.126	0.142	0.169	0.159	0.159	0.092	0.085	0.089	0.085	0.086	0.086	0.036	0.087	0.032	0.032
2	0.126	0.142	0.169	0.159	0.159	0.092	0.085	0.088	0.087	0.088	0.085	0.036	0.084	0.032	0.032
3	0.029	0.050	0.061	0.054	0.054	0.072	0.076	0.082	0.072	0.067	0.077	0.050	0.071	0.060	0.060
4	0.126	0.056	0.063	0.066	0.066	0.083	0.076	0.079	0.080	0.077	0.076	0.047	0.078	0.058	0.058
5	0.126	0.056	0.063	0.066	0.066	0.072	0.067	0.062	0.050	0.060	0.066	0.051	0.051	0.059	0.059
6	0.126	0.138	0.158	0.149	0.149	0.083	0.076	0.083	0.080	0.086	0.078	0.042	0.086	0.081	0.081
7	0.015	0.021	0.020	0.020	0.020	0.032	0.039	0.045	0.039	0.047	0.052	0.084	0.052	0.089	0.089
8	0.010	0.011	0.013	0.014	0.014	0.032	0.039	0.044	0.039	0.046	0.051	0.084	0.054	0.087	0.088
9	0.029	0.042	0.041	0.042	0.042	0.061	0.058	0.060	0.059	0.079	0.046	0.078	0.079	0.062	0.062
10	0.059	0.042	0.048	0.050	0.050	0.061	0.058	0.060	0.059	0.072	0.046	0.077	0.065	0.070	0.069
11	0.029	0.046	0.044	0.056	0.056	0.052	0.058	0.055	0.053	0.043	0.045	0.070	0.079	0.085	0.086
12	0.059	0.045	0.041	0.047	0.047	0.052	0.058	0.054	0.053	0.042	0.045	0.069	0.064	0.072	0.071
13	0.059	0.043	0.035	0.040	0.040	0.052	0.058	0.053	0.052	0.041	0.040	0.068	0.061	0.071	0.070
14	0.059	0.043	0.042	0.047	0.047	0.042	0.049	0.040	0.092	0.044	0.039	0.058	0.048	0.060	0.060
15	0.015	0.021	0.019	0.020	0.020	0.042	0.049	0.034	0.038	0.041	0.033	0.092	0.046	0.090	0.090
16	0.010	0.008	0.012	0.012	0.012	0.021	0.029	0.024	0.030	0.066	0.028	0.094	0.046	0.092	0.091

Таблиця 9

Фінальний вектор ваг моделей

№	Модель	Глобальна вага
1	Claude 3	0.080878
2	GPT-4o	0.080368
3	GPT-4.5	0.079751
4	Gemini	0.071257
5	DeepSeek R1	0.066910
6	Llama 3.1	0.064397
7	Nemotron-4 340B	0.063148
8	Qwen2.5	0.061516
9	Grok-3	0.060414
10	PaLM 2	0.057583
11	LaMDA 3	0.056168
12	BLOOM	0.052256
13	Megatron-Turing NLG 530B	0.051330
14	Mistral 7B	0.051305
15	Phi-2	0.050585
16	LLaMA 2	0.046557

лінгвісти особливо відзначили її чутливість до дискурсивних маркерів, гнучкість у стилістичних трансформаціях, а також уважність до граматичних деталей. Крім того, наукові працівники високо оцінили здатність Claude 3 дотримуватись фактологічної достовірності, що є критично важливим у завданнях академічного письма та інтерпретації текстів.

GPT-4o продемонстрував надзвичайну контекстуальну глибину та широку функціональність у лінгвістичних застосуваннях, зокрема у мультимодальних інтерфейсах і роботі з багатомовним контентом. Його перевага полягала не лише в точності мовного аналізу, а й у високій швидкості генерації, адаптивності та стабільності інтеграції – чинниках, які активно підтримувались з боку IT-фахівців. Лінгвісти, зі свого боку, відзначили силу GPT-4o у розпізнаванні коференції, побудові аргументативних структур та узгодженні стилістичних регістрів.

GPT-4.5 посів третю позицію, що підтверджує його баланс між якістю генерації та технічною ефективністю. Він дещо поступається GPT-4o в мультимодальних аспектах, однак демонструє стабільно високі результати у завданнях синтаксичного аналізу, реферування та стилістичного перефразування. Особливу увагу звернули на нього ті експерти, хто працює з навчальними текстами, корпусами та автоматизованим оцінюванням студентських робіт.

Висновки. Проведене дослідження засвідчило зростаючу значущість великих мовних моделей у сучасному лінгвістичному дискурсі та дало змогу системно проаналізувати їх ефективність у вирішенні складних філологічних завдань. Вперше для такого аналізу було застосовано груповий метод аналізу ієрархій, який дозволив не лише зіставити технічні характеристики моделей, а й інтегрувати експертні оцінки представників різних дисциплін: лінгвістів, гуманітаріїв і IT-фахівців. Це забезпечило об'єктивність результатів і дозволило сформувати зважену багатовимірну модель оцінювання.

Результати показали, що найбільш критичними характеристиками LLM у контексті лінгвістичних завдань є здатність моделі до глибокого розуміння контексту, якість текстової генерації, точність відтворення фактологічної інформації та коректність граматичних конструкцій. Саме ці параметри отримали найвищі ваги в експертній ієрархії, що цілком відповідає природі філологічної роботи, де мовна чутливість і змістова достовірність є визна-

чальними. Натомість технічні аспекти, такі як апаратні вимоги, швидкість генерації або можливість API-інтеграції, хоча й мають значення для практичного використання, виявилися вторинними з точки зору гуманітарних задач.

Аналіз моделей, серед яких були GPT-4o, Claude 3, Gemini, DeepSeek R1, Mistral 7B, Grok-3, LLaMA 3.1 та інші, дозволив встановити чіткий профіль кожної з них у контексті мовної ефективності. Наприклад, GPT-4o вирізняється високим рівнем мультимодальності, блискавичною швидкістю реакції та контекстною точністю, що робить її надзвичайно зручною для інтерактивних освітніх і дослідницьких середовищ. Claude 3, у свою чергу, продемонструвала виняткову якість у задачах академічного письма, стилістичного аналізу та інтерпретації складних дискурсивних структур. Серед відкритих моделей, DeepSeek R1 та Mistral 7B показали хороший баланс між продуктивністю та доступністю, а Phi-2 – високу ефективність при мінімальних обчислювальних потребах, що робить її придатною для використання в локальних освітніх середовищах.

Значущим результатом дослідження стало підтвердження цінності поєднання LLM із онтологічними структурами знань. Таке інтегрування не лише зменшує ризики семантичних викривлень, а й забезпечує більшу структурованість та узгодженість при виконанні складних лінгвістичних завдань, зокрема семантичного аналізу, стилометрії, жанрової класифікації або інтертекстуальних досліджень. У цьому контексті великі мовні моделі постають не як конкуренти класичної лінгвістики, а як інтелектуальні інструменти її поглиблення й розвитку.

Не менш важливою є роль міждисциплінарного підходу до оцінювання якості LLM. Метод Group AHP дозволив гармонізувати судження представників трьох ключових груп експертів – лінгвістів, гуманітаріїв та інженерів – і сформувати спільну систему пріоритетів. Такий підхід забезпечив глибоку методологічну обґрунтованість результатів, дав змогу врахувати як мовну, так і аналітичну та технічну складові, що особливо важливо в умовах зростаючої інтеграції штучного інтелекту в гуманітарний простір.

Практична цінність дослідження полягає в тому, що отримані результати можуть бути безпосередньо використані у філологічній освіті, наукових дослідженнях, редакційній діяльності, а також у створенні цифрових архівів і гуманітарних інформаційних систем. Вони дозволяють не лише обґрунтовано обрати ту чи іншу модель для конкретного завдання, але й визначити її слабкі сторони та можливості для донавчання або вдосконалення. У цьому сенсі запропонована модель оцінювання перетворюється на інструмент не лише аналітики, а й стратегічного планування розвитку ШІ в гуманітарних дисциплінах.

Підсумовуючи, варто наголосити, що великі мовні моделі вже сьогодні здатні не лише автоматизувати рутинні лінгвістичні процедури, а й відкривати нові горизонти для філологічного аналізу, міжмовних зіставлень, літературознавчої інтерпретації та лексикографічної розвідки. Водночас їх ефективне й відповідальне використання потребує чіткого методологічного підґрунтя, міждисциплінарного підходу та системи контролю якості. Саме таке поєднання – технічної потужності, мовної чутливості й гуманітарної обачності – є запорукою успішної інтеграції LLM у сучасну філологію.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Пасічник В. В., Яромич М. В. Великі мовні моделі та онтології у філологічних дослідженнях: аналітичний огляд джерел. *Актуальні питання гуманітарних наук. Мовознавство. Літературознавство*. 2025. Вип. 83, № 3. С. 236–250. DOI: https://doi.org/10.24919/2308-4863/83-3-35.

2. Кунанець Н. Е., Яромич М. В. Виділення концептів у літературних текстах із використанням великих мовних моделей. *Вісник науки та освіти*. 2025. Вип. 32, № 2. С. 343–357. DOI: https://doi.org/10.52058/2786-6165-2025-2(32)-343-357.

3. Пасічник В. В., Яромич М. В. Автоматизоване формування технічної документації в галузі ІТ з використанням великих мовних моделей. *Studia Methodologica*. 2025. Вип. 59. С. 250–273. DOI: https://doi.org/10.32782/2307-1222.2025-59-22.

4. Пасічник В. В., Яромич М. В. Жанрова класифікація літератури за метриками за допомогою великих мовних моделей. *Наукові праці Міжрегіональної академії управління персоналом. Філологія*. 2025. Вип. 1, № 15. С. 60–68. DOI: https://doi.org/10.32689/maup.philol.2025.1.11.

5. Пасічник В. В., Яромич М. В. Особливості жанрової класифікації літератури за допомогою великих мовних моделей. 2025. *Folium*. Вип. 6. С. 132–144.

6. Anthropic Team. The Claude 3 Model Family. URL: https://www.anthropic.com/claude-3-model-card.

7. BigScience & HuggingFace. BLOOM – BigScience Large Open Multilingual Model. URL: https://bigscience.huggingface.co/blog/bloom.

8. DeepSeek-AI, Zhu Q., Guo D. DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence. 2024. DOI: https://doi.org/10.48550/arXiv.2406.11931.

9. Giglou H.B., D'Souza J., Auer S. LLMs4OL: Large Language Models for Ontology Learning. *The Semantic Web – ISWC*. 2023. P. 408–427. DOI: 10.1007/978-3-031-47240-4_22.

10. Georgiev P., Lei V. I., Burnell R.. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. 2024. DOI: https://doi.org/10.48550/arXiv.2403.05530.

11. LaMDA: Our breakthrough conversation technology. URL: https://blog.google/technology/ai/lamda/.

12. Introducing PaLM 2. URL: https://blog.google/technology/ai/google-palm-2-ai-large-language-model/.

13. Touvron H. ,Martim L., Stone K. Llama 2: Open Foundation and Fine-Tuned Chat Models. 2023. DOI: https://doi.org/10.48550/arXiv.2307.09288.

14. Phi-2: The surprising power of small language models. URL: https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/.

15. Mu Y., Dong C., Bontcheva K., Song X. Large Language Models Offer an Alternative to the Traditional Approach of Topic Modelling. 2024. DOI: https://doi.org/10.48550/arXiv.2403.16248.

16. Mukanova A., Milosz M., Dauletkaliyeva A. LLM-powered natural language text processing for ontology enrichment. *Applied Sciences*. 2024. Vol. 14. No 13. P. 5860–5875. DOI: 10.3390/app14135860.

17. Smith S., Patwary M., Morick B. Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model. 2022. DOI: https://doi.org/10.48550/arXiv.2201.11990.

18. Nemotron-4 340B Technical Report. URL: https://arxiv.org/html/2406.11704v1.

19. Neuhaus F. Ontologies in the era of large language models – a perspective. Applied Ontologyю. 2024. Vol. 18. No 4. P. 399–407. DOI: 10.3233/AO-230072.

20. Opara C. StyloAI: Distinguishing AI-Generated Content with Stylometric Analysis. *25th International Conference on Artificial on Artificial Intelligence in Education*. 2024. arXiv.2405.10129. DOI: 10.48550/arXiv.2405.10129.

21. Introducing GPT-4o and more tools to ChatGPT free users. URL: https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/.

22. OpenAI Platform. URL: https://platform.openai.com/docs/api-reference/introduction.

23. Qwen2-0.5B-Instruct. URL: https://huggingface.co/Qwen/Qwen2-0.5B-Instruct/blob/main/README.md.

24. Саатi Т. Л. Прийняття групових рішень та аналітичний ієрархічний процес. *European Journal of Operational Research*. 1989. Т. 48, № 1. С. 9–26.

25. Saaty T. L. The Analytic Hierarchy Process. New York : McGraw-Hill, 1980. 287 c.

26. Саатi Т. Л. Прийняття рішень за допомогою аналітичного ієрархічного процесу *International Journal of Services Sciences*. 2008. Т. 1, № 1. С. 83–98.

27. Schreiber H. Genre Ontology Learning: Comparing Curated with Crowd-Sourced Folksonomies. *Proceedings of the 17th International Society for Music Information Retrieval Conference*. New York City, USA. Duration: August 7-11, 2016. P. 400–406. URL: https://archives.ismir.net/ismir2016/paper/000074.pdf

28. Tkachenko O. I., Tkachenko K. O., Tkachenko O. A. Linguistic Ontologies: Designing and Using in the Educational Intellectual Systems. *Digital Platform Information Technologies in Sociocultural Sphere*. 2021. Vol. 4. № 1. С. 97–111. DOI: 10.31866/2617-796X.4.1.2021.236950.

29. Umphrey R., Roberts J., Roberts L. Investigating Expert-in-the-Loop LLM Discourse Patterns for Ancient Intertextual Analysis. *The 4th International Conference on Natural Language Processing for Digital Humanities (NLP4DH 2024)*. Miami, USA. Duration: Nov 16 2024. USA, 31–41. URL: https://arxiv.org/abs/2409.01882.

30. Grok. URL: https://x.ai/grok.

31. Mistral 7B. URL: https://mistral.ai/news/announcing-mistral-7b.

REFERENCES

1. Pasichnyk, V. V., & Yaromych, M. V. (2025). Velyki movni modeli ta ontolohii u filolohichnykh doslidzhenniakh: analitychnyi ohliad dzherel [Large language models and ontologies in philological research: An analytical review of sources]. *Aktualni pytannia humanitarnykh nauk. Movoznavstvo. Literaturoznavstvo* Vol. 83. №. 3. P. 236–250. https://doi.org/10.24919/2308-4863/83-3-35

2. Kunanets, N. E., & Yaromych, M. V. (2025). Vydilennia kontseptiv u literaturnykh tekstakh iz vykorystanniam velykykh movnykh modelei [Extracting concepts from literary texts using large language models]. *Visnyk nauky ta osvity* Vol. 32. №. 2. P. 343–357. https://doi.org/10.52058/2786-6165-2025-2(32)-343-357

3. Pasichnyk, V. V., & Yaromych, M. V. (2025). Avtomatyzovane formuvannia tekhnichnoi dokumentatsii v haluzi IT z vykorystanniam velykykh movnykh modelei [Automated generation of technical documentation in the IT field using large language models]. *Studia Methodologica*. Vol. 59. P. 250–273. https://doi.org/10.32782/2307-1222.2025-59-22

4. Pasichnyk, V. V., & Yaromych, M. V. (2025). Zhanrova klasyfikatsiia literatury za metrykamy za dopomohoiu velykykh movnykh modelei [Genre classification of literature based on metrics using large language models]. *Naukovi pratsi Mizhrieghionalnoi akademii upravlinnia personalom. Filolohiia* Vol. 1. №.15. P. 60–68. https://doi.org/10.32689/maup.philol.2025.1.11

5. Pasichnyk, V. V., & Yaromych, M. V. (2025). Osoblyvosti zhanrovoi klasyfikatsii literatury za dopomohoiu velykykh movnykh modelei [Features of genre classification of literature using large language models]. *Folium.* Vol. 6. P. 132–144.

6. Anthropic Team. The Claude 3 Model Family. Website. Retrieved from: https://www.anthropic.com/claude-3-model-card.

7. BigScience & HuggingFace. BLOOM – BigScience Large Open Multilingual Model. Website. Retrieved from: https://bigscience.huggingface.co/blog/bloom.

8. DeepSeek-AI, Zhu Q., Guo D. DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence (2024). DOI: https://doi.org/10.48550/arXiv.2406.11931.

9. Giglou, H.B., D'Souza, J., & Auer, S. (2023). LLMs4OL: Large Language Models for Ontology Learning. The Semantic Web – ISWC. P. 408–427. DOI: 10.1007/978-3-031-47240-4_22.

10. Georgiev P., Lei V. I., Burnell R.. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context (2024). DOI: https://doi.org/10.48550/arXiv.2403.05530.

11. LaMDA: Our breakthrough conversation technology. Website. Retrieved from: https://blog.google/technology/ai/lamda/.

12. Introducing PaLM 2. Website. Retrieved from: https://blog.google/technology/ai/google-palm-2-ai-large-language-model/.

13. Touvron H., Martim L., Stone K. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. DOI: https://doi.org/10.48550/arXiv.2307.09288.

14. Phi-2: The surprising power of small language models. Website. Retrieved from: https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/.

15. Mu Y., Dong C., Bontcheva K., Song X. (2024). Large Language Models Offer an Alternative to the Traditional Approach of Topic Modelling. DOI: https://doi.org/10.48550/arXiv.2403.16248.

16. Mukanova A., Milosz M., Dauletkaliyeva A. (2024). LLM-powered natural language text processing for ontology enrichment. Applied Sciences. Vol. 14. No 13. P. 5860–5875. DOI: 10.3390/app14135860.

17. Smith S., Patwary M., Morick B. (2022). Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model. DOI: https://doi.org/10.48550/arXiv.2201.11990.

18. Nemotron-4 340B Technical Report. Website. Retrieved from: https://arxiv.org/html/2406.11704v1.

19. Neuhaus F. (2024). Ontologies in the era of large language models – a perspective. *Applied Ontology*. Vol. 18. No 4. P. 399–407. DOI: 10.3233/AO-230072.

20. Opara C. (2024) StyloAI: Distinguishing AI-Generated Content with Stylometric Analysis. *25th International Conference on Artificial on Artificial Intelligence in Education*. arXiv.2405.10129. DOI: 10.48550/arXiv.2405.10129.

21. Introducing GPT-4o and more tools to ChatGPT free users. Website. Retrieved from: https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/.

22. OpenAI Platform. Website: https://platform.openai.com/docs/api-reference/introduction.

23. Qwen2-0.5B-Instruct. Website. Retrieved from: https://huggingface.co/Qwen/Qwen2-0.5B-Instruct/blob/main/README.md.

24. Saaty, T. L. (1989). Group decision making and the analytic hierarchy process. *European Journal of Operational Research.* Vol. 48. №. 1. P. 9–26.

25. Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill. 287 p.

26. Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences. V*ol. 1. №. 1. P. 83–98.

27. Schreiber H. Genre Ontology Learning: Comparing Curated with Crowd-Sourced Folksonomies. *Proceedings of the 17th International Society for Music Information Retrieval Conference*. New York City, USA. Duration: August 7-11, 2016. P. 400–406. URL: https://archives.ismir.net/ismir2016/paper/000074.pdf

28. Tkachenko O. I., Tkachenko K. O., Tkachenko O. A. (2021). Linguistic Ontologies: Designing and Using in the Educational Intellectual Systems. *Digital Platform Information Technologies in Sociocultural Sphere*. Vol. 4. № 1. С. 97–111. DOI: 10.31866/2617-796X.4.1.2021.236950.

29. Umphrey R., Roberts J., Roberts L. (2024) Investigating Expert-in-the-Loop LLM Discourse Patterns for Ancient Intertextual Analysis. *The 4th International Conference on Natural Language Processing for Digital Humanities (NLP4DH 2024)*. Miami, USA. Duration: Nov 16 2024. USA, 31–41. URL: https://arxiv.org/abs/2409.01882.

30. Grok. Website. Retrieved from: https://x.ai/grok.

31. Mistral 7B. Website. Retrieved from: https://mistral.ai/news/announcing-mistral-7b.