Mozart AI: Browser-Based AI Music Co-Production

Pascual Merita
Arthos
pascual@arthos.inc

Petr Ivan
Arthos
petr.ivan@arthos.inc

Immanuel Rajadurai
Arthos
immanuel@arthos.inc

Sundar Arvind
Arthos
sundar.arvind@arthos.inc

Arjun S. Khanna
Arthos
arjunskhanna@arthos.inc

Abstract

We present Mozart AI[1], a browser-based digital audio workstation (DAW) augmented with artificial intelligence. Mozart AI acts as a creative co-producer, empowering users to generate loops, melodies, arrangements, lyrics, and mixdowns via natural language interactions. By architecting AI capabilities as first-class DAW components rather than external dependencies, Mozart AI eliminates context-switching overhead and enables deeper iterative refinement of generated content.

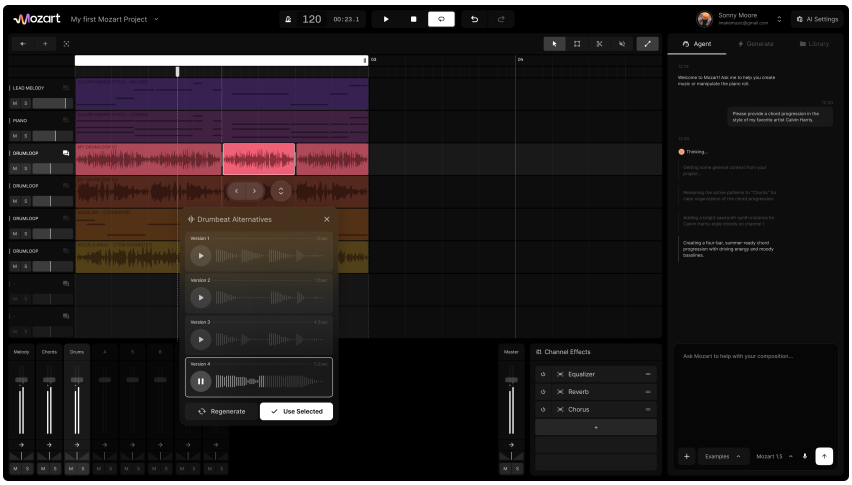

Figure 1: Mozart AI's DAW interface, demonstrating track-based audio in-painting.

1 Introduction

The landscape of artificial intelligence in music production has evolved rapidly, yet remains fundamentally fragmented across different user experience levels and production workflows. Current AI music tools target distinct extremes of the production spectrum: complete beginners benefit from end-to-end song generation platforms like Suno AI, Udio and others Gong et al. (2025); Yuan et al. (2025), which enable users to generate full tracks from simple text prompts. Meanwhile, professional producers rely on specialised AI plugins and models integrated into traditional Digital

[1] https://getmozart.ai

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

Audio Workstations (DAWs) for specific tasks like melody generation Mixed in Key (2024); Guo and Dixon (2025); Qu et al. (2024), loop creation Copet et al. (2024); Agostinelli et al. (2023), semantic sample search Elizalde et al. (2022), or mixing iZotope, by Native Instruments (2024).

This polarisation leaves intermediate music producers without adequate support, as they often require tools from both extremes. End-to-end song generation platforms lack DAW integration entirely, while DAW plugins fail to ffer seamless AI-assisted workflows, forcing users to context-switch between external services and their primary production environemnt

Recent attempts to bridge this gap through web-based solutions have shown promise but fallen short of comprehensive integration. WavTool pioneered the concept of an AI-enhanced browser-based DAW with GPT-4 integration, featuring a "Conductor" assistant capable of generating MIDI and controlling effects through natural language. However, the platform was discontinued in November 2024 before additional AI components could be integrated. Other newer solutions such as Riff, offer semantic search, conductor assistance, and loop generation, but lack sophisticated digital signal processing capabilities and beginner-friendly workflows for end-to-end song generation and editing.

2 Mozart AI

We present Mozart AI, a browser-based digital audio workstation that fundamentally reimagines the relationship between artificial intelligence and music production. Rather than treating AI as an external tool or plugin, Mozart AI architects AI features as first-class DAW components, enabling seamless transitions between traditional music production techniques and AI-assisted generation. This approach supports both “top-down” workflows familiar to beginners—such as prompt-to-song generation—and “bottom-up” professional workflows including loop matching, melody completion, automated mixing, and collaborative arrangement.

Mozart AI’s technical architecture encompasses several novel contributions:

- A >1B parameter custom autoregressive transformer for MIDI-based next-bar prediction and inpainting, trained on a combination of public Project Los Angeles (2024) and proprietary datasets;
- A fine-tuned CLAP model for semantic audio search with planned 3D exploration interfaces, similar to XLN Audio (2024);
- A diffusion-based loop generation model conditioned on both text and audio inputs, with multi-granularity descriptions generated through multimodal large language models;
- A diffusion-based model for text-conditioned synthesiser generation, similar to Fadr (2024). It currently generates a short audio samples that are sampled across the frequency spectrum, but a possible extension would be to generate the actual synthesiser parameters directly.
- Proprietary stem separation and audio-to-MIDI conversion;
- A main-agent and track-based agent system utilizing Model Control Protocol (MCP) Anthropic (2024) architecture to provide contextual tool access to all the DAW, including specialized DSP workflows for humming-based input;
- A partnership with ElevenLabs ElevenLabs (2024) for end-to-end song generation and voice agent integration.

While we leverage third-party services for certain functions—such as end-to-end song generation and audio inpainting through partnerships with providers like ElevenLabs—the core innovation lies not solely in our individual models but in their deep integration within a unified production environment. This integration preserves creative flow by eliminating context-switching overhead and enables iterative refinement of AI-generated content within the familiar DAW paradigm.

References

Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. 2023. MusicLM: Generating Music From Text. arXiv:2301.11325 [cs.SD] https://arxiv.org/abs/2301.11325

Anthropic. 2024. Model Context Protocol. https://modelcontextprotocol.io/ Accessed August 2025.

Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2024. Simple and Controllable Music Generation. arXiv:2306.05284 [cs.SD] https://arxiv.org/abs/2306.05284

ElevenLabs. 2024. ElevenLabs Music. https://elevenlabs.io/music Accessed August 2025.

Benjamin Elizalde, Soham Deshmukh, Mahmoud Al Ismail, and Huaming Wang. 2022. CLAP: Learning Audio Concepts From Natural Language Supervision. arXiv:2206.04769 [cs.SD] https://arxiv.org/abs/2206.04769

Fadr. 2024. SynthGPT. https://fadr.com/synthgpt Accessed August 2025.

Junmin Gong, Sean Zhao, Sen Wang, Shengyuan Xu, and Joe Guo. 2025. ACE-Step: A Step Towards Music Generation Foundation Model. arXiv:2506.00045 [cs.SD] https://arxiv.org/abs/2506.00045

Zixun Guo and Simon Dixon. 2025. Moonbeam: A MIDI Foundation Model Using Both Absolute and Relative Music Attributes. arXiv:2505.15559 [cs.SD] https://arxiv.org/abs/2505.15559

iZotope, by Native Instruments. 2024. Neutron 5. https://www.izotope.com/en/products/neutron Accessed August 2025.

Mixed in Key. 2024. Captain Melody. https://mixedinkey.com/captain-plugins/captain-melody/ Accessed August 2025.

Project Los Angeles. 2024. Los Angeles MIDI Dataset. https://huggingface.co/datasets/projectlosangeles/Los-Angeles-MIDI-Dataset Accessed August 2025.

Xingwei Qu, Yuelin Bai, Yinghao Ma, Ziya Zhou, Ka Man Lo, Jiaheng Liu, Ruibin Yuan, Lejun Min, Xueling Liu, Tianyu Zhang, Xinrun Du, Shuyue Guo, Yiming Liang, Yizhi Li, Shangda Wu, Junting Zhou, Tianyu Zheng, Ziyang Ma, Fengze Han, Wei Xue, Gus Xia, Emmanouil Benetos, Xiang Yue, Chenghua Lin, Xu Tan, Stephen W. Huang, Jie Fu, and Ge Zhang. 2024. MuPT: A Generative Symbolic Music Pretrained Transformer. arXiv:2404.06393 [cs.SD] https://arxiv.org/abs/2404.06393

Riff. 2024. Riff. @misc{wavtool,title={{WavTool,acquiredbySunoAI}},author={{WavTool}},year={2024},url={https://wavtool.com/},note={AccessedAugust2025}} Accessed August 2025.

Suno AI. 2024. Suno AI Music Generation Platform. https://suno.ai Accessed August 2025.

Udio. 2024. Udio: AI Music Creation. https://udio.com Accessed August 2025.

WavTool. 2024. WavTool, acquired by Suno AI. https://wavtool.com/ Accessed August 2025.

XLN Audio. 2024. XO. https://www.xlnaudio.com/products/xo Accessed August 2025.

Ruibin Yuan, Hanfeng Lin, Shuyue Guo, Ge Zhang, Jiahao Pan, Yongyi Zang, Haohe Liu, Yiming Liang, Wenye Ma, Xingjian Du, Xinrun Du, Zhen Ye, Tianyu Zheng, Yinghao Ma, Minghao Liu, Zeyue Tian, Ziya Zhou, Liumeng Xue, Xingwei Qu, Yizhi Li, Shangda Wu, Tianhao Shen, Ziyang Ma, Jun Zhan, Chunhui Wang, Yatian Wang, Xiaowei Chi, Xinyue Zhang, Zhenzhu Yang, Xiangzhou Wang, Shansong Liu, Lingrui Mei, Peng Li, Junjie Wang, Jianwei Yu, Guojian Pang, Xu Li, Zihao Wang, Xiaohuan Zhou, Lijun Yu, Emmanouil Benetos, Yong Chen, Chenghua Lin, Xie Chen, Gus Xia, Zhaoxiang Zhang, Chao Zhang, Wenhu Chen, Xinyu Zhou, Xipeng Qiu, Roger Dannenberg, Jiaheng Liu, Jian Yang, Wenhao Huang, Wei Xue, Xu Tan, and Yike Guo. 2025. YuE: Scaling Open Foundation Models for Long-Form Music Generation. arXiv:2503.08638 [eess.AS] https://arxiv.org/abs/2503.08638