

ВИКОРИСТАННЯ  
МАШИННОГО НАВЧАННЯ  
ДЛЯ СИНТЕЗУ ЗВУКУ НА  
ОСНОВІ ТЕКСТОВИХ  
ОПИСІ В

Виконав: студент 4 р. н. (КН)  
Письменний Антон

Науковий керівник: Медвідь С. О.



# Вступ

- **Успішна інтеграція Ш** в інструменти створення інших медіа матеріалів та програмного коду
- **Відсутність зручних Ш-інструментів** для відкриття нових можливостей перед недосвідченим музикантом



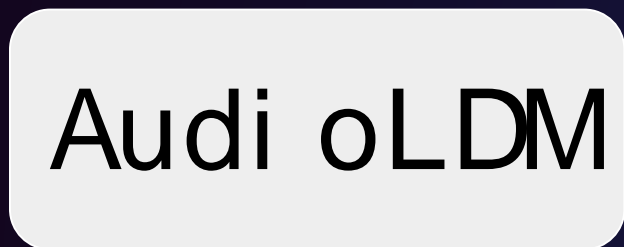
Маємо дослідити перспективи використання Ш для синтезу та звуку в цифрових звукових робочих станціях (ЦЗРС)

# Постановка задачі

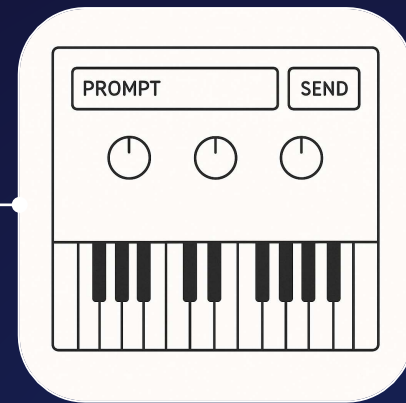
- Дослідити наявні рішення інтеграції Ш в інструментарій музиканта
- Проаналізувати наявні технології й підходи
- Створити демонстраційні програмні застосунки

## Огляд наявних рішень

- Обмежена якість сигналу
- Відсутність інтеграції в ЦЗРС
- Налаштування моделей Ш приховані



# Час еволюці онувати



# Два підходи

## True text-to-sound

- Універсальність
- Низька якість зразків

## LLM + synth

- Певні обмеження діапазону можливих звуків
- Висока якість сигналу

# Взаємодія плагінів з моделями Ш

\* REST-API all the way, адже:

- потрібно багато ресурсів
- локальне розміщення != портативність
- часові витрати мережеві на запити - мізерні в порівнянні



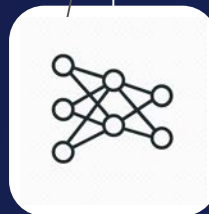
\*



\*



\*



# Використані технології

- JUCE
- Fast API
- Gemini
- AudioLDM 2
- HuggingFace Spaces
- Gradio



Результат простіше  
продемонструвати :)

- Current project
- Recent files
- Plugin database
  - Effects
  - Generators
    - Installed
      - Effects
        - Fruity
        - VST
        - VST3
      - Generators
        - CLAP
        - Fruity
        - VST
        - VST3
        - OTT
        - tapSynth
        - Vital
        - Wingman2
  - Plugin presets
  - Channel presets
  - Mixer presets
  - Scores
  - Backup
  - Clipboard files
  - Demo projects
  - Envelopes
  - Impulses
  - Misc
  - My projects
  - Packs
  - Project bones
  - Recorded
  - Rendered
  - Sliced audio
  - Soundfonts
  - Sounds

The screenshot shows the FL Studio Playlist - Arrangement view. The interface is dark-themed. On the left, there is a track list with the following items: Wingman2 (blue), TapSynth (red), Track 3, Track 4, Track 5, Track 6, Track 7, Track 8, Track 9, Track 10, Track 11, Track 12, Track 13, Track 14, Track 15, and Track 16. The main area is a grid with 16 tracks and 28 time steps. The 'Kick' track (blue) has a single blue block in the first time step. The 'bass' track (red) has a red block in the first time step. The 'Wingman2' track (blue) has a blue block in the first time step. The rest of the tracks are empty. A mouse cursor is visible over Track 5, time step 13.

# Можливі покращення

- Уніфікація інтеграції з LLM
- Дотренування/заміна AudioLDM 2 і Gemini 1.5 Flash
- Реалізація навчального потенціалу моделей Ш

# Результати роботи

- Проаналізовано існуючі засоби Ш для синтезу звуку.
- Досліджено сучасні інструменти, що можна використати для інтеграції Ш в ЦРС.
- Розроблено два програмних модулі для синтезу звуку з використанням Ш на базі текстового опису.

Дякуємо за увагу

<https://github.com/BIVrp/diploma-2025>