

ШТУЧНИЙ ІНТЕЛЕКТ ЯК ІННОВАЦІЙНИЙ ІНСТРУМЕНТ СОЦІАЛЬНИХ ЗМІН: ВИКЛИКИ, ПЕРСПЕКТИВИ, РИЗИКИ ТА СОЦІАЛЬНІ НАСЛІДКИ

Андрій Жорняк,

Центральне міжрегіональне управління Державної служби з питань праці

Анотація. У статті розглянуто та проаналізовано напрями досліджень у галузі штучного інтелекту. Визначено актуальні аспекти щодо обмеження та перспектив удосконалення, а також сучасні виклики щодо впровадження цієї технології в різних галузях. Підкреслено, що впровадження цієї технології може викликати перетворення в організаційних структурах різних сфер суспільства. Акцентовано, що застосування штучного інтелекту може бути ефективним інструментом для покращення добробуту суспільства, а також зазначено можливі загрози та ризики, пов'язані з його застосуванням.

Розглянуто особливості організації систем штучного інтелекту, зокрема архітектуру, принципи функціонування, алгоритми та моделі ухвалення рішень, на прикладі ChatGPT. Виявлено нерівномірність у практичному впровадженні цієї технології. Вказано на брак уніфікованого стандарту впровадження систем штучного інтелекту в різні сфери суспільного життя.

У статті наведено пропозиції щодо подальших досліджень у сфері впровадження систем штучного інтелекту.

Ключові слова: штучний інтелект; інтелектуальні комп'ютерні системи; кіберпростір; інтелектуальні алгоритми; ChatGPT; кібербезпека.

ВСТУП

Штучний інтелект (далі — ШІ) є однією з перспективних інновацій сучасності. Маючи значний потенціал, ця технологія здатна спричинити кардинальні зміни та перетворення в різних сферах суспільного життя. Розвиток та впровадження ШІ відбувається експоненціально швидко та має як позитивні наслідки, так і можливі потенційні ризики та загрози. Важливо відзначити, що на міжнародному рівні ще немає єдиного погляду на питання, яким може бути набір правил щодо регулювання ШІ.

ШІ заслуговує пильної уваги. Спостерігається збільшений інтерес науковців, дослідників та практиків різних галузей до ШІ. Упровадження та застосування цієї технології вже привело до змін у різних сферах суспільного життя. Питання, пов'язані з розвитком ШІ, виходять за межі традиційної наукової діяльності, ставлячи нові виклики,

продиктовані необхідністю розуміти механізми процесів навчання, сенсорного сприйняття, природу мови та механізми дії взаємопов'язаних нейронів. Зараз питання полягає не лише в тому, як автоматизувати рутинну роботу людини та велику кількість операцій, але й у тому, як за допомогою технологій ШІ наділити інтелектом машину-робота.

Науковцями, які займаються розробкою ШІ, одночасно досліджується процес функціонування людського розуму.

ШІ розглядається науковою спільнотою як інтелектуальні комп'ютерні системи, які здатні інтерпретувати складні завдання й самостійно розробляти відповіді на рівні, достатньо близькому до мисленневих психологічних здібностей людини, а також виконувати творчі функції, які притаманні людині.

Дослідження ШІ повинні бути спрямовані на забезпечення загального добробуту суспільства, збереження соціальної стабіль-

ності та безпеки, цілісності економічних систем, а також захисту прав і законних інтересів громадян, охоплюючи аспекти, які мають вплив як на внутрішні (національні), так і на міжнародні сфери.

Мета дослідження — дослідити вплив штучного інтелекту (ШІ) як інноваційного інструмента на соціальні зміни, аналізуючи його виклики, перспективи, ризики та соціальні наслідки.

Завдання для реалізації мети:

- дослідити принципи організації та напрями розвитку ШІ;
- проаналізувати алгоритми та моделі прийняття рішень ШІ на прикладі ChatGPT;
- визначити обмеження, перспективи та можливості впровадження ШІ в різних галузях;
- надати пропозиції щодо подальших досліджень у сфері впровадження систем ШІ.

МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

Дослідження принципів організації ШІ, процесу та механізмів прийняття рішень на прикладі моделі ChatGPT, обмежень у роботі та перспектив розвитку проводилось із застосуванням таких методів: теоретичний аналіз (вивчення наукових публікацій, що стосуються ШІ та моделей прийняття рішень технології ChatGPT); порівняльний аналіз (порівняння різних підходів та алгоритмів у сфері ШІ, визначення переваг та недоліків ChatGPT); емпіричний аналіз (оцінка ефективності ШІ у вирішенні різних завдань, оцінка якості алгоритмів та моделей прийняття рішень ChatGPT, дослідження конкретних випадків використання ChatGPT в різних галузях); статистичний аналіз (аналіз показників ефективності: точності, швидкості відповіді, кількості помилок тощо); оцінка ризиків та перспектив (аналіз та оцінка сильних і слабких сторін ШІ, можливостей та загроз, перспектив розвитку ШІ, виявлення потенційних ризиків); прогнозування (оцінка майбутніх тенденцій у розвитку ШІ, аналіз наявних досліджень, інновацій та потенційних напрямів розвитку тощо).

АНАЛІЗ ПУБЛІКАЦІЙ

Дослідження практичного застосування ШІ активно проводиться як українськими, так і зарубіжними науковцями. В Україні різні аспекти впровадження систем ШІ у різних галузях аналізуються такими дослідниками, як В. Биков, О. Спірін, Т. Вакалюк, С. Литвинова, О. Овчарук, О. Пінчук, О. Соколюк та іншими. Значна увага приділяється методологічним підходам, що висвітлюються в працях М. Мар'єнко, С. Семерікова, О. Маркової. Серед актуальних проблем, що виявляються в дослідженнях, особлива увага приділяється необхідності підвищення професійної компетенції користувачів і працівників для ефективного застосування ШІ. Теоретичне обґрунтування можливостей та перспектив ШІ представлено в роботах В. Коваленка, М. Мар'єнка. Важливим напрямом є уточнення понятійного апарату, що включає осмислення й удосконалення визначень базових термінів. Дослідження демонструють потенціал ШІ в різних галузях, акцентуючи увагу на перспективах автоматизації виробничих процесів. У своїх працях Д. Лубко та С. Шаров детально розглядають теоретичні, практичні й методологічні основи проектування систем ШІ. Особливий інтерес викликають методи розробки, впровадження та застосування таких систем, орієнтованих на обробку даних за допомогою інтелектуальних алгоритмів. Хоча ШІ є відносно новою технологією, зарубіжні вчені, серед яких Е. Постма, А. Нанне, М. Антеуніс, С. Вуббен, Г. ван Ноорт, Г. Аль Муршиді, А. Капуза, Т. Хагендорф, Т. Адігузед та інші, також активно досліджують цю сферу. Результати їхніх досліджень підтверджують значний потенціал ШІ, водночас демонструючи необхідність врахування викликів і ризиків, пов'язаних із його застосуванням. Зокрема, актуальними залишаються питання забезпечення відповідального використання ШІ. Дослідники акцентують увагу на важливості досягнення балансу між використанням переваг ШІ та запобіганням потенційним ризикам, забезпечення людського контролю й збереження фундаментальних людських цінностей.

РЕЗУЛЬТАТИ Й ОБГОВОРЕННЯ ДОСЛІДЖЕННЯ

Дослідження принципів організації та напрямів розвитку ШІ. Сучасність характеризується глобалізаційними та інтеграційними процесами, потужність яких зростає під значним впливом інноваційних ІТ-технологій. Одним із ключових аспектів сьогодення є впровадження ШІ в різні сфери суспільного життя. «Досягнення безпеки та співробітництва в розколотому світі», «Створення зростання і робочих місць для нової ери», «Штучний інтелект як рушійна сила економіки та суспільства» та «Довгострокова стратегія клімату, природи та енергії» — основні дискусійні питання, порушені в межах події світового рівня (Давос-2024: Всесвітній економічний форум, Великобританія, 02.11.2023р.; далі — Форум), у якій узяли участь понад 100 представників урядів з усього світу, політичні діячі, представники основних міжнародних організацій, бізнесу, зокрема генеральний директор OpenAI Сем Олтмен, засновник корпорації Microsoft Білл Гейтс, а також власник SpaceX, Tesla Ілон Маск. Крім того, ця подія об'єднала лідерів громадянського суспільства, провідних експертів різноманітних галузей, молодь та ЗМІ (World Economic Forum, 2024).

Уряди 28 країн світу (Великої Британії, США, ЄС, Австралії, Китаю та ін.) за підсумками Форуму досягли домовленостей щодо спільної діяльності в галузі досліджень і в галузі безпеки ШІ, засвідчивши цей намір підписанням першої в історії міжнародної декларації (декларація Блетчлі), у якій зазначено, що ШІ може становити потенційно катастрофічний ризик для існування людства (Delo.ua, 2023).

Свою думку стосовно того, як ШІ змінить життя суспільства, на Форумі висловив засновник корпорації Microsoft Білл Гейтс, зауваживши, що «в найближчі п'ять років технологія штучного інтелекту стане революційною для всіх, оскільки полегшить життя кожної людини» (World Economic Forum, 2024). Прем'єр-міністр Британії Ріші Сунак зазначив: «Немає нічого більш перетворюваль-

ного для майбутнього наших дітей і онуків, ніж такі технологічні досягнення, як штучний інтелект. Однак ми зобов'язані забезпечити їм безпечний і відповідальний розвиток ШІ, усвідомлюючи ризики, які він становить, на ранніх стадіях процесу» (World Economic Forum, 2024). Заступник міністра науки і технологій КНР У. Чжаохуей також зазначив: «Ми підтримуємо принципи взаємної поваги, рівності та взаємної вигоди. Країни, незалежно від їхнього розміру і масштабу, мають рівні права на розробку і використання ШІ» (World Economic Forum, 2024). Водночас Ілон Маск зазначив: «Уперше ми зіткнулися із ситуацією, коли відбувається щось, що буде значно розумнішим за найрозумнішу людину. Мені не ясно, чи можемо ми насправді контролювати таку річ». На його думку, розвиток штучного інтелекту може призвести до загибелі людства. Він сказав: «Я думаю, є певна ймовірність, що він (ШІ) призведе до загибелі людства. Можливо, я погоджуся з оцінкою британсько-канадського вченого, першопрохідця в галузі штучного інтелекту Джеффри Хінтона, що вона становить приблизно 10-20% або щось близько цього» (World Economic Forum, 2024).

Брак уніфікованого підходу до розуміння сутності та особливостей використання ШІ створює сприятливі умови для зловживань із боку недобросовісних користувачів і кіберзлочинців. Необхідність дослідження процесів впровадження та застосування ШІ стає надзвичайно актуальною, оскільки ці процеси вже активно відбуваються в глобальному просторі. Пріоритетним завданням для держав, які прагнуть успішно інтегрувати та використовувати технології ШІ, є розробка та реалізація заходів, що спрямовані на забезпечення етичного використання цієї технології в межах чинного законодавства.

Наявні визначення інтелекту базуються на різних підходах. Визначено, що інтелект — це здатність вирішувати складні завдання; здатність до навчання, узагальнення та аналізу; здатність сприймати та зберігати інформацію для адаптивної поведінки в середовищі; здатність виконувати творчі функції, які зазвичай належать виключно до людських

здібностей, для інших інтелект може бути лише біологічним явищем (Karwowski, 2015). Різниця в поглядах та визначеннях зумовлює застосування науковцями та дослідниками ШІ різних підходів, серед яких такі: семіотичний підхід, орієнтований на створення експертних систем, що імітують високорівневі психологічні процеси; біологічний підхід, який зосереджується на вивченні нейронних мереж та еволюційних обчислень, що моделюють інтелектуальну поведінку на основі біологічних елементів (OpenAI, n.d.).

У 1956 р. в Дартмутському університеті відбувся науковий семінар, який зібрав провідних фахівців у таких галузях, як теорія управління, автоматизація, нейронні мережі, теорія ігор та дослідження ШІ. Основною метою заходу було дослідження можливості моделювання процесів міркування, інтелектуальної діяльності та творчості за допомогою обчислювальних пристроїв (History of Data Science, n.d.). Учасники ґрунтовно обговорювали різноманітні аспекти концепції ШІ. Зокрема, Джон Маккарті запропонував нове визначення терміна «штучний інтелект», яке не співвідносилось з традиційним поняттям людського інтелекту. На думку науковця, основною проблемою було те, що на той момент неможливо було чітко визначити, які саме обчислювальні процеси варто вважати інтелектуальними. Він зазначав: «Ми маємо певне розуміння деяких механізмів інтелекту, однак інші залишаються поза нашим знанням. У межах цієї науки інтелект визначається як обчислювальний компонент здатності досягати цілей у навколишньому середовищі». Він також наголошував: «Наші дослідження базуються на припущенні, що кожен характеристичу інтелекту можна описати настільки точно, щоб вона могла бути відтворена обчислювальною машиною. Ми прагнемо зрозуміти, як навчити машини опрацьовувати природні мови, створювати абстракції та концепції, вирішувати задачі, що досі під силу винятково людям, а також самовдосконалюватися» (History of Data Science, n.d.). Водночас під час семінару Марвін Мінський звернув увагу на важливість роз-

робки адаптивних систем навчання, підкресливши: «Необхідно створити машину, яка демонструватиме здатність до навчання. Така система повинна бути оснащена вхідними та вихідними каналами, що забезпечують різноманітність реакцій на запити. Цей підхід можна охарактеризувати як «метод спроб і помилок», що полягає у формуванні набору функцій для введення та виведення. У разі відповідного налаштування середовища й визначення критеріїв успіху чи невдачі така машина може проявляти цілеспрямовану поведінку» (History of Data Science, n.d.).

Сьогодні ШІ розглядається як технологія, що дозволяє імітувати когнітивні функції людини з метою отримання результатів, порівнянних із результатами інтелектуальної діяльності людини. Центральним елементом таких систем є алгоритми, здатні здійснювати аналіз, прогнозування, розпізнавання та інтерпретацію даних, які ґрунтуються на математичних, фізичних і комп'ютерних принципах. Діяльність таких систем варіюється від автоматизації раціональних дій до реалізації інтуїтивних процесів. Ключовим аспектом є здатність ШІ адаптуватися до змінних умов середовища та генерувати відповідні реакції залежно від обставин. Методологія ШІ активно застосовується в різних галузях науки та техніки, зокрема в машинному навчанні, розробці генетичних алгоритмів та використанні нечіткої логіки. Ця технологія, яка перебуває на етапі інтенсивного розвитку, потребує значних обчислювальних ресурсів, обробки великих обсягів даних і вдосконалених алгоритмів, що вимагає подальшого наукового дослідження й практичного впровадження (Gozalo-Brizuela & Garrido-Merchan, 2023).

Розвиток ШІ висуває вимогу забезпечення прозорості в роботі систем, які повинні бути не лише засобом для вирішення завдань, але й демонструвати зрозумілість процесу отримання результатів. Важливо, щоб такі системи не діяли як «чорні скриньки», а забезпечували ефективну й обґрунтовану підтримку в прийнятті рішень (Gozalo-Brizuela & Garrido-Merchan, 2023). Ключовий

аспект сучасного розвитку ШІ — це створення керованих людиною автоматизованих систем. Для досягнення цієї мети науковці визначають такі стратегічні напрями: розробка математичних основ методів інтелектуальної обробки та аналізу даних для різних прикладних завдань; створення методологій комп'ютерного моделювання та аналізу фізичних процесів; інтеграція інтелектуальних технологій у реальні досліджувані процеси та явища; інтерпретація й аналіз результатів моделювання за допомогою інтелектуальних підходів (Gozalo-Brizuela & Garrido-Merchan, 2023). Основою роботи ШІ є здатність систематизувати й аналізувати великі обсяги даних за допомогою інтелектуальних алгоритмів, які здатні автоматично навчатися, виявляючи закономірності та ознаки у вхідних даних, щоб формувати прогнози та пропонувати оптимальні рішення в багатопараметричних завданнях. У цьому процесі широко використовуються штучні нейронні мережі, які побудовані за аналогією до біологічних нейронних мереж, а також методи статистичного аналізу та дослідження операцій (Gozalo-Brizuela & Garrido-Merchan, 2023).

Одним із центральних напрямів розвитку ШІ є машинне навчання, яке забезпечує здатність систем до самостійного отримання знань у процесі виконання завдань. Цей клас методів базується на застосуванні математичної статистики, оптимізаційних методів, теорії ймовірностей, графів та інших підходів до обробки даних. Є два основні підходи до навчання: навчання на прикладах (виявлення емпіричних закономірностей у даних) та дедуктивне навчання (формалізація знань експертів і їх інтеграція в комп'ютерні системи у вигляді баз знань) (OpenAI API, 2023). Машинне навчання знаходить застосування в численних галузях, включаючи науку, виробництво, бізнес, медицину та транспорт. Зростання обсягів даних внаслідок інформатизації сприяє розширенню сфери його використання, дозволяючи вирішувати завдання прогнозування, управління й прийняття рішень, які раніше розв'язувалися іншими методами (Bang et al., 2023).

Аналіз алгоритму та моделі прийняття рішень ШІ на прикладі ChatGPT. Сучасний етап технологічного прогресу та впровадження інноваційних рішень сприяв появі новітніх підходів і методів у сфері ШІ. Особливу увагу привертають алгоритми «глибокого навчання», які стали каталізатором підтримки розвитку ШІ з боку урядів багатьох країн (Gozalo-Brizuela & Garrido-Merchan, 2023). Галузь досліджень у цій сфері тісно переплітається з іншими науковими дисциплінами, такими як теоретична та прикладна математика, фізика, аналіз сигналів, машинне навчання, комп'ютерний зір, психологія, лінгвістика, біологія тощо. Ця взаємодія створює умови для міждисциплінарного розвитку й сприяє розробці інноваційних технологій.

Розглянемо модель ChatGPT, яка базується на технології штучного інтелекту (ШІ). ChatGPT створеної на основі мовної моделі ШІ та розробленої компанією OpenAI. Офіційне представлення ChatGPT як прототипу відбулося 30 листопада 2022 року, але її доступність для громадськості було оголошено 30 січня 2023 року (OpenAI, n.d.). Висока технологічність ChatGPT сприяли її швидкому поширенню. Вже за перший тиждень після запуску платформа нарахувала понад мільйон користувачів (Grant & Metz, 2022).

Ця модель стала важливим інструментом для широкого спектра застосувань, таких як машинне навчання, генерація тексту та прогнозування контенту, створення і модифікацію зображень на основі текстових описів (Gozalo-Brizuela & Garrido-Merchan, 2023), перетворення аудіо в текст (OpenAI, 2023), формування числових представлень тексту (Brennan & Lesage, 2023) (OpenAI, n.d.). Ці моделі знайшли застосування як в дослідницькій діяльності так і в комерційних проектах. (OpenAI, n.d.), (Brown, Mann, & Ryder, 2020), (OpenAI, 2023). „(OpenAI, n.d.)в й (Gozalo-Brizuela & Garrido-Merchan, 2023)(OpenAI, 2023)(OpenAI, 2023)межПід час першого етапу модель навчалася на значних обсягах текстових даних, використовуючи методи неконтрольованого навчання, такі як мовне моделювання та масковане мовне моделю-

вання. Основною метою цього етапу було сформувані в моделі глибоке розуміння природної мови, зокрема її структури та взаємозв'язків між словами й реченнями. Наступний етап передбачав точне налаштування моделі на основі завдань, таких як завершення тексту, відповіді на запитання та генерація діалогів. Для цього використовували марковані набори даних (Brownlee, 2021). Завдяки цьому процесу вдалося створити універсальну мовну модель, здатну ефективно виконувати завдання обробки природної мови та генерувати відповіді, які максимально наближені до людських (OpenAI, n.d.). Завдяки високій продуктивності ChatGPT здобула репутацію однієї з провідних мовних моделей (Chen et al., 2021).

Основною метою її розробки було забезпечення точних і релевантних відповідей на запити користувачів. Архітектура ChatGPT базується на складних алгоритмах глибокого навчання. Навчальна база моделі включає різноманітні джерела, зокрема понад 150 мільярдів текстових елементів, створених людиною, таких як книги, наукові статті, блоги, відгуки та записи онлайн-розмов (Dowling & Lusey, 2023). Такий різноплановий набір даних дозволяє моделі брати участь у складних діалогах і забезпечувати точну інформацію з багатьох тем (OpenAI API, 2023). Серія GPT-3.5 складається з моделей, що були навчені на комбінованих текстових і кодових даних, зібраних до кінця четвертого кварталу 2021 р. Завдяки великому обсягу текстових даних, на яких проводилося навчання, модель здатна аналізувати структуру, закономірності та нюанси людської мови. Мовне моделювання ChatGPT має значний потенціал для трансформації взаємодії людини з комп'ютерами, забезпечуючи природне та інтуїтивне спілкування (Wahde & Virgolin, 2022). Гнучкість ChatGPT зумовлена здатністю навчатися на структурованих і неструктурованих даних, що робить його універсальним інструментом для різних завдань розмовного ШІ (Wei et al., 2022). Його вдосконалена нейронна архітектура забезпечує обробку множинних вхідних даних і генерацію персоналізованих відпові-

дей (Zhang et al., 2022). Окрім того, модель демонструє здатність адаптуватися до вподобань і стилів спілкування користувачів із часом, що значно підвищує її ефективність (OpenAI, n.d.). Уміння генерувати контекстуальні та зв'язні відповіді кількома мовами сприяє подоланню мовних бар'єрів і розширює можливості міжкультурної комунікації. Завдяки високій якості згенерованих текстів модель має широкий спектр застосувань у творчих галузях. Однією з ключових переваг ChatGPT є здатність розуміти контекст введених даних і забезпечувати релевантні відповіді (Susnjak, 2022). Тексти, створені моделями GPT, вирізняються високою якістю, через що їх складно відрізнити від текстів, створених людиною. Окрім того, ChatGPT здатна надавати відповіді на запитання, що виходять за межі можливостей стандартного інтернет-пошуку.

Основою архітектури ChatGPT є спеціалізована нейронна мережа, розроблена для моделювання мови та генерації тексту (OpenAI, 2023). Її структура базується на використанні серії блоків із механізмами самонавчання, що дозволяють моделі концентрувати увагу на різних аспектах вхідного тексту. Це забезпечує розуміння складних нелінійних кореляцій між вхідними та вихідними даними (OpenAI, n.d.). ChatGPT використовує адаптований варіант архітектури (OpenAI, n.d.), що включає велику кількість параметрів. Це дає змогу моделі ефективно аналізувати складні взаємозв'язки в даних (OpenAI, 2023). Попередня обробка текстових даних є важливим компонентом процесу навчання. Вона впливає на якість і придатність вхідної інформації, яка подається до моделі (OpenAI, 2023). Процес попередньої обробки включає кілька ключових етапів, а саме: токенизація (текст поділяється на дискретні одиниці, або токени, які стають базовими елементами для подальшої обробки) (Rivas & Zhao, 2023); кодування підслів'їв (текст розбивається на дрібніші частини, або підслова, що покращує ефективність моделі під час виконання різноманітних завдань обробки природної мови) (OpenAI, 2023), (OpenAI API, 2023)];

очищення даних (від тексту відсіюється нерелевантна або «зашумлена» інформація, що покращує якість даних і сприяє кращій адаптації моделі до значущих закономірностей у тексті) (OpenAI, n.d.). Навчання ChatGPT здійснюється за допомогою неконтрольованого підходу, який базується на моделюванні мови (OpenAI API, 2023). Модель навчається передбачати наступне слово в тексті, використовуючи попередній контекст. Основною метою цього процесу є мінімізація функції втрат шляхом точного передбачення слів у текстовій послідовності.

Процес навчання складається з наступних етапів:

- налаштування базових параметрів моделі;
- попереднє навчання (нейронна мережа навчається на великих обсягах немаркованих текстових даних. На цьому етапі виділяють дві фази: неконтрольовану, що базується на аналізі тексту за допомогою трансформаторної архітектури, і контрольовану, яка передбачає тонке налаштування на основі маркованих даних) (OpenAI, n.d.);
- точне налаштування (на цьому етапі модель адаптується до конкретних завдань, таких як текстова класифікація чи генерація відповідей. Це досягається шляхом використання меншого корпусу маркованих даних);
- тонке налаштування (цей етап забезпечує додаткову оптимізацію роботи моделі. Він включає підготовку даних, модифікацію архітектури та оптимізацію параметрів (OpenAI API, 2023). На цьому етапі модель навчається на позначених даних, які проходять ті самі процедури попередньої обробки, що й немарковані дані. Архітектура адаптується до специфіки завдання, а параметри моделі оптимізуються для мінімізації функції втрат, що забезпечує високоякісну обробку природної мови (OpenAI API, 2023).

Обмеження, перспективи та можливості впровадження ШІ в різних галузях. ChatGPT демонструє значний прогрес у сфе-

рі моделювання природної мови. Її здатність генерувати зв'язні та контекстуально релевантні відповіді відкриває нові перспективи для розробки та впровадження технологій ШІ в різних галузях (OpenAI, 2023). Історичний і технічний аналіз ChatGPT, проведений дослідниками-науковцями, підтверджує можливість його застосування в різних галузях, зокрема в медицині, освіті та наукових дослідженнях. Встановлено, що ChatGPT демонструє здатність аналізувати великі обсяги даних, упорядковувати великий обсяг інформації, що є його значною перевагою. Водночас підкреслюються його обмеження, такі як імовірність помилок та можливі неточності в логічних побудовах, математичних розрахунках та викладі фактів, а також обмежене знання спеціалізованої термінології. За результатами досліджень, максимальний обсяг вхідних даних, які може обробити ChatGPT, становить 5000 текстових токенів.

Застосування ChatGPT відкриває значний потенціал для інновацій, але також створює низку викликів, пов'язаних із контекстуальним розумінням, упередженістю, пояснюваністю та чутливістю до формулювань вхідних даних. Для ефективного впровадження моделі ChatGPT важливо враховувати його сильні та слабкі сторони. Одним із ключових викликів є можливість демонстрації упередженості у відповідях моделі. Оскільки ChatGPT навчається на великому обсязі даних, що можуть містити упередження, ці тенденції можуть впливати на результати, підтверджуючи стереотипи або призводячи до нерівного ставлення. Тому необхідно розробляти методи для виявлення, моніторингу та мінімізації упередженості, щоб забезпечити справедливі відповіді незалежно від походження користувачів. Ще одна проблема полягає у створенні дезінформації чи небезпечного контенту через сприйнятливості ChatGPT до формулювань вхідних даних. Це може спричинити поширення шкідливих порад чи неправдивих фактів. Запобігти цьому можна шляхом впровадження механізмів модерації та фільтрації контенту, що зменшить ризики неналежних відповідей і

захистити користувачів від потенційної шкоди. Мовні моделі, зокрема ChatGPT, можуть обробляти конфіденційну інформацію користувачів, тому пріоритетним є захист даних та конфіденційності. Необхідно впроваджувати стандарти, як-от анонімізація, наскрізне шифрування та диференційована конфіденційність, щоб запобігти витоку даних і їх неправильному використанню (Gozalo-Brizuela & Garrido-Merchan, 2023). Прозорість моделей є ключовим фактором довіри до ШІ. Оскільки ChatGPT генерує складні відповіді, потрібні дослідження для створення механізмів пояснення його рішень, що підвищить зрозумілість і довіру до моделі (OpenAI, n.d.). Інтеграція ChatGPT готових рішень завдання, таких як генерація контенту чи підтримка клієнтів, має як позитивний потенціал, так і ризики витіснення робочих місць. Важливо забезпечити баланс між участю людини й ШІ, розвивати співпрацю та сприяти перекваліфікації осіб, які можуть постраждати від автоматизації (OpenAI, 2023). Питання відповідальності та підзвітності ШІ є критичним. З огляду на значні набори даних, на яких навчається ChatGPT, необхідно розробити фреймворки для визначення відповідальності за результати моделі та чіткі рекомендації для користувачів і розробників (OpenAI, n.d.). Етичні аспекти є важливими для вдосконалення ChatGPT. Вирішення цих питань сприятиме створенню відповідальних і надійних систем, які узгоджуються з людськими цінностями (OpenAI, n.d.). Ці аспекти є необхідними для етичного та безпечного використання мовних моделей у сучасному суспільстві.

ChatGPT забезпечує широкі можливості завдяки розумінню природної мови, здатності генерувати текст, схожий на людський, та підтримувати контекст. Ця особливість робить його ефективним інструментом у проєктах обробки природної мови (OpenAI, 2023). Серед спеціальних функцій варто виділити простоту взаємодії: модель дозволяє користувачам без технічного досвіду використовувати складні технології, забезпечуючи універсальність у застосуванні. ChatGPT підходить для різних сфер, зокрема медицини, освіти,

підтримки клієнтів, творчого письма та перекладу (OpenAI, 2023). Ефективність ChatGPT полягає в здатності швидко створювати контент і відповідати на запити, що значно скорочує час виконання завдань. Крім того, його можна адаптувати для конкретних потреб шляхом тонкого налаштування, що робить модель гнучкою та зручною у використанні (OpenAI API, 2023). Завдяки здатності до трансферного навчання ChatGPT може працювати з великими наборами даних, демонструючи багатомовні можливості. Це сприяє глобальній доступності та підтримці різних мов (Gozalo-Brizuela & Garrido-Merchan, 2023). Модель також підтримує навчання з нульовими або мінімальними даними, дозволяючи вирішувати нові завдання без значної підготовки. Її адаптивність до специфічних доменів забезпечує релевантність відповідей у різних галузях (Rivas & Zhao, 2023). ChatGPT є інструментом, здатним покращувати освітній процес, співпрацювати у творчих проєктах, підтримувати психічне здоров'я, а також надихати інновації в обробці природної мови. Його здатність створювати взаємодію, схожу на людську, робить модель привабливою для інтерактивних додатків (Gozalo-Brizuela & Garrido-Merchan, 2023). Водночас слід зважати на недоліки цієї технології, серед яких обмеженість розуміння контексту, яка призводить до того, що в довготривалих розмовах модель може втрачати зв'язність, що породжує абсурдні або нерелевантні відповіді. ChatGPT може відображати упередження, наявні в навчальних даних, посилюючи соціальні стереотипи; модель здатна генерувати відповіді, які мають переконливий вигляд, але є фактично хибними; ChatGPT лише імітує розуміння, що може створювати оманливе враження в користувачів; використання моделі для створення фейкових відгуків чи дезінформації також викликає занепокоєння, тому що ChatGPT може бути використаний для створення шкідливого контенту або атак соціальної інженерії; обмін даними через модель може загрожувати конфіденційності користувачів; результати ChatGPT сильно залежать від якості та різноманітності даних,

використаних у навчанні; нездатність до причинно-наслідкового міркування (генерація відповідей ґрунтується на шаблонах, а не на реальному розумінні причинно-наслідкових зв'язків). Незважаючи на численні переваги, описані недоліки вказують на необхідність постійного вдосконалення мовних моделей для забезпечення їх ефективності, етичності та безпеки використання.

У сучасних наукових дослідженнях приділяється значна увага оцінці ефективності ChatGPT у контексті академічного письма, зокрема в галузі біомедичних наук (Kumar, 2023). Дослідження, присвячені використанню ChatGPT у сфері охорони здоров'я (Kung et al., 2023), акцентують увагу на його потенціалі як інструмента для вдосконалення медичної освіти (Dahmen et al., 2023). Інструмент (ШІ) оцінюється за критеріями часу реакції, якості створюваного контенту та надійності відповідей на визначену кількість випадково обраних тем із відповідного переліку запитів. Науковці пропонують використовувати ChatGPT для навчання студентів навичкам академічного письма, акцентуючи увагу на його потенціалі для покращення навчального процесу та підкреслюючи необхідність подальшого вивчення цього напрямку. Особливу увагу приділено можливостям застосування ШІ в стоматологічній освіті, де досліджується точність ChatGPT у відповіді на стоматологічні запити. Вчені розглядають перспективи вдосконалення стоматологічної освіти за допомогою ChatGPT, зокрема в теоретичній підготовці та практичних навичках (Thurzo et al., 2023). Крім того, проводяться дослідження щодо здатності ШІ аналізувати дані пацієнтів, сприяти віддаленому моніторингу пацієнтів, генерувати навчальні плани, створювати віртуальні навчальні середовища та оцінювати студентів у практичних завданнях. Водночас вчені звертають увагу на етичні аспекти використання ChatGPT, наголошуючи на важливості дотримання етичних принципів, упроваджуючи ШІ в медичну освіту. Вони підкреслюють необхідність забезпечення належного регулювання та етичного контролю, що має супроводжувати розвиток і застосу-

вання таких технологій (OpenAI, 2023). Також є наукові дослідження, присвячені оцінці діагностичної точності списків диференційної діагностики, сформованих ChatGPT-3, для типових клінічних випадків (Hirosawa et al., 2023). У ході аналізу було встановлено, що ChatGPT демонструє високу точність у визначенні діагнозів, яка перевищує 90%. Це свідчить про значний потенціал технології в медичній практиці. Проте для розширення можливостей ШІ пропонується проводити тести зі складнішими клінічними випадками та вивчати нові аспекти його застосування в системі охорони здоров'я. Є дослідження, у яких порівнюється ефективність технологій на основі ШІ, таких як ChatGPT, і традиційних методів діагностики, моніторингу та оцінки прогресу під час ортодонтичного лікування (Strunga et al., 2023). Встановлено, що ШІ дозволяє оптимізувати роботу ортодонтів, підвищуючи кількість пацієнтів, яких можна обслуговувати без зниження якості лікування. Окрім цього, використання програмного забезпечення на основі ШІ значно спрощує процес планування ортогнатичної хірургії порівняно з традиційними діагностичними підходами. Водночас науковці висловлюють занепокоєння щодо надійності таких технологій через можливі помилки та упередження, які можуть призвести до ускладнень у лікувальному процесі. Дослідження підкреслюють, що впровадження ШІ в ортодонцію має значний позитивний вплив, але вирішення наявних обмежень у діагностиці та плануванні лікування потребує подальших глибоких і досконалих досліджень (Rivas & Zhao, 2023).

В Україні на платформі соціального краудфандингу «Соціальна країна» (Finance.ua, н.д.) з'явився проєкт зі створення AI-застосунку, який допоможе дітям із порушеннями мовлення спілкуватися. Це рішення підтримує дітей із аутизмом, ДЦП чи затримкою психомовного розвитку, допомагаючи долати бар'єри в спілкуванні. Проєкт вже здобув широку підтримку та перебуває на етапі розробки. Технологія працюватиме завдяки відстеженню погляду дитини, що дозволить й

обирати слова та формувати фрази (Соціальна країна, 2023), (Sotsialna Kraina, 2023).

Було проведене дослідження, спрямоване на оцінку здатності ChatGPT генерувати стислий огляд теми «системи доставки ліків на основі ліпідів» за допомогою методів діалогового навчання (Zhu et al., 2023). Аналіз відповідей ChatGPT показав, що він здатний створювати узагальнені огляди, які можуть слугувати основою для подальших досліджень. Проте обмеження, зумовлені браком посилань на надійні джерела, ставлять під сумнів точність і послідовність таких оглядів. Науковці дійшли висновку, що ChatGPT може надавати лише загальну інформацію з обраної теми, але не забезпечує детального аналізу. Водночас науковці досліджують ефективність ChatGPT у клінічному прийнятті рішень шляхом аналізу точності його відповідей на стандартизовані клінічні запити (Nastasi et al., 2023). Результати показали, що рівень точності ChatGPT становить 71,7%, що підтверджує його потенціал як допоміжного інструмента. Проте ChatGPT не може бути використаний як єдиний засіб прийняття рішень, оскільки його відповіді інколи можуть бути недостатньо точними. У дослідженнях впливу ChatGPT на діяльність стоматологічних асистентів і медичних сестер в ортодонтії встановлено, що використання ШІ може підвищити точність і персоналізацію лікування. При цьому наголошується на необхідності подальших досліджень щодо впливу ШІ на робочі процеси у сфері охорони здоров'я. Вчені рекомендують розробити чіткі керівні принципи для розв'язання етичних і правових питань, що постають у процесі інтеграції таких технологій. ШІ має потенціал для покращення медичного догляду та підвищення результативності лікування пацієнтів, але його впровадження потребує адаптації фахівців і врахування особливостей галузі.

Науковці-дослідники також активно вивчають потенційні наслідки впровадження ChatGPT у педіатричних дослідженнях, аналізуючи як позитивні, так і негативні аспекти його застосування (Corsello & Santangelo, 2023). До переваг відносять покращення

клінічних рішень, прискорення розробки лікарських засобів і підвищення ефективності лікування. Водночас порушуються питання безпеки та етичних викликів. Науковці зазначають, що, попри очевидні ризики, мовні моделі можуть зробити вагомий внесок у розвиток педіатрії. Вирішення ж викликів можливе шляхом співпраці лікарів, дослідників і фахівців зі ШІ, спрямованої на реалізацію повного потенціалу цієї технології для покращення здоров'я дітей.

У контексті вищої освіти науковці досліджують вплив ШІ, зокрема ChatGPT, на навчальний процес (Rudolph & Tan, 2023). Аналіз показав, що для студентів ці технології забезпечують покращення інтелектуальних систем підтримки, а для викладачів — автоматизацію завдань, таких як оцінювання, перевірка на плагіат тощо. Проте науковці рекомендують закладам вищої освіти та студентам впроваджувати заходи, спрямовані на мінімізацію можливих недоліків використання таких інструментів. Питання доброчесності в онлайн-екзаменах також стало об'єктом уваги науковців (Susnjak, 2022). Дослідження показали, що ChatGPT може генерувати текст високої якості з й демонструвати розвинені аналітичні здібності, що підвищує ризик шахрайства. Застосування наукових методів для аналізу здатностей мислення виявило, що згенеровані відповіді були актуальними й відповідали заданим темам. Серед запропонованих рішень — створення тестових запитань, недоступних у відкритих джерелах, повернення до усних екзаменів і використання моделей для виявлення текстів, створених GPT (OpenAI, 2023). Ці результати підкреслюють необхідність подальших досліджень і розробки ефективних заходів для збереження академічної доброчесності.

У сучасних дослідженнях увага приділяється можливим наслідкам впровадження ChatGPT у сферу освіти (Tlili et al., 2023). Аналіз, що включав три етапи — вивчення соціальних мереж, контент-аналіз інтерв'ю та дослідження досвіду користувачів, — показав, що ChatGPT сприймається як потужний

інструмент для трансформації освітнього процесу. Водночас були висловлені застереження, що стосуються запобігання шахрайству, забезпечення якості контенту та створення рівного доступу до технології. У результаті дослідники запропонували інтегрувати ChatGPT у навчальні практики та надали рекомендації щодо напрямів майбутніх досліджень для вирішення виявлених викликів (Rivas & Zhao, 2023). Це підкреслює важливість усвідомлення як ризиків, так і переваг ШІ в освітньому середовищі. Окремий напрям досліджень стосується використання ChatGPT для навчання англійської мови (Kovačević, 2023). Науковці пропонують застосовувати цю технологію для створення навчальних матеріалів, таких як генерування текстового контенту, розробка вправ на збагачення словникового запасу, засвоєння граматичних конструкцій, а також як інструмент віртуального помічника-вчителя чи інтерактивного навчального ресурсу. Виявлено значний потенціал ChatGPT у підвищенні ефективності освітнього процесу. Однак зазначено, що його застосування може бути складним для викладачів, які не мають знань у програмуванні чи машинному навчанні. У зв'язку із цим науковці та дослідники рекомендують організувати навчання для педагогів, що дозволить ефективніше впроваджувати ChatGPT у навчальну діяльність (Tlili et al., 2023).

Наукові дослідження, присвячені вивченню сприйняття ChatGPT студентами, виявили загалом їх позитивне ставлення до цього інструмента. Студенти відзначили його ефективність, але водночас висловили застереження щодо точності відповідей. Було підкреслено, що для ефективного використання ChatGPT необхідні попередні знання, а також запропоновано педагогам навчати студентів технікам формулювання запитів і перевірки отриманих відповідей (OpenAI, 2023). Дослідники рекомендували розробникам зосередитися на підвищенні точності цієї технології. Інше дослідження, яке аналізувало використання ChatGPT для написання наукових статей, висвітлює як переваги,

так і ризики (Castellanos-Gomez, 2023). Було виявлено, що ChatGPT може сприяти створенню якісного тексту для неангломовних авторів, допомагаючи з перевіркою лексики та граматики. Водночас було зазначено, що є небезпека створення матеріалів низької якості через надмірну залежність від інструмента. У межах досліджень можливостей та обмежень ChatGPT у мисленні, фактичності, генерації вигаданої інформації та взаємодії порівняно з іншими мовними моделями було встановлено, що ChatGPT демонструє впевненість у дедуктивному та абдуктивному мисленні, але стикається з труднощами під час вирішення семантичних завдань, які не мають текстової основи (Bang et al., 2023)]. Дослідники також виявили, що ChatGPT може створювати вигадану інформацію (галюцинації), що обмежує його надійність як інструмента для критичного мислення. Дослідження ефективності ChatGPT у виправленні помилок, яке оцінювало його продуктивність на основі аналізу набору даних і порівняння з іншими методами, показало, що ChatGPT значно перевершує традиційні алгоритми глибокого навчання (Sobania et al., 2023). Отримані результати підтверджують потенціал цієї технології для автоматизованого редагування тексту, хоча й потребують подальших удосконалень для забезпечення більшої точності та універсальності.

У США створено першу державну школу, де викладання повністю здійснюється за допомогою штучного інтелекту. Онлайн-школа Unbound Academy, розташована в штаті Арізона, приймає учнів з 4-го до 8-го класів. Інноваційний підхід, який раніше тестували в приватних школах Техасу та Флориди, використовує інтерактивні платформи, такі як EdTech, IXL і Khan Academy. Штучний інтелект адаптується до індивідуального темпу та стилю навчання кожного учня, посилюючи увагу до успіхів і допомагаючи долати труднощі. Завдяки цьому підходу школярі можуть вивчати предмети на різних рівнях, наприклад, читати на рівні 8-го класу й опановувати математику за програмою 3-го класу. Роль викладачів зводиться до наставництва,

тоді як академічні дисципліни повністю викладаються штучним інтелектом. Заняття тривають до двох годин на день, решту часу учні займаються розвитком життєвих навичок, зокрема критичним мисленням, фінансовою грамотністю, публічними виступами та підприємництвом. На перший рік навчання школа планує зарахувати близько 200 учнів. У перспективі програма буде розширена для дітей молодших вікових груп, починаючи з дитячого садочка (Finance.ua, н.д.).

Наукові дослідження, присвячені етичним аспектам використання ChatGPT та інших мовних моделей (Zhuo et al., 2023), підкреслюють, що ці технології можуть спричинити соціальні упередження та токсичність, що створює моральні та соціальні ризики (Moradi, Dorffner, & Samwald, 2020). Аналіз етичних проблем, таких як упередженість, надійність, токсичність та стійкість, акцентує увагу на необхідності розробки етичних принципів для роботи з ChatGPT. Дослідники висловили низку пропозицій для майбутнього вдосконалення мовних моделей, які спрямовані на виявлення та мінімізацію потенційних ризиків. Їхні висновки мають значний внесок у розвиток етичних підходів до ШІ. Оцінюючи ефективність ChatGPT у виконанні завдань з розуміння природної мови, таких як аналіз настроїв, перефразування, визначення текстової подібності та лінгвістична прийнятність, дослідники зауважили, що ChatGPT демонструє високі результати (Zhong et al., 2023). Однак порівняльний аналіз показав, що модель не перевершує інші передові технології у вузькоспеціалізованих завданнях, що вказує на наявність обмежень. У дослідженнях можливостей ChatGPT для генерації наукових статей із фінансової тематики було виявлено, що ця модель є дієвим інструментом для формування наукових ідей завдяки доступу до значного обсягу текстових даних. Однак якість оглядів літератури та запропонованих методик оцінювалася як недостатня для самостійного створення висококласних публікацій. Це підкреслює важливість залучення експертів із відповідної галузі для аналізу й адаптації згенерованих текстів. У

висновках досліджень наголошується на доцільності використання ChatGPT як допоміжного інструмента в поєднанні з експертизою людини-фахівця.

Дослідження використання ChatGPT для оцінювання знань з основних економічних принципів на базовому рівні за стандартизованими форматами продемонструвало його здатність успішно справлятися з такого роду завданнями завдяки широкому доступу до знань, що часто є важкодоступними для студентів. Водночас дослідники наголосили на важливості поєднання таких підходів із традиційними методами оцінювання для забезпечення комплексної перевірки знань (Geerling et al., 2023). Емпіричний аналіз можливостей ChatGPT у режимі нульового навчання на основі двадцяти широко застосовуваних наборів даних у сфері обробки природної мови, що охоплювали сім основних категорій завдань (Qin et al., 2023), засвідчив його переваги. Зокрема, ChatGPT продемонстрував високу ефективність у завданнях, що вимагають арифметичного мислення та формулювання точних відповідей для задач із виведення природної мови. Дослідники зазначили значний потенціал ChatGPT для обробки природної мови та можливість подальшого вдосконалення цієї моделі (Rivas & Zhao, 2023). У ході досліджень, спрямованих на створення моделі машинного навчання для розрізнення текстів, написаних людиною, і тих, що були згенеровані ChatGPT (Mitrović, Andreoletti, & Ayoub, 2023), особливу увагу приділили коротким текстам, які є складними для автоматичної класифікації. Результати вказали на труднощі, з якими стикається модель машинного навчання під час розпізнавання перефразованого тексту, створеного ChatGPT, порівняно з оцінками за методами на основі характеристик тексту. Було встановлено, що текст, згенерований ChatGPT, вирізняється меншою послідовністю та емоційністю порівняно з текстами, створеними людиною (Mitrović, Andreoletti, & Ayoub, 2023). Науковці детально описали, як ChatGPT може сприяти автоматизації численних завдань у науково-дослідницькій

діяльності, включаючи очищення даних, тренування моделей та інтерпретацію результатів. Відзначено значний потенціал цієї мовної моделі в роботі з неструктурованими даними, а також її здатність надавати нові інсайти для вдосконалення процесів прийняття рішень (Hassani & Silva, 2023). Водночас зазначено, що ефективність ChatGPT у виконанні конкретних завдань залежить від попереднього навчання на відповідних даних, а отримані результати можуть бути складними для інтерпретації. Проте головною перевагою ChatGPT є його здатність суттєво підвищувати продуктивність і точність наукових методологій, що робить його важливим інструментом у сфері науки про дані. Дослідники наголосили на необхідності розв'язання проблем, пов'язаних із плагіатом, та запропонували проведення додаткових досліджень для розробки етичних норм використання ChatGPT у науковій діяльності (OpenAI, 2023).

У межах дослідження математичних можливостей ChatGPT було встановлено, що його математичні здібності поступаються людським. Проте однією з переваг є здатність моделі шукати математичні концепції на основі наданої інформації. Це підкреслює її потенціал у доповненні людського досвіду під час розв'язання математичних задач. Дослідження також акцентує увагу на можливостях створення нового набору даних для порівняння ChatGPT з іншими мовними моделями в контексті розуміння складних математичних концепцій.

Дослідження, присвячене політичній упередженості у відповідях ChatGPT на тести політичної орієнтації (Rozado, 2023), підкреслює необхідність забезпечення неупередженості в системах ШІ, особливо у випадках, коли надається інформація на запити, що мають нормативний характер. Висновки вказують на те, що ШІ, який демонструє політичну упередженість, може мати серйозний вплив на суспільство, і тому такий феномен слід вважати проблематичним. Також вчені попередили, що завдяки таким можливостям, як генерування текстових шаблонів на

конкретні теми, ChatGPT може бути використаний як інструмент для формування та створення компонентів фішингових атак та маніпуляцій із застосуванням методів соціальної інженерії (OpenAI API, 2023). Відповідно до результатів дослідження, науковці розробили та запропонували методологію протидії фішинговим атакам та окреслили механізми захисту від атак соціальної інженерії.

Експериментальні дослідження, що стосуються використання ChatGPT у додатках робототехніки (Vemprala, Bonatti, Bucker, & Kapoor, 2023), показали, що людський нагляд є необхідним для забезпечення контролю та втручання у випадку виникнення несподіваної поведінки системи, спричиненої роботою ChatGPT. Дослідження, які аналізують вплив ChatGPT у темах мистецтва, зокрема в контексті покращення художньої творчості, демонструють, що модель може бути використана для надання лінгвістичних мистецьких знань і текстових інтерфейсів для взаємодії людини та машини (Guo et al., 2023). Вчені припустили, що ChatGPT може генерувати рішення, базуючись на інформації про мистецтво, отриманій з інтернету, що сприяє підвищенню різноманітності та точності експериментів у малюванні. Крім того, науковцями було проведено дослідження зосереджені на здатності ChatGPT надавати чіткі й детальні вказівки щодо змісту малюнка, організації елементів і розуміння абстрактних художніх форм, таких як стилі живопису та емоційний контекст. На основі цих результатів науковці зробили висновок, що ChatGPT здатний ефективно контролювати вміст живопису, що може потенційно сприяти розвитку креативності та підвищенню продуктивності в мистецтві.

Дослідники Google та Стенфордського університету створили агентів штучного інтелекту, які імітують особистість людей, на основі двогодинних інтерв'ю. У межах експерименту було розроблено 1052 моделі, кожна з яких навчалася на даних, отриманих під час інтерв'ю з відповідним учасником. Для перевірки точності ці агенти пройшли особистісні тести, соціальні опитування та логічні ігри, результати яких порівнювалися з відповідя-

ми учасників. Репліки ШІ досягли 85% відповідності людським аналогам. Однак точність залежала від типу завдань: моделі демонстрували високі результати у відтворенні відповідей на особисті опитування та визначенні соціальних установок, але були менш ефективними у прогнозуванні поведінки в економічних іграх із соціальною динамікою. Дослідники відзначають, що така технологія може розширити можливості вивчення людської поведінки, роблячи доступними нові методи аналізу (Sud.ua, 2023).

Отже, сучасні наукові дослідження в галузі ШІ зосереджуються на кількох основних напрямках: символічному моделюванні процесів мислення; обробці мов; представленні, обробці та використанні інформації; машинному навчанні; біологічному моделюванні в системах ШІ; інтелектуальній робототехніці; машинній творчості тощо. Одним із перших напрямів розвитку ШІ є моделювання мислення, яке охоплює такі аспекти, як доказ теорем, прийняття рішень, планування тощо. Це передбачає створення символічних систем для вирішення формалізованих задач, які не мають простих або ефективних алгоритмів вирішення. Іншим важливим напрямом є обробка природної мови, що включає як аналіз (розуміння мови та сприйняття смислу тексту), так і синтез (генерація тексту). Розуміння мови ШІ вимагає величезних знань про навколишній світ та здатності до взаємодії з ним. Розробники активно працюють над створенням алгоритмів, які дозволяють програмам сприймати та обробляти тексти, виправляючи фактичні та логічні помилки. Аналіз наявних досліджень вказує на очевидну прогалину в здатності ChatGPT ефективно вирішувати завдання машинного навчання. Щодо здатності ChatGPT до програмування, дослідження підтверджують, що він має відповідні навички. Однак науковці зазначають, що для ефективного складання завдань і запитів до ChatGPT необхідний певний рівень знань програмування. Крім того, адаптація ChatGPT до конкретних вимог вимагає досвіду в програмуванні. Таким чином, ChatGPT може бути корисним інструментом для допо-

моги розробникам у створенні коду, але він не є заміною кваліфікованих розробників. Водночас ChatGPT може бути використаний для допомоги вченим у розумінні фундаментальних концепцій певних галузей, хоча інструменти ШІ повинні бути вдосконалені, щоб забезпечити знання, що відповідають рівню досвіду.

ВИСНОВКИ ТА ПРОПОЗИЦІЇ ЩОДО ПОДАЛЬШИХ ДОСЛІДЖЕНЬ У СФЕРІ ВПРОВАДЖЕННЯ СИСТЕМ ШІ

Безперервний прогрес мовних моделей, прикладом якого є ChatGPT, відкрив майбутні горизонти для досліджень і розробок у галузі ШІ. Ураховуючи сучасну ситуацію, а також актуальність та важливість впровадження ШІ в різні галузі суспільного життя, вважаю, що деякі аспекти зазначеної тематики залишаються недостатньо вивченими й вимагають поглибленого аналізу та конкретизації. Такий підхід дозволить враховувати особливості застосування моделей ШІ у різних контекстах і розробляти стратегії, спрямовані на досягнення балансу між інноваціями, безпекою та суспільними інтересами.

Одним із ключових напрямів досліджень має бути вдосконалення технічних характеристик і розширення функціональних можливостей ChatGPT. Пріоритетним завданням залишається підвищення рівня контекстуального розуміння та ефективності управління діалогами. Незважаючи на значні досягнення моделі в генерації когерентних відповідей, залишається необхідність покращення роботи із тривалими розмовами, що включають кілька рівнів контексту. Інтеграція методів навчання з підкріпленням і архітектур із розширеною пам'яттю може сприяти здатності моделі зберігати й ефективно використовувати довгостроковий контекст. Розробка нових механізмів утримання контексту діалогу й оптимізації пам'яті також може забезпечити глибше розуміння цілей користувача та підвищити відповідність створюваних відповідей. Ще одним перспективним напрямом є дослідження впровадження динамічних підказок і методів спільного навчання, що

надають користувачам змогу безпосередньо впливати на поведінку моделі під час взаємодії. Нині можливості користувачів щодо керування результатами роботи моделі є обмеженими. Завдяки інтерактивним механізмам і адаптивним підказкам можливо забезпечити більш персоналізовану комунікацію та вдосконалення відповіді моделі через інтеграцію механізмів зворотного зв'язку. Водночас одним із напрямів майбутніх досліджень має бути розширення функціоналу ChatGPT поза межами текстових даних. Включення мультимодальних вхідних даних, таких як аудіо й зображення, сприятиме більшій інтерактивності й експресивності взаємодії. Це дозволить моделі генерувати відповіді, враховуючи не лише текстову інформацію, але й додаткові графічні чи аудіальні дані, що є важливим для таких сфер, як віртуальні асистенти й генерація контенту. Окрему увагу також слід приділити розширенню можливостей ChatGPT для роботи з різними мовами. Це сприятиме зниженню бар'єрів у використанні моделі для представників мовних груп із недостатніми навчальними даними, розширюючи її доступність.

Забезпечення справедливості та мінімізація упереджень у відповідях ChatGPT, є одним із ключових викликів та завдань етичного характеру. У цьому контексті майбутні дослідження мають бути спрямовані на вдосконалення алгоритмів виявлення упереджень і розробку методів навчання, які враховують принципи рівності. Це сприятиме зменшенню нерівномірності результатів для різних категорій користувачів. Окрім цього, важливим аспектом є впровадження підходів, що сприяють інтерпретації роботи моделі, дозволяючи користувачам зрозуміти механізми прийняття рішень, а також ідентифікувати та усувати можливі упередження. Розширення можливостей контролю над генерацією тексту також залишається перспективною сферою майбутніх досліджень. Використання умовного декодування та навчання з підкріпленням дозволить створювати відповіді, які відповідають стилю, настрою чи змісту, заданим користувачем. Це може стати важ-

ливим кроком до персоналізації контенту та його адаптації до різних сфер застосування, включаючи творчі завдання. Також особливу увагу слід приділити питанням ефективності обчислень і управління пам'яттю. Оптимізація обчислювальних ресурсів, застосування методів контекстного обрізання та стиснення моделі є критично важливими для забезпечення роботи ChatGPT в умовах обмежених ресурсів, особливо під час застосування в реальному часі.

В умовах роботи з конфіденційними даними користувачів важливим напрямом є розробка технологій захисту приватності. Використання таких підходів, як диференціальна конфіденційність, дозволяє забезпечити захист інформації, з якою взаємодіє модель, зберігаючи при цьому високий рівень безпеки. Перспективним напрямом є інтеграція моделей ШІ з людською співпрацею. Гібридні системи, у яких людина має можливість контролювати відповіді моделі, особливо важливі для роботи в критичних ситуаціях, забезпечуючи високу точність і надійність результатів. Важливим аспектом розвитку ChatGPT є вдосконалення роботи з тривалими діалогами та складними контекстами. Майбутні дослідження мають бути спрямовані на створення інноваційних методів, які сприяють збереженню релевантності й зв'язності під час тривалих обговорень, забезпечуючи при цьому краще розуміння змісту складних речень. Крім того, важливим є забезпечення доступності моделі для користувачів із обмеженими ресурсами. Розробка методів стиснення моделі та оптимізації її роботи дозволить використовувати ChatGPT на пристроях із низькою обчислювальною потужністю, сприяючи більшій інтеграції технології у повсякденне життя.

Загалом майбутні дослідження мають охоплювати широкий спектр питань, пов'язаних із розвитком ChatGPT, щоб забезпечити його адаптивність, ефективність і етичну відповідальність. Це відкриє значний потенціал для інтеграції моделі у різноманітні сфери діяльності, сприяючи інноваціям і соціальному прогресу.

REFERENCES

1. Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., et al. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *Future Internet*, 15(1), 192-203.
2. Brennan, R. W., & Lesage, J. (2023). Exploring the implications of OpenAI Codex on education for Industry 4.0. In A. Author & B. Author (Eds.), *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future* (pp. 254-266). Springer International Publishing.
3. Brown, T. B., Mann, B., & Ryder, N. (2020). Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165. <https://arxiv.org/abs/2005.14165>
4. Brownlee, J. (2021). How to Develop a GPT-2 Text Generator in Python. *Machine Learning Mastery*. <https://machinelearningmastery.com/how-to-develop-a-generative-model-for-text-generation-in-python/>
5. Castellanos-Gomez, A. (2023). Good practices for scientific article writing with ChatGPT and other artificial intelligence language models. *Nanomanufacturing*, 3(1), 135-138.
6. Chen, M., Tworek, J., Jun, H., Yuan, Q., de Oliveira Pinto, H. P., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., & et al.** (2021). Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374. <https://arxiv.org/abs/2107.03374>
7. Corsello, A., & Santangelo, A. (2023). May Artificial Intelligence Influence Future Pediatric Research?—The Case of ChatGPT. *Children*, 10(5), 757.
8. Dahmen, J., Kayaalp, M. E., Ollivier, M., Pareek, A., Hirschmann, M. T., Karlsson, J., & Winkler, P. W. (2023). Artificial Intelligence Bot ChatGPT in medical research: The potential game changer as a double-edged sword. *Knee Surgery, Sports Traumatology, Arthroscopy*, 31, 1187–1189.
9. Delo.ua. (2023, грудень 20). Декларація Блетчлі зі ШІ: Як Україні скористатися глобальними трендами. Delo.ua. <https://delo.ua/telecom/deklaraciya-bletcli-z-si-yak-ukrayini-skoristatisya-globalnimi-trendami-426749/#:~:text=>
10. Dowling, M., & Lucey, B. (2023). ChatGPT for (Finance) Research: The Bananarama conjecture. *Financial Research Letters*, 53, 103662.
11. Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). Is ChatGPT smarter than a student in principles of economics? *SSRN Electronic Journal*.
12. Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). CHATGPT is not all you need: A state of the art review of large generative AI models. arXiv. <https://doi.org/10.48550/arXiv.2301.04655>
13. Grant, N., & Metz, C. (2022, December 21). A new chat bot is a 'code red' for Google's search business. *The New York Times*. <https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html> 1
14. Guo, C., Lu, Y., Dou, Y., & Wang, F.-Y. (2023). Can ChatGPT boost artistic creation: The need of Imaginative Intelligence for parallel art. *IEEE/CAA Journal of Automatica Sinica*, 10, 835-838.
15. Hassani, H., & Silva, E. S. (2023). The role of ChatGPT in data science: How AI-assisted conversational interfaces are revolutionizing the field. *Big Data and Cognitive Computing*, 7(1), 62.
16. Hirose, T., Harada, Y., Yokose, M., Sakamoto, T., Kawamura, R., & Shimizu, T. (2023). Diagnostic accuracy of differential-diagnosis lists generated by Generative Pre-trained Transformer 3 chatbot for clinical vignettes with common chief complaints: A pilot study. *International Journal of Environmental Research and Public Health*, 20(5), 3378.
17. History of Data Science. (n.d.). Dartmouth Summer Research Project: The birth of artificial intelligence. Retrieved January 4, 2025, from <https://www.historyofdatascience.com/dartmouth-summer-research-project-the-birth-of-artificial-intelligence/>
18. Karwowski, M. (2015). Intelligence and creativity: A complex but important relationship. ResearchGate. https://www.researchgate.net/publication/276166527_Intelligence_and_creativity_A_complex_but_important_relationship
19. Kovačević, D. (2023). Use of ChatGPT in ESP teaching process. In *Proceedings of the 2023 22nd International Symposium INFOTEH[1]JAHORINA (INFOTEH)* (pp. 125–132). East Sarajevo, Bosnia and Herzegovina: University of East Sarajevo.

20. Kumar, H. S. A. (2023). Analysis of ChatGPT tool to assess the potential of its utility for academic writing in biomedical domain. *Biological Engineering & Medical Science Reports*, 9, 24-30.
21. Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., et al. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(1), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
22. Mitrović, S., Andreoletti, D., & Ayoub, O. (2023). ChatGPT or human? detect and explain. Explaining decisions of machine learning model for detecting short ChatGPT-generated text. *arXiv preprint arXiv:2301.13852*.
23. Moradi, M., Dorffner, G., & Samwald, M. (2020). Deep contextualized embeddings for quantifying the informative content in biomedical text summarization. *Computers in Biology and Medicine*, 184, 105117.
24. Nastasi, A. J., Courtright, K. R., Halpern, S. D., & Weissman, G. E. (2023). Does ChatGPT provide appropriate and equitable medical advice? A vignette-based, clinical evaluation across care contexts. *medRxiv*.
25. OpenAI. (n.d.). OpenAI Blog. Retrieved April 26, 2023, from <https://openai.com/blog/>
26. OpenAI. (2023). Data usage policies. Retrieved April 26, 2023, from <https://platform.openai.com/docs/data-usage-policies>
27. OpenAI. (2023). Embeddings. OpenAI Platform Documentation. <https://platform.openai.com/docs/guides/embeddings>
28. OpenAI. (2023). Introducing Whisper. <https://openai.com/research/whisper>
29. OpenAI API. (2023). Model Index for Researchers. <https://platform.openai.com/docs/model-index-for-researchers> (accessed 26 April 2023).
30. OpenAI. (2023). Moderation Model. OpenAI Platform Documentation. <https://platform.openai.com/docs/guides/moderation/overview>
31. OpenAI. (n.d.). OpenAI models. Retrieved April 26, 2023, from <https://platform.openai.com/docs/models/overview>
32. Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., & Yang, D. (2023). Is ChatGPT a general-purpose natural language processing task solver? *arXiv preprint arXiv:2302.06476*.
33. Rivas, P., & Zhao, L. (2023). Marketing with ChatGPT: Navigating the ethical terrain of GPT-based chatbot technology. *AI*, 4, 375-384.
34. Sobania, D., Briesch, M., Hanna, C., & Petke, J. (2023). An analysis of the automatic bug fixing performance of ChatGPT. *arXiv preprint arXiv:2301.08653*.
35. Rudolph, J., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6.
36. Rozado, D. (2023). The political biases of ChatGPT. *Social Sciences*, 12, 148.
37. Strunga, M., Urban, R., Surovková, J., & Thurzo, A. (2023). Artificial Intelligence Systems Assisting in the Assessment of the Course and Retention of Orthodontic Treatment. *Healthcare*, 11(4), 683.
38. Susnjak, T. (2022). ChatGPT: The end of online exam integrity? *arXiv preprint arXiv:2212.09292*. <https://arxiv.org/abs/2212.09292>
39. Thurzo, A., Strunga, M., Urban, R., Surovková, J., & Afrashtehfar, K. I. (2023). Impact of Artificial Intelligence on Dental Education: A Review and Guide for Curriculum Update. *Education Sciences*, 13, 150.
40. Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15.
41. Vemprala, S., Bonatti, R., Bucker, A., & Kapoor, A. (2023). ChatGPT for Robotics: Design Principles and Model Abilities. Microsoft Autonomous Systems and Robotics Research. Retrieved April 26, 2023, from https://www.microsoft.com/en-us/research/uploads/prod/2023/02/ChatGPT____Robotics.pdf 1
42. Wahde, M., & Virgolin, M. (2022). Conversational agents: Theory and applications. *arXiv preprint arXiv:2202.03164*. <https://arxiv.org/abs/2202.03164>

43. Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M., & Le, Q. V. (2022). Finetuned language models are zero-shot learners. arXiv preprint arXiv:2109.01652. <https://arxiv.org/abs/2109.01652>
44. World Economic Forum. (2024). World Economic Forum Annual Meeting 2024. World Economic Forum. <https://www.weforum.org/meetings/world-economic-forum-annual-meeting-2024/>
45. Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J., & Dolan, B. (2022). DialogPT: Large-scale generative pre-training for conversational response generation. arXiv preprint arXiv:1911.00536. <https://arxiv.org/abs/1911.00536>
46. Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023). Can ChatGPT understand too? A comparative study on ChatGPT and fine-tuned BERT. arXiv preprint arXiv:2302.10198.
47. Zhu, Y., Han, D., Chen, S., Zeng, F., & Wang, C. (2023). How can ChatGPT benefit pharmacy: A case report on review writing. Preprints, 2023020324.
48. Zhuo, T. Y., Huang, Y., Chen, C., & Xing, Z. (2023). Exploring AI ethics of ChatGPT: A diagnostic analysis. arXiv preprint arXiv:2301.12867.
49. Finance.ua. (н.д.). U SSHA stvoryly pershu derzhavnu shkolu, de vykladatyme shtuchnyy intelekt. <https://wrt.sud.ua/uk/news/abroad/320202-v-ssha-poyavilas-gosudarstvennaya-shkola-gde-budet-prepodavat-isklyuchitelno-iskusstvennyy-intellekt>
50. Sotsial'na krayina. (2023, 15 travnya). AI-zastosunok, shcho daruye holos dityam, yaki ne mozhut hovoryty. <https://welfare.green/initiative/ai-zastosunok-shho-darueh-golos-dityam-yaki-ne-mozhut-govoriti/>
51. Sud.ua. (2023). Sud.ua. (2023). Iskusstvennyy intelekt mozhe stvoryty tsyfrovu kopiyyu osobystosti lyudyny vs'oho za 2 hodyny. Retrieved from <https://sud.ua/uk/news/abroad/320107-iskusstvennyy-intellekt-mozhet-sozdat-tsifrovuyu-kopiyu-lichnosti-cheloveka-vsego-za-2-chasa>
52. Sotsialna Kraina. (2023, May 15). AI-zastosunok, shcho darue holos ditiam, yaki ne mozhut hovoryty. <https://welfare.green/initiative/ai-zastosunok-shho-darueh-golos-dityam-yaki-ne-mozhut-govoriti/>

ARTIFICIAL INTELLIGENCE AS AN INNOVATIVE TOOL FOR SOCIAL CHANGE: CHALLENGES, PROSPECTS, RISKS AND SOCIAL CONSEQUENCES

Abstract. The article reviews and analyses the areas of research in the field of artificial intelligence. The article identifies relevant aspects of limitations and prospects for improvement, as well as current challenges to the implementation of this technology in various industries. It is emphasised that the introduction of this technology can cause transformations in the organisational structures of various spheres of society. It is emphasised that the use of artificial intelligence can be an effective tool for improving the well-being of society, and the possible threats and risks associated with its use are also indicated.

The features of the organisation of artificial intelligence systems, in particular, the architecture, principles of functioning, algorithms and decision-making models, are considered on the example of ChatGPT. Unevenness in the practical implementation of this technology is revealed. It is pointed out that there is no unified standard for the implementation of artificial intelligence systems in various spheres of public life.

The article provides suggestions for further research on the implementation of artificial intelligence systems.

Keywords: artificial intelligence; intelligent computer systems; cyberspace; intelligent algorithms; ChatGPT; cybersecurity.

ІНФОРМАЦІЯ ПРО АВТОРА

Жорняк Андрій — доктор філософії в галузі публічного управління та адміністрування, головний державний інспектор відділу з питань безпеки праці південного напрямку управління інспекційної діяльності у Київській області Центрального міжрегіонального управління Державної служби з питань праці; Україна, м. Київ, вул. Вавілових, 10, 04060; +38 (067) 956-95-99; e-mail: dr.andrij.zhorniak@gmail.com; ORCID: 0000-0001-9515-0180

INFORMATION ABOUT THE AUTHOR

Zhorniak Andrii — PhD, Doctor of Philosophy in Public Administration and Management, Chief State Inspector of the Labour Safety Department of the Southern Division of the Inspection Department in Kyiv Region of the Central Interregional Directorate of the State Labour Service of Ukraine, 10 Vavilovykh Street, Kyiv, Ukraine, 04060; +38 (067) 956-95-99; e-mail: dr.andrij.zhorniak@gmail.com; ORCID: 0000-0001-9515-0180