

<https://doi.org/10.15407/csc.2024.04.010>  
UCD 514.18

**В.Д. МІНЕНКО**, засновник, Twigames Inc.,  
3422 Old Capitol Trail, Suite# 241, Wilmington, DE 19808, US,  
Вілмінгтон, штат Делавер, США,  
ORCID: <https://orcid.org/0009-0003-5299-6786>,  
[valerii@twigames.net](mailto:valerii@twigames.net)

## **АНАЛІЗ ЗАСТОСУВАННЯ ШІ-ГЕНЕРАТОРІВ ДЛЯ РОЗВ'ЯЗАННЯ СКЛАДНИХ БІЗНЕС-ЗАДАЧ**

Стаття присвячена аналізу поточного стану розвитку технологій ШІ-генерації візуального контенту, наявних моделей, їхніх функціональних можливостей, переваг, обмежень і недоліків. Мета дослідження полягає, з одного боку, у виявленні місця ефективного застосування технологій ШІ-генерації у вирішенні складних бізнес-задач, а з іншого — у виокремленні проблем наявних моделей генеративного штучного інтелекту з огляду на подальшу інтеграцію їх у ланцюжок реалізації бізнес-процесу та визначення можливих шляхів їх вирішення.

**Ключові слова:** генеративний штучний інтелект, обробка природної мови, ШІ-генерація, нейронна мережа, нейромережева модель, модель text-to-text, модель text-to-image, модель text-to-video, генеративна змагальна мережа, машинне навчання.

### **Вступ**

Останніми роками спостерігається значне зростання кількості застосунків генеративного штучного інтелекту (ШІ), що створюють візуальні дані (зображення чи відео) на основі текстових описів. Це відкриває нові можливості в удосконаленні інформаційних систем у різних прикладних галузях, і це не лише яскраві форми представлення чи виведення інформації, а й можливість вирішення складних практичних задач на зовсім іншому рівні. А можливості збереження історії візуалізацій за темою, та візуалізації даних “не з 0”, а з урахуванням попередньої версії, які надаються вже багатьма наявними застосунками генерації зображень та відеокон-

тенту, дають змогу відслідковувати динаміку змін певної інформації в часі, що може забезпечити глибину аналізу інформації та інший рівень прийняття рішень і формування стратегій розвитку в складній бізнес-системі.

Але, слід зауважити, що попри досить тривалу історію розвитку галузі штучного інтелекту, що бере початок від перших експертних систем, ШІ-генерація зображень та відео виникла досить нещодавно. Її стрімкий розвиток спостерігається лише за останні кілька років та на відміну від впевненого покращення технічної якості генерованого візуального контенту (розмір, роздільна здатність тощо), достовірність вхідної інтерпретації та семантична якість генерації досі залишається проблемою, що ви-

Cite: Міненко В.Д. Аналіз застосування ШІ-генераторів для розв'язання складних бізнес-задач. *Control Systems and Computers*, 2024, 4, 10—18. <https://doi.org/10.15407/csc.2024.04.010>

© Видавець ВД «Академперіодика» НАН України, 2024. Стаття опублікована на умовах відкритого доступу за ліцензією CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

значається не лише вибором самої моделі генерації, а й точністю вхідних інструкцій. Заміна одного слова може кардинально змінити результат генерації, а, отже, і рішення задачі.

Цю роботу присвячено аналізу поточного стану розвитку технологій ШІ-генерації візуального контенту, вже створених рішень, їхніх функціональних можливостей, переваг та недоліків, та має на меті виокремлення наявних проблем в аспекті подальшого застосування в ланцюжку реалізації бізнес-процесу та визначення можливих шляхів подолання цих проблем.

## Генеративний штучний інтелект. Методи створення зображень та відео

Як уже було зазначено, всі системи ШІ-генерації отримують на вході текстовий опис, який мають перетворити на новий контент потрібного формату (текст, зображення чи відео) для досягнення мети вирішення задачі. Отже, всі застосунки такого типу вимагають вирішення задачі обробки природної мови (NLP). Якісно нові можливості обробки природної мови відкривають Великі мовні моделі (LLM), які працюють на основі штучного інтелекту, що використовує методи глибокого навчання.

До їхніх ключових характеристик зараховують [1, 2]:

- **Токенізація** [2] як важливий етап попередньої обробки під час навчання LLM, що розділяє текст на нерозкладні одиниці (токени). Це можуть бути символи, підслова, слова, залежно від застосованого процесу токенизації.

- **Трансформерна архітектура** (архітектура глибокого навчання) дає змогу ефективно (паралельно та незалежно одне від одного) обробляти послідовності вхідних даних. Трансформер використовує механізм уваги для взаємодії з різними частинами вхідних даних та забезпечує привертання уваги до токенів у тексті. Слід зазначити, що інтегрованим компонентом трансформаторів є нормалізація.

- **Навчання.** Зазвичай, LLM вимагають масштабного розподіленого навчання [2] на величезних обсягах текстових даних. Це дає змогу

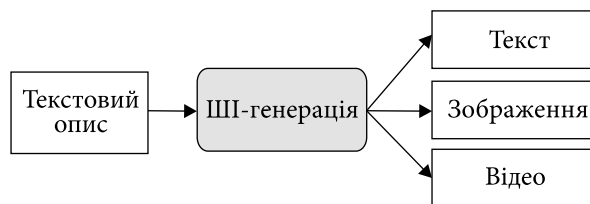


Рис.1. Класифікація систем ШІ-генерації за отриманим контентом

моделі засвоювати широкий спектр лінгвістичних особливостей і зв'язків між словами та лексемами.

- **Механізм генерації** нового мовного контенту на базі набутих знань.

- Механізми **зосередження (self-attention) та уваги (attention)** дають акцентувати увагу моделі на тих чи інших токенах за допомогою приписування певної ваги відповідним вхідним токенам на підставі їхньої значущості.

- **Функції активації** [2] відіграють вирішальну роль у підгонці кривих нейронних мереж.

- Механізми **контролю сприйняття** дають змогу користувачам впливати на згенерований контент за допомогою регулювання параметрів моделі або додаткових вказівок.

- **Кодувальний та декодувальний шари** моделі відповідають за обробку, відповідно, вхідної та вихідної (виведеної) інформації.

Таким чином, LLM є сучасним рішенням для задач, які вимагають розуміння та генерації тексту. Головна відмінність систем генеративного штучного інтелекту від великих мовних моделей полягає в тому, що вони можуть створювати нові дані будь-якого типу, тобто генерувати нові форми контенту.

Умовно вирізняють три основні групи за формою отриманого контенту (*text-to-text*, *text-to-image* та *text-to-video*).

Спільним для них є те, що в якості вхідних даних вони використовують текстову підказку, на основі якої і здійснюється генерація нового контенту, а також застосування тих чи інших технологій машинного навчання для реалізації. Розглянемо ці класи генерації трохи детальніше.

Задача генерації тексту (*text-to-text*) має на меті створення нового текстового опису, відпо-

відного до вхідних текстових інструкцій. У літературі використовується також поняття-синонім “генерування природної мови” (NLG). Для розв’язання задач даної групи часто використовуються процеси Маркова чи глибокі генеративні моделі.

Процес генерації тексту, зазвичай, можна розділити на три етапи:

- вибір контенту — мовна модель описує зміст та обирає інформацію для перетворення на стандартизовану форму, наприклад, групує пов’язані фрази чи створює таблицю з позначеними стовпцями;
- планування речень — розділення контенту на абзаци та речення, заміна частин мови, наприклад, використання займенників та сполучників;
- реалізація речення — перевірка граматичної точності та уточнення послідовності слів.

Тобто, штучний інтелект або моделі машинного навчання генерують мову відповідно до правил граматики, синтаксису чи вибору лексики. Модель генерації починається на концептуальному рівні (вибір контенту), потім зменшується до досконалих правил мови, таких як коректне написання, грамика та вибір слів (планування речення), і, нарешті, створює ланцюжок слів у формі речення (реалізація речення).

Технологія III-генерації зображення з тексту (*text-to-image*) використовує штучний інтелект для створення візуальних представлень на основі текстових описів. Простіше кажучи, це штучний інтелект, який має/вміє “прочитати” фрагмент тексту та створити зображення, яке досить точно відобразить його зміст і контекст.

Зазвичай, цей процес включає три основні етапи:

1) “Читання” та аналіз введеного тексту. III ідентифікує та витягує критично важливу інформацію, таку як ключові слова, поняття та контекст (зазвичай, для цього використовуються методи обробки природної мови NLP).

2) Синтез зображення: III, використовуючи алгоритми дескриптивних логік (DL), створює зображення на основі інформації, що отрима-

на в результаті аналізу вхідного тексту. Даний етап, зазвичай, включає генеративні змагальні мережі (GAN). GAN [3] є специфічним класом алгоритмів III, які використовуються у навчанні без учителя та реалізовані системою двох штучних нейронних мереж: генератора та дискримінатора, які працюють разом (змагаються) для створення реалістичних зображень. Генератор створює зображення на основі вхідного тексту, тоді як дискримінатор оцінює згенеровані зображення та надає зворотний зв’язок генератору.

3) Удосконалення та оптимізація полягає в точному налаштуванні згенерованих зображень для забезпечення більшої точності та візуальної привабливості. Система III може переглядати кілька версій зображення, удосконалюючи та оптимізуючи його, поки воно не досягне бажаного рівня якості та деталізації.

Отже, більшість сучасних сервісів генерації *text-to-image* є поєднанням моделі, що обробляє природну мову (NLP) та генеративно змагальної мережі GAN.

Наявні технології перетворення тексту на відео (*text-to-video*) [5] також працюють на базі штучного інтелекту і, зазвичай, охоплюють кілька його підгалузей: обробку природної мови, комп’ютерний зір та машинне навчання.

Технологія T2V (*text-to-video*) дає змогу створювати відео-ролики на підставі підказки. Стрімкий розвиток, який останнім часом відбувається у цій галузі, що спричинило суттєве покращення якості візуалізацій, здебільшого пов’язаний з розвитком дифузійних моделей (SD). T2V, заснована на дифузійній моделі, розуміє значення введеного користувачем тексту та генерує відповідні візуальні ефекти й анімацію. Вхідні тексти можуть складатися з величезної кількості слів. Додавання/видалення слів в/з опис(у) здатне кардинально змінити результат перетворення, бо слова мають ключову роль у створенні відео.

Суть алгоритму T2V полягає в наступній послідовності кроків. Спочатку система аналізує та інтерпретує вхідний текст, застосовуючи методи токенизації, визначає його контекст і значення. Потім система обирає відповідні ві-

зуальні ефекти й анімацію і планує на основі підказки відеоконтент. Наступним кроком, алгоритм створює візуальні елементи на кшталт 3D-моделей або анімації. Для цього можуть бути використані GAN моделі, або ці елементи можуть бути витягнуті просто з наявної бібліотеки відеоматеріалів. Нарешті, ШІ з отриманих візуальних об'єктів будує послідовність, яка відповідає вхідному тексту, додаючи переходи та, можливо, синхронізуючі їх зі звуком.

Слід зазначити, що існує ще дуже багато невирішених задач та обмежень, навіть технічних. Так, наприклад, досі залишається серйозною проблемою генерація руху між відеокадрами.

## Аналіз наявних сервісів

Сервіси *text-to-text* — це так звані генератори тексту, або, простіше, інструменти перетворення тексту на текст.

Найвідомішими на сьогодні є, розроблені OpenAI [6], моделі серії GPT (*Generative Pretrained Transformer*), такі як GPT-4, GPT-3, GPT-2. Вони використовують трансформерну архітектуру для попереднього тренування на великому обсязі даних та генерації тексту. GPT-3, зокрема, має 175 мільярдів параметрів.

Слід також виокремити нейромережеву модель *StableLM*, засновану на моделях *Transformer* та GPT. Це доволі потужна та ефективна модель для генерації тексту природною мовою, головною її перевагою є стабільність її внутрішнього стану та можливість генерувати тексти високої якості з мінімальними витратами на навчання. На *Github* представлено серію доступних на цей час мовних моделей *StableLM*, що постійно доповнюється. Некомерційні версії моделі для особистого використання та дослідження є безкоштовними. Так, модель із відкритим кодом *StableLM-3B-4E1T* [5] досягає найсучаснішої продуктивності. Вона використовує три мільярди параметрів та є попередньо-навченою в багато-епохальному режимі на наборах даних великого масштабу з відкритим кодом. Це застосунок на базі штучного інтелекту з відкритим кодом, який може стати безкоштовною альтернативою *ChatGPT*.

Щодо генераторів зображень та відео [8]: попри те, що ця галузь ШІ виникла досить нещодавно, на сьогодні вже є чимало застосунків. Вони відрізняються не лише ціною та якістю результуючої візуалізації, а й мисленням (зокрема, мисленням штучної нейронної мережі). У [7] було здійснено порівняльний аналіз десяти найкращих (на погляд авторів) нейромереж за цілою низкою критеріїв, що стосуються і функціональності, і якості зображення. Найбільшу кількість балів набрала нейромережа *Midjourney*, як у плані функціональності, так і у плані якості генерацій. *Midjourney* дозволяє швидко та якісно працювати з генеративним контентом, але вона є доволі дорогою (має найбільшу вартість серед усіх проаналізованих) і доступ до нейромережі відкритий лише через сервер *Discord*, де вводиться текстовий запит.

Далі розглядається декілька найвідоміших альтернативних безкоштовних нейромереж *text-to-image*.

*DALL-E-2* — це перша нейронна мережа, що працює за моделлю *text-to-image*, представлена компанією OpenAI. Це доволі потужний генератор зображень з тексту, побудований на комбінації *ChatGPT* (що виступає в якості трансформера) та варіації генеративно змагальної мережі *VQ-GAN*. Така інтеграція дає змогу будувати цікаві текстові запити та опрацьовувати їх відео-зміни для генерації власних зображень чи маніпулювання наявними. Забезпечує досить непогану якість зображень. Підтримує велику кількість мов, включно з українською. Остання версія моделі *DALL-E-2* здатна змішувати різні концепції з двох зображень в одному. Вона є офіційно безкоштовною, при чому з нескінченною кількістю генерацій, але це є умовним через інтеграцію з *ChatGPT*. Крім цього, зараз *DALL-E 2* доступна розробникам тільки на спеціальній платформі, на яку можна потрапити лише за запрошенням OpenAI.

Друга професійна нейромережа класу *text-to-image* — *Imagen*, запропонована Google, працює за схожим принципом. Порівняння показало, що *Imagen* має перевагу перед *DALL-E 2* і в якості зображень, і у змістовій відповідності між описом та зображенням.

*Stable Diffusion* — III генератор, який надає можливість використання готових шаблонів. Він відомий тим, що має відкритий код. Але базова модель (безкоштовна версія застосунку) є доволі обмеженою за кількістю генерацій та картинок, і якість генерації в ній не є досконалою. Нова версія — *Stable Diffusion XL* — здатна генерувати зображення зі швидкістю друку. Дозволяє налаштовувати кількість кроків обробки, точність відповідності текстовому запиту та інші параметри генерації, домальовувати готові зображення, а не тільки створювати картинку з нуля, додавати нові деталі тощо. Ця версія доступна за некомерційною дослідницькою ліцензією, що обмежує її використання особистими некомерційними цілями.

*NightCafe Studio* — один із найпопулярніших генераторів зображень у світі. Він привертає увагу тим, що інтегрує в собі одразу декілька методів генерації зображень, а саме: *Stable Diffusion*, *DALL-E 2*, *CLIP-Guided Diffusion*, *VQGAN+CLIP* і *Neural Style Transfer*, та надає користувачеві можливість вибору найвідповіднішого. Застосунок добре розуміє українську мову. Втім, безкоштовно *NightCafe* генерує тільки п'ять наборів зображень щодня. Ще один недолік безкоштовної версії — досить обмежений розмір отриманої картинки. Для використання ширших професійних можливостей (відсутність реклами, розширений пошук, нормальний розмір зображень тощо) необхідна передплата.

*VQGAN+CLIP*, що вже згадувався, фактично є поєднанням двох окремих алгоритмів машинного навчання, які можуть бути використані разом для створення зображення на базі текстового запиту чи підказки. *VQGAN* — це генеративно-змагальна нейронна мережа, яка доволі непогано генерує зображення, що схожі на інші (але не за текстовим запитом), а *CLIP* — інша нейромережа, яка здатна визначати, наскільки підказка відповідає зображенню. Окрім того, що можна отримати доступ до *VQGAN+CLIP* з *NightCafe Studio*, є можливість його використання в хмарному середовищі програмування *Google Colab*. *Google Colab* дозволяє запускати розроблений на *Python* код на серверах, які мають доступ до швидких процесорів

(*GPU*). Це є дуже важливим, оскільки *VQGAN+CLIP* (і машинне навчання загалом) потребує чималої процесорної потужності.

Генерація відео-контенту за текстовим описом (*text-to-video*) — це напрям машинного навчання, який є логічним продовженням досліджень, що стосуються проблеми генерації зображень. Цей напрям виник не так давно, але протягом останніх двох років у цій галузі спостерігається шалений прогрес.

Генерація відео є складнішою задачею, ніж генерація статичного зображення, тому що крім самої картинки треба передбачити, як вона змінюватиметься у часі. У 2022 р. компанії *Meta* та *Google* майже одночасно представили одразу три інноваційних алгоритми.

Інноваційна мережа *Make-A-Video*, представлена компанією *Meta*, здатна генерувати відео за текстовим описом. Вона працює за принципом *DALL-E 2*, але замість статичних зображень генерує невеликі (не більше 5 секунд) відео-роліки без звуку, але з купою підказок. Система була розроблена для «оживлення» зображень. Нейромережа використовує зображення з текстовими описами, як джерело знань про те, як виглядає світ і як його часто описують, а також — відео без позначок, щоб дізнатися, як світ рухається. Завдяки цим даним *Make-A-Video* надає можливості створювати відео на підставі текстових описів, додавати динаміку до статичних зображень та отримувати нові інтерпретації вже наявного відео. За аналогічним сценарієм працюють і моделі *Phenaki* (безкоштовна) та *Imagen Video*. Але, слід зазначити, що якість генерації всіх трьох моделей є доволі низькою.

На початку 2024 р. компанія *OpenAI* представила *text-to-video* модель *Sora* (на сьогодні безкоштовна), що є розвитком попередніх досліджень *DALL-E* та *GPT*. Це дифузійна генеративна модель, що створює досить якісне відео за його текстовим описом. Процес починається з випадкового шуму, який поступово перетворюється на зображення. *Sora* використовує архітектуру трансформера, як і інші моделі *GPT*. Але замість tokenів, базових текстових блоків, що використовуються в мовних моделях, візу-

альна інформація в *Sora* розбивається на фрагменти, які містять дані про частину кадру, та як вона змінюється в часі (просторово-часові патчі). Модель отримує вхідні зашумлені патчі разом з текстовими інструкціями щодо візуального контенту, потім намагається передбачити вихідні патчі, що відповідають вхідному тексту. Модель генерує відео роликів тривалістю до 1 хвилини. Окрім цілковито нового контенту, що отримується тільки за текстовими інструкціями, *Sora* вміє вносити зміни в наявні відео-ролики, додавати до них нові кадри, змінювати напрямок відтворення та поєднувати один з одним, а також генерувати статичні зображення або перетворювати готові зображення на анімацію. Модель вміє точно інтерпретувати текстові підказки та розуміє, як описані об'єкти існують в фізичному світі. *Sora* надає можливість створювати складні сцени з кількома персонажами, з точними деталями об'єкта та фону, а також певними типами рухів.

Також, слід відзначити безкоштовну *text-to-video* модель з відкритим кодом *Zeroscope AI* від *Cerspense*. Ця модель є удосконаленням безкоштовної моделі з відкритим кодом *Modelscope* та спирається на її фундамент, забезпечуючи більш вишуканий і професійний досвід створення відео. Фактично це перша високоякісна модель перетворення тексту на відео з відкритим кодом. Її навчання передбачало впровадження зміщеного шуму у велику кількість відео роликів і тегованих кадрів. Ця техніка забезпечила краще розуміння розподілу даних моделлю та, як наслідок, дозволяє генерувати на основі вхідних текстових описів більш різноманітний діапазон реалістичних відео.

### **Аспекти використання ШІ-генераторів у ланцюжках вирішення складних бізнес-задач**

Визначені вище характеристики роблять системи ШІ-генерації ефективними у розв'язанні різноманітних задач у сферах обробки природної мови й створення нового цифрового контенту різних форм, які на сьогодні користуються чималим попитом, як окремі прикладні

задачі. Але подібна функціональність може складати частину спільного процесного ланцюжка більш широко спеціалізованої задачі, що дозволило б вивести вирішення складних бізнес-задач на принципово новий інтелектуальний рівень, зокрема, вдосконалити якість прийняття рішень в складній системі.

Використання ШІ-генерації може забезпечити інтелектуалізацію тренінгових систем, проведення моніторингу змін у реальному часі, отримання швидкої візуалізації результатів аналізу, динамічне формування/коригування стратегії в реальному часі, проведення візуального аналізу інформації з метою виявлення протиріч і ще багато можливостей інтелектуалізації рішень в складних інформаційних системах.

Проведений аналіз дозволив виокремити спільні характеристики існуючих систем та зробити наступні висновки в аспекті їх інтеграції та застосування як складової автоматизованої інформаційної системи:

1. Всі існуючі системи ШІ-генерації базуються на вхідному текстовому описі, від якості якого залежить семантична адекватність результату генерації. Введення текстової підказки на сьогодні здебільшого здійснюється вручну, наприклад, за допомогою строки підказки. Адекватність одержаного результату генерації досягається багаторазовим змінням вхідних текстових інструкцій.

2. Різні системи, безумовно, дають різну семантичну якість генерації (як і технічну). Якість та семантичний зміст отриманого результату напряму залежить від вхідного текстового опису, і може кардинально змінюватися при зміні окремого слова в описі. Але не існує чіткого алгоритму формування цих текстових інструкцій чи, навіть, вимог до структури та змісту вхідного текстового опису, що ґрунтовно убезпечив би смислову якість результату генерації.

3. Більшість реалізації ШІ-генераторів, як правило, можна розглядати лише як “чорну скриньку” з входом та виходом, що ускладнює їх використання у спільному процесі вирішення задачі та робить їх непривабливими для розробників у цьому аспекті, але існують мо-

делі з відкритим кодом, як наприклад, модель відео генерації *Zeroscope AI*.

4. Зрозуміло, що ще однією характеристикою, важливою для розробників складних систем, є вартість моделі генерації. Тут, слід зазначити, що існує широкий діапазон цін. Зазвичай, безкоштовно, або за умовну ціну, надаються базові, не комерційні версії моделей з обмеженою функціональністю. Наприклад, з обмеженням числа генерацій на добу, розміру результуючого контенту чи просто для не комерційного використання. Але існує чимало безкоштовних моделей, з якими можна працювати і розвивати. Так, наприклад, вище згадана модель генерації з відкритим кодом *Zeroscope AI* є не лише досить розвинутою, але й безкоштовною. Нещодавно оновила ліцензію на свої продукти генеративного ШІ спільнота *Stability AI* і тепер вона стала безоплатною для дослідницького, не комерційного і навіть комерційного використання з доходом, що не перевищує \$1 млн.

5. Дуже важливий момент — мультимовність моделі. На сьогодні, при створенні складних систем з обробкою природної мови в середині, зазвичай, слід передбачати щонайменше білінгвальність таких систем, тобто необхідність підтримки української та англійської природних мов. Здійснений аналіз показав, що англійську мову підтримують переважна більшість існуючих моделей, а українську — далеко не всі. Тож при виборі базових моделей потрібно звернути на це увагу та обирати або мультимовну модель з підтримкою української мови, або передбачати додаткову обробку природної мови, а саме переклад. З описаних вище моделей генерації українську мову добре розуміють *DALL-E-2* та *NightCafe Studio*.

6. Також, слід зауважити, що генераційні моделі побудовані на системах машинного навчання чи глибокого навчання. Тренування та використання таких моделей потребує значних обчислювальних ресурсів, а самі моделі — великого обсягу різноманітних даних для навчання. Їх використання пов'язано з певними ризиками — моделі можуть створювати та поширювати небажані погляди, які могли опи-

нитися у тренувальних вибірках. В деяких випадках, можуть генерувати зайвий контент, надавати не очікувані результати при зміні контексту, тощо.

Щоб інструменти ШІ-генерації могли стати складовою автоматизованого ланцюжка вирішення бізнес-задачі, вони мають бути інтегровані в загальний процес, мати можливість автоматичного виклику, а, отже, отримувати файл з чіткими вхідними текстовими інструкціями зі змістом та структурою, що забезпечуватиме семантично адекватний результат генерації, який може підлягати подальшій обробці.

## Висновки

У роботі здійснено аналіз та наведено стислий опис технологій генеративного штучного інтелекту та найбільш вживаних на сьогодні застосунків ШІ-генерації. Стрімкий розвиток галузі генеративного ШІ дозволяє не лише вдосконалити вирішення лінгвістичних чи інших задач обробки природної мови та створювати на їх базі інформаційні контенті різних форм і форматів, а й по-новому подивитися на вирішення композитних задач, які генеративний ШІ може вивести на принципово інший рівень.

Однак, зовсім молода галузь генеративного штучного інтелекту, особливо щодо створення цифрових форм контентів таких як відео та зображення, містить чимало проблем, що потребують вирішення для забезпечення ефективності застосування даних технологій. Проведений в ході досліджень аналіз існуючих засобів дозволив виокремити спільні характеристики існуючих систем, дійти певних висновків щодо критеріїв вибору рішень для подальшої інтеграції в розробку складної системи та визначити напрями подальших досліджень наступним чином:

1) формування системи рекомендацій щодо вибору найбільш відповідних вимогам прикладної розробки генеративних моделей на основі більш глибокого аналізу функціональних та не функціональних характеристик;

2) на базі обраних моделей, вивчення залежностей семантичної якості результату генерації від структури та вмісту вхідного текстового опису;

3) вивчення особливостей та ступеня впливу специфіки прикладної задачі на формування вимог до вхідних інструкцій;

4) вироблення методологічних основ формування вхідних текстових описів для моделей ШІ-генерації класів *text-to-image* та *text-to-video*.

## ЛІТЕРАТУРА

1. Що таке велика мовна модель (Large Language Model, LLM)? URL:<https://thetransmitted.com/adlucem/shho-take-velyka-movna-model-large-language-4model-llm/> (дата звернення 20.04.2024).
2. Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, Ajmal Mian. A Comprehensive Overview of Large Language Models (2024). URL: <https://arxiv.org/pdf/2307.06435>
3. Іванов А., Онищенко В. Методи генерації зображень з використанням мереж GAN. Методи генерації автоматичного управління. Том 1 №42 (2023). URL: <https://asac.kpi.ua/article/view/279109>
4. Generative video model: Types, tasks, development and implementation. URL: <https://www.leewayhertz.com/create-generative-video-model/> (дата звернення 10.01.2023).
5. Tow, J., Bellagente, M., Mahan, D., Carlos Riquelme Ruiz, C.R. StableLM-3B-4E1T. Technical report for StableLM-3B-4E1T. URL: <https://stability.wandb.io/stability-llm/stable-lm/reports/StableLM-3B-4E1T--VmllDzoYMjU4?accessToken=u3zujipenkx5g7rtcj9qojjgxpconyjktkli2po09nffrfdhhchq045vp0wyfo> (дата звернення 10.01.2023).
6. Video generation models as world simulators. URL: <https://openai.com/index/video-generation-models-as-world-simulators/> (дата звернення 10.01.2023).
7. Text to Video Without Text-Video Training Data. Make-A-Video, an AI System from Meta, Generates Video from Text. URL: <https://www.deeplearning.ai/the-batch/ai-system-make-a-video-generates-video-from-text/8>. Порівняння графічних нейромереж. Топ-10. URL: <https://cases.media/en/article/porivnyannya-grafichnikh-neiromerezh-top-10> (дата звернення 10.01.2023).

Надійшла 29.07.2024

## REFERENCES

1. What is a large language model (Large Language Model, LLM)? [online]. Available at: <<https://thetransmitted.com/adlucem/shho-take-velyka-movna-model-large-language-4model-llm/>> [Accessed 20 Apr. 2024].
2. Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., ... & Mian, A. (2023). "A comprehensive overview of large language models". <https://doi.org/10.48550/arXiv.2307.06435>
3. Ivanov, A., Onishchenko V. (2023). "Image generation methods using GAN networks". Automatic control generation methods. T. 1, no 42. DOI: <https://doi.org/10.20535/1560-8956.42.2023.279109>
4. Generative video model: Types, tasks, development and implementation.[online], Available at: <<https://www.leewayhertz.com/create-generative-video-model/>> [Accessed 10 Jan. 2023].
5. Tow, J., Bellagente, M., Mahan, D., Carlos Riquelme Ruiz, C.R. StableLM-3B-4E1T. Technical report for StableLM-3B-4E1T. [online], Available at: <<https://stability.wandb.io/stability-llm/stable-lm/reports/StableLM-3B-4E1T--VmllDzoYMjU4?accessToken=u3zujipenkx5g7rtcj9qojjgxpconyjktkli2po09nffrfdhhchq045vp0wyfo>> [Accessed 10 Jan. 2023].
6. Video generation models as world simulators. [online], Available at: <<https://openai.com/index/video-generation-models-as-world-simulators/>> [Accessed 10 Jan. 2023].
7. Text to Video Without Text-Video Training Data. Make-A-Video, an AI System from Meta, Generates Video from Text. [online], Available at: <<https://www.deeplearning.ai/the-batch/ai-system-make-a-video-generates-video-from-text/>> [Accessed 10 Jan. 2023].
8. Comparison of graphical neural networks. Top 10. [online], Available at: <<https://cases.media/en/article/porivnyannya-grafichnikh-neiromerezh-top-10>> [Accessed 10 Jan. 2023].

Received 29.07.2024



V.D. Minenko, Founder, Twigames Inc.,  
Wilmington, Delaware, USA,  
3422 Old Capitol Trail, Suite# 241, Wilmington, DE 19808, US,  
ORCID: <https://orcid.org/0009-0003-5299-6786>,  
valerii@twigames.net

## ANALYSIS OF THE APPLICATION OF AI GENERATORS TO SOLVE COMPLEX BUSINESS PROBLEMS

**Introduction.** In recent years, there has been significant growth in the use of Generative Artificial Intelligence (Generative AI) applications that create visual data based on textual descriptions. This opens up new possibilities for enhancing information systems across various applied fields. It goes beyond just vivid forms of representation or information display; it offers the potential to address complex practical tasks at a fundamentally different level. The ability to store the history of visualizations allows tracking the dynamics of changes in specific information over time, providing deeper information analysis and a higher level of decision-making and strategy formulation in complex business systems.

However, the semantic quality of generating visual content remains a challenge, influenced not only by the choice of the generation model itself but also by the accuracy of the input instructions.

This article is dedicated to analyzing the current state of Generative AI technologies, existing models, their functional capabilities, advantages, limitations, and drawbacks.

**Purpose.** The research aims, on one hand, to identify effective applications of these technologies in solving complex business tasks, and on the other hand, to highlight issues with existing models regarding their integration into the business process chain and possible ways to address them.

**Results and conclusion.** The analysis conducted allowed:

Identifying common features of existing systems and highlighting characteristics significant for choosing a model in terms of its integration into a complex system, such as the semantic accuracy of generation results, openness of code, price, multilinguality, etc.

Formulating directions for further research, primarily focused on developing methodological foundations for automatically generating input textual descriptions for Generative AI models in video content generation.

**Keywords:** *generative artificial intelligence, natural language processing, AI generation, neural network, neural network model, text-to-text model, text-to-image model, text-to-video model, generative adversarial network, machine learning, business system, open-source models, semantic representation quality.*