

О. А. БОБАРЧУК, канд. техн. наук

Т. В. ЗЛОТКІВСЬКА, студентка

ІННОВАЦІЙНІ МОЖЛИВОСТІ РОЗВИТКУ ШТУЧНОГО ІНТЕЛЕКТУ ЗА ДОПОМОГОЮ МУЛЬТИМОДАЛЬНОГО ПІДХОДУ

Резюме. Публікація присвячена дослідженню мультимодального штучного інтелекту в порівнянні з одно-модальними системами. У статті розглянуто математичні аспекти функціонування мультимодальних систем, висновки щодо потенціалу і переваг у порівнянні з одномодальними.

Мультимодальний штучний інтелект має значний потенціал для покращення життя людей у таких галузях, як медицина, освіта, бізнес і розваги. Дослідження нових та інноваційних застосувань є дуже актуальним. Зазначена технологія використовується для покращення ефективності та точності чинних систем. Мультимодальний штучний інтелект може використовуватися для розв'язання складних задач, які неможливо вирішити за допомогою одномодальних систем, що робить вказану систему досить актуальною для наукових досліджень та інновацій.

За результатами дослідження сформовано висновок, що мультимодальні системи штучного інтелекту мають потенціал бути більш зручними для користувачів і надавати споживачам детальне розуміння складних реальних даних у порівнянні з унімодальними системами. Сучасні досягнення у сферах мультимодального представлення інформації та методів управління великими мультимодальними наборами даних розширюють звичний функціонал для розв'язання описаних проблем і надають ширший спектр можливостей сучасного одномодального штучного інтелекту.

Ключові слова: штучний інтелект, мультимодальний, одномодальний, унімодальний, автоматизація, генерація, модальності, звукова обробка, продуктивний дизайн.

ВСТУП

Впровадження штучного інтелекту привело до значних змін у сферах життя, бізнесу та технологій. Штучний інтелект здійснив світову революцію в багатьох галузях, дозволивши компаніям автоматизувати процеси, отримувати цінну інформацію та приймати рішення на основі даних. З урахуванням темпів розвитку цієї технології з'являється нова концепція, яка має потенціал для подальшої трансформації — мультимодальний штучний інтелект.

Його поява є революційною як для працівників цифрових робочих груп, які прагнуть оптимізувати діяльність, так і для клієнтів, які бажають персоналізованої та ефективної взаємодії з організаціями для вирішення складних завдань, наповнених великою кількістю різноманітних даних. Нові технології відкривають нові можливості для працівників, організацій і клієнтів [1].

ПОСТАНОВКА ПРОБЛЕМИ

Дослідження робить комплексний внесок у розвиток мультимодального штучного інтелекту, який має потенціал для розв'язання широкого спектру проблем у таких галузях, як освіта, медицина, бізнес і сфера розваг. Головною проблемою, розгляду якої присвячене дослідження,

постає застосування мультимодального штучного інтелекту в нетипових сферах людської життєдіяльності та практичне дослідження наявних сервісів на основі описаної технології.

Мета статті — аналіз значення мультимодального штучного інтелекту, визначення відмінних рис між одномодальним і мультимодальними системами, опис математичного підґрунтя функціонування технології, аналіз сфер застосування (медицина, освіта, бізнес і сфера розваг), визначення нових можливостей використання мультимодального штучного інтелекту у звуковій обробці та продуктовому дизайні, здійснення порівняльного аналізу під час використання наявних сервісів із мультимодальним штучним інтелектом.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

У сучасній науковій діяльності питання дослідження можливостей мультимодального штучного інтелекту набули значного розвитку. У працях західних і вітчизняних науковців (Е. М. Бендер, Т. Гебру, О. Ледано, Дж. Лоутон, А. МакМіллан-Мейджор, М. МакНейсбі, М. Рауз, М. Роллінс, А. Так'яр, А. Хаген, Ш. Шмітцелл) розглядалися питання переваг застосування

технології, головні принципи роботи, сфери застосування, перспективи розвитку та наявні практичні приклади для використання.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

I. Визначення поняття “мультимодальний штучний інтелект”

Мультимодальний штучний інтелект використовує різні види даних та інформації з різних джерел для розуміння та взаємодії з оточуючим середовищем, подібно до того, як це робить людський мозок. Головна ідея полягає в тому, щоб моделі штучного інтелекту могли обробляти інформацію з таких джерел, як текст, зображення, звук, відео, для здійснення більш широкого та комплексного розуміння ситуації та завдання.

Більшість систем штучного інтелекту сьогодні є одномодальними (унімодальними). Вони розроблені та побудовані для роботи виключно з одним типом даних і використовують алгоритми, що пристосовані для цієї модальності. Наприклад, унімодальна система ChatGPT-3 використовує алгоритми обробки природної мови для розуміння та вилучення сенсу з текстового контенту і єдиним типом вихідних даних, який може створити чат-бот, є текст. Натомість мультимодальні архітектури, які можуть інтегрувати та обробляти декілька модальностей одночасно, мають потенціал для створення більш ніж одного типу вихідних даних [7].

Система з мультимодальним штучним інтелектом може розуміти і спілкуватися декількома “мовами” — форматами даних (текст, візуальні зображення, аудіо, відео). Вона поєднує в собі сильні сторони різних типів моделей штучного інтелекту для обробки різних форматів даних. Наприклад, можна використовувати обробку природної мови для аналізу тексту, комп’ютерний зір для розшифровки зображень і розпізнавання мови для введення аудіо [2].

Мультимодальний штучний інтелект працює через процес тренування і навчання, який передбачає доступ моделі штучного інтелекту до наборів даних, що містять приклади з різних модальностей, наприклад, парні зображення і текстові описи. Окрім того, процес навчання вчить модель: розпізнавати закономірності та асоціації між різними типами даних; пов’язувати те, що бачить на зображенні, зі словами, які його описують. Після такого навчання штучному інтелекту можна дати нову картинку, а він видасть опис того, що зображено на картинці, або ввести декілька слів, а він створить відповідну картинку.

Зазвичай архітектура мультимодального штучного інтелекту складається з трьох ключових компонентів: вхідного модуля, модуля

злиття і вихідного модуля. Модуль введення (як очі та вуха) розглядає різні типи інформації (текст, зображення тощо) і готує їх для мозку штучного інтелекту. Модуль синтезу об’єднує всю інформацію воедино. Він бере те, що побачили/почули очі/вуха, і розумно змішує її. Це може бути як просте складання речей разом, так і використання інноваційних методів, наприклад, звернення уваги на важливі деталі. Модуль виведення дає остаточну відповідь. Після того, як мозок (модуль синтезу) опрацював усю надану інформацію, модуль виводу видає те, що штучний інтелект вважає правильною відповіддю [3].

Мультимодальні моделі покладаються на методи глибокого навчання, які включають складні нейронні мережі. Кодерний рівень моделі перетворює необроблені сенсорні дані у високорівневе абстрактне представлення, яке можна аналізувати та порівнювати в різних модальностях. Шар змішувача вводу/виводу поєднує інформацію з різних сенсорних входів для створення комплексного представлення даних, тоді як шар декодера генерує вихід, який представляє прогнозований результат на основі введених даних [4].

II. Порівняльна характеристика мультимодальної та одномодальної систем

Здійснено порівняльний аналіз різноманітних характеристик одномодальних і мультимодальних систем, результати якого наведено в табл. 1.

III. Математичне підґрунтя для функціонування системи

Мультимодальна модель пропонує більш просунутий, інтелектуальний підхід до вирішення, оскільки працює безпосередньо зі значенням. Наприклад, інформація — людина. Цю інформацію можна передати декількома способами — текстово (слово “людина”), візуально (за допомогою зображення) (рис. 1).

Принцип роботи побудований подібно до text-to-image моделі, що додають модальність тексту для керування процесом створення зображення. Мультимодальний штучний інтелект використовує моделі вбудовування для перетворення тексту та зображень у вектори, які фіксують їх значення. Вбудовані моделі у сфері обробки природної мови використовуються для перетворення слова у векторну форму, що виражає семантичні особливості в просторі чисел. Слова з подібним значенням, контекстом мають схожі векторні представлення. Вбудовані моделі дозволяють представити слова у вигляді векторів із числовими значеннями, які відображають їхнє значення в контексті. Це дає комп’ютерам змогу розуміти семантику слів та їхнє взаємо-

Таблиця 1

Порівняльна характеристика одноmodalьних і мультимodalьних систем

	Одноmodalьні	Мультимodalьні
Обсяг даних	Обробка одного типу даних, таких, як зображення, текст або аудіо	Інтеграція декількох джерел даних, включаючи зображення, текст, аудіо та відео
Складність	Менш складні, оскільки потрібно обробляти лише один тип даних	Складна архітектура для інтеграції та аналізу декількох джерел даних одночасно
Контекст	Зосереджені на обробці одного типу даних, їм бракує контексту і допоміжної інформації, які є вирішальними для створення точних прогнозів	Інтегрують дані з різних джерел і можуть надавати більше контексту та інформації, що приводить до більш точних прогнозів
Продуктивність	Можуть добре виконувати завдання, що пов'язані з конкретною сферою, можуть мати проблеми в процесі вирішення завдань, що вимагають ширшого розуміння	Інтегрують декілька джерел даних і можуть запропонувати більш комплексний і нюансований аналіз, що приводить до більш точних прогнозів
Вимоги до даних	Для ефективного навчання потрібні великі обсяги даних, оскільки покладаються на один тип даних	Навчаються на більших обсягах даних, оскільки інтегрують із джерел, що є більш надійними і адаптивними

відношення, що є важливим для ефективної обробки природної мови в алгоритмах машинного навчання.

Інакше кажучи, це різновид методу вивчення особливостей у машинному навчанні, коли високорозмірні дані перетворюються на низькорозмірні вектори зі збереженням релевантної інформації. Цей процес зменшення розмірності допомагає спростити дані та полегшити їхню обробку алгоритмами машинного навчання [5].

Кожна координата вектора відображає певний аспект семантики. Вбудовані моделі намагаються знаходити такі векторні представлення слів, щоб слова зі схожим семантичним значенням були близькими до вираження зображення

у векторному просторі. Вектори надають можливість визначити ступінь схожості між словами. Схожі вектори вказують на семантичну близькість. Наведемо приклад: три текстові значення “кошик зелених яблук”, “пухнастий білий кіт”, “червона ручка” та зображення, що представляють значення. Для тексту введено вектори T_a , T_c , T_p , для зображень — I_a , I_c , I_p (рис. 2).

Схожість між векторами можна визначити за допомогою косинусної схожості. Косинусна схожість між двома векторами визначається як косинус кута між ними. Якщо вектори мають близькі напрямки, то їхня косинусна схожість велика, а якщо вектори вказують у різні напрямки, то схожість менша. У результаті можна

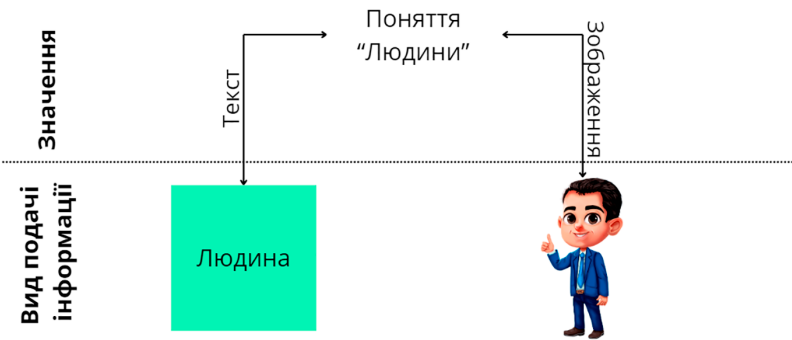


Рис. 1. Два способи подання поняття “людина”

отримати числові значення, які вказують на ступінь схожості між текстом і зображенням. Чим вище значення, тим більша схожість між відповідними векторами, що може вказувати на семантичну близькість між текстовими та візуальними представленнями об'єктів (рис. 3).

Мультимодальна модель призначена для одночасної обробки декількох форм сенсорного введення, подібно до того, як людина пізнає світ, на відміну від традиційних одномодальних систем штучного інтелекту, що навчені виконувати певне завдання на основі однієї вибірки даних. Навчання передбачає об'єднання розрізнених даних, зібраних із різних датчиків і вхідних даних, в єдине ціле (рис. 4).

IV. Аналіз можливих сфер застосування мультимодального штучного інтелекту

Під час аналізу іноземних джерел було виділено основні "класичні" сфери застосування

описаної технології, а саме такі — медицина (аналіз медичних зображень для поєднання рентгенівських знімків, комп'ютерної томографії та даних про пацієнта, персоналізована медицина — генетика, дані про спосіб життя та історію хвороби, роботизована хірургія для поєднання візуальних і сенсорних даних); освіта (персоналізовані навчальні матеріали, регулювання темпу і типу подачі контенту, можливість отримання ефективного зворотного зв'язку); бізнес та розповсюдження (швидка обробка різного типу інформації, автономізація рутинних і ресурсозатратних процесів, вирішення складних клієнтських кейсів, точні та персоналізовані рекомендації, доступність 24/7, віртуальна візуалізація); сфера дозвілля (доступні та зручні додатки, сервіси на основі мультимодального штучного інтелекту, створення персоналізованого контенту, розширення можливостей).

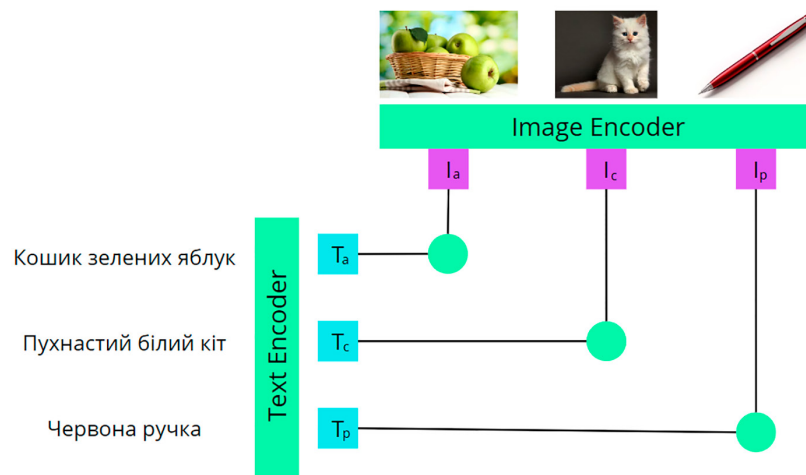


Рис. 2. Відповідність текстових значень і зображень

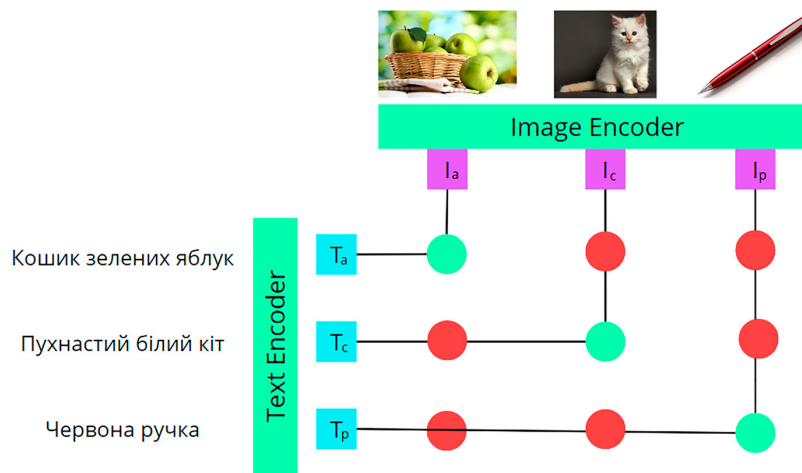


Рис. 3. Матриця косинусної схожості текстів та зображень (зелений — висока схожість, червоний — низька схожість)

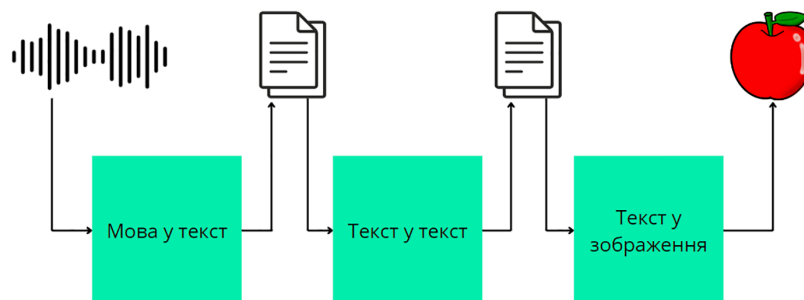


Рис. 4. Алгоритм передачі інформації під час машинного навчання

Мультимодальний генеративний штучний інтелект може допомогти з фінансовими даними, профілями клієнтів, статистикою магазинів, географічною інформацією, пошуковими тенденціями та маркетинговою інформацією — усе це зберігається в різних формах, включаючи зображення, графіки, текст, голос і діалоги. Мультимодальний штучний інтелект може автоматично знаходити зв'язки між різними наборами даних, що представляють такі об'єкти, як клієнти, об'єкти та процеси [8].

Варто виокремити дві не такі явні сфери застосування — звукова обробка та продуктивний дизайн. Одномодальний штучний інтелект уже здійснив часткову революцію в цих двох напрямках, проте в цій статті розглянуто те, яким чином переваги мультимодального штучного інтелекту можуть відобразити принципово новий метод до роботи. Головні переваги полягають у можливості роботи з різними типами інформації, персоналізації контенту, автоматизації, покращенні аналізу даних.

Мультимодальний штучний інтелект, який поєднує аналіз звуку з іншими типами даних, може революціонізувати сферу обробки звукової інформації.

Можливість роботи з різними типами інформації вплине на якість розпізнавання мови (використання візуальної інформації, наприклад, рухів губ, для розпізнавання в умовах шуму, аналіз для розпізнавання емоцій, намір мовця). Персоналізація контенту вплине на можливість персоналізації звукового досвіду (адаптація звучання створеного запису до унікальних комбінацій залежно від текстового чи візуального запиту, створення 3D-аудіо, яке враховує акустику середовища користувача під певний запит із використанням візуальної або текстової інформації). Автоматизація вплине на можливість створення звукових ефектів. Зокрема це може бути автоматичне створення звукових ефектів (генерація звукових ефектів на основі візуальних сцен у фільмах, іграх, адаптація звукових ефектів до емоційного контексту контенту).

Концепція мультимодального штучного інтелекту у сфері продуктового дизайну полягає в застосуванні мультисенсорного підходу. Традиційний дизайн продукту фокусується на тому, як продукт виглядає та відчувається, використовуючи візуальні елементи та ергономіку. Мультимодальний підхід додає ще один рівень, аналізуючи дані користувацького досвіду, які охоплюють різні органи чуття (вираз обличчя, жестикуляція, голос). Аналіз цих даних за допомогою штучного інтелекту допомагає створювати продукти, які залучають більше, ніж просто зір і дотик.

Також варто враховувати використання цієї технології в процесі створення продуктів. У цьому контексті варто згадати персоналізацію (створення інтерактивних продуктів, які реагують на дії користувача, поєднання різних модальностей для кращої передачі ідеї замовнику) і автоматизацію (використання нових інструментів у ПЗ, генерація прототипів, тестування та створення технічної документації з метою звільнення часу на виконання рутинних завдань).

V. Ризики та недоліки

Після аналізу широкого спектру переваг і можливостей, що відкриваються перед користувачами для роботи, варто окреслити певну кількість недоліків і ризиків, що можуть спричинити ненормоване та неконтрольоване зловживання застосування сервісу на основі мультимодального штучного інтелекту.

Упередженість даних, ризик зловживання інформацією, створення фейків, технічні ризики, обмеження інтерпретації, проблеми масштабування сервісу — усе це можливі неконтрольовані наслідки. Усунення ризиків вимагає цілісного підходу, що передбачає етичні міркування, надійні технічні рішення, а також постійний моніторинг і оцінку мультимодальних систем штучного інтелекту протягом усього їх життєвого циклу. Розробники, освітні установи та політики мають працювати разом, щоб пом'якшити ці потенційні проблеми та забезпечити відповідальне розгортання штучного інтелекту [10].

VI. Практична частина дослідження

Для практичної частини дослідження було обрано онлайн-ресурси, що використовують мультимодальну систему. На основі вхідних даних було виконано об'єднання та перетворення в інший вид подачі інформації та отримані проміжні результати. На основі виконаного практичного дослідження було сформульовано висновки, сформовано основні особливості кожної обраної системи, мету її застосування (табл. 2).

ВИСНОВКИ

Згідно з аналітичними даними провідних світових компаній, мультимодальна модель штучного інтелекту є глобальним викликом і тенденцією розвитку. Оскільки моделі використовують синергію різних типів даних, щоб отримати більш докладну інформацію та вирішувати складніші завдання. Наприклад, мультимодальний інтелект може одночасно аналізувати зображення та текст, щоб краще розуміти

зміст, що допомагає створювати ефективні та контекстно орієнтовані системи.

Математична модель мультимодального штучного інтелекту складається з різних компонентів, які об'єднують інформацію з різних джерел. Ключова ідея полягає в тому, що кожен тип даних проходить через свій власний модуль обробки. Наприклад, для тексту — це модель обробки природної мови, для зображень — згорткові нейронні мережі, а для звуку — аудіорозпізнавання.

Математична природа векторів дає змогу ефективно використовувати їх у численних задачах обробки природної мови та розуміння семантики тексту. У контексті вбудованих моделей використовуються векторне представлення слів, де кожна координата вектора відображає певний аспект семантики чи контексту слова. Вбудовані моделі намагаються знаходити такі векторні представлення слів, щоб слова зі схожим семантичним значенням були близькими одне до одного у векторному просторі.

Таблиця 2

Результати практичного дослідження деяких мультимодальних ресурсів

Назва онлайн-ресурсу	Модальність	Особливості	Застосування
CLIPBERT	Текст, зображення, відео	Попередньо навчені кодери зображень і тексту, мультимодальні нейрони	text-to-image, video-to-text
GLIDE	Текст, зображення, аудіо	Використовує модель дифузії (класом моделей прихованих змінних) для генерації зображень, ранжує зображення	text-to-image
VLMo	Бачення (vision), текст	Модульна трансформаторна мережа зі спільним самоконтролем	Візуальна відповідь на питання
METER	Бачення (vision), текст	Загальна основа для трансформаторів мови бачення	Візуальна відповідь на питання
SwinBERT	Текст, відео	Використовує архітектуру Swin Transformer для відеозадач	Відповіді на відеозапитання, пошук відео, субтитрування відео
VIOLET	Текст, відео	Масковане візуально-знакове моделювання	Відеовідповіді на запитання, пошук відео
ALIGN	Бачення (vision), текст	Використовує зашумлені дані альтернативного тексту для навчання окремих кодерів	Пошук за допомогою мультимодальних систем
ClipBERT	Текст, відео	Поєднання CNN і трансформаторної моделі, наскрізне	Завдання з відеомови: відповіді на запитання, пошук, субтитрування

Мультимодальні системи штучного інтелекту мають потенціал бути більш зручними для користувачів, аніж одномодальні системи, і надавати споживачам детальне розуміння складних реальних даних. Постійні дослідження та досягнення в таких сферах, як мультимодальне представлення, методи злиття та управління великими мультимодальними наборами даних, допомагають розв'язувати ці проблеми та розширювати межі можливостей сьогоdnішнього одномодального штучного інтелекту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *McNasby M.* How Multimodal Capabilities Can Revolutionize AI Models. Discover the impact of multimodal AI on marketing and content strategies. [Electronic resource] / M. McNasby // Spisework. — January 29, 2024. — Access mode: <https://www.spiceworks.com/tech/artificial-intelligence/guest-article/leveraging-multimodal-capabilities-to-revolutionize-ai-models/>.
2. What is Multimodal AI? Combining Tools for Business Impact [Electronic resource] // Pecan. — December 28, 2023. — Access mode: <https://www.pecan.ai/blog/what-is-multimodal-ai-business/#:~:text=Multimodal%20AI%20relies%20on%20combining,need%20to%20be%20seamlessly%20integrated.>
3. *Takyar A.* Multimodal Models: Architecture, workflow, use cases and development [Electronic resource] / A. Takyar // Leeway Hertz. — Access mode: <https://www.leewayhertz.com/multimodal-model/>.
4. *Anglen J.* The Future of AI: How Multimodal Models Are Leading the Way [Electronic resource] / J. Anglen // Rapid Innovation. — Access mode: <http://surl.li/zhroux>.
5. Embeddings in Machine Learning: Types, Models, and Best Practices [Electronic resource] // Swimm. — Access mode: <https://swimm.io/learn/large-language-models/embeddings-in-machine-learning-types-models-and-best-practices>.
6. Растрові зображення (людини, кошика з яблуками, червоної ручки, білого кота, червоного яблука) [Електронний ресурс] // Freepik. — Режим доступу: <https://www.freepik.com/>.
7. *Rouse M.* Multimodal AI (Multimodal Artificial Intelligence) [Electronic resource] / M. Rouse // Techopedia. — July 4, 2023. — Access mode: <https://www.techopedia.com/definition/multimodal-ai-multimodal-artificial-intelligence>.
8. *Lawton G.* Explore real-world use cases for multimodal generative AI [Electronic resource] / G. Lawton // TechTarget. — February 12, 2024. — Access mode: <https://www.techtarget.com/searchenterpriseai/feature/Explore-real-world-use-cases-for-multimodal-generative-AI>.
9. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? [Electronic resource] / E. M. Bender, A. McMillan-Major, T. Gebru, Sh. Shmitchell //

Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery. — 2021. — P. 610–623. DOI: <https://doi.org/10.1145/3442188.3445922>.

10. *Mark Rollins M.Sc., B. Sc., Cert. Ed, PGDip.* Multimodal Generative AI in Education [Electronic resource] / Mark Rollins M. Sc., B. Sc., Cert. Ed, PGDip // LinkedIn. — 2023. — Access mode: <https://www.linkedin.com/pulse/multimodal-generative-ai-education-mark/>.

REFERENCES

1. McNasby, M. (2024). How Multimodal Capabilities Can Revolutionize AI Models. Discover the impact of multimodal AI on marketing and content strategies. *Spisework*. Retrieved from: <https://www.spiceworks.com/tech/artificial-intelligence/guest-article/leveraging-multimodal-capabilities-to-revolutionize-ai-models/>.
2. (2023). What is Multimodal AI? Combining Tools for Business Impact. *Pecan*. Retrieved from: <https://www.pecan.ai/blog/what-is-multimodal-ai-business/#:~:text=Multimodal%20AI%20relies%20on%20combining,need%20to%20be%20seamlessly%20integrated.>
3. Takyar, A. Multimodal Models: Architecture, workflow, use cases and development. *Leeway Hertz*. Retrieved from: <https://www.leewayhertz.com/multimodal-model/>.
4. Anglen, J. The Future of AI: How Multimodal Models Are Leading the Way. *Rapid Innovation*. Retrieved from: <http://surl.li/zhroux>.
5. Embeddings in Machine Learning: Types, Models, and Best Practices. *Swimm*. Retrieved from: <https://swimm.io/learn/large-language-models/embeddings-in-machine-learning-types-models-and-best-practices>.
6. Raster images (of a person, a basket of apples, a red pen, a white cat, a red apple). *Freepik*. Retrieved from: <https://www.freepik.com/>.
7. Rouse, M. (2023). Multimodal AI (Multimodal Artificial Intelligence). *Techopedia*. Retrieved from: <https://www.techopedia.com/definition/multimodal-ai-multimodal-artificial-intelligence>.
8. Lawton, G. (2024). Explore real-world use cases for multimodal generative AI. *TechTarget*. Retrieved from: <https://www.techtarget.com/searchenterpriseai/feature/Explore-real-world-use-cases-for-multimodal-generative-AI>.
9. Bender, E. M., McMillan-Major, A., Gebru, T., & Shmitchell, Sh. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery*, P. 610–623. Retrieved from: <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>.
10. Rollins, Mark M. Sc., B. Sc., Cert. Ed, PGDip. (2023). Multimodal Generative AI in Education. *LinkedIn*. Retrieved from: <https://www.linkedin.com/pulse/multimodal-generative-ai-education-mark/>.

O. A. BOBARCHUK, PhD in Engineering
T. V. ZLOTKIVSKA, Student

INNOVATIVE POSSIBILITIES OF DEVELOPING ARTIFICIAL INTELLIGENCE USING A MULTIMODAL APPROACH

Abstract. The publication is devoted to the study of multimodal artificial intelligence in comparison with unimodal systems. The paper discusses the mathematical aspects of multimodal systems, conclusions on the potential and advantages over unimodal systems.

Multimodal artificial intelligence has a significant potential to improve people's lives in areas such as medicine, education, business, and entertainment. Research into new and innovative applications is highly relevant. This technology is used to improve the efficiency and accuracy of existing systems. Multimodal artificial intelligence can be used to solve complex problems that cannot be solved by unimodal systems, which makes this system highly relevant for research and innovation.

The study concludes that multimodal AI systems have the potential to be better suited to users than unimodal systems and provide a detailed understanding of complex real-world data to consumers. The ongoing research and advancements in the fields of multimodal representation, fusion methods, and management of large multimodal datasets are contributing to solving these challenges and extending the capabilities of the current unimodal AI.

Keywords: artificial intelligence, multimodal, unimodal, unimodal, automation, generation, modalities, sound processing, product design.

ІНФОРМАЦІЯ ПРО АВТОРІВ

Бобарчук Олександр Антонович — канд. техн. наук, Національний авіаційний університет, просп. Любомира Гузара, 1, Київ, Україна, 03058; a.bobarchuk@gmail.com; ORCID: 0000-0003-3176-7231

Злотківська Тетяна Володимирівна — студентка, Національний авіаційний університет, просп. Любомира Гузара, 1, Київ, Україна, 03058; 7395409@stud.nau.edu.ua; ORCID: 0009-0009-0661-8956

INFORMATION ABOUT THE AUTHORS

Bobarchuk O. A. — PhD in Engineering, National Aviation University, 1, Lubomyr Huzar Ave., Kyiv, Ukraine, 03058; a.bobarchuk@gmail.com; ORCID: 0000-0003-3176-7231

Zlotkivska T. V. — Student, National Aviation University, 1, Lubomyr Huzar Ave., Kyiv, Ukraine, 03058; 7395409@stud.nau.edu.ua; ORCID: 0009-0009-0661-8956



ДО УВАГИ АВТОРІВ:

До друку приймаються статті українською та англійською мовами.

Відповідальність за достовірність поданих даних несуть автори матеріалів.

Редакція може не поділяти думки авторів, викладені у статтях.

У разі передруку матеріалів — посилання на журнал "Наука, технології, інновації" обов'язкове.

Адреса редакції: вул. Антоновича, 180, м. Київ, Україна, 03150.

Контакти редакції: тел.: +38 (044) 521-00-39.

e-mail: journal@uintei.kiev.ua

Умови для публікації викладено на сайті: <http://nti.ukrintei.ua>.

З питань придбання та розміщення реклами: тел. +38 (044) 521-00-39.

e-mail: uintei.ua@gmail.com або sale@uintei.kiev.ua