

<https://doi.org/10.31891/2307-5732-2025-357-59>

УДК 004.89

ПРИЙМА АНДРІЙ

Хмельницький національний університет

<https://orcid.org/0009-0001-7272-007X>

e-mail: andriipriyma@gmail.com

МАНЗЮК ЕДУАРД

Хмельницький національний університет

<https://orcid.org/0000-0002-7310-2126>

e-mail: eduard.em.km@gmail.com

ПАСІЧНИК ОЛЕКСАНДР

Хмельницький національний університет

<https://orcid.org/0000-0002-8760-4688>

e-mail: o.a.pasichnyk@gmail.com

СКРИПНИК ТЕТЯНА

Хмельницький національний університет

<https://orcid.org/0000-0002-8531-5348>

e-mail: tskripnik1970@gmail.com

МЕТОД ГЕНЕРАЦІЇ БАГАТОТРЕКОВИХ СИМВОЛІЧНИХ КОМПОЗИЦІЙ ЗА ДОПОМОГОЮ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ

В роботі наведено результат дослідження методу генерації багатотрекових символічних композицій за допомогою генеративного штучного інтелекту. Досліджено метод генерації композицій та проведено експерименти щодо його ефективності та якості моделей при різних умовах генерації.

Ключові слова: штучний інтелект, музика, музична композиція, партитура, piano-roll, модель.

PRYIMA ANDRII, MANZIUK EDUARD, PASICHNYK OLEKSANDR, SKRYPNYK TETIANA

Khmelnytskyi National University

A METHOD FOR GENERATION OF MULTITRACK SYMBOLIC COMPOSITIONS USING GENERATIVE ARTIFICIAL INTELLIGENCE

This study explores a method for symbolic music generation using artificial intelligence techniques. Unlike traditional rule-based or statistical approaches, the proposed method employs deep generative models to automatically create multi-track, polyphonic musical compositions.

As part of the research, a comparative experiment was conducted to assess the quality of music generation by different architectural variants of the models. The main task of this stage was to determine the ability of the models to reproduce key characteristics of musical logic.

The focus is placed on capturing the temporal structure of music and ensuring harmonic and rhythmic coherence across multiple instrument tracks. A model architecture is introduced that generates music in the form of piano-roll representations, treating musical bars as fundamental units. The system supports both autonomous generation and human-AI collaboration, where the model can accompany a user-provided melody with additional tracks. A specially prepared dataset of symbolic music was used for training, and a series of objective metrics were developed to evaluate the quality of the generated output.

The proposed architecture ensures the formation of coherent musical phrases while preserving rhythm, harmony, and inter-track interaction. Analysis of intermediate results showed a gradual increase in the quality of generation: from chaotic sounds to ordered structures, which indicates the effectiveness of training. Results from both quantitative analysis and a subjective user study demonstrate that the method is capable of producing musically coherent and structurally consistent compositions. This research contributes to the development of AI-assisted tools for creative music production and expands the possibilities of automated composition in the symbolic domain.

Keywords: artificial intelligence, music, musical composition, score, piano-roll, model.

Стаття надійшла до редакції / Received 11.07.2025

Прийнята до друку / Accepted 15.08.2025

Постановка проблеми

У сучасному світі штучний інтелект (ШІ) дедалі активніше проникає у сферу творчості, зокрема у створення музики. Генерація музичних композицій за допомогою ШІ відкриває нові можливості для автоматизації творчих процесів, підтримки музикантів та створення інноваційного музичного контенту. Проте, попри значний прогрес у галузі генеративних моделей, завдання створення якісної, багатотрекової та поліфонічної символічної музики залишається складним і недостатньо вирішеним. Основна проблема полягає в тому, що музика є не лише послідовністю звуків у часі, але й структурованим набором взаємопов'язаних елементів – гармонії, ритму, мелодії та інструментального балансу. Більшість моделей демонструють успіх у генерації простих мелодій або одноголосних треків, однак не враховують повною мірою складну взаємодію між різними інструментальними партіями. Це призводить до створення фрагментованих або гармонічно несумісних результатів.

Актуальність теми зумовлена потребою в розвитку інтелектуальних систем, здатних не лише імітувати, але й творчо доповнювати процес музичної композиції. Дослідження в цій сфері сприяють розширенню можливостей для музичної індустрії, освіти, інтерактивних медіа та персоналізованого аудіоконтенту. У контексті активного розвитку генеративних нейромереж тема є перспективною як у

науковому, так і в практичному аспектах.

Аналіз останніх джерел

У роботі [1] розглядається тема автоматизованої генерації музики за допомогою штучного інтелекту як одна з найбільш привабливих та значущих у сфері ШІ, музики та мультимедіа. Автори підкреслюють, що сучасне створення музики за участю ШІ охоплює широкий спектр задач, зокрема генерацію мелодій, написання пісень, створення акомпанементу, аранжування, виконання, генерацію тембру, синтез звуку та навіть синтез співу. У роботі пропонується систематизований підхід до теми: спочатку подається вступ до основ музичної теорії та технік глибокого навчання, потім розглядаються основні складові системи ШІ-композиції (створення партитур, виконання та звукогенерація), а також розширені теми, як-от моделювання музичної структури, стилю, емоцій та тембру. Завершується робота аналізом викликів і можливих напрямів розвитку у галузі. Основна ідея – надати широкому колу дослідників та практиків уявлення про поточний стан і майбутнє штучного інтелекту в музиці.

Робота [2] починається з констатації того, що хоча межі штучного інтелекту ще не повністю визначені, комп'ютери вже здатні виконувати завдання, які колись вважались притаманними виключно людям-музикантам. Після короткого історичного вступу автор наголошує на важливості застосування ШІ у чотирьох основних напрямках музичних досліджень: композиція, виконання, музикознавство та цифрова обробка звуку. Далі подається огляд сучасних досягнень у цих галузях, що демонструють практичне використання ШІ. Завершальна частина статті розмірковує над тим, як планування та навчання в системах штучного інтелекту можуть сприяти накопиченню знань і покращенню поведінки інтелектуальних музичних систем, тобто розширенню їхніх можливостей у створенні та інтерпретації музики.

У роботі [3] розглядаються три основні типи комп'ютерних музичних систем, які базуються на ШІ: системи для композиції, імпровізації та виконання. Після короткого опису прикладів кожного типу основна увага приділяється темі музичної виразності – однієї з найскладніших задач комп'ютерного музикування. Автори зазначають, що головним викликом у моделюванні виразності є відтворення «дотику виконавця», тобто його інтерпретаційного знання. Попередні методи, які спиралися на формальні правила, виявилися недостатніми. Натомість пропонується підхід, що наближається до людського навчання – наслідування шляхом спостереження, з використанням прикладів реальних виконавців. Описано систему SAXEX, яка здатна створювати якісні, емоційно забарвлені джазові соло, базуючись на записах людських виконавців, використовуючи методику Case-Based Reasoning.

Метою роботи: покращення якості та різноманітності генерованих музичних композицій шляхом застосування методів генеративного штучного інтелекту.

Виклад основного матеріалу

Розроблений метод використовує поетапну генерацію музичних композицій для створення однієї цілої композиції у вигляді послідовності тактів. Схему методу наведено на рисунку 1.

Увесь процес організовано у вигляді чотирьох логічно взаємопов'язаних етапів, кожен з яких виконує свою специфічну роль у створенні фінального музичного фрагмента. На вході система отримує шумові вектори, які виступають джерелом стохастичності та варіативності для майбутньої композиції. Крім того, система може (але не зобов'язана) отримати попередньо підготовлений трек, який слугує умовою для генерації акомпанементу. Ще одним обов'язковим елементом вхідних даних є параметри, що визначають кількість тактів у композиції та перелік інструментів, які будуть задіяні у процесі генерації.

розвиток музичної ідеї з часом.

Другий етап пов'язаний із генерацією внутрішніх шумових векторів для кожного інструментального треку. Система ініціалізує окремі вектори для кожного треку, беручи до уваги кількість інструментів і їхні типи, які були задані у параметрах. Для кожного інструмента створюється індивідуальний набір латентних шумових векторів, які мають передати характер гри цього інструмента, стилістичні й ритмічні особливості. Особливістю цього етапу є те, що усі вектори зберігаються у стабільній, узгодженій формі, яка буде використана на наступних етапах генерації.

На третьому етапі здійснюється комбінування різних типів векторів для створення повноцінних музичних тактів. Для цього об'єднуються часові вектори (отримані на першому етапі), вектори, пов'язані з окремими треками (створені на другому етапі), а також загальні шумові вектори у єдину умову генерації. Цей об'єднаний вектор подається до генератора відповідного інструментального треку для кожного окремого такту. Генератор формує фрагмент у форматі *ripo-roll*, який є поширеним представленням символічної музики, де кожна нота відображається у часі та по висоті. Таким чином, утворюються окремі музичні фрагменти, які поєднують інформацію про структуру, стиль, інструмент та момент виконання. Цей етап повторюється для кожного наступного такту, поки не буде охоплено всю задану довжину композиції.

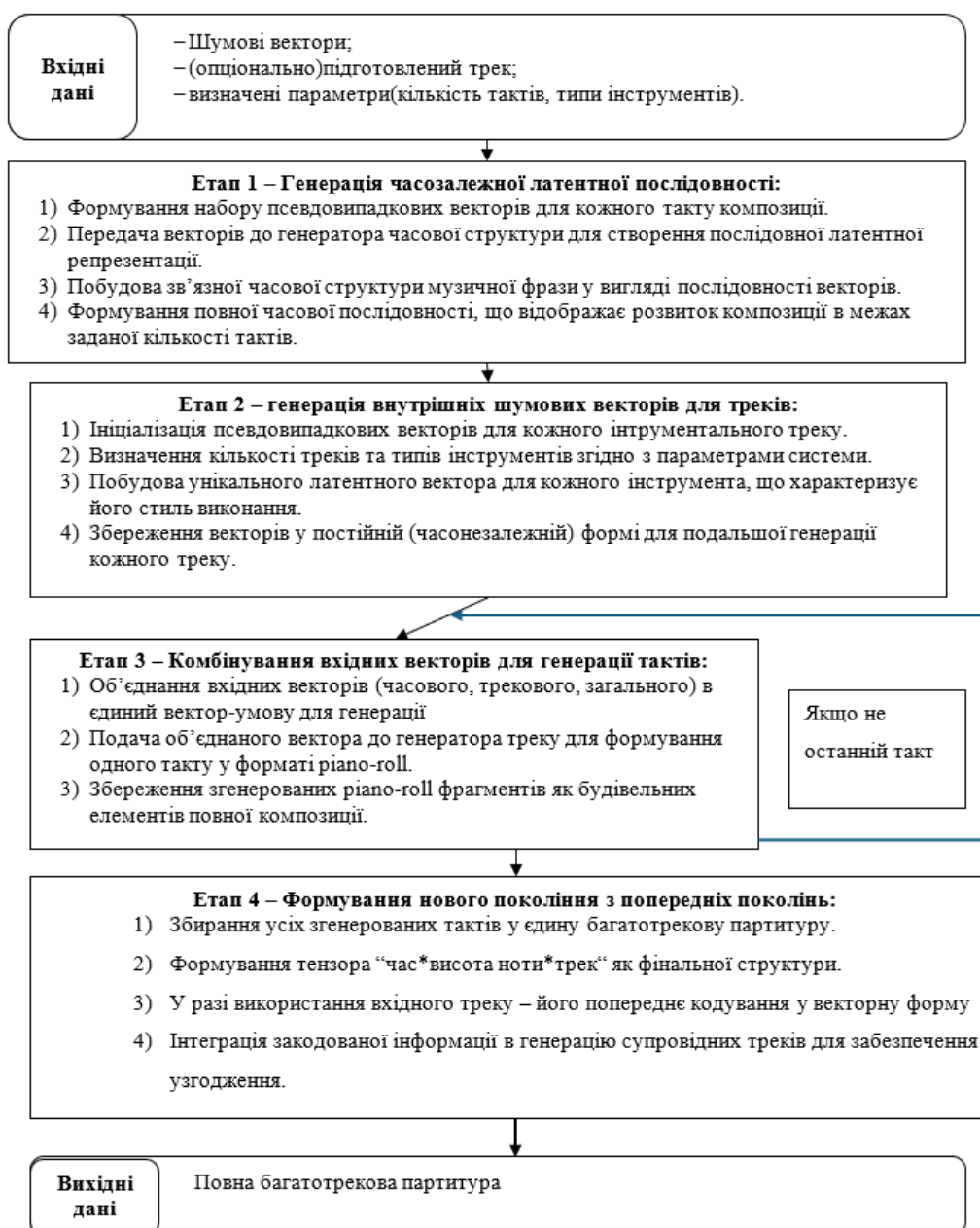


Рис. 1. Схема методу генерації багатотрекових символічних композицій

Перший етап – це генерація узгодженої латентної послідовності. Його суть полягає у формуванні набору векторів, які представляють собою часову структуру музичної фрази. Кожен такт композиції отримує відповідний вектор, що описує його положення у музичному часі, відображає ритмічну та структурну організацію. Ці вектори передаються до генератора часової структури, який формує повну послідовність – вона слугує “скелетом” композиції, її загальною формою. Результатом цього етапу є серія узгоджених векторів, які відповідають кількості заданих тактів і відображають

Четвертий і фінальний етап присвячений збиранню окремих тактів у завершену багатотрекову партитуру. Усі згенеровані фрагменти інтегруються у єдину музичну структуру, де враховується розташування нот у часі, розподіл між треками, а також їхня послідовність. У випадку, якщо на вхід була подана умова у вигляді попередньо підготовленого треку, він кодується у векторну форму і також береться до уваги під час формування фінального результату. Одним з важливих аспектів цього етапу є інтеграція інформації про часову структуру, яка була сформована на початку, для забезпечення логічної та музичної узгодженості між усіма треками.

У результаті роботи всього методу на виході формується повна багатотрекова партитура, яка може містити кілька інструментальних партій, розподілених у часі, та бути використаною для подальшого відтворення, аналізу або музичної обробки. Уся система побудована з урахуванням музичної структури, ритму, гармонії та варіативності, що забезпечує гнучкість і здатність до творчої генерації музики з урахуванням заданих параметрів і умов.

Також, у методі є свої групи функцій, які зображені на рисунку 2. На цій схемі представлено логічно структурований процес генерації музичної композиції із застосуванням штучного інтелекту.

Весь процес поділено на чотири послідовні групи функцій, які описують повний цикл – від ініціалізації генерації до виведення фінального результату у форматі, придатному для прослуховування або збереження.

Перший блок – **ініціалізація параметрів генерації** – є підготовчим етапом, у якому система отримує початкові умови для формування майбутньої композиції. Тут визначається кількість тактів, що визначає тривалість музичного фрагменту. Потім обирається кількість інструментальних треків – це дозволяє системі знати, скільки різних інструментів братиме участь у композиції. Далі здійснюється вибір типів інструментів: це може бути набір із таких типових партій, як бас, ударні, гітара, фортепіано тощо. Нарешті, ініціалізуються випадкові шумові вектори – набір числових параметрів, які вводять стохастичність у процес генерації і забезпечують різноманітність музичних результатів.



Рис. 2. Функціональна схема методу

Другий блок – **генерація структури композиції** – відповідає за створення абстрактного «каркасу» майбутньої музики. На цьому етапі формується часова послідовність латентних векторів, які визначають порядок розвитку композиції, зокрема ритмічну структуру і фразування. Паралельно генеруються незмінні шумові вектори, які залишаються сталими протягом усього процесу генерації і забезпечують загальну стилістичну цілісність композиції. Потім усі вхідні вектори (структурні, трекові, загальні) комбінуються в єдину структуру, яка буде використана як умова для побудови конкретних музичних елементів.

У третьому блоці – **побудова музичних тактів** – система переходить до генерації конкретного вмісту композиції. Зкомбіновані вектори подаються до генератора тактів, який буде віртуальне представлення нот у вигляді piano-roll – матриці, де зафіксовано моменти початку, тривалість і висоту нот для кожного інструментального треку. Далі всі згенеровані такти з'єднуються у межах кожного треку, утворюючи безперервну музичну лінію. Результатом цього етапу є формування повної багатотрекової партитури – повноцінної музичної структури, готової до збереження чи прослуховування.

Четвертий блок – **збереження та виведення результату** – виконує фінальні операції, що перетворюють згенеровану структуру у форму, зручну для користувача. Спершу всі треки об'єднуються в єдиний фінальний результат. Потім відбувається конвертація у формат MIDI – універсальний цифровий формат для збереження символічної музики, сумісний із багатьма музичними редакторами та відтворювачами. Сформований файл можна зберегти на фізичному носії, або вивести безпосередньо для подальшого використання, наприклад, у музиці, навчанні або навіть у професійній творчості.

Експериментальна частина

Щоб дослідити ефективність методу генерації багатотрекових символічних композицій, було проведено ряд експериментів. Перший експеримент присвячений перевірці чіткості вихідної музики у вигляді piano-roll на різних етапах навчання. Метою експериментального дослідження було проаналізувати динаміку навчання генеративної моделі шляхом фіксації її результатів на різних етапах тренування. Для цього було обрано п'ять контрольних точок: 0, 700, 2500, 6000 та 7900 ітерацій. На кожному з цих етапів модель зупинялась, і на основі поточного стану ваг здійснювала повну генерацію нового музичного фрагмента з нуля. Згенеровані результати зберігались у форматі piano-roll – візуальному представленні музики, де вісь X відповідає часовій шкалі, а вісь Y – висоті нот.

а) Перший етап (0 кроків): результат генерації мав випадковий і хаотичний характер: відсутні ритмічні та гармонійні зв'язки, що є очікуваним наслідком необроблених випадкових параметрів.

б) Другий етап (700 кроків): композиція залишається малоструктурованою, але спостерігається формування первинних ознак впорядкованості.

в) Третій етап (2500 кроків): модель демонструє значний прогрес. Незважаючи на це, у

результати все ще присутні окремі недоліки.

г) Четвертий етап (6000 кроків): генерація суттєво наближається до цілісної музичної композиції.

д) П'ятий етап (7900 кроків): найвищий рівень якості генерації.

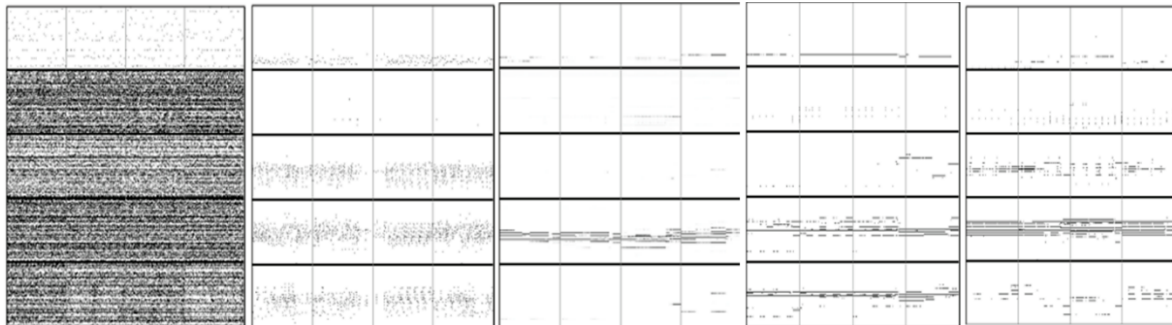


Рис. 3. Зафіксовані дані на першому етапі при а)0, б)700, в)2500 кроків, г)6000 та д)7900 кроків

Отже, на початку експерименту модель не може сформувати повноцінний структурований трек, але чим більше ітерацій впроваджено, тим краще стає трек.

У межах дослідження було проведено порівняльний експеримент з метою оцінити якість генерації музики різними архітектурними варіантами моделей. Основним завданням цього етапу було визначення здатності моделей відтворювати ключові характеристики музичної логіки: структурованість, ритмічну цілісність, гармонійну узгодженість та відповідність окремих треків загальній композиційній структурі. Для оцінки результатів були застосовані як внутрішньотрекові (intra-track), так і міжтрекові (inter-track) метрики.

Таблиця 1

Результати експерименту

		Порожні такти (EB; %)					Використані класи висоти тону (UPC)				Кваліфіковані ноти (QN; %)				DP (%)
		B	D	G	P	S	B	G	P	S	B	G	P	S	
Навчальні дані		8.06	8.06	19.4	24.8	10.1	1.71	3.08	3.28	3.38	90.0	81.9	88.4	89.6	88.6
З нуля	глушіння	6.59	2.33	18.3	22.6	6.10	1.53	3.69	4.13	4.09	71.5	56.6	62.2	63.1	93.2
	компонований	0.01	28.9	1.34	0.02	0.01	2.51	4.20	4.89	5.19	49.5	47.4	49.9	52.5	75.3
	гібрид	2.14	29.7	11.7	17.8	6.04	2.35	4.76	5.45	5.24	44.6	43.2	45.5	52.0	71.3
	аблений	92.4	100	12.5	0.68	0.00	1.00	2.88	2.32	4.72	0.00	22.8	31.1	26.2	0.0
Умовний трек	глушіння	4.60	3.47	13.3	-	3.44	2.05	3.79	-	4.23	73.9	58.8	-	62.3	91.6
	компонований	0.65	20.7	1.97	-	1.49	2.51	4.57	-	5.10	53.5	48.4	-	59.0	84.5
	гібрид	2.09	4.53	10.3	-	4.05	2.86	4.43	-	4.32	43.3	55.6	-	67.1	71.8

Інтра-трекова оцінка охоплювала такі показники: частка порожніх тактів (Empty Bars), кількість використовуваних класів висот нот (Used Pitch Classes), частка кваліфікованих нот (Qualified Notes), що свідчать про фрагментованість звукових подій, та відповідність ритмічним шаблонам ударних (Drum Patterns). Для оцінки гармонійної взаємодії між треками застосовувався показник Tonal Distance, який відображає ступінь гармонійної узгодженості між окремими інструментальними партіями.

У Таблиці 1 наведено результати для двох основних режимів генерації: повністю автономної генерації з нуля (from scratch) та умовної генерації на основі заздалегідь визначеного треку (track-conditional). Для кожного з цих режимів розглядалися три архітектури, а саме Jamming (кожен трек створюється незалежним генератором), Composer (один загальний генератор відповідає за формування всіх треків одночасно) і Hybrid (поєднує підходи, використовуючи як спільні, так і приватні латентні простори). Також було протестовано обмежену версію моделі composer без нормалізації пакетів (аблятивна модель), яка виступає як умовна нижня межа якості генерації.

Результати показали, що jamming-модель генерує активні треки з невеликою кількістю порожніх тактів, що свідчить про ефективне ритмічне наповнення. Однак у composer- та hybrid-моделях частка порожніх тактів, особливо для ударних та фортепіано, є суттєво вищою (20–30%), що може вказувати на менш стійку ритмічну структуру. Ablated модель майже повністю втрачає здатність до генерації, результати є переважно порожні.

Аналіз Used Pitch Classes показав, що jamming-модель відтворює мінімальну кількість висотних класів для басу, що узгоджується з його функцією як монофонічного інструмента. Інші треки демонструють ширший діапазон, але в composer- та hybrid-моделях він вищий, ніж у навчальних даних, що може свідчити про перенасичення висотними класами та менш контрольовану гармонійну логіку.

Показник Qualified Notes виявився чутливим до структури вихідних тензорів: значення у навчальному наборі перевищували 80–90%, тоді як згенеровані дані мали значно нижчі результати, особливо в composer- та hybrid-моделях. Це вказує на наявність коротких, нефункціональних нот, що можуть виникати внаслідок особливостей бінаризації.

Для метрики Drum Patterns jamming-модель показала високу відповідність (понад 90%) типовим ритмічним шаблонам, тоді як composer- та hybrid-виконання дещо поступались. У ablated-моделі ця метрика дорівнювала нулю, що підтверджує відсутність осмисленої генерації ударної партії.

У режимі умовної генерації (track-conditional) результати в цілому покращуються. Зокрема, jamming-модель демонструє високу ритмічну активність і зменшену кількість порожніх тактів. Hybrid-і composer-архітектури зберігають середні показники, при цьому hybrid досягає порівнянного або навіть кращого рівня якості порівняно з composer, зберігаючи більшу гнучкість.

Оцінка міждрекової гармонії за допомогою показника Tonal Distance показала, що composer- і hybrid-моделі мають найнижчі значення, що свідчить про високу узгодженість між партіями. Jamming-модель, яка не моделює міждрекову взаємодію, демонструє значення, ближчі до навчальних, але й менш гармонійно цілісні результати.

Загалом, результати експерименту свідчать про те, що hybrid-модель демонструє найбільш збалансовану продуктивність – вона поєднує гнучкість архітектури з досить високими показниками як на внутрішньотрековому, так і міждрековому рівнях. Composer-модель краще справляється з гармонійною координацією треків, однак втрачає якість у ритмічному наповненні. Jamming-модель, у свою чергу, ефективна у генерації окремих партій, але менш придатна до формування узгоджених ансамблевих структур. Ablated-модель виступає як приклад неефективного архітектурного рішення, що підтверджує важливість повноцінного дизайну моделі та застосування нормалізації у процесі навчання.

Висновки

У дослідженні було реалізовано метод генерації багатотрекової символічної музики на основі штучного інтелекту. Запропонована архітектура забезпечує формування цілісних музичних фраз із збереженням ритміки, гармонії та міждрекової взаємодії. Аналіз проміжних результатів показав поступове зростання якості генерації: від хаотичних звуків до впорядкованих структур, що свідчить про ефективність навчання. Порівняння трьох архітектур виявило переваги hybrid-моделі як найбільш збалансованої за всіма метриками. Composer-модель забезпечує кращу гармонійну узгодженість, а jamming – вищу якість окремих треків, проте з гіршою інтеграцією між ними. Спрощена ablated-модель підтвердила критичну важливість повноцінної архітектури.

Отримані результати демонструють перспективність використання генеративних моделей для створення музики та закладають основу для подальшого розвитку у напрямку покращення виразності та музичної форми.

Література

1. Tan X. A Tutorial on AI Music Composition [Electronic resource] / Xu Tan, Xiaobing Li // MM '21: ACM Multimedia Conference, Virtual Event China. – New York, NY, USA, 2021. doi: <https://doi.org/10.1145/3474085.3478875>
2. Roads C. Research in music and artificial intelligence [Electronic resource] / Curtis Roads // ACM Computing Surveys. – 1985. – Vol. 17, no. 2. – P. 163–190. doi: <https://doi.org/10.1145/4468.4469>
3. Shreya Singh. Music composition with AI [Electronic resource] / Shreya Singh, Sachin Ramling Jadhav // World Journal of Advanced Research and Reviews. – 2025. – Vol. 25, no. 3. – P. 1031–1037. doi: <https://doi.org/10.30574/wjarr.2025.25.3.0723>
4. Zulić H. How AI can Change/Improve/Influence Music Composition, Performance and Education: Three Case Studies [Electronic resource] / Harun Zulić // INSAM Journal of Contemporary Music, Art and Technology. – 2019. – No. 2. – P. 100–114. doi: <https://doi.org/10.51191/issn.2637-1898.2019.2.2.100>
5. Human, I wrote a song for you: An experiment testing the influence of machines' attributes on the AI-composed music evaluation [Electronic resource] / Joo-Wha Hong [et al.] // Computers in Human Behavior. – 2022. – Vol. 131. – P. 107239. doi: <https://doi.org/10.1016/j.chb.2022.107239>
6. Dash A. AI-Based Affective Music Generation Systems: A Review of Methods and Challenges [Electronic resource] / Adyasha Dash, Kathleen Agres // ACM Computing Surveys. – 2024. doi: <https://doi.org/10.1145/3672554>
7. Kaliakatsos-Papakostas M. Artificial intelligence methods for music generation: a review and future perspectives [Electronic resource] / Maximos Kaliakatsos-Papakostas, Andreas Floros, Michael N. Vrahatis // Nature-Inspired Computation and Swarm Intelligence. – [S. l.], 2020. – P. 217–245. doi: <https://doi.org/10.1016/b978-0-12-819714-1.00024-5>
8. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models [Electronic resource] / Ryan Louie [et al.] // CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu HI USA. – New York, NY, USA, 2020. doi: <https://doi.org/10.1145/3313831.3376739>