

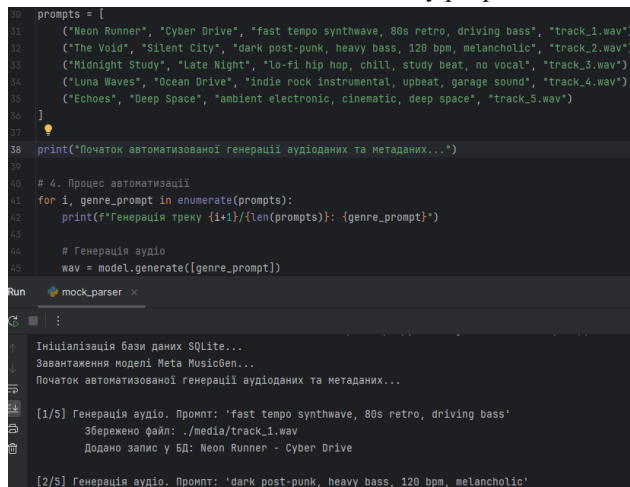
УДК 004.7

*Васянович О.А., здобувач,
Морозов А.В., к.т.н., доцент
Державний університет «Житомирська політехніка»*

АВТОМАТИЗОВАНА ГЕНЕРАЦІЯ МОСК-ДАНИХ У РОЗРОБЦІ СТРІМІНГОВИХ СЕРВІСІВ

Розробка та навантажувальне тестування сучасних стрімінгових сервісів вимагає наявності великих масивів медіаданих [1]. Для перевірки роботи кешування, швидкості віддачі контенту та коректності роботи баз даних під навантаженням необхідні тисячі унікальних аудіофайлів з валідними метаданими. Використання захищених авторським правом композицій створює юридичні ризики, а відкриті бібліотеки не забезпечують достатньої варіативності та стандартизації тегів для тестування систем рекомендацій.

Практичне вирішення цієї проблеми реалізовано під час розробки інформаційної системи організації прослуховування аудіоконтенту (проект «IndieStream»). Ефективним інженерним підходом для автоматизованого наповнення тестового середовища (mock-даних) обрано використання відкритої генеративної нейромережі Meta MusicGen [2]. На відміну від комерційних веб-сервісів, ця модель дозволяє розгорнути процес локально та керувати генерацією програмно, без ліцензійних обмежень для етапу розробки.



```
30 prompts = [  
31     ("Neon Runner", "Cyber Drive", "fast tempo synthwave, 80s retro, driving bass", "track_1.wav"),  
32     ("The Void", "Silent City", "dark post-punk, heavy bass, 120 bpm, melancholic", "track_2.wav"),  
33     ("Midnight Study", "Late Night", "lo-fi hip hop, chill, study beat, no vocal", "track_3.wav"),  
34     ("Luna Waves", "Ocean Drive", "indie rock instrumental, upbeat, garage sound", "track_4.wav"),  
35     ("Echoes", "Deep Space", "ambient electronic, cinematic, deep space", "track_5.wav")  
36 ]  
37  
38 print("Початок автоматизованої генерації аудіоданих та метаданих...")  
39  
40 # 4. Процес автоматизації  
41 for i, genre_prompt in enumerate(prompts):  
42     print(f"Генерація треку {i+1}/len(prompts): {genre_prompt}")  
43  
44     # Генерація аудіо  
45     wav = model.generate(genre_prompt)  
46  
Run mock_parser x  
:  
:  
Ініціалізація бази даних SQLite...  
Завантаження моделі Meta MusicGen...  
Початок автоматизованої генерації аудіоданих та метаданих...  
[1/5] Генерація аудіо. Промпт: 'fast tempo synthwave, 80s retro, driving bass'  
Збережено файл: ./media/track_1.wav  
Додано запис у БД: Neon Runner - Cyber Drive  
[2/5] Генерація аудіо. Промпт: 'dark post-punk, heavy bass, 120 bpm, melancholic'
```

Рис. 1. Скрипт та процес генерації аудіоданих у терміналі.

Процес створення та інтеграції синтетичних даних у платформу IndieStream повністю автоматизовано. Розроблений Python-скрипт використовує бібліотеку Audioscraft для ініціалізації моделі MusicGen. Алгоритм формує масив вхідних параметрів (жанр, темп, настрої), які слугують текстовими запитами для нейромережі. Під час генерації аудіофайлу формату WAV скрипт паралельно створює для нього фіктивні текстові метадані (назву треку та ім'я неіснуючого виконавця) за допомогою бібліотеки Faker [3].

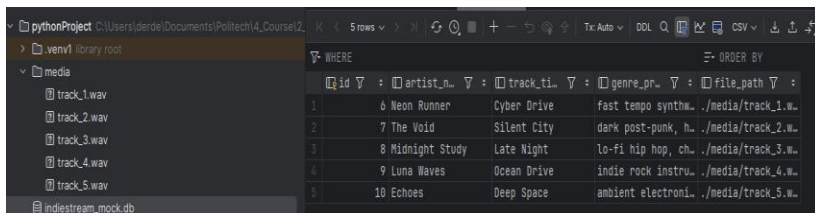


Рис. 2. Згенеровані аудіофайли та відповідні їм записи у реляційній базі даних.

Створені аудіофайли зберігаються у локальну директорію сховища, а скрипт автоматично формує SQL-запити для запису фізичних шляхів до файлів та їхніх згенерованих характеристик у реляційну базу даних проекту. Застосування відкритих моделей штучного інтелекту дозволяє вибудувати повністю керований процес створення синтетичних медіаданих. Такий підхід забезпечує розробника необхідним обсягом стандартизованих даних для стрес-тестування бекенду та налаштування рекомендаційних алгоритмів платформи, мінімізуючи витрати часу на пошук та підготовку тестової бази.

Список використаних джерел:

1. Соммервілл І. Інженерія програмного забезпечення : навч. посіб. пер. з англ. Київ : ВІЛЬЯМС, 2002. 680 с..
2. Офіційний репозиторій бібліотеки Audioscraft / Meta Research. GitHub, 2023. URL: <https://github.com/facebookresearch/audioscraft> (дата звернення: 22.03.2026).
3. Офіційна документація бібліотеки Faker для генерації тестових даних. Read the Docs. 2024. URL: <https://faker.readthedocs.io/> (дата звернення: 22.03.2026).