

The Ghost in the Keys: A Disklavier Demo for Human-AI Musical Co-Creativity



Louis Bradshaw, Alexander Spangher, Stella Biderman, Simon Colton



Interactive systems for music

Many popular AI models enabling musical co-creation are **asynchronous** and, consequently, are fundamentally at odds with the **embodied, responsive** nature of instrumental performance.

AIM: Adapt a generative model for piano (*Aria*) into a *creatively meaningful tool* by focusing on:

CONTROLLABILITY: **Musicians are willing to cede some control; however, they want to retain creative agency.**

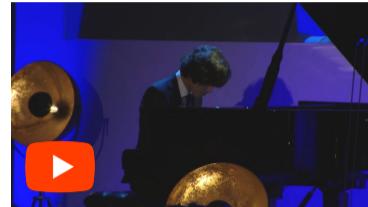
EMBODIMENT: **Enable musicians to interface with AI models directly using their instruments, rather than text prompting.**

RESPONSIVENESS: **Restore familiar feedback loops by removing friction and latency from the co-creative process.**

We introduce ***Aria-Duet***, an interactive system for **real-time musical duets** between a pianist and a generative model, using a Yamaha Disklavier as a shared interface.

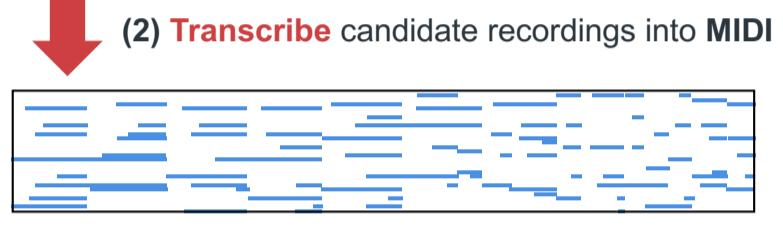
*In this paradigm, the model acts as a real-time, low-latency, and controllable **compositional autopilot** for the pianist.*

How we created Aria-Duet

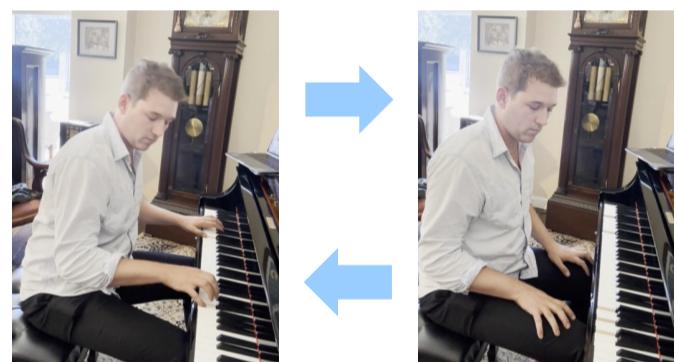


id: jj3A...
id: 9C2v...
id: Heo4...
...

(1) Crawl public piano recordings (~100k hours)

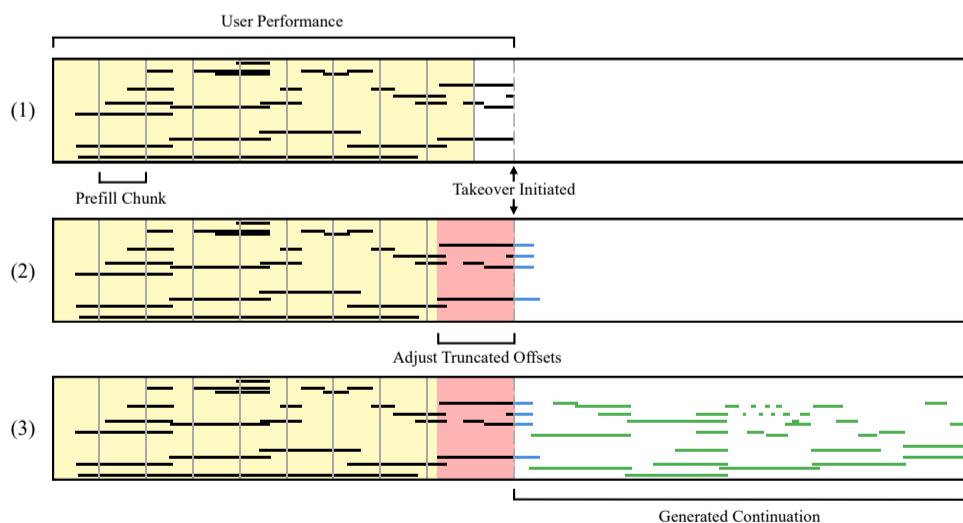


(3) Train generative model via **next-note-prediction** (*Aria*)



(4) Pianist and model take turns performing on Disklavier

Technical Challenges



- To mitigate **prefill-induced transition latency**, we continuously fill chunks into the KV-Cache as the pianist performs.

- At the point of transition, **truncated offsets**, e.g., hanging notes, are speculatively re-calculated and backfilled into the KV-Cache.

- We implement a real-time Disklavier MIDI I/O streaming layer ensuring **accurate playback** without incurring any latency penalties.

Our system achieves remarkably **low transition latency** (time-to-first-note) whilst **preserving generation & playback quality**:

Apple Silicon (**M-Series**): **200-300ms**

Consumer GPU (**RTX 4090/5090**): **10-20ms**

NOT A PIANIST? NO PROBLEM!

Use **playback mode** to trigger the AI takeover from a pre-loaded track with a keystroke.

More information **Dataset** 😊

