

An Open Agentic System for Multimodal Music Understanding and Generation

Emmanouil Karystinaios^{1,2}

{firstname}.{lastname}@jku.at

¹ Institute of Computational Perception, Johannes Kepler University Linz, Austria ² Human-centered Artificial Intelligence (HCAI), Linz Institute of Technology, Austria.







Overview

WeaveMuse is an open-source agentic system for music understanding & generation:

- ☐ Bridges text, symbolic notation, and audio in one workflow
- ☐ Supports analysis, transformation, and creative generation
- ☐ Efficient & accessible: runs locally or via hosted APIs
- ☐ Enables flexible, multimodal interaction for musicians & researchers
- ☐ Build on top of Smolagents
- ☐ WeaveMuse still in development with new tools coming up, supporting the innovation of the Open MIR community

System Preview WeaveMuse | Audio Files Music Agent Chat music piece generated by the notagen function is in ABC notation format. It captures the essence of Frédéric Chopin's style, characterized by intricate and flowing melodies, rich harmonies, and expressive dynamics. The piece is designed to evoke a **Uploaded Files** sense of peace and introspection, making it suitable for solo piano performance. The composition adheres to the Romantic period, No audio files uploaded emulating Chopin's renowned nocturnes and preludes, which are known for their lyrical beauty and emotional depth. Clear Audio Files 3. Additional context (if relevant): Example Prompts The notagen function has successfully generated a composition that aligns with the specified parameters. The piece can be further Compose a peaceful piano piece in the style of Chapir refined or converted to various formats, such as PDF, for visual display and performance. The use of ABC notation ensures full Analyze the harmony in my uploaded audio conversion capabilities and supports various musical styles and instrumentation. Generate a slow blues guitar solo Generated Sheet Music (3 pages) Explain modal harmony Check the download files below for complete score and audio Audio Support Supported formats: WAV, MP3, FLAC, OGG Max duration: 10 minutes Sample rates: 16kHz - 48kHz Send Ask me anything **Download Generated Files** Ask me anything about music, upload audio, or request compositions. Clear Chat Click the Send button or use CTRL+Shift. Generated Files (PDF, Audio, MIDI, etc.) 99.9 KB 1 × 20250818_105944_Romantic_Chopin_Piano.pdf ₽ Upload Audio File Agent Generated Audi 291.4 KB 1 > 20250818_105944_Romantic_Chopin_Piano.xml 16.9 KB 1 × 20250818_105944_Romantic_Chopin_Piano.mid 20250818_105944_Romantic_Chopin_Piano.mp3 1.5 mB 1 > 20250818_105944_Romantic_Chopin_Piano.abc 5.4 KB 1 × Drop Audio Here Sheet Music PDF Click to Upload (i) [1x] Generated Sheet Music Analyze This Audio M Download 1 0 - User Caption: Compose a piece in the style of Chopin

Pipeline

- **System:** Uses Notagen Tool, generates ABC, transforms to

audio, musicXML and PDF and returns the output files.

Functional View

(What it Does)

- Accepts multimodal inputs: text,
 symbolic scores (abc for now), audio
- Understands music through analysis tools (harmony, rhythm, structure)
- Generates new material: symbolic notation, audio, hybrid transformations
- Transforms across modalities
 (e.g., score → audio, text → score)

Architectural View

(How it Works)

- User Input Layer: text prompts, uploaded scores, audio files
- Manager Agent: plans, selects tools, and verifies results (perceive-plan-act loop)
- Specialized Tools:
 - o AudioFlamingo, NotaGen, StableAudio
- Score renderers → PDF/visual outputs
- Output Layer: multimodal results (scores, audio, structured metadata)

Deployment / Resource View

(Where it Runs)

- Lightweight local setup: efficient execution on CPU/GPU laptops
- Hybrid mode: local orchestration + hosted APIs for heavy models
- Efficiency optimizations: quantization, caching, memory offloading, dynamic precision, lazy tool loading.
- Scalable setup: reproducible deployment across hardware tiers

Resources

Implemented Tools and Models

- ☐ ChatMusician text-to-symbolic music generation
- □ NotaGen symbolic generation in ABC notation
- ☐ Stable Audio Open text-to-audio synthesis
- □ sympy / Partitura / Parsers parsing and symbolic feature analysis
- ☐ Score Visualization symbolic-to-PDF rendering **EngravingGNN**
- ☐ Audio Understanding spectral, rhythmic, context and harmonic analysis (QwenAudio and Audio Flamingo)
- ☐ Gradio Interface



Read Our Paper



Github Repo: manoskary/weavemuse

