

Spawning the Groove: OBSIDIAN-Neural as a Live Performance Instrument for AI-Mediated Improvisation

Anthony Charretier
Independent Researcher & Musician
Saint-Étienne, France
anthony.charretier@etik.com

Abstract

This paper presents OBSIDIAN-Neural, a VST3 plugin whose core framework is open-source, integrating a closed-source local inference module running Stable Audio 3 Medium for live electronic music performance. Running 100% locally on consumer-grade CPU hardware, the system generates audio samples in as little as 10 seconds from text prompts, directly within any Digital Audio Workstation, or standalone with Ableton Link support. Rather than positioning AI as a compositional replacement, OBSIDIAN-Neural recasts generative AI as a performance instrument. For this demonstration, the plugin is used live within Bitwig Studio alongside a MIDI controller and a complementary VST synthesizer, combining on-the-fly AI sample generation with conventional performance elements. Presented at AES AIMLA 2025 (Queen Mary University London) and at AES Europe 2026 (Technical University of Denmark, Copenhagen), the project challenges dominant notions of AI music as automated or cloud-dependent production, proposing instead a model of mediated musicianship rooted in embodied, improvisatory practice.

1 INTRODUCTION

The so-called generative turn in music has largely been framed around post-hoc production: AI systems generating finished tracks, stems, or loops that composers later arrange and edit. This model, exemplified by platforms such as Suno and Udio (Udio, 2024), presupposes a separation between generation and performance - the machine produces, the human curates.

OBSIDIAN-Neural proposes a different relationship. By embedding generative AI inference directly into the VST3 plugin format - the standard interface of professional DAW environments - the system allows a performer to trigger AI sample generation mid-performance, weaving the latency and unpredictability of the model into the fabric of live improvisation. Generation becomes gesture. The inference window, rather than a technical limitation, functions as a structural constraint that shapes performance decisions: the musician must anticipate, prepare, and commit.

This paper accompanies a live performance for the AIMC 2026 concert programme. The performance, *Substrat*, is a 15-minute solo improvisation in the psychedelic electro-techno idiom, performed at 127 BPM, combining a MIDI controller and a complementary VST synthesizer with real-time AI sample generation via OBSIDIAN-Neural, all run live within Bitwig Studio.

2 PRACTICE FIELD: AI IN LIVE ELECTRONIC PERFORMANCE

The question of liveness and mediation in electronic music performance has a long critical history (Small, 1998; Kim-Boyle, 2008). From laptop performance to modular synthesis, practitioners have continually negotiated the tension between technological mediation and the perceived authenticity of live sound production.

The introduction of machine learning into this space has predominantly taken two forms. The first is *generative*

accompaniment: AI systems that respond to a performer in real time, producing complementary melodic or rhythmic content. The second is *offline augmentation*: the use of AI tools to generate material that is subsequently integrated into a prepared performance context.

OBSIDIAN-Neural occupies a third position: *deferred generation as instrument*. The system does not respond in real time to the performer's actions; rather, it receives a prompt and returns a generated audio sample within a predictable latency window. This predictability is itself musically productive. Like a hardware sampler requiring manual loading, the workflow imposes a preparation phase that structures improvisation rather than eliminating it.

The use of consumer-grade, locally-run inference (Stable Audio 3 Medium (Stability AI, 2026)) further distinguishes the system from cloud-dependent approaches. No network connection is required; the model runs on the same CPU as the DAW. This is not merely a technical detail: it has implications for sovereignty, reliability in live contexts, and the relationship between the performer and the generative system. In a club or festival setting, where internet access is often unavailable or unreliable, local inference is not a stylistic choice but a practical necessity for any performer wishing to generate sounds with AI during a live set.

3 SYSTEM DESCRIPTION AND WORKFLOW

OBSIDIAN-Neural is implemented as a VST3 plugin using the JUCE C++ framework. The plugin's core framework is open-source; the module handling local inference of Stable Audio 3 Medium is closed-source. The model is loaded directly within the plugin via ONNX Runtime, avoiding any external server or network dependency. The model generates stereo audio samples of variable duration from text prompts. A standalone version of the project is also available, supporting Ableton Link for tempo synchronization with other software and hardware.

The performance workflow for *Substrat* is deliberately non-prescriptive. OBSIDIAN-Neural operates as the primary sound source: rhythmic elements, pads, and bass lines are generated via text prompts and distributed across the plugin's 8 stereo output tracks, each comprising 4 pages of generated content. Each track features its own onboard effects chain and a glitch sequencer, along with 8 modulators per track whose targets can be manually assigned, allowing generated material to be reshaped and reprocessed directly within the plugin before being routed independently into Bitwig Studio for further processing. A MIDI controller and a complementary VST synthesizer enter the mix as additional layers, played and shaped alongside the generated material depending on performance choices made in the moment.

There is no fixed order of operations. A performance may begin with an AI-generated rhythmic foundation, or with the MIDI-driven VST synthesizer, or with both simultaneously. The performer shapes each generated element through the plugin's onboard effects chain and glitch sequencer, further reprocessed through Bitwig Studio's own effects, treating the raw AI output as raw material rather than finished sound. Text prompts are submitted throughout the performance, anticipating textural transitions whose exact character remains unknown until generation completes.

The system was first performed publicly at the AES AIMLA 2025 workshop, Queen Mary University of London (Charretier, 2025), where it was presented as a Late Breaking Demo Paper. It was subsequently presented at the AES Europe 2026 Convention, held at the Technical University of Denmark, Copenhagen. The plugin has since accumulated 239+ GitHub stars and received coverage in 8 countries across specialist music technology press.

4 ARTISTIC AND THEMATIC CONTEXT

Substrat engages directly with this year's conference theme, *The Generative Turn: Mediated Musicianship in a Hyper-reproductive Age*. The performance does not attempt to hide or aestheticise the AI generation process; rather, it foregrounds the seams. Generated samples retain their characteristic timbral signatures - the slightly inhuman attack transients, the spectral smoothness of diffusion-model audio - and are placed in deliberate contrast with the MIDI-driven VST synthesizer layer.

This friction is intentional. The performance asks: what does it mean to improvise *with* a generative system that cannot hear you, that responds only to language, and that takes ten seconds to answer? The answer proposed by *Substrat* is not anxiety but attunement - a practice of anticipatory listening, of shaping prompts as one shapes a phrase, of treating latency as rhythm.

In this sense, OBSIDIAN-Neural engages with the broader critical discourse on mediated musicianship (Small, 1998), while remaining firmly grounded in the specific aesthetic traditions of psychedelic electronic music: repetition, transformation, the dissolution of foreground and background.

5 PROGRAM NOTES

Substrat (2025–2026) is a solo live performance for MIDI controller, VST synthesizer, and AI sample generation, run entirely within Bitwig Studio. Running at 127 BPM in an open improvisatory structure, the work explores the

threshold between machine-generated texture and human rhythmic intention. Samples generated by OBSIDIAN-Neural during performance - drawn from text prompts typed in real time - emerge from the mix as found objects: slightly wrong, slightly familiar, always provisional.

Duration: 15 minutes.

ETHICS STATEMENT

OBSIDIAN-Neural uses Stable Audio 3 Medium, a model released under the Stability AI Community License for commercial and non-commercial use. No additional training data was collected or used in the preparation of this performance. The model's training data provenance is governed by Stability AI's published terms. The performer is the sole author of all prompts and performance decisions.

ACKNOWLEDGEMENTS

The author thanks the AES AIMLA 2025 programme committee at Queen Mary University of London, and the AES Europe 2026 programme committee at the Technical University of Denmark, Copenhagen, for the opportunity to present earlier versions of this work.

REFERENCES

- Charretier, A. (2025). OBSIDIAN-Neural: Real-time AI music generation plugin for live performance. In *Proceedings of the AES International Conference on Artificial Intelligence and Machine Learning for Audio*, London, UK. Late Breaking Demo Paper.
- Kim-Boyle, D. (2008). Network musics: Play, engagement and the democratization of performance. In *Proceedings of the 2008 International Computer Music Conference*, Belfast, Ireland.
- Small, C. (1998). *Musicking: The Meanings of Performing and Listening*. Wesleyan University Press.
- Stability AI (2026). Stable audio 3 medium. <https://github.com/Stability-AI/stable-audio-3>. Accessed 2026.
- Udio (2024). Udio: Ai music generation platform. <https://www.udio.com>. Accessed 2026.