Hack-Nation

Global AI Hackathons & Venture Incubation Program

in collaboration with MIT Club of Northern California

# Music to My Ears - Building the AI Backbone of the Super Bowl Halftime Music Industry

VC Track

---

## 1. Goals and Motivation

Generate ideas 100× faster. Produce music played 100× more often. Deliver experiences with 100× better vibes. This challenge is an invitation to become an artist — not an app builder, not a prompt engineer — an artist of a new kind. But more than that, it is an invitation to think bigger: not just becoming a single creator, but becoming part of the generative engine that powers how music itself is produced and discovered.

For most of history, music came from instruments. Then studios. Then laptops. Now it can come from life itself — a photo, a voice note, a memory, a moment never written down. Multimodal signals are becoming instruments, and AI systems are increasingly the interpreters translating them into sound.

At planetary scale, music is already shaped by invisible infrastructure. Algorithms influence discovery, generative tools accelerate creation, and audience signals shape evolution — quietly determining what rises through the ecosystem and reaches global cultural stages like the Super Bowl halftime show. Behind those moments sits an enormous economy: recorded music generates roughly $30B annually, the broader industry exceeds $100B, and forecasts suggest it may approach $200B within the next decade. Meanwhile, more than 100,000 songs are uploaded daily, competing for attention where discovery — not creation — is the bottleneck.

What reaches the top is not just art. It is infrastructure.

This challenge explores the frontier of building that infrastructure. Rather than producing isolated music outputs, participants are encouraged to prototype components of the **AI backbone of the music industry** — systems that ingest multimodal signals, generate content, adapt through interaction, and scale creative workflows. You are not only becoming an artist. You are designing the engine that could empower millions of artists.

---

## 2. Creative Objective

Music communicates what language cannot. A song can preserve memory, encode emotion, and transform experience into something shared. Your goal is to design systems capable of translating multimodal inputs — images, text, voice, video, motion — into music that retains narrative meaning and emotional coherence.

This is about inventing the **multimodal artist**, but also exploring what happens when that artist becomes a platform capability rather than an individual role. Consider the creative flywheel: signals become generation, generation meets audiences, feedback reshapes models, and iteration accelerates cultural emergence. Your system should engage with some portion of this loop.

The moonshot is amplification — exploring how musical interpretation can multiply resonance, engagement, or emotional reach relative to the originating moment.

---

## 3. Technical Direction — Your Instruments

Participants are free to use any stack, but the following directions provide a strong foundation.

### Generative Music Models

**MusicGen (Meta)**
https://huggingface.co/docs/transformers/model_doc/musicgen

MusicGen provides a controllable baseline for text- or audio-conditioned generation. Participants are encouraged not merely to invoke pretrained models but to adapt and direct them.

Relevant techniques include:

- **LoRA / PEFT fine-tuning** for lightweight specialization
- **Cross-modal conditioning pipelines** mapping external signals into latent musical control variables

- **Instruction-following evaluation** measuring semantic alignment
- **Reinforcement learning from user feedback** where listener responses guide generation updates
- **Preference-based optimization loops** shaping model behavior toward perceived resonance

Exploring how controllable, steerable, or responsive your system becomes is central to the challenge.

### Suggested Datasets

MusicCaps — text/music alignment
https://huggingface.co/datasets/google/MusicCaps

VGGSound — audiovisual relationships
https://www.robots.ox.ac.uk/~vgg/data/vggsound/

Other datasets to research: Sound-of-Story (SoS), Free Music Archive (FMA)

Teams are encouraged to curate focused datasets aligned with creative direction. Intentional data often drives stronger stylistic outcomes than scale.

---

## 4. Thinking Beyond Generation — Tooling & Infrastructure

Participants are encouraged to think not only about outputs, but about tooling primitives that would enable large-scale creative ecosystems. This may include interfaces for multimodal capture, pipelines for conditioning signals, mechanisms for evaluating controllability, feedback aggregation systems, or workflows that allow iterative refinement across user populations.

The goal is to explore what infrastructure might look like if AI-mediated music creation became foundational rather than experimental.

---

## 5. Evaluation Lens

Projects will be evaluated holistically based on:

- Depth of model direction or adaptation
- Effectiveness of multimodal transformation
- Creativity and intentional authorship
- Exploration of tooling or infrastructure concepts
- Evidence of audience engagement signals
- Originality of insight or system framing

We especially value work that surfaces new questions about the role of AI in creative ecosystems.

---

## 6. Why It Matters

We live in a world overflowing with data yet starving for meaning. Music may become the layer that translates raw experience into emotional artifacts — compressing memory into sound and distributing it across networks. Understanding how intelligent systems participate in this transformation is key to shaping future creative economies.

The halftime show symbolizes the visible peak of cultural music influence.
This challenge explores the machinery beneath it.

Build boldly.
Design systems.
Shape the engine.
And take the stage.