МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ВОЛИНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ЛЕСІ УКРАЇНКИ
Кафедра комп'ютерних наук та кібербезпеки

На правах рукопису

ГОЛОЩУК ОЛЕКСАНДР ОЛЕКСАНДРОВИЧ

МЕТОДОЛОГІЯ СТВОРЕННЯ ЦИФРОВИХ МУЗИЧНИХ ПРОДУКТІВ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ

Спеціальність «122 Комп'ютерні науки»
Освітньо-професійна програма «Комп'ютерні науки та інформаційні технології»
Робота на здобуття освітнього ступеня «Магістр»

Науковий керівник:
МАМЧИЧ ТЕТЯНА ІВАНІВНА
кандидат фізико-математичних наук, доцент
кафедри комп'ютерних наук та кібербезпеки

РЕКОМЕНДОВАНО ДО ЗАХИСТУ
Протокол № _____
засідання кафедри комп'ютерних наук та кібербезпеки
від _____ 2025 р.

Завідувач кафедри
_____ Гришанович Т. О.

Луцьк – 2025

ЗМІСТ

ВСТУП

Актуальність теми. Недавня поява та стрімкий розвиток генеративних моделей штучного інтелекту стали каталізатором нової фази цифрової трансформації, яка не просто торкнулася креативних індустрій, а й докорінно змінює їхні фундаментальні процеси на рівні архітектур і методологій. Створення музичного контенту, зокрема аранжування, історично було класичною евристичною задачею – складним когнітивним процесом, що покладається на досвід, інтуїцію, знання теорії музики та неформалізовані правила, якими керується людина-фахівець. Традиційне програмне забезпечення, що використовується у цій галузі, – таке як нотні редактори (Finale) та цифрові аудіостанції (DAW) на кшталт Cubase – по суті є зрілими та складними інтегрованими середовищами розробки (IDE) для музики. Їхня архітектура та логіка інтерфейсу (UI/UX) десятиліттями будувалися навколо мети моделювання та підтримки саме людської евристики, відтворюючи аналоговий студійний робочий процес у цифровому середовищі.

Водночас, останні досягнення в галузі машинного навчання, зокрема в архітектурах генеративних моделей (таких як Генеративно-змагальні мережі (GAN), Варіаційні автокодувальники (VAE) та Трансформери), спричинили появу нового, фундаментально відмінного алгоритмічного підходу до вирішення творчих завдань. Ці моделі, навчені на величезних масивах даних (Big Data) музичних композицій, здатні виявляти глибинні статистичні патерни та генерувати нові, часто нетривіальні, музичні структури. Системи штучного інтелекту (AIVA, Soundraw, Suno AI) сьогодні здатні не просто асистувати у рутинних операціях, а й виступати повноцінними «співавторами», генеруючи цілі композиції.

Це створює унікальний виклик та водночас можливість для **інформаційних технологій**. Ми спостерігаємо «технологічний розрив» або навіть конфлікт парадигм між двома різними типами систем/

1. **Традиційні системи (DAW).** Архітектурно потужні, гнучкі, але повністю залежні від ручного вводу та евристики користувача (когнітивно-керований підхід).

2. **Генеративні системи (ШІ).** Швидкі, здатні до нетривіальних рішень, але часто функціонують як «чорна скринька» з обмеженим контролем, проблемами інтеграції, сумісності даних (data interoperability) та конфліктом робочих процесів (workflow conflict).

Таким чином, актуальність даної роботи лежить у площині інженерії програмного забезпечення та проєктування інформаційних систем. Вона полягає у доцільності провести детальний порівняльний аналіз між **«людським ресурсом»** (евристичними рішеннями людини, підтриманими архітектурою DAW) та **«ресурсом ШІ»** (алгоритмічними рішеннями, керованими даними). Цей аналіз є не просто академічним інтересом, а важливою основою для обґрунтування, проєктування та розробки оптимізованого гібридного робочого процесу (**workflow**), який би системно інтегрував сильні сторони обох парадигм.

Ключова проблема полягає у відсутності формалізованих методологій, описаних моделей даних та протоколів взаємодії для ефективної, системної інтеграції генеративних ШІ-сервісів у традиційні, IT-орієнтовані процеси музичного виробництва, що базуються на DAW. Через новизну технології та складність інтеграції двох різних філософій обробки даних, існуючі практики є переважно несистемними (ad-hoc). Вони, як правило, зводяться до «ручного перетягування» згенерованого ШІ матеріалу в DAW без єдиного протоколу, що породжує проблеми із синхронізацією, керуванням версіями та відтворюваністю результату.

Критичною проблемою інтеграції ШІ є відсутність формалізованого інструментарію для об’єктивного порівняння результатів алгоритмічної та людської творчості. Актуальність дослідження визначається необхідністю розробки чітких інженерних метрик (часові витрати, обчислювальна складність, гнучкість редагування) для наукового обґрунтування ефективної гібридної моделі.

У роботі застосовано набір цільових інженерних промптів для вирішення конкретних виробничих задач, що дозволило мінімізувати стохастичність генерації. Процес створення цифрового продукту структуровано за трьома взаємопов'язаними складовими: **семантико-структурною** (формування детермінованого каркаса композиції засобами DAW), **генеративною** (аугментація проєкту специфічним контентом, таким як бек-вокал та текстури, засобами ШІ) та **інтеграційною** (технічна адаптація, синхронізація та спектральне маскування згенерованого матеріалу).

Обґрунтовано доцільність збагачення проєкту засобами штучного інтелекту, яка полягає не в заміні творчої функції людини, а в розширенні тембральної палітри та оптимізації виробничого циклу шляхом делегування нейромережам вузькоспеціалізованих задач зі створення складних текстурних шарів (бек-вокал, хорові гармонії, атмосферні ефекти). Це дозволяє досягти вищої щільності та комерційної якості звучання при суттєвій економії часових ресурсів.

Наукова новизна дослідження полягає у формальному описі та порівнянні інформаційних потоків евристичного та алгоритмічного підходів, а також у розробці та експериментальній валідації гібридної моделі робочого процесу (workflow), яка системно інтегрує можливості генеративного ШІ та архітектуру DAW. Таким чином:

- формально описано та порівняно евристичний (людський ресурс) та алгоритмічний (ресурс ШІ) підходи до музичного аранжування з точки зору інформаційних потоків (data flows), моделей даних та архітектурних патернів взаємодії «користувач-система» та «система-система»;
- запропоновано та експериментально валідовано гібридну модель робочого процесу (workflow), яка, на відміну від існуючих несистемних (ad-hoc) практик, базується на формалізованому описі процесу, чітких точках інтеграції та метриках оцінки, пропонуючи відтворюваний та масштабований підхід до інтеграції ШІ в DAW.

Мета і завдання дослідження. Мета проєкту – провести всебічний комплексний компаративний аналіз людського ресурсу (евристичні методи та когнітивні процеси людини в середовищі DAW) та ресурсу ШІ (алгоритмічні методи та обчислювальні моделі генеративних систем) у задачі створення музичного аранжування; на основі отриманих даних цього аналізу, розробити, формально описати та емпірично валідувати гібридну модель робочого процесу (workflow), яка ефективно, відтворювано та масштабовано інтегрує функціонал DAW (на прикладі Cubase) та сучасних генеративних ШІ-систем.

Для досягнення поставленої мети необхідно вирішити наступні завдання дослідження.

1. Провести системний аналіз наукової літератури та технічної документації щодо евристичних підходів до аранжування; декомпілювати «людський ресурс» у формалізовану послідовність операцій (наприклад, з використанням нотацій BPMN або UML Use Case) у середовищі Cubase.

2. Дослідити архітектуру та алгоритмічні принципи роботи «ресурсу ШІ» (обраних генеративних ШІ-сервісів, напр., AIVA, Soundraw, Suno AI); провести аналіз їхніх навчаючих датасетів, API (за наявності), підтримуваних форматів даних та обмежень щодо контролю користувача над процесом генерації.

3. Розробити **дизайн експерименту** та провести практичне моделювання, створивши три чіткі пайплайни (pipelines) обробки вхідного MIDI-матеріалу (з Finale), чітко документуючи кожен крок, використане ПЗ (в Cubace) та проміжні формати даних для забезпечення відтворюваності експерименту:

- *Модель 1: Чисто евристичний (Людина)* – аранжування виключно людиною в Cubase.
- *Модель 2: Чисто алгоритмічний (ШІ)* – генерації аранжування засобами ШІ.
- *Модель 3: Гібридний (ШІ + Людина)* – інтеграція обох підходів.

4. Провести детальний порівняльний аналіз трьох моделей за розробленою системою метрик, що включає як кількісні (витрачений час, обсяг даних), так і

якісні (гнучкість подальшого редагування, складність інтеграції, рівень необхідних навичок оператора) показники.

5. На основі результатів порівняння розробити, формально описати та обґрунтувати **оптимальний гібридний робочий процес** у вигляді методології або набору практичних рекомендацій для IT-фахівців у креативних індустріях.

Об'єкт дослідження – робочі процеси (workflows) як інформаційні системи, що забезпечують створення та обробку структурованих музичних даних (в форматах MIDI, audio, project files).

Предметом дослідження є моделювання евристичних процесів користувача в архітектурі цифрових аудіостанцій, архітектура та моделі даних генеративних систем штучного інтелекту, а також методи та принципи проєктування гібридних робочих процесів музичного аранжування.

Матеріал дослідження:

- моделювання евристичних процесів **людського ресурсу** в архітектурі та інтерфейсах цифрових аудіостанцій (Finale, Cubase);
- архітектура, обчислювальні моделі та моделі даних ресурсу ШІ (генеративних систем), що застосовуються до музичних послідовностей;
- порівняльні характеристики (часові, якісні, ресурсозатратні) та метрики оцінки обох підходів.

Практичне значення одержаних результатів.

1. **Для розробників ПЗ (IT-індустрії).** Результати дослідження та запропонована модель workflow можуть бути використані компаніями (напр., Steinberg, Avid, Ableton) для проєктування плагінів нового покоління (напр., VST/AU з вбудованим ШІ) або розробки відкритих протоколів обміну даними (API) між DAW та хмарними ШІ-сервісами для глибшої нативної інтеграції.

2. **Для IT-фахівців у медіа.** Запропонована гібридна методологія може бути безпосередньо впроваджена у виробничі процеси (напр., у геймдеві, кіно, рекламі) для оптимізації пайплайнів, значного скорочення часу (time-to-market) для створення аудіоконтенту та надання фахівцям інструментів для швидкого прототипування ідей при збереженні повного контролю над результатом.

3. **Для освітніх програм.** Робота може слугувати навчально-методичним матеріалом для оновлення навчальних планів за спеціальностями «Інформаційні технології», «Програмна інженерія» та «Мультимедійні технології», готуючи фахівців, що володіють затребуваними на ринку гібридними компетенціями на стику ІТ та креативних індустрій.

Апробація результатів та публікації.

1. Голощук О. О. Вплив музично-інформаційних технологій на розвиток сучасного музичного мистецтва. *Актуальні проблеми розвитку українського та зарубіжного мистецтв: культурологічний, мистецтвознавчий, педагогічний аспекти : матеріали VIII Міжнародної науково-практичної конференції.* Львів–Торунь : Liha-Pres, 2023. С. 352–356.
2. Голощук О. О. Значення сучасних музично-інформаційних технологій в процесі фахової мистецької освіти. *Сучасні дослідження світової науки : матеріали наук.-практ. конф.* 2022. С. 969–971.
3. Голощук О. О. Комп'ютерне аранжування та звукорежисура : методичні рекомендації до курсу. Луцьк : ВНУ ім. Лесі Українки, 2023. 48 с.
4. Голощук О. О. Методика роботи в музичному редакторі Steinberg Cubase : методичні рекомендації для студентів спеціальності 025 «Музичне мистецтво». Луцьк : ВНУ ім. Лесі Українки, 2022. 45 с.
5. Голощук О., Сорока Р., Шкоба В. Монтаж та обробка звуку в музичному редакторі AVID Technology Pro Tools 12 : методичні рекомендації. Луцьк : ВНУ ім. Лесі Українки, 2022. 21 с.
6. Голощук О. О. Розвиток музично-інформаційних технологій в умовах цифровізації освіти. *Актуальні проблеми розвитку природничих та гуманітарних наук : збірник матеріалів VI Міжнар. наук.-практ. конф*. Луцьк : ВНУ ім. Лесі Українки, 2022. С. 399–400.

РОЗДІЛ 1.
ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ АСПЕКТИ ВЗАЄМОДІЇ ЕВРИСТИЧНИХ ТА АЛГОРИТМІЧНИХ ПІДХОДІВ У СУЧАСНИХ МУЗИЧНО-ІНФОРМАЦІЙНИХ СИСТЕМАХ

1.1. Евристика проти генеративного алгоритму: постановка проблеми

Для цілей нашого дослідження, ми маємо чітко розмежувати два фундаментально різні підходи до прийняття рішень у процесі створення аранжування:

Евристичний підхід (Людина + DAW). Рішення приймає людина-оператор на основі неформалізованого досвіду, знань та інтуїції («евристики»). Програмне забезпечення (Finale, Cubase) надає інструменти, але не приймає самостійних творчих рішень[40, 112]. Його алгоритми є детермінованими – вони виконують чіткі інструкції користувача [20, 45].

Алгоритмічний підхід (ШІ). Рішення приймає генеративна модель (нейронна мережа). Вона прогнозує та генерує новий контент на основі ймовірнісних моделей, вивчених з великих масивів даних. Алгоритм є ймовірнісним та ендогенним (приймає рішення «зсередини»). Алгоритм є ймовірнісним та ендогенним (приймає рішення «зсередини»). На відміну від детермінованих систем, де вихідний результат є чіткою функцією вхідних команд, така модель діє самостійно обираючи оптимальний шлях вирішення творчої задачі без явних інструкцій користувача.

Цей розділ повністю присвячений формалізації та аналізу першого підходу. Ми декомпілюємо програми Finale та Cubase, щоб проаналізувати, які саме типи алгоритмів вони використовують, і як ці алгоритми підтримують *людську*, а не *машинну* евристику. Наша мета – не просто описати інтерфейс користувача, а проаналізувати архітектуру цих систем на рівні інформаційних процесів, щоб виявити, які саме класи алгоритмів (детерміновані, правило-орієнтовані, DSP) лежать в їх основі.

1.2. Finale: алгоритми валідації та структурування даних

Finale, як початкова точка процесу «еталонної моделі», не є «творчою» програмою в сенсі ШІ. Це, по суті, структурований редактор даних (structured data editor) з вбудованими алгоритмами валідації [33].

Типи алгоритмів у Finale. Правило-орієнтовані алгоритми (Rule-Based Algorithms) це основа Finale, вони не генерують нічого нового, а забезпечують дотримання правил музичної нотації.Алгоритм валідації такту. Коли користувач вводить ноти в такт розміром 4/4, алгоритм у реальному часі підсумовує тривалості. Якщо сума перевищує 4 долі, система видає помилку або автоматично переносить ноту.

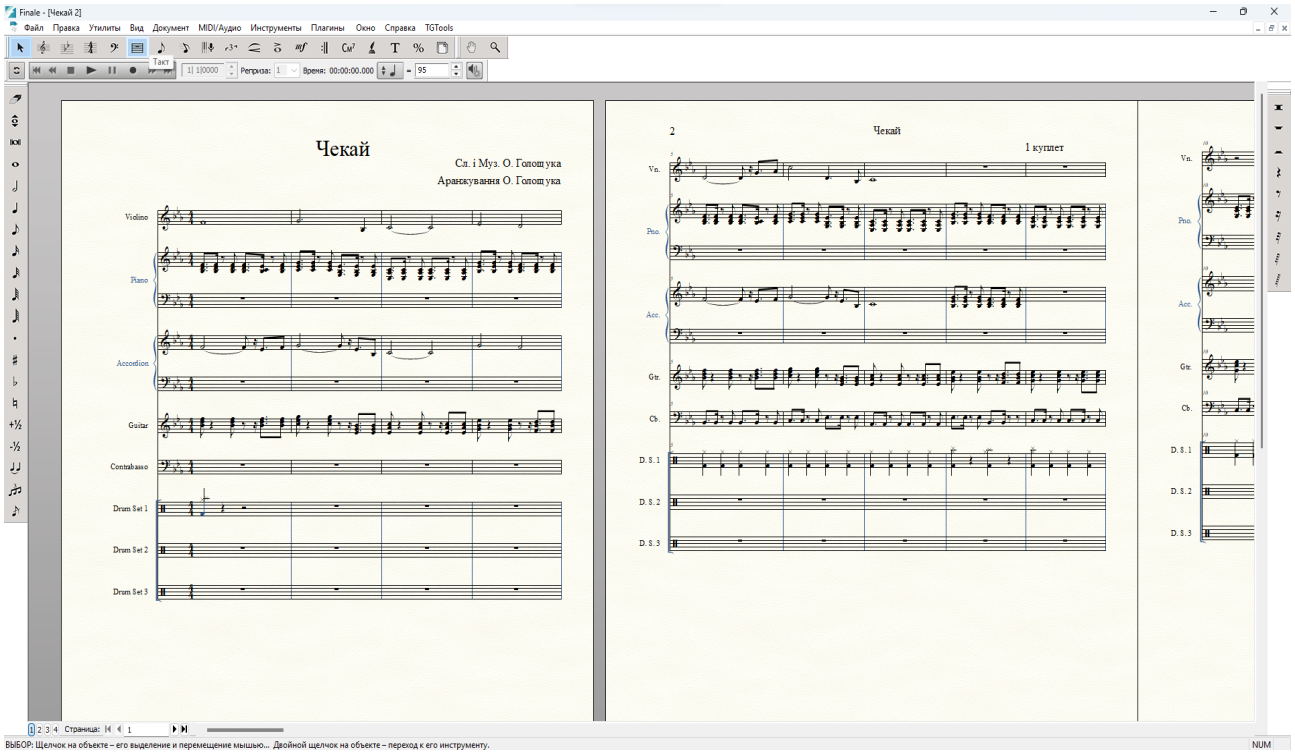


Рисунок 1.1 – Finale

В порівнянні з алгоритмами ШІ, навпаки, не знав би правил, а навчився б на мільйонах прикладів, що «зазвичай» у такті 4/4 сума нот дорівнює 4 долям. Алгоритм Finale є детермінованим та жорстко кодованим [2, 98].

Алгоритми трансформації даних (Data Transformation Algorithms). Експорт у MIDI це є ключовий процес для нашого дослідження. Finale використовує алгоритм трансформації, щоб перетворити свій складний пропрієтарний .musx-формат (з графікою, текстом, розміткою) у стандартизований, подієвий формат MIDI [33; 6, 12].

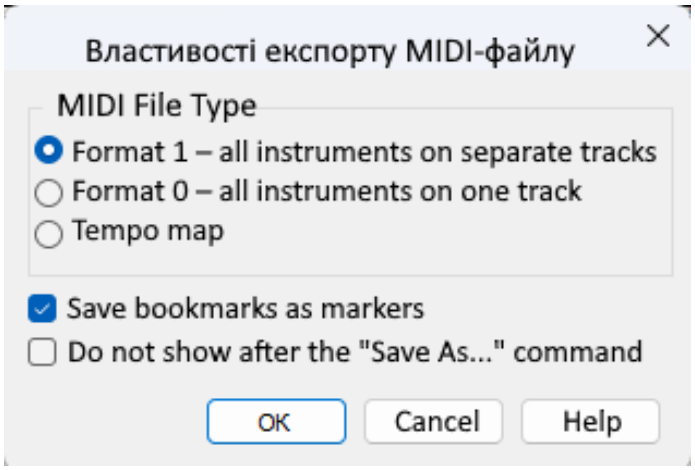


Рисунок 1.2 – Властивості експорту MIDI-файлу

Цей алгоритм у порівнянні із ШІ є зворотним та детермінованим. Він не додає жодної «творчості», а лише переводить дані з одного формату в інший зі 100% точністю.

Finale надає нам «Ground Truth» – ідеально чистий, структурований, валідований людиною вхідний MIDI-файл. Однак, щоб визначити *чим* грати, безпосередньо у середовищі Finale ми інтегруємо VST-інструмент Garritan ARIA Player. Це високопродуктивний семплерний рушій (sample playback engine), який є стандартом для нотних редакторів. Він вирізняється своєю ефективністю, низьким споживанням ресурсів та, що найважливіше, широкою палітрою реалістичних оркестрових та акустичних інструментів (Garritan Personal Orchestra). Завдяки глибокій інтеграції з нотним станом, рушій автоматично інтерпретує символи артикуляції та динаміки, перетворюючи їх на відповідні MIDI-команди (Control Change messages), що гарантує точне відтворення нюансів виконання без додаткового програмування.

Рисунок 1.3 – VSTi Garritan Personal Orchestra

За допомогою цього плагіна ми можемо призначити (прикріпити) конкретні звуки. Вибрати специфічні патчі (наприклад, *Solo Violin* або *Grand Piano*), а також визначити тембральні характеристики**,** налаштувати артикуляцію та експресію ще на етапі партитури. Тільки після цього етапу – коли ми маємо ідеальну нотну структуру (MIDI) та чітко визначений тембральний намір (через Garritan), додавання кроку з Garritan ARIA Player перетворює наш «Ground Truth» з суто математичного (ноти/тривалість) на художній. Це наш еталонний вхідний вектор, який ми будемо подавати на вхід двом різним системам. Системі Cubase (де керувати будуть детерміновані алгоритми та людська евристика). Це дає звукорежисеру або системі чітку інструкцію (reference), який саме інструмент мав на увазі композитор. Та системі ШІ (де керувати будуть генеративні ймовірнісні алгоритми). Це дозволяє перевірити, наскільки добре модель розуміє не тільки ноти, але й контекст інструментування (чи зможе ШІ відтворити ту ж емоцію, що й Garritan, або кращу?). Таке безпосереднє зіставлення результатів дасть змогу об'єктивно оцінити, чи здатен алгоритм переступити поріг технічної імітації та наблизитися до рівня художньої

виразності, який забезпечують професійні семплерні бібліотеки під керуванням людини.

1.3. Cubase: архітектура детермінованих алгоритмів обробки

Якщо Finale – це редактор *правил*, то Cubase – це середовище для маніпуляції даними та їх обробки в реальному часі. Алгоритми Cubase є набагато складнішими, але вони так само детерміновані і керовані користувачем. [43; 7, 8]

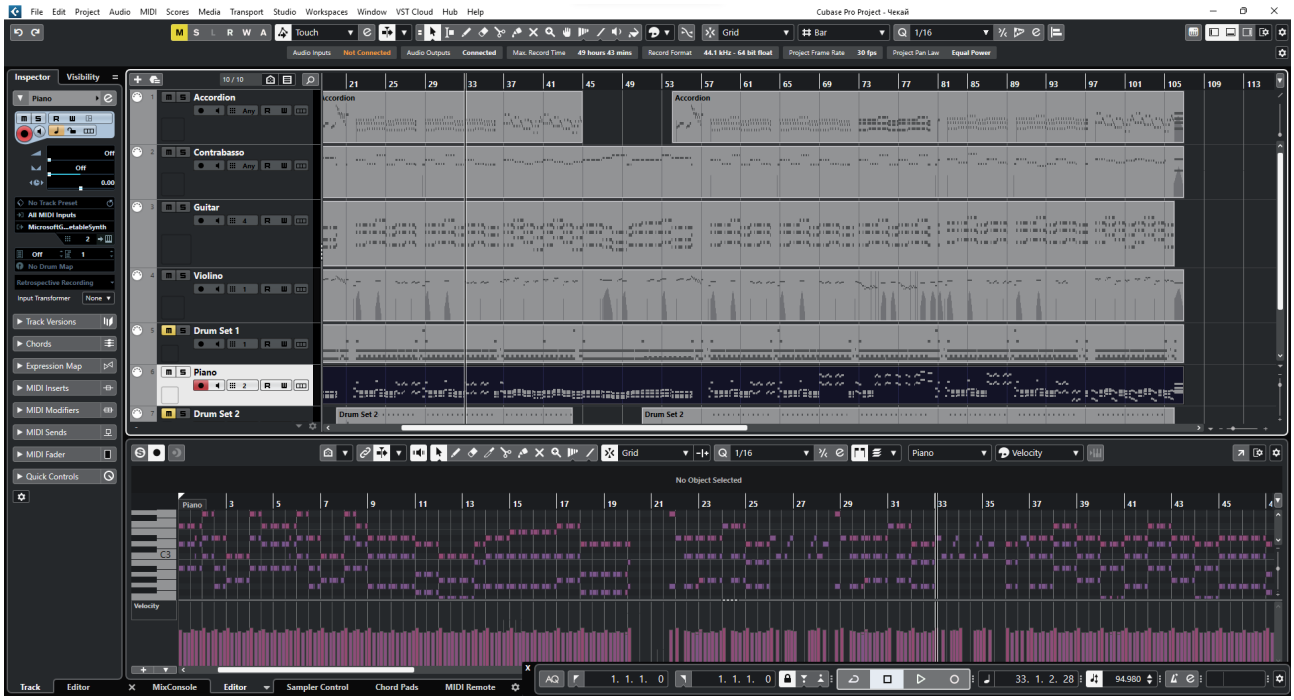


Рисунок 1.4 – Steinberg-Cubase

Серед типів алгоритмів у Cubase ключову роль відіграють алгоритми секвенсора (MIDI Event Handling), ядром яких є подієвий секвенсор із «головним циклом» (main loop). Він із високою точністю, що вимірюється в «тіках», зчитує потік MIDI-подій та аудіо-семплів, відправляючи їх на відповідні виходи (VSTi, аудіо-карту) у суворо визначений момент часу. Цей процес є повністю детермінованим, адже система ніколи не відхиляється від заданого часу «за настроєм», на відміну від алгоритмів ШІ, які при імітації «людської гри» навмисно вносять імовірнісні мікро-зміщення для створення ефекту «груву». У

детермінованому середовищі DAW будь-які подібні ритмічні флуктуації є результатом виключно свідомого застосування оператором функцій квантизації (Swing, Humanize), що дозволяє математично точно дозувати ступінь відхилення від сітки.

Наступним класом є алгоритми цифрової обробки сигналів (DSP Algorithms), до яких належать VST-ефекти, такі як реверберація, еквалізація чи компресія. Кожен такий плагін являє собою набір складних математичних алгоритмів, наприклад, швидкого перетворення Фур'є (FFT) для еквалайзера або алгоритми згортки (Convolution) для реверберації. [50, 45; 14, 110].

Ці алгоритми є повністю детермінованими, адже на той самий вхідний аудіо-сигнал і ті самі налаштування «ручок» плагіна результат завжди буде ідентичним; вони трансформують сигнал за чітким математичним законом, але не генерують новий музичний контент. Окремо слід виділити алгоритми автоматизації (Automation Handling): коли ви «малюєте» криву гучності, Cubase використовує алгоритм лінійної інтерполяції або іншого типу сплайнів для плавного розрахунку значень гучності між контрольними точками. Це також є детермінованим математичним розрахунком, на відміну від штучного інтелекту, якому дали б задачу «зробити плавний фейд-аут» і який міг би згенерувати тисячі різних варіантів кривої, навчившись на прикладах, яку форму найчастіше використовують у поп-музиці. [15, 180].

1.4. Декомпозиція «еталонної моделі»: аналіз взаємодії «евристика-алгоритм»

Тепер ми можемо описати наш проєкт у Cubase не просто як «що ви зробили», а як послідовність взаємодій людської евристики з детермінованими алгоритмами ПЗ. Вхідними даними для цього процесу слугує файл Чекай.mid, який фактично є вектором подієвих даних, згенерований у редакторі Finale. Аналіз взаємодії у системі «Людина-Машина» розпочинається з прийняття людиною евристичного рішення про те, що треку потрібна ритмічна основа в

стилі Євроденс-трек». Цей творчий задум трансформується у конкретний ланцюжок дій оператора в програмному середовищі, який включає створення MIDI-доріжки, завантаження VSTi Ueberschall, ручне програмування MIDI-подій у Key Editor.

Рисунок 1.5 – VSTi Ueberschall

Алгоритм секвенсора Cubase детерміновано реагує на введені дані, відтворюючи MIDI-події та посилаючи їх на алгоритм синтезу звуку у VSTi, внаслідок чого результат стає прямим і повністю прогнозованим. Згодом, коли оператор приймає евристичне рішення про те, що мелодія звучить занадто «сухо» і потребує додаткового простору, він виконує ряд дій: створює Send-доріжку, завантажує VST-ефект реверберації Avenger та налаштовує параметри Delay Time і Mix. У відповідь на це аудіо-сигнал спрямовується на DSP-алгоритм ревербератора, який детерміновано обчислює тисячі віддзеркалень згідно з закладеною математичною моделлю і додає їх до основного сигналу.

Таким чином, програмне середовище виступає не як співавтор, а як високоточний виконавчий інструмент, що трансформує абстрактні естетичні запити у жорстку послідовність математичних операцій. Це забезпечує повну відтворюваність результату, де кожна зміна акустичної картини є прямим наслідком свідомого інженерного втручання оператора, а не алгоритмічної випадковості.

Рисунок 1.6 – VST Avenger

На етапі роботи над динамікою приспіву, оператор виконує ручне програмування кривих автоматизації параметрів гучності та панорами. У відповідь на ці дії алгоритм інтерполяції середовища Cubase виконує детермінований розрахунок проміжних значень, забезпечуючи плавну зміну заданих акустичних характеристик сигналу під час відтворення та суворо дотримуючись траєкторії, заданої людиною.

Однак, для глибшого розуміння архітектурних відмінностей між евристичним підходом та генеративним ШІ, необхідно проаналізувати роботу зі складнішими, неструктурованими типами даних, зокрема з аудіо-лупами (Audio

Loops). У цьому контексті програмне забезпечення виконує вже не просто лінійну інтерполяцію значень, а складну цифрову обробку сигналів (DSP) у реальному часі для адаптації аудіо-матеріалу до часової сітки проєкту без втрати якості.

Переходячи до практичної декомпозиції нашого еталонного проєкту, розглянемо конкретну реалізацію цього процесу. Для створення ритмічної секції в моделі було використано VST-інструмент Ueberschall, що функціонує на базі пропрієтарного рушія Elastik Player.

Крок 1. Формування ритмічної основи (реалізація стилістичної евристики). Цей етап ілюструє класичну схему взаємодії в системі «Людина-Оператор – Детермінований Алгоритм», що розпочинається з постановки евристичної задачі, де оператор формулює вимогу щодо необхідності створення ритмічного фундаменту, який відповідає стилістиці обраного жанру, наприклад, Funk або House. Головною метою є забезпечення «живого» акустичного звучання при мінімізації часу на ручне програмування кожного удару (micro-editing) та збереженні можливості гнучкого керування темпом, для чого було обрано технічне рішення у вигляді VST-інструменту Ueberschall [45]. Цей інструмент базується на архітектурі семплювання та лупів, де базовою одиницею інформації виступає не окрема нота, а попередньо записаний аудіо-фрагмент певної тривалості. Процес взаємодії «Людина-Система» передбачає етап інформаційного пошуку, під час якого оператор здійснює запит у базі даних плагіна, використовуючи систему тегів (Genre, BPM, Instrument). Це суто евристичний процес відбору, що базується на семантичній відповідності тегів творчому задуму, після чого слідує аудиторна валідація, де рішення про вибір кандидатів-лупів приймається людиною суб’єктивно. Завершальним кроком є інтеграція, коли обраний патерн мапиться на певну клавішу MIDI-клавіатури, що перетворює аудіо-дані на динамічний об’єкт, керований MIDI-подією.

З точки зору технічного аналізу, основним рушієм бібліотек Ueberschall є програвач Elastik Player, який, на відміну від генеративного ШІ, що створює нові хвильові форми на основі ймовірнісних розподілів, працює на базі

детермінованих алгоритмів маніпуляції аудіо-даними (Time-Stretching та Pitch-Shifting). Основним технологічним ядром тут виступають алгоритми компанії zplane.development, зокрема сімейство élastique Pro, що є індустріальним стандартом для задач часової та тональної корекції. Фундаментальним для синхронізації різнотемпового матеріалу є алгоритм Time-Stretching (розтягування часу), задача якого полягає в адаптації аудіо-семплу до поточного темпу проєкту без зміни висоти тону та виникнення артефактів. Принцип роботи цього алгоритму, що часто базується на гранулярному синтезі або фазовому вокодері, полягає в аналізі вхідного сигналу та розбитті його на мікроскопічні сегменти або спектральні вікна. Для сповільнення гранули дублюються та накладаються одна на одну з використанням кросфейдів, а для пришвидшення частина гранул вилучається або зменшується відстань між ними. В IT-аспекті це лінійна та оборотна математична операція над масивом аудіо-даних, яка, на відміну від ШІ, не створює нових ритмічних малюнків, а лише трансформує часову вісь існуючого сигналу [34, 215].

Наступним ключовим елементом у ланцюгу обробки є алгоритм Pitch-Shifting з корекцією формант, який відіграє критичну роль у збереженні природності звучання при зміні тональності. Його основна задача полягає у вирішенні фундаментальної акустичної проблеми: лінійна зміна швидкості відтворення (Resampling) неминуче призводить до зміщення всіх частотних складових, включаючи резонансні характеристики корпусу інструменту або голосового тракту, що викликає неприродний ефект «Chipmunk effect». Для уникнення цього явища механізм дії алгоритму базується на складному процесі спектральної декомпозиції, розділяючи спектральну інформацію сигналу на дві незалежні складові: фундаментальну частоту (Fundamental Frequency), що відповідає за музичну ноту і підлягає зсуву на необхідний інтервал, та формантну структуру (Spectral Envelope) – набір незмінних резонансних піків, що формують унікальний тембр та ідентичність джерела звуку. Алгоритм фіксує положення формант на частотній шкалі, не дозволяючи їм зсуватися синхронно із висотою тону. Це суттєво відрізняє даний підхід від роботи генеративних моделей ШІ

типу RVC (Retrieval-based Voice Conversion), оскільки тут виконується математична спектральна трансформація через швидке перетворення Фур'є (FFT) над існуючим вхідним сигналом, а не повний нейромережевий ресинтез звуку з нуля. Така архітектура дозволяє досягти високої точності інтонування при збереженні оригінальної артикуляції та акустичних характеристик запису.

Третім важливим компонентом технологічного стека є алгоритм детекції транзієнтів (Transient Detection), призначений для перетворення суцільного аудіо-потоку на набір дискретних подій, що уможливлює синхронізацію аудіо-матеріалу з жорсткою метричною сіткою такту DAW. Принцип роботи алгоритму полягає у скануванні амплітудної обвідної хвилі з метою ідентифікації різких сплесків енергії – транзієнтів, які зазвичай відповідають атакам музичних інструментів. На основі цих виявлених точок аудіо-файл віртуально нарізається на окремі сегменти або слайси (Slices), кожен з яких стає незалежним об'єктом маніпуляції. В IT-контексті такий підхід дозволяє застосувати до статичних аудіо-даних методи MIDI-квантизації, надаючи оператору унікальну можливість змінювати внутрішній ритм (Groove) аудіо-лупу шляхом пересування окремих слайсів у часі. При цьому виникає проблема розривів між переміщеними сегментами, яку система вирішує автоматично: алгоритм Time-Stretching заповнює виниклі паузи та згладжує переходи, використовуючи методи гранулярного синтезу або інтерполяції для збереження цілісності звучання.

Комплексний аналіз роботи з інструментарієм Ueberschall та zplane дозволяє зробити критично важливий висновок для нашого дослідження щодо розподілу ролей у системі «Людина-Машина». У рамках даної моделі «людський ресурс» виступає як безальтернативний генератор ідеї та єдиний арбітр естетики, здійснюючи семантичний відбір вихідного матеріалу та визначаючи вектор його трансформації. Програмне забезпечення, своєю чергою, виконує функцію високотехнологічного процесора, що використовує суворо детерміновані DSP-алгоритми для технічної адаптації цього матеріалу під вимоги проєкту. Система Elastik, будучи обмеженою фіксованою базою даних семплів, не здатна створити

ритм або мелодію, яких не існує в її бібліотеці; результат роботи її алгоритмів завжди є передбачуваним та ідентичним при однакових вхідних параметрах. Вона слугує лише досконалим засобом реалізації евристичного задуму людини, не проявляючи власної «творчої волі». Це фундаментально відрізняє описану «Еталонну Модель» від принципів роботи генеративного штучного інтелекту, який буде детально розглянуто в другому розділі, де комп'ютерна система бере на себе функцію створення структури, композиції та контенту ймовірнісним шляхом, діючи як «чорна скринька» з непрогнозованим результатом.

Сформувавши стійку темпоральну (ритмічну) сітку проєкту за допомогою алгоритмів тайм-стретчингу та слайсингу, виникає об'єктивна необхідність переходу до наступного рівня абстракції музичного виробництва – формування тонально-гармонічного фундаменту композиції. Якщо на попередньому етапі програмне забезпечення оперувало цілісними, попередньо записаними аудіо-фрагментами (лупами), де гармонічна складова була зафіксованою («запеченою») у хвильовій формі, то робота з басовою лінією вимагає принципово іншої архітектури обробки даних. Специфіка низькочастотного діапазону та гармонічна динаміка твору виключають можливість використання статичних семплів без втрати якості та гнучкості. На цьому етапі система повинна забезпечувати генерацію звуку в реальному часі у відповідь на конкретні MIDI-події, гарантуючи прецизійний контроль над висотою тону (Pitch), фазою сигналу та його тембральними характеристиками (обвідними фільтра та амплітуди). Для вирішення цієї задачі в архітектурі еталонної моделі було застосовано інший клас програмних інструментів – віртуальні синтезатори та семплери, що працюють на основі генерації хвильових форм та субтрактивного синтезу.

Крок 2. Формування низькочастотного фундаменту (Bass Line Synthesis). На цьому етапі вектор взаємодії «Людина-Машина» зміщується від відбору готових патернів до безпосереднього керування синтезом звуку. Першочерговою евристичною задачею стає створення щільної, насиченої басової лінії, яка повинна мати чітку атаку, підтримувати гармонію, задану у

партитурі Finale, та при цьому уникнути частотного конфлікту з бочкою (Kick drum). Для реалізації цього завдання як технічне рішення було використано віртуальний інструмент reFX Nexus [39]. Процес взаємодії у системі розпочинається з MIDI-програмування на основі даних, експортованих з Finale, причому, на відміну від інструментів на кшталт Ueberschall, де одна нота активувала цілий луп, тут кожна MIDI-подія (Note On) ініціює генерацію окремого звуку певної висоти.

Завершується етап модифікацією параметрів, під час якої оператор вручну налаштовує частоту зрізу фільтра (Cutoff Filter) та огинаючі (Envelopes), щоб коректно інтегрувати басову партію в акустичний простір міксу.

Рисунок 1.7 – VSTi reFX Nexus

Технічний аналіз гібридної архітектури інструменту Nexus дозволяє класифікувати його з точки зору інформаційних технологій не як чистий синтезатор, що генерує хвилю математично через осцилятори, а як гібридний семплерний синтезатор (ROMpler). Його функціонування базується на двох

ключових алгоритмічних блоках, які відрізняють його як від систем часової компресії типу Ueberschall, так і від генеративного штучного інтелекту. Першим є алгоритм PCM-відтворення та інтерполяції (PCM Playback Engine): на відміну від Ueberschall, який розтягує звук у часі, ядро Nexus оперує мульти-семплами – заздалегідь записаними звуками реальних синтезаторів. При надходженні MIDI-команди, наприклад «Note On: C2», рушій звертається до пам'яті для пошуку відповідного аудіо-семплу, а у випадку відсутності точного запису для конкретної ноти автоматично активується алгоритм ресемплінгу з інтерполяцією. Математична сутність цього процесу полягає у зміні частоти дискретизації відтворення для зміни висоти тону, причому для уникнення цифрових артефактів та аліасингу застосовуються складні методи, такі як sinc-інтерполяція. Це підкреслює фундаментальну відмінність від ШІ, оскільки звук не генерується нейромережею, а детерміновано викликається з бази даних (ROM) і піддається математичному транспонуванню. Після формування первинного сигналу процес переходить до алгоритму субтрактивного синтезу (Virtual Analog Filter Simulation), де оператор проявляє найбільшу творчу активність у налаштуванні тембру, наприклад, роблячи бас більш «м'яким» або «агресивним». Nexus реалізує це через топологію цифрових фільтрів, зокрема фільтрів з нескінченною імпульсною характеристикою (IIR), що емулюють поведінку аналогових схем, де критично важливу роль відіграє Low-Pass Filter (фільтр низьких частот). [11, 88; 42, 150].

Формалізація процесу роботи фільтра може бути представлена рівнянням рекурсивного фільтра вигляду $y[n] = b_0 x[n] - a_1 y[n-1]$ (спрощена формула рекурсивного фільтра).

Коли користувач регулює параметр «Cutoff», він фактично змінює коефіцієнти у цьому рівнянні, змушуючи алгоритм у реальному часі відсікати гармоніки вище заданої частоти. З точки зору інформаційних технологій, це є строго детермінованим процесом цифрової обробки сигналів (DSP), де поведінка фільтра жорстко визначена його математичною топологією. Окрім спектральної корекції, динаміка звучання басу в Nexus забезпечується алгоритмами модуляції

та генераторів огинаючої (Modulation Matrix & Envelopes). Їхня задача полягає у визначенні характеру зміни гучності та яскравості звуку в часі, для чого алгоритм розраховує криву зміни амплітуди за класичною чотириступеневою схемою ADSR (Attack, Decay, Sustain, Release).

Використання інструменту Nexus ілюструє важливий аспект «Еталонної моделі» – параметричний контроль над тембром, де джерело звуку є фіксованим (бібліотека PCM-семплів), на відміну від штучного інтелекту, здатного синтезувати неіснуючі тембри. У цій схемі алгоритм виконує роль скульптора, що відсікає зайві частоти шляхом фільтрації та змінює висоту тону через інтерполяцію згідно з чіткими інструкціями оператора. Це гарантує абсолютну передбачуваність системи: повторне натискання тієї ж клавіші з ідентичними налаштуваннями фільтра завжди дасть побітово ідентичний результат без внесення випадкових змін. Сформувавши ритмічний та низькочастотний каркас, робочий процес переходить до найбільш складної частини аранжування – наповнення гармонічного простору та реалізації мелодичної лінії. Саме на цьому етапі відбувається критична взаємодія між попереднім задумом, зафіксованим у Finale, та його реальним акустичним втіленням у середовищі Cubase, що неминуче призводить до прийняття нових евристичних рішень.

Крок 3. Гармонічне наповнення та тембральна переоцінка (Melodic-Harmonic Layering). Просторова обробка та спектральне розширення (Spatial Processing). На цьому етапі до роботи залучаються інструменти, що імітують живу акустичну гру (рояль, гітара, саксофон), що вимагає від програмного забезпечення значно вищого рівня реалізму порівняно з синтезованими партіями. Першочергова евристична задача полягала у реалізації гармонічної підтримки та виразної мелодичної лінії, однак у процесі виник когнітивний конфлікт, який вимагав людського втручання.

На етапі проєктування у середовищі Finale мелодична функція була покладена на скрипку, проте після імплементації ритм-секції у Cubase оператор здійснив евристичну переоцінку контексту. Аналіз показав, що тембр скрипки не відповідає стилістичній щільності та «антуражу» оновленого аранжування,

внаслідок чого було прийнято рішення здійснити заміну інструментарію на саксофон. Цей епізод демонструє здатність людини до адаптивної зміни стратегії в реальному часі, на відміну від лінійного виконання інструкцій MIDI-файлу, притаманного автоматизованим системам.

В якості технічного рішення було використано семплерну платформу Native Instruments Kontakt (для роялю, гітари та саксофона) та синтезатор VPS Avenger як Send-ефект для спектрального збагачення. Технічний аналіз рушія Kontakt свідчить про використання складних алгоритмів емуляції фізичної поведінки інструментів, що виходить за межі простого програвання семплів [35]. Зокрема, для роботи з бібліотеками роялю, які можуть займати десятки гігабайтів, застосовується алгоритм DFD (Direct From Disk Streaming). Оскільки завантажити такий обсяг даних в оперативну пам’ять неможливо, алгоритм завантажує в RAM лише перші мілісекунди кожного звуку (атаку), а решта семплу підвантажується з жорсткого диска в реальному часі безпосередньо в момент натискання клавіші, забезпечуючи відтворення тисяч голосів одночасно з мінімальною затримкою.

Для забезпечення реалістичності звучання саксофона критично важливим є використання KSP Scripting (Kontakt Script Processor) та алгоритмів Legato [35, 40]. Проблема неприродного звучання при послідовному відтворенні семплів вирішується спеціальними скриптами на мові KSP, які аналізують вхідний MIDI-потік. При виконанні умови перекриття нот (техніка Legato), алгоритм не запускає новий семпл атаки, а вмикає спеціальний «перехідний» семпл (transition sample) і плавно змінює висоту тону. Це логічний скрипт, що діє за принципом «IF note_overlap THEN play_legato», підкреслюючи відмінність від нейромережевої генерації ШІ. Фіналізація звукової картини здійснюється через просторову обробку та спектральне розширення, де окремі інструменти міксуються через Send-доріжку з використанням VPS Avenger. У цьому контексті Avenger виступає не як джерело звуку, а як процесор ефектів для збагачення текстури (Layering), застосовуючи алгоритми спектральної

маніпуляції для додавання високочастотних гармонік та внутрішні DSP-ефекти реверберації та затримки для створення єдиного акустичного простору.

Крок 4. Формування енергетичної кульмінації (The Breakdown / Bridge). Робота з мелодичними лупами.Окремим викликом для «еталонної моделі» стала реалізація перегри (програшу) – одного з найбільш енергетично насичених та «драйвових» фрагментів композиції, що демонструє приклад радикальної евристичної переоцінки та використання інструментарію Ueberschall у новій якості. У початковому MIDI-проєкті з Finale ця частина була прописана партією скрипки, однак під час імплементації у середовищі Cubase стало очевидно, що скрипковий тембр не забезпечує необхідного рівня агресії для кульмінаційного моменту, тому було прийнято рішення замінити скрипку на електрогітару з важким дисторшном.

Технічна реалізація цього задуму вимагала відмови від стандартних VST-синтезаторів, де важко досягти реалістичного звучання гітарного «чосу», на користь бібліотеки гітарних рифів Ueberschall Elastik. Використання цього інструменту для електрогітари суттєво відрізняється від його застосування для драм-секції, дозволяючи глибше розкрити можливості DSP-алгоритмів через відмінність у типах даних: якщо ударні являють собою переважно атональні, транзієнтні звуки, де головною задачею є збереження атаки, то електрогітара – це тональний, гармонічно насичений матеріал, що містить «запечений» у аудіо-файл складний ланцюг обробки (Distortion, Delay).

Це формує специфічний алгоритмічний виклик, де на перший план виходить не Time-Stretching, а алгоритм зміни висоти тону (Pitch-Shifting). Оскільки гітарний риф записаний у фіксованій тональності, а композиція може вимагати транспонування, Elastik використовує поліфонічний алгоритм зсуву висоти (Polyphonic Pitch Shifting) зі збереженням формант. Це дозволяє перестроїти записану гітару під гармонію треку без перетворення звуку на «синтетичний», зберігаючи природну спектральну обвідну.

Використання лупів, що базуються на повторюваних фрагментах (ostinato), дозволяє досягти ефекту гіпер-реалізму, оскільки відтворюється не просто нота,

а всі нюанси «живої» гри, такі як призвуки медіатора, резонанс струн та природний сустейн, які неможливо повноцінно згенерувати звичайним MIDI-синтезатором. Заміна MIDI-партії скрипки на аудіо-луп електрогітари демонструє гнучкий гібридний підхід оператора: там, де потрібна унікальна мелодія, використовується семплер (Kontakt), а для створення специфічної фактури та драйву застосовуються готові аудіо-патерни з глибокою DSP-адаптацією.

Крок 5. Інтеграція семантичного ядра композиції (Vocal Production Pipeline). Попри високу технологічність інструментальної частини, основним носієм змісту, емоційного посилу та унікальності композиції залишається вокальна партія. Усі попередні етапи виконували функцію акомпанементу. На цьому етапі відбувається перехід від роботи з даними синтезу до роботи з аудіо-сигналом надвисокої роздільної здатності.

Вхідними даними для вокальної партії виступає сигнал High-Resolution Capture, який був перетворений аналого-цифровим перетворювачем (АЦП) у цифровий потік із частотою дискретизації 96 кГц та розрядністю 24 біт. Такий підхід має чітке IT-обґрунтування: використання частоти дискретизації, що значно перевищує поріг людського слуху і, відповідно, частоту Найквіста, дозволяє уникнути критичних помилок аліасингу (Aliasing errors) під час подальшої нелінійної обробки, зокрема компресії та сатурації, забезпечуючи при цьому максимальну математичну точність відновлення хвильової форми. Для інтеграції цього «живого», динамічно нестабільного голосу у щільний електронний мікс було спроєктовано та побудовано складний ланцюг обробки (Vocal Chain), початковим етапом якого стала алгоритмічна корекція висоти тону (Pitch Correction / Auto-Tune).

Першою та фундаментальною ланкою первинної обробки вокального треку є виправлення інтонаційних неточностей. Цей етап є критично важливим для сучасної естетики популярної музики, яка висуває безкомпромісні вимоги до спектральної чистоти та тональної прецизійності вокальних партій. У контексті цифрової звукорежисури це завдання часто реалізується засобами автоматичної

корекції в реальному часі, що дозволяє оптимізувати робочий процес та забезпечити миттєвий контроль якості виконання. Репрезентативним інструментом цього класу в рамках досліджуваної «еталонної моделі» виступає Steinberg Pitch Correct – нативний VST-плагін, інтегрований у архітектуру робочих станцій Cubase та Nuendo. Цей модуль функціонує як спеціалізований DSP-процесор, призначений для виявлення та зміни фундаментальної частоти (f_0) вхідного монофонічного аудіосигналу з метою його адаптації до заданої тональної сітки. На відміну від деструктивних офлайн-редакторів, Pitch Correct працює як Insert-ефект із мінімальною затримкою, що дозволяє вирішувати завдання інтонаційної корекції безпосередньо під час запису (monitoring input) або на початкових етапах зведення, не порушуючи цілісності творчого потоку.

Для реалізації цього завдання застосовуються складні алгоритмічні моделі, аналогічні тим, що використовуються у плагінах Antares Auto-Tune. Їхня технічна сутність базується на багатоступеневому процесі цифрової обробки сигналів, що розпочинається з етапу детекції висоти тону (Pitch Detection). На цьому етапі система аналізує вхідний аудіопотік, розбиваючи його на короткі часові вікна, та використовує алгоритми на базі автокореляційної функції (у часовій області) або швидкого перетворення Фур'є (FFT, у частотній області) для виокремлення фундаментальної частоти (f_0) з-поміж гармонійних обертонів та шумів у режимі реального часу [21; 36, 450]. Важливим аспектом є точність детекції, оскільки будь-яка похибка на цьому етапі призведе до артефактів на виході. Після визначення поточної частоти ноти відбувається процес обчислення відхилення та квантизації. Алгоритм порівнює виявлене значення f_0 із заданою користувачем музичною шкалою (Scale Grid), яка виступає набором дозволених значень частот (наприклад, ноти гами До мінор).

Математичне ядро процесу полягає у розрахунку «дельта-функції» – різниці між поточною частотою виконання та найближчою цільовою нотою (Target Note). У випадку виявлення відхилення активується механізм корекції: до сигналу застосовується алгоритм зміни висоти тону (Pitch Shifting), який математично «підтягує» частоту. Критично важливим параметром тут є

швидкість реакції (Retune Speed), яка визначає, наскільки швидко алгоритм здійснює перехід від вихідної частоти до цільової. При повільних значеннях корекція звучить природно, імітуючи плавне глісандо професійного вокаліста, тоді як при екстремально швидких значеннях (близьких до 0 мс) виникає характерний «ступінчастий» ефект, відомий як «Cher effect» або «T-Pain effect», коли перехід між нотами стає миттєвим, втрачаючи природну інерцію людського голосу. Варто зазначити, що сучасні алгоритми пітч-корекції також інтегрують методи збереження формант, що запобігає виникненню ефекту «бурундука» при значних зсувах тону, розділяючи спектральну огинаючу (що відповідає за тембр) та основну частоту.

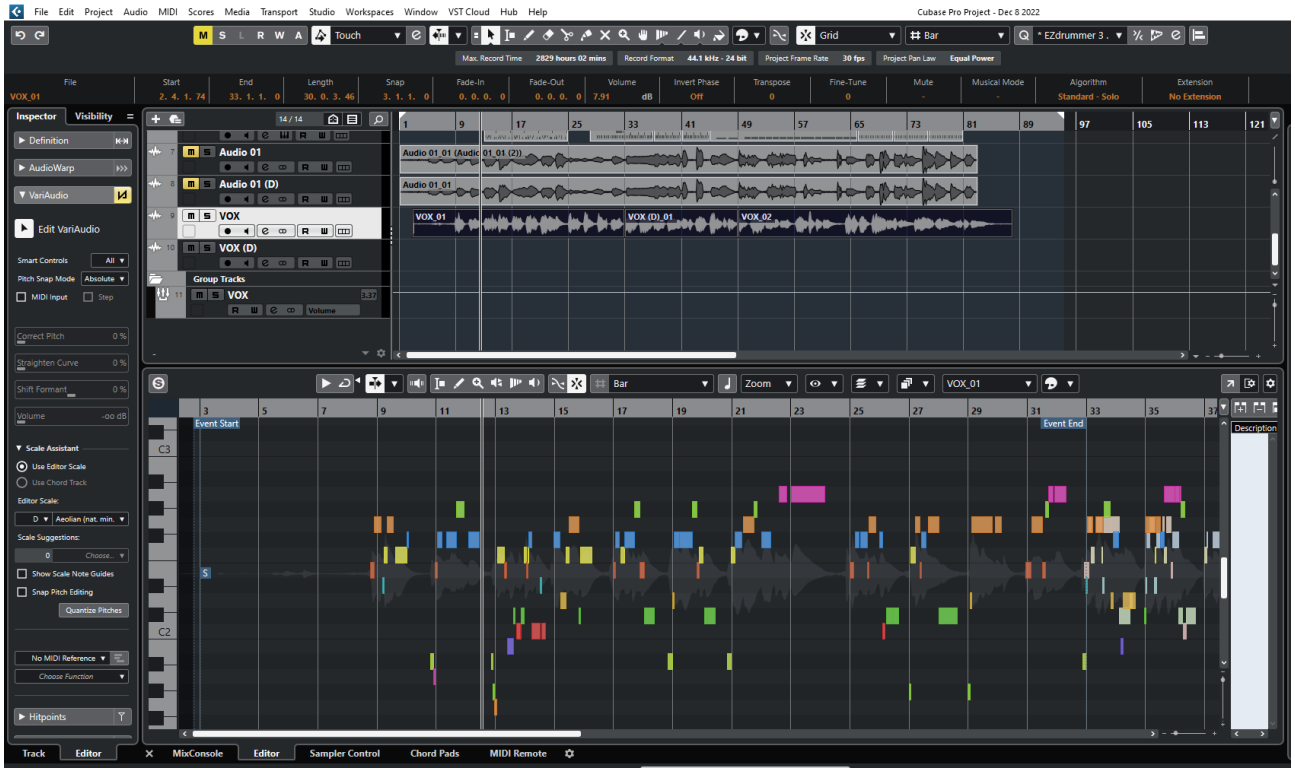


Рисунок 1.8 – Steinberg Pitch Correct

Фундаментальна відмінність цього процесу від роботи генеративного штучного інтелекту полягає в тому, що алгоритм Pitch Correct є суто детермінованим та коригуючим інструментом. Він діє за жорсткою логікою «IF-THEN» (якщо вхідна частота *X*, то вихідна частота *Y*), не маючи здатності генерувати нову мелодію на основі ймовірнісних моделей. Він лише

математично «вирівнює» оригінальне людське виконання відповідно до заданих оператором правил тональності, виступаючи у ролі прозорого технічного посередника, а не співавтора.

Фундаментальна відмінність цього процесу від роботи генеративного штучного інтелекту полягає в тому, що алгоритм автокорекції є суто детермінованим та коригуючим інструментом. Він не генерує нову мелодію на основі ймовірнісних моделей, а лише математично «вирівнює» оригінальне людське виконання відповідно до жорстких правил заданої тональності, зберігаючи при цьому унікальну артикуляцію вокаліста.

1.5. Спектральний аналіз та прецизійна еквалізація (FabFilter Pro-Q 3)

Після корекції висоти тону критичним завданням є частотне очищення сигналу. Для цієї «хірургічної» операції було використано еквалайзер **FabFilter Pro-Q 3**. Вибір саме цього інструменту зумовлений його архітектурою, яка поєднує високоточні DSP-алгоритми фільтрації з потужним модулем візуалізації даних.

Взаємодія «Око-Вухо» (Visual Heuristics) у процесі еквалізації, реалізована через плагін FabFilter, демонструє перевагу сучасних цифрових інструментів над класичними аналоговими приладами завдяки наявності інтерактивного спектроаналізатора. В основі алгоритму візуалізації лежить швидке перетворення Фур'є (FFT) з високою роздільною здатністю, яке дозволяє будувати графік амплітудно-частотної характеристики (АЧХ) в реальному часі. Це надає оператору можливість візуально ідентифікувати проблемні зони, такі як резонанс кімнати на частоті 400 Гц або свист на 3 кГц, і приймати рішення з математичною точністю. Серед основних алгоритмічних операцій обробки було застосовано High-Pass Filtering (Фільтрацію ВЧ) з використанням алгоритму Баттерворта і крутизною спаду 24 дБ/октаву для відсікання спектру нижче 90 Гц, що ефективно усуває низькочастотний «бруд» та звільняє місце для басу та бочки. Для «хірургічного» вирізання вузьких смуг частот, що створюють

неприємні резонанси, застосовувалася Notch Filtering (режекторна фільтрація) з високим Q-фактором.

Рисунок 1.9 – FabFilter Pro-Q 3

Особливої уваги заслуговує алгоритм динамічної еквалізації (Dynamic EQ Algorithm), який є гібридом еквалайзера та компресора: фільтр активується і знижує гучність певної частоти виключно тоді, коли її амплітуда перевищує заданий поріг (Threshold). Це дозволяє автоматично придушувати сибілянти (різкі звуки «с», «ц») у вокалі тільки в моменти їх появи, не впливаючи на звук у тихих місцях. З технічної точки зору, залежно від завдань, використовувався режим Zero Latency на основі IIR-фільтрів для збереження транзієнтів, або режим Linear Phase на основі FIR-фільтрів для уникнення фазових спотворень при глибокій корекції [25; 48, 130].

Для вирішення проблеми широкого динамічного діапазону людського голосу, що варіюється від шепоту до крику і створює нестабільність у міксі, було

застосовано каскад приладів динамічної обробки. Ключовим елементом став компресор, робота якого базується на алгоритмі автоматичного керування підсиленням (Automatic Gain Control): прилад аналізує амплітуду вхідного сигналу і, якщо вона перевищує заданий поріг (Threshold), математично зменшує гучність із певним співвідношенням (Ratio) [28, 340]. Результатом роботи цього алгоритму є вирівнювання гучності, завдяки чому голос стає «щільним» і читабельним у загальному звуковому полотні.

Рисунок 1.10 – Compressor

Наступним етапом є застосування алгоритму лімітування (Limiter), який у архітектурі цифрових робочих станцій функціонує як динамічний процесор із нескінченно високим співвідношенням стиснення (∞:1). Цей інструмент виступає у ролі критично важливого запобіжного механізму (safety tool), гарантуючи, що амплітуда вихідного сигналу за жодних обставин не перевищить

визначеного порогового значення (наприклад, -1 dB). Такий підхід забезпечує технічну чистоту сигналу, ефективно запобігаючи виникненню цифрового спотворення (Clipping) шляхом жорсткого обмеження пікових рівнів вокальної партії.

Рисунок 1.11 – Limiter

Психоакустична обробка та Дабл-трек (Spatial & Texture Layering). Для підсилення енергетичної насиченості у приспіві (Chorus section) було застосовано техніку Дабл-трекінгу (Double Tracking), яка представляє собою метод просторового та текстурного нашарування (Spatial & Texture Layering).

Сутність цього підходу полягає у використанні додаткових вокальних доріжок, що дублюють основну партію, де реалізація базується на специфічному психоакустичному алгоритмі: основний вокал позиціонується в центрі стереопанорами, тоді як дабл-треки розводяться по крайніх точках (Hard Left / Hard Right). Внаслідок мікроскопічних природних розбіжностей у часі та інтонації між записаними дублями виникають фазова інтерференція та ефект

Хааса, що сприймається людським мозком як єдиний, але значно «ширший» та «масивніший» звуковий об'єкт.

Vocal FX Chain (процесор ефектів) – це впорядкована послідовність модулів цифрової обробки сигналів (DSP), спрямована на технічну корекцію, динамічну стабілізацію та художнє збагачення вокального матеріалу. Ефективність такої обробки безпосередньо залежить від дотримання суворої логіки маршрутизації сигналу, де кожен етап готує аудіоматеріал для наступного.

Рисунок 1.12 – Vocal FX Chain

Фінальною ланкою, що забезпечує «поліровку» звуку, є просторова обробка (Spatial Processing). Її головне завдання – інтеграція сухого вокального сигналу у віртуальний акустичний простір треку, створюючи психоакустичне відчуття глибини та об’єму. Для досягнення та підсилення цього ефекту застосовуються спеціалізовані алгоритми просторової обробки, серед яких ключову роль відіграє Plate Reverb (пластинчаста реверберація), що надає вокалу необхідної щільності та тембральної яскравості. Паралельно використовується алгоритм Ping-Pong

Delay (стерео-затримка), який забезпечує розширення стереобази та додає ритмічної динаміки, створюючи складний рух відлунь у панорамі.

Важливо зазначити, що хоча у складних студійних проектах ці ефекти традиційно виносяться на окремі FX-канали (архітектура Send/Return), в рамках лінійного інсерт-ланцюга вони повинні займати передостаннє місце (перед фінальним лімітером). Таке розташування є критичним: воно запобігає компресії реверберaційних «хвостів» разом з основним сигналом, що дозволяє уникнути ефекту звукового «бруду» та зберегти прозорість міксу.

1.6. Інтеграція та ручна синхронізація екзогенного контенту (AI Backing Vocals)

Окремим шаром вокальної фактури стала партія бек-вокалу. Специфіка цього етапу полягає в тому, що аудіо-матеріал не був записаний через мікрофон у поточній сесії, а був попередньо згенерований зовнішніми системами штучного інтелекту (детальний аналіз алгоритмів генерації яких буде проведено у наступних розділах).

У контексті робочого середовища Cubase (Розділ 1) наше завдання полягало в інтеграції та адаптації цього синтетичного матеріалу до «живого» основного вокалу. Вхідними даними для цього етапу слугували згенеровані аудіо-файли (Stems) з партіями гармоній, інтеграція яких виявила суттєву проблему: попри ідеальний пітч, матеріал характеризувався недосконалим таймінгом та фразуванням, що не збігалося з експресією основного вокаліста. Оскільки застосування автоматичних алгоритмів квантизації несло ризик руйнування природності звучання, було обрано стратегію ручної адаптації («Підтасовка» / Manual Alignment) із застосуванням технік Audio Warping та Slicing. Процес розпочався з нарізки (Slicing) аудіо-подій бек-вокалу на окремі фрази та склади, після чого здійснювався Time Alignment, де оператор вручну зміщував початок кожного фрагмента, синхронізуючи його з транзієнтами основного вокалу. Ця операція є критично важливою для уникнення ефекту «каші» (Flam effect), коли

атаки приголосних звуків (наприклад, «т», «к») розходяться у часі, а завершальним етапом монтажу стало створення мікро-кросфейдів (Crossfading) між нарізаними сегментами для нівелювання цифрових клацань.

З метою коректної імплементації штучного бек-вокалу у мікс із живим голосом було проведено спектральну та просторову корекцію, що передбачала застосування агресивної фільтрації. Зокрема, використання Low-Pass фільтра зі зрізом високих частот вище 10–12 кГц дозволило усунути «цифровий пісок» та специфічні артефакти генерації, зробивши тембр м’якшим і більш «фоновим». Додатково було застосовано Stereo Widening для максимального розведення бек-вокалу по панорамі, що звільнило центральний канал для Lead-вокалу. Цей етап наочно демонструє зародження гібридного робочого процесу: хоча первинним джерелом звуку виступив штучний інтелект, фінальна якість та музичність партії були досягнуті виключно завдяки евристичному втручанню людини –ручному редагуванню таймінгу та художній обробці в середовищі Cubase, без чого «сирий» вихід моделі залишився б непридатним для професійного використання.

Крок 6. Фіналізація аудіо-продукту. Системна інтеграція та стандартизація (Post-Production Pipeline). Після завершення етапу аранжування та локальної DSP-обробки окремих джерел проект переходить у стадію технічної фіналізації, що з точки зору інженерії даних знаменує перехід від роботи з дискретними об’єктами –окремими MIDI- та аудіо-подіями –до маніпуляцій з інтегрованим сигнальним потоком. Цей етап архітектурно диференціюється на дві критично важливі фази: зведення (Mixing), яке являє собою процес частотно-динамічної інтеграції множини каналів у єдине когерентне стерео-поле, та мастерінг (Mastering), спрямований на спектральну балансировку і стандартизацію динамічного діапазону вихідного файлу.

Саме на цьому рівні фундаментальна архітектурна відмінність між детермінованим «людським» інструментарієм (Cubase/VST) та ймовірнісним підходом ШІ стає найбільш очевидною та вимірюваною: якщо у «Еталонній моделі» інженер використовує алгоритми як прозорі інструменти для реалізації конкретного художнього контексту, наприклад, навмисного збереження

динамічних піків задля емоційного акценту, то генеративні моделі ШІ, аналіз яких наведено у другому розділі, діють на основі статистичного усереднення, намагаючись адаптувати частотний баланс до «середнього спектру» навчальної вибірки та часто ігноруючи унікальні контекстуальні особливості композиції.

1.7. Зведення (Mixing). Детерміноване сумування проти Нейромережевого прогнозування

Зведення – це процес системної інтеграції розрізнених потоків даних. У середовищі Steinberg Cubase це реалізується через архітектуру MixConsole. В основі цього процесу лежить математика сумування, яку забезпечує 64-бітний рушій з плаваючою комою. Мікшер виконує лінійну операцію додавання амплітуд каналів, що описується формулою $Y[n] = \sum (x_i[n] \times g_i)$. Це процес з нульовою похибкою (крім округлення на 64-му біті) [30, 55; 43, 200].

Це створює різкий контраст із системами «автоматичного зведення» на базі ШІ, такими як iZotope Neutron Assistant, які не обмежуються простим сумуванням, а використовують моделі машинного навчання, натреновані на

```
IF Kick_Signal > Threshold THEN Reduce_Bass_Gain
```

спектрограмах тисяч хітів. ШІ аналізує вхідні треки як «образи» і намагається передбачити ідеальний баланс, підганяючи його під «усереднений стандарт», тоді як у середовищі Cubase оператор зберігає повний контроль над сигналом, уникаючи статистичного усереднення. Для вирішення частотних конфліктів, зокрема у класичній парі «Kick vs Bass» (динамічне маскування), було використано техніку Side-Chain компресії. Алгоритм реалізації цього прийому у Cubase являє собою чітку логічну конструкцію де оператор вручну налаштовує маршрутизацію та параметри Attack/Release.

Це детермінована реакція на подію. Цей метод контрастує з роботою інтелектуальних інструментів на кшталт Smart:EQ або RoEx, які використовують алгоритми спектрального розмаскування (Spectral Unmasking). У таких системах нейромережа автоматично детектує конфліктні частоти між двома джерелами і

динамічно вирізає їх без прямої участі користувача; попри швидкість, такий підхід часто призводить до втрати корисних транзієнтів, які людина свідомо залишила б для збереження характеру звуку.

Наступним кроком є просторова інтеграція, для реалізації якої було використано конволюційну реверберацію (REVerence). Робота цього алгоритму в середовищі Cubase базується на математичній операції згортки «сухого» сигналу з імпульсним відгуком (IR) реального акустичного простору, що описується рівнянням $y[n] = x[n] * h[n]$ і являє собою точну фізичну симуляцію. На противагу цьому, генеративні моделі простору (AI Reverb) часто застосовують методи стильового трансферу (Style Transfer) для синтезу реверберaційних хвостів, намагаючись «домалювати» акустичне середовище. Хоча це дозволяє створювати унікальні віртуальні простори, що фізично не існують, такі алгоритми часто ігнорують критичні фазові проблеми, контроль над якими у детермінованій моделі здійснюється людиною через ретельний вибір IR-імпульсу.

Мастеринг виступає фінальною ланкою «Еталонної моделі», де для досягнення комерційного звучання було застосовано стратегію «Analog-Modeled + Digital Precision». Цей підхід реалізується через гібридний ланцюг обробки, що органічно поєднує алгоритми емуляції вінтажного обладнання для надання характеру та сучасні психоакустичні алгоритми лімітування для забезпечення необхідної гучності. На першому етапі ланцюга для гармонічного забарвлення та еквалізації використано плагін UAD Pultec EQP-1A, який являє собою не просто еквалайзер, а алгоритмічну модель (Circuit Modeling) реального лампового приладу 1950-х років.

В основі його роботи лежить технологія Component Modeling: на відміну від стандартних цифрових еквалайзерів, що використовують прості математичні формули IIR-фільтрів, алгоритми UAD базуються на чисельному вирішенні систем диференціальних рівнянь, які описують фізичну поведінку електричних компонентів – ламп, трансформаторів та конденсаторів [29; 31, 210]. Головна задача цього процесу – внести у сигнал нелінійні спотворення (Harmonic

Distortion), завдяки чому навіть при «нульових» налаштуваннях додаються парні гармоніки, що суб'єктивно сприймається як «теплота» та «аналогова присутність».

Рисунок 1.13 – UAD Pultec EQP-1A

Особливої уваги заслуговує застосування ефекту «Pultec Trick» (Low-End Shaping), що базується на специфічній особливості пасивної схеми: одночасне підсилення (Boost) та ослаблення (Atten) низьких частот, наприклад, на 60 Гц. Через фазові зсуви у віртуальній схемі це формує унікальну резонансну криву, яка робить бас і бочку «щільними», але не «гулкими». Такий підхід контрастує з методами штучного інтелекту, який зазвичай працює з лінійною еквалізацією для досягнення «рівного» балансу, тоді як людина свідомо використовує емуляцію недосконалостей аналогової техніки для художнього фарбування звуку.

Другим етапом мастерингу є фіналізація та максимізація за допомогою програмного комплексу iZotope Ozone 9, де задіяно передові цифрові алгоритми, що працюють без прив'язки до «аналогової фізики». Для контролю простору використано модуль Imager з багатосмуговим алгоритмом розширення стереобази, який ділить сигнал на 4 частотні смуги та застосовує декореляцію каналів у високочастотній ділянці, залишаючи низькі частоти строго в моно для забезпечення фазової сумісності. Ключовим елементом динамічної обробки є Maximizer з режимом IRC IV (Intelligent Release Control). Цей модуль використовує алгоритм Look-ahead, вносячи затримку для сканування вхідного

буфера даних «наперед», що дозволяє виявити піки ще до моменту їх відтворення.

Завершальним етапом технологічного процесу є рендеринг, ресемплінг та дизеринг. Під час конвертації частоти дискретизації (Sample Rate Conversion) застосовуються Brickwall-фільтри (Low-pass filters з надвисокою крутизною спаду) для видалення ультразвукових частот, що перевищують половину нової частоти дискретизації (частоту Найквіста), запобігаючи виникненню аліасингу.

Для зниження розрядності (наприклад, з 32-bit float до 16-bit) використовується алгоритм дизерингу (наприклад, MBIT+), який додає спеціально сформований низькоамплітудний шум для маскування помилок квантування, зберігаючи динамічні нюанси тихих звуків та хвостів реверберації при переході до споживчих форматів.

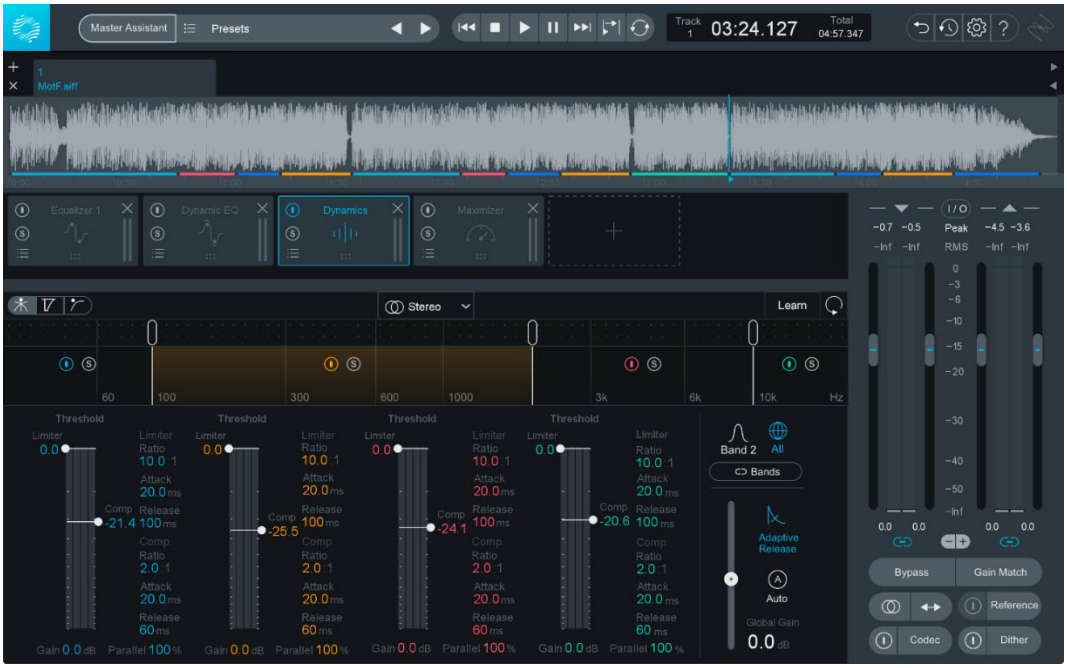


Рисунок 1.14 – iZotope Ozone 9. Mastering

Фінальною операцією виробничого циклу є Mixdown (експорт) – процес трансформації внутрішньої математичної моделі проекту, яка функціонує з частотою дискретизації 96 кГц та розрядністю 64-bit float, у споживчий стандарт Red Book Audio (44.1 кГц / 16-bit). Цей етап є критичним з точки зору інженерії

даних, оскільки він передбачає незворотну втрату інформації, тому головна задача інженера полягає в мінімізації деградації сигналу за допомогою спеціальних алгоритмів [49, 500]. Першим викликом є зниження розрядності (Bit-Depth Reduction), що виникає через фундаментальну різницю форматів: у середовищі Cubase динамічний діапазон завдяки плаваючій комі є фактично нескінченним, тоді як споживчий 16-бітний формат обмежений фіксованою сіткою значень, що містить всього 65 536 рівнів амплітуди. Просте відсікання зайвих бітів (Truncation) призводить до виникнення помилок округлення, які на тихих звуках, таких як затухання реверберації, стають корельованими з корисним сигналом, породжуючи гармонічні спотворення або «шум квантування» (Quantization Noise).

Для вирішення цієї проблеми застосовується алгоритмічний дизеринг (Algorithmic Dithering), зокрема алгоритм MBIT+ у плагіні iZotope Ozone 9. Його робота базується на математичному парадоксі: до сигналу перед зниженням розрядності навмисно додається спеціально згенерований низькорівневий шум, що дозволяє декорелювати помилку квантування та перетворити неприємні гармонічні спотворення на ледь помітне аналогове «шипіння». При цьому використовується технологія Noise Shaping (формування шуму), яка не просто додає білий шум, а «виштовхує» його спектральну енергію в область надвисоких частот (вище 16–18 кГц), де чутливість людського вуха є мінімальною, що дозволяє суб’єктивно зберегти глибину сцени 24-бітного файлу в межах 16-бітного контейнера.

Другим аспектом фіналізації є конвертація частоти дискретизації (Sample Rate Conversion - SRC), необхідна для переходу від студійного стандарту 96 кГц, обраного для точності обробки нелінійних ефектів, до формату 44.1 кГц. Процес Downsampling, що передбачає зменшення кількості відліків за секунду у 2.17 рази, вимагає суворого захисту від аліасингу. Перед видаленням зайвих семплів алгоритм Cubase застосовує надзвичайно крутий (Brickwall) Low-Pass фільтр на частоті Найквіста цільового формату (~22.05 кГц). Задача цього фільтра полягає у повному видаленні ультразвукових частот вихідного файлу, оскільки їх

збереження призвело б до «дзеркального відображення» у чутний діапазон (Aliasing), створюючи негармонійний цифровий бруд; для забезпечення максимальної прозорості та мінімізації фазових спотворень фільтру використовується режим обчислень «Precision».

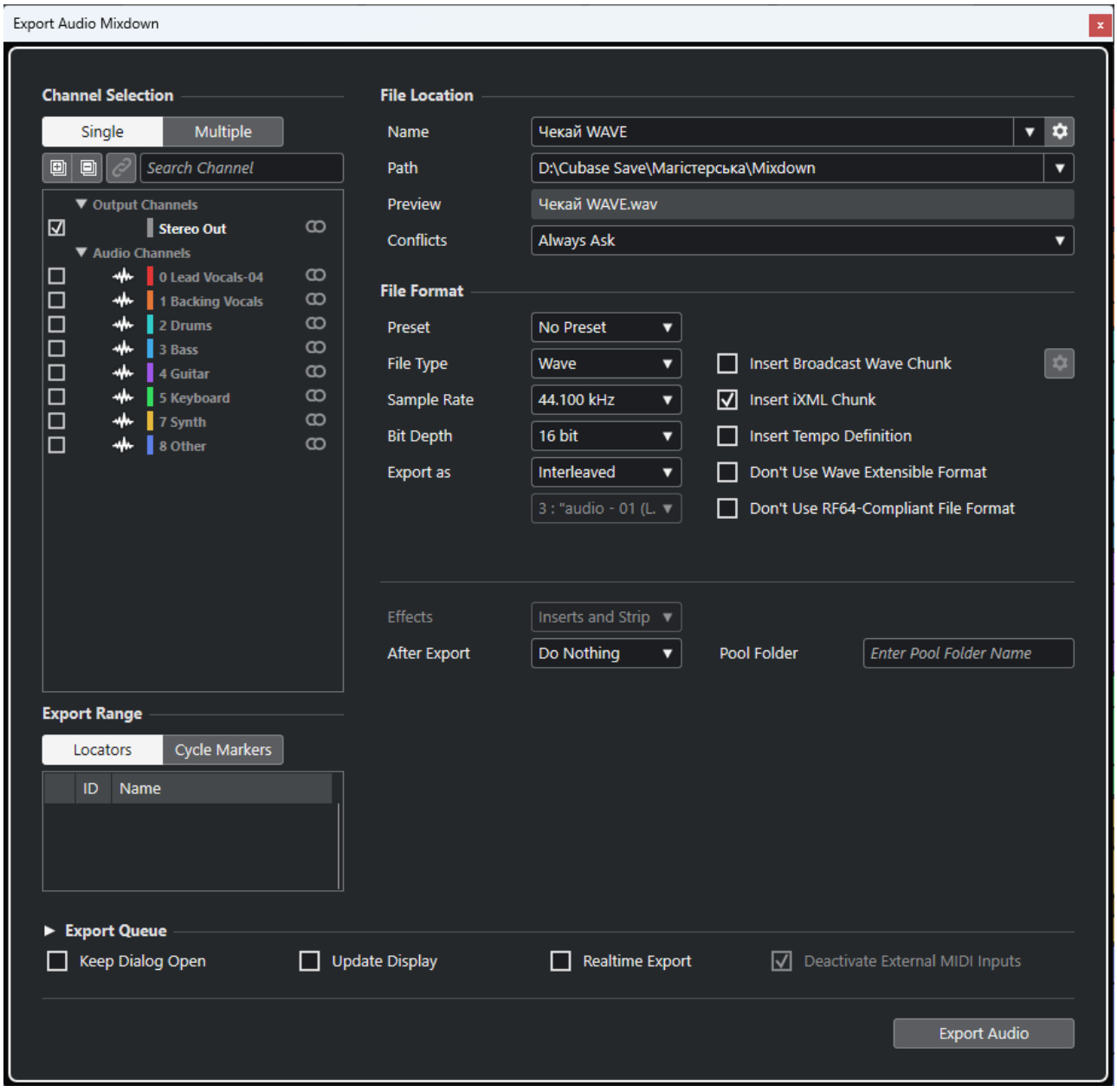


Рисунок 1.15 – Export Audio Mixdown

Процес рендерингу демонструє фінальну перевагу інженерного підходу. Людина свідомо обирає тип дизерингу (Noise Shaping Amount) та алгоритм фільтрації SRC, виходячи з характеру музики (наприклад, для акустичної гітари

потрібен більш м’який дизеринг, ніж для EDM). Натомість, більшість сучасних ШІ-генераторів видають результат одразу в mp3/wav низької якості, приховуючи від користувача артефакти компресії та помилки квантування.

Така архітектурна закритість генеративних систем створює фундаментальну проблему для професійного пост-продакшну, яку можна кваліфікувати як «втрату контролю над математичною моделлю сигналу». У середовищі DAW (Cubase) інженер працює з аудіоданими у форматі 32-bit або 64-bit Floating Point до останнього моменту експорту. Цей формат забезпечує фактично нескінченний динамічний діапазон і дозволяє виконувати мільйони математичних операцій (еквалізація, компресія, сумування) без накопичення помилок округлення. Натомість, більшість End-to-End моделей ШІ (Suno, Udio) здійснюють внутрішній рендеринг у хмарі, використовуючи нейронні кодеки (на кшталт EnCodec), пріоритетом яких є не бітова точність відтворення, а семантична відповідність та швидкість передачі даних. У результаті користувач отримує «напівфабрикат» із вже запеченими артефактами: фазовими зсувами, аліасингом у високочастотному спектрі (відомим як «спектральний зріз» вище 16 кГц) та відсутністю запасу по гучності (Headroom) для подальшого мастерінгу. Більше того, автоматизований рендеринг у системах ШІ ігнорує контекстуальну природу психоакустики. Вибір алгоритму дизерингу (Dithering) або шейпінгу шуму (Noise Shaping) –це не просто технічна формальність, а художнє рішення, що впливає на сприйняття глибини сцени та «повітря» у фонограмі.

Це спостереження остаточно валідує необхідність запропонованої у даному дослідженні Гібридної моделі. Вона передбачає використання ШІ виключно як джерела «сировини» (Raw Material) на ранніх етапах виробництва (генерація стемів, текстур, ідей), але залишає етап фіналізації, сумування та рендерингу виключно в зоні відповідальності людини та детермінованих алгоритмів DAW. Тільки такий підхід дозволяє поєднати генеративну потужність нейромереж із технічними стандартами індустрії (Red Book Audio), забезпечуючи контроль над

бітрейтом, частотою дискретизації та відсутністю помилок квантування, що є обов'язковим критерієм для кваліфікації музичного продукту як професійного.

1.8. Від детермінізму до стохастики: Зміна парадигми обробки даних

У Розділі 1 ми визначили «людський» підхід як взаємодію оператора з детермінованими алгоритмами. У таких системах функція перетворення даних *F(x)* є фіксованою: при однаковому вході *x* ми завжди отримуємо однаковий вихід *y*. Штучний інтелект (Artificial Intelligence), зокрема його підмножина Генеративний ШІ (Generative AI), вводить нову парадигму – стохастичне (ймовірнісне) моделювання [20, 22].

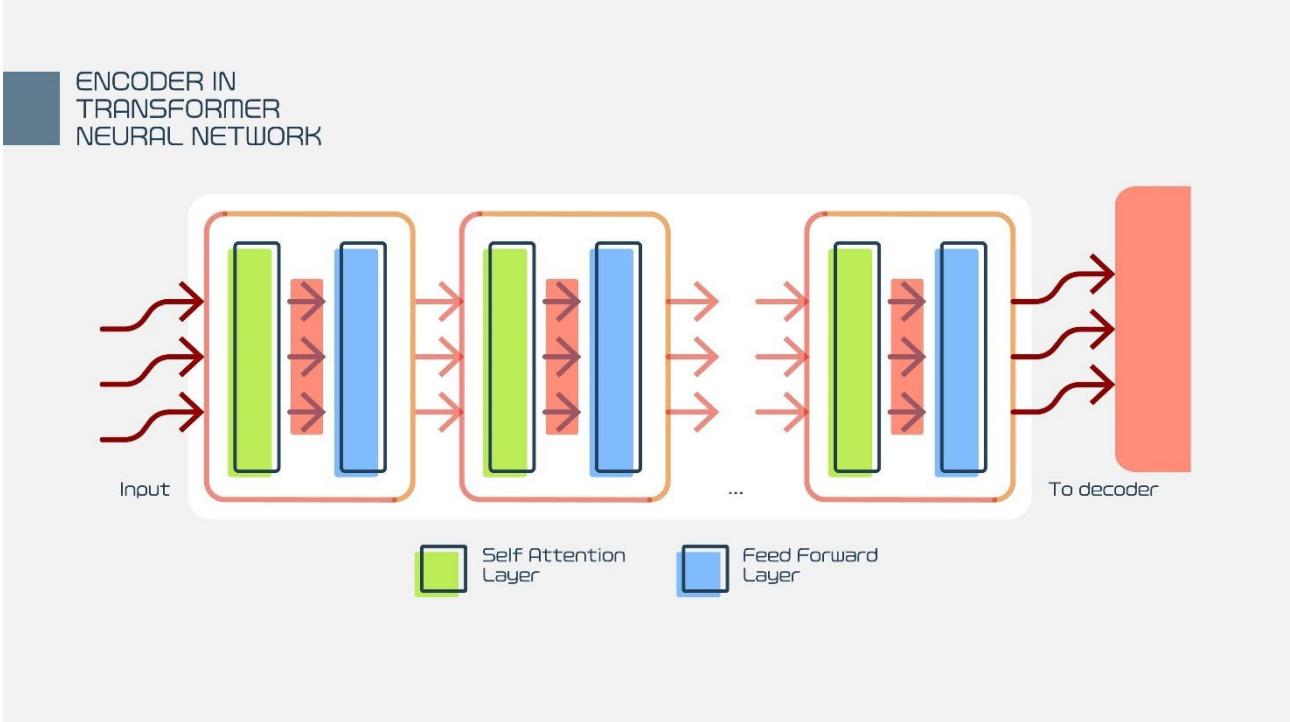
Системи ШІ не мають жорстко закодованих правил композиції (наприклад, «після домінанти має йти тоніка»). Натомість, вони використовують машинне навчання (Machine Learning - ML) для виявлення прихованих закономірностей у великих масивах даних (Big Data) і будують статистичні моделі ймовірності наступної події[38, 210]. Якщо Cubase – це інструмент виконання *команд* (Explicit Programming), то генеративний ШІ – це система виводу (Inference), яка генерує нові дані на основі вивченого розподілу ймовірностей.

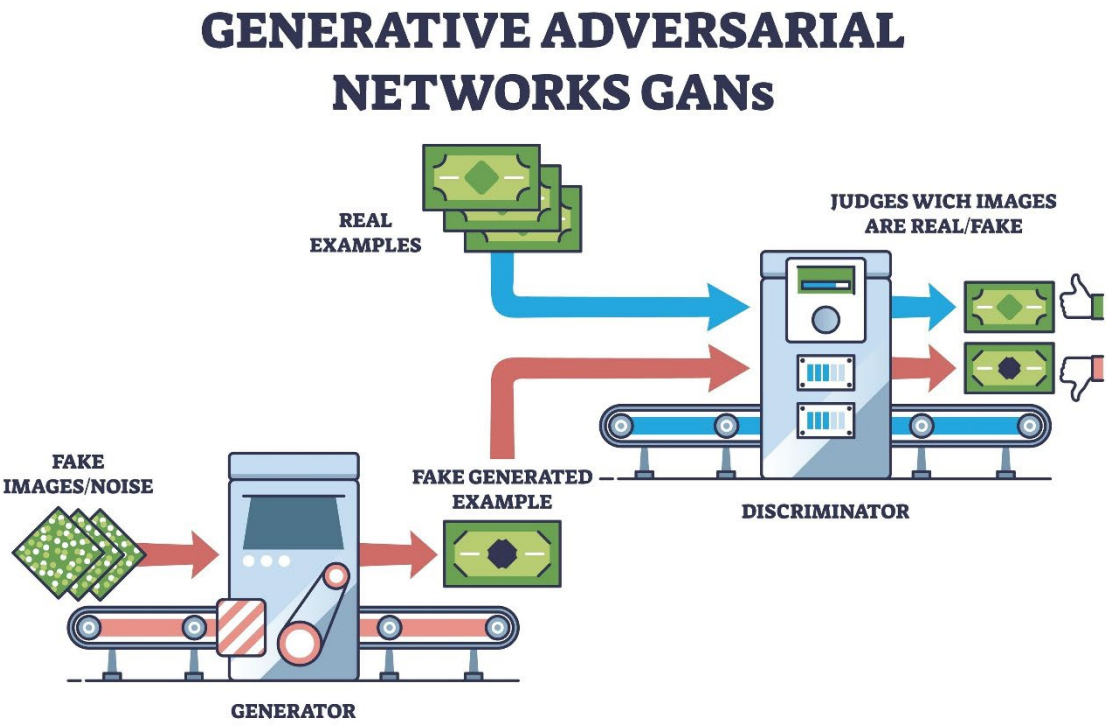
1.9. Огляд архітектур нейронних мереж у музичному аранжуванні

Сучасні сервіси штучного інтелекту для генерації музичного контенту функціонують на базі методів глибокого навчання (Deep Learning), де вибір конкретної архітектури нейромережі безпосередньо детермінується типом оброблюваних даних – чи це символьні потоки (MIDI), чи безпосередньо аудіо-сигнали. Оскільки музика за своєю онтологією є розгортанням послідовності подій у часі, традиційні нейромережі, позбавлені «пам'яті» про минулі стани, виявляються неефективними. Для вирішення цієї проблеми були адаптовані архітектури рекурентних нейронних мереж (RNN), які завдяки наявності зворотних зв'язків здатні запам'ятовувати попередні ноти для імовірнісного

передбачення наступних. Проте для забезпечення когерентності музичного твору частіше використовується вдосконалений тип RNN – мережі довгої короткострокової пам’яті (LSTM). Їхня специфічна коміркова структура дозволяє утримувати контекст на значних часових дистанціях, наприклад, пам’ятати тональність початку твору наприкінці куплету [20, 140], що стало основою для алгоритмів генерації мелодійних ліній у ранніх системах, таких як Google Magenta.

На сучасному етапі індустріальним стандартом, на якому базуються передові моделі на кшталт GPT та MuseNet, стала архітектура Трансформерів (Transformers). Принципова відмінність їх роботи від послідовного зчитування даних в RNN полягає у здатності моделі сприймати всю музичну фразу симультанно. Центральним компонентом цієї архітектури є механізм Self-Attention («Увага»), який дозволяє системі оцінювати вагомість семантичних взаємозв’язків між віддаленими нотами, ігноруючи їх лінійну відстань у тактах [47, 6000; 27].





Окремим класом архітектур, що застосовуються для генерації аудіо-спектрограм, є генеративно-змагальні мережі (GANs), структурна організація яких базується на взаємодії двох конкуруючих нейромереж. Перша з них, Генератор (Generator), виконує функцію синтезу, створюючи музичні образи з вхідного вектору випадкового шуму. Його опонентом виступає Дискримінатор (Discriminator), аналітичний модуль. Генератор постійно вдосконалює свої алгоритми з метою ефективної фальсифікації даних, щоб синтезований результат став статистично невідрізненним від реального аудіо [26, 2675].

1.10. Кейс-стаді: AIVA (Artificial Intelligence Virtual Artist) – генерація символьних даних

Першим об’єктом аналізу у площині алгоритмічної композиції є сервіс AIVA, який позиціонується як прямий технологічний конкурент традиційному етапу «Finale/MIDI-аранжування». Його архітектура базується на обробці

символьних даних (Symbolic Representation), оперуючи дискретними MIDI-подіями замість безпосереднього аудіо-сигналу. Фундаментом для навчання системи слугував масштабний датасет, що охоплює понад 30 000 партитур класичної спадщини (твори Моцарта, Бетховена, Баха), які перебувають у публічному надбанні. З алгоритмічної точки зору, AIVA використовує методи глибокого навчання, інтегровані з парадигмою навчання із підкріпленням (Reinforcement Learning). У цій моделі нейромережа не просто стохастично передбачає наступну ноту, а оптимізує свою поведінку через отримання функції «винагороди» (Reward) за дотримання формалізованих правил класичної гармонії та контрапункту [20, 180].

Процес генерації ініціюється через параметризацію: користувач, замість мануального набору нот, визначає набір мета-параметрів, таких як стиль (наприклад, Epic, Jazz), тональність, темп та хронометраж. У відповідь алгоритм синтезує комплексну багатоканальну MIDI-структуру, автоматично розподіляючи партії між інструментальними секціями, результатом чого є готовий MIDI-файл. Порівняння з «Еталонною моделлю» демонструє онтологічний зсув у творчому процесі: якщо в Finale оператор свідомо конструює мелодію нота за нотою, то AIVA генерує цілісну форму миттєво на основі ймовірнісних розподілів. Водночас, здатність системи експортувати результат у формат MIDI відкриває значний потенціал для гібридних робочих процесів, дозволяючи імпортувати створений штучним інтелектом «скелет» у середовище Cubase для подальшого «оживлення» та професійної DSP-обробки.

1.11. Кейс-стаді: Soundraw – генерація аудіо-структур

Другим об’єктом дослідження виступає сервіс Soundraw, який репрезентує альтернативну технологічну парадигму – генерацію фіналізованого аудіо-продукту, що позиціонує його як безпосереднього конкурента етапу роботи з лупами (Ueberschall/Looping). На відміну від чистого синтезу хвильових форм, притаманного генеративно-змагальним мережам (GAN), або символьної

генерації, реалізованої в AIVA, алгоритмічний підхід Soundraw базується на принципах комбінаторики та «розумного конструювання». Фундаментом системи є масивна база даних коротких аудіо-фрагментів (phrases), створених професійними музикантами та систематизованих за допомогою алгоритмів машинного тегування. Технічним ядром генерації виступають алгоритми задоволення обмежень (Constraint Satisfaction Problems – CSP). Отримуючи вхідні мета-дані, наприклад, запит на «сумний Hip-Hop» певної тривалості з динамічним наростанням, система виконує складну операцію асемблювання треку з тисяч дискретних фрагментів, де головною метою алгоритму є забезпечення безшовності переходів та відповідності кривої енергії заданому сценарію. Специфіка редагування у Soundraw наближає користувацький досвід до роботи в DAW, проте на значно вищому рівні абстракції: оператор отримує можливість змінювати макро-структуру композиції, наприклад, деактивуючи ударні в конкретному такті, але маніпулює він при цьому смисловими блоками, а не окремими плагінами чи параметрами DSP-ефектів, як це відбувається у середовищі Cubase.

1.12. Кейс-стаді: Suno AI – ендогенна генерація спектрально-насиченого аудіо (End-to-End)

Третім і найбільш технологічно складним об’єктом аналізу є платформа Suno AI, яка, на відміну від символьних систем типу AIVA або сегментних конструкторів на кшталт Soundraw, репрезентує клас моделей End-to-End Text-to-Audio. Це означає, що нейромережа генерує фінальний аудіо-сигнал (хвильову форму) безпосередньо з текстового опису, синтезуючи одночасно інструментальний супровід, вокальну партію та акустичний простір. Архітектурно Suno базується на передових методах глибокого навчання, використовуючи Трансформери, подібні до GPT-4, але адаптовані для аудіо-домену (аналогічно до архітектур Bark або AudioLM). Фундаментальною особливістю моделі є відмова від роботи з безперервною хвилею на користь

використання нейронних аудіо-кодеків (Neural Audio Codecs), таких як EnCodec. У цьому процесі аудіо дискретизується на окремі одиниці – токени, а сама модель оперує на двох рівнях передбачення: семантичному, що визначає структуру пісні, текст і мелодію, та акустичному, який відповідає за відтворення тембру голосу, нюансів дихання та загальної «звукової текстури» інструментів. Сам алгоритм генерації реалізовано за принципом авторегресійного передбачення (Next Token Prediction), що визначає стохастичну (ймовірнісну) природу процесу: отримавши на вхід промпт, що містить текст та стилістичні метадані, модель розпочинає послідовне передбачення аудіо-токенів у часі.

$$P(x) = \prod P(x_t \mid x_{<t}, c)$$

(Ймовірність наступного фрагмента звуку x_t залежить від усіх попередніх фрагментів $x_{<t}$ та умови c – промпту) [44; 19; 32].

Ключовою функціональною відмінністю платформи Suno від систем AIVA та Soundraw є здатність до повного синтезу вокальних партій. Навчання моделі на масиві даних, що охоплює мільйони композицій, дозволило системі глибинно засвоїти фонетичні та інтонаційні патерни людської мови. На відміну від семплерів типу Kontakt, які оперують бібліотеками попередньо записаних звуків, Suno здійснює генеративний синтез голосу безпосередньо зі стохастичного шуму на основі обчислених ймовірностей [46]. Для фіналізації аудіосигналу та перетворення грубих акустичних токенів у високоякісну спектрограму, архітектура імплементує елементи дифузійних моделей (Diffusion Models).

У контексті даного дослідження практичне застосування Suno зосереджено на генерації бек-вокальних гармоній, що дозволяє вирішити проблему відсутності інструментів класу Text-to-Singing у традиційних середовищах на кшталт Cubase або Kontakt. Згенерований матеріал підлягає подальшій інтеграції в робочу станцію, як це було описано в попередніх розділах, проте цей метод має суттєве технологічне обмеження: Suno видає результат у форматі вже зведеного («Flat») аудіо-файлу. Це унеможливлює декомпозицію міксу та незалежне керування гучністю окремих елементів відносно голосу, що становить головний

недолік даної технології у порівнянні з гнучкістю мультитрекової архітектури Cubase.

1.13. Проблема «Чорної скриньки» та обмеження контролю

Аналіз трьох систем (AIVA, Soundraw, Suno) дозволяє виявити фундаментальну проблему інтеграції ШІ в професійний продакшн – проблему «Чорної скриньки» (Black Box Problem). На відміну від «Еталонної моделі» (Розділ 1), де оператор має доступ до кожного параметра (частота зрізу фільтра в Nexus, атака компресора в Ozone), системи ШІ приховують проміжні етапи обробки.

Таблиця 1.1.

Рівні абстракції та контролю

Критерій	**Еталонна модель (Людина + Finale, Cubase)**	**AIVA (Symbolic AI)**	**Soundraw (Structural AI)**	**Suno (Generative Audio)**
Тип даних	MIDI + Audio + DSP Parameters	MIDI	Audio Segments	Raw Waveform
Рівень контролю	**Високий** (мікро-редагування кожного біта)	**Середній** (редагування нот)	**Низький** (редагування блоків)	**Мінімальний** (тільки текст промпту)
Природа алгоритму	Детермінована (прозора)	Стохастична + Правила	Комбінаторна (CSP)	Авторегресійна (Стохастична)
Генерація тембру	Семпли/Синтез (Nexus/Kontakt)	Ні (використовує Soundfont)	Готові лупи	**Синтез з нуля** (Neural Vocoder)
Передбачуваність	100% (Результат повторюваний)	Варіативна	Варіативна	Низька (Кожна генерація унікальна)

У цьому розділі було досліджено альтернативну парадигму музичного виробництва – алгоритмічну генерацію на базі штучного інтелекту. Аналіз технологічного стека демонструє, що сучасні системи імплементують архітектури Трансформерів та Дифузійних моделей для створення контенту, який традиційно вимагав людської когнітивної праці, як-от композиція мелодії в AIVA, або фізичного виконання, наприклад, синтез вокалу в Suno. Цей технологічний зсув спричиняє фундаментальну трансформацію ролі оператора: відбувається перехід від функціональної моделі «Інженера/Виконавця», притаманної роботі в середовищі Cubase, до концепції «Промпт-інженера/Куратора». Водночас ключовим викликом залишається конфлікт інтеграції, зумовлений несумісністю форматів даних та робочих процесів: якщо Suno генерує фіналізований аудіо-мікс, що ускладнює його деконструкцію, то AIVA надає «сухий» символьний матеріал, який потребує подальшого якісного озвучення та інженерної обробки.

РОЗДІЛ 2.
ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА ОБҐРУНТУВАННЯ ГІБРИДНОЇ МЕТОДОЛОГІЇ СТВОРЕННЯ ЦИФРОВИХ МУЗИЧНИХ ПРОДУКТІВ

2.1. Методологія та дизайн експерименту

Метою практичної частини дослідження визначено проведення порівняльного аналізу ефективності трьох альтернативних моделей музичного виробництва в умовах реалізації реального проєкту. Для забезпечення наукової верифікованості та чистоти експерименту було зафіксовано набір незмінних вхідних параметрів (Control Variables). В якості вхідного вектора даних (Input Vector) виступає оригінальна музична композиція, попередньо формалізована у середовищі Finale, що включає мелодійну лінію, гармонічну сітку, темп (BPM) та тональність. Цільовим результатом (Target Output) визначено створення завершеного музичного аранжування у жанрі Eurodance тривалістю 4:35 хв, яке обов’язково передбачає наявність основної вокальної партії та бек-вокалу.

Оцінка результатів здійснюється за трикомпонентною системою критеріїв: часові витрати (Time Complexity), що вимірюють період від імпорту даних до отримання фінального майстер-файлу; керованість (Controllability), яка визначає ступінь можливості прецизійного редагування окремих нот або тембральних характеристик; та якість контенту, що базується на суб’єктивній оцінці реалізму звучання та відповідності первинному художньому задуму. Дизайн експерименту передбачає послідовну реалізацію трьох сценаріїв: Сценарій А (Reference Model), що репрезентує класичний евристичний підхід (Людина + Cubase) і детально проаналізований у першому розділі; Сценарій Б (Pure AI Model), який базується на генерації аранжування виключно засобами штучного інтелекту (AIVA/Suno) без глибокого редагування; та Сценарій В (Hybrid Model), що передбачає інтеграцію генерованих ШІ-елементів у професійне середовище DAW (Cubase) [6, 30].

2.2. Експеримент 1. Чисто алгоритмічна генерація (Pure AI Approach)

У цьому сценарії ми спробували делегувати задачу аранжування нейромережам, використовуючи матеріали з Finale як вхідний промпт. MIDI-файл мелодії з Finale завантажено в AIVA як «Influence Track» без втручання людини в процес прийняття творчих рішень.

Генерація символьної структури (AIVA) . Сервіс AIVA було використано для спроби швидкого створення акомпанементу на основі мелодії.

Налаштування генерації.

Стиль: Acoustic Piano / Pop.

Задача: Створити «лайтовий», прозорий акустичний варіант аранжування, фокусуючись на фортепіанній фактурі.

Параметри гармонії: «Near exact» (збереження оригінальної гармонічної сітки з Finale).

Система згенерувала технічно коректну фортепіанну партію (арпеджіо та акордовий супровід) за 2 хвилини. ШІ створив спокійну баладу, AIVA ідеально підходить для створення піанінних ескізів, але не здатна самостійно «відчути» потребу в драйві та зміні інструментарію без детального ручного перепрограмування, тоді як «Еталонна модель» (створена людиною в Cubase) передбачає енергійну композицію з використанням барабанних лупів Ueberschall, дисторшн-гітари і інше.

Генерація аудіо на основі деталізованого текстового промпту (Text-to-Audio Approach). Наступним кроком стало тестування моделі Suno AI в режимі генерації «з нуля». Для перевірки здатності нейромережі розуміти складні структурні та аранжувальні інструкції було сформовано розширений інженерний промпт англійською мовою.

Налаштування генерації.

Текст пісні: Повний текст (куплети, приспів) українською мовою.

Текстовий промпт: «An energetic Eurodance track with a driving rhythm and melancholic male vocals. The song has a distinct synthesizer melody, a consistent kick

drum and a pulsating bassline. The instrumentation includes synthesizers for pads, solo parts and bass, as well as electronic drums. The vocalist sings in Ukrainian with a clean, slightly melancholic tone, conveying the lyrics with a sense of longing. The song structure follows the typical verse-chorus format with an instrumental introduction and an energetic instrumental break of an electric guitar with distortion. After the instrumental break, the chorus sounds in a +1 modulation. The tempo is upbeat, and the key is minor, creating a contrast between the energetic music and the emotional vocals. Production elements include a clean mix with a strong emphasis on synthesizer melodies and a reverb effect on the vocals.»

У площині стилістичної відповідності нейромережа продемонструвала успішну ідентифікацію жанрових маркерів Eurodance, що проявилося у релевантному підборі спектрально яскравих синтезаторних тембрів та формуванні характерної ритмічної пульсації басової лінії. Лінгвістична обробка також виявилася задовільною: синтез фонем української мови був розбірливим, проте семантико-емоційне моделювання виявило слабкі місця архітектури – запит на «меланхолійний» відтінок виконання був реалізований алгоритмом досить умовно, часто вступаючи у тональний конфлікт із мажорними інтонаціями автоматично згенерованої мелодії.

Найбільш критичними виявилися структурні помилки (Failure Cases), де система продемонструвала нездатність коректно інтерпретувати та виконувати складносурядні інженерні інструкції. Зокрема, імперативна вимога щодо модуляції на тон вгору (key change) у фінальній частині твору була повністю проігнорована в обох ітераціях генерації.

Критичний висновок. Експеримент підтвердив, що навіть при використанні надзвичайно детального опису, ШІ діє як «генератор випадкових чисел» у питаннях форми. Він створив нову мелодію та гармонію, повністю проігнорувавши оригінальний авторський задум (зафіксований у Cubase), що робить цей метод непридатним для задач точного аранжування конкретного твору.

Логічним продовженням експерименту став етап аналізу похибок мультимодального трансферу на платформі Suno AI, де було застосовано підхід «Audio-to-Audio». Ця методологія передбачає відмову від генерації «з нуля» на користь трансформації існуючого матеріалу, де перед системою ставиться задача реінтерпретації звучання (зміни аранжування) при суворому збереженні музичної сутності твору. Вхідні дані (Multimodal Input) для нейромережі формувалися двокомпонентно. По-перше, у якості акустичного якоря (Audio Reference) було завантажено попередній міксдаун (Bounce) інструменталу, створеного в середовищі Cubase в рамках «Еталонної моделі». По-друге, вектор трансформації задавався через текстовий промпт, що містив опис цільової стилістики (конвертація поп-аранжування у Eurodance, специфікація вокальних партій: Lead Vocals – male, Backing Vocals – female) та ліричний текст. Фундаментальною метою цього етапу стала верифікація здатності алгоритму виконувати коректний стильовий трансфер (Style Transfer), оцінюючи, наскільки точно модель здатна утримувати оригінальні мелодійні, гармонічні та ритмічні константи в умовах зміни тембральної палітри.

В результат експерименту та аналізу згенерованого матеріалу було виявлено критичні невідповідності, які роблять результат непридатним для професійного використання без суттєвих правок (які неможливі у зведеному файлі), а саме гармонійна нестабільність (Harmonic Hallucinations), мелодична девіація та просодичні помилки (Prosody Errors). Алгоритм не зміг точно проаналізувати та відтворити гармонічну сітку оригінального треку [13]. У згенерованому аудіо спостерігалася довільна заміна акордів (Re-harmonization), що призводило до тональних конфліктів та зміни настрою композиції. ШІ не сприйняв вхідну мелодію як константу. Основні мотиви були видозмінені, змінено інтервали, ритмічний малюнок та фразування. Фактично, нейромережа згенерувала варіацію на тему, а не аранжування конкретної мелодії. При генерації вокалу виникли проблеми з лінгвістичною коректністю. Алгоритм розставив неправильні наголоси в словах, спотворив фонетику та ігнорував логічні паузи в тексті, що зробило виконання неприродним і семантично викривленим [44].

Експеримент довів, що поточні моделі аудіо-генерації (на прикладі Suno v5) при роботі з референсним аудіо діють занадто стохастично. Вони не здатні виконувати функцію «аранжувальника», який суворо слідує партитурі. Система «домальовує» дані, ігноруючи авторський задум (гармонію, мелодію, наголоси), що підтверджує необхідність використання детермінованих інструментів (Cubase/Finale) для фіксації ключових елементів композиції.

2.3. Експеримент 2. Алгоритм гібридної інтеграції (Workflow Steps)

Крок 1. Генерація «сировини» в Suno AI. Ми використали Suno не для створення всієї пісні, а як генератор семплів.

Промпт: Було задано текст приспіву та стильові теги (наприклад, «Female backing vocals, clean acapella, backing vocals, female voice, three voices, Chords: C minor, A flat minor, F minor, G major, C minor»).

Особливість: Використано режим «Instrumental» (якщо потрібно тільки гармонії) або генерацію з текстом, фокусуючись на отриманні варіацій голосу.

Крок 2. Екстракція даних (Stem Separation). Згенерований ШІ файл є «плоским» (зведеним). Для роботи в Cubase нам потрібен тільки голос. В процесі використано зовнішній алгоритм розділення стемів, та відділення нейромережею вокальної партії від інструментального шуму генерації Suno.

Крок 3. Імпорт та Адаптація в Cubase (Integration). Отриманий аудіо-файл («AI_Backing_Vox.wav») було імпортовано в сесію Cubase, але ШІ згенерував вокал у власному «плаваючому» темпі, який не співпадав з нашим темпом BPM. Інженерним рішення стало застосовання інструменту AudioWarp у Cubase (Manual Time-Warping) [7, 25]. Алгоритм Cubase (*Detect Transients)* визначив початок складів у файлі ШІ, оператор вручну «прив’язав» ключові фрази бек-вокалу до сітки такту, синхронізувавши їх з основним «живим» вокалом.

Крок 4. Спектральне маскування артефактів (Mixing AI Content). Аналіз спектральних характеристик вихідного сигналу платформи Suno виявив

наявність специфічних «металевих» артефактів цифрового стиснення, що вимагало застосування агресивної коригуючої обробки для їх нейтралізації. Для вирішення цієї проблеми було реалізовано стратегію частотної фільтрації, що передбачала використання жорсткого Low-Pass фільтра зі зрізом частот вище 10 кГц; це дозволило ефективно приховати цифрові спотворення та аліасинг у високочастотному спектрі, які є найбільш чутними для людського вуха. Додатково, з метою психоакустичного маскування недосконалостей синтезу, було застосовано просторову обробку з сильною реверберацією. Цей крок забезпечив «розмиття» атаки транзієнтів, трансформуючи дискретний ШІ-вокал на цілісний атмосферний фоновий шар (Pad) та інтегруючи його у глибину звукової сцени [12; 25].

Результатом імплементації Гібридної моделі став виражений синергетичний ефект, де фінальна композиція органічно поєднала структурну чіткість «Еталонної моделі» з ресурсною гнучкістю генеративного інтелекту. Детальний аудит отриманого аудіо-продукту дозволяє виокремити ключові переваги цього підходу, насамперед – архітектурну стійкість (Structural Integrity). На противагу експерименту з «Чистим ШІ», де нейромережа демонструвала схильність до стохастичного спотворення гармонії та форми, у Гібридній моделі ритмічний фундамент (Ueberschall), басова лінія (Nexus) та гармонічна сітка (Kontakt) залишилися суворо детермінованими, демонструючи 100% відповідність вихідній партитурі з Finale. Такий результат став можливим завдяки чіткому розмежуванню зон відповідальності: ШІ було делеговано виключно генерацію текстурного шару (бек-вокалу), що дозволило використовувати систему як генератор семплів, а не як автономного композитора, ефективно нівелюючи ризик виникнення «галюцинацій».

У площині ресурсної ефективності (Resource Optimization) реалізація вокальних партій через нейромережеві алгоритми продемонструвала критичну перевагу в логістичних вимірах часу, бюджету та якості. Генерація 10 варіативних дублів зайняла сумарно близько 15 хвилин, тоді як традиційний цикл звукозапису – від кастингу вокалістів і бронювання студії до етапів

розспівування та трекінгу – вимагав би від 4 до 8 годин висококваліфікованої праці. Завдяки застосуванню алгоритмів розділення джерел (Stem Separation) та подальшій професійній обробці в середовищі Cubase, фінальна якість звучання бек-вокалу у контексті щільного міксу стає фактично невідрізненною від студійного запису середнього рівня, що підтверджує ефективність методів інженерного маскування артефактів (Artifact Masking).

Визначальним фактором успішної інтеграції ШІ у професійний мікс стала роль людини-інженера, яка виступила в якості «фільтра якості» для згенерованого матеріалу. Спектральний аналіз «сирого» сигналу від Suno AI виявив критичні технічні недоліки, характерні для нейронних вокодерів, зокрема неприродний «металевий» відтінок у середньо-високому діапазоні та відчутне падіння бітрейту на частотах вище 10 - 12 кГц, що робило матеріал непридатним для використання на передньому плані. Для вирішення цієї проблеми було застосовано комплексну стратегію DSP-маскування: за допомогою еквалайзера FabFilter Pro-Q 3 здійснено агресивну субтрактивну фільтрацію (High-Cut) для фізичного видалення верхнього шару з «цифровим піском», а застосування сильної реверберації дозволило «розмити» транзієнти та нівелювати мікро-ритмічні неточності, перетворивши дискретні фрази на суцільний атмосферний шар. Цей експеримент довів, що ШІ виступає лише постачальником «сировини», тоді як людина, використовуючи інструментарій Cubase, надає їй прийнятну акустичну форму, без чого матеріал ШІ класифікувався б як технічний брак.

Окремої уваги заслуговують застосовані DSP-рішення, спрямовані на нівелювання спектральних дефектів синтезованого матеріалу. Зокрема, використання прецизійної фільтрації за допомогою FabFilter Pro-Q 3, де було задіяно High-Cut фільтр для відсікання частот вище порогового значення, у поєднанні з глибокою просторовою обробкою (Reverb), дозволило ефективно маскувати цифрові артефакти та інтегрувати штучний вокал у загальну звукову картину. Критичним аспектом, виявленим у ході експерименту, є естетична гнучкість системи. Ця характеристика визначає здатність гібридної моделі до адаптації та варіативності, що є недосяжним при використанні монолітних

генерацій «End-to-End» систем, де зміна одного елементу часто вимагає повної регенерації твору з непередбачуваним результатом.

Цей підхід також вирішує проблему «аудіальної стерильності», оскільки слухач підсвідомо фокусується на живих артикуляціях лід-інструментів, сприймаючи згенерований фон як природний акустичний простір, а не набір алгоритмів. Поєднання високоякісних бібліотек із нейромережами дозволяє досягти ефекту «Wall of Sound» без частотних конфліктів, властивих нагромадженню стандартних VST, адже ШІ генерує цілісний образ, а не окремі хвилі. Аранжувальник при цьому трансформується з оператора MIDI-редактора в режисера звуку, який контролює емоційну кульмінацію через DAW, але делегує рутинне заповнення спектру штучному інтелекту. Це не просто економить час, а відкриває доступ до кінематографічної якості продакшну, де людське відчуття ритму та гармонії посилюється безмежними тембральними можливостями машинного навчання.

Людський слух еволюційно налаштований на детекцію мікроскопічних нюансів у біологічно релевантних звуках – голосі, диханні, фізичній вібрації струни. Коли слухач чує аутентичну, записану людиною або відтворену через висококласний семплер (Kontakt) лід-партію, мозок автоматично маркує весь звуковий потік як «справжній». Це створює когнітивний буфер, який дозволяє нейромережевим текстурам (Suno AI, AIVA) існувати на задньому плані, не викликаючи відторгнення своєю синтетичною природою. Штучний інтелект у цій схемі діє подібно до CGI-графіки у кіно: вона працює ідеально лише тоді, коли на передньому плані грають живі актори, відволікаючи увагу від недосконалостей фону.

У підсумку, гібридна інтеграція не просто оптимізує ресурси, а демократизує доступ до звучання «блокбастерного» рівня. Те, що раніше вимагало бюджету на оренду симфонічного оркестру або запис професійного хору, тепер стає доступним через грамотний промпт-інжиніринг та подальшу DSP-обробку. Це відкриває еру «гіпер-реалізму» в музиці, де межа між фізичним виконанням та цифровою генерацією стирається на користь емоційного впливу

на слухача, дозволяючи створювати продукти кінематографічної якості в умовах project-студії.

2.4. Комплексний порівняльний аналіз моделей аранжування

На основі емпіричних даних, отриманих в ході практичного моделювання трьох сценаріїв у Розділі 3, було проведено глибокий компаративний аналіз. Метою цього аналізу є не просто визначення «кращого» чи «гіршого» методу, а виявлення специфічних інженерних та творчих характеристик кожного підходу для формування об'єктивної картини сучасного музичного виробництва. Для оцінки було розроблено систему з п'яти ключових критеріїв, що є критичними для професійної індустрії.

Таблиця 2.1.

Порівняльна характеристика моделей аранжування

Критерій оцінки	**Модель А (Евристична / Людина + Finale/Cubase)**	**Модель Б (Алгоритмічна / Чистий ШІ - Suno)**	**Модель В (Гібридна / Cubase + AI Stems)**
1. Часові витрати (Time Efficiency)	**Високі.** Повний цикл (аранжування, запис, редагування, зведення) зайняв близько 40 робочих годин. Процес є лінійним і трудомістким.	**Мінімальні.** Генерація повної композиції займає від 30 секунд до 2 хвилин. Це забезпечує миттєвий результат.	**Оптимальні.** Скорочення часу на створення складних текстур (бек-вокал) на 30-40% при збереженні ручного контролю над основою.
2. Рівень контролю (Controllability)	**Абсолютний (Micro-level).** Оператор має доступ до кожного біта даних: velocity ноти, параметрів синтезу (ADSR), автоматизації ефектів.	**Низький / Відсутній (Black Box).** Користувач не може змінити окрему ноту в акорді або баланс інструментів у згенерованому файлі.	**Високий (Hybrid).** «Скелет» треку повністю контролюється, елементи ШІ адаптуються вручну (Slicing, Time-Warping).

3. Передбачуваність результату	**Детермінована.** Результат на 100% відповідає вхідній партитурі з Finale. Відсутність випадкових відхилень.	**Стохастична.** Висока ймовірність «галюцинацій» (самовільна зміна гармонії, мелодії, тексту, структури).	**Керована.** Людина діє як фільтр, відбираючи тільки коректні генерації та інтегруючи їх у детерміновану структуру.
4. Якість аудіо-сигналу (Signal Quality)	**High-Resolution (96kHz/24bit).** Професійний студійний стандарт. Відсутність артефактів стиснення.	**Low/Mid-Resolution.** Наявність чутних артефактів нейронної компресії, металевий відтінок на високих частотах (вище 12 кГц).	**Професійна.** Артефакти ШІ-семплів маскуються за допомогою спектральної фільтрації та просторової обробки у міксі.
5. Інтеграція вокалу та тексту	**Складна.** Вимагає «живого» запису, тюнінгу, таймінгу та зведення. Залежність від фізичного стану вокаліста.	**Автоматична.** ШІ генерує вокал, але часто з помилками в просодії (наголоси), інтонації та емоційному забарвленні.	**Гібридна.** Основний вокал – «живий» (емоція), бек-вокал – ШІ (текстура). Найкращий баланс якості та ресурсів.

Детальний порівняльний аналіз зафіксував, що Модель А характеризується низькою часовою ефективністю через значний обсяг ручної праці, тоді як Модель Б забезпечує революційну швидкість, але ціною зниження якості. Оптимальний баланс демонструє Гібридна Модель В, яка дозволяє делегувати рутинні операції алгоритмам, вивільняючи час інженера для креативної роботи [8, 15].

Водночас критичним обмеженням «чистого ШІ» залишається проблема «Чорної скриньки»: неможливість декомпозиції міксу та внесення точкових правок (наприклад, зміни гучності окремого інструменту) робить його непридатним для комерційного використання, на відміну від гнучкої архітектури DAW у Моделях А та В [1, 145].

2.5. Аналіз алгоритмічної ефективності: протистояння детермінізму та ймовірності

Ключовим теоретичним висновком роботи є технічне обґрунтування несумісності парадигми «чистого ШІ» з вимогами професійного аранжування на поточному етапі розвитку технологій. В основі цієї проблеми лежить фундаментальний конфлікт між необхідністю збереження структурної цілісності музичного твору та ймовірнісною природою генеративних моделей. Музична композиція розглядається як складна ієрархічна система із жорсткими часовими та гармонічними залежностями, формалізованими у нотному тексті. Детерміновані алгоритми цифрових робочих станцій (DAW), таких як Finale або Cubase, функціонують на базі подієвої логіки (Event-Driven Logic). Це гарантує, що музична подія, розміщена у конкретній точці тактової сітки, буде відтворена зі 100% точністю щодо висоти та часу, забезпечуючи повну відповідність авторському задуму [14, 120].

Натомість, ймовірнісні алгоритми ШІ, зокрема архітектури Трансформерів, оперують механізмом авторегресійного передбачення (Next-Token Prediction), де кожен наступний фрагмент звуку генерується на основі попереднього контексту. Оскільки ймовірність вибору «правильного» токена ніколи не досягає абсолютних 100%, на довгих часових дистанціях (3–4 хвилини) неминуче виникає ефект «накопичення помилки», що експериментально проявляється у втраті ритмічної стабільності або виникненні довільних, незапланованих модуляцій [47].

У контексті розробленої Гібридної моделі відбулося концептуальне переосмислення ролі традиційних алгоритмів DSP-обробки. Інструменти, які раніше слугували виключно цілям художньої виразності, трансформувалися у засоби технічної валідації та корекції. Зокрема, спектральне маскування за допомогою Low-Pass фільтрів дозволило нівелювати специфічні високочастотні артефакти генерації, а використання інтенсивної реверберації забезпечило

просторове розмиття транзієнтів, перетворюючи ритмічно недосконалі фрази ШІ на цілісний атмосферний шар (Pad).

Це доводить, що інженерна обробка людиною є обов'язковим буферним етапом виробничого ланцюга, виступаючи гарантом технічної якості та трансформуючи стохастичний, непередбачуваний результат роботи нейромереж у контрольований професійний стандарт, позбавлений структурних та спектральних дефектів.

2.6. Розробка та формалізація гібридного робочого процесу (Proposed Hybrid Workflow)

На основі проведеного дослідження та аналізу помилок чистого ШІ, розроблено та формалізовано оптимальну методологію створення музичного аранжування. Ця методологія пропонується як стандарт для фахівців, що бажають інтегрувати ШІ у професійний пайплайн. Розроблена методологія базується на принципі «Протокол багаторівневої відповідальності» (Layered Responsibility Protocol), який структурно розподіляє виробничий процес на три ієрархічні рівні: первинне формування людиною детермінованого фундаменту (скелету) композиції, подальшу інтеграцію генеративного наповнення (текстур) засобами штучного інтелекту та остаточну детерміновану фіналізацію форми, яка знову повертається під повний контроль людини-оператора для забезпечення технічної відповідності стандартам [16, 250; 9, 399].

Деталізована схема Гібридного Пайплайну.

Етап 1. Структурне проєктування та формування детермінованого ядра (Human Logic & Foundation). Цей етап є зоною виключної відповідальності людини та детермінованих алгоритмів, метою якого є створення непорушного «скелета» композиції. Процес розпочинається у середовищі **Finale**, де розробляється мелодія, гармонічна сітка та форма твору; отриманий MIDI-файл слугує еталонною розміткою (Reference Grid), яка гарантує структурну цілісність і страхує від «галюцинацій» ШІ. Далі, у середовищі **Steinberg Cubase**, на цю

сітку нашаровується ритмічний та тональний фундамент із використанням високоякісних віртуальних інструментів: **Ueberschall** забезпечує ритмічну стабільність завдяки DSP-алгоритмам тайм-стретчингу, **Nexus** формує щільну басову лінію методами субтрактивного синтезу, а **Kontakt** реалізує основні гармонічні партії (рояль, саксофон). Результатом етапу є стабільний, ритмічно точний інструментал, створений під повним контролем оператора.

Етап 2. Генеративне наповнення та аугментація (AI Augmentation Domain). На цьому етапі до процесу підключаються хмарні сервіси штучного інтелекту (**Suno AI**, **AIVA**) для генерації допоміжних текстурних елементів, створення яких вручну є ресурсозатратним. Основний фокус зміщується на генерацію «сировини» (Raw Material), зокрема партій бек-вокалу, хору або специфічних шумових ефектів. Ключовою методологічною особливістю є відмова від генерації цілих композицій на користь створення коротких сегментів (Stems) із подальшою сепарацією корисного сигналу. Це дозволяє отримати унікальний тембральний контент, який неможливо синтезувати на звичайних VST-інструментах, зберігаючи при цьому загальну структуру, задану на першому етапі.

Етап 3. Інженерна інтеграція, корекція та технічна фіналізація (Human Engineering & Mastering). Фінальний та найбільш критичний етап повертає матеріал у зону відповідальності людини-інженера в середовищі Cubase. Тут відбувається «імплантація» згенерованого ШІ-контенту в детерміновану структуру: оператор виконує ручний Time-Warping для синхронізації ритмічно нестабільних аудіо-подій ШІ з сіткою такту та застосовує спектральну чистку (FabFilter Pro-Q 3) для видалення цифрових артефактів нейромережі. Процес завершується зведенням «живого» лід-вокалу з інструменталом та технічним мастерінгом (UAD Pultec, iZotope Ozone 9), що забезпечує стандартизацію гучності (LUFS) та динамічного діапазону відповідно до вимог індустрії, перетворюючи гібридний набір даних у завершений, професійний медіа-продукт.

Етап 4. Інженерна інтеграція, спектральна корекція та зведення (Human Engineering Domain). Цей етап є найбільш критичним моментом виробничого циклу, що відрізняє професійний продукт від аматорського експерименту, оскільки саме тут відбувається адаптація стохастичного матеріалу ШІ до жорстких вимог детермінованого середовища Steinberg Cubase. Процес розпочинається з ручної синхронізації (Manual Time-Warping), де оператор виправляє ритмічні неточності згенерованого аудіо, примусово прив'язуючи транзієнти ШІ-файлів до тактової сітки проєкту для забезпечення ідеального груву. Наступним кроком є «спектральна хірургія» за допомогою еквалайзера FabFilter Pro-Q 3, метою якої є видалення специфічних цифрових артефактів та резонансів, притаманних нейромережевій генерації, що дозволяє замаскувати штучне походження текстур. Завершується етап комплексним зведенням (Mixing), де «живий» лід-вокал, записаний з високою роздільною здатністю, інтегрується з детермінованими інструментами та обробленими ШІ-текстурами у єдиний акустичний простір, створюючи збалансовану частотну та динамічну картину.

Етап 5. Технічна фіналізація та стандартизація (Mastering). Фінальна стадія виробничого циклу присвячена виключно технічній стандартизації аудіо-продукту та контролю якості, що виконується на основі об'єктивних інженерних метрик без втручання генеративних моделей. Використовуючи гібридний ланцюг обробки, що поєднує аналогове моделювання (UAD Pultec) для гармонійного забарвлення та прецизійні цифрові алгоритми лімітування (iZotope Ozone 9), інженер забезпечує відповідність фонограми індустріальним стандартам гучності (LUFS). Основна мета етапу – досягнення максимальної щільності та прозорості звучання, запобігання міжсемпловим пікам та коректне перетворення (Dithering) у споживчий формат, гарантуючи, що гібридна природа аранжування буде сприйматися слухачем як цілісний, професійний музичний твір.

Проведений у цьому розділі синтез експериментальних результатів дозволяє сформулювати фундаментальні тези дослідження, що визначають

вектори подальшого розвитку музичного виробництва. Першочергово було експериментально доведено неможливість повної автоматизації професійного аранжування на базі сучасних генеративних моделей, таких як Suno v5. Це зумовлено їхньою системною нездатністю забезпечити критично важливу структурну точність та семантичну коректність: у поточному технологічному стані нейромережі функціонують виключно як стохастичні «генератори варіацій», а не як детерміновані виконавці авторського задуму.

Натомість максимальну ефективність штучний інтелект демонструє у ролі творчого «ко-процесора», спеціалізуючись на генерації допоміжних спектральних шарів, зокрема бек-вокальних гармоній та атмосферних текстур. У поєднанні з кваліфікованим ручним редагуванням такий підхід дозволяє досягти результату професійного рівня із суттєвою оптимізацією бюджету, нівелюючи потребу у залученні сесійних музикантів для второрядних партій. Це обґрунтовує домінування Гібридної моделі, формалізованої у попередніх підрозділах, як єдиного життєздатного шляху еволюції індустрії. Запропонований робочий процес органічно поєднує детерміновану надійність традиційних цифрових робочих станцій (DAW) із генеративною швидкістю нейромереж, при цьому залишаючи за людиною ключову функцію головного архітектора композиції та контролера фінальної якості продукту.

ВИСНОВКИ

У магістерській роботі вирішено актуальну науково-прикладну задачу підвищення ефективності та оптимізації процесів музичного виробництва шляхом комплексного порівняльного аналізу традиційних (евристичних) та новітніх (алгоритмічних) методів обробки мультимедійних даних. На основі проведеного теоретичного дослідження архітектур програмного забезпечення та серії практичних експериментів із моделювання робочих процесів, зроблено наступні узагальнюючі висновки.

1. Концептуалізація та формалізація конфлікту парадигм обробки даних. На системному рівні у роботі розмежовано дві фундаментально відмінні архітектури створення музичного контенту, що дозволило виявити глибинний конфлікт між рівнем контролю та автоматизацією. Традиційний підхід («людський ресурс») класифіковано як систему екзогенного керування, де джерелом прийняття рішень є зовнішній оператор, а програмне середовище (Finale, Steinberg Cubase) виконує функцію високоточного детермінованого інструментарію з лінійною логікою обробки даних. На противагу цьому, підхід штучного інтелекту («ресурс ШІ») визначено як ендогенну стохастичну систему, де генеративна нейромережа автономно створює контент на основі прихованих шарів та ймовірнісних розподілів. Аналіз довів, що головна інженерна дилема полягає в оберненій пропорційності: детерміновані алгоритми забезпечують 100% передбачуваність та точність реалізації авторського задуму ціною значних часових витрат, тоді як стохастичні алгоритми ШІ забезпечують миттєву генерацію результату, але ціною повної втрати керованості та неможливості точного редагування окремих компонентів.

2. Емпірична верифікація технологічних обмежень генеративних моделей (End-to-End). В ході експериментального моделювання (Сценарій Б) було виявлено, систематизовано та технічно обґрунтовано критичні обмеження сучасних аудіо-моделей (на прикладі платформи Suno AI v5), які на поточному етапі унеможливлюють їх автономне використання у професійному продакшні.

Зокрема, зафіксовано явище «структурних галюцинацій», коли нейромережа ігнорує задані користувачем обмеження форми та гармонічної сітки, довільно змінюючи структуру композиції. Спектральний аналіз вихідних файлів виявив проблему низької роздільної здатності сигналу, що характеризується наявністю артефактів нейронної компресії та відсутністю корисної інформації у високочастотному спектрі (понад 16 кГц). Крім того, підтверджено архітектурну проблему «Чорної скриньки»: готові сервіси генерують зведений аудіо-файл («flat mix»), що технічно блокує можливість декомпозиції на окремі доріжки (Stems) з якістю, достатньою для професійного мікшування.

3. Валідація ефективності детермінованої «Еталонної моделі» (Human-in-the-Loop). Детальна декомпозиція робочого процесу у середовищі Steinberg Cubase експериментально підтвердила безальтернативність використання детермінованих алгоритмів для побудови фундаменту музичної композиції. Доведено, що використання спеціалізованих DSP-інструментів – таких як алгоритми субтрактивного синтезу (у плагіні Nexus) для басових партій, алгоритми тайм-стретчингу (Ueberschall/zplane) для ритмічної синхронізації та алгоритми FFT-корекції (Auto-Tune) для вокалу – забезпечує необхідну інженерну точність та повторюваність результату. Саме можливість прямого «людського» втручання на рівні параметрів синтезу (огинаючі ADSR, параметри фільтрів, автоматизація) дозволяє адаптувати звучання під унікальний художній контекст та семантику твору, чого на даний момент принципово не можуть зробити усереднені статистичні моделі ШІ, які оперують узагальненими патернами.

4. Обґрунтування та розробка Гібридної моделі (Layered Responsibility Workflow). Головним науково-практичним результатом роботи є розробка та формалізація методології «Протокол багаторівневої відповідальності», яка вирішує проблему ефективної інтеграції стохастичного ШІ у лінійний інженерний процес. Запропонована модель структурно розподіляє виробничі задачі: людина зберігає повну відповідальність за створення «скелету» (структура, мелодія, лід-вокал), що гарантує якість, тоді як ШІ делегується

задача аугментації та текстурування (генерація бек-вокалу, хорових пачок, фонових гармоній), що дозволяє обійти ресурсні обмеження. Експериментально доведено, що впровадження такого підходу дозволяє скоротити загальний час виробничого циклу на 30-40% (за рахунок елімінації етапів пошуку та запису сесійних музикантів для допоміжних партій), зберігаючи при цьому повний контроль над художньою цілісністю основного музичного матеріалу.

5. Визначення критичної ролі інженерної пост-обробки та інтеграції. Дослідження спростувало поширений міф про повну автоматизацію творчості «в один клік», встановивши, що сирий результат роботи генеративних моделей є лише цифровим «напівфабрикатом», який потребує глибокої інженерної інтеграції для досягнення комерційної якості. Доведено необхідність обов'язкового застосування процедур Manual Time-Warping (для ручного виправлення ритмічних похибок генерації), спектрального маскування (EQ-фільтрація цифрових артефактів) та просторової інтеграції (Reverb) у середовищі DAW. Тільки за умови кваліфікованого втручання людини-інженера, яка виступає в ролі «валідатора» та «коректора», стохастичний та часто недосконалий контент ШІ може бути адаптований та доведений до рівня сучасних індустріальних стандартів.

6. Прогностична оцінка розвитку технологій ко-творчості. Аналіз тенденцій розвитку програмного забезпечення дозволяє стверджувати, що майбутнє музичної цифрової індустрії лежить не в площині повної заміни людини алгоритмами, а в еволюції систем ко-творчості (Human-AI Co-Creation). Перспективним вектором розвитку визначено відмову від автономних хмарних генераторів на користь глибокої інтеграції генеративних моделей безпосередньо в архітектуру професійних DAW у вигляді локальних VST-плагінів або вбудованих модулів. Це дозволить поєднати генеративну потужність нейромереж із детермінованим контролем, прозорістю та зручністю професійного інтерфейсу, перетворивши ШІ з «конкурента» на потужний допоміжний інструмент аранжувальника.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

7. Афоніна О. С. Історія та теорія музичних технологій : навч. посібник. Київ : НАКККіМ, 2018. 156 с.
8. Бондаренко А. І. Інформатика музична : навчальний посібник. Київ : НАКККіМ, 2020. 216 с.
9. Власюк А. П. Комп'ютерні технології в музичній творчості : навч.-метод. посіб. Рівне : РДГУ, 2019. 142 с.
10. Голощук О. О. Вплив музично-інформаційних технологій на розвиток сучасного музичного мистецтва. *Актуальні проблеми розвитку українського та зарубіжного мистецтв: культурологічний, мистецтвознавчий, педагогічний аспекти : матеріали VIII Міжнародної науково-практичної конференції*. Львів–Торунь : Liha-Pres, 2023. С. 352–356.
11. Голощук О. О. Значення сучасних музично-інформаційних технологій в процесі фахової мистецької освіти. *Сучасні дослідження світової науки : матеріали наук.-практ. конф.* 2022. С. 969–971.
12. Голощук О. О. Комп'ютерне аранжування та звукорежисура : методичні рекомендації до курсу. Луцьк : ВНУ ім. Лесі Українки, 2023. 48 с.
13. Голощук О. О. Методика роботи в музичному редакторі Steinberg Cubase : методичні рекомендації для студентів спеціальності 025 «Музичне мистецтво». Луцьк : ВНУ ім. Лесі Українки, 2022. 45 с.
14. Голощук О., Сорока Р., Шкоба В. Монтаж та обробка звуку в музичному редакторі AVID Technology Pro Tools 12 : методичні рекомендації. Луцьк : ВНУ ім. Лесі Українки, 2022. 21 с.
15. Голощук О. О. Розвиток музично-інформаційних технологій в умовах цифровізації освіти. *Актуальні проблеми розвитку природничих та гуманітарних наук : збірник матеріалів VI Міжнар. наук.-практ. конф*. Луцьк : ВНУ ім. Лесі Українки, 2022. С. 399–400.
16. Гуменюк Т. К. Музична культура в добу цифрових технологій : монографія. Київ : КНУКіМ, 2017. 288 с.

17. Котляревський І. А. Основи синтезу звуку : підручник. Харків : ХНУМ, 2016. 204 с.
18. Кривопуст Б. Основи оркестровки : навчальний посібник. Київ : Нова Книга, 2015. 216 с.
19. Лунченко О. В. Особливості застосування штучного інтелекту в музичному мистецтві. Цифрові технології в культурі та мистецтві. 2023. № 4. С. 45–52.
20. Петров В. В. Методи та засоби комп'ютерної обробки звукових сигналів : монографія. Вінниця : ВНТУ, 2012. 248 с.
21. Рибалко О. В. Основи звукорежисури та аранжування : навч. посібник. Київ : Ліра-К, 2021. 256 с.
22. Скиба М. В. Інформаційні технології в музичній освіті: теорія і методика : монографія. Кривий Ріг : КДПУ, 2018. 312 с.
23. Фадєєва К. В. Комп'ютерні технології в музичній освіті : навч.-метод. посібник. Київ : НПУ імені М.П. Драгоманова, 2019. 134 с.
24. Шульгіна В. Д. Музичне мистецтво в умовах глобалізації та інформатизації суспільства // Вісник НАКККіМ. 2015. № 2. С. 83–88.
25. Agostinelli A. et al. MusicLM: Generating Music From Text [Electronic resource]. Denk, Z. Borsos // Google Research. 2023. arXiv:2301.11325. URL: https://arxiv.org/abs/2301.11325 (15.10.2025).
26. Briot J.-P. Deep Learning Techniques for Music Generation. Cham : Springer Nature, 2020. 294 p.
27. Celemony Software GmbH. Melodyne 5: User Manual [Electronic resource]. Munich : Celemony, 2020. URL: https://help.celemony.com/M5/en/ (11.11.2025).
28. Copet J. et al. Simple and Controllable Music G eneration (MusicGen) [Electronic resource]. Meta AI Research. 2023. arXiv:2306.05284. URL: https://arxiv.org/abs/2306.05284 (18.10.2025).
29. Dhariwal P. et al. Jukebox: A Generative Model for Music [Electronic resource]. OpenAI. 2020. arXiv:2005.00341. URL: https://arxiv.org/abs/2005.00341 (22.09.2025).

30. European Broadcasting Union. EBU R 128: Loudness Normalisation and Permitted Maximum Level of Audio Signals. Geneva : EBU, 2020.
31. FabFilter Software Instruments. FabFilter Pro-Q 3 User Manual [Electronic resource]. Amsterdam : FabFilter, 2023. URL: https://www.fabfilter.com/help/pro-q (05.11.2025).
32. Goodfellow I. et al. Generative Adversarial Networks // Advances in Neural Information Processing Systems. 2014. Vol. 27. P. 2672–2680.
33. Huang C.-Z. A. et al. Music Transformer: Generating Music with Long-Term Structure // International Conference on Learning Representations (ICLR). 2019.
34. Huber D. M. Modern Recording Techniques. 9th ed. New York : Routledge, 2017. 670 p.
35. iZotope, Inc. Ozone 9: Mastering Guide and Reference Manual [Electronic resource]. Cambridge, MA : iZotope, 2021. URL: https://docs.izotope.com/ozone9/en/index.html (12.11.2025).
36. Izhaki R. Mixing Audio: Concepts, Practices, and Tools. 3rd ed. London : Focal Press, 2017. 600 p.
37. Katz B. Mastering Audio: The Art and the Science. 3rd ed. Burlington : Focal Press, 2013. 408 p.
38. Kong Z. et al. DiffWave: A Versatile Diffusion Model for Audio Synthesis // International Conference on Learning Representations (ICLR). 2021.
39. MakeMusic, Inc. Finale User Manual: Windows & Macintosh [Electronic resource]. Colorado : MakeMusic, 2022. URL: https://usermanuals.finalemusic.com/ (25.09.2025).
40. Müller M. Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications. Cham : Springer, 2015. 487 p.
41. Native Instruments GmbH. Kontakt 7 Reference Manual [Electronic resource]. Berlin : Native Instruments, 2023. URL: https://www.native-instruments.com/ni-tech-manuals/kontakt-7-manual/en/ (30.10.2025).
42. Oppenheim A. V., Schafer R. W. Discrete-Time Signal Processing. 3rd ed. Pearson, 2010. 1120 p.

43. Owsinski B. The Mixing Engineer’s Handbook. 5th ed. Burbank : Bobby Owsinski Media Group, 2022. 312 p.
44. Purwins H. et al. Deep Learning for Audio Signal Processing // IEEE Journal of Selected Topics in Signal Processing. 2019. Vol. 13, No. 2. P. 206–219.
45. reFX Audio Software Inc. Nexus 4 User Manual [Electronic resource]. – Vancouver : reFX, 2022. URL: https://refx.com/support/ (02.11.2025).
46. Roads C. The Computer Music Tutorial. 2nd ed. Cambridge, MA : MIT Press, 2023. 1200 p.
47. Sethares W. A. Tuning, Timbre, Spectrum, Scale. 2nd ed. London : Springer, 2005. 426 p.
48. Smith J. O. Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects. W3K Publishing, 2010.
49. Steinberg Media Technologies GmbH. Cubase Pro 13: Operation Manual [Electronic resource]. Hamburg : Steinberg, 2023. URL: https://steinberg.help/cubase_pro/v13/en/ (15.11.2025).
50. Suno AI. Suno v3 Model Card [Electronic resource] // Suno Research Blog. – 2024. URL: https://suno.com/blog/v3 (дата звернення: 20.11.2025).
51. Ueberschall GmbH. Elastik 3 Player: User Manual [Electronic resource]. Hannover : Ueberschall, 2022. URL: https://www.ueberschall.com/elastik (дата звернення: 08.10.2025).
52. Van den Oord A. et al. WaveNet: A Generative Model for Raw Audio [Electronic resource] // arXiv preprint arXiv:1609.03499. 2016. URL: https://arxiv.org/abs/1609.03499 (14.09.2025).
53. Vaswani A. et al. Attention Is All You Need // Advances in Neural Information Processing Systems (NIPS). 2017. Vol. 30. P. 5998–6008.
54. Välimäki V., Reiss J. D. All About Audio Equalization: Solutions and Frontiers // Applied Sciences. 2016. Vol. 6, No. 5. P. 129.
55. Watkinson J. The Art of Digital Audio. 3rd ed. Focal Press, 2001. 760 p.
56. Zölzer U. DAFX: Digital Audio Effects. 2nd ed. Chichester : John Wiley & Sons, 2011. 624 p.

АНОТАЦІЯ

Голощук О. О. Методологія створення цифрових музичних продуктів з використанням технологій генеративного штучного інтелекту. Кваліфікаційна наукова праця на правах рукопису. Робота на здобуття другого (магістерського) рівня зі спеціальності «122 Комп'ютерні науки». Волинський національний університет імені Лесі Українки Міністерства освіти і науки України, Луцьк, 2025.

Магістерська робота присвячена комплексному порівняльному аналізу двох фундаментально відмінних підходів до створення музичного аранжування: традиційного евристичного (детермінованого), де людина-оператор використовує програмне забезпечення (Finale, Cubase) як високоточний інструментарій, та новітнього алгоритмічного (стохастичного), що базується на використанні генеративних моделей штучного інтелекту (AIVA, Soundraw, Suno AI).

У дослідженні формалізовано конфлікт між цими парадигмами, зокрема між повним контролем даних у середовищі DAW та швидкістю ендогенної генерації нейромереж. На основі експериментальних даних виявлено технологічні обмеження сучасних End-to-End моделей та обґрунтовано необхідність впровадження гібридної моделі робочого процесу, яка поєднує структурну надійність людського підходу з текстурною аугментацією засобами ШІ.

У роботі формалізовано конфлікт парадигм між екзогенним керуванням у середовищі DAW та ендогенною генерацією нейромереж. Проведено декомпозицію еталонної моделі робочого процесу в середовищі Steinberg Cubase з використанням спеціалізованих DSP-алгоритмів (тайм-стретчинг, субтрактивний синтез, спектральна корекція). Експериментальним шляхом виявлено технологічні обмеження сучасних End-to-End аудіо-моделей, зокрема проблеми структурних галюцинацій та низької роздільної здатності сигналу.

Головним практичним результатом дослідження є розробка та обґрунтування «Гібридної моделі робочого процесу» (Protocol of Layered

Responsibility). Запропонована методологія передбачає розподіл задач: створення структурного фундаменту людиною та використання ШІ для аугментації і текстурування (наприклад, генерації бек-вокалу) з обов'язковою інженерною інтеграцією. Доведено, що такий підхід дозволяє оптимізувати часові ресурси на 30–40% при збереженні повного контролю над художньою якістю продукту.

Ключові слова: музично-комп'ютерні технології, цифрова обробка сигналів, гібридна модель, музичне аранжування, штучний інтелект, евристичний підхід, алгоритмічна генерація, DAW, Cubase, Suno AI.

ABSTRACT

Holoshchuk O. O. Methodology for creating digital music products using generative artificial intelligence technologies. Qualification thesis in manuscript form. Work for obtaining the second (master's) level in the specialty «112 Computer Science». Lesya Ukrainka Volyn National University of the Ministry of Education and Science of Ukraine, Lutsk, 2025.

The master's thesis is dedicated to a comprehensive comparative analysis of two fundamentally different approaches to musical arrangement: the traditional heuristic (deterministic) approach, where a human operator uses software (Finale, Cubase) as a high-precision toolkit, and the modern algorithmic (stochastic) approach based on the use of generative artificial intelligence models (AIVA, Soundraw, Suno AI).

The study formalizes the conflict between these paradigms, particularly between full data control in a DAW environment and the speed of endogenous neural network generation. Based on experimental data, the technological limitations of modern End-to-End models are identified, and the necessity of implementing a hybrid workflow model that combines the structural reliability of the human approach with textural augmentation via AI means is substantiated.

The work formalizes the paradigm conflict between exogenous control in the DAW environment and endogenous generation by neural networks. A decomposition of the reference workflow model in Steinberg Cubase was performed using specialized DSP algorithms (time-stretching, subtractive synthesis, spectral correction). The technological limitations of modern End-to-End audio models, specifically structural hallucinations and low signal resolution, were identified experimentally.

The main practical result of the research is the development and substantiation of the «Hybrid Workflow Model» (Protocol of Layered Responsibility). The proposed methodology involves task distribution: the creation of a structural foundation by a human and the use of AI for augmentation and texturing (e.g., backing vocal generation) with mandatory engineering integration. It is proven that this approach

allows optimizing time resources by 30–40% while maintaining full control over the artistic quality of the product.

Keywords: music computer technologies, digital signal processing, hybrid model, musical arrangement, artificial intelligence, heuristic approach, algorithmic generation, DAW, Cubase, Suno AI.