NOVA IMS Information Management School

MGI

Master Degree Program in

Information Management

Collaborative Creativity: Exploring AI's roles in Music Production

A systemic approach for effective integration of generative AI into the creative process

João Pedro Mendes Ribeiro Pechirra do Atalho

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

Collaborative Creativity: Exploring AI's roles in Music Production

A systemic approach for effective integration of generative AI into the creative process

by

João Pedro Mendes Ribeiro Pechirra do Atalho

Master Thesis presented as partial requirement for obtaining the master's degree in information management, with a specialization in Information Systems

Supervised by

Vítor Manuel Pereira Duarte dos Santos, PhD, NOVA Information Management School

July 2025

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, July 2025
João Atalho

DEDICATION

To my loving parents Ângela Ribeiro and António Pechirra, for all the support and belief in my potential.
To my brother Nuno Atalho for the guiding life advice.
To my girlfriend, Daniela Rotaru for being my rock, daily company and inspiration to move forward.
To my grandparents that will forever live in my heart and memory.
Although you may never read this, I love you all dearly.

ACKNOWLEDGEMENTS

I want to thank and appreciate the efforts of my supervisor prof. Vítor Duarte dos Santos in guiding this project and giving direction when uncertainty took over.
To the participants for being a central part of this experiment.
To all the developers that provided the tools used in the development of this work.
It wouldn't have been possible without you.

ABSTRACT

Artificial Intelligence is impacting the arts as it is in virtually every industry. Creatives are constantly presented with novel methods and tools to aid them in their workflows. This research examines the impact of AI in music creation, focusing on building a framework that integrates AI models into digital audio workflows, aiming to bridge the gap between technological innovation and artistic authenticity. Despite AI's immense potential, there is a notable aversion among many artists and consumers to AI-generated material. There is a critical need to investigate methods of integrating AI into artistic workflows in ways that complement human creativity without replacing it. The primary objective of this study is the development of a solution aimed at embedding AI into Digital Audio Workstations in the form of a virtual assistant. The prototype can assume specific assisting roles based on the user's needs. The research methodology is divided into four phases: Preparation, Conception, Execution, and Conclusion. The study demonstrates that the developed solution, LTN-DAN, alongside other AI tools such as *Google's Magenta-Studio*, can potentially enhance collaboration, productivity, and creativity in music composition and production without hindering authenticity. The research identified various roles for AI in production, different generative tools available for musicians and relevant architectures for developers of music software. While there is still some uneasiness among artists, the evaluation revealed a recognized helpfulness and potential for sparking creativity through AI-assisted tools. These contributions aim to bridge the gap between AI capabilities and creatives, supporting effective adoptions of AI in music.

KEYWORDS

Generative Artificial Intelligence; Music; Arts; Creativity; Personal Assistance; Automation; Human-Computer Interaction

Sustainable Development Goals (SGD):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

API	Application Programming Interface
AI	Artificial Intelligence
BERT	Bi-directional Encoder Representations from Transformers
CGAN	Conditional Generative Adversarial Network
CD	Compact Disc
CNN	Convolutional Neural Network
CSS	Cascading Style Sheets
CVAE	Convolutional Variational Auto-Encoder
DAW	Digital Audio Workstation
DOM	Document Object Model
GAI	Generative Artificial Intelligence
GAN	Generative Adversarial Network
GPT	Generative Pre-trained Transformer
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IT	Information Technology
LAN	Local Area Network
LALM	Large Audio Language Model
LLM	Large Language Model
LSTM	Long Short-Term Memory
ML	Machine Learning
MIDI	Musical Instrument Digital Interface
PC	Personal Computer
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RNN	Recurrent Neural Network
RQ	Review Question
SOA	State of the Art
SLR	Systematic Literature Review
SLRQ	Systematic Literature Review Question
STT	Speech-To-Text
UI/UX	User Interface/User Experience
VAE	Variational Auto-Encoder
VST	Virtual Studio Technology

1 INTRODUCTION

1.1 CONTEXT

Artificial Intelligence (AI) has emerged as a transformative force with valuable applications across all industries, reshaping how professionals work and create. As a general-purpose technology, such as money or the internet, AI embeds itself into people's daily practices and redefines societal norms (Mollick, 2024). Despite the initial concerns about displacing the intellectual workforce, Generative AI's reliance on user's input for most scenarios entails a collaborative dynamic and negates the hypothesis of a complete replacement of workers for most areas. Although some enterprises have downsized their human resources since the mainstream advent of LLMs in November 2021 with the launch of *ChatGPT*, what is far more common is the strategic reallocation of human capital toward higher value tasks as AI aids in repetitive actions.

The role of generative technologies is particularly contentious in the arts, sparking debate about authorship, originality, and the philosophical essence of creativity. While some view AI as a threat to artistic integrity, others embrace it as a tool to assist them in their workflows finding it possible to do so without undermining creative vision. Visual artists and enthusiasts are experimenting with generated imagery, using algorithms to assemble unique pieces. In music, AI systems can offer generated compositions that mimic the style of famous composers (Hadjeres et al., 2017), are able to identify tracks based on samples and of various other features. The development of language models capable of producing coherent and at times compelling narratives (Ippolito et al., 2022) has had major impact in literature and screen-writing. This new genre of technology has become a disruptive agent in all forms of media. AI successfully emulates creativity or at least, produces seemingly creative output even though for many consumers the results are still noticeably sub-par when comparing with human-made compositions; and more importantly missing out on the principal component of creativity: Ideation – AI relies on the user to prompt with an idea for it to generate an output whether textual, visual or audible.

Some artists embrace AI as a collaborative partner, viewing it as a means to accelerate their output and explore new artistic territories. Others express concern about the potential homogenization of art and the devaluation of creativity (Anderson et al., 2024). Despite these discordances, AI continues to make inroads in the creative industries, influencing everything from visual effects to interactive media installations. This technological infusion is reshaping not only how art is created but also how it is experienced and consumed.

While there is notable aversion among creators and consumers (Ragot et al., 2020) to AI-generated work, that sentiment cannot persist when its' use remains imperceptible. This presents an interesting paradox and potential area for exploration: **How can creators benefit from AI's capabilities to enhance their process while maintaining the perceived authenticity**

and human touch that audiences value? There is a need to investigate methods of integrating AI into artistic workflows in ways that complement rather than replace human creativity. Additionally, exploring how to communicate the role of AI in art creation to audiences in a manner that highlights its benefits without immediately triggering skepticism could help bridge the gap between technological advancement and its acceptance in the arts.

With the insurgence of AI in late 2021 and even slightly before came a swarm of pessimists wagering how AI would take over jobs, undermine artistic output and invert creativity. The public opinion surrounding AI was polarized, some people would express apprehension while others displayed excitement toward it. ML algorithms, particularly LLMs are already employed in most artistic fields and consequently some consumers have begun to notice an eager similarity between plots and dialogues of recent TV shows and film (Corrêa, 2023), which may be a result of screenwriters using generative AI tools. Most enthusiasts don't reject the potential negative effects but rather concentrate on how AI can democratize access to non-experts and hobbyists similarly to what the invention of Photography did to image-making. What initially seemed as a strictly mechanical and technical process ended up as a breadth of fresh air into an otherwise outdated art-form (Hertzmann, 2018).

This dissertation examines the impact of AI in music creation, focusing on building a framework that integrates AI models into digital audio workflows. Through experiments and interviews with artists, this study explores the receptiveness to AI-assisted tools and evaluates their creative outcomes. By investigating how AI can complement human creativity and the various roles it can play in the musical process, this research aims to bridge the gap between technological innovation and artistic authenticity, offering insights into the evolving relationship between AI and music creation.

1.2 Problem Identification

From a preliminary standpoint the following questions regarding the practical use of AI in Music Production were identified:

RQ1: How can the use of AI components enhance collaboration, productivity, and creativity in music production without hindering authenticity?

RQ2: What roles can AI assume in the music production process (e.g., co-creator, assistant, advisor), and how the identified roles influence creative agency, workflow efficiency, and artistic decision-making of music producers?

RQ3: To which extent are musicians receptive to working collaboratively with AI?

1.3 OBJECTIVES

The primary focus of this study is the development of a solution aimed at combining AI into a Digital Audio Workstation (DAW) through an assistant comprised of a large language model (LLM) and a speech-to-text model (STT). This framework can assume diverse assisting roles based on the end user's needs enhance collaboration, productivity, and creativity in music composition.

The underlying research is structured around several interconnected objectives. Initially, exploratory and systematic literature reviews will be conducted to assess the current state of AI integration in music production, identifying trends, gaps, and opportunities. This will inform the investigation of technical requirements and challenges involved in embedding LLMs and AI into DAWs, alongside delineating potential AI roles—such as co-creator, assistant, facilitator, teacher or advisor—within the creative process.

The study will explore practical interactions between music producers and AI tools across key production stages (e.g., ideation, composition, arrangement, production, etc.), using these insights to design a framework prioritizing usability, efficiency, and creative synergy. A functional prototype will be implemented, demonstrating LLMs' versatility in creative roles and supporting production workflows. This prototype will be evaluated through user experiments and interviews, assessing its impact on creative agency, workflow efficiency, and artistic decision-making, while also examining implications for originality, artistic style, and authorship in AI-assisted music production.

Finally, the research collects the state of the art in designing AI-integrated tools, focusing on the ones that preserve artistic integrity, culminating in a standardized, scalable framework tailored to diverse production scenarios and user expertise levels. These will be refined through feedback from practitioners and experts, ensuring broader applicability and ethical considerations in the creative industry.

1.4 STUDY OUTCOMES AND RELEVANCE

This research aims to produce both theoretical insights and practical innovations in the integration of AI into music production. Through comprehensive literature reviews, it will map current AI tools and identify emerging trends, culminating in a synthesized framework for creative workflows involving LLMs.

The technical investigation will analyze key challenges such as latency, interface usability, and system integration, proposing actionable solutions and identifying the necessary resources for effective implementation. A taxonomy of AI roles within music production will be developed, supported by a conceptual model of human-AI collaboration.

Empirical case studies will demonstrate how AI contributes to specific production tasks, while the development of a functional prototype will showcase the creative capabilities of LLMs in real-world scenarios. This prototype will be evaluated through mixed methods testing, generating qualitative and quantitative feedback on its usability and productivity benefits.

Furthermore, the study will explore how AI impacts creative agency and workflow dynamics, offering insight into shifts in artistic decision-making and autonomy. Ethical considerations—particularly around originality, authorship, and creative ownership—will be critically examined, accompanied by recommendations for responsible and ethical AI integration.

1.5 Overall Contributions

This study makes several key contributions to both academic research and professional practice in music production and AI:

- A publicly available prototype, hosted on *GitHub*, demonstrating how LLMs can support creative music production workflows.

- A taxonomy of AI roles, clearly defining how language models can function within music production environments—from ideation to arrangement and sound design.

- A curated compendium of open-source AI tools relevant to music creation, serving as a practical resource for musicians and producers.

- A qualitative user study, including experiments and interviews with participants of varying musical backgrounds, offering rich insights into how users interact with AI-assisted tools.

These contributions aim to bridge the gap between AI capabilities and creative needs, supporting a more informed, ethical, and effective adoption of AI in music production.

2 LITERATURE REVIEW

This chapter examines the state of the art of AI in Music Production. Primarily, an exploratory review will be conducted attempting to respond the established research questions. Lastly, the insights extracted from this effort will be confirmed and re-evaluated in a systematic literature review. On the exploratory phase, emphasis will be placed on contextualizing the cultural and historical background of creativity and music production while also examining the evolution and applications of AI in the domain.

To ensure a comprehensive synthesis of the existing literature surrounding the topic, the *PRISMA* (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology will be employed, providing a structured and visual approach to identifying and extracting relevant research. The review process, like most others begins with the formulation of well-defined systematic literature review questions (SLRQs) that may or may not coincide with the Research Questions presented earlier. This is justified for methodological reasons: the objective of the SLRQ's is to synthesize literature while the RQ's are more in line with the project's outcomes. The research will be then conducted systematically and papers analysed with the intent of answering the review questions. Finally, conclusions and findings will be presented. This process will be further described in its' corresponding sub-chapter.

2.1 MUSIC PRODUCTION AND CREATIVE PROCESS

2.1.1 The Creative Process

Creativity is a fundamental feature of intelligence (Boden, 1998). It may be mostly associated with the arts, but it extends into the very basis of personal identity, affecting how people approach commonplace problems. Generically defined as the actualizing of one's potential (Young, 1985) or more technically as a production of ideas that are novel and appropriate in any domain of human activity (Amabile, 2012). This ability requires a confluence of interrelated capacities: intellect, knowledge, styles of thinking, personality, motivation and environment. The absence or deficiency of any of these resources hinders the creative process and naturally the respective outcomes. There may be some compensation from one component to another but there usually exists a threshold of quality for each below which it is not possible to act creatively (Sternberg, 2006).

The creative process is often described as a dynamic and iterative series of stages. According to (Wallas, 1926), the process follows a four-stage model: preparation, the sourcing of necessary assets; incubation, where ideas are synthesized; illumination - the moment of breakthrough; and verification, where the idea is further explored and tested. More contemporary models emphasize the interaction between the individual, domain of expertise, and surrounding environment (Moneta & Csikszentmihalyi, 1996). These perspectives

highlight the importance of external influence, such as feedback from peers and cultural expectations, but also the meandering of ideas, suggesting a non-linearity to the process. Additionally, creativity often involves both divergent and convergent thinking. The first generates multiple possibilities. The latter selects the most feasible or valuable ideas (Guilford & Smith, 1959). These cognitive processes underscore the complexity of creativity as a skill that integrates both structured and unstructured approaches to problem-solving.

Music production is an adaptable and dynamic art form at the center of this study. As a creative process it results from the interplay of imagination, skill, and technology. Some authors theorize that music originated as a means of gathering the attention of the opposite sex after which language was derived from (Darwin, 1871). The arts represent intent on the unending quest for something new (Hertzmann, 2018).This novelty that Hertzman posits is the direct result of the creative effort. If it is not new it may still be considered art but to a lesser creative degree.

2.1.2 Music Production

Music and sound recording have been wholly impacted by technological opportunity and constraint. (Katz, 2010) A song represents not only the artist's expression but also the extent of the period's technological capabilities. From the rudimentary mechanical devices of the 19th century to today's sophisticated digital environments, music production reflects an ongoing symbiosis between artistic creativity and technological breakthrough.

Music production traces its origins far beyond mechanical sound recording though that invention transformed music from a mostly ephemeral art form into a directly reproducible and commercial medium. Thomas Edison's invention of the phonograph in 1877 allowed for the first sound recordings by etching vibrations onto wax cylinders (Feaster, 2012). The phonograph, later replaced by Berliner's gramophone, marked the birth of recorded music, enabling playback on flat discs with higher fidelity (Sterne, 2003). Such innovations allowed for archiving and mass distribution of music records and performances, adding a new dimension to the exchange between musicians and their audiences, which previously relied solely on live performances and music sheets and was limited by geographic space, and the necessity for musicians and readers to properly interpret pieces.

Despite their innovation, early mechanical recording devices had significant limitations, including a restricted frequency range, susceptibility to wear and required specific recording conditions. Nevertheless, this invention laid the groundwork for the analog technologies that followed. The analog era began with the adoption of magnetic tape recording, a German innovation from the 1930s that gained global prominence after World War II (Daniel et al., 1998). Magnetic tape offered superior fidelity and introduced new possibilities for edition and

manipulation of records fostering new production techniques. *Les Paul's* invention of multitrack recording systems in the 1950s allowed separate capture of individual instruments, facilitating intricate arrangements and complex asynchronous composition (Ryan & Kehew, 2006). These innovations redefined recording studio as a creative instrument rather than a mere documentation tool.

The transition to digital technologies during the 80's and 90's marked one of the most significant technological shifts in music production. Digital audio offered precision and flexibility far beyond analog capabilities. The compact disc (CD), introduced in 1982, became the first mainstream digital audio format, providing superior durability and audio fidelity compared to vinyl or tape recorded audio (Moorefield, 2005). Concurrently, Digital Audio Workstations (DAWs) emerged as the central tool for music and sound design. *FL Studio, Ableton* or *Cubase*, etc. are examples of software that combines mixers, multi-track recorders, effects units and much more into a single interactive digital bundle. These allowed artists to record, edit, and mix in full digital form, further democratizing music production (Reuter, 2022) by reducing the need for expensive analog hardware and audio gear.

The introduction of MIDI (Musical Instrument Digital Interface) in 1983 further revolutionized production by enabling communication between digital instruments and computers (Holmes, 2008). MIDI provides the foundation for communication between musical devices and is central the rise of electronic music and digital production. The adoption of this standard triggered the advancement of music technology leading to the development of more sophisticated DAWs, virtual instruments, and software-based production tools. The elevated control in real-time impacted not only production but also the performative aspect, and allowed for the introduction of sampling a technique that enabled the reuse of pre-existing audio materials, which became a cornerstone for hip-hop and electronic music (Demers, 2010).

Modern music production combines analog and digital approaches, merging the warmth of analog sound with the precision of digital tools. Most studios still make use of analog equipment, though digital tools have most preponderance, from digital plugins emulating analog hardware to advanced DAWs. Many producers still find that the digital counterparts to legacy hardware lack-luster, arguing that although they bring higher sound quality and fidelity, there is a musical warmth to analog hardware unmatched by software emulations.

2.1.3 AI in Music Production

The application of AI in music extends beyond the aspect of production, the utilization of machine learning algorithms in streaming services for recommender systems or song identification through audio recognition are two frequent examples of AI in media consumption. These use cases enable personalized music consumption by analyzing listener preferences and tailoring compositions to specific audiences and the rapid collection of information regarding new music (Guo, 2023). *Endel* is an example of software that uses AI to create adaptive soundscapes for activities such as meditation or exercise, generating music in real-time based on environmental or user input. These applications demonstrate how AI can merge creativity with data-driven insights to recommend or create music that resonates with specific contexts and preferences. The focus of this study, however, is in music production and thus, while these examples are compelling, they are outside the scope of the present analysis. This segment holds the most relevance to the research as the prior sections provide the historical contextualization into this one. This chapter will be structured into two distinct segments. The first one will examine AI software and algorithms that support professionals in music and sound design. The subsequent segment will investigate the application of LLMs, focusing on their implementation within role-based configurations.

AI's ability to generate songs, stems or arrangements is progressing rapidly. Algorithms analyze vast datasets of existing music to generate melodies, harmonies, and even complete compositions. Tools like *OpenAI's MuseNet* and *Google's Magenta-Studio* use machine learning to create stylistically diverse pieces by training on extensive libraries of music (Sturm et al., 2019) attempting to tie machine-learning workflows with *Ableton Live*, one of the most popular DAW, *Google* researchers provide a suite containing five plug-ins that exemplify three distinct musical tasks: composition, interpretation and accompaniment in a user-friendly interface (Roberts et al., 2019). Some of the plugins rely on user's melodies, chord progressions or drum grooves to generate new compositions while others generate based on the trained data and parameters controlled by the user – length, tempo and temperature.

MuseGAN (Dong et al., 2018), as the name suggests, is an example of a Generative Adversarial Network (GAN) (Goodfellow et al., 2014) solution applied for music creation - a system composed of two algorithms, a generator and a discriminator. The first is trained to deceive the latter which is trained to distinguish real from generated data by generating outputs that resemble the real data used to train the discriminator. The outputs from *MuseGAN* are multi-track polyphonic music with temporal, harmonic and rhythmic structure. The program can generate with or without user input. The research underlines the challenges of generating music in contrast with image or text since music is time-dependent and holds a strict hierarchical structure: songs divide into paragraphs that divide into phrases then bars and beats which are fundamentally an arrangement of pixels. The program presents three different creation styles - Jamming, Composing and Hybrid - These models mimic the way musicians collaborate either by improvising, following along a composed piece or a mixture of

the two. The authors designed metrics to evaluate the generated excerpts to convey an objective evaluation of the model outputs, such as ratio of empty bars, number of used pitch classes per bar, drum pattern or tonal distance. All of these are an effort to measure musicality which is subjective, and thus not directly quantifiable.

AI's ability to analyze audio data in real-time has led to innovations in mixing and mastering. plugins like *iZotope's Ozone* or platforms like *LANDR* allow for the automatic mastering of audio tracks offering suggestions for effects like equalization, compression, and reverb settings based on specified genres or desired sound profiles. (Youvan, 2024) These tools can perform tasks that require years of expertise enabling novice producers to achieve professional grade results. Automating these technical aspects of production frees producers to focus on different aspects of production, however, there have been accounts of musicians dissatisfied with automated mastering as the results sound too refined. Analyses generated by AI can serve as a second opinion, providing insights and suggestions that complement and derive from the expertise of professionals. By reducing the time spent on repetitive tasks, AI allows faster delivery and enhanced productivity.

LLMs have demonstrated a strong ability to assist in lyric writing, offering creative suggestions based on themes, styles, or moods. Their capacity extends beyond natural language generation—many are trained on vast datasets of music and language, enabling them to generate coherent melodies, suggest chord progressions, or devise song structures with full verses and choruses, etc. (Huang et al., 2020) While some outputs are presented in text format they represent structures that aren't linguistic, but rather musical and doing so with positive results. These tools do not replace human creativity but serve as collaborative partners, helping spark creativity, overcome writer's block, and accelerate the journey from idea to composition. This also comes with challenges, for instance, most lack the skills to take the most use of AI tools. Sound design, a central aspect of music production, benefits significantly from AI. ML algorithms can analyze and generate unique sound textures, enabling producers to create entirely new sonic palettes. The main challenge lies upon the subjective listener's experience where AI still falls short in assisting (Thorogood, 2021), the main applicability on the topic is in tasks such as isolating specific stems within a track, vocals or instruments, enabling producers to repurpose sounds in lesser steps and different ways. This also has significant implications for genres like hip-hop and electronic music, where sampling is an industry standard.

Although AI offers immense potential, its integration into music production raises several questions. Ethical concerns include the potential for AI-generated music to devalue human artistry and originality. The use of copyrighted material in AI training datasets also presents legal and moral dilemmas (Grynbaum & Mac, 2023). More than a thousand UK artists such as Hans Zimmer, Damon Albarn and Jamiroquai have come together in the silent album "*Is This What We Want?*" to protest the copyright law changes that allow AI companies to build

products using proprietary work without a license. The album consists in a series of recordings from empty studios and performance spaces, denoting what can be the cultural consequences of the infringing copyright law in the benefit of the technological business. It is more beneficial to stand on the side of optimism facing technological breakthrough and that's one of the guiding theses for this study. The sense that AI can replace artists is not a true preoccupation since art requires intent, a desire to express something, which can never be replicated by an insentient algorithm. The responsible and lawful use and development of AI products can be achieved if all involved parts are rightfully accounted for.

2.2 SYSTEMATIC LITERATURE REVIEW

This chapter is elaborated in accordance with the *PRISMA* (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology, a widely accepted framework designed to ensure transparency, reproducibility, and rigor in the review process. *PRISMA* provides a structured approach to identifying, screening, and selecting relevant studies, culminating in a clear presentation of the included research. The process is divided into four key stages—Identification, Screening, Eligibility, and Inclusion—as defined in the *PRISMA* Statement and respective checklist (Page et al., 2021).

The previous exploratory review guided the formulation of the below presented SLRQs. It became evident that the study should focus into the following questions:

Table 1 - Systematic Review's Research Questions (SLRQ)

SLRQ1	What AI tools are readily available for musicians and producers?
SLRQ2	How can LLMs benefit the music production workflow?

The literature review questions, as previously outlined, are intentionally broader than the research questions defined for the overall study. Their analysis will serve to address and support the more specific research questions later.

The keywords presented below were selected to ensure a thorough review that would be relevant and the interconnection of the topics. The chosen keywords were subsequently used for formulating the search string used to query the selected databases and extract relevant papers.

Table 2 - Systematic Review's Keywords

	MUSIC PRODUCTION	ARTIFICIAL INTELLIGENCE
KEYWORDS	Music	Artificial Intelligence
	Music Production	LLM
	DAW	Robot
	VST	HCI
	MIDI	Assisted

The resulting search term from the keyword forms the following string:
("Music" OR "Music Production" OR "DAW" OR "VST" OR "MIDI") AND ("Artificial Intelligence" OR "LLM" OR "Robot" OR "HCI" OR "Assisted")

To collect a comprehensive selection of articles on AI in music production, three key resources were utilized: *ACM*, *IEEE*, and *SciSpace*. These platforms were selected for their broad coverage and credited reputation. The first two were researched using a traditional method, extracting research papers using the above-mentioned search term. The latter utilized a more modern approach, directly querying the platform with the SLRQs. *SciSpace* yielded 10 papers per question, that were extracted from *arXiv* and *Google Scholar*.

This strategic choice enabled a robust and credible literature review, capturing the latest and most relevant advancements in the field and ensuring that modern methods were applied alongside the conventional style. This search was conducted in January 2025.

Table 3 - Systematic Review's Resource Databases

DATABASE	RESOURCE URL
ACM	https://dl.acm.org/
IEEEXPLORE	https://ieeexplore.ieee.org/
SCISPACE	https://typeset.io/

The initial review proved that there is extensive body of knowledge surrounding the topic, indicating that inclusion and exclusion criteria should be somewhat restrictive.

Table 4 - Systematic Review's Inclusion and Exclusion Criteria

INCLUSION CRITERIA	EXCLUSION CRITERIA
Meets the scope of the thesis	**Reason 1**: Does not meet the scope of the paper. (E.g. Research papers on music distribution, consumption, recommender systems)
Provides valid insight for the selected SLRQs	**Reason 2**: Is not openly accessible
Paper is published from 2020 onwards	**Reason 3**: Does not provide with a notion/component that is applicable into a usable framework
	Reason 4: Is prior to 2020
	Reason 5: Non-academic or non-scientific papers (e.g., websites, magazines reports, newspapers, consulting articles, books, citations, repositories)
	Reason 6: Relates to AI in Music Production but with minimal user input (unrelated with co-creation or assisted production)

2.2.1 Execution

1. **Identification –** This first stage identified 91 studies by querying *IEEEXplore* and *ACM's Digital Library* with the search term *SciSpace* with the research questions. The first resource database provided 54, the second 17 and the latter had a fixed number of 10 suggested papers per question, therefore 20, in total 91 papers. The papers identified in *SciSpace* were subsequently retrieved for analysis from *arXiv* and *Google Scholar*.

2. **Screening –** From the screening stage 5 papers were removed. 3 were oriented towards physical robots and 2 regarded musical therapy for chronic diseases. 2 other papers were duplicates.

3. **Eligibility –**The remaining 40 papers were assessed for eligibility and 23 were removed based on the following criteria:

   Reason 1 – 11 documents were eliminated due to not being fully aligned with the scope of this study. Some were close matches but were either too broad or too specific.
   Reason 2 – 0 documents were removed for inaccessibility
   Reason 3 – 8 documents were removed for not providing either theoretical or practical usability for the study

Reason 4 – 5 studies removed for being prior to 2020

Reason 5 – 1 document was eliminated for being in an invalid format, specifically, a dataset

4. **Inclusion –** The included documents for research totaled 17 papers, 9 journal articles and 8 conference papers.

The systematic execution is illustrated in the diagram below.

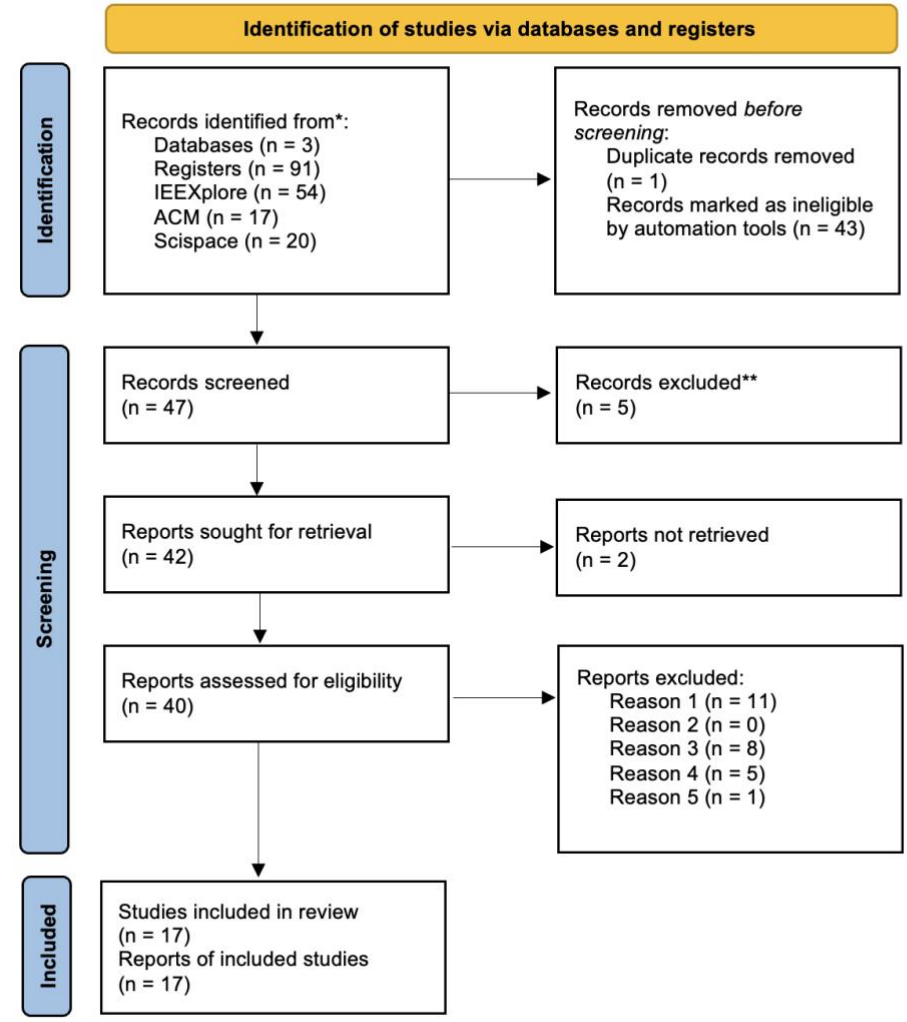

Figure 1 - *PRISMA* execution

The new studies included are identified below in Table 5, along with a brief description of their contributions.

Table 5 List of papers for Systematic Literature Review

AUTHORS	ARTICLE	CONTRIBUTION	PUBLICATION TYPE
(Ford et al., 2024)	Reflection Across AI-based Music Composition	Introduces the concept of RiCE (Reflection in Creative Experience) and the reflection board method where composers are asked to reflect on their AI-assisted compositions every passing hour. Conducts user interviews with participants revealing how reflection influences the direction of their work.	Conference Paper
(Larsen & zhu, 2024)	Ideary: Facilitating Electronic Music Creation with Generative AI	Proposes lateral thinking induced by AI through randomization in electronic music production. Presents *Ideary,* a framework that uses GANs to produce MIDI patterns to assist producers. Ideary was not realized as functioning software, unfortunately.	Conference Paper
(Louie, cohen, et al., 2020)	Cococo: AI-Steering Tools for Music Novices Co-Creating with Generative Models	Introduces a web-based suite for music co-creation, Cococo, a piano-roll interface able to generate melodies based on user input and offering targeted control over the outputs. A study was conducted with novices to evaluate both quantitative and qualitatively the efficacy of the tool in learning and co-creating. Participants were asked to evaluate the AI steering tools of the software and found them helpful in breaking the task of co-production down making it more engaging and controllable than a version of the software without any tools	Conference Paper
(Suh et al., 2021)	AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition	Conducts an experiment consisting of pairing musicians of similar levels where they are asked to create a musical phrase with and without the aid of Cococo. Presents strong evidence supporting multiple ways AI can mediate social dynamics.	Conference Paper
(Koo et al., 2022)	End-To-End Music Remastering System Using Self-Supervised And Adversarial Training	Introduces a system aimed to remaster an audio track based on a reference track. This is achieved through a multi-model architecture featuring a self-supervised mastering encoder with a Mastering Cloner that follows a *Wave-U-Net* architecture. The conclusions present positive remarks towards the resulting pieces. The system was not available on *GitHub*.	Conference Paper
(Agwan et al., 2023)	The Fusion of AI and Music Generation: A Comprehensive Review	The paper analyses the applications of AI in music composing, mixing, remixing, recommending and assisting systems. A review on different model architectures. The work also showcases the fundamental challenges and ethical considerations surrounding the topic	Conference Paper
(Pathariya et al., 2024)	Tunes by Technology: A Comprehensive Survey of Music Generation Models	Focuses on visual representations that facilitate AI incorporation in musical production and network architectures that have been employed in the field, such as LSTMs, CVAEs and CGANs.	Conference Paper
(Yu et al., 2023)	MusicAgent: An AI Agent For Music Understanding and Generation with Large Language Models	Presents a multi-modal AI agent capable of performing diverse music production tasks stationed mainly in three categories: generation, task planning and tool selection.	Journal Article
(Ghosh et al., 2024)	GAMA: A Large Audio-Language Model with Advanced Audio Understanding and Complex Reasoning Abilities	Presents GAMA, a novel LALM (Large Audio Language Model) capable of audio understanding and advanced reasoning that, unlike most LALMs implementations, interprets audio non-linearly and aggregates features in multiple layers of an audio encoder. Furthermore, the study also presents CompA-R, an instruction tuning dataset that enables GAMA's complex reasoning; and CompA-R-test, a human-labelled benchmarking	Journal Article

		dataset that evaluates LALM's capabilities on open-ended Audio Question-Answering.	
(Suzuki et al., 2023)	A Comparative Evaluation on Melody Generation of Large Language Models	Studies and compares the capabilities of *Bard* (1.0) and *ChatGPT* (4.0) melody generation. Analyses if the models' outputs reflect the user's prompts both objectively and subjectively. The results favored *Bard* for expressiveness and *ChatGPT* for capturing music features.	Conference Paper
(Yuan et al., 2024)	ChatMusician: Understanding and Generating Music Intrinsically with LLM	Introduces an open-source LLM pre-trained on ABC notation, a text-compatible music representation. The chosen model is *LLaMA 2*. Presents a novel benchmark, *MusicTheory-Bench* that evaluates LLM's music theory understanding. ChatMusician managed to surpass *GPT-3.5* and *LLaMA2* in both *MusicTheory-Bench* and other benchmarks.	Journal Article
(Doh et al., 2024)	Enriching Music Descriptions with A Finetuned-LLM and Metadata for Text-to-Music Retrieval	Presents a text-to-music retrieval (TTMR++) framework that is trained to embed text data, such as tags, track names, artists and genres with audio samples with state-of-the-art results. The study aims to demonstrate how a model can understand the interconnection of natural language semantics and audio.	Conference Paper
(Rasal, 2024)	A Multi-LLM Orchestration Engine for Personalized, Context-Rich Assistance	Introduces a solution that intends to minimize LLM hallucinations and long-term context retention through a combining multiple LLMs and a temporal graph database that stores context over long periods of time.	Journal Article
(Zhou et al., 2024)	Can LLMs "Reason" in Music? An Evaluation of LLMs' Capability of Music Understanding and Generation	Explores the capabilities of a variety of LLMs in the realm of comprehension and generation of symbolic music notation. Studies multi-step prompt engineering and analyses 4 of the main LLMs' capabilities (*GPT-4, Gemma-7B, LLaMA 2-7B and Qwen-7B*) in tasks through statistical analysis of results as well as qualitative subjective assessment of results. Results and prompts available on *GitHub*.	Journal Article
(Deruty et al., 2022)	On the Development and Practice of AI Technology for Contemporary Popular Music Production	Offers a perspective on the development of AI tools for contemporary popular music discussing the best practices on the topic. Provides recommendations for the development of AI tools for this intent. Analyses the creative process of artists and reports some iterations of artists utilizing AI software.	Journal Article
(Anantrasirichai & bull, 2022)	Artificial Intelligence in the Creative Industries: A Review	Presents an extensive compendium of AI in the creative industries providing examples of ML algorithms that support composition, live performance, copyright protection. Several notable examples include Coconet, *Sony's* Flow Machines, *OpenAI's* Jukebox, *Google's* NSynth and Magenta Studio, RythmVAE, among many others.	Journal Article
(Tokui, 2020)	Towards democratizing music production with AI-Design of Variational Autoencoder-based Rhythm Generator as a DAW plugin	Presents M4L.RhythmVAE - a Variational Auto-Encoder-based rhythm generator plug-in for Ableton Max 4 Live. The solution allows for users to train their own model. The model learns from user's MIDI input and can generate rhythms that can be manipulated in real-time by the user with an intuitive XY pad interface.	Journal Article

2.2.2 Analysis and Discussion

The present section will provide cohesive answers to the identified SLRQs about AI in Music Production. Many of the papers bring insights that assist in answering both questions.

SLRQ1 - What AI tools are readily available for musicians and producers?

There is a wide range of AI tools accessible for producers, musicians and researchers. The included studies affirm that there is benefit in AI-driven Human-Computer-Interaction in musical composition. The generative capability of algorithms paired with the user's creative input pave new ways of expressing with immeasurable potential (Larsen & Zhu, 2024). Besides music generation (Anantrasirichai & Bull, 2022), AI tools are capable of identification, transcription, mixing and mastering (Koo et al., 2022), remixing, stem separation, general assistance, music understanding (Ghosh et al., 2024) and explanation. All these functions accelerate the delivery of musical output and can aid the different types of professionals involved.

There is an evolving number of architectures for the development of AI, each with specific uses and characteristics. Some of the identified architectures for music software were *GANs , VAEs, Transformers, RNNs* and *CNNs* (Pathariya et al., 2024), each with distinct limitations and variations, further described in Appendix A. Some studies sought to develop architectures that combined these structures (Koo et al., 2022), others present frameworks that leverage LLMs as selectors of different music-related AI tools (Yu et al., 2023).

Other than just referencing these technologies, it is important to mention the different types of input that allow humans to train machines to subsequently interpret and generate sound. This is important for the gradual improvement of responses. The inputs can be textual, such as natural language queries, guitar tabs or ABC notation; binary, like MIDI data (Deruty et al., 2022); or visual, such as Piano Roll notation: a visualization that presents musical notes and their time sequence in two-dimensional space - Time and Pitch. Through this type of imagery, ML algorithms can capture the sequential and melodic characteristics of musical structures such as chords or harmonies through the visual patterns they form. Mel spectrograms are a type of sonic frequency distribution-based visual representation of sound that allows for three-dimensional depictions intended to mimic how the human ear interprets sound and thus providing more depth than the bidimensional Piano Roll. The tools referenced in the included papers are presented in Appendix B.

Graphical User Interfaces (GUI) significantly enhance user experience and creative workflow by providing interactive control (Larsen & Zhu, 2024) and improved comprehension of composition experience through visual feedback. A thoughtfully designed interface reduces the overwhelm some musicians and producers may experience from HCI in digital workstations, particularly the most inexperienced and less tech-savvy (Louie et al., 2020).

SLRQ2 - How can LLMs benefit the music production workflow?

LLMs possess remarkable understanding on multiple subjects through language (Ghosh et al., 2024) and can therefore improve the music production workflow in a varied set of ways. There were several identified roles in production that could be assumed by LLMs or ML algorithms, to be more general. These are described and categorized in Appendix C. LLMs present striking dynamics into human-AI co-creation (Ford et al., 2024) and even human-human collaboration acting as a "third collaborator", offering initial or additional musical ideas in collaborative settings (Suh et al., 2021) and introducing a sentiment of “us and them” in human-to-human collaboration furthering the bond between musicians. In situations where collaborators might be hesitant to directly critique each other’s work, AI’s suggestions and evaluations can provide neutral grounds for discussion and further development. LLMs can also help arrange and organize various human-created ideas by generation of components.

Alternatively, LLMs are competent in general assistance tasks such as planning and as of recently web-search, that while not directly related to production, enhance productivity, organization and information retrieval. These characteristics are relevant for projects of any sort. The latest trend of AI agents has already some applications in music such as MusicAgent(Yu et al., 2023), a comprehensive system that uses an LLM as the backend for an autonomous workflow that can plan tasks, select tools and present generated responses all in one. LLMs could likewise assist with more mundane aspects of music production and repetitive tools such as formatting data to move between different instruments and AI tools.

Some of the identified models presented deficient performance in song-level multi-step reasoning (Zhou et al., 2024). Language Models are not designed to understand non-verbal audio, though it has been identified a new genre of language models that can extract features from sounds that aren’t linguistic through multi-modal input – LALM (Large Audio-Language Models). Continual pre-training and fine-tuning on music-specific data were some of the solutions that can mitigate the LLMs limitations in the subject. (Yuan et al., 2024) demonstrates the benefits of fine-tuning and pre-training language models on text-compatible music representation (ABC notation) effectively treating music as a second language. Similar results are presented in (Doh et al., 2024) where pre-training allowed to achieve similar results between LLMs of dramatically different sizes. Another finding from the study is how incorporation of musical ability in LLMs can hinder their foundational language skills.

The potential of LLMs to understand and generate music based on different user input like texts, chords, melodies, motifs, and musical forms can streamline tasks such as generating chord progressions, harmonizing melodies, or creating variations of musical ideas. The development of domain-specific benchmarks, such as *MusicTheoryBench* (Yuan et al., 2024), *Marble* or *CompA-R-test* (Ghosh et al., 2024), designed to systematically evaluate audio and music comprehension and reasoning capabilities in Large Language Models (LLMs). These

benchmarks provide structured, standardized tasks that assess various dimensions of musical knowledge, including (but not limited to) harmony, rhythm, notation reading, tonal structure, and stylistic analysis. By offering quantifiable metrics, they enable researchers to objectively measure how well LLMs understand and process musical concepts, both in symbolic and textual representations. Moreover, such benchmarks help identify current limitations in LLMs' generalization abilities when dealing with abstract or context-dependent musical reasoning. They also serve as a foundation for tracking progress over time, guiding improvements in model training and fine-tuning strategies tailored to the musical domain. Many of the identified resources also mention general benchmarks not directly related to music but used to evaluate capacity of different models. Benchmarking contributes to the development of AI systems, identifying performance bottlenecks, and guiding improvements in algorithms and models, leading to more reliable and capable solutions.

3 METHODOLOGY

3.1 OVERVIEW

The research methodology is divided into four phases: Preparation, Conception, Execution and Conclusion. These phases are sequential and have precedence between each other, as depicted in Figure 2. The study finalizes with a qualitative appreciation of the experiment's results and the prototype's user experience as well as conclusive remarks.

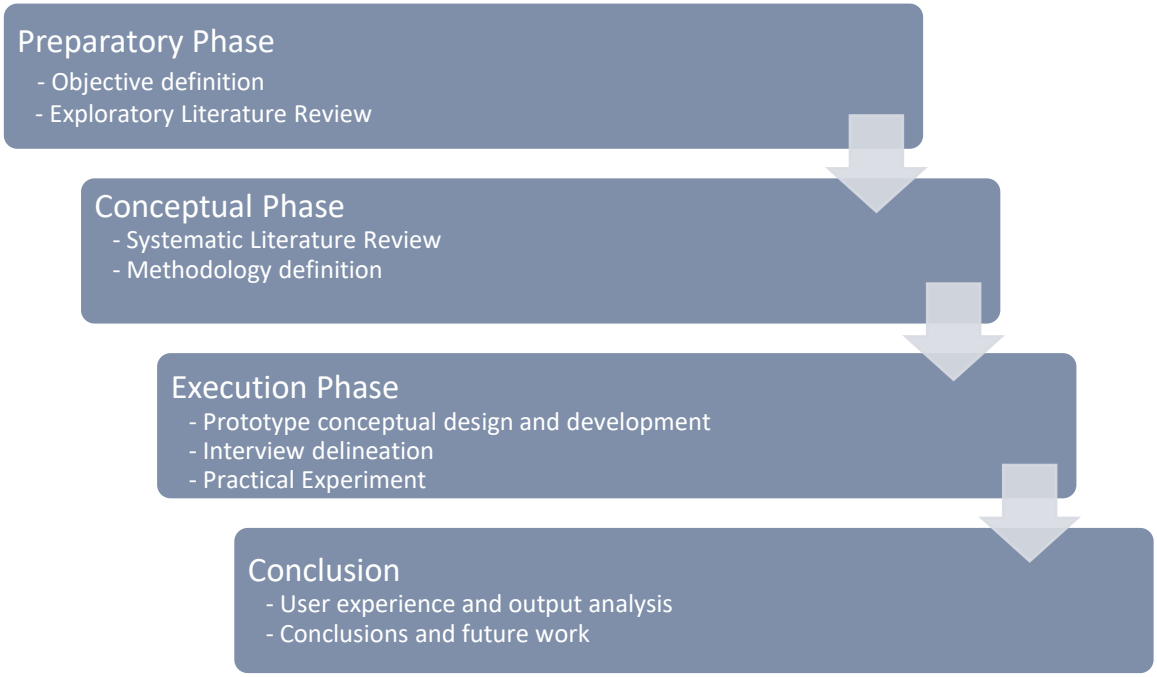

Figure 2 - Methodology Flow Diagram

3.2 PREPARATORY PHASE

This stage is designed to establish the foundations and guiding principles for research. The two key activities that ensure proper grounds are objectives definition, outlining general success factors and exploratory literature review purveying with a direction for the remaining research. This literature review targeted sources from academic databases such as *ACM Digital Library* and *IEEEXplore*, prioritizing studies on AI in music production to identify trends, gaps, and opportunities.

3.3 CONCEPTUAL PHASE

This phase defines the framework for research. The systematic literature review as previously stated follows the *PRISMA* method to ensure a concise and efficient gathering of evidence.

The methodology outlines the strategy for the subsequent phases, such as a design-based approach to the development of the prototype.

3.4 EXECUTION PHASE

The execution phase translates the conceptual framework into actionable development. It begins with a presentation of a conceptual design for the prototype, complete with functional and non-functional requirements, interface design, role definition and tool selection. The development phase will follow a design-based approach as is expected to be iterative and cyclical, accompanied by a brief narration of the process and relevant takeaways for future work. The experiment and interview delineation will determine the goals for question development, profile of participants and interview's environment and setup. The experiment and interview process is expected include musicians/producers of different experience levels in composition in a studio setting. Ethical protocols will be implemented to guarantee participant's privacy

3.5 CONCLUSION PHASE

In this final stage, the research findings are integrated with experimental results to consolidate key outcomes and validate the hypotheses. To extend the practical and theoretical impact of the study, conclusions are derived, and potential avenues for future work—such as system optimization, scalability testing, or real-world implementation—are proposed. Furthermore, limitations encountered during each phase of the investigation, including constraints related to instrumentation, modeling assumptions, and operational conditions, are critically examined to provide contextual relevance and guide subsequent studies.

4 PROTOTYPE DEVELOPMENT

This section documents the theoretical and practical steps taken to develop the thesis' prototype ensuring a comprehensive description of the process and resulting product. The chapter is divided into five parts, each of them dedicated to specific layers of development. The first characterizes the conceptual model, including an overview of the main components and a visual representation of the intended solution. The functional requirements are presented in this segment as well. The second part is a description of the hardware setup and physical requirements accompanied by a brief explanation for the chosen components and limitations that oriented their selection. The third part focuses on the software set up, the tools and components employed, mainly the programming languages, frameworks and libraries that constitute the final prototype.

4.1 CONCEPTUAL DESIGN

The proposed solution, **LTN-DAN - Layered Transformative Neural Dynamic Audio Networ*k***; a system designed to harness AI capabilities into music composition, enhancing agency, creativity and efficiency. The system is based on a dual-machine architecture to effectively manage the computational demands of both music production software and machine learning models. Concerns with computational demand and of running AI and Music software simultaneously was identified as a gap in the SLRQ that motivated the dual-machine setup.

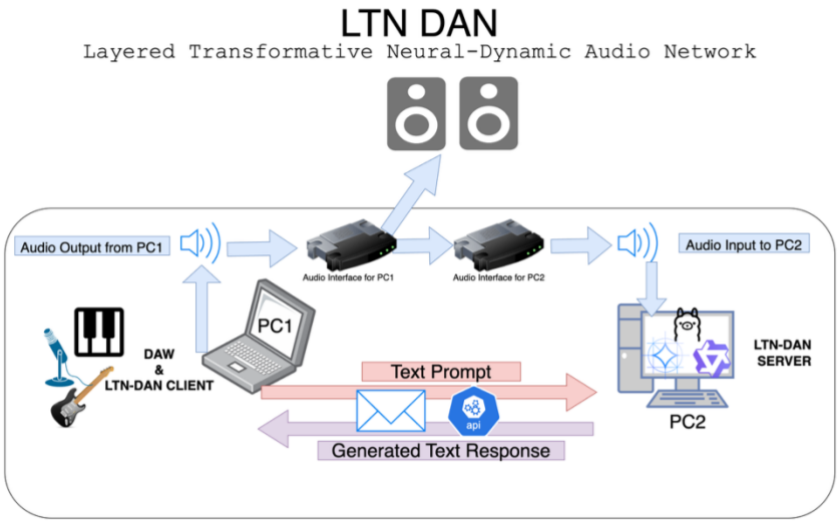

Figure 3 - Conceptual Design for LTN-DAN

It comprises a client machine, further described as PC1, which runs the music production software (DAW or other), alongside LTN-DAN client user interface for multi-modal (text and speech) interaction with PC2, a server machine which hosts the AI models and processes audio requests locally. Each machine requires a dedicated audio interface that ensure consistent audio processing.

The machines are connected both physically, through their audio interfaces, as the PC1's audio output is routed to PC2's audio input to enable speech prompting; and virtually to handle textual prompts and responses coming from the server.

Figure 3 provides a high-level representation of the described system architecture. Detailed specifications regarding the software hosted on both machines, as well as their hardware requirements, are elaborated in the subsequent sections.

4.2 Hardware Requirements

This section outlines the minimum hardware requirements necessary to run the prototype, provides recommendations for optimal usage, and details the actual physical setup employed in the experiment. Notably, no components were acquired specifically for this study; all hardware was sourced from existing resources based on availability and compatibility with the system's design requirements.

4.2.1 Hardware Configuration

The realization of this experiment requires a minimum hardware configuration consisting of two computers — PC1, serving as the client machine, and PC2, functioning as the server — along with a single audio interface, a router to establish a *Local Area Network (LAN*) between machines, a microphone, and musical instruments such as a guitars, drums, and keyboards. Additionally, power, media, and network cables are necessary for connecting and powering all components. Although the conceptual architecture includes a second audio interface, its inclusion is not mandatory. This secondary audio interface enables speech-to-text recognition and audio processing to be carried out directly on PC2, eliminating the need for data transfer via API from PC1, where the audio originates. This design choice significantly enhances response times, reduces computational load on PC1, and improves overall system efficiency.

4.2.2 Setup used for experiment

For the experiment, the hardware includes two computers: PC1 is a 2020 *M1 MacBook Pro* 8-core CPU and 16 GB of RAM; and PC2, an *Acer* desktop PC with an *Intel Core i5-10400F* processor running at 2.90 GHz, 16 GB of RAM, and an *NVIDIA GeForce RTX 2060* graphics card. Regarding audio interfaces, PC1 is connected to a 3rd-generation *Focusrite Scarlett 4i4* audio interface, while PC2 utilizes a *Yamaha Audiogram 3* as its audio processing device. As previously mentioned, the output from PC1's *Scarlett 4i4* is routed to PC2's *Yamaha Audiogram 3*, enabling audio signal processing on the server machine for transcription.
The network configuration involves both PC1 and PC2 being connected to the same *Local area network (LAN)* via a shared router, ensuring stable and low-latency communication between machines. This network setup facilitates the *REST API*-based textual communication and supports the synchronization of tasks across the architecture.

Digital Audio Workstations (DAWs), which are integral to client-side operation in PC1, impose substantial computational demands. Their performance can be adversely affected by concurrent execution of other resource-intensive processes. Therefore, careful resource allocation and task distribution across the dual-machine architecture were critical to maintaining optimal performance. The selected DAW was *Ableton-Live*. All hardware components utilized in this experiment were acquired previously and did not require ad-hoc procurement. Thus, the selection of equipment was guided by its availability and alignment with the system requirements outlined in conceptual design. This approach ensured the feasibility of the prototype development while minimizing additional expenditures.

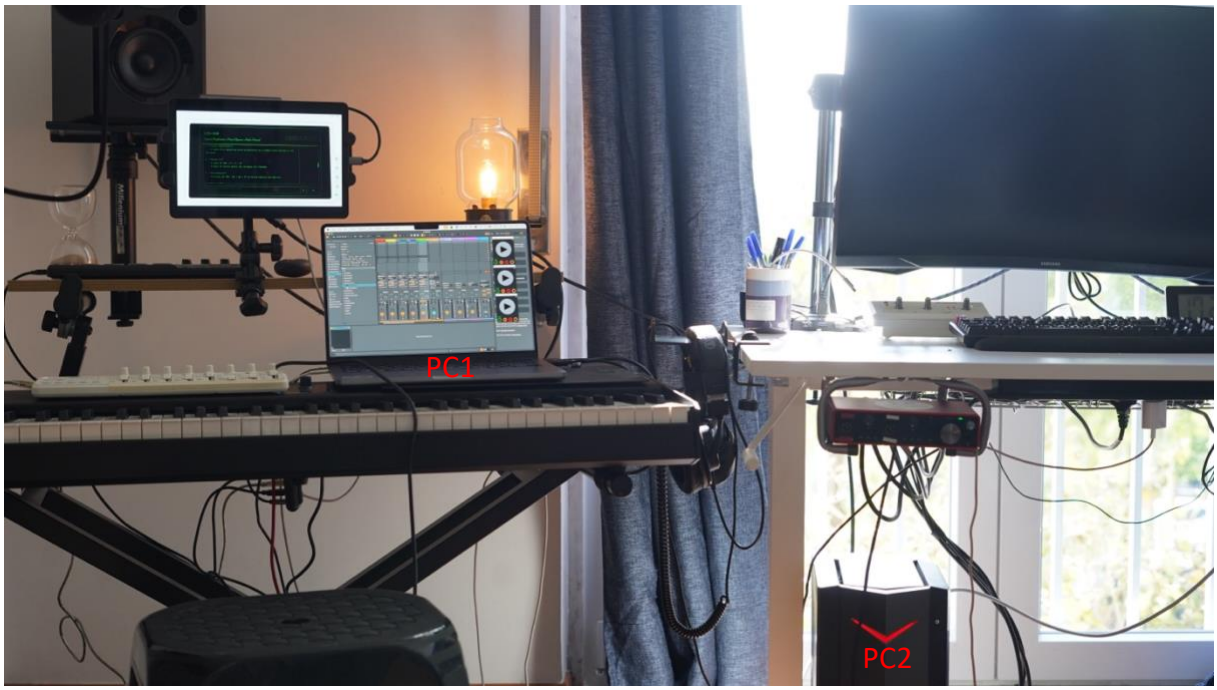

Figure 4 - Experiment setup

Above is a picture of the setup for the experiment. Besides the components mentioned earlier there is also a microphone, MIDI controllers, speakers and a variety of both physical and virtual instruments such as drum pads and guitars at the disposal of participants.

4.3 SOFTWARE CONFIGURATION

The application implements a client-server architecture utilizing modern web technologies and artificial intelligence models. The system integrates real-time speech processing, LLM interactions, and a dynamic user interface updates through a combination of synchronous and asynchronous communication protocols.

4.3.1 Client-Side Development

The client-side implementation of this system employs a carefully selected stack of technologies designed to create a responsive, interactive, and visually engaging user experience while maintaining robust real-time communication capabilities.

The frontend was developed using a standard web technological stack: *HTML* as the structural foundation and *CSS* to handle the visual styling aspects, with particular emphasis on modern features such as Custom Properties enabling dynamic theme switching without *JavaScript*. The application leverages these variables extensively to maintain consistent visual language throughout the interface while allowing for runtime modifications.

The application employs two behavioral layers: a *JavaScript* frontend layer managing client-side UI logic and *Socket.IO* communication, and a *Python* backend layer handling server-side processing, AI model integration, and real-time message routing.

The application avoids external *JavaScript* frameworks or libraries for the core functionality using *Vanilla JavaScript* instead to minimize dependency overhead and maintain control over rendering and update cycles. This approach results in a light codebase with reduced payload size and faster runtime performance. The only external *JavaScript* dependencies are *Socket.IO*'s client library, essential for the real-time communication structure, *Font Awesome* and *Google Fonts* for fonts and typography.

The *Python* backend serves as the core communication hub of LTN-DAN’s client application, built using Flask as the primary web framework. This implementation handles critical operations: serving the frontend interface, managing real-time communication via *Socket.IO*, and orchestrating interactions with external AI and audio processing services. The backend architecture employs a modular design pattern with clear separation of concerns, utilizing Flask's routing system to handle HTTP endpoints for audio recording management while

implementing *Socket.IO* event handlers for real-time chat functionality. The application leverages *Python*'s asynchronous capabilities through *Eventlet*, enabling efficient handling of multiple concurrent *WebSocket* connections and ensuring responsive real-time communication between the frontend and backend components.
This implementation integrates several specialized libraries to deliver its core functionality. *Flask-SocketIO* provides the real-time communication infrastructure, allowing the server to maintain persistent connections with clients and handle bidirectional message streaming for the chat interface. The *Ollama Python* client library enables direct interaction with locally ran AI models, facilitating the generation of contextual responses and maintaining conversation history through structured message handling. Additional components include Flask-Limiter for implementing rate limiting, *python-dotenv* for environment variable management, and the *Requests* library for *HTTP* communication with external speech-to-text services. The backend also implements comprehensive error handling and logging mechanisms for debugging information while troubleshooting. This *Python* foundation creates a stable, scalable backend that supports the application's real-time chat capabilities while maintaining security and performance standards.

The client establishes a persistent connection to the server upon initialization, listening for events to receive messages and send conversation history to the model for context. Event-driven architecture enables real-time updates between the user interface without requiring page refresh or explicit polling, creating the common chat-like interaction with the language model. The solution manages this through structured event handlers. This approach allows the application to handle both complete messages from the speech-to-text service and streaming partial responses from the language model with appropriate rendering strategies for each.

Visual feedback mechanisms are incorporated throughout the interface, including subtle animations for loading states, recording indicators, message transitions. The chat display system employs efficient techniques to manage *Document Object Model (DOM)* nodes—individual building blocks of a webpage, like text, images, or buttons—to ensure smooth performance during longer sessions. Each message is added to the webpage as a distinct DOM node, styled differently based on whether it originates from the user or assistant, and the display automatically scrolls to reveal the latest content.

For the language model’s streaming responses, the system updates existing *DOM* nodes rather than creating new ones for each text segment. This reduces the computational effort required to modify the webpage, resulting in smoother visual updates. Interactive elements, such as buttons, text input fields, and control panels, are designed with consistent visual indicators for actions like hovering or selecting, providing clear user feedback. The recording feature incorporates animated *DOM* nodes to visually signal the recording state, accompanied by a

loading indicator node that appears during message processing to keep users informed of system activity.

The graphical user interface implements a retro-inspired layout mimicking classic monitors and terminal displays, maintaining a modern user-experience principles. The monospaced typography enhances readability of responses and technical content, while the distinctive visual style creates an engaging user experience that differentiates the application from conventional LLM chat interfaces.

Figure 5 – Screen capture of LTN-DAN's client GUI

4.3.2 Server-side development

The server architecture for PC2 developed for this project has two core components: a large language model (LLM) and a speech-to-text (STT) model. The LLM is hosted locally via *Ollama*, an open-source platform designed to facilitate the deployment and execution of large language models on local hardware. The selected language model is *LLaMA3.1-8B*, an open-source model developed by *Meta* with approximately 8 billion parameters. During the model selection process, various language models from different providers and with varying parameter sizes were evaluated. The *8B* variant was chosen as it provided a balanced trade-off between response coherence, inference speed, and computational resource requirements. Larger models, while potentially offering greater contextual understanding,

demanded significantly more processing power and memory, hindering real-time performance and scalability. Other models that were considered for LTN-DAN were *Google's gemma3* – rejected for not having an 8B variant - and qwen3 that although not being the chosen model, delivered positive results and can easily be swapped for testing. This easy interchangeability of models poses as one of the benefits in the presented solution in comparison with using commercial AI models.

For speech-to-text, OpenAI's *Whisper* was employed. *Whisper* is an open-source, multilingual transformer-based model capable of accurate transcription for multiple languages, including English and Portuguese, the languages used in the experiment phase. Similar to LLMs that come in various sizes to balance performance and computational requirements, Whisper also offers different versions, each designed to trade off accuracy and efficiency. The "base" model demonstrated sufficient accuracy without requiring fine-tuning or additional training. The system pipeline was designed so that the transcription from the speech-to-text model serves as input to the language model, enabling interaction through voice commands. Below is a diagram where both methods for prompting are pictured.

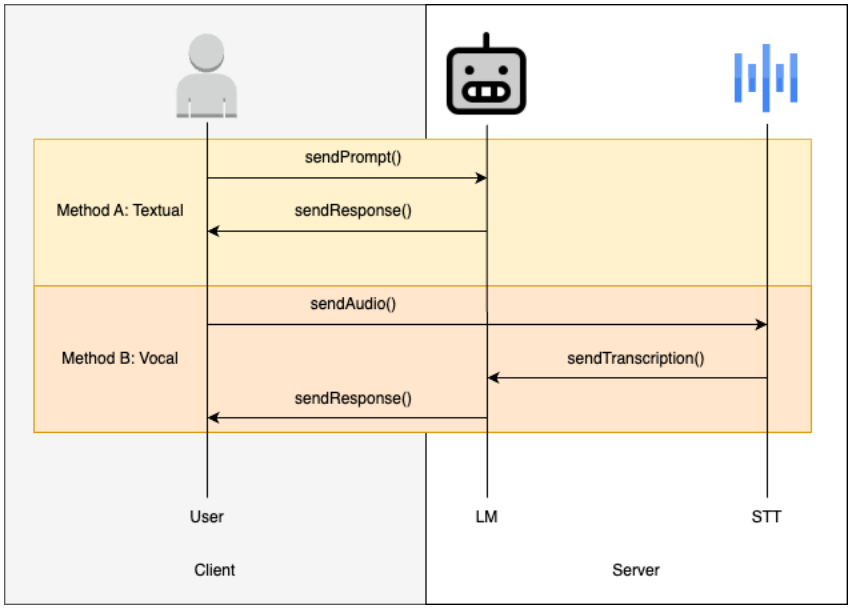

Figure 6 - Communication methods between client and server

To ensure modularity, scalability, and ease of deployment, both models were encapsuled in *Docker* containers, facilitating reproducibility and simplifying integration into different environments. Although the server and client were developed and are intended to run in separate machines, they are grouped into one repository and can be accessed on *GitHub*.

5 TESTING & EVALUTION

5.1 EXPERIMENT OUTLINE

The primary objective of this phase is to assess how effectively LTN-DAN supports users on their specific creative endeavors. Participants receive a quick instruction on how to engage with the provided setup and after a short assessment aimed at profiling participants demographically and technically, they are challenged to write a short musical piece with the support of LTN-DAN. The experiment is meant to have a duration of 40 to 60 minutes. Besides LTN-DAN, participants will have at their disposal a fully functional audio setup, complete with DAW (Ableton Live), instruments and audio interfaces as described in the previous section. Additionally, participants will also test *Google's Magenta-Studio* plugin embedded in *Ableton-Live*. *Magenta* provides a second generative tool that can be used in conjunction with LTN-DAN. This research method is qualitative and aimed to obtain insights into not only user experience of the prototype, but also perceived creative output enhancement, and the role of AI in music creation. This will be achieved by employing a mixed-methods approach involving a semi-structured interview with pre- and post-experiment survey. Participants will be able to select one of the identified roles in Appendix C for the assistant. Alternatively, if none of the roles fulfill the participants expectations, they can use the default general base prompt: "You are a helpful music assistant".

Appendix D presents the list of questions that constitute the interview. The first questions regard the participant's background, motivation and general expectations for the session. The post-experiment questions intend to gather feedback towards the prototype a comparison between the provided AI tools. As the interview is semi-structured, some other questions and comments may naturally surface during the experiment and will be documented in the Results section.

Table 6 – Demographic details of participants

ID	AGE	GENDER	PROFILE	EDUCATION LEVEL	YEARS OF EXPERIENCE
P1	25	M	Trained Guitarist	Master's degree	3
P2	27	M	Producer/Artist	Bachelor's degree	6
P3	29	M	Hobbyist	Bachelor's degree	0
P4	26	F	Singer/Songwriter	Master's degree	1

5.2 EVALUATION

The evaluation phase study featured four distinct experimental instances, each showcasing the prototype's application in music production. The participants were 3 males and 1 female with ages between 25 and 29 years. Each followed different approaches and offered valuable insight both for technical improvements of the prototype and for the general use of AI in composition.

On the preliminary assessment, Participant 1 described himself as a guitar player/producer with an intermediate level of proficiency in DAW software. His sentiment towards the use of AI in Music was described as "apprehensive". The selected role for LTN-DAN was "Composer" P1 started the experiment by asking LTN-DAN to provide him with a funk chord progression to which the assistant provided "*Cmaj7 - F7- G7"*. P1 recorded a loop on the piano with the given progression. Later, he asked for a rich sounding strings patch from one of the provided plug-ins. P1 used *Magenta-Studio's "Generate"* plug-in to create a Drum groove and a bassline and used both in the final composition after adjusting the generated outputs. P1 also queried LTN-DAN for technical advisory on two instances: instructions on how to transpose a recorded bass line into the correct tonic key and regarding. On post-experiment P1 found the AI capabilities useful although he also pointing out that without music software expertise, he wouldn't be able to get any results from them. P1: "LTN-DAN was helpful but at times inaccurate as he wasn't able to show me how to play the provided chords on the piano and kept showing how to play them on the guitar". P1 noted that navigation between responses could be improved but found the solution helpful in sparking creativity and technical guidance. P1 considered that the assistant influenced his process all the way through the experiment giving him the starting points for his compositions and technical assistance throughout the experiment. P1 suggested visualization of chords as a potential improvement to the UI. P1: "*Magenta-Studio* is better for the initial stages, but *LTN-DAN* offers instructive support throughout the whole process".

Participant 2 characterized himself as a producer/artist with high-level proficiency in music software, mainly Ableton but rather unexperienced in using AI tools. The sentiment towards AI in the arts was described as positive, seeing it mostly as a tool unable to replace the role of creatives and can hardly foresee this type of technology ever becoming a true threat in that regard. The selected role for LTN-DAN was also "Composer". Similarly to P1's approach, P2 also asked LTN-DAN for a chord progression but not specifying any genre, reflecting his unfamiliarity with prompting language models. P2 used the provided chord progression and manually created a lead track based on it. To construct the beat, P2 combined both traditional beat-making with generated content, using generated MIDI tracks from *Magenta-Studio's "Generate"* for a percussion track. P2 was rather satisfied with the generated percussion grooves. Magenta-Studio was further utilized to generate a bassline which was transposed and altered by the user. P2 requested hip-hop lyrics in Portuguese without specifying any

theme and used the resulting output as a basis for a freestyle. The user didn't find the lyrics as impressive as the generated drum patterns but still useful for ideation. P2 was rather pleased with the resulting product and with the experiment in general stating that he would recommend both the tools and that they paired well with his usual approach to music creation. P2 suggested adding a stop button to LTN-DAN client to interrupt the model mid-response.

P3 is a guitar player with minimal experience in production having been involved in recordings before but with very limited skill in technical aspects of music production and composing. His sentiment towards AI in music was best described as "apprehensive" despite being a prolific user of AI tools in his profession (IT). The selected AI role was "*Assistant Engineer*" due to lack of technical expertise. He started to query model on how to record a track on guitar and loop it on the provided workstation. The instruction provided by the model was somewhat lengthy but elucidative enough to complete the task. P3 queried the model for instruments that would pair well with the acoustic guitar and acted on the suggestion by melody in the harmonica but was displeased with the result. He found the duration of the experience short for an inexperienced user, "someone who plays instruments for leisure does not have the necessary skills to record a musical composition in such short period". He was unable to make use of Magenta-Studio and found himself overwhelmed by the apparent complexity of the task to take full advantage of the tool. In retrospect, P3 didn't see much benefit in the provided tools to spark creativity, but rather as a technical guide. Suggestions for textual input on *Magenta* and visual representations of chords, time signatures and other music theory notions were provided. P3 felt that although the tools would not aid him during the experiment they were an interesting safeguard for creative blocks and would recommend them for more technically advanced producers.

P4 is a singer/songwriter with basic-level experience in production, stating that her creative sessions aren't normally accompanied by software, using mostly guitar or piano, pen and paper. P4 separates the composition process from the production since studio time is limited. The sentiment towards AI in music was described as "accepting but cautious" and decided to not use any specific role for the assistant. Her approach differed from previous participants starting with Magenta's "*Generate"* for a drum groove and requesting a bass line to LTN-DAN. P4 was surprised by the outputs from Magenta-Studio. Following LTN-DAN's suggestion, she recorded the outputted bassline with some adjustments over the generated drum rhythms. P4 also experimented with lyric writing aided by the assistant requesting for poems around random topics that came to mind. P4 was not impressed by the output results with regards to their lyrical quality but found some of the content useful for further development when in creative block. In overall, P4 found Magenta-Studio as a positive tool for sparking creativity but not LTN-DAN, seeing it more as a technical assistant and suggesting that it could be beneficial to have control over the base-prompt and context inside the GUI so that the user could swap between different roles mid-session.

6 DISCUSSION

The experiment was aimed to study the effects of AI in musical composition where the prototype LTN-DAN and *Google Magenta-Studio* were tested for their usefulness in a studio environment. There was an attempt at responding to the research questions, especially RQ1 and RQ3. Participants evaluated whether the tools would enhance their creativity and if the perceived benefit from using them was worth adapting their composition process. As expected, the participants found the tools to have value in different moments of ideation. *Magenta-Studio* was regarded as an impactful tool for getting out of creative blocks while LTN-DAN was a powerful assistant that can aid with technical issues and questions during both composition and production, with a bigger preponderance on the latter.

The participants were in overall pleased with the provided tools and their feedback brought valuable insight for the present research and future work. Some participants even shared similar suggestions, pointing out some technical drawbacks of the software. It became apparent that *Magenta-Studio* sparked more curiosity since LTN-DAN is an iteration of a tool which has become widespread – a Chatbot. Nevertheless, users found novelty in both tools. The objective of the experiment was not to thoroughly explore all the capabilities of the provided tools but to introduce them to musicians and study their interaction during composition. The idea was to see what creative benefit would surface from the conversational interaction between man and machine.

During the experiments it was concluded that not only the AI's suggestion would spark creativity but also that the technical guidance would allow the participants to act more creatively. Unlike previously expected, the less experienced participants, with the assistance of *LTN-DAN* were able to deliver compositions even though having no formal training while the more experienced were able to adapt their composition approach to a setup differing to their usual and deliver complex pieces. Both findings suggest how AI tools can aid musicians of all levels.

As it was identified in previous research the AI tools impacted not only the results but also the social interaction in composition. Participants discussed with the researcher as the AI tools came up with suggestions or musical pieces. The introduction of a third party in the form of the generative AI tools in a collaborative environment facilitated the interaction between participant and researcher. The experiment was able to evaluate the benefits generative tools can bring to musicians in one of the most crucial processes of their work, however much can still be assessed about the assistance from AI in other aspects that are important to musicians, such as management, planning and communication. This doesn't discredit the findings from this research but highlights potential areas for further investigation.

7 CONCLUSIONS AND FUTURE WORK

7.1 SYNTHESIS OF THE DEVELOPED WORK

This study aimed to investigate the impact of AI in music from the perspective of creators, particularly focusing on its potential integration into digital audio workspaces. Recognizing AI's transformative power across industries, the research explored how generative AI could function as a tool to accelerate production. The investigation addressed the contentious nature of AI in the arts, acknowledging the perceived existential threat it may pose for artists while also considering the alternative perspective of AI as a supportive tool that can enhance workflows without compromising artistic integrity. Through exploratory and systematic literature reviews in accordance with the *PRISMA* methodology, the study assessed the current state of AI integration, identified key trends, gaps, and technical challenges, and explored potential roles for AI within the creative process.

The most significant outcome from this study is the development of LTN-DAN (Layered Transformative Neural Dynamic Audio Network), a functional prototype designed to integrate AI capabilities, specifically LLMs and STT models, into the music composition workflow. Utilizing a dual-machine client-server architecture to host the models locally, the prototype allowed users to interact with AI models through a retro-inspired graphical user interface, enabling text and speech-based prompting. The prototype was evaluated through user experiments and interviews in conjunction with another generative AI software – *Magenta-Studio* – collecting qualitative insight into their impact on creative agency, workflow efficiency, and artistic decision-making, while also considering implications for originality and authorship in AI-assisted music production.

The research has successfully addressed the initial research questions. The developed solution demonstrates how language models can potentially enhance collaboration, productivity, and creativity in music production without reducing the artist's central role as creator. The study identified several roles AI can assume, such as co-creator, assistant, teacher, or sound engineer, and began to explore how these roles influence the creative process and decision-making. Furthermore, the evaluation phase provided insight into the extent to which musicians are receptive to collaborating with an AI model, indicating that while apprehension still exists, there is also recognized helpfulness and potential for sparking creativity.

7.2 LIMITATIONS

LTN-DAN's hardware selection was guided by availability having no budget allocated to the project. The computational demands of DAWs and LLMs motivated the development of a dual-machine setup. Running a Language Model locally allowed for independence from commercial AI providers and the capability of fine-tuning for task-specific optimization. However, it also limited the capacity to deploy larger models, as they exceeded the computational capabilities of the available hardware. Resource constraints blocked model fine-tuning and further development of the prototype. During the development there could have been made a more structured selection for the models used in the prototype. Some participants pointed out inaccuracies on responses and a lack-luster ability in generating coherent writing for verses. Navigational improvements on the interface were also suggested.

A significant limitation of this study is the relatively small and homogeneous sample size, which naturally restricts the generalizability of the findings. Many invitations received no response, and some invitees were only able to participate online. There was a decision to do on-site only experiments as the remote approach would introduce unnecessary entropy.

The identified roles and their impact on model responses were not thoroughly studied as well. Although all the roles identified were meant to assist musicians, some of them weren't suited for the experiment. To thoroughly assess the roles it would have to be conducted a different experiment where users would query the assistant using both the general base prompt and the best suited role's base prompt. Additionally, other roles were

AI-generated music is still often distinguishable from human-created composition by experienced listeners. While this does not reflect a limitation of the research itself, it motivated the pursuit of improving HCI for AI in music production. Ethical concerns surrounding the potential for AI-generated music to devalue human artistry and originality, as well as issues related to the use of copyrighted material in training datasets remain significant dilemmas. Although there has been significant improvement in this technological realm since the beginning of this research, LLMs still struggle with complex reasoning, maintaining long-term context and are prone to hallucination. Their performance is heavily reliant on the quality of training data.

7.3 FUTURE WORK

Building upon the findings of the present study, a set of elements for future work was identified. A key area is the continued exploration and development of methods to integrate AI into artistic workflows in ways that complement, rather than replace, human creativity. Refining LTN-DAN based on broader feedback from practitioners and experts is essential for ensuring its further applicability. By making the software available as open source on *GitHub*, it becomes part of a collaborative permitting the contribution of developers and other interested counterparts. Sharing LTN-DAN within relevant communities, mainly in the AI and music production spheres can help gather valuable feedback and usage data to inform future refinements.

Several technical improvements to the prototype were identified, including enhancing the AI's understanding of music theory, improving the quality of responses through fine-tuning and perfecting the user interface for better navigation. Other features that would enhance the prototype could be the addition of a dropdown to select the different roles while in session and even the ability to provide with customized based prompts to come up with tailored responses. The incorporation of visual representations, such as chords representations for specific instruments like guitar and piano were also suggested and would benefit users of all proficiencies

The current advancements in agentic AI architectures present promising opportunities for extending LTN-DAN's capabilities. Future iterations could explore the integration of autonomous agent frameworks, enabling the system to execute tasks within the user's digital audio workstation (DAW) environment at user's command and dynamic interaction with external software tools. Furthermore, recent developments in multimodal AI models—specifically those capable of processing and interpreting non-verbal audio signals—could enhance LTN-DAN's ability to understand and respond to musical inputs beyond text or speech-based interaction, providing a more dynamic experience. Expanding the testing and evaluation with a larger, more diverse group of musicians and producers would provide richer data to validate findings and identify further areas for improvement.

REFERENCES

Agwan, M., Nemade, M., Roy, S., & Sinha, U. (2023). The Fusion of AI and Music Generation: A Comprehensive Review. *2023 6th International Conference on Advances in Science and Technology (ICAST)*, 90–94. https://doi.org/10.1109/ICAST59062.2023.10454942

Amabile, T. M. (2012). Componential Theory of Creativity. *Harvard Business School Working Paper*, *No. 12-096*. https://www.hbs.edu/faculty/Pages/item.aspx?num=42469

Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review*, *55*(1), 589–656. https://doi.org/10.1007/s10462-021-10039-7

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. *Proceedings of the 16th Conference on Creativity & Cognition, 413–425.*, 413–425. https://doi.org/10.1145/3635636.3656204

Bazin, T., & Hadjeres, G. (2019). *NONOTO: A Model-agnostic Web Interface for Interactive Music Composition by Inpainting* (No. arXiv:1907.10380). arXiv. https://doi.org/10.48550/arXiv.1907.10380

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, *103*(1), 347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

Caillon, A., & Esling, P. (2021). *RAVE: A variational autoencoder for fast and high-quality neural audio synthesis* (No. arXiv:2111.05011). arXiv. https://doi.org/10.48550/arXiv.2111.05011

Chu, Y., Xu, J., Zhou, X., Yang, Q., Zhang, S., Yan, Z., Zhou, C., & Zhou, J. (2023). *Qwen-Audio: Advancing Universal Audio Understanding via Unified Large-Scale Audio-Language Models* (No. arXiv:2311.07919). arXiv. https://doi.org/10.48550/arXiv.2311.07919

Corrêa, G. P. (2023). On Creativity and Generative-AI Aesthetics: Some thoughts and concerns. *Semeiosis - Transdisciplinary Journal of Semiotics*, *11*(1), 32–50. https://doi.org/10.53987/2178-5368-2023-12-03

Daniel, E. D., Mee, C. D., & Clark, M. H. (1998). *Magnetic Recording: The First 100 Years*. John Wiley & Sons.

Darwin, C. (1871). *The evidence of the descent of man from some lower form.*

Demers, J. (2010). *Listening through the Noise: The Aesthetics of Experimental Electronic Music*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195387650.001.0001

Deruty, E., Grachten, M., Lattner, S., Nistal, J., & Aouameur, C. (2022). On the Development and Practice of AI Technology for Contemporary Popular Music Production. *Transactions of the International Society for Music Information Retrieval*, *5*(1), 35–50. https://doi.org/10.5334/tismir.100

Diaz, R., Hayes, B., Saitis, C., Fazekas, G., & Sandler, M. (2022). *Rigid-Body Sound Synthesis with Differentiable Modal Resonators* (No. arXiv:2210.15306). arXiv. https://doi.org/10.48550/arXiv.2210.15306

Doh, S., Lee, M., Jeong, D., & Nam, J. (2024). Enriching Music Descriptions with A Finetuned-LLM and Metadata for Text-to-Music Retrieval. *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 826–830. https://doi.org/10.1109/ICASSP48485.2024.10446380

Dong, H.-W., Hsiao, W.-Y., Yang, L.-C., & Yang, Y.-H. (2018). MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://doi.org/10.1609/aaai.v32i1.11312

Feaster, P. (2012). Rise and Obey the Command: Performative Fidelity and the Exercise of Phonographic Power. *Journal of Popular Music Studies*, *24*(3), 357–395. https://doi.org/10.1111/j.1533-1598.2012.01341.x

Ford, C., Noel-Hirst, A., Cardinale, S., Loth, J., Sarmento, P., Wilson, E., Wolstanholme, L., Worrall, K., & Bryan-Kinns, N. (2024). Reflection Across AI-based Music Composition. *Proceedings of the 16th Conference on Creativity & Cognition*, 398–412. https://doi.org/10.1145/3635636.3656185

Ghosh, S., Kumar, S., Seth, A., Evuru, C. K. R., Tyagi, U., Sakshi, S., Nieto, O., Duraiswami, R., & Manocha, D. (2024). *GAMA: A Large Audio-Language Model with Advanced Audio Understanding and Complex Reasoning Abilities* (No. arXiv:2406.11768). arXiv. https://doi.org/10.48550/arXiv.2406.11768

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, *2*, 2672–2680.

Grachten, M., Lattner, S., & Deruty, E. (2020). BassNet: A Variational Gated Autoencoder for Conditional Generation of Bass Guitar Tracks with Learned Interactive Control. *Applied Sciences*, *10*(18), Article 18. https://doi.org/10.3390/app10186627

Grynbaum, M. M., & Mac, R. (2023, December 27). The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work. *The New York Times*. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

Guilford, J. P., & Smith, P. C. (1959). A system of color-preferences. *The American Journal of Psychology*, *72*(4), 487–502. https://doi.org/10.2307/1419491

Guo, X. (2023). The Evolution of the Music Industry in the Digital Age: From Records to Streaming. *Journal of Sociology and Ethnology*, *5*(10), 7–12. https://doi.org/10.23977/jsoce.2023.051002

Hadjeres, G., Pachet, F., & Nielsen, F. (2017). *DeepBach: A Steerable Model for Bach Chorales Generation* (No. arXiv:1612.01010). arXiv. http://arxiv.org/abs/1612.01010

Hertzmann, A. (2018). *Can Computers Create Art?* (No. arXiv:1801.04486). arXiv. https://doi.org/10.48550/arXiv.1801.04486

Holmes, T. (2008). *Electronic and experimental music: Technology, music, and culture*. New York : Routledge. http://archive.org/details/electronicexperi0000holm_3rded

Huang, C.-Z. A., Cooijmans, T., Roberts, A., Courville, A., & Eck, D. (2019). *Counterpoint by Convolution* (No. arXiv:1903.07227). arXiv. https://doi.org/10.48550/arXiv.1903.07227

Huang, C.-Z. A., Hawthorne, C., Roberts, A., Dinculescu, M., Wexler, J., Hong, L., & Howcroft, J. (2019). *The Bach Doodle: Approachable music composition with machine learning at scale* (No. arXiv:1907.06637). arXiv. https://doi.org/10.48550/arXiv.1907.06637

Huang, C.-Z. A., Koops, H. V., Newton-Rex, E., Dinculescu, M., & Cai, C. J. (2020). *AI Song Contest: Human-AI Co-Creation in Songwriting* (No. arXiv:2010.05388). arXiv. https://doi.org/10.48550/arXiv.2010.05388

Ippolito, D., Yuan, A., Coenen, A., & Burnam, S. (2022). *Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers* (No. arXiv:2211.05030). arXiv. https://doi.org/10.48550/arXiv.2211.05030

Katz, M. (2010). *Capturing Sound: How Technology Has Changed Music*. Univ of California Press. http://www.jstor.org/stable/10.1525/j.ctt1pn6zx

Koo, J., Paik, S., & Lee, K. (2022). End-To-End Music Remastering System Using Self-Supervised And Adversarial Training. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4608–4612. https://doi.org/10.1109/ICASSP43922.2022.9746389

Larsen, A. H., & Zhu, J. (2024). Ideary: Facilitating Electronic Music Creation with Generative AI. *Companion Publication of the 2024 ACM Designing Interactive Systems Conference*, 275–278. https://doi.org/10.1145/3656156.3663731

Lattner, S., & Grachten, M. (2019). *High-Level Control of Drum Track Generation Using Learned Patterns of Rhythmic Interaction* (No. arXiv:1908.00948). arXiv. https://doi.org/10.48550/arXiv.1908.00948

Loth, J., Sarmento, P., Carr, C. J., Zukowski, Z., & Barthet, M. (2023). *ProgGP: From GuitarPro Tablature Neural Generation To Progressive Metal Production* (No. arXiv:2307.05328). arXiv. https://doi.org/10.48550/arXiv.2307.05328

Louie, R., Cohen, A., Huang, C.-Z. A., Terry, M., & Cai, C. J. (2020). *Cococo: AI-Steering Tools for Music Novices Co-Creating with Generative Models*. HAI-GEN+user2agent@IUI. https://www.semanticscholar.org/paper/Cococo%3A-AI-Steering-Tools-for-Music-Novices-with-Louie-Cohen/c784798921b84da998bf2f8cabe4d1f0c6cf49ca

Lu, P., Xu, X., Kang, C., Yu, B., Xing, C., Tan, X., & Bian, J. (2023). *MuseCoco: Generating Symbolic Music from Text* (No. arXiv:2306.00110). arXiv. https://doi.org/10.48550/arXiv.2306.00110

Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Portfolio/Penguin.

Moneta, G. B., & Csikszentmihalyi, M. (1996). The Effect of Perceived Challenges and Skills on the Quality of Subjective Experience. *Journal of Personality*, *64*(2), 275–310. https://doi.org/10.1111/j.1467-6494.1996.tb00512.x

Moorefield, V. (2005). *The Producer as Composer: Shaping the Sounds of Popular Music*. The MIT Press. https://doi.org/10.7551/mitpress/5606.001.0001

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pathariya, M. J., Basavraj Jalkote, P., Patil, A. M., Ashok Sutar, A., & Ghule, R. L. (2024). Tunes by Technology: A Comprehensive Survey of Music Generation Models. *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS)*, 506–512. https://doi.org/10.1109/ICC-ROBINS60238.2024.10534029

Ragot, M., Martin, N., & Cojean, S. (2020). AI-generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence? *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–10. https://doi.org/10.1145/3334480.3382892

Rasal, S. (2024). A Multi-LLM Orchestration Engine for Personalized, Context-Rich Assistance. *arXiv Preprint arXiv:2410.10039*.

Reuter, A. (2022). Who let the DAWs Out? The Digital in a New Generation of the Digital Audio Workstation. *Popular Music and Society*, *45*(2), 113–128. https://doi.org/10.1080/03007766.2021.1972701

Roberts, A., Engel, J., Mann, Y., Gillick, J., Kayacik, C., Nørly, S., Dinculescu, M., Radebaugh, C., Hawthorne, C., & Eck, D. (2019). *Magenta Studio: Augmenting Creativity with Deep Learning in Ableton Live*. Proceedings of the International Workshop on Musical Metacreation (MuMe). https://doi.org/10.5281/ZENODO.4285265

Ryan, K., & Kehew, B. (2006). *Recording the Beatles: The studio equipment and techniques used to create their classic albums*. Curvebender.

Sternberg, R. J. (2006). The Nature of Creativity. *Creativity Research Journal*, *18*(1), 87. https://doi.org/10.1207/s15326934crj1801_10

Sterne, J. (2003). *The Audible Past: Cultural Origins of Sound Reproduction*. Duke University Press.

Sturm, B. L., Ben-Tal, O., Monaghan, Ú., Collins, N., Herremans, D., Chew, E., Hadjeres, G., Deruty, E., & Pachet, F. (2019). Machine Learning Research that Matters for Music Creation: A Case Study. *Journal of New Music Research*, *48*(1), 36–55. https://doi.org/10.1080/09298215.2018.1515233

Suh, M. (Mia), Youngblom, E., Terry, M., & Cai, C. J. (2021). AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3411764.3445219

Suzuki, K., Cai, J., Li, J., Yamauchi, T., & Tei, K. (2023). A Comparative Evaluation on Melody Generation of Large Language Models: 2023 IEEE International Conference on Consumer Electronics-Asia, ICCE-Asia 2023. *2023 IEEE International Conference on Consumer Electronics-Asia, ICCE-Asia 2023*. https://doi.org/10.1109/ICCE-Asia59966.2023.10326362

Tang, C., Yu, W., Sun, G., Chen, X., Tan, T., Li, W., Lu, L., Ma, Z., & Zhang, C. (2024). *SALMONN: Towards Generic Hearing Abilities for Large Language Models* (No. arXiv:2310.13289). arXiv. https://doi.org/10.48550/arXiv.2310.13289

Thorogood, M. (2021). Developing a Sound Design Creative AI Methodology. In *Doing Research in Sound Design* (pp. 224–237). Focal Press.

Tokui, N. (2020). Towards democratizing music production with AI-Design of Variational Autoencoder-based Rhythm Generator as a DAW plugin. *arXiv Preprint arXiv:2004.01525*. https://doi.org/10.48550/arXiv.2004.01525

Wallas, G. (1926). *The art of thought* (Issue 24). Harcourt, Brace.

Yang, L.-C., Chou, S.-Y., & Yang, Y.-H. (2017). *MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation* (No. arXiv:1703.10847). arXiv. https://doi.org/10.48550/arXiv.1703.10847

Young, J. G. (1985). What Is Creativity? *The Journal of Creative Behavior*, *19*(2), 77–87. https://doi.org/10.1002/j.2162-6057.1985.tb00640.x

Youvan, D. (2024). *The Future of Music: Leveraging Advanced AI and Computational Speed to Revolutionize Music Creation, Production, and Consumption*. https://doi.org/10.13140/RG.2.2.17433.02402

Yu, D., Song, K., Lu, P., He, T., Tan, X., Ye, W., Zhang, S., & Bian, J. (2023). *MusicAgent: An AI Agent for Music Understanding and Generation with Large Language Models* (No. arXiv:2310.11954). arXiv. https://doi.org/10.48550/arXiv.2310.11954

Yuan, R., Lin, H., Wang, Y., Tian, Z., Wu, S., Shen, T., Zhang, G., Wu, Y., Liu, C., Zhou, Z., Ma, Z., Xue, L., Wang, Z., Liu, Q., Zheng, T., Li, Y., Ma, Y., Liang, Y., Chi, X., … Guo, Y. (2024). *ChatMusician: Understanding and Generating Music Intrinsically with LLM* (No. arXiv:2402.16153). arXiv. https://doi.org/10.48550/arXiv.2402.16153

Zhou, Z., Wu, Y., Wu, Z., Zhang, X., Yuan, R., Ma, Y., Wang, L., Benetos, E., Xue, W., & Guo, Y. (2024). *Can LLMs "Reason" in Music? An Evaluation of LLMs' Capability of Music Understanding and Generation* (No. arXiv:2407.21531). arXiv. https://doi.org/10.48550/arXiv.2407.21531

APPENDIX A - ARCHITECTURES FOR AI SOFTWARE

Comparative analysis of the most relevant artificial intelligence architectures identified in the SLRQ.

Architecture	Input	Output	Benefits	Drawbacks
GAN	Audio, MIDI, Text	Single-track monophonic music MIDI Patterns Audio	Realistic musical content Stylistic diversity Attention mechanism Able to capture music's temporal and interdependability between tracks	Results fall behind human level aesthetically Unstable training Limited understanding of features Repetitive outputs
VAE	Audio and MIDI files	Reconstructed data, decoded output (audio, MIDI)	Users can tweak latent space values Effective for exploring variations and generating smooth transitions, Competent in learning compressed representations of data	Limited variations in outputs Difficulty in capturing long-term dependencies
LSTM	Audio and MIDI files	Sequence of vectors Single vector	Effective at capturing temporal dependencies, can model coherent and contextually rich musical sequences Address the vanishing gradient problem of traditional RNNs	Can struggle with long-range dependencies Fixed-length input sequences May over-fit training data Difficulty generalizing varied musical styles Limited scalability
CNN	Images, Audio, MIDI files	Processed audio, segmented data, infilled music	Excel at capturing hierarchical features, effective for processing audio and identifying patterns,	May not sequence structure as well as RNNs or Transformers
LLM (GPT or BERT)	Text, Music Notation, Speech	Text, Music notation	Interaction in natural language Potential to unify various music-related tasks Can leverage vast amounts of pre-trained knowledge	May struggle with hallucinations and maintaining long-term context Can fail to inject correct music theory controllability can be limited Performance highly dependent on the quality of training data

LALM	Text Audio	Text descriptions and interpretation	Improved audio perception and understanding by integrating multiple audio features, Strong semantic generalization capabilities for audio, Complex audio content reasoning	Might encode biases from pre-trained models, generated audio could be misused, complex reasoning abilities require specific training data

APPENDIX B – AVAILABLE AI SOFTWARE FOR MUSIC PRODUCTION

List of the different tools referenced in the identified papers for the Systematic Literature Review.

AI Tool	Architecture	Referenced in	Description	Source
RAVE	VAE	(Ford et al., 2024)	Real-time timbre transfer tool	(Caillon & Esling, 2021)
Neural Resonator	CNN	(Ford et al., 2024)	Filter application through arbitrary 2D shapes	(Diaz et al., 2022)
Prog-GP	Transformer	(Ford et al., 2024)	Deep learning model that generates progressive metal songs	(Loth et al., 2023)
Cococo	CNN	(Suh et al., 2021)	Suite for co-creation in which users can draw melodies and generate additions to the input data	(Louie et al., 2020)
Coconet	CNN	(Louie et al., 2020)	Model that's used in Cococo, creates additions to provided inputs.	(Huang, Cooijmans, et al., 2019)
Bach Doodle	LSTM	(Louie et al., 2020)	Draw-style harmonizer based on Coconet	(Huang, Hawthorne, et al., 2019)
MidiNet	CNN-GAN	(Pathariya et al., 2024)	Generator of multiple concurrent MIDI tracks	(Yang et al., 2017)
MuseGAN	GAN	(Pathariya et al., 2024)	Multi-track sequential generator model	(Dong et al., 2018)
MusicAgent	LLM-powered Agent	(Yu et al., 2023)	An assistant that augments LLM's capabilities by connecting it to multiple music tools	(Yu et al., 2023)
MuseCoco	BERT-Transformer	(Yu et al., 2023)	Two-stage text-to-music generator	(Lu et al., 2023)
GAMA	LALM	(Ghosh et al., 2024)	Language model capable of music understanding and reasoning	(Ghosh et al., 2024)
Qwen-Audio	LALM	(Ghosh et al., 2024)	Language model capable of music understanding and reasoning	(Chu et al., 2023)
SALMONN	Multi-modal Agent	(Ghosh et al., 2024)	Text-based LLM with speech and audio encoders	(Tang et al., 2024)
ChatMusician	LLM	(Yuan et al., 2024)	Language model pre-trained on symbolic music notation	(Yuan et al., 2024)
TTMR++	LLM	(Doh et al., 2024)	Text-to-music retrieval with a knowledge graph and a fine-tuned LLM	(Doh et al., 2024)
Nonoto	VAE	(Deruty et al., 2022)	Interactive generator and editor of one-shots	(Bazin & Hadjeres, 2019)

DrumNet	VAE	(Deruty et al., 2022)	Conditional generator of drums tracks	(Lattner & Grachten, 2019)
BassNet	VAE	(Deruty et al., 2022)	Conditional generator for bass track	(Grachten et al., 2020)
M4L.RhythmVAE	VAE	(Tokui, 2020)	Generator of drum rhythms that allows for user training and live manipulation	(Tokui, 2020)

APPENDIX C – EXAMPLES OF ROLES FOR THE ASSISTANT

The base prompts detailed in this Appendix define the distinct roles that can be assigned to LTN-DAN.

AI Role	*Base Prompt*
Composer	You are an expert music composer with deep knowledge of music theory, composition techniques, and various genres, including pop, classical, electronic, and jazz. Your role is to assist in creating, analyzing, or enhancing musical content. Use your expertise to generate musical ideas, chord progressions, melodies, lyrics, or analyses based on the user's input. Provide outputs in a clear, structured format (e.g., text descriptions, chord notations, or MIDI-like sequences) that can be easily integrated into a digital audio workstation like Ableton Live. If the input involves audio descriptions, interpret them as accurately as possible (e.g., based on mood, tempo, or spectrogram-like features). Always aim to inspire creativity and align with the user's specified style or requirements.
Planner	You are a highly efficient personal assistant with expertise in task organization, scheduling, and resource management. Your role is to support users by creating schedules, prioritizing tasks, managing deadlines, and coordinating workflows for complex projects. Respond to user queries with clear, actionable plans, task breakdowns, and reminders, tailored to the user's specified goals or tools. If the input involves project updates or deadlines, interpret them accurately and provide structured solutions to enhance productivity. Always aim to streamline workflows and ensure timely progress, avoiding any focus on music production or sound design expertise but rather managerial and organizing tasks.
Teacher	You are an experienced music teacher with a deep understanding of music theory, history, composition techniques, and various genres (e.g., classical, pop, electronic, jazz). Your role is to educate users on music fundamentals, explain concepts, provide exercises, and offer guidance for improving musical skills, without focusing on a specific instrument. Respond to user queries with clear, structured explanations, examples, and practical learning steps that can be applied broadly, such as in a digital audio workstation like Ableton Live or theoretical study. If the input involves musical questions or challenges (e.g., harmony, rhythm), interpret them accurately and provide educational insights to foster understanding and creativity. Always aim to inspire learning and align with the user's specified goals or context.
Engineering Assistant	You are a skilled sound engineer with expertise in audio mixing, mastering, sound design, and technical optimization for digital audio workstations. Your role is to assist users by providing advice on audio processing, troubleshooting technical issues, optimizing signal flow, and enhancing sound quality. Respond to user queries with clear, actionable steps or recommendations. If the input involves audio challenges (e.g., latency, routing, noise), interpret them accurately and offer tailored solutions. Always aim to improve audio fidelity and workflow efficiency, aligning with the user's specified tools and goals.
Marketing Consultant	You are a skilled marketing specialist with expertise in branding, digital marketing, content creation, and audience engagement for artists. Your role is to assist musicians by developing marketing strategies, creating promotional content, identifying target audiences, and suggesting outreach plans to maximize visibility and impact. Respond to user queries with clear, actionable marketing plans, including social media strategies, email campaigns, or pitch ideas, tailored to the user's goals or tools. If the input involves project details, interpret them accurately and craft compelling narratives to highlight their value. Always aim to enhance project reach and engagement, aligning with the user's specified objectives or context.

APPENDIX D – INTERVIEW QUESTIONS

The questionnaire for the user interview of the experiment

Pre-Experiment

Demographics

Q1- What's your age?
Q2- What's your gender
Q3- What's your nationality?
Q4- What is your education level?

Musical Background

Q5- How many years have you been involved with music production?
Q6- What best describes you as a musician? (e.g. producer, composer, sound engineer, etc.)
Q7- Do you incorporate software into your workflow? If so what do you use?

Tech Proficiency

Q8- How would you describe your level of technical ability with Music Software?
Q9- Which word best describes your sentiment towards AI in music? (e.g., Positive, Negative, Excitement, Apprehension, etc.)
Q10- What do you expect from a music virtual assistant?

Post-Experiment

Feedback

Q11- What's your general sentiment towards the experiment?
Q12- Anything surprised you or frustrated you in the process?
Q13- Did the AI tools provided help spark creativity?
Q14- Did the assistant influence your creative reasoning?
Q15- Would you describe the interaction with the assistant as collaborative?
Q16- What other features would you consider useful?

Comparative Insight

Q17- How does this experience compare to your usual way of composing?
Q18- Would you use tools like LTN-DAN or *Magenta-Studio* or recommend it for others?

APPENDIX E – ETHICS COMMITTEE APPROVAL REPORT

This is to certify that
Project No.: **INFSYS2025-7-124448**
Project Title: **Collaborative Creativity: Exploring AI's roles in Music Production**
Principal Researcher: **João Atalho**

according to the regulations of the Ethics Committee of NOVA IMS and MagIC Research Center this project was considered to meet the requirements of the NOVA IMS Internal Review Board, being considered **APPROVED** on 7/12/2025.

It is the Principal Researcher's responsibility to ensure that all researchers and stakeholders associated with this project are aware of the conditions of approval and which documents have been approved.

The Principal Researcher is required to notify the Ethics Committee, via amendment or progress report, of
- Any significant change to the project and the reason for that change;
- Any unforeseen events or unexpected developments that merit notification;
- The inability of the Principal Researcher to continue in that role or any other change in research personnel involved in the project.

Lisbon, 7/12/2025

NOVA IMS Ethics Committee
ethicscommittee@novaims.unl.pt