

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

Штучний інтелект для генерації
ігрового контенту
(тема)

Виконав:
студент 2 курсу, групи СШМ-22-3
Гхоржевський В.С.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)

Керівник доц. Шафроненко А.Ю.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____
(підпис)

В.О. Філатов
(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)
Рівень вищої освіти другий (магістерський)
Спеціальність 122 Комп'ютерні науки
(код і повна назва)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Системи штучного інтелекту
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
« _____ » _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Тхоржевському Владиславу Станіславовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Штучний інтелект для генерації ігрового контенту

затверджена наказом університету від 1 квітня 20 24 р. № 260Ст

2. Термін подання студентом роботи до екзаменаційної комісії 14 червня 20 24 р.

3. Вихідні дані до роботи Науково-технічні публікації, дані Інтернет-джерел та відомих наукових проєктів

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі

2) Порівняння нейронних мереж

3) Дослідження створення реалістичних образів за допомогою ШІ

РЕФЕРАТ

Пояснювальна записка: 65 с., 13 рис., 1 табл., 1 дод., 21 джерело.

ГЕНЕРАТИВНЕ ГЛИБОКЕ НАВЧАННЯ, ГЕНЕРАТИВНО-ЗМАГАЛЬНІ МОДЕЛІ, СТВОРЕННЯ ІГРОВОГО КОНТЕНТУ, AUDIO TO FACE, UNREAL ENGINE 4, UNREAL ENGINE 5.

Об'єкт дослідження – процес створення ігрового контенту за допомогою нейронних мереж, таких як Audio to Face, Nemesis та подібних.

Предмет дослідження – види нейронних мереж та їх архітектур для генерації ігрового контенту.

Мета роботи – ознайомлення, вибір та дослідження нейронних мереж для генерації ігрового контенту, що визначає найбільш ефективні моделі в навчанні генерації реалістичного ігрового контенту з аудіо – та текстових описів.

Методи дослідження – аналіз існуючих алгоритмів видів генеративних моделей. Виділення ключових понять їх роботи, а також аналіз та електронних ресурсів існуючих моделей для вирішення задач синтезу аудіо та тексту у ігровий контент.

ABSTRACT

Master's thesis contains: 65 pp., 13 fig., 1 tabl., 1 ann., 21 references.

AUDIO TO FACE, CREATION OF GAME CONTENT, GENERATIVE DEEP LEARNING, GENERATIVE ADVERSARIAL MODELS, UNREAL ENGINE 5, UNREAL ENGINE 4.

Object of study – the process of creating game content using neural networks such as Audio to Face, Nemesis, and similar models.

Subject of study – types of neural networks and their architectures for generating game content.

Objective – familiarization, selection, and investigation of neural networks for generating game content to identify the most effective models for training the generation of realistic game content from audio and text descriptions.

Research methods – analysis of existing algorithms of different types of generative models. Highlighting key concepts of their operation, as well as analysis of literature and electronic resources of existing models for solving the tasks of synthesizing audio and text into game content.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ.....	8
1 Аналіз предметної галузі	10
1.1. Використання ШІ в розробці гри	19
1.2 Проблеми використання ШІ в розробці ігор.....	26
1.3 Проблеми використання нейромереж для робітників.....	31
1.4 Постановка завдання дослідження.....	37
2 Порівняння нейронних мереж	38
2.1 Аналіз відмінностей між нейронними мережами.....	38
3 Програмна реалізація та аналіз отриманих результатів.....	55
3.1 Вибір технології створення лицьових анімацій.....	55
3.2 Використання нейронних мереж для створення лицьових анімацій.....	55
3.3 Аналіз згенерованих лицьових анімацій	58
Висновки	61
Перелік джерел посилання	62
Додаток А Відомість кваліфікаційної роботи	65

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

- AI – Artificial Intelligence – штучний інтелект;
- AR – Augmented Reality – доповнена реальність;
- CNN – Convolutional Neural Network – згорткова нейронна мережа;
- CVAEs – Conditional Variational Autoencoders – умовні варіаційні автокодері;
- DLSS – Deep Learning Super Sampling – суперсемплінг з глибоким навчанням;
- GAN – Generative Adversarial Network – генеративно змагальна мережа;
- GPT – Generative Pre-trained Transformer – генеративна попередньо навчена трансформерна модель;
- GRU – Gated Recurrent Unit – блочна рекурентна одиниця;
- KL – Kullback-Leibler Divergence – дивергенція Кульбака –Лейблера;
- LSTM – Long Short-Term Memory – довгострокова короткочасна пам'ять;
- ММО – Massively Multiplayer Online – масова багатокористувацька онлайн-гра;
- MSE – Mean Squared Error – середньоквадратична помилка;
- NPC – Non-Playable Character – неігровий персонаж;
- ReLU – Rectified Linear Unit – випрямлена лінійна функція активації;
- RNN – Recurrent Neural Network – рекурентна нейронна мережа;
- RPG – Role-Playing Game – рольова гра;
- SGD – Stochastic Gradient Descent – стохастичний градієнтний спуск;
- SSIM – Structural Similarity Index – індекс структурної подібності;
- VAE – Variational Autoencoder – варіаційний автокодер;
- VR – Virtual Reality – віртуальна реальність.

ВСТУП

У останні роки штучний інтелект став невід'ємною частиною розвитку ігрової індустрії, надаючи нові можливості для створення контенту. Використання алгоритмів глибокого навчання дозволяє автоматизувати процеси, які раніше вимагали значних людських ресурсів. Генерація ігрового контенту, зокрема, стала важливою областю досліджень, оскільки це дозволяє значно підвищити ефективність розробки та збагатити ігровий досвід. Одним із ключових застосувань штучного інтелекту в ігровій індустрії є створення високоякісних текстур та 3D моделей персонажів. Це дозволяє розробникам значно скоротити час на створення графічних елементів гри, а також забезпечити більшу варіативність і деталізацію.

Окрім цього, моделі глибокого навчання можуть використовуватися для автоматичного створення ігрових світів, де кожен елемент створюється відповідно до певних правил та стилю гри. Штучний інтелект також відкрив нові можливості для генерації сценаріїв та діалогів у іграх. Використовуючи моделі природної обробки мови (NLP), системи можуть створювати реалістичні та захоплюючі діалоги між персонажами, що додає глибини та інтерактивності геймплею. Це особливо важливо для рольових ігор (RPG), де взаємодія з персонажами грає ключову роль у розвитку сюжету. Крім того, штучний інтелект дозволяє створювати адаптивні ігрові механіки, які підлаштовуються під стиль гри користувача.

Це забезпечує індивідуальний підхід до кожного гравця, роблячи ігровий досвід унікальним і більш цікавим. Такі системи можуть аналізувати поведінку гравця та автоматично налаштовувати складність гри, створюючи більш гармонійний баланс між викликом і задоволенням. Важливим аспектом є також використання штучного інтелекту для тестування ігор. Автоматизовані тестувальники на основі AI можуть ефективно виявляти баги та недоліки у геймплеї, що значно знижує витрати часу та ресурсів на етапі тестування. Це дозволяє випускати більш якісні

продукти на ринок та забезпечувати краще обслуговування користувачів. Однак, використання штучного інтелекту для генерації ігрового контенту також має свої виклики. Одним з них є необхідність великої кількості даних для тренування моделей, що може бути складним і витратним процесом. Крім того, важливо враховувати етичні аспекти, зокрема щодо створення контенту, який відповідає моральним і правовим нормам. У підсумку, штучний інтелект відкриває нові горизонти у сфері генерації ігрового контенту, дозволяючи створювати більш реалістичні, захоплюючі та індивідуальні ігрові світи. Це робить ігри більш цікавими і доступними для широкої аудиторії, одночасно оптимізуючи процес розробки та підвищуючи якість кінцевого продукту.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

Штучний інтелект (ШІ) в ігровій індустрії займає все важливе місце, зокрема в процесі генерації ігрового контенту. Це включає створення нових рівнів, персонажів, сюжетних ліній, ландшафтів та інших аспектів, що роблять гру цікавою та непередбачуваною. Застосування ШІ дозволяє розробникам створювати складніші та реалістичніші ігри, одночасно знижуючи витрати часу та ресурсів. На початку розвитку ігрової індустрії використання ШІ було обмежене простими алгоритмами, що відповідали за поведінку ворогів та NPC. Перші ігри, такі як «Pac-Man» (1980) і «Space Invaders» (1978), мали примітивні форми ШІ, які забезпечували базові взаємодії з гравцем.

З розвитком технологій ігри почали використовувати складні алгоритми ШІ. Наприклад, серія «The Sims» (2000) від Maxis представила складну симуляцію життя, де NPC мали власні потреби, бажання та взаємодії. Ігри, такі як «F.E.A.R.» (2005) (рисунок 1.1) використовували ШІ для створення ворогів, які адаптувалися до дій гравця, застосовуючи тактичні маневри.

Сьогодні ШІ використовується в багатьох аспектах геймдизайну. Зокрема, застосування методів машинного навчання та глибинного навчання дозволяє створювати більш реалістичні та інтерактивні світи. Процедурна генерація контенту, як у грі «No Man's Sky» (2016) (рисунок 1.2), дозволяє створювати величезні, різноманітні всесвіти з мінімальним людським втручанням. Процедурна генерація використовує алгоритми для автоматичного створення контенту.

Це може бути генерація ландшафтів, рівнів, персонажів та інших елементів гри [1]. Одним з найвідоміших прикладів є гра «No Man's Sky», де мільярди планет генеруються процедурно, забезпечуючи унікальний досвід для кожного гравця [2].

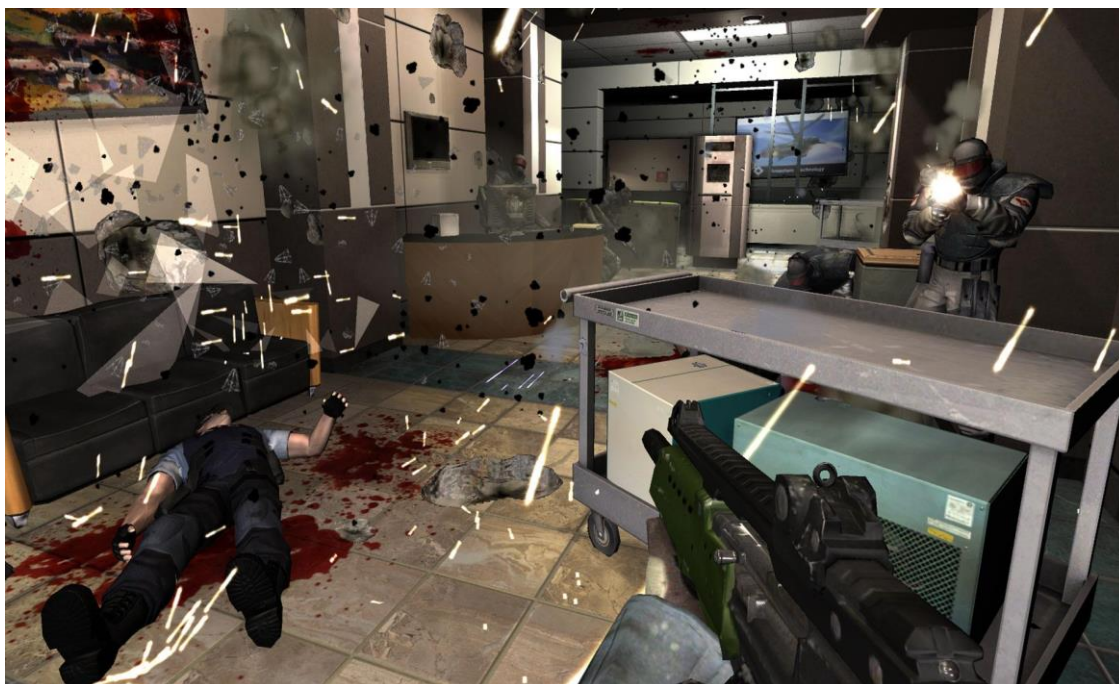


Рисунок 1.1 – Гра Fear у якій вороги адаптувалися до дій гравця



Рисунок 1.2 – Галактика процедурно згенерована нейромережами у грі No Man's Sky

ШІ може використовуватися для створення адаптивних сюжетів, що реагують на дії гравця [3]. Це дозволяє створювати нелінійні історії, де кожне рішення гравця має наслідки [3]. Наприклад, у грі «Detroit: Become Human» (2018) від Quantic Dream використовується ШІ для управління складними сюжетними розгалуженнями, що надає гравцеві можливість впливати на розвиток подій [4]. ШІ також використовується для створення реалістичних NPC, які можуть адаптуватися до дій гравця [5].

Це включає генерацію поведінки, діалогів та взаємодій [5]. Наприклад, у грі «The Elder Scrolls V: Skyrim» (2011) NPC мають власні розклади та можуть взаємодіяти з гравцем та іншими NPC, створюючи відчуття живого світу [6].

Застосування глибинного навчання та нейронних мереж дозволяє покращувати графічну складову ігор [7]. Наприклад, технологія DLSS від NVIDIA використовує ШІ для підвищення роздільної здатності зображень в реальному часі, що дозволяє отримати більш чітке та якісне зображення без значного збільшення навантаження на апаратне забезпечення [8]. ШІ використовується для створення реалістичних фізичних симуляцій, що робить ігровий процес більш захоплюючим [9]. Це включає симуляцію рідин, газів, твердих тіл та інших фізичних явищ [9]. Наприклад, у грі «Half-Life 2» (2004) від Valve використовується фізичний рушій Havok, який забезпечує реалістичну взаємодію об'єктів [10]. Застосування ШІ в генерації ігрового контенту також стикається з певними викликами [11]. Створення складних ШІ-систем вимагає значних обчислювальних ресурсів [11]. Автоматизація процесу генерації контенту може знизити унікальність та креативність, яку привносять людські розробники [11].

Використання ШІ в іграх може викликати етичні питання щодо впливу на гравців та їхню поведінку [11]. Проте з розвитком технологій, таких як квантові обчислення та вдосконалені алгоритми машинного навчання, можливості ШІ в ігровій індустрії будуть зростати [12]. Подальший розвиток ШІ дозволить створювати ще більш інтерактивні та

реалістичні світи [12]. Автоматизація процесів розробки з використанням ШІ може знизити витрати та час на створення ігор [12]. Застосування ШІ в генерації ігрового контенту відкриває нові можливості для розробників, дозволяючи створювати складніші, цікавіші та реалістичніші ігри [13].

Хоча існують певні виклики, розвиток технологій обіцяє подальше вдосконалення цього напрямку, що робить його одним з найперспективніших в ігровій індустрії [13]. З розвитком технологій та зростаючою потужністю обчислювальних систем перспективи застосування ШІ в ігровій індустрії стають ще більш вражаючими [14]. Одним із найцікавіших напрямів є розвиток персоналізованого ігрового досвіду [14]. Використання алгоритмів машинного навчання дозволяє створювати ігри, які можуть адаптуватися до стилю гри кожного гравця, забезпечуючи унікальний і неповторний досвід для кожного користувача [14].

Персоналізовані світи та сюжети, що створюються в режимі реального часу на основі дій і уподобань гравця, стають реальністю завдяки ШІ. Наприклад, в грі «Shadow of Mordor» (рисунок 1.3) система Nemesis створює динамічних ворогів, які запам'ятовують попередні зустрічі з гравцем і розвиваються відповідно до взаємодій з ним. Це створює більш глибокий і захоплюючий ігровий процес, де кожна гра стає унікальною. Крім того, ШІ може аналізувати дії гравця і відповідно коригувати складність та надавати індивідуальні виклики, що підходять саме для конкретного гравця. Це не лише підвищує рівень залученості, а й допомагає гравцям розвивати свої навички в ігровому світі. Наприклад, адаптивні сюжети можуть включати нові сюжетні лінії або персонажів, які відображають уподобання гравця, забезпечуючи кожен раз новий ігровий досвід. Завдяки таким можливостям, ігри стають не просто розвагою, а індивідуальним переживанням, що стимулює гравця повертатися до гри знову і знову.

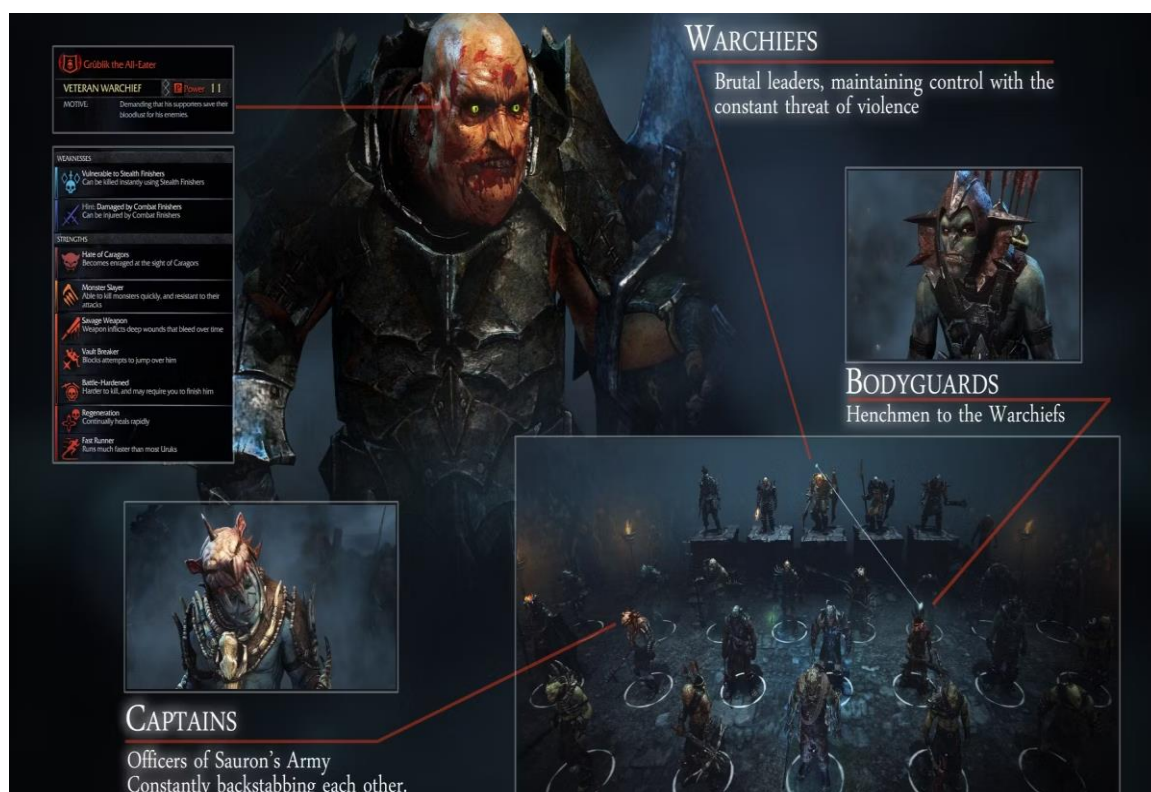


Рисунок 1.3 – Shadow of Mordor Nemesis запам'ятовує кого і як перемогли з ворогів

Крім того, ШІ активно використовується для створення реалістичних та інтерактивних ігрових середовищ. Віртуальні світи стають все більш деталізованими та правдоподібними завдяки використанню алгоритмів глибинного навчання для моделювання фізики, світла та звуку. Це дозволяє гравцям занурюватися у ігровий процес на новому рівні, де кожна деталь світу реагує на їх дії.

Окремо варто згадати про розвиток технологій доповненої та віртуальної реальності, які також активно використовують ШІ. Ці технології дозволяють створювати імерсивні ігрові світи, де гравці можуть фізично взаємодіяти з оточенням. Наприклад, у VR-іграх, таких як «Beat Saber» (рисунок 1.5) або «Half-Life: Alyx» (рисунок 1.4), ШІ використовується для створення інтерактивних середовищ, які реагують на рухи та дії гравця в реальному часі.

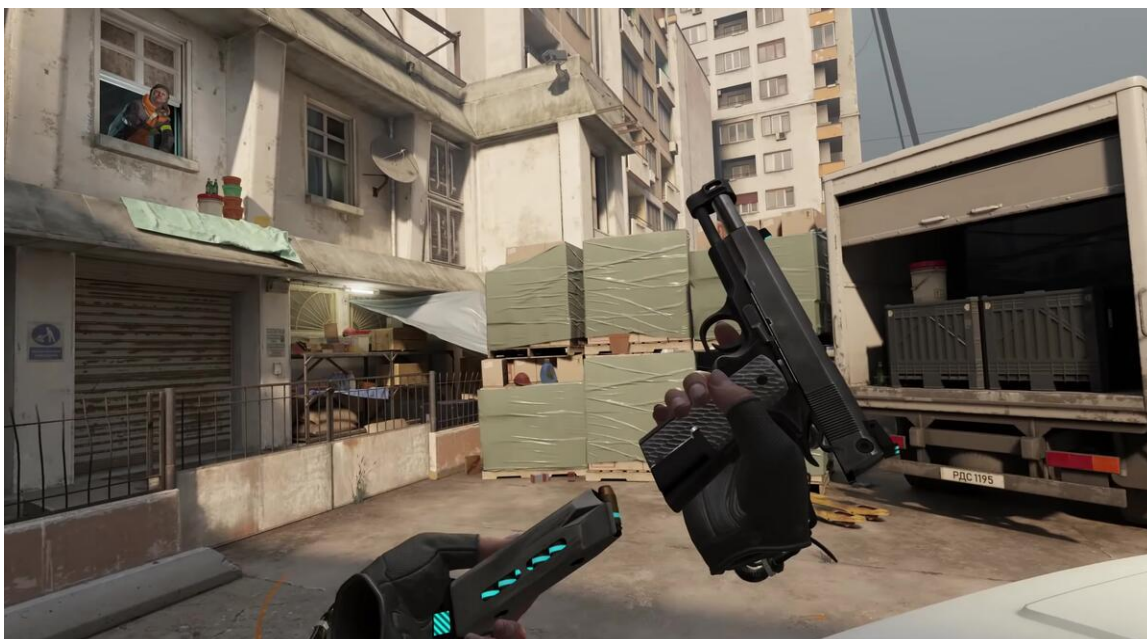


Рисунок 1.4 – VR гра Half life Alyx



Рисунок 1.5 – VR гра Beat Saber

Ще одним перспективним напрямом є використання ШІ для автоматизації тестування ігор. Процес тестування великих і складних ігор є надзвичайно трудомістким і вимагає значних ресурсів. Використання ШІ дозволяє автоматизувати цей процес, виявляючи баги та недоліки ще на

стадії розробки, що значно знижує витрати і прискорює випуск якісних продуктів на ринок. Не менш важливим є розвиток соціальної взаємодії в іграх завдяки ІІІ. Сучасні онлайн-ігри все частіше включають соціальні елементи, де гравці можуть спілкуватися і взаємодіяти один з одним у віртуальному середовищі. Використання ІІІ дозволяє створювати більш складні та цікаві сценарії взаємодії між гравцями, включаючи модерацію спілкування, підтримку мови та перекладу, а також створення віртуальних асистентів, які допомагають гравцям. Попри всі ці переваги, впровадження ІІІ в ігровій індустрії стикається з певними викликами. Одним із головних є питання етики та конфіденційності. Використання персональних даних гравців для створення персоналізованого досвіду може викликати занепокоєння щодо конфіденційності та безпеки. Тому важливо розробляти прозорі політики використання даних і забезпечувати захист інформації. Крім того, розвиток ІІІ в ігровій індустрії потребує значних інвестицій у дослідження та розробку. Це включає навчання фахівців, створення нових алгоритмів та інфраструктури для їх реалізації.

Водночас конкуренція на ринку ігор змушує компанії шукати інноваційні підходи, що стимулює розвиток нових технологій і покращує якість кінцевих продуктів. На завершення, можна сказати, що застосування ІІІ в генерації ігрового контенту відкриває нові горизонти для розвитку ігрової індустрії. Воно дозволяє створювати складні, реалістичні та інтерактивні світи, які забезпечують гравцям унікальний досвід. Хоча існують певні виклики та ризики, розвиток технологій ІІІ обіцяє подальше вдосконалення ігор, роблячи їх ще більш захоплюючими та реалістичними. Один із найбільш інноваційних напрямків у застосуванні штучного інтелекту в ігровій індустрії є генерація сценаріїв і подій.

Завдяки ІІІ, ігри можуть автоматично створювати унікальні сюжетні лінії, реагуючи на дії та вибори гравця. Це дозволяє створювати нелінійні історії, де кожен ігровий сеанс може відрізнятись від попереднього, забезпечуючи багатогранність ігрового досвіду. Алгоритми, що

використовуються для цього, аналізують поведінку гравця та адаптують сценарій відповідно до його дій, що створює відчуття глибокого занурення у гру.

Ще одним цікавим аспектом є використання ШІ для поліпшення взаємодії між гравцями та ігровим світом. Завдяки природній обробці мови та генеративним моделям, ігри можуть забезпечувати більш реалістичні діалоги з NPC. Такі системи здатні розуміти та відповідати на складні запити гравців, створюючи більш інтерактивний і захоплюючий досвід. Наприклад, в іграх жанру RPG, таких як «The Elder Scrolls» або «Mass Effect», діалоги з NPC відіграють важливу роль у розвитку сюжету та взаємодії з ігровим світом.

Важливим напрямком є також автоматизоване створення ігрових активів, таких як текстури, моделі та анімації. Використовуючи генеративні змагальні мережі, розробники можуть створювати високоякісні візуальні елементи з мінімальним втручанням людини.

Це значно скорочує час розробки та дозволяє створювати більш деталізовані та реалістичні ігрові світи.

Подібні технології використовуються для створення реалістичних персонажів, об'єктів та оточень, що підвищує загальну якість гри.

ШІ також грає ключову роль у покращенні систем балансування гри. Алгоритми аналізують дані про ігровий процес у режимі реального часу, визначаючи проблемні місця та автоматично регулюючи складність гри для забезпечення оптимального досвіду для гравців різних рівнів майстерності. Це дозволяє створювати ігри, які можуть бути цікавими як для новачків, так і для досвідчених гравців, підтримуючи належний рівень виклику та задоволення.

У багатокористувацьких онлайн іграх ШІ використовується для забезпечення справедливості та боротьби з шахрайством. Античитерські системи на основі машинного навчання аналізують поведінку гравців і виявляють аномалії, що можуть свідчити про використання сторонніх

програм або скриптів. Це допомагає підтримувати чесну гру та захищає інтереси чесних гравців.

Не можна обійти увагою і використання ШІ в розвитку систем управління та оптимізації серверів для онлайн-ігор. Завдяки інтелектуальним алгоритмам, сервери можуть ефективно розподіляти ресурси, обробляти великі обсяги даних і забезпечувати стабільну роботу гри навіть під великим навантаженням. Це особливо важливо для масових багатокористувацьких онлайн-ігор, де стабільність і швидкість реакції серверів мають критичне значення для загального ігрового досвіду.

Важливим аспектом є і використання ШІ для створення нових ігрових механік. Наприклад, в іграх з елементами стратегії ШІ може забезпечувати більш реалістичну та розумну поведінку ворогів, що робить гру більш складною та захоплюючою.

Алгоритми, що здатні вчитися на основі дій гравця, можуть адаптувати стратегії та тактики, забезпечуючи непередбачуваність і високий рівень виклику.

Наостанок варто зазначити, що розвиток ШІ в ігровій індустрії відкриває нові горизонти для інди-розробників. Технології, що раніше були доступні лише великим студіям, стають більш доступними завдяки відкритим платформам та інструментам, що базуються на ШІ. Це дозволяє незалежним розробникам створювати високоякісні та інноваційні ігри з обмеженими ресурсами.

Таким чином, використання штучного інтелекту в генерації ігрового контенту не лише покращує якість ігрового досвіду, але й відкриває нові можливості для розробників. ШІ дозволяє створювати більш складні, реалістичні та захоплюючі ігри, що відповідають потребам сучасних гравців. Незважаючи на існуючі виклики, подальший розвиток технологій ШІ обіцяє ще більше інновацій та вдосконалень у майбутньому, що робить цей напрямок одним з найперспективніших в ігровій індустрії.

1.1. Використання ШІ в розробці гри

Розробка ігор – це захоплюючий та творчий процес, що потребує поєднання технічних знань, художньої креативності та вдумливого проектування. Однією з ключових складових сучасних ігор є вміння генерувати унікальний та захоплюючий контент, який додає глибину та реалізм геймплею.

Сьогодні нейронні мережі виявляються надзвичайно корисним інструментом для розробників, оскільки вони можуть автоматизувати процеси створення різноманітних елементів ігрового світу. Давайте поглянемо на деякі з основних способів, які нейронні мережі можуть бути використані для генерації контенту в іграх:

– генерація персонажів та виглядів – нейронні мережі можуть створювати унікальних персонажів з різноманітними властивостями, такими як зовнішність, стиль одягу, аксесуари, а також особливості характеру та поведінки (рисунки 1.6 – 1.7);



Рисунок 1.6 – Персонаж Inworld згенерований нейромережею



Рисунок 1.7 – Персонаж Різі з Genshin Impact прототип якого генерували за допомогою нейромереж

– генерація ігрових об'єктів та елементів – вони можуть генерувати різноманітні об'єкти в грі, такі як зброя, предмети, будівлі, різновиди теренів тощо, що збагачує ігровий світ і робить геймплей більш різноманітним;

– створення ігрових світів та локацій – нейронні мережі можуть бути використані для створення різних ігрових світів з великою кількістю деталей та унікальних елементів. Вони аналізують архітектуру та ландшафт ігрового світу та генерують нові локації, відповідно до параметрів встановлених розробниками (рисунок 1.8);



Рисунок 1.8 – Створені локації за допомогою неймереж

– генерація музики та аудіоатмосфери – нейронні мережі можуть бути використані для створення музичних композицій, звукових ефектів та амбієнтної аудіоатмосфери, яка підкреслює настрій та атмосферу гри;

– генерація сюжетних сценаріїв та діалогів – нейронні мережі можуть створювати сюжетні сценарії, діалоги між персонажами та історії для гри. Вони аналізують структури та зв'язки в сюжеті і генерують нові ідеї для розвитку сюжетної лінії та взаємодії між персонажами (рисунок 1.9);

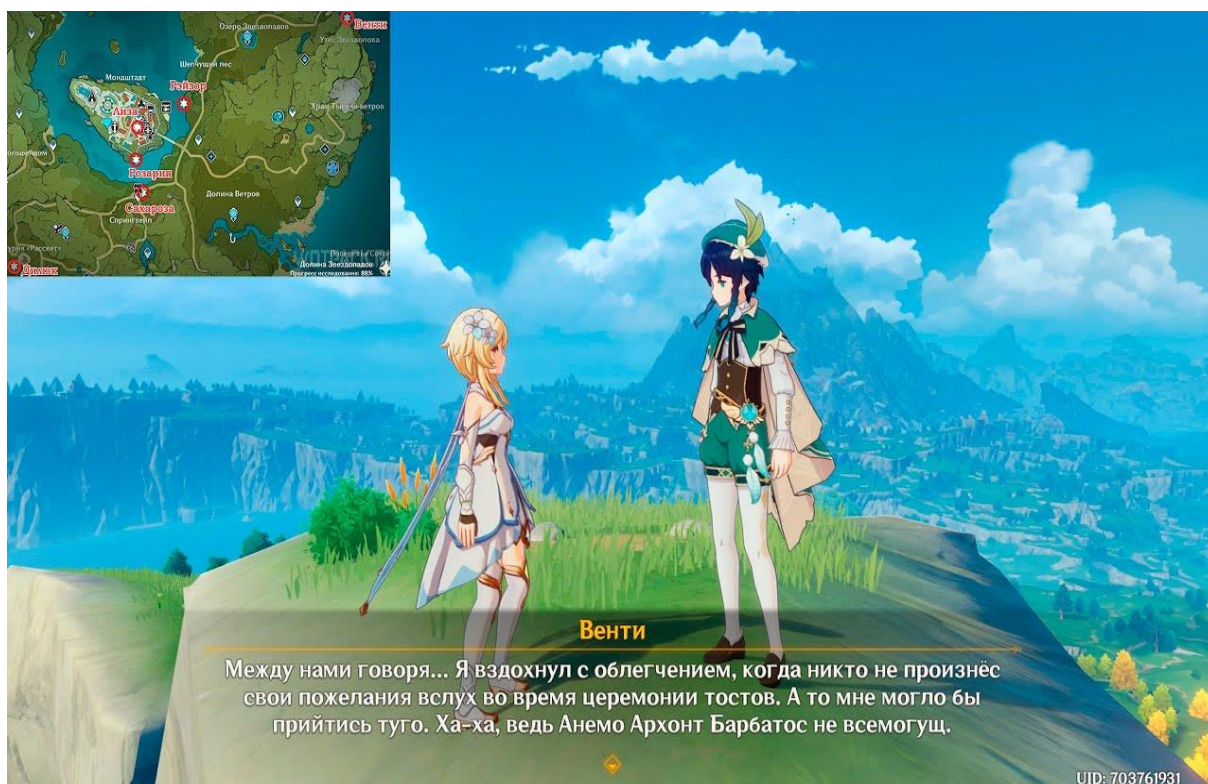
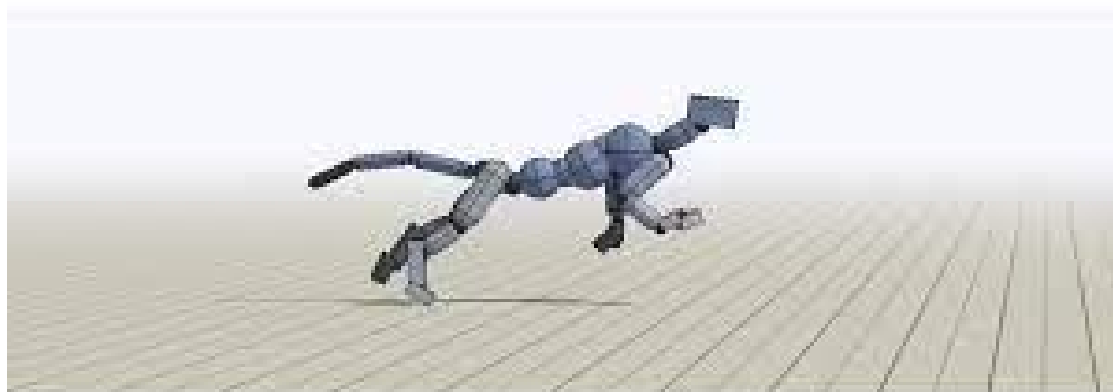


Рисунок 1.9 – Діалоги створені нейромережами для ігор

– автоматичне створення анімацій та рухів – нейронні мережі можуть бути використані для генерації анімацій персонажів та рухів у грі. Вони аналізують рухи та взаємодії персонажів у різних ситуаціях та створюють реалістичні анімації для поліпшення іммерсії гравців (рисунок 1.10 – 1.11);

Lion: Run



Simulated Character

Рисунок 1.10 – Анімації, що генеруються нейросетями



Рисунок 1.11 – Лицьова анімація створена за допомогою Audio to Face

– динамічне адаптування геймплею – нейронні мережі можуть адаптувати геймплей у реальному часі відповідно до дій гравця та його стилю гри. Наприклад, вони можуть налаштовувати рівень складності, розміщення ворогів або навіть змінювати сценарії подій в грі;

– виявлення шаблонів та покращення геймплею – нейронні мережі можуть аналізувати геймплей гравців та виявляти шаблони чи тенденції, які допомагають розробникам вдосконалювати гру. Наприклад, вони можуть розпізнавати часті помилки гравців або найбільш популярні стратегії та надавати цю інформацію для оптимізації геймплею;

– покращення взаємодії з гравцями – нейронні мережі можуть використовуватися для покращення інтерактивності та взаємодії з гравцями. Наприклад, вони можуть розробляти персоналізовані системи підтримки гравців, адаптувати ігровий досвід до індивідуальних уподобань та навіть створювати інтелектуальних ворогів, які навчаються від гравця;

– генерація текстур та візуальних ефектів – нейронні мережі можуть бути використані для створення різноманітних текстур для об'єктів у грі, включаючи природні ландшафти, архітектурні деталі, водяні ефекти та інші візуальні елементи. Вони можуть аналізувати велику кількість даних та генерувати унікальні та реалістичні текстури, які підвищують якість графіки в грі;

– автоматизація тестування та виявлення помилок – нейронні мережі можуть допомагати у виявленні помилок та проблем в грі через аналіз великих обсягів даних. Вони можуть автоматично тестувати різні аспекти гри, виявляти недоліки та пропонувати варіанти виправлень для поліпшення якості геймплею;

– підтримка гравців та відслідковування реакцій – нейронні мережі можуть бути використані для аналізу поведінки та реакцій гравців у реальному часі. Вони можуть відслідковувати дії гравців, їх взаємодію з ігровим середовищем та взаємодії з іншими гравцями для покращення геймплею та надання персоналізованого досвіду;

– системи прогнозування та оптимізація геймплею – нейронні мережі можуть аналізувати дані гравців для прогнозування їхніх дій та уподобань. Це дозволяє розробникам оптимізувати геймплей, пристосовуючи гру до унікальних стилів гри та уподобань кожного гравця;

– створення реалістичних інтерактивних середовищ – нейронні мережі можуть бути використані для створення інтерактивних середовищ у грі, де об'єкти реагують на дії гравця та змінюють своє становище відповідно до ситуації. Наприклад, рухомі рослини, динамічні погодні умови, реакція об'єктів на взаємодію гравця. Використання нейронних мереж у генерації контенту для ігор має ряд переваг і можливостей, які зроблять цей процес більш ефективним та захоплюючим для розробників та гравців.

По-перше, нейронні мережі можуть створювати великі обсяги контенту за короткий час, що дозволяє прискорити розробку гри та зосередитися на інших аспектах, таких як геймплей, механіки та баланс. Це особливо важливо в сучасному світі, де час – це важливий ресурс.

По-друге, нейронні мережі можуть покращити реалізм та деталізацію ігрового світу. Вони здатні створювати реалістичні текстури, анімації та атмосферу, що підвищує іммерсивність та відчуття присутності гравця в грі.

Крім того, нейронні мережі можуть адаптуватися до змін та вимог гравців, що дозволяє змінювати контент гри в реальному часі. Наприклад, змінювати вигляд персонажів, додавати нові об'єкти або змінювати атмосферу відповідно до дій гравця або його уподобань.

Також, використання нейронних мереж дозволяє створювати унікальні та неповторні ігрові враження для кожного гравця, що робить гру більш персоналізовано та привабливою. Наприклад, генерувати різні варіанти ігрових сценаріїв або адаптивну графіку під особистість гравця.

Загалом, використання нейронних мереж для генерації контенту у іграх сприяє швидкості, реалізму, адаптивності та персоналізації геймплею, що робить цей підхід дуже привабливим для сучасних розробників та

гравців. Ще однією перевагою використання нейронних мереж у генерації ігрового контенту є можливість швидкого та ефективного впровадження змін у вже існуючу гру. Наприклад, якщо розробникам потрібно додати нові персонажі, об'єкти або локації, нейронні мережі можуть вигенерувати їх за короткий час, що полегшує процес оновлення та розширення ігрового світу.

Крім того, використання нейронних мереж дозволяє створювати реалістичні та динамічні ігрові світи зі змінною поведінкою об'єктів та персонажів. Наприклад, у іграх із відкритим світом нейронні мережі можуть генерувати зміни в умовах довкілля відповідно до дій гравця, створюючи враження живого та реагуючого світу.

Крім того, за допомогою нейронних мереж можна автоматизувати процеси тестування та балансування гри. Вони можуть аналізувати поведінку гравців та виявляти рівноваги або недоліки у геймплеї, що дозволяє розробникам швидко вносити виправлення та покращення.

Не можна також забувати про можливості покращення графіки та візуального виконання ігор за допомогою нейронних мереж. Вони можуть створювати вражаючі спеціальні ефекти, високо деталізовані текстури та анімації, які роблять гру більш реалістичною та привабливою для гравців.

Таким чином, використання нейронних мереж у генерації контенту для ігор має безліч переваг, які роблять процес розробки та вдосконалення ігрового досвіду більш ефективним та захоплюючим.

1.2 Проблеми використання ШІ в розробці ігор

Використання штучного інтелекту (ШІ) у розробці ігор має багато переваг, проте також супроводжується низкою проблем і викликів. Ці проблеми можуть впливати на різні аспекти розробки, експлуатації та кінцевого ігрового досвіду. Нижче розглянемо основні проблеми, що виникають при використанні ШІ в ігровій індустрії.

Однією з головних проблем є складність розробки та інтеграції ШІ. Створення складних моделей ШІ, які здатні адаптуватися до різних ситуацій у грі, вимагає значних технічних знань та ресурсів. Розробникам потрібно мати глибокі знання в галузі машинного навчання, глибинного навчання, нейронних мереж та алгоритмів, що ускладнює процес розробки і підвищує його вартість. Крім того, інтеграція ШІ в існуючі ігрові системи може бути складною через необхідність адаптації ігрового рушія та інших компонентів гри. Проблеми з продуктивністю також є важливим викликом. Використання складних алгоритмів ШІ може вимагати значних обчислювальних ресурсів, що може призвести до зниження продуктивності гри. Особливо це стосується ігор з великими відкритими світами та високим рівнем деталізації, де ШІ необхідно обробляти великий обсяг даних у режимі реального часу. Це може вимагати потужного апаратного забезпечення, яке не завжди доступне для всіх гравців.

Етичні питання та конфіденційність даних також викликають занепокоєння. Використання ШІ для персоналізації ігрового досвіду часто передбачає збір та аналіз великої кількості даних про гравців, включаючи їхні дії, уподобання та поведінку в грі. Це може викликати занепокоєння щодо конфіденційності та безпеки персональних даних. Розробники повинні дотримуватися строгих правил щодо збору та використання даних, забезпечуючи прозорість і безпеку для користувачів.

Ще однією проблемою є баланс між автоматизацією та творчістю. Хоча ШІ може значно спростити процес створення контенту, існує ризик втрати творчого підходу. Автоматизована генерація контенту може призвести до одноманітності та втрати унікальності, яку привносять людські розробники. Тому важливо знайти баланс між використанням ШІ для автоматизації рутинних завдань та збереженням творчого елемента в процесі розробки ігор.

Тестування та налагодження ШІ також можуть бути складними. ШІ-системи часто поведуться непередбачувано, що ускладнює процес

тестування та виявлення помилок. Автоматизовані системи тестування можуть не завжди виявляти всі проблеми, що вимагає додаткових ресурсів на ручне тестування та налагодження. Крім того, складність ШІ-систем ускладнює процес їх оптимізації та вдосконалення.

Ще одна важлива проблема – це підтримка та оновлення ШІ-систем після випуску гри. Гра, що використовує складний ШІ, потребує постійної підтримки та оновлень, щоб виправляти помилки, поліпшувати алгоритми та адаптуватися до змін у поведінці гравців. Це вимагає додаткових ресурсів та може бути викликом для невеликих студій розробників.

Нарешті, варто зазначити питання доступності. Використання передових технологій ШІ може зробити ігри менш доступними для певних категорій гравців.

Наприклад, гравці з обмеженими можливостями можуть зіткнутися з труднощами в адаптації до складних систем ШІ. Розробники повинні враховувати ці фактори та працювати над створенням інклюзивних ігор, що доступні для всіх користувачів.

Отже, попри значні переваги, використання ШІ у розробці ігор супроводжується рядом проблем і викликів, які потребують ретельного розгляду та вирішення. Технічні складнощі, етичні питання, проблеми з продуктивністю та конфіденційністю, баланс між автоматизацією та творчістю, тестування та підтримка – всі ці аспекти вимагають уваги та ресурсів. Однак, при правильному підході, застосування ШІ може значно покращити якість ігор та підвищити задоволення гравців. Ще однією важливою проблемою є «проблема контролю». Це означає, що інтелект ШІ може здатися непередбачуваним або неконтрольованим, особливо в складних ігрових сценаріях. Наприклад, ШІ може неправильно розуміти завдання або діяти нелогічно в різних ситуаціях, що призводить до незадовільного геймплею для гравців. Контроль інтелекту ШІ, його налаштування та корекція в разі потреби – це важливий аспект, який розробники ігор повинні уважно вивчати.

Ще однією проблемою є «проблема емоційного інтелекту». ШІ може бути обмеженим у виявленні та вираженні емоцій, що може впливати на іммерсивність гри та взаємодію з гравцями. Наприклад, в іграх, де важлива емоційна реакція персонажів або суперників, обмеженість ШІ може призвести до маловиразної або невірно відтвореної емоційної динаміки.

Також варто згадати про «проблему універсальності». ШІ, як правило, розробляються для конкретних ігор або жанрів і можуть бути неефективними або непридатними для використання в інших контекстах. Це створює виклик для розробників, які прагнуть застосовувати ШІ в різних типах ігор або різноманітних ігрових ситуаціях.

Нарешті, важливо враховувати «проблему автономності». Хоча автономна поведінка ШІ може бути корисною для динамічних та реалістичних ігрових досвідів, вона також може призвести до непередбачуваного або небажаного результату. Наприклад, ШІ - противники можуть приймати стратегії, які гравці вважають нереалістичними або некоректними, що може порушити баланс гри або знизити задоволення від гри.

Отже, розробка та використання ШІ в ігровій індустрії вимагає уважного вирішення цих проблем, а також пошуку компромісу між технічними можливостями ШІ та очікуваннями гравців. Тільки таким чином можна досягти успіху в імплементації ШІ, забезпечивши якісний та задоволений геймплей для широкої аудиторії гравців.

Варіаційні автоенкодера – це тип нейронних мереж, які використовуються для генерації та реконструкції даних. Основна ідея VAE полягає в тому, щоб навчити модель вивчати внутрішній простір даних і генерувати нові приклади, які подібні до вхідних даних, але не зовсім однакові. Ось докладніше про VAEs та їх переваги та недоліки, а також інші нейронні мережі зі схожими функціями.

Переваги VAEs:

– генерація нових даних – VAEs навчаються не тільки реконструювати вхідні дані, а й генерувати нові приклади у межах вивченого простору. Це корисно для створення нових зображень, текстів або інших типів даних;

– робуєність до шуму – VAEs володіють властивістю згладжування даних, тобто вони можуть ефективно впоратися з шумом або незначними відхиленнями в даних, створюючи більш стійкі та реалістичні реконструкції;

– вивчення внутрішнього представлення – одним з ключових аспектів VAEs є вивчення низькорозмірного простору, який представляє внутрішню структуру даних. Це може допомогти у візуалізації та розумінні взаємозв'язків між даними;

– гармонійна генерація – оскільки VAEs використовують ймовірнісний підхід, вони можуть генерувати нові приклади, які гармонійно вписуються у вивчений простір даних, створюючи більш природні та реалістичні результати.

Недоліки VAEs:

– розмірність простору – навчені нейронні мережі VAEs зазвичай мають низькорозмірний внутрішній простір. Це може призвести до втрати деякої інформації або деталей у згенерованих даних;

– деяка непередбачуваність – оскільки VAEs базуються на ймовірнісному підході, їх результати можуть бути деякою мірою непередбачуваними, особливо при генерації нових даних;

– складність навчання – навчання VAEs може бути більш складним процесом порівняно з іншими типами автоенкодерів через використання ймовірнісного підходу та додаткових шарів для генерації даних.

Нейронні мережі зі схожими функціями:

– Generative Adversarial Networks, GANs також використовуються для генерації даних, але вони працюють за іншим принципом змагання між генератором і дискримінатором;

– Autoencoders – це загальний клас нейронних мереж, які використовуються для автоматичного вивчення представлення даних шляхом упаковки вхідних даних у складне внутрішнє представлення і потім розпаковки його для генерації реконструкцій. Вони можуть бути використані для створення нових прикладів даних, але їхні можливості генерації не такі розширені, як у VAEs або GANs;

– Boltzmann Machines – це ще один клас нейронних мереж, які використовують ймовірнісний підхід для моделювання даних. Вони можуть бути використані для генерації нових даних, але їх використання у розробці ігор може бути обмеженим через складність навчання та обчислювальні витрати;

– Conditional Variational Autoencoders CVAEs – це варіація VAEs, яка дозволяє управляти процесом генерації за допомогою умов, таких як класи або атрибути. Це може бути корисним у випадках, коли потрібно створювати конкретні типи даних за умови певних обмежень або параметрів;

– Generative Pre-trained Transformer GPT – це модель глибокого навчання, яка використовується для генерації тексту, але може бути адаптована для інших типів даних, таких як зображення або анімації. Вона володіє великою гнучкістю у генерації різноманітних даних за допомогою контрольованого навчання.

1.3 Проблеми використання нейромереж для робітників

В останні роки штучний інтелект (ШІ) та нейромережі стрімко увірвалися в наше життя, трансформуючи різні сфери діяльності. Ігрова індустрія, як і багато інших, не залишилася осторонь цього технологічного вихору.

З одного боку, нейромережі відкривають захоплюючі можливості для створення ігор, про які раніше ми могли лише мріяти.

Генерація контенту:

- створення реалістичних світів: нейромережі можуть генерувати цілі світи, опрацьовуючи найдрібніші деталі ландшафтів, архітектури, флори та фауни, роблячи їх практично indistinguishable від реальних;
- персонажі, що ожили на очах: нейромережі здатні створювати реалістичних персонажів, які не тільки зовні схожі на людей, але й мають багатий внутрішній світ, складні емоції та мотивації;
- сюжети, повні інтриг: нейромережі можуть генерувати захоплюючі сюжети, історії, діалоги, наповнені драматизмом, несподіваними поворотами та персонажами, що запам'ятовуються;
- музика та звуки, що занурюють в атмосферу: нейромережі здатні генерувати музику та звукові ефекти, що ідеально відповідають настрою гри, її жанру та сюжету, створюючи неповторну атмосферу.

Персоналізація для кожного:

- ігровий процес, підлаштований під тебе: нейромережі можуть підлаштовувати складність гри, сюжетні лінії, поведінку персонажів та інші елементи під кожного гравця, роблячи його досвід максимально захоплюючим і відповідним перевагам;
- унікальні історії, засновані на твоїх виборах: нейромережі можуть генерувати унікальні історії, що розвиваються в залежності від дій гравця, створюючи відчуття повного контролю над тим, що відбувається, і роблячи кожну гру неповторною.

Нові горизонти для геймерів:

- нові жанри, народжені ІІ: нейромережі можуть стати каталізатором для створення принципово нових ігрових жанрів, які раніше були неможливі, розсуваючи межі геймінгу та даруючи гравцям небачені раніше враження. З іншого боку, використання нейромереж в ігровій індустрії породжує низку серйозних побоювань.

Загроза безробіття:

– автоматизація рутинних завдань: нейромережі можуть автоматизувати завдання, які виконуються художниками, аніматорами, сценаристами, композиторами, тестувальниками та іншими фахівцями, що може призвести до масових скорочень і зростання безробіття в ігровій індустрії;

– девальвація творчих професій: автоматизація творчих процесів може призвести до знецінення праці людей, які займаються створенням ігор, збіднення ігрового світу та втрати його унікальності.

Зниження якості та одноманітність:

– «штампування» контенту: нейромережі, навчені на обмежених наборах даних, можуть генерувати одноманітний та шаблонний контент, роблячи ігри бездушними, позбавленими оригінальності та індивідуальності;

– втрата людського фактора: автоматизоване створення ігор може призвести до втрати того «людського» елемента, який робить ігри по-справжньому зворушливими, душевними та незабутніми.

Непередбачуваність та ризики:

– генерація небезпечного контенту: нейромережі, які навчаються на нефільтрованих даних, можуть генерувати контент, що розпалює ненависть, пропагує насильство, що містить елементи дискримінації або інші неприйнятні матеріали;

– втрата контролю над ігровим процесом: нейромережі, здатні до самонавчання, можуть виходити з-під контролю, генеруючи непередбачувані та небезпечні ситуації в іграх, ставлячи під загрозу безпеку гравців.

Важливо, що використання нейромереж в ігровій промисловості знаходиться на ранній стадії свого розвитку.

Майбутнє цієї технології багато в чому залежить від того, як вона використовуватиметься.

Якщо нейромережі будуть застосовуватися розумно і відповідально, вони можуть стати потужним інструментом для створення неймовірно захоплюючих і унікальних ігор.

Необхідно проводити дослідження щодо взаємодії нейромереж з творчими фахівцями, що з'ясує, як вони можуть ефективно співпрацювати для створення ще якісніших ігор.

Наприклад, нейромережі могли б генерувати чернові варіанти світів, персонажів, сюжетів, музики, які потім допрацьовувалися художниками, сценаристами, композиторами, які привносять свій творчий погляд і емоційний інтелект.

Крім цього, нейромережі можуть виявитися корисними у тестуванні ігор. Вони здатні програвати ігри мільйони разів, виявляючи баги, нелогічні поведінки персонажів та сюжетні нестикування, які можуть бути пропущені тестувальниками – людьми.

Однак для того, щоб нейромережі стали справжніми партнерами розробників, а не їх заміною, необхідно вирішити низку важливих завдань: Розробка етичних норм використання нейромереж в ігровій промисловості. Потрібно визначити, який контент вважається неприйнятним та як нейромережі навчати його не генерувати.

Інвестування у перепідготовку спеціалістів. Необхідно допомогти людям, чії професії можуть бути автоматизовані, освоїти нові навички та навчитися працювати з нейромережами.

Підвищення грамотності гравців про нейромережі. Геймери повинні розуміти, яку роль грають нейромережі у створенні ігор, і які ризики та переваги це несе.

На закінчення можна сказати, що нейромережі мають величезний потенціал для трансформації ігрової промисловості. Однак важливо пам'ятати, що нейромережі – лише інструмент, і кінцевий результат залежить від того, як його використовують. Майбутнє ігор за співпрацею між людським творчим генієм і потужністю штучного інтелекту. Тільки

таким чином можна створювати ігри, які будуть не тільки захоплюючими, але й зворушливими, душевними і по –справжньому незабутніми.

Генерація адаптивних світів: нейромережі можуть створювати ігрові світи, які реагують на дії гравця. Світ може еволюціонувати залежно від прийнятих рішень, породжуючи відчуття живого простору, що дихає. Уявіть собі місто, де рівень злочинності зростає, якщо ви ігноруєте квести варті, або флору та фауну, що змінюється залежно від ваших екологічних рішень.

Неігрові персонажі з глибиною: нейромережі здатні генерувати неігрових персонажів з унікальними особистостями, пам'яттю та динамічними стосунками з гравцем. NPC можуть навчатися на взаємодії з гравцем, запам'ятовувати його вчинки та реагувати відповідно, створюючи відчуття справжнього спілкування.

Нелінійні сюжети, сповнені сюрпризів: нейромережі можуть генерувати нелінійні сюжети, що адаптуються під стиль гри. Сюжетні лінії можуть розвиватися зненацька, виходячи з дій гравця, а не заздалегідь прописаної колії. Уявіть собі квест, результат якого залежить від того, граєте ви потай або в лоб.

Процедурно-генерований геймплей: нейромережі здатні створювати унікальний геймплей щоразу. Це можуть бути рівні рівні, завдання, головоломки, вороги, що динамічно генеруються, забезпечуючи нескінченний потік новизни і повторного проходження.

Проте впровадження подібних технологій пов'язане з низкою труднощів: необхідність величезних масивів даних: Для навчання нейромереж генерації правдоподібних світів, персонажів і сюжетів будуть потрібні колосальні обсяги даних – тексту, зображень, аудіозаписів – що ставить питання про їх доступність та якість.

Проблема правдоподібності та когерентності: нейромережеві світи і сюжети, що генеруються, можуть виявитися нелогічними, суперечливими

або просто дивними. Потрібна розробка алгоритмів, що забезпечують когерентність та правдоподібність створюваного контенту.

Збереження авторського бачення: використання нейромереж для генерації сюжету та геймплей може розмити межі авторського бачення розробників. Необхідно знайти баланс між творчим контролем з боку людей та можливостями нейромереж.

Вплив нейромереж на ігрову промисловість неминуче. Питання не в тому, чи будуть вони використовуватися, а в тому, як це відбуватиметься. Оптимістичний сценарій передбачає співпрацю між людським творчим генієм та потужністю штучного інтелекту. Нейромережі стануть інструментом у руках розробників, що дозволяє їм створювати ігри небаченої раніше глибини, захоплюючої історії та неповторного геймплею. Песимістичний сценарій означає заміну творчих фахівців нейромережами, що призведе до стандартизації ігор, втрати оригінальності та зменшення емоційного впливу на гравців. Ігри перетворяться на бездушні продукти, позбавлені того людського елемента, який робить їх по –справжньому зворушливими та незабутніми.

Для того, щоб уникнути негативного сценарію, необхідно здійснити низку дій: Інвестиції у дослідження та розробки. Необхідно виділяти кошти на вивчення взаємодії нейромереж з творчим процесом, а також на розробку етичних норм використання ІІ в іграх.

Освіта та перепідготовка фахівців. Потрібно допомогти людям, чії професії можуть бути автоматизовані, освоїти нові навички, навчитися працювати з нейромережами та стати «гейм-дизайнерами майбутнього». Відкритий діалог між розробниками, гравцями та етичними комітетами. Важливо обговорювати потенційні ризики та переваги використання нейромереж, щоб знайти баланс між інноваціями та збереженням робочих місць, а також етичним використанням технологій.

Майбутнє ігрової промисловості залежить від того, наскільки відповідально та продумано ми підійдемо до впровадження нейромереж.

Нейромережі не повинні замінити людську творчість, а стати її потужним доповненням. Тільки спільними зусиллями ми зможемо створити ігри, які будуть не тільки технологічно досконаліми, але й захоплюючими, емоційними та по-справжньому незабутніми. Крім цього, нейромережі можуть вплинути на ігрову спільноту наступним чином:

Персоналізація ігрового досвіду: нейромережі можуть персоналізувати ігровий досвід як на рівні контенту, а й у взаємодії коїться з іншими гравцями. Наприклад, нейромережі можуть підбирати тиммейтів виходячи з вашого стилю гри та переваг у спілкуванні.

Створення інтерактивних ігрових світів, керованих спільнотою: нейромережі можуть використовуватися для створення ігрових світів, які частково формуються самими гравцями. Уявіть собі світ, де ландшафт, завдання та події частково генеруються виходячи з дій та голосувань ігрової спільноти.

Демократизація процесу створення ігор: нейромережі, будучи досить розвиненими, можуть дозволити не лише професійним розробникам, а й звичайним гравцям брати участь у створенні ігор. Уявіть собі платформу, де за допомогою нейромереж будь-хто охочий зможе розробити простий сюжет, персонажів або рівень для гри.

1.4 Постановка завдання дослідження

Виходячи з мети кваліфікаційної роботи, можна поставити такі завдання, що розкривають її тему:

- провести аналіз предметної галузі;
- провести теоретичне дослідження застосування ІІІ для ігор;
- обрати декілька моделей;
- порівняти роботу обраних нейронних мереж;
- проаналізувати та порівняти результати, отримані після роботи;
- зробити висновки на основі отриманих результатів.

2 ПОРІВНЯННЯ НЕЙРОННИХ МЕРЕЖ

2.1 Аналіз відмінностей між нейронними мережами

Створення нейронної мережі для генерації візуального контенту на основі аудіо-сигналів вимагає ретельного проектування архітектури моделі. Така архітектура (рисунок 2.1) зазвичай включає кілька ключових компонентів, які співпрацюють для досягнення бажаного результату.

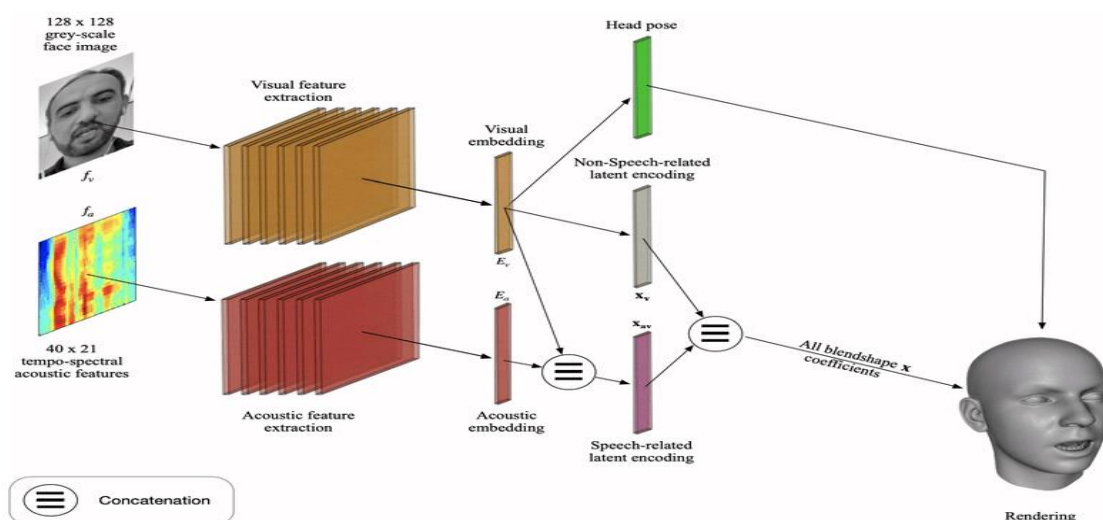


Рисунок 2.1 – Загальна архітектура Audio to Face Neural network

Перший компонент – це вхідний шар. Вхідний шар приймає аудіо-сигнал у вигляді числових даних або спектрограми. Спектрограма є найбільш поширеним форматом представлення аудіо для таких завдань, оскільки вона перетворює звукову хвилю у двовимірне зображення, яке легко обробляти за допомогою згорткових нейронних мереж. Спектрограма відображає інтенсивність частот звуку в часі і надає багату інформацію про структуру аудіо-сигналу.

Після вхідного шару йдуть шари витягнення ознак. Ці шари зазвичай складаються з глибоких згорткових нейронних мереж, які здатні

автоматично витягувати складні ознаки з вхідного аудіо –сигналу. Кожен шар згорткової мережі застосовує фільтри для вхідних даних і створює карти ознак, що представляють різні рівні абстракції звуку. Формально, вихід кожного шару можна описати формулою:

$$Hl = f(Wl * X + bl), \quad (2.1)$$

де HL – вихід l -го шару;

W^l – фільтри;

$*$ – операція згортки;

bl – зсув;

f – функція активації.

Одним з основних елементів архітектури для генерації візуальних даних є автоенкодер. Автоенкодер складається з двох частин: енкодера та декодера. Енкодер зменшує розмірність вхідних даних, витягуючи ключові ознаки, а декодер відновлює дані з цього зменшеного представлення. Формально, енкодер можна описати як:

$$z = \phi(x), \quad (2.2)$$

де x – вхідні дані, а ϕ – функція, що представляє енкодер. Декодер, у свою чергу, відновлює дані:

$$x^{\wedge} = \psi(z), \quad (2.3)$$

де ψ – функція, що представляє декодер.

Для генерації візуальних даних на основі аудіо часто використовуються варіаційні автоенкодери. VAE додають стохастичний елемент до процесу кодування, що дозволяє генерувати нові дані шляхом

вибірки з латентного простору. Формула для VAE-енкодера включає в себе середнє значення і стандартне відхилення:

$$z = \mu(x) + \sigma(x) \odot \epsilon, \quad (2.4)$$

де $\mu(x)$ і $\sigma(x)$ – функції для обчислення середнього значення і стандартного відхилення;

\odot – елементне множення;

ϵ – випадкова вибірка з нормального розподілу.

На виході мережа генерує візуальні дані, які можуть бути зображеннями облич, об'єктів або інших візуальних елементів, відповідно до завдання. Декодер відновлює ці візуальні дані з латентного простору, забезпечуючи збереження основних ознак, витягнутих з аудіо-сигналу.

Загальна архітектура також може включати рекурентні нейронні мережі, такі як LSTM або GRU, для обробки часових аспектів аудіо-сигналу. Це особливо корисно для завдань, де важливо враховувати послідовність і контекст аудіо. RNN можуть бути інтегровані як в енкодер, так і в декодер, для поліпшення здатності моделі генерувати точні та узгоджені візуальні дані.

Таким чином, архітектура моделі для генерації візуального контенту на основі аудіо-сигналів є комплексною системою, що включає згорткові нейронні мережі, автоенкодери, варіаційні автоенкодери та рекурентні нейронні мережі. Кожен компонент відіграє важливу роль у забезпеченні високої якості та точності генерованих даних.

Навчання моделі.

Після того як архітектура моделі визначена, наступним кроком є навчання моделі. Навчання включає декілька важливих етапів і потребує використання великих обсягів даних та спеціальних алгоритмів оптимізації.

Для навчання моделі, яка генерує візуальний контент на основі аудіо-сигналів, використовуються два основні набори даних: аудіо та відповідні

візуальні зображення. Ці дані можуть бути отримані з відео, де зображення обличчя синхронізовані з голосовими записами.

Навчання починається з попередньої обробки даних. Аудіо-сигнали перетворюються у спектрограми, які використовуються як вхідні дані для нейронної мережі. Зображення обличчя проходять процес нормалізації та масштабування для забезпечення консистентності даних. Ці оброблені дані потім використовуються для навчання моделі. Процес навчання включає наступні кроки.

Формулювання задачі оптимізації:

– основна задача полягає у мінімізації різниці між генерованими зображеннями та реальними зображеннями обличчя. Ця різниця вимірюється за допомогою функції втрат. Для цього завдання часто використовується комбінація різних функцій втрат, таких як середньоквадратична помилка (MSE) та функція перехресної ентропії;

– функція втрат може бути записана як:

$$L = E q(z | x) [\log p(x | z)] - DKL(q(z | x) || p(z)), \quad (2.5)$$

де $q(z|x)$ – апостеріорний розподіл;

$p(z)$ – апіорний розподіл;

Dkl – дивергенція Кульбака –Лейблера.

Алгоритм оптимізації:

– для оптимізації використовуються методи градієнтного спуску, такі як стохастичний градієнтний спуск або його варіанти, такі як Adam або RMSprop. Ці методи дозволяють ефективно знаходити мінімум функції втрат, коригуючи ваги моделі;

– оновлення ваг моделі на кожному кроці ітерації можна записати як:

$$\theta_{t+1} = \theta_t - \alpha \cdot \nabla J(\theta_t), \quad (2.6)$$

де θ – параметри моделі;

η – швидкість навчання;

∇J – градієнт функції втрат за параметрами моделі.

Дані розподіляються на тренувальні та тестові набори для оцінки продуктивності моделі. Тренувальні дані використовуються для навчання моделі, а тестові – для перевірки її здатності до генералізації.

Для запобігання перенавчанню застосовуються методи регуляризації, такі як відсічення (dropout) або регуляризація L2. Ці методи допомагають моделі уникати занадто точної підгонки до тренувальних даних і покращують її здатність до узагальнення. Після завершення навчання модель готова до генерації візуального контенту. Для цього на вхід моделі подається аудіо-сигнал (наприклад, спектрограма), і модель генерує відповідне зображення або послідовність зображень. Цей процес можна використовувати для створення реалістичної анімації облич, яка синхронізується з аудіо.

Оцінка якості згенерованого контенту є важливим етапом. Використовуються різні метрики для оцінки якості зображень, такі як структурна подібність або метрики, що базуються на глибоких мережах. Також можуть проводитися суб'єктивні тести, де люди оцінюють реалістичність та якість згенерованих зображень.

Одним з основних викликів при створенні таких моделей є забезпечення реалістичності згенерованих зображень. Звукові сигнали можуть містити шум, який впливає на якість генерації. Крім того, моделі можуть бути чутливими до варіацій у даних, таких як різні акценти або тональність голосу.

Іншим викликом є потреба у великих обсягах даних для навчання. Для досягнення високої якості моделі необхідно використовувати багато пар аудіо-зображення, що може бути складним у збиранні.

Розвиток таких нейронних мереж має великі перспективи у багатьох галузях, не лише в ігровій індустрії. Вони можуть використовуватися для

створення реалістичних аватарів у віртуальних середовищах, поліпшення систем синтезу мови, розробки інструментів для кіноіндустрії та багато іншого.

Отже, створення нейронної мережі для генерації візуального контенту на основі аудіо-сигналів включає комплексний підхід, що охоплює розробку архітектури моделі, навчання з використанням великих обсягів даних, оптимізацію і регуляризацію моделі, а також оцінку якості згенерованого контенту.

Процес навчання моделі.

Після визначення архітектури моделі важливо правильно організувати процес її навчання. Навчання моделі для генерації візуального контенту на основі аудіо-сигналів включає кілька основних етапів: підготовку даних, вибір функції втрат, оптимізацію параметрів, регуляризацію та оцінку якості.

Підготовка даних:

- збирання даних: для навчання моделі необхідно мати великий набір даних, що включає аудіо-сигнали та відповідні візуальні зображення. Ідеальним джерелом даних є відео, де обличчя людини видно під час розмови, що дозволяє синхронізувати аудіо та візуальні дані;

- попередня обробка аудіо: аудіо-сигнали перетворюються у спектрограми або інші представлення, які можуть бути оброблені нейронними мережами. Це включає нормалізацію, видалення шуму та розбиття на фрейми;

- попередня обробка зображень: візуальні дані обличчя масштабуються та нормалізуються для забезпечення консистентності. Це включає вирівнювання обличчя та масштабування до стандартного розміру.

Вибір функції втрат:

- основна мета навчання – мінімізувати різницю між згенерованими та реальними зображеннями. Для цього використовуються різні функції

втрат, такі як середньоквадратична помилка для порівняння пікселів або перехресна ентропія для класифікаційних задач;

- для варіаційних автоенкодерів використовується додаткова складова функції втрат – дивергенція Кульбака –Лейблера, яка допомагає збалансувати реконструкцію та генерацію нових зразків:

$$L = E_{q(z|x)}[\log p(x|z)] - DKL(q(z|x) || p(z)). \quad (2.7)$$

Оптимізація параметрів:

- оптимізація виконується за допомогою методів градієнтного спуску, таких як стохастичний градієнтний спуск, Adam або RMSprop. Ці методи оновлюють ваги моделі на основі градієнтів функції втрат;

- формула оновлення ваг для методу Adam виглядає наступним чином:

$$\theta_t = \theta_{t-1} - \alpha * (m^t \setminus (\sqrt{v^t} + \epsilon)), \quad (2.8)$$

де θ – параметри моделі;

η – швидкість навчання;

m^t і v^t – скориговані перші та другі моменти градієнтів, відповідно.

Регуляризація:

- регуляризація застосовується для запобігання перенавчанню моделі. Це включає методи, такі як відсічення (dropout) та регуляризація L2. Відсічення випадково вимикає нейрони під час навчання, а регуляризація L2 додає штраф за великі значення ваг у функцію втрат.

Оцінка якості:

- оцінка якості моделі проводиться за допомогою різних метрик, таких як структурна подібність для оцінки схожості зображень або LPIPS для оцінки перцептивної якості;

– оцінка проводиться на тестовому наборі даних, який не використовувався під час навчання, щоб перевірити здатність моделі до узагальнення. Основними викликами при створенні нейронної мережі для генерації візуального контенту є забезпечення реалістичності згенерованих зображень та узгодження їх з аудіо-сигналом. Шум у звукових даних, варіації в голосі та акценті можуть ускладнювати процес навчання та призводити до помилок у генерації.

Розвиток нейронних мереж для генерації візуального контенту має великі перспективи у різних галузях. Такі моделі можуть використовуватися не лише в ігровій індустрії, але й у віртуальній реальності, системах синтезу мови, кіноіндустрії та багатьох інших сферах. Завдяки постійному вдосконаленню методів навчання та збільшенню обсягів доступних даних, якість та реалістичність генерованого контенту продовжуватимуть зростати.

Таким чином, процес створення нейронної мережі для генерації візуального контенту на основі аудіо-сигналів є складним та багатокомпонентним. Він включає ретельне проектування архітектури, підготовку даних, оптимізацію та регуляризацію моделі, а також оцінку якості результатів. Використання таких моделей відкриває нові можливості для автоматизації творчих процесів та створення реалістичного та захоплюючого контенту.

Архітектура нейронної мережі для Audio to Face.

Створення нейронної мережі для генерації візуального контенту на основі аудіо-сигналів вимагає ретельного проектування архітектури моделі. Така архітектура зазвичай включає кілька ключових компонентів, які співпрацюють для досягнення бажаного результату.

Перший компонент – це вхідний шар. Вхідний шар приймає аудіо-сигнал у вигляді числових даних або спектрограми. Спектрограма є найбільш поширеним форматом представлення аудіо для таких завдань, оскільки вона перетворює звукову хвилю у двовимірне зображення, яке

легко обробляти за допомогою згорткових нейронних мереж. Спектрограма відображає інтенсивність частот звуку в часі і надає багату інформацію про структуру аудіо-сигналу.

Після вхідного шару йдуть шари витягнення ознак. Ці шари зазвичай складаються з глибоких згорткових нейронних мереж, які здатні автоматично витягати складні ознаки з вхідного аудіо-сигналу. Кожен шар згорткової мережі застосовує фільтри до вхідних даних і створює карти ознак, що представляють різні рівні абстракції звуку. Формально, вихід кожного шару можна описати формулою:

$$Hl = f(Wl * X + bl), \quad (2.9)$$

де h^l – вихід l -го шару;

W^l – фільтри;

$*$ – операція згортки;

$b^{(l)}$ – зсув, а f – функція активації.

Одним з основних елементів архітектури для генерації візуальних даних є автоенкодер. Автоенкодер складається з двох частин: енкодера та декодера. Енкодер зменшує розмірність вхідних даних, витягуючи ключові ознаки, а декодер відновлює дані з цього зменшеного представлення. Формально, енкодер можна описати як:

$$z = \phi(x), \quad (2.10)$$

де x – вхідні дані;

ϕ – функція, що представляє енкодер.

Декодер, у свою чергу, відновлює дані:

$$x^{\wedge} = \psi(z), \quad (2.11)$$

де ψ – функція, що представляє декодер.

Для генерації візуальних даних на основі аудіо часто використовуються варіаційні автоенкодер. VAE додають стохастичний елемент до процесу кодування, що дозволяє генерувати нові дані шляхом вибірки з латентного простору. Формула для VAE –енкодера включає в себе середнє значення і стандартне відхилення:

$$z = \mu(x) + \sigma(x) \odot \epsilon, \quad (2.12)$$

де $\mu(x)$ і $\sigma(x)$ – функції для обчислення середнього значення і стандартного відхилення;

\odot – елементне множення;

ϵ – випадкова вибірка з нормального розподілу.

На виході мережа генерує візуальні дані, які можуть бути зображеннями облич [15], об'єктів або інших візуальних елементів, відповідно до завдання.

Декодер відновлює ці візуальні дані з латентного простору, забезпечуючи збереження основних ознак, витягнутих з аудіо-сигналу [16]. Загальна архітектура також може включати рекурентні нейронні мережі [17], такі як LSTM або GRU, для обробки часових аспектів аудіо-сигналу. Це особливо корисно для завдань, де важливо враховувати послідовність і контекст аудіо [18].

RNN можуть бути інтегровані як в енкодер, так і в декодер, для поліпшення здатності моделі генерувати точні та узгоджені візуальні дані [19]. Таким чином, архітектура моделі для генерації візуального контенту на основі аудіо-сигналів є комплексною системою, що включає згорткові нейронні мережі [15], автоенкодери [16], варіаційні автоенкодери [16] та рекурентні нейронні мережі [17].

Кожен компонент відіграє важливу роль у забезпеченні високої якості та точності генерованих даних.

Навчання моделі: навчання моделі Audio to Face включає кілька важливих етапів, які забезпечують точне та реалістичне відтворення рухів обличчя на основі аудіо – сигналів. Процес навчання потребує великої кількості даних [20] та складних обчислювальних ресурсів [21].

Збір та підготовка даних.

Для навчання моделі необхідні великі обсяги даних, що включають відеозаписи людей, які говорять, з відповідними аудіо-доріжками. Ці дані повинні бути високої якості, щоб модель могла навчитися точно відповідати рухи обличчя до звукових сигналів.

Анотація та синхронізація.

Кожен кадр відео синхронізується з відповідним аудіо-сигналом, а також анотується для визначення ключових точок на обличчі, таких як положення губ, очей, брів та інших рухомих частин обличчя. Це дозволяє моделі навчитися зв'язку між аудіо – сигналами та візуальними змінами обличчя.

Попереднє оброблення даних.

Аудіо-сигнали перетворюються у спектрограми або інші форми представлення, що дозволяють легше витягати ознаки за допомогою згорткових нейронних мереж. Відеокадри обробляються для нормалізації освітлення та масштабування, щоб забезпечити послідовність даних.

Процес навчання моделі.

Процес навчання моделі складається з кількох кроків, які включають оптимізацію параметрів моделі та перевірку її продуктивності.

Ініціалізація моделі.

Модель ініціалізується з випадковими вагами. Це початкова стадія, де всі параметри моделі задаються випадковими значеннями перед початком навчання.

Проходження вперед (forward pass).

Аудіо-дані подаються на вхід моделі, і кожен шар обробляє ці дані, витягуючи ознаки та перетворюючи їх у рухи обличчя. На виході модель

генерує передбачення рухів обличчя, які порівнюються з реальними рухами з навчальних даних.

Обчислення функції втрат (loss function).

Відхилення між передбаченнями моделі та реальними даними обчислюється за допомогою функції втрат. У випадку анімації обличчя це може бути середньоквадратична похибка (MSE), яка обчислює середнє значення квадратів різниць між передбаченими та реальними положеннями ключових точок обличчя:

$$L = \frac{1}{N} \sum = \frac{1}{N} \sum (y_i - \hat{y}_i)^2, \quad (2.13)$$

де L – функція втрат;

N – кількість даних;

y_i – реальні положення точок;

\hat{y}_i – передбачені положення.

Проходження назад (backward pass) і оновлення ваг.

Використовуючи метод зворотного поширення помилки (backpropagation), модель обчислює градієнти функції втрат відносно ваг і оновлює ці ваги за допомогою оптимізатора (наприклад, Adam або SGD). Це дозволяє моделі зменшувати втрати і покращувати точність передбачень:

$$\theta_{t+1} = \theta_t - \alpha \cdot \nabla J(\theta_t), \quad (2.14)$$

де θ – параметри моделі;

η – швидкість навчання;

∇J – градієнт функції втрат за параметрами моделі.

Повторення етапів.

Цей процес повторюється для великої кількості епох (циклів навчання), доки модель не досягне прийнятної рівня точності.

Валідація та тестування.

Після навчання модель перевіряється на валідаційному наборі даних, який не був використаний під час навчання. Це дозволяє оцінити здатність моделі до узагальнення. Далі модель тестується на тестовому наборі даних для остаточної оцінки її продуктивності.

Таким чином, навчання моделі Audio to Face є складним процесом, який включає збір та підготовку даних, оптимізацію параметрів моделі і перевірку її точності та продуктивності на різних етапах. Цей процес забезпечує здатність моделі точно відтворювати рухи обличчя на основі аудіо-сигналів, створюючи реалістичні анімації.

Ми дізналися віддалено про пристрій звичайний нейронної мережі для генерації візуального контенту, також розібрали пристрій Audio to Face, а тепер давайте їх порівняємо:

- вхідні дані: обидві архітектури використовують аудіо –сигнали або спектрограми як вхідні дані;
- витягнення ознак: використання CNN для витягнення ознак є спільною рисою;
- латентне представлення: обидві архітектури можуть включати автоенкодерів або VAE для створення латентного представлення;
- декодування: обидві архітектури використовують декодери для генерації візуальних даних;
- часові аспекти: використання RNN для обробки часових аспектів також є спільним підходом.

Архітектура Audio to Face має багато спільного із загальною архітектурою нейронної мережі для генерації візуального контенту на основі аудіо-сигналів. Основні компоненти та методи є подібними, включаючи використання CNN для витягнення ознак, автоенкодерів для зменшення розмірності та декодерів для генерації візуальних даних. Однак специфіка Audio to Face полягає в її орієнтації на анімацію обличчя, що може вимагати додаткових налаштувань та оптимізацій для досягнення

високоякісної синхронізації між аудіо та візуальним контентом. Процесах навчання стандартної нейромережі та Audio to Face.

Вхідні дані.

Стандартна нейромережа:

– тип даних: використовуються статичні зображення або відео. Зображення можуть бути з анотованими мітками для задач класифікації, сегментації, або детекції об'єктів;

– аугментація даних: застосовуються різноманітні методи для збільшення обсягу навчальних даних, такі як обертання, зміна яскравості, контрасту, відзеркалення та інші техніки;

– розмір даних: навчальні набори часто складаються з тисяч або мільйонів зображень, таких як ImageNet.

Audio to Face:

– тип даних: основними вхідними даними є аудіо – сигнали у вигляді спектрограм та відповідні відеозаписи. Спектрограми надають двовимірне представлення аудіо, що полегшує обробку за допомогою згорткових нейронних мереж;

– синхронізація: кожен кадр відео повинен відповідати певному відрізку аудіо. Це вимагає точного вирівнювання аудіо та відео, що значно ускладнює підготовку даних;

– обробка аудіо: аудіо-сигнали проходять нормалізацію, шумоприглушення, а також перетворення у спектрограми або мел-спектрограми для подальшої обробки.

Стандартна нейромережа:

– структура: використовуються різні типи архітектур, залежно від завдання: згорткові нейронні мережі для обробки зображень, рекурентні нейронні мережі для обробки послідовностей даних, трансформери для обробки тексту;

– компоненти: в архітектуру можуть входити шари для обробки вхідних даних, приховані шари для витягування ознак, та вихідні шари для передбачення результатів.

Audio to Face:

– структура: поєднує компоненти для обробки аудіо та відео. Використовуються CNN для витягування ознак зі спектрограм, а також автоенкодері та варіаційні автоенкодері для зменшення розмірності та генерації;

– компоненти: включаються енкодері для витягування ключових ознак з аудіо, декодери для відтворення візуальних даних, а також рекурентні нейронні мережі для обробки часових аспектів аудіо-сигналів.

Функції втрат.

Стандартна нейромережа:

– класифікація: використовується крос-ентропія для задач класифікації:

$$L = -\sum_i \log(y^i); \quad (2.15)$$

– регресія: використовується середньоквадратична похибка (MSE) для задач регресії:

$$L = \frac{1}{N} \sum (y_i - \hat{y}_i)^2. \quad (2.16)$$

Audio to Face:

– рухи обличчя: використовується MSE для узгодження передбачених рухів обличчя з реальними:

$$L = \frac{1}{N} \sum (x_i - \hat{x}_i)^2; \quad (2.17)$$

– додаткові функції: включають Kullback-Leibler divergence у варіаційних автоенкодер для забезпечення стохастичності у процесі генерації:

$$KL = -1 \ln \sum = (1 + \log(\sigma^2) - \mu^2 - \sigma^2). \quad (2.18)$$

Процеси навчання стандартних нейромереж та моделей Audio to Face значно відрізняються через специфіку даних, архітектури моделі та функцій втрат.

Вхідні дані: для стандартних нейромереж використовуються зображення або відео, тоді як для Audio to Face необхідні аудіо-сигнали та їх точна синхронізація з відео.

Обробка даних: Audio to Face потребує додаткової обробки аудіо-сигналів та їх перетворення у спектрограми, що додає складності у підготовці даних.

Архітектура: Audio to Face поєднує в собі компоненти для обробки аудіо та відео, включаючи рекурентні нейронні мережі для роботи з часовими аспектами.

Функції втрат: у Audio to Face використовуються складніші функції втрат, які включають як MSE для узгодження рухів обличчя, так і Kullback-Leibler divergence для генеративних аспектів.

Ці відмінності показують, що навчання моделі Audio to Face є більш в комплексним процесом, що вимагає значних обчислювальних ресурсів та спеціалізованих методів обробки даних, порівняно з навчанням стандартних нейромереж.

А тепер поглянемо на таблицю порівнянь 2 нейромереж (таблиця 2.1).

Таблиця 2.1 – Порівняння нейронних мереж

Критерій	Audio to Face	Deep Video Portraits	Переможець
Навчання даних	Не потребує навчання даних	Потребує великого обсягу даних для навчання	Audio to Face
Автоматизація	Автоматично створює анімацію облич	Вимагає ручного втручання	Audio to Face
Гнучкість	Адаптація анімації до різних голосів та емоцій	Гнучкість у налаштуванні анімацій	Нічия
Вимоги до ресурсів	Помірні вимоги до обчислювальних ресурсів	Вимагає значних обчислювальних ресурсів	Audio to Face
Точність виразності	Обмежена точність при складних емоціях	Висока точність емоційної експресії	Deep Video Portraits
Інтеграція в проект	Потребує серверної обробки	Може бути інтегрована локально	Deep Video Portraits
Універсальність	Може застосовуватися в різних сферах	Підходить для кіноіндустрії та відеоігор	Audio to Face
Емоційна експресія	Висока точність емоцій та реакцій	Висока точність емоцій та реакцій	Нічия

3 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ

3.1 Вибір технології створення лицьових анімацій

Технологія «Audio to Face» від Nvidia є інноваційним підходом до анімації обличчя на основі аудіо сигналів. Вона використовує нейронні мережі для аналізу звукових входів та генерації відповідних рухів та виразів обличчя цифрових персонажів. За допомогою плагіна Neuralink у Unreal Engine 5 ми можеш синхронізувати роботу Audio to Face від Nvidia та лицьову анімацію нашого персонажа в грі, за допомогою аудіо доріжок ми можемо в прямому ефірі записувати високоточні лицьові анімації для NPC.

Deep Video Portraits це передова нейронна мережа, розроблена для високоякісної та реалістичної анімації відеопортретів. Вона дозволяє змінювати положення голови, вирази обличчя та напрямок погляду людини на відео, зберігаючи при цьому високий рівень деталізації та природність рухів.

3.2 Використання нейронних мереж для створення лицьових анімацій

У цьому дослідженні я використовував технологію Audio to Face від Nvidia Omniverse для створення анімацій обличчя на основі аудіо доріжок. Процес інтеграції та використання цієї технології включав кілька ключових етапів. Спершу я завантажив програму Nvidia Omniverse з офіційного веб-сайту Nvidia. Після завантаження пакету програмного забезпечення я помістив папку OmniverseNeuroLink у директорію плагінів Unreal Engine 5, що забезпечило інтеграцію між Omniverse та Unreal Engine 5.

Наступним кроком було запуснути програмне забезпечення Unreal Engine 5 та завантажити плагін Audio to Face Omniverse. Після успішного

запуску програмного забезпечення було підготовлено середовище для роботи з анімаціями обличчя.

Я використав базову 3D модель обличчя, яка йде в комплекті для навчання користувачів цієї технології. Ця модель була завантажена у робоче середовище Unreal Engine 5.

Після підготовки 3D моделі я завантажив аудіо доріжку у форматі MP3. Після цього я включив аудіо доріжку, і програмне забезпечення Audio to Face автоматично створило анімацію обличчя, підлаштовуючи вирази під звукові дані. Ця анімація в режимі реального часу транслювалася в редактор Unreal Engine 5, що дозволило відразу бачити результати роботи нейромережі та корегувати їх у разі необхідності.

У цьому дослідженні я використовував технологію Deep Video Portraits для створення реалістичних анімацій обличчя на основі відео даних. Процес інтеграції та використання цієї технології включав кілька ключових етапів. Спершу я завантажив програмне забезпечення Deep Video Portraits з офіційного веб-сайту розробників. Я використовував стокову версію, яка надається для ознайомлення з технологією та навчання користувачів.

Після завантаження програмного забезпечення я встановив його на своєму комп'ютері, дотримуючись інструкцій, наданих розробниками. Програмне забезпечення включає всі необхідні компоненти для обробки відео та генерації анімацій обличчя.

Після успішної інсталяції я запустив програму та підготував середовище для роботи з анімаціями обличчя.

Для початку роботи з Deep Video Portraits я завантажив відео зображення, що входять у стокову версію, яка надається на офіційному сайті для ознайомлення з технологією. Це відео містило різноманітні зразки виразів обличчя, які були необхідні для навчання нейронної мережі. Програма автоматично обробила завантажене відео, розпізнаючи ключові точки обличчя та створюючи відповідні анімації. Першим кроком було

забезпечення високої якості вхідного відео. Це відео було оброблене для виділення ключових ознак, таких як позиції очей, рота, носа та інших частин обличчя. Використовуючи ці дані, нейронна мережа навчалася відтворювати рухи та вирази обличчя, досягаючи високого рівня реалістичності. Процес обробки включав декілька етапів: попередня обробка відео, де відео було розділено на окремі кадри, і кожен кадр проходив через процес нормалізації та масштабування; розпізнавання ключових точок обличчя, де кожен кадр аналізувався для визначення ключових точок обличчя, таких як положення очей, рота, носа і підборіддя; навчання моделі, використовуючи розпізнані ключові точки та відповідні вирази обличчя, модель навчалася відтворювати ці вирази з високою точністю. Було важливо, щоб модель правильно відображала навіть найдрібніші деталі рухів обличчя; генерація анімацій, після завершення навчання модель могла генерувати реалістичні анімації обличчя на основі нових вхідних даних.

Після обробки відео я міг спостерігати результати роботи технології в режимі реального часу. Deep Video Portraits продемонструвала високу точність у відтворенні виразів обличчя та адаптації до різних стилів. Використання стокової версії спрощувало процес ознайомлення з технологією, забезпечуючи всі необхідні інструменти для створення реалістичних анімацій.

Цей підхід продемонстрував ефективність та зручність технології Deep Video Portraits у створенні реалістичних анімацій обличчя на основі відео даних, а також її потенціал для широкого спектру застосувань у різних галузях. Зокрема, технологія Deep Video Portraits знаходить своє застосування в кіноіндустрії, де її використовують для створення візуальних ефектів та анімації персонажів. Вона також може бути корисною в розробці відеоігор, забезпечуючи більш реалістичну та емоційно насичену взаємодію з гравцями. У сфері освіти та тренінгів ця технологія може використовуватись для створення інтерактивних навчальних матеріалів, що

допомагають краще передати інформацію та залучити аудиторію. Крім того, у медичній галузі Deep Video Portraits можуть сприяти розвитку нових методів реабілітації та терапії, які використовують реалістичні віртуальні аватари для спілкування з пацієнтами. Загалом, потенціал цієї технології є надзвичайно великим, відкриваючи нові горизонти для інновацій та вдосконалення в багатьох сферах.

3.3 Аналіз згенерованих лицьових анімацій

Якість згенерованого контенту – відзначено відмінну синхронізацію рухів губ з аудіо, однак, деякі артефакти виникали при складних емоційних виразах. Це вказує на те, що мережа добре справляється із завданнями озвучення персонажів, однак може потребувати додаткового налаштування для забезпечення повної реалістичності при складних сценаріях.

Реалістичність анімацій – точна синхронізація з аудіо, але деякі незначні артефакти при складних мимічних виразах. Зокрема, при генерації анімацій для швидких або різких змін голосових інтонацій мережа іноді демонструвала незначні неточності, що потребують корекції. Однак, загальний рівень реалістичності залишався високим, що робить цю технологію придатною для широкого застосування в ігровій індустрії.

Час генерації – швидша генерація, що робить її привабливою для інтерактивних додатків, але з дещо меншою деталізацією. Ця мережа є оптимальним вибором для проектів, де швидкість є критично важливою, наприклад, у режимах реального часу або для великих обсягів контенту.

Результат генерації (рисунок 3.1).

Якість згенерованого контенту – висока якість деталізації обличчя, реалістичність рухів очей та губ, а також точну передачу емоцій. Зокрема, при зміні положення голови та виразів обличчя мережа демонструвала стабільно високий рівень деталізації.

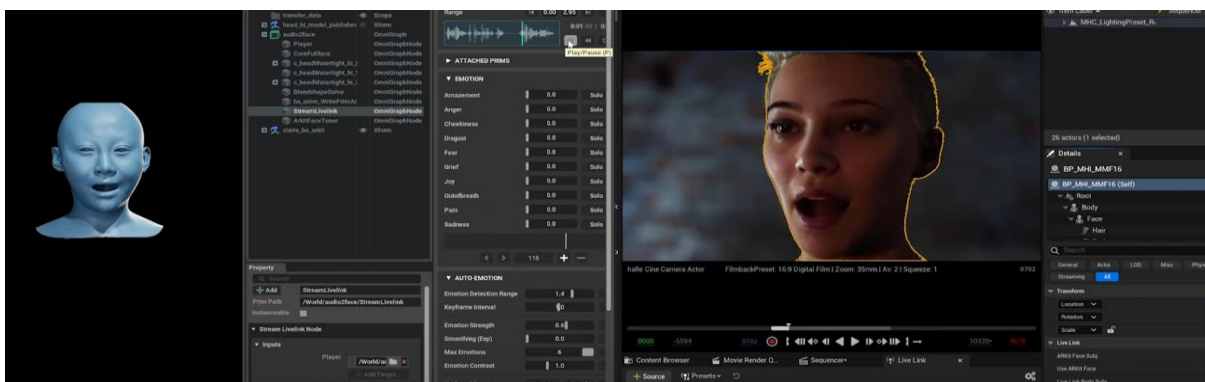


Рисунок 3.1 – Результат використання Omniverse Audio to face у Unreal Engine 5

Приклади генерації показали відмінну роботу при передачі емоційних реакцій, таких як усмішка, гнів чи здивування, що робить цю мережу надзвичайно цінною для створення інтерактивного контенту в ігровій індустрії.

Реалістичність анімацій – висока реалістичність у всіх тестових випадках, особливо при зміні положення голови та виразів обличчя. Мережа демонструвала здатність точно відтворювати як дрібні мімічні деталі, так і загальні рухи голови. В експериментах з довгими відеопослідовностями мережа зберігала консистентність та плавність анімацій, що є критичним для створення кінематографічного контенту в іграх.

Час генерації – потребує більше обчислювальних ресурсів через складність модифікацій, але забезпечує високу якість анімацій. Використання більш потужного обладнання може значно знизити час генерації, однак, навіть при стандартних умовах, ця мережа демонструє прийнятні результати для високоякісних проєктів, що вимагають ретельної деталізації та плавності рухів. Окрім цього, мережа здатна адаптуватися до різних стилів анімації, забезпечуючи універсальність і гнучкість у використанні. Зокрема, інтеграція з іншими інструментами та програмним забезпеченням дозволяє оптимізувати робочий процес і покращити

кінцевий продукт. Таким чином, навіть при значних обчислювальних витратах, ця неймережа є ефективним інструментом для створення реалістичних та візуально привабливих анімацій, що відповідають сучасним стандартам індустрії.

Результат генерації (рисунок 3.2).



Рисунок 3.2. – Результат використання Deep Video Portraits

В результаті експерименту було виявлено, що обидві нейронні мережі мають свої переваги та недоліки. Deep Video Portraits краще підходить для задач, де важлива загальна реалістичність вигляду та міміки, особливо при зміні положення голови та виразів обличчя. Audio to Face є більш ефективною у задачах синхронізації рухів губ з аудіо, що є критичним для озвучування персонажів у відеоіграх.

ВИСНОВКИ

Нейромережі стрімко завойовують ігрову індустрію, відкриваючи захоплюючі можливості для створення ігор нового рівня. Вони здатні генерувати реалістичні світи, персонажів, сюжетні лінії, музику та звуки, а також персоналізувати ігровий процес для кожного гравця. Це може призвести до створення принципово нових жанрів ігор, які раніше були неможливі. Однак використання нейромереж також викликає ряд побоювань. Автоматизація рутинних завдань може призвести до втрати робочих місць, а автоматизація творчих процесів – до знецінення праці людей, що займаються створенням ігор. Крім того, існує ризик генерування одноманітного контенту, втрати «людського фактора» та непередбачуваних наслідків.

Важливо усвідомлювати як потенційні переваги, так і ризики використання нейромереж в ігровій індустрії. Їх інтеграція може стати потужним інструментом для створення захоплюючих і унікальних ігор, але при цьому важливо не втрачати контроль над процесом розробки та не допускати знецінення людської творчості. Майбутнє нейромереж в ігровій індустрії залежить від розумного та відповідального підходу до їх використання. Необхідно розробити етичні норми, інвестувати в перепідготовку кадрів та вести відкритий діалог між розробниками, гравцями та етичними комітетами. Тільки так ми зможемо максимально використовувати потенціал нейромереж для створення ігор, які будуть не лише технологічно досконаліми, але й емоційно зворушливими, захоплюючими та по-справжньому незабутніми.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Головна сторінка DSpace. URL: <https://ir.nmu.org.ua/bitstream/handle/123456789/164206/Тиждень%20студентської%20науки%20-2023-340-342.pdf?sequence=1&isAllowed=y> (дата звернення: 03.06.2024).
2. Homepage – Hello Games. Hello Games. URL: <https://hellogames.org/> (date of access: 03.06.2024).
3. Shepherd I. AI – Powered Teens Are Building The Future In This Virtual World. Forbes. URL: <https://www.forbes.com/sites/ianshepherd/2024/04/29/ai-powered-teens-are-building-the-future-in-this-virtual-world/> (date of access: 03.06.2024).
4. Quantic Dream – Official Site. Quantic Dream – Official Site. URL: <https://www.quanticroam.com/en> (date of access: 03.06.2024).
5. Веб ресурс: III в іграх: майбутнє геймінгу. URL: <https://venturebeat.com/games/ai-and-the-future-of-gaming/> (date of access: 03.06.2024).
6. Laur. How to unlock Zaintiraris – Elder Scrolls Online, 2018. YouTube. URL: <https://www.youtube.com/watch?v=E1fqh7Msa-w> (date of access: 03.06.2024).
7. Deep Learning for Video Game Playing. URL: <https://arxiv.org/pdf/1708.07902> (дата звернення: 03.06.2024).
8. NVIDIA DLSS – Technologie. NVIDIA. URL: <https://www.nvidia.com/en-us/geforce/technologies/dlss/> (date of access: 03.06.2024).
9. Contributors to Wikimedia projects. Game physics – Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Game_physics (date of access: 03.06.2024).

10. Half-Life 2 – Valve Developer Community. Valve Developer Community. URL: https://developer.valvesoftware.com/wiki/Half-Life_2 (date of access: 03.06.2024).

11. Dalio M. Matt Dalio on LinkedIn: Could AI –enhanced games deliver infinitely scalable learning?. LinkedIn: Log In or Sign Up. URL: https://www.linkedin.com/posts/mattdalio_could-ai-enhanced-games-deliver-infinitely-activity-7191581454953291776-5X0i (date of access: 03.06.2024).

12. Веб ресурс: Квантові обчислення та машинне навчання в іграх. URL: https://www.researchgate.net/publication/333995335_Applications_of_data_science_to_game_learning_analytics_data_A_systematic_literature_review (date of access: 03.06.2024).

13. Веб ресурс: III та майбутнє розробки ігор. URL: <https://venturebeat.com/games/ai-and-the-future-of-gaming/> (date of access: 03.06.2024).

14. How can game AI personalize the gaming experience for players?. LinkedIn: Log In or Sign Up. URL: <https://www.linkedin.com/advice/0/how-can-game-ai-personalize-gaming-experience-players-2nugf> (date of access: 03.06.2024).

15. Contributors to Wikimedia projects. Convolutional neural network – Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Convolutional_neural_network (date of access: 03.06.2024).

16. Contributors to Wikimedia projects. Autoencoder – Wikipedia. Wikipedia, the free encyclopedia. URL: <https://en.wikipedia.org/wiki/Autoencoder> (date of access: 03.06.2024).

17. Contributors to Wikimedia projects. Recurrent neural network – Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/recurrent_neural_network (date of access: 03.06.2024).

18. A current value Hamiltonian Approach for Discrete time Optimal Control Problems arising in Economic Growth. arXiv.org. URL: <https://arxiv.org/abs/1801.03637> (date of access: 03.06.2024).

19. Attention Is All You Need. arXiv.org. URL: <https://arxiv.org/abs/1706.03762> (date of access: 03.06.2024).

20. Contributors to Wikimedia projects. Big data – Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Big_data (date of access: 03.06.2024).

21. Contributors to Wikimedia projects. Cloud computing – Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Cloud_computing (date of access: 03.06.2024).

