Generative AI Music Beyond Text-to-Music

Nick Bryan
Head of Music AI
Adobe Research

Adobe Research

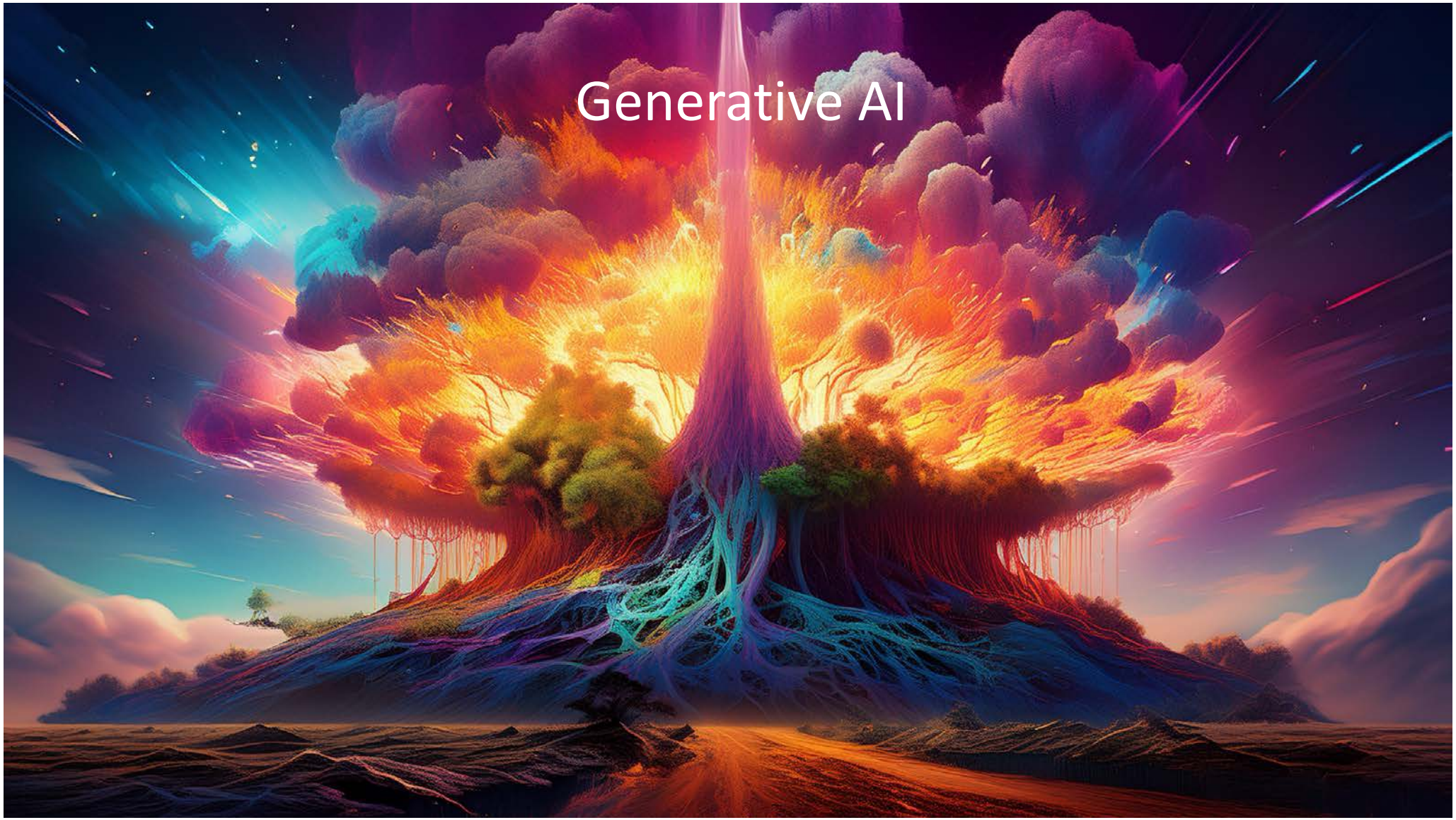
Audio

Human Computer Interaction

Artificial Intelligence & Machine Learning

Computer Vision, Imaging & Video

Natural Language Processing


Generative AI

Why is Generative Music AI Interesting?

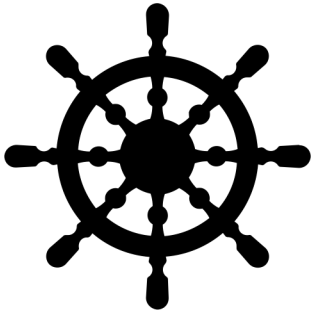

“The Act of Creation”


“The Act of Creation?”

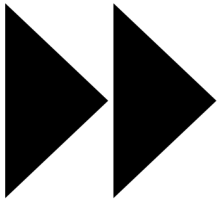

Control


Iteration

Technical Challenges in Music Gen AI



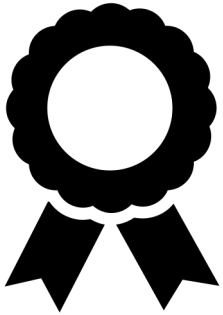
1. Control



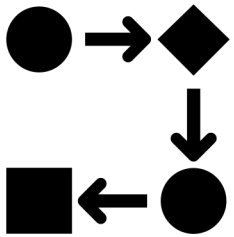
2. Speed

3. Efficiency

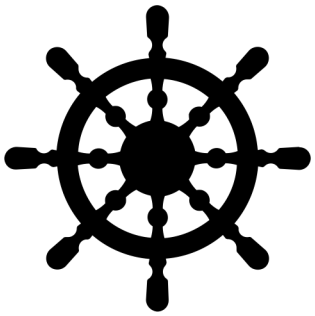
4. Data



5. Quality



6. Workflows

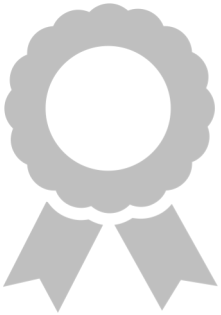


1. Control

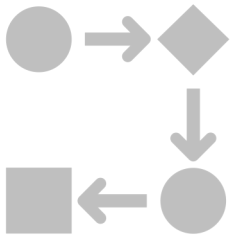
2. Speed

3. Efficiency

4. Data

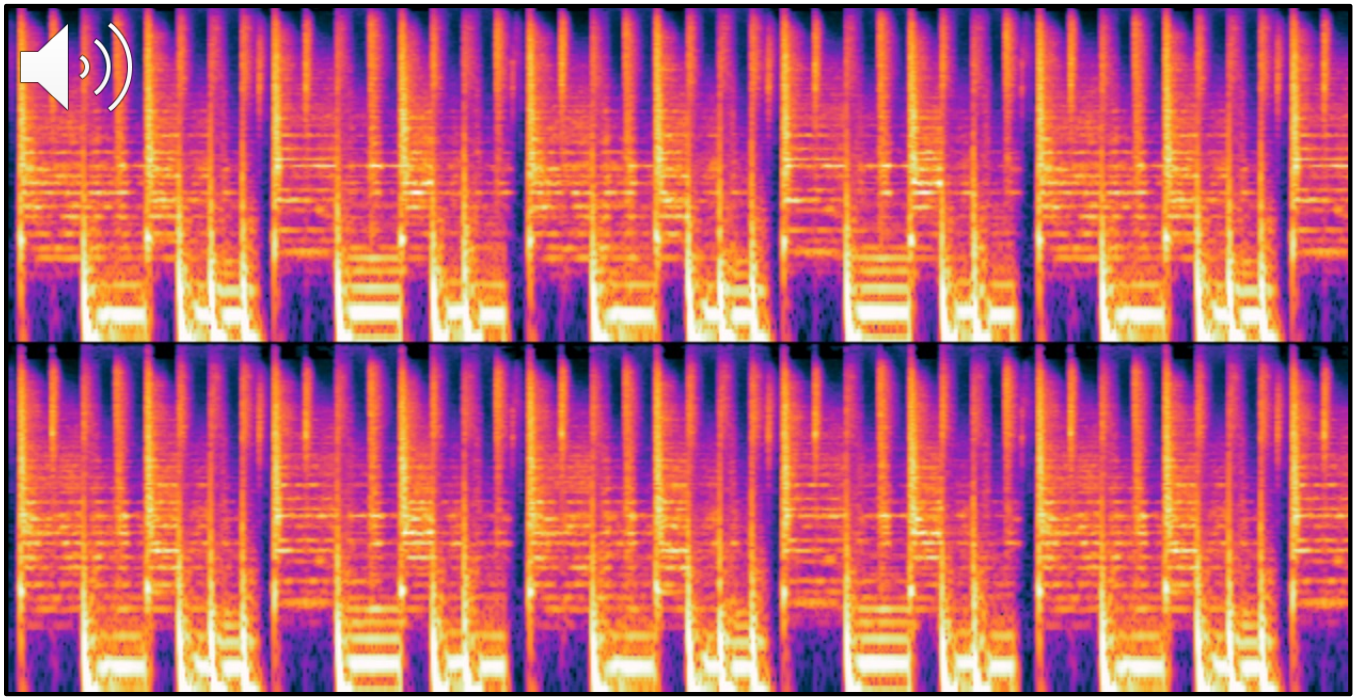


5. Quality



6. Workflows

Hip-hop with prominent kick drum and snappy snare

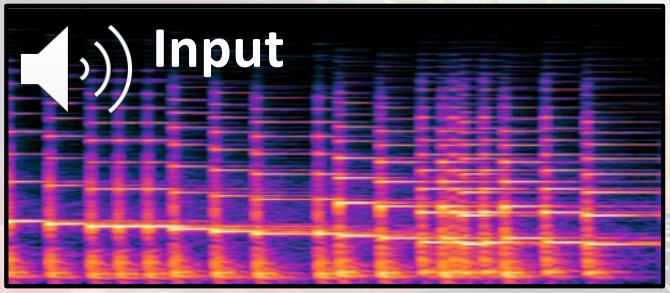


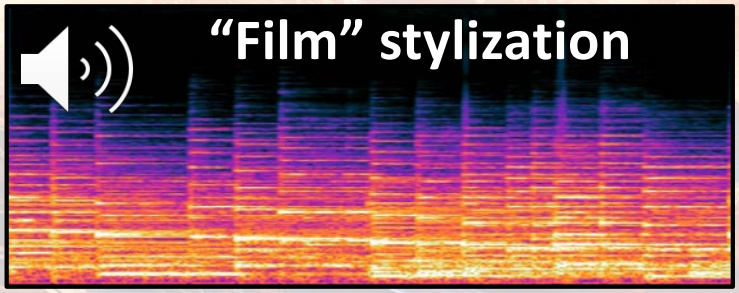
An Incomplete Way of Creating Music

“This song contains digital drums playing a simple groove along with two guitars. One strumming chords along with the snare the other one playing a melody on top. An e-bass is playing the footnote while a piano is playing a major and minor chord progression. A trumpet is playing a loud melody alongside the guitar. All the instruments sound flat and are being played by a keyboard. There are little bongo hits in the background panned to the left side of the speakers. Apart from the music you can hear eating sounds and a stomach rumbling. This song may be playing for an advertisement.” – Music Caps Dataset ”0xzrMun0Rs”

Beyond Text-to-Music

Melody Control

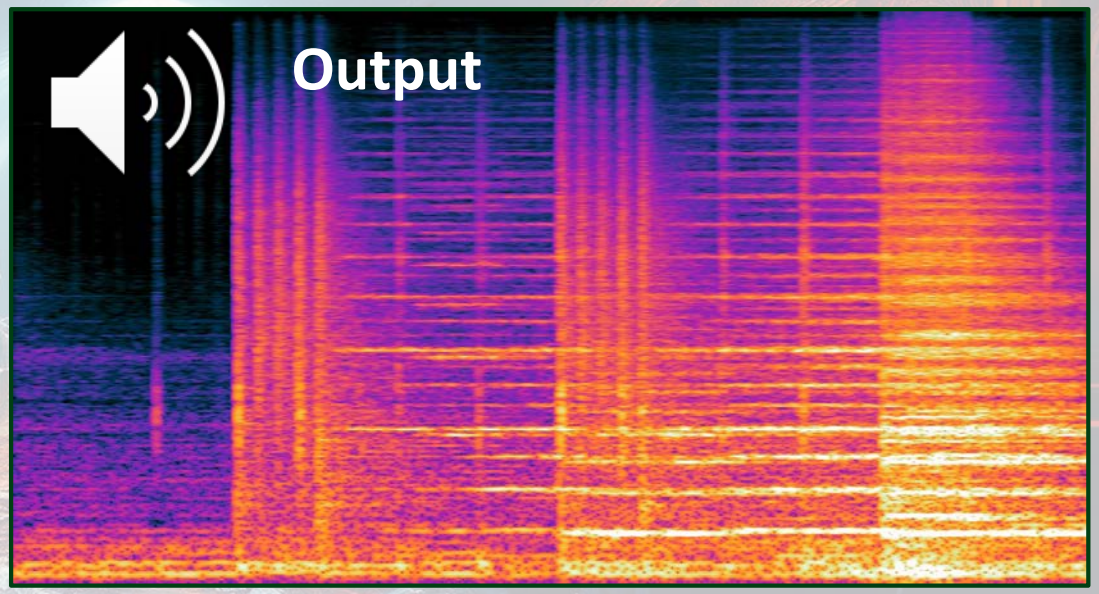




S.L. Wu, C. Donahue, S. Watanabe, N. J. Bryan “Music ControlNet: Multiple Time-varying Controls for Music Generation,” IEEE TASLP 2024.

Intensity Control

Input intensity

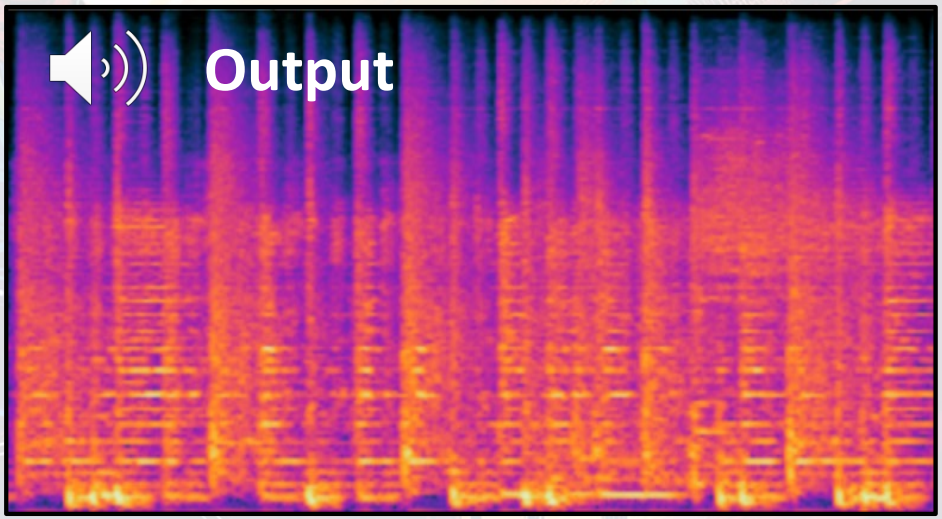


S.L. Wu, C. Donahue, S. Watanabe, N. J. Bryan "Music ControlNet: Multiple Time-varying Controls for Music Generation," IEEE TASLP 2024.

Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. "DITTO: Diffusion Inference-time T Optimization for Music Generation." ICML 2024.

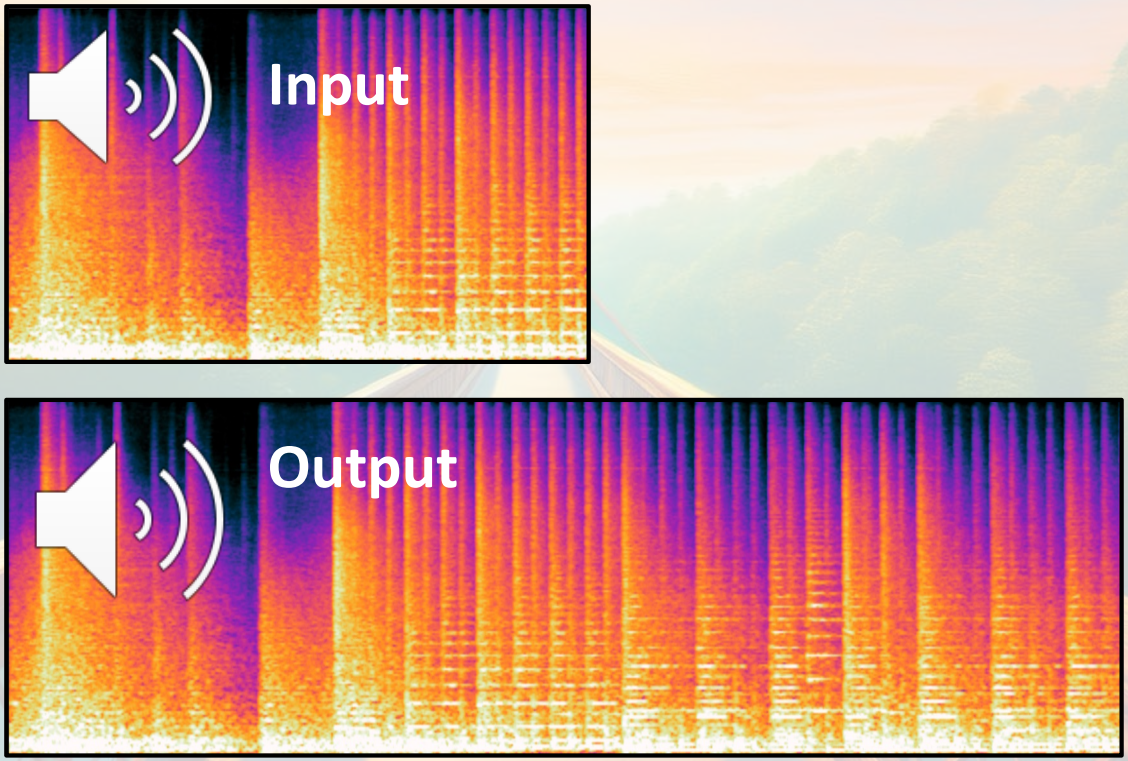
Rhythm Control

Input click



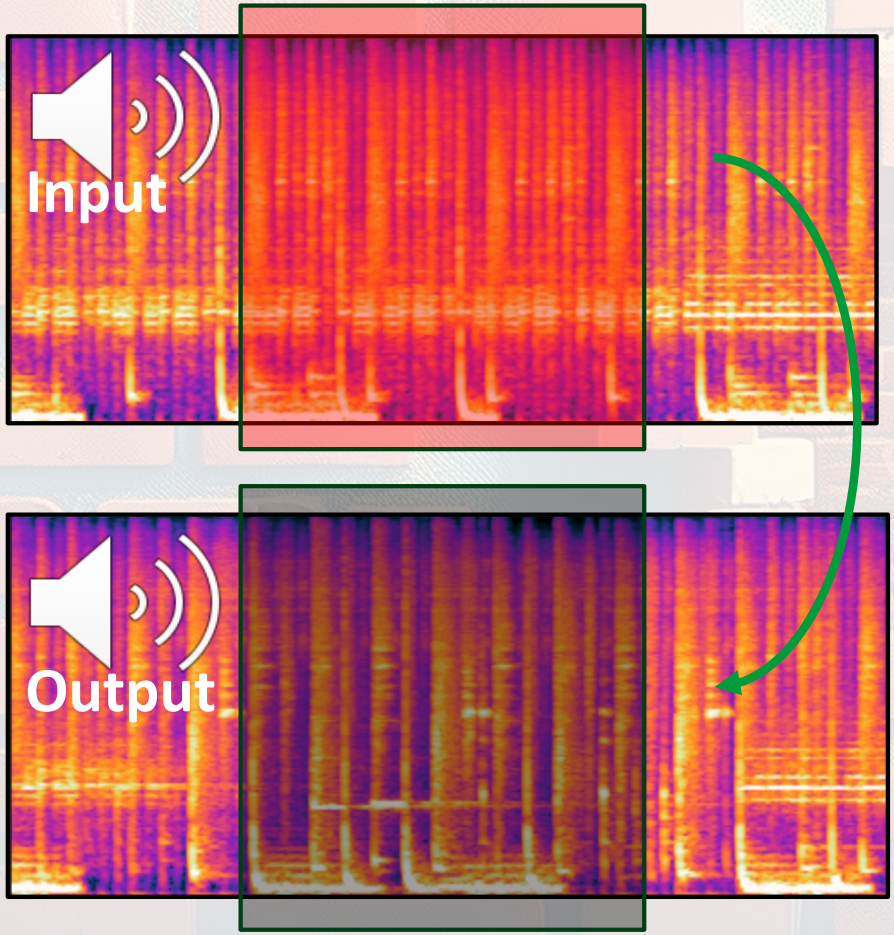
S.L. Wu, C. Donahue, S. Watanabe, N. J. Bryan "Music ControlNet: Multiple Time-varying Controls for Music Generation," IEEE TASLP 2024.

Length Extension



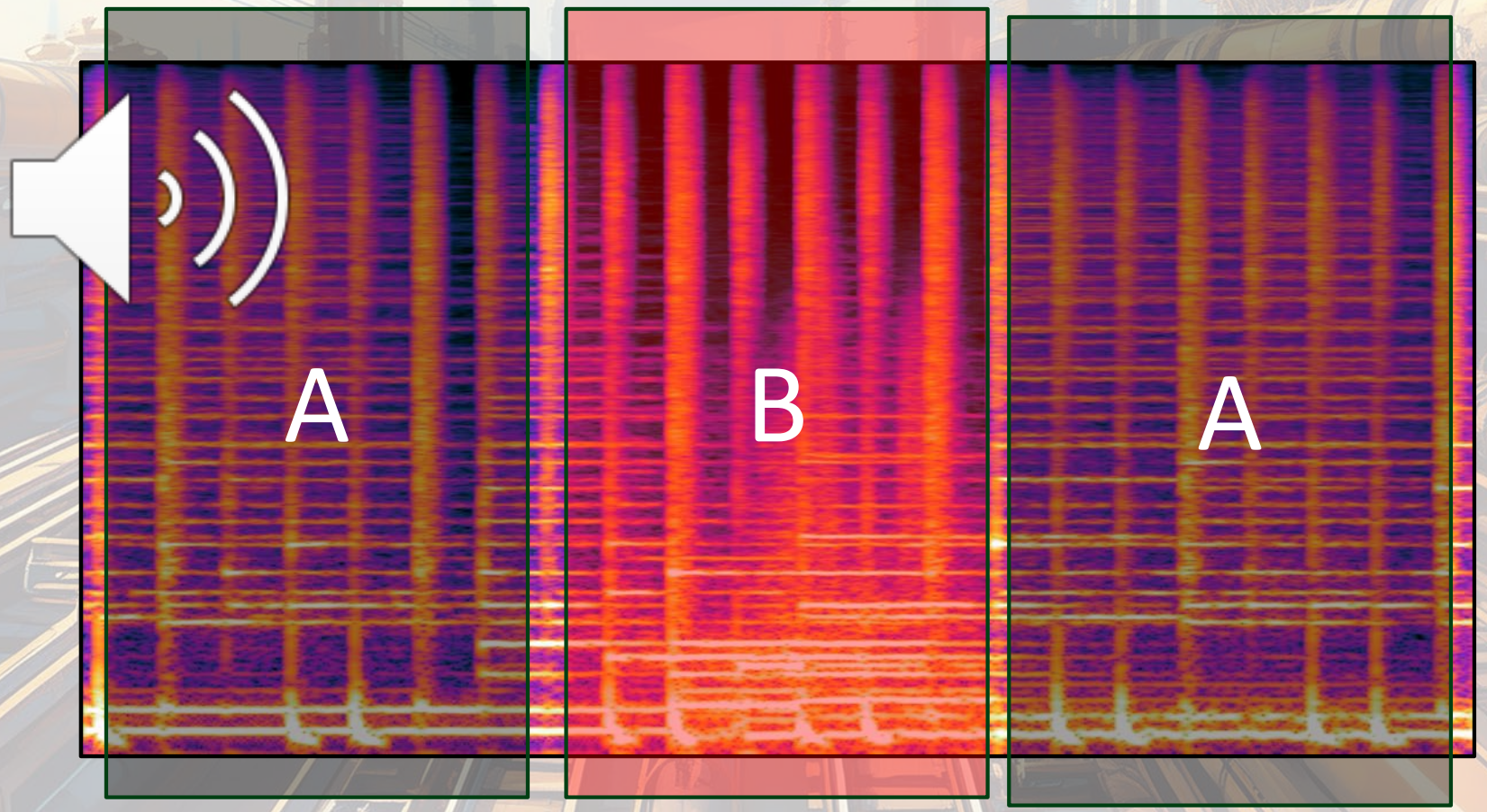
Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. "DITTO: Diffusion Inference-time T Optimization for Music Generation." ICML 2024.

Region Replacement



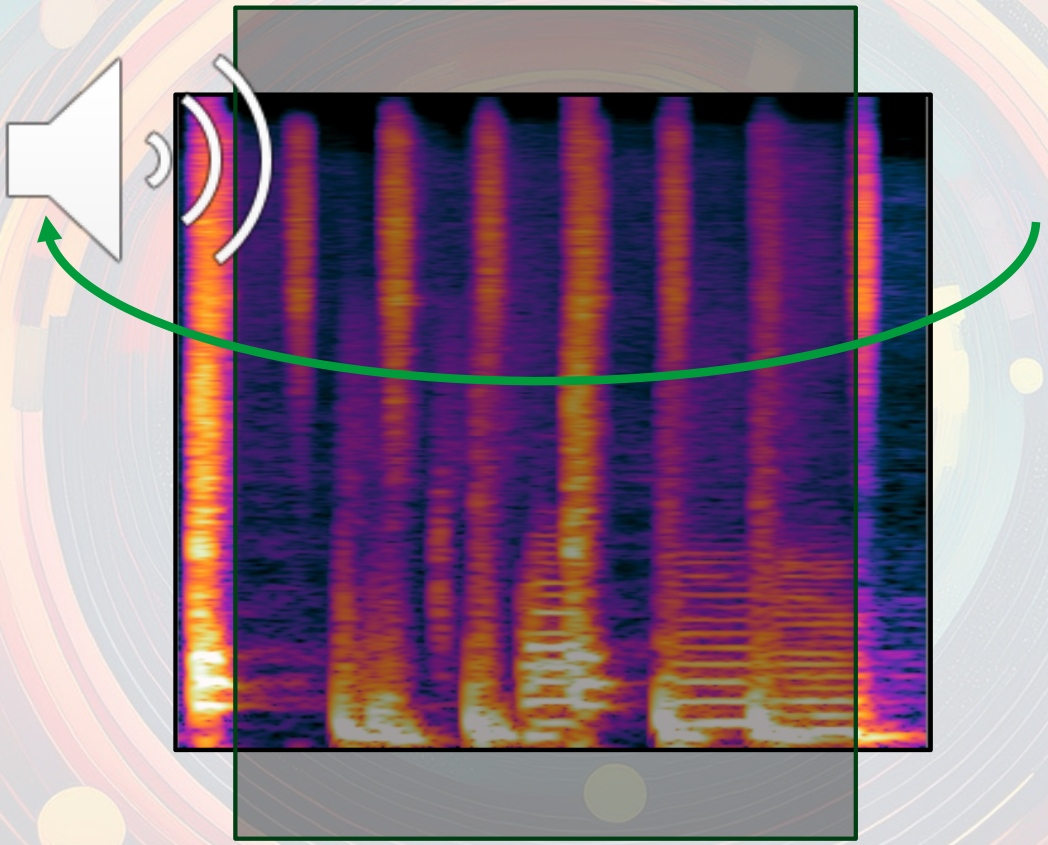
Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. "DITTO: Diffusion Inference-time T Optimization for Music Generation." ICML 2024.

Structure



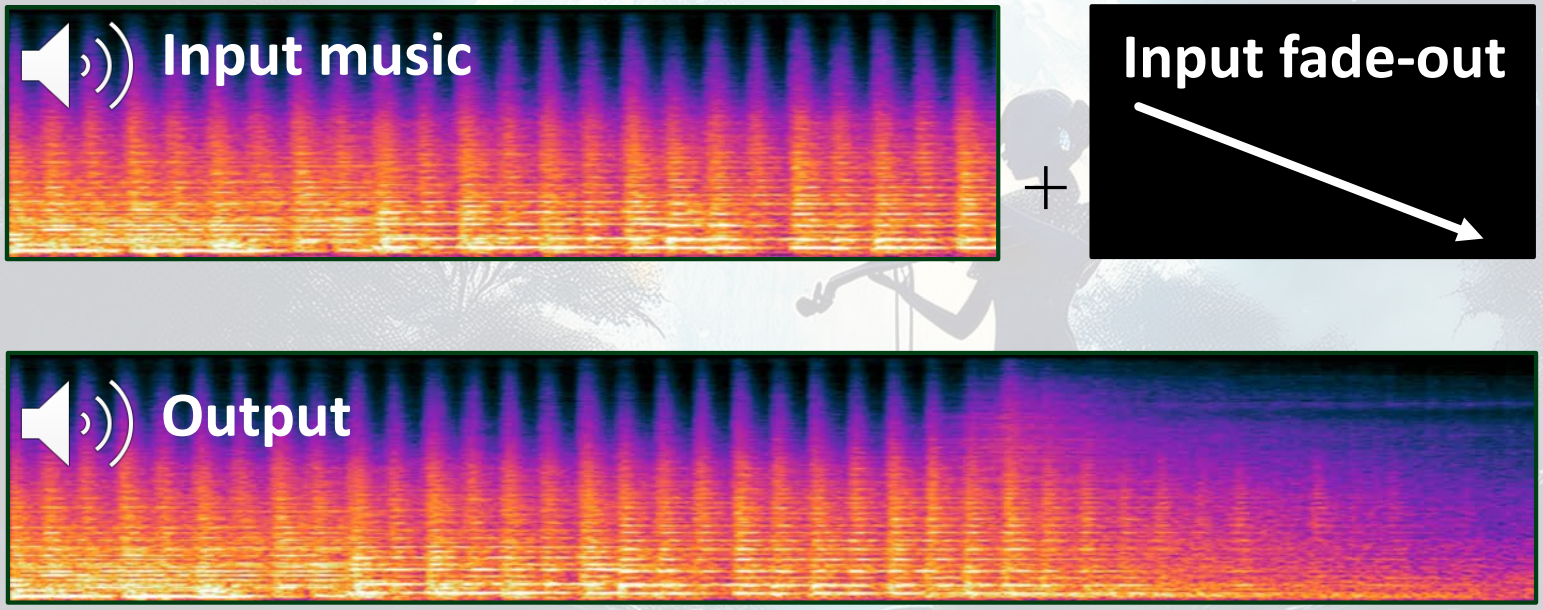
Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. “DITTO: Diffusion Inference-time T Optimization for Music Generation.” ICML 2024.

Looping



Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. “DITTO: Diffusion Inference-time T Optimization for Music Generation.” ICML 2024.

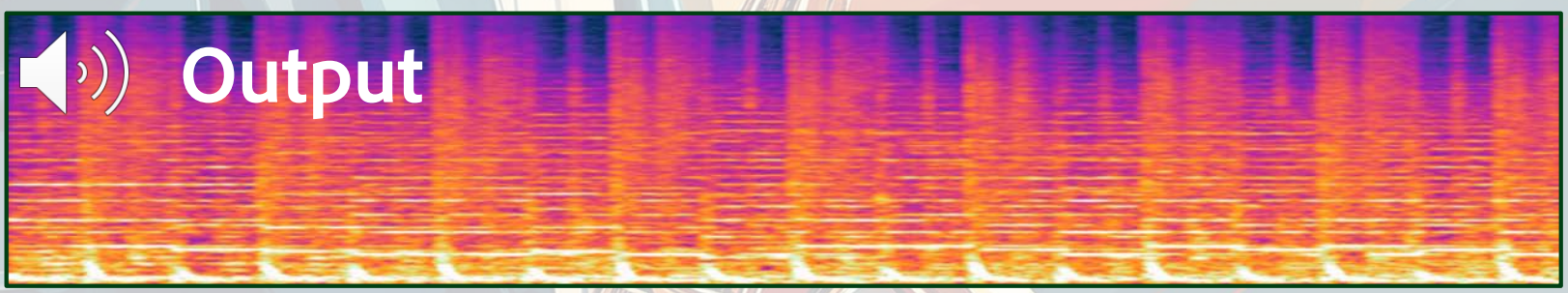
Composite Operators



Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. "DITTO: Diffusion Inference-time T Optimization for Music Generation." ICML 2024.

Partial or Improvisation Control

Improvised



S.L. Wu, C. Donahue, S. Watanabe, N. J. Bryan "Music ControlNet: Multiple Time-varying Controls for Music Generation," IEEE TASLP 2024.

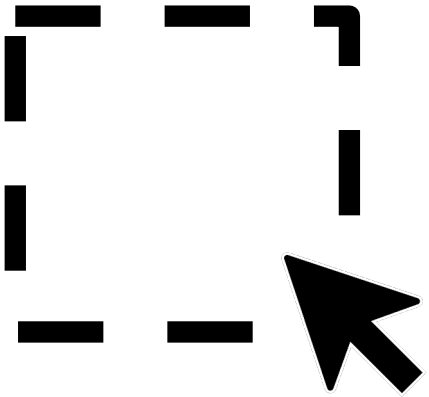
A New Design Language for Music Co-Creation

Control

- Text
- Genre
- Mood
- Instrument
- Descriptions
-

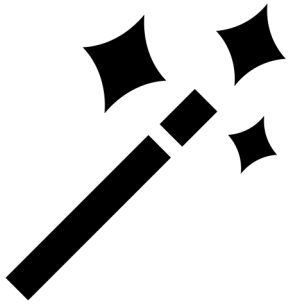
Musical Controls

- Melody
- Intensity
- Rhythm
- Structure
-



Editing

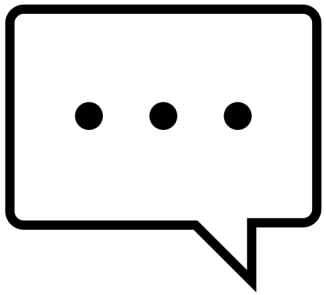
- Looping
- Length Extension
- Region Replacement
- …



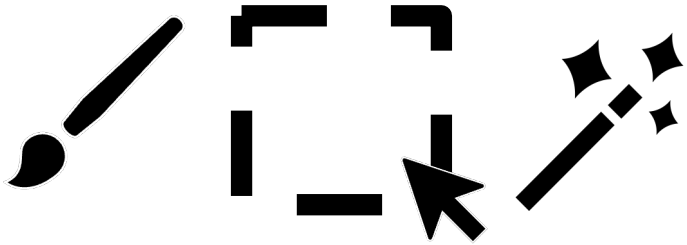
Special

- Composite operators
- Partial controls
- …

High-level to Low-Level Controls



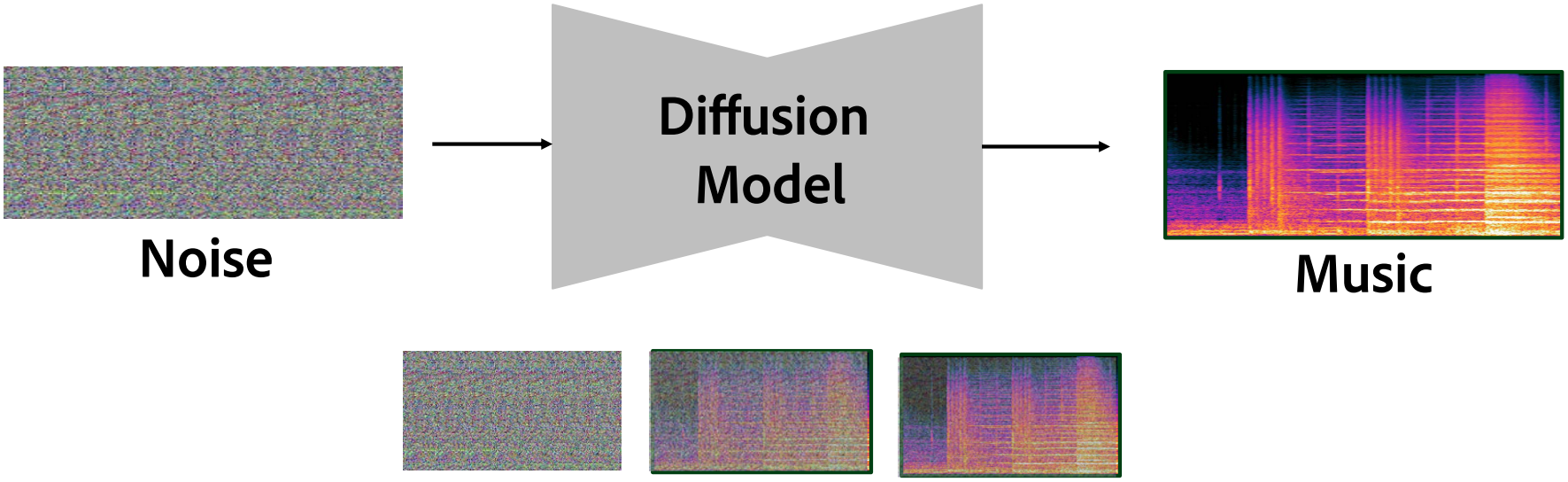
High-level



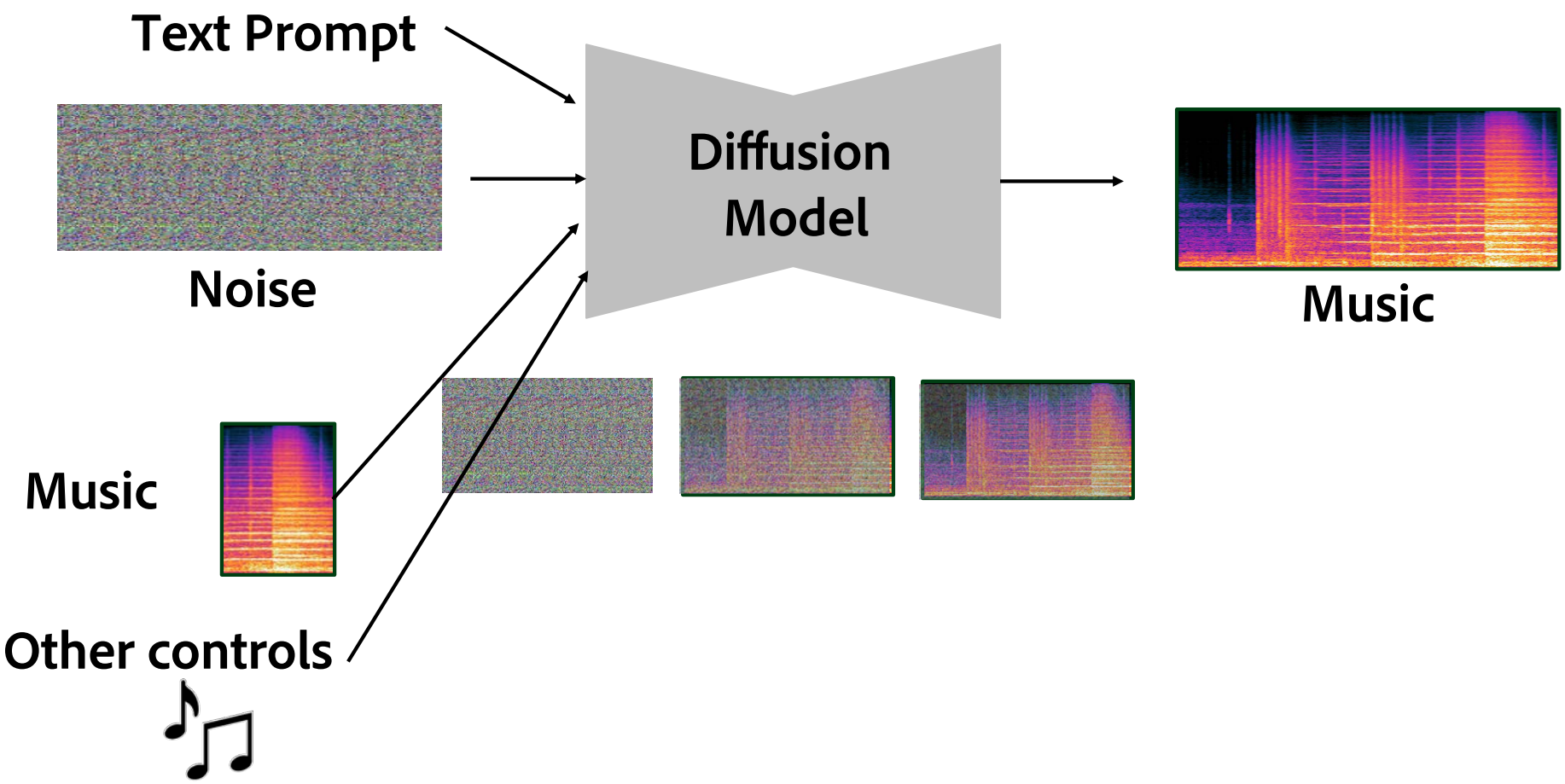
Mid-level

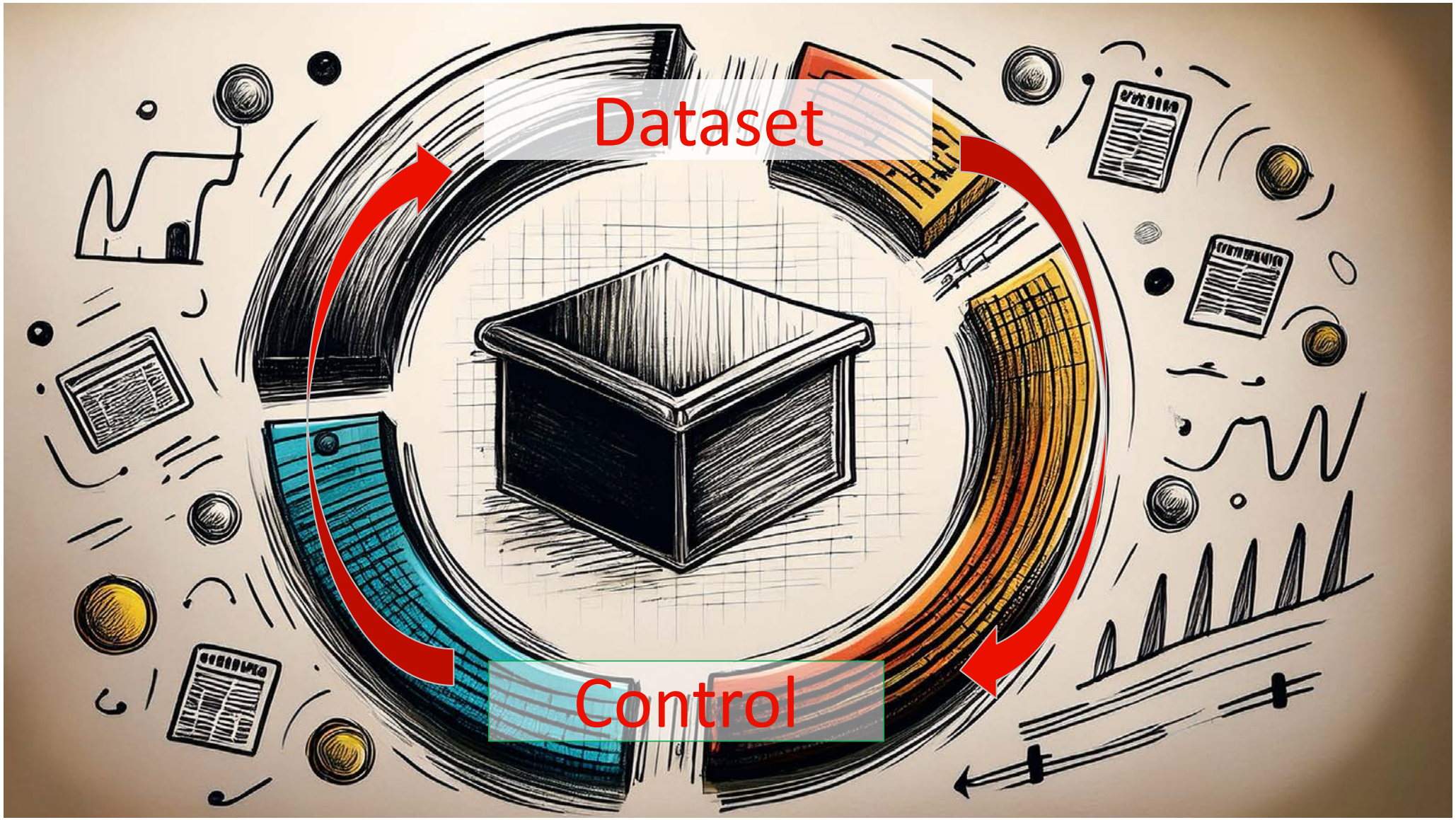
Low-level

Diffusion Model



Diffusion Model w/Controls

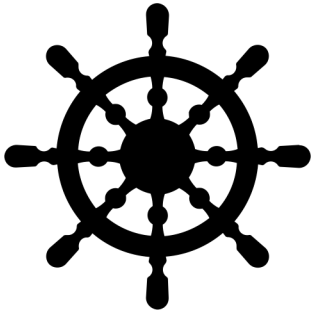



Dataset
Control

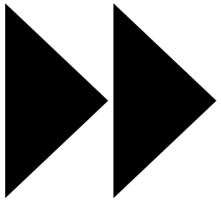

Learn from messy data
Synthesize w/structure

Generative Models for Downstream Tasks

- Emotion Recognition
- Instrument Classification
- Melody Transcription
- Beat Detection
- Musical Structure Segmentation
- …



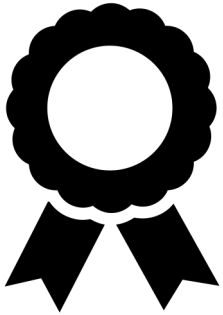
1. Control



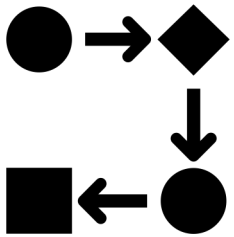
2. Speed

3. Efficiency

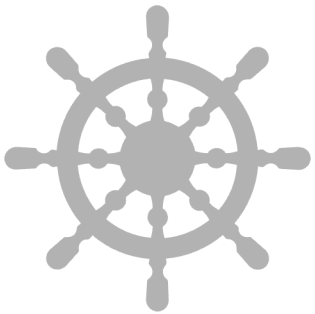
4. Data



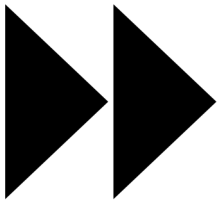
5. Quality



6. Workflows



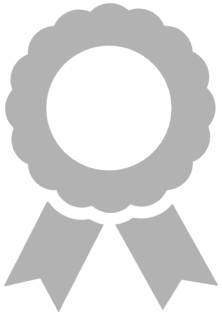
1. Control



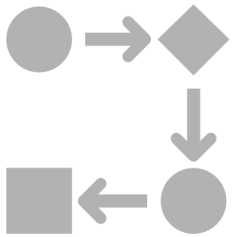
2. Speed

3. Efficiency

4. Data



5. Quality



6. Workflows

Presto! Distilling Steps and Layers for Accelerating Music Generation

Zachary Novack[♯♭*], Ge Zhu[♭], Jonah Casebeer[♭]

Julian McAuley[♯], Taylor Berg-Kirkpatrick[♯], Nicholas J. Bryan[♭]

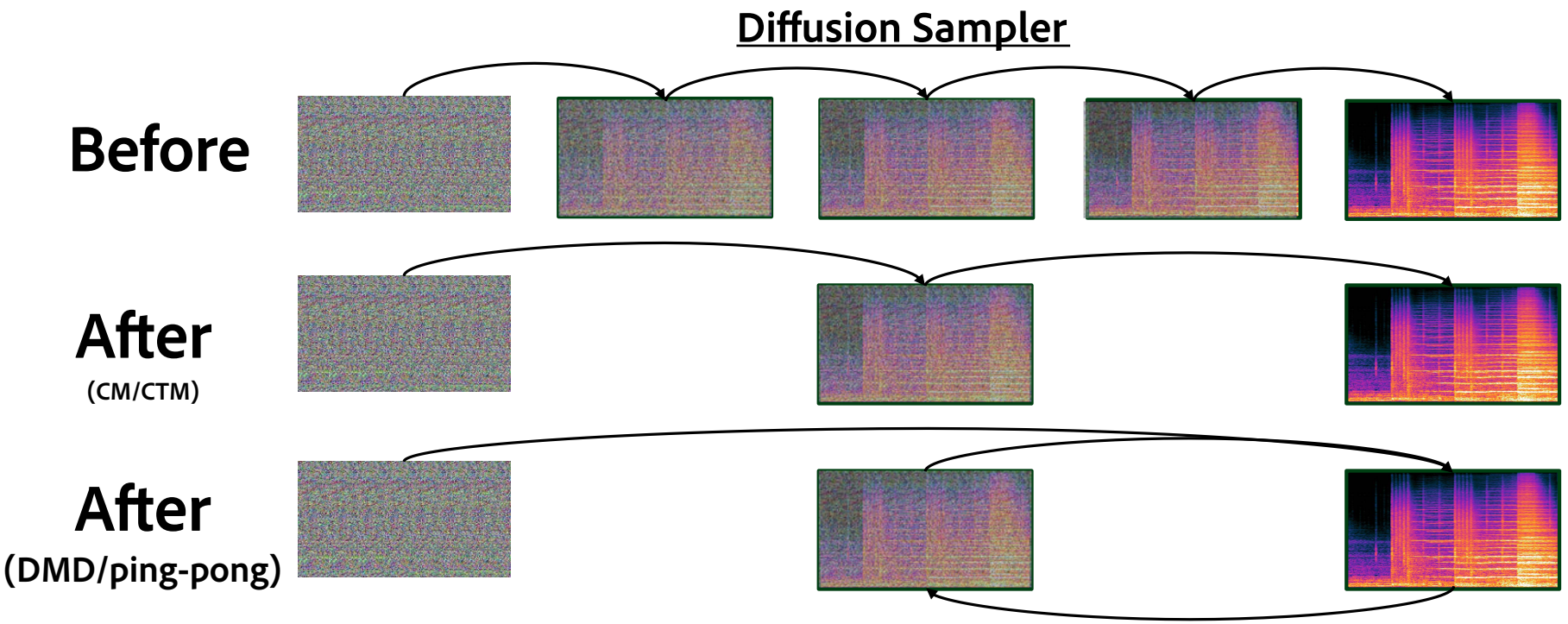
[♯]UCSD

[♭]Adobe Research

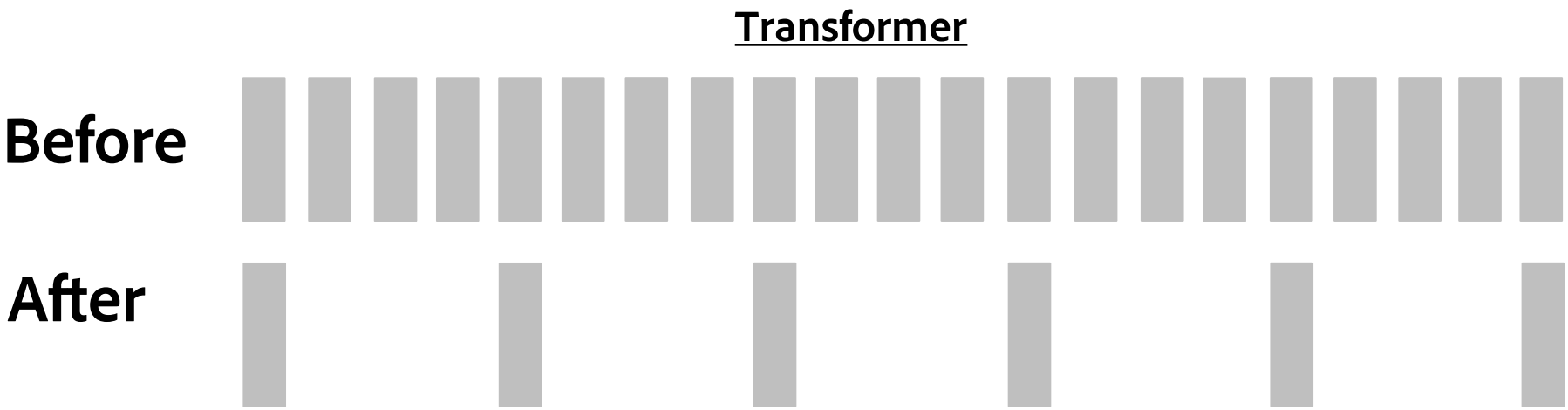
* Work done during an internship at Adobe Research.

(Listen with headphones)

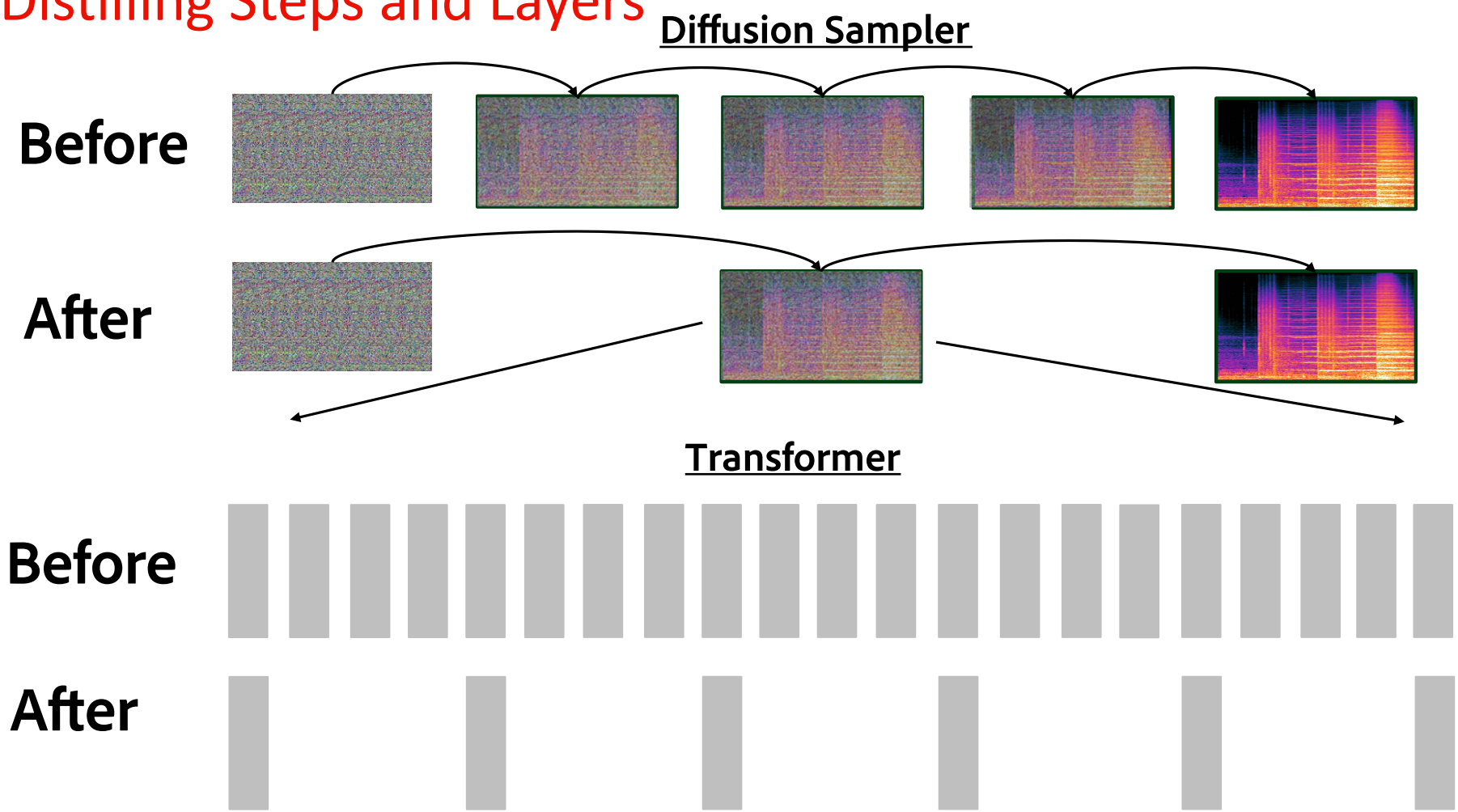
Distilling Steps

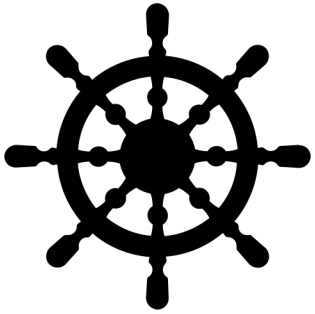


Distilling Layers (of a Transformer)

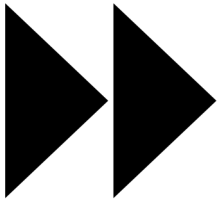


Distilling Steps and Layers





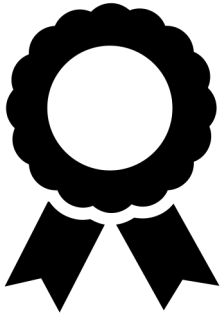
1. Control



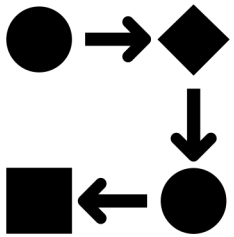
2. Speed

3. Efficiency

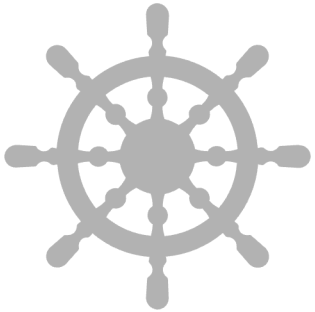
4. Data



5. Quality



6. Workflows

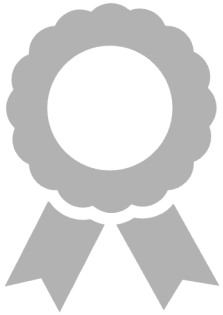


1. Control

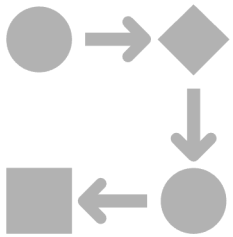
2. Speed

3. Efficiency

4. Data



5. Quality

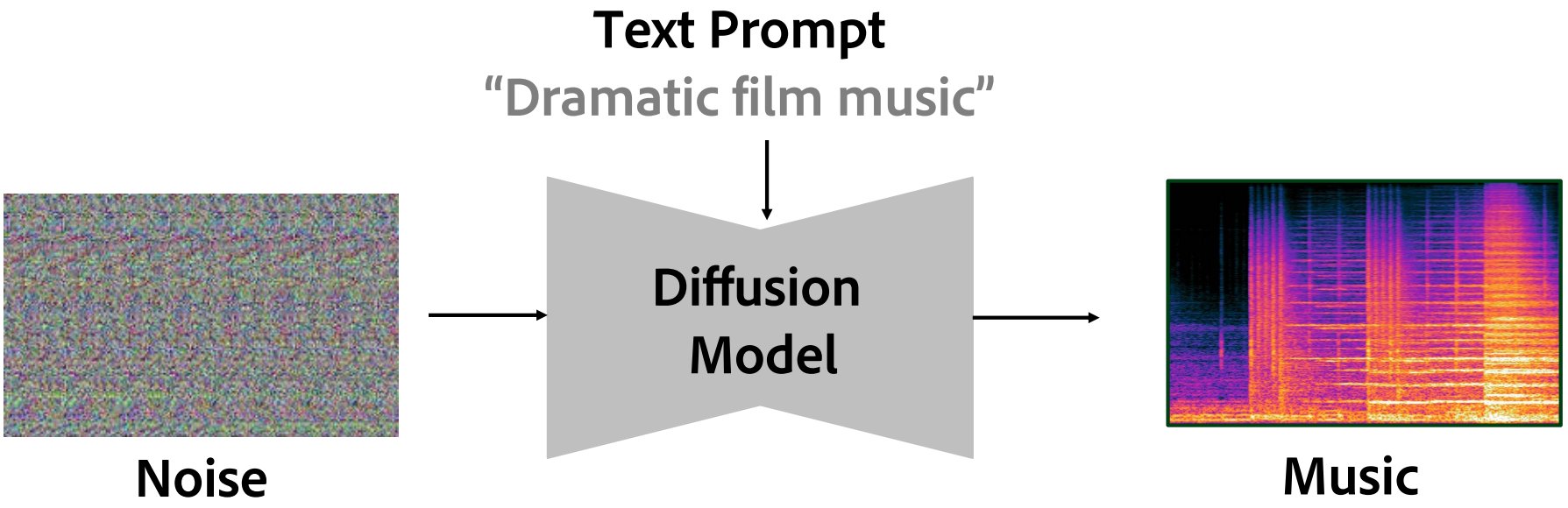


6. Workflows

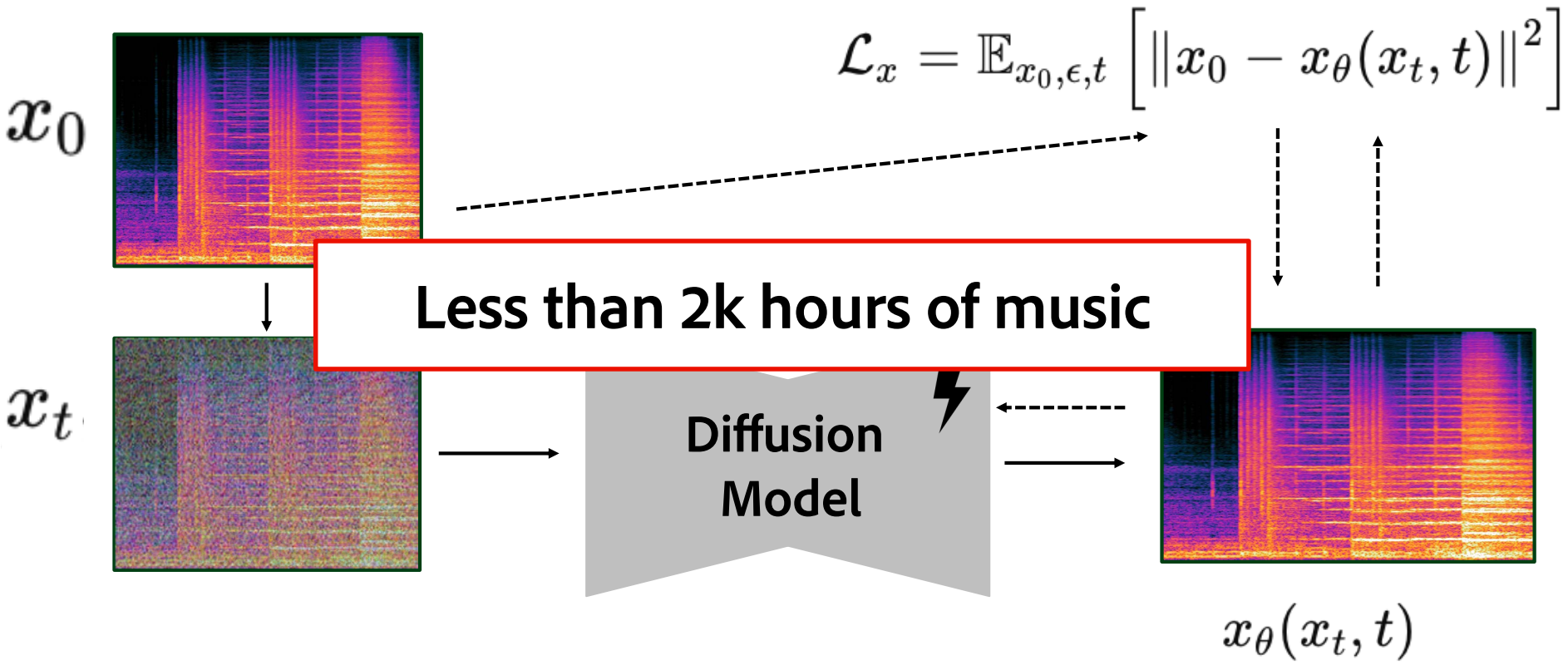
Training Methods for Efficiency

- Train a new model
- Fine-tine a model
- Inference-time control a pre-trained model
- Train a new model faster…

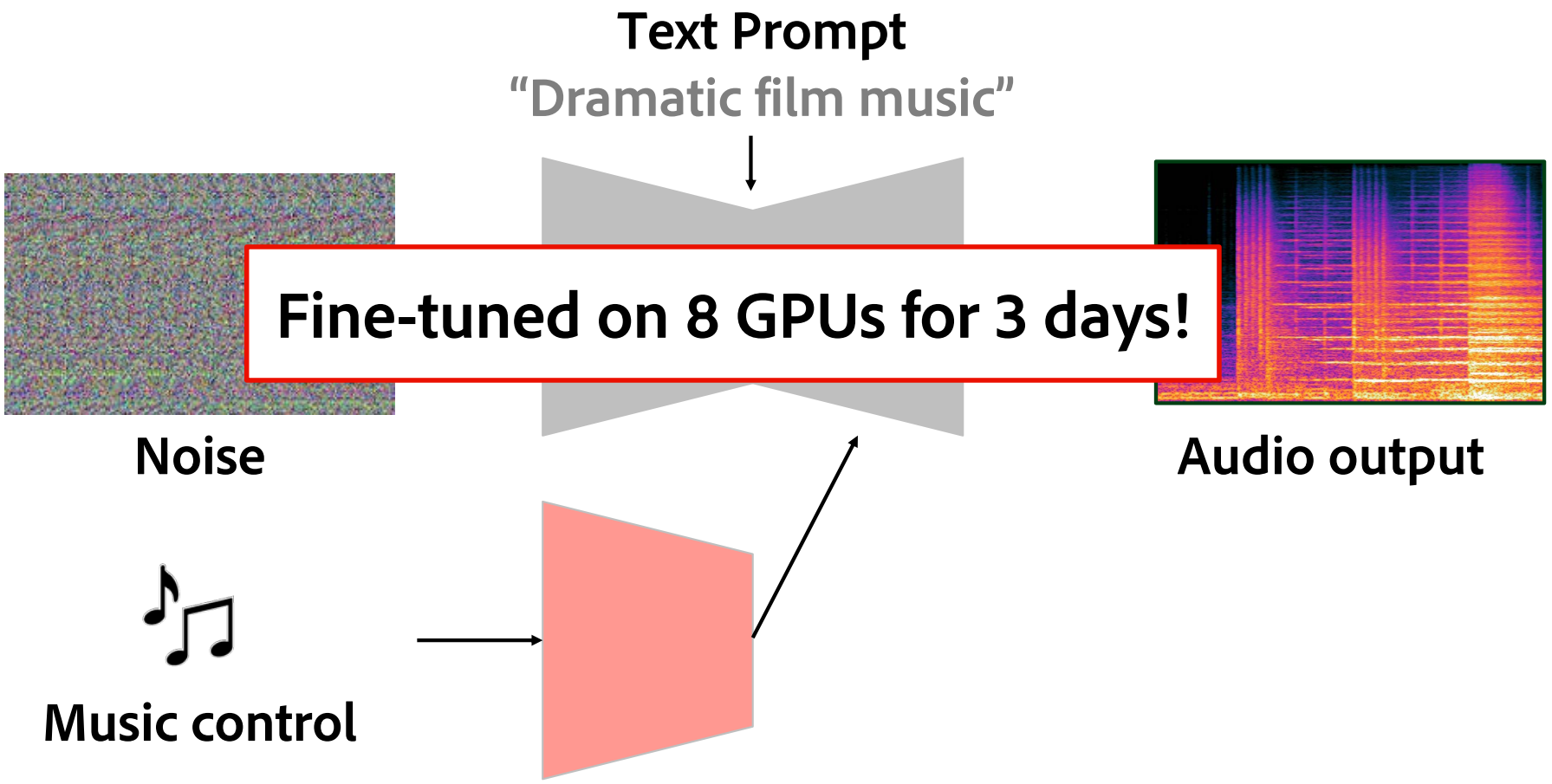
Train a New Model



Train a New Model

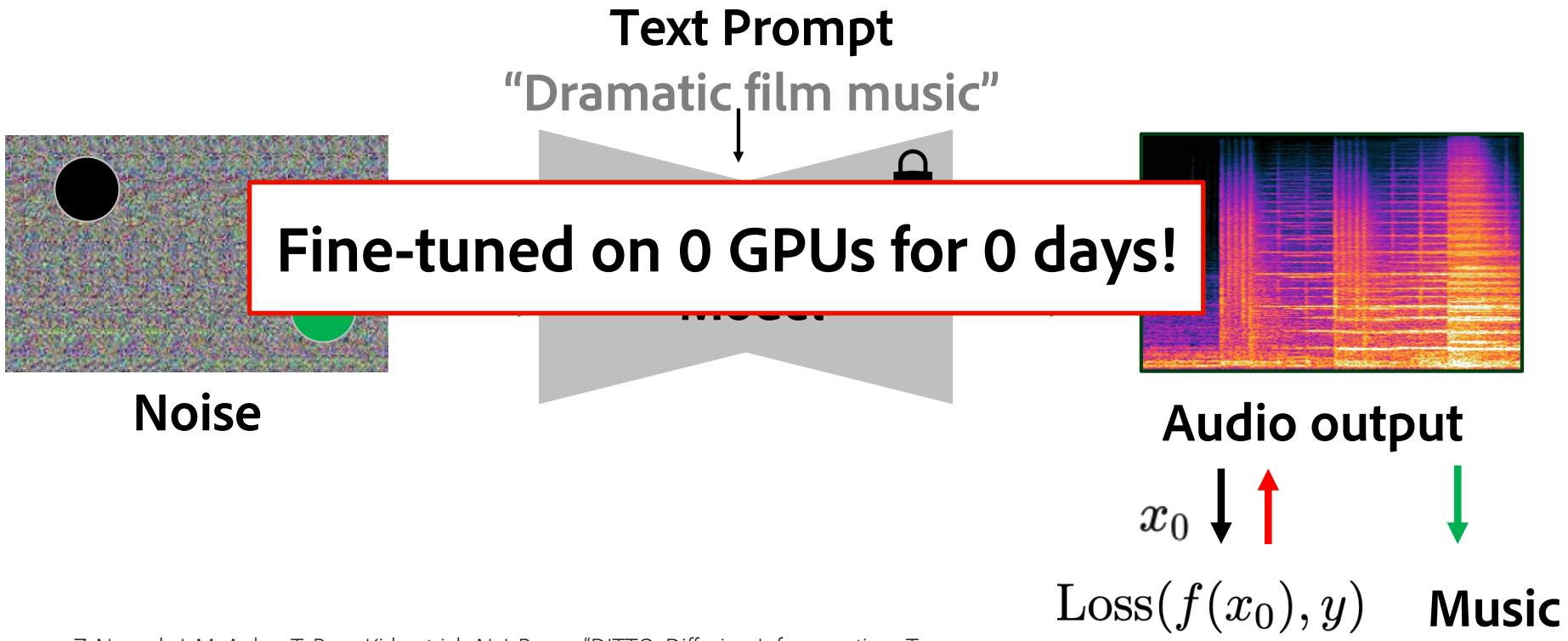


Fine-Tune a Pre-trained Model



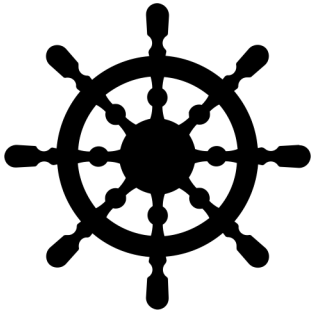
S.L. Wu, C. Donahue, S. Watanabe, N. J. Bryan “Music ControlNet: Multiple Time-varying Controls for Music Generation,” IEEE TASLP 2024.

Control a Pre-trained Model

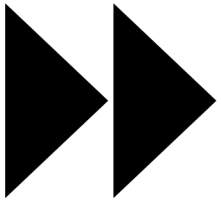


Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. "DITTO: Diffusion Inference-time T Optimization for Music Generation." ICML 2024.

Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. "DITTO-2: Distilled Diffusion Inference-Time T-Optimization for Music Generation" ISMIR 2024



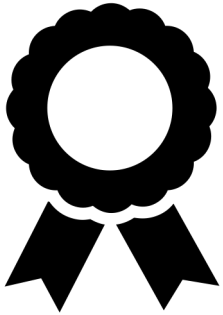
1. Control



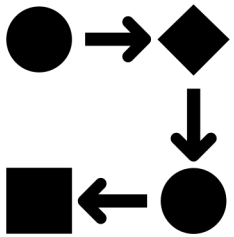
2. Speed

3. Efficiency

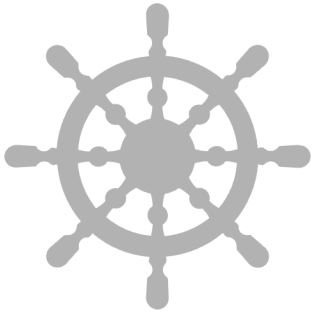
4. Data



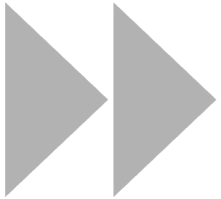
5. Quality



6. Workflows



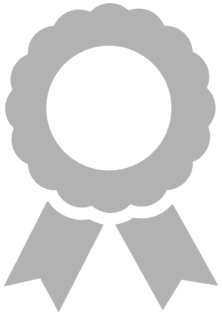
1. Control



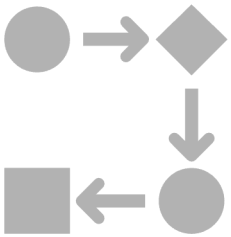
2. Speed

3. Efficiency

4. Data

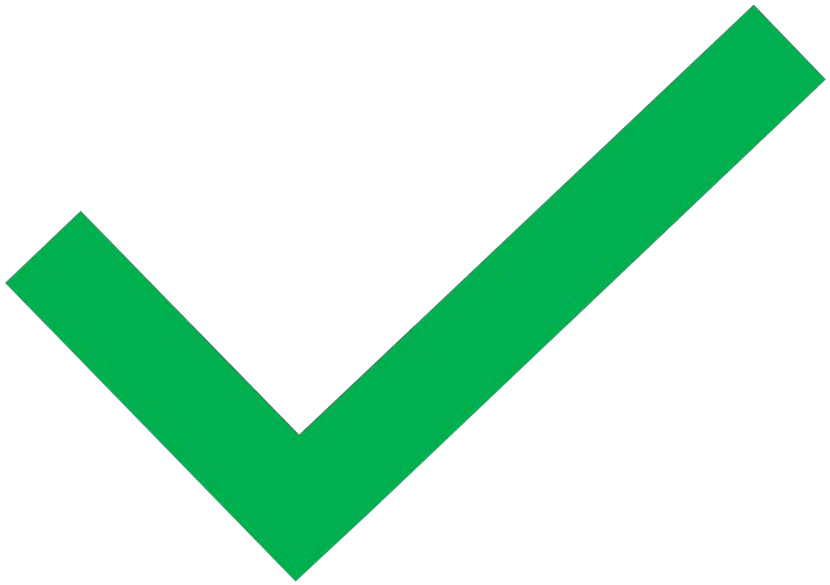


5. Quality



6. Workflows

UNLICENSED



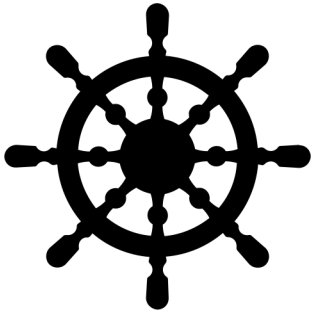
Grace Yee
10-14-2024

Reflecting on our five-year journey with our AI Ethics principles

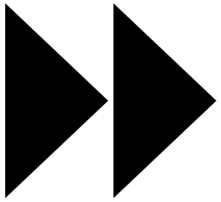
We do not and have never trained Adobe Firefly on customer content.

We only train Adobe Firefly on content where we have permission or rights to do so.

We compensate creators who contribute to Adobe Stock for use of their content in training Adobe Firefly.



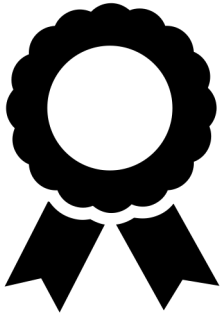
1. Control



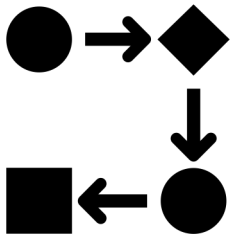
2. Speed

3. Efficiency

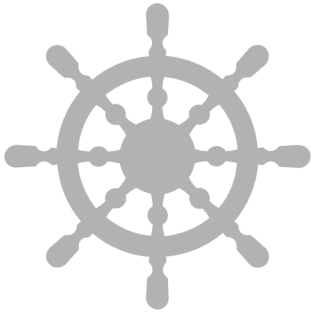
4. Data



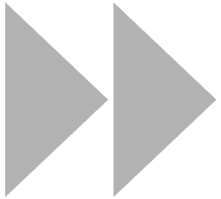
5. Quality



6. Workflows



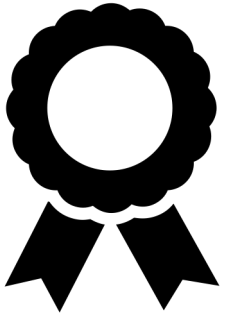
1. Control



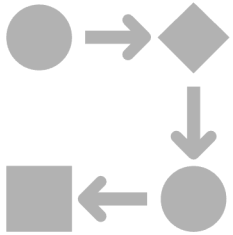
2. Speed

3. Efficiency

4. Data

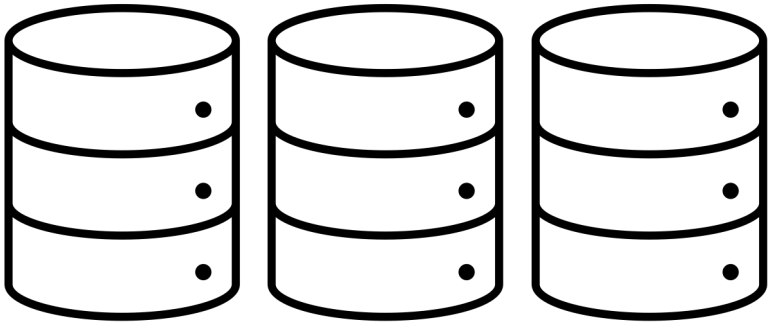


5. Quality

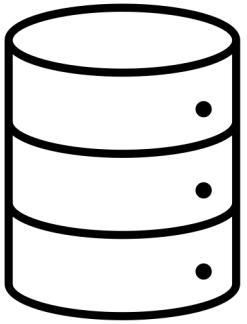


6. Workflows

Big Data vs. Small Data



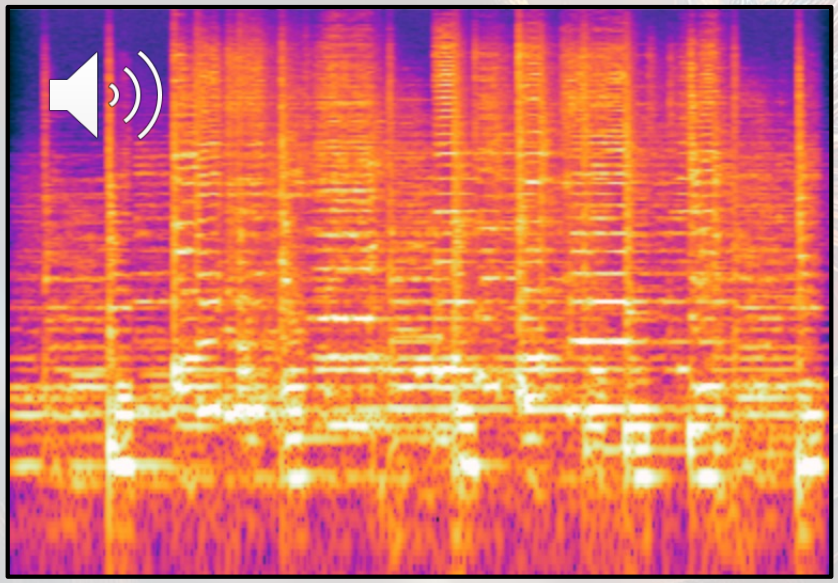
Big Data



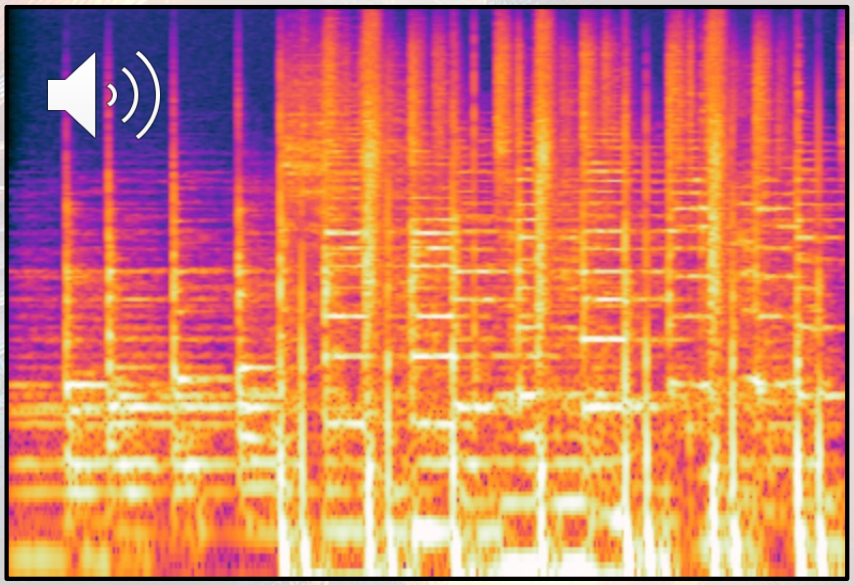
Small Data


Aesthetics

DRAGON: Distributional Quality Metrics To Enhance Diffusion Models

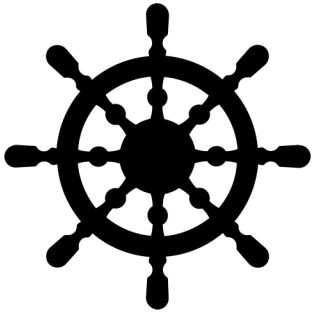


Before

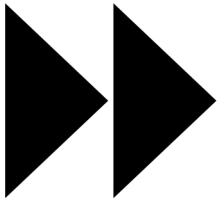


After

[Coming Soon!] Y. Bai, J. Casebeer, S. Sojoudi, N. J. Bryan, "DRAGON: Optimizing Distributional Quality Metrics To Enhance Diffusion Models"



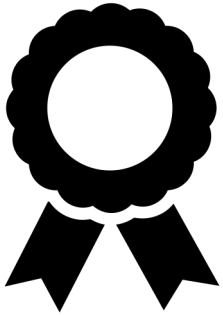
1. Control



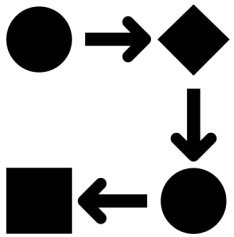
2. Speed

3. Efficiency

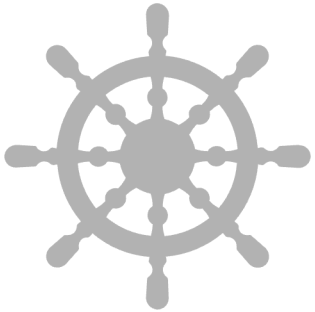
4. Data



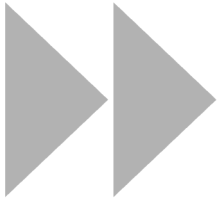
5. Quality



6. Workflows



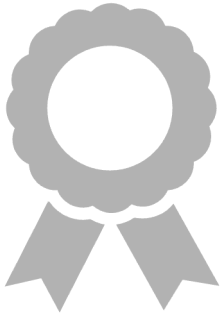
1. Control



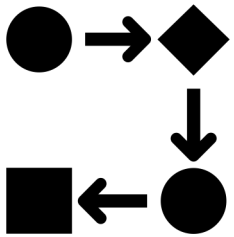
2. Speed

3. Efficiency

4. Data



5. Quality



6. Workflows

Conclusions

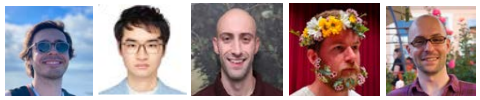
References & Collaborators

S.L. Wu, C. Donahue, S. Watanabe, N. J. Bryan “Music ControlNet: Multiple Time-varying Controls for Music Generation,” IEEE TASLP 2024.

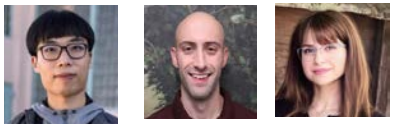
Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. “DITTO: Diffusion Inference-time T Optimization for Music Generation.” ICML 2024.

Z. Novack, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. “DITTO-2: Distilled Diffusion Inference-Time T-Optimization for Music Generation” ISMIR 2024

G. Zhu, J. P. Caceres, Z. Duan, N. J. Bryan, “Music HiFi: Fast High-Fidelity Stereo Vocoding”, IEEE SPL 2024.



Z. Novack, G Zhu, J. Casebeer, J. McAuley, T. Berg-Kirkpatrick, N. J. Bryan. “Presto! Distilling Steps and Layers for Accelerating Music Generation” ICLR 2024.



[Soon!] Y. Bai, J. Casebeer, S. Sojoudi, N. J. Bryan, “DRAGON: Optimizing Distributional Quality Metrics To Enhance Diffusion Models”

Generative AI Music: Beyond Text-to-Music

Nick Bryan
Head of Music AI
Adobe Research

Web: https://njb.github.io
Email: njb@ieee.org

- Let's go beyond text-to-music!
- Control, speed, efficiency, data, quality, and workflows
- A new design language for music co-creation
- Speed amplifies the power of control
- Different ways to train models efficiently, but needs more work
- Data and quality of outputs are linked
- Real iterative workflows will be an exciting next step