

Generative AI for Music: Challenges & Opportunities

Hao-Wen (Herman) Dong

Department of Performing Arts Technology
School of Music, Theatre & Dance
University of Michigan
hermandong.com

July 8, 2026

Music & Technology Co-evolves



Hildegard Dodel, Public domain, via Wikimedia Commons.
Taken at Hamamatsu Museum of Musical Instruments, August 2019.
yan, [CC BY-SA 4.0](#), via Wikimedia Commons.

Music & AI

(Source: Yamaha)



(Source: Sankei Shimbun)



(Source: Robot Gizmos)



(Source: NBC DFW)

yamaha.com/en/news_release/2018/18013101/
sankei.com/article/20240113-CQCOSQHJWFIYPJJKZDCITRTRVI/
roboticgizmos.com/shimon-musical-robot-deep-learning/
nbcdfw.com/entertainment/the-scene/how-verdigris-ensemble-is-using-ai-to-create-a-new-concert-experience/3366031/

Generative AI for Content Creation



Universitaetsmedizin, [CC BY-SA 4.0](#), via Wikimedia Commons
[uploadvr.com/iron-man-vr-breaks-free-from-cords-load-screens-on-quest-2/](#)
[descript.com/blog/article/what-is-the-best-audio-interface-for-recording-a-podcast](#)
[denverpost.com/2019/08/02/colorado-symphony-movie-scores-harry-potter-star-wars/](#)
[dailybruin.com/2023/08/04/theater-review-the-musical-les-misrables-offers-stellar-displays-and-impassioned-vocals](#)

Art challenges Technology




Music

**Augmenting Human Creativity
with AI**


AI

Technology inspires the Art

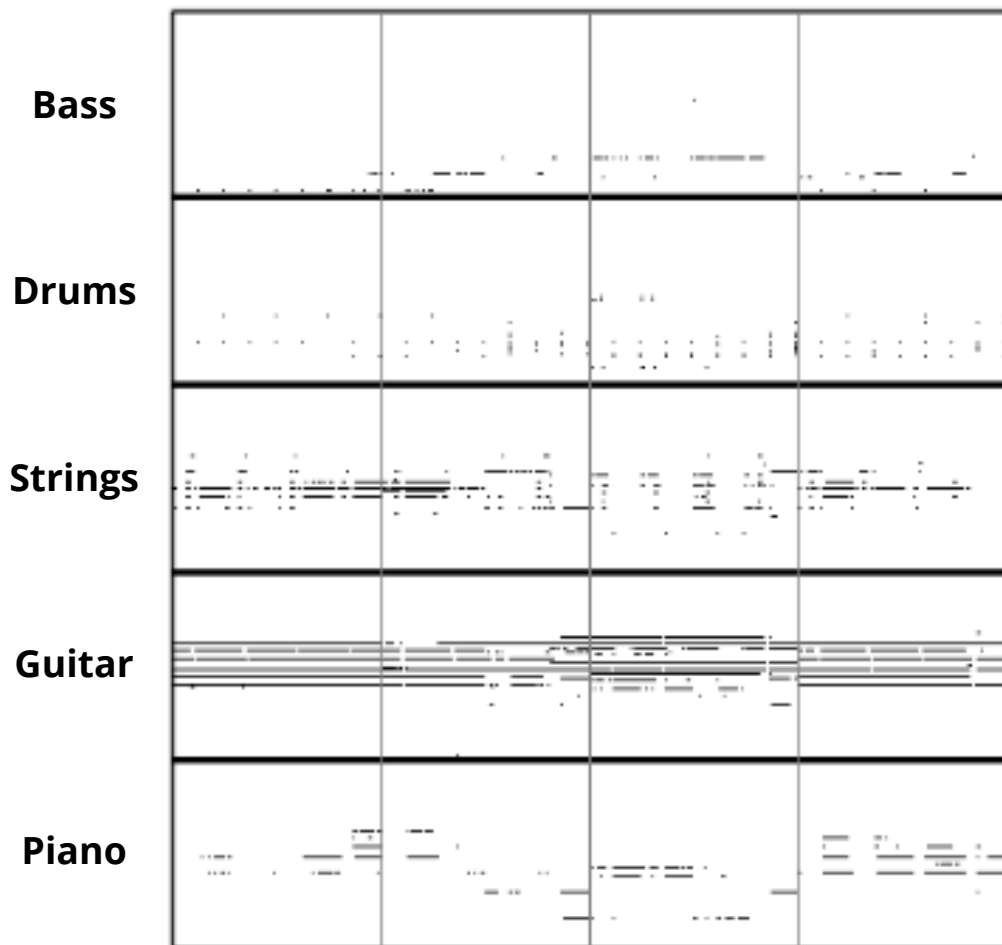


| My Research on AI for Music

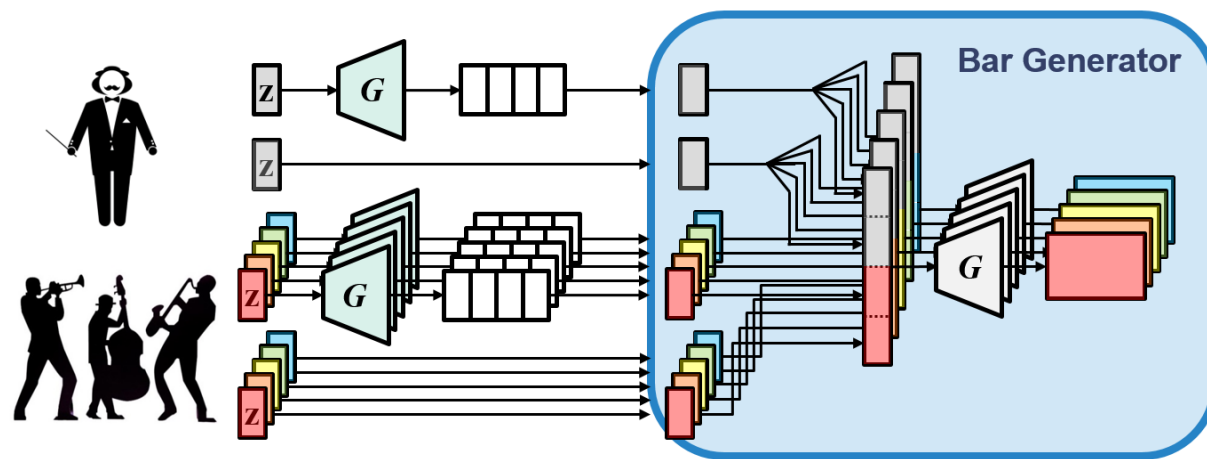
- **Multitrack music generation** (ISMIR 2024 , ICASSP 2023, ISMIR 2020 , ISMIR 2018, AAAI 2018, ISMIR LBD 2017)
- **Expressive violin performance synthesis** (ICASSP 2025, ICASSP 2022)
- **Text-to-music generation** (ISMIR 2025, AIMG 2024, ISMIR LBD 2024)
- **Video-to-music generation** (ISMIR 2025, ISMIR LBD 2025)
- **Music arrangement** (ISMIR 2021, JNMR 2020)
- **Music LLM** (ICASSP 2025, NLP4MusA 2024)
- **Choral music separation** (ISMIR LBD 2025, ISMIR 2022)
- **Optical music recognition** (ISMIR 2021)

Generating Multi-instrument Music using GANs (AAAI 2018)

Multitrack Piano Roll



MuseGAN Generator



Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang, "MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment," AAAI, 2018.

MuseGAN Featured in AWS DeepComposer (2020)

AWS DeepComposer > Models > Train a model

Train a model

Generative algorithm [Info](#)

Choose a generative algorithm to train a model

☒ **MuseGAN**
GAN algorithm often used for complex music structures

☐ **U-Net**
U-Net is the distinguishing

512

Input noise

Volume

32-key, 2-octave keyboard

The screenshot shows the AWS DeepComposer 'Train a model' interface. The 'Generative algorithm' section has 'MuseGAN' selected, with a note that it is a GAN algorithm often used for complex music structures. The 'U-Net' option is also visible. The background features a large 32-key, 2-octave keyboard and a video of a man playing the keyboard.

amazon.com/dp/B07YGZ4V5B/

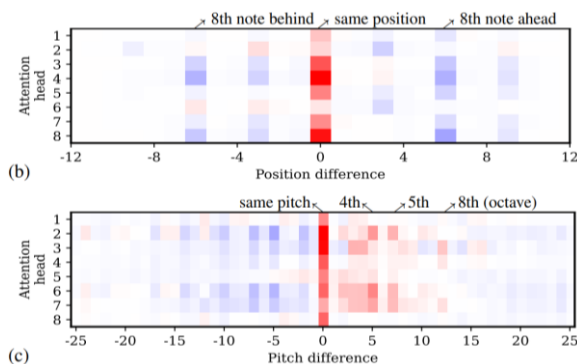
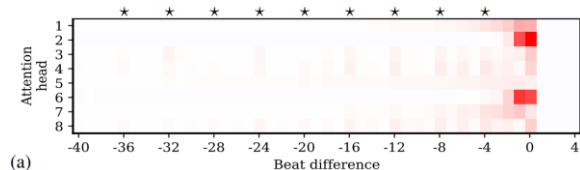
Julien Simon, "AWS DeepComposer – Now Generally Available With New Features," AWS News Blog, April 2, 2020.

Generating Multitrack Music with Transformers (ICASSP 2023)

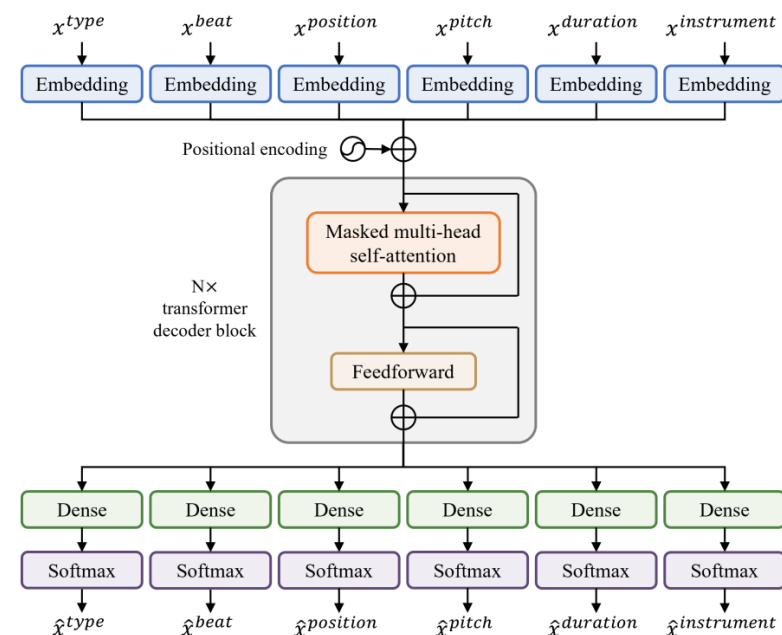
Multitrack Music Representation

(0, 0, 0, 0, 0, 0)	Start of song
(1, 0, 0, 0, 0, 15)	Instrument: accordion
(1, 0, 0, 0, 0, 36)	Instrument: trombone
(1, 0, 0, 0, 0, 39)	Instrument: brasses
(2, 0, 0, 0, 0, 0)	Start of notes
(3, 1, 1, 41, 15, 36)	Note: beat=1, position=1, pitch=E2, duration=48, instrument=trombone
(3, 1, 1, 65, 4, 39)	Note: beat=1, position=1, pitch=E4, duration=12, instrument=brasses
(3, 1, 1, 65, 17, 15)	Note: beat=1, position=1, pitch=E4, duration=72, instrument=accordion
(3, 1, 1, 68, 4, 39)	Note: beat=1, position=1, pitch=G4, duration=12, instrument=brasses
(3, 1, 1, 68, 17, 15)	Note: beat=1, position=1, pitch=G4, duration=72, instrument=accordion
(3, 1, 1, 73, 17, 15)	Note: beat=1, position=1, pitch=C5, duration=72, instrument=accordion
(3, 1, 13, 68, 4, 39)	Note: beat=1, position=13, pitch=G4, duration=12, instrument=brasses
(3, 1, 13, 73, 4, 39)	Note: beat=1, position=13, pitch=C5, duration=12, instrument=brasses
(3, 2, 1, 73, 12, 39)	Note: beat=2, position=1, pitch=C5, duration=36, instrument=brasses
(3, 2, 1, 77, 12, 39)	Note: beat=2, position=1, pitch=E5, duration=36, instrument=brasses
...	...
(4, 0, 0, 0, 0, 0)	End of song

Musical Self-attention



Multitrack Music Transformer



UC San Diego

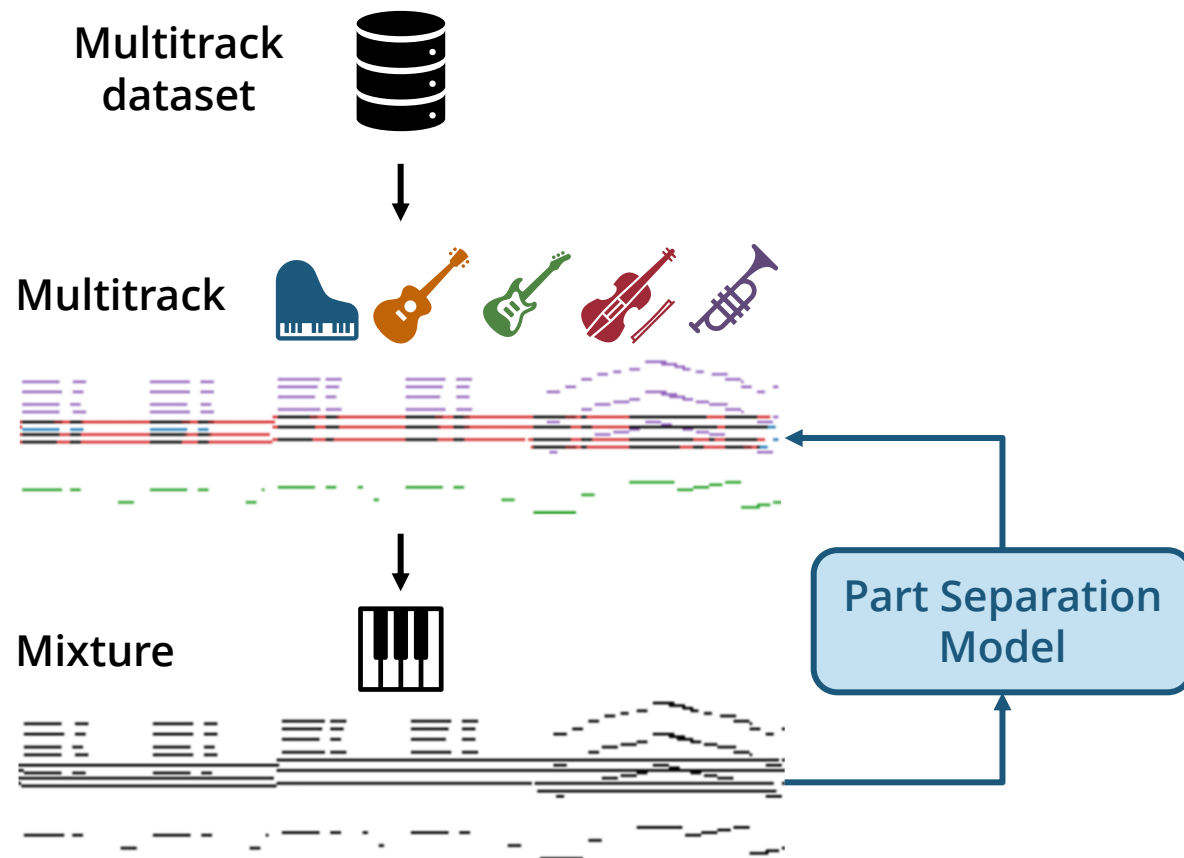
Automatic Instrumentation (ISMIR 2021)



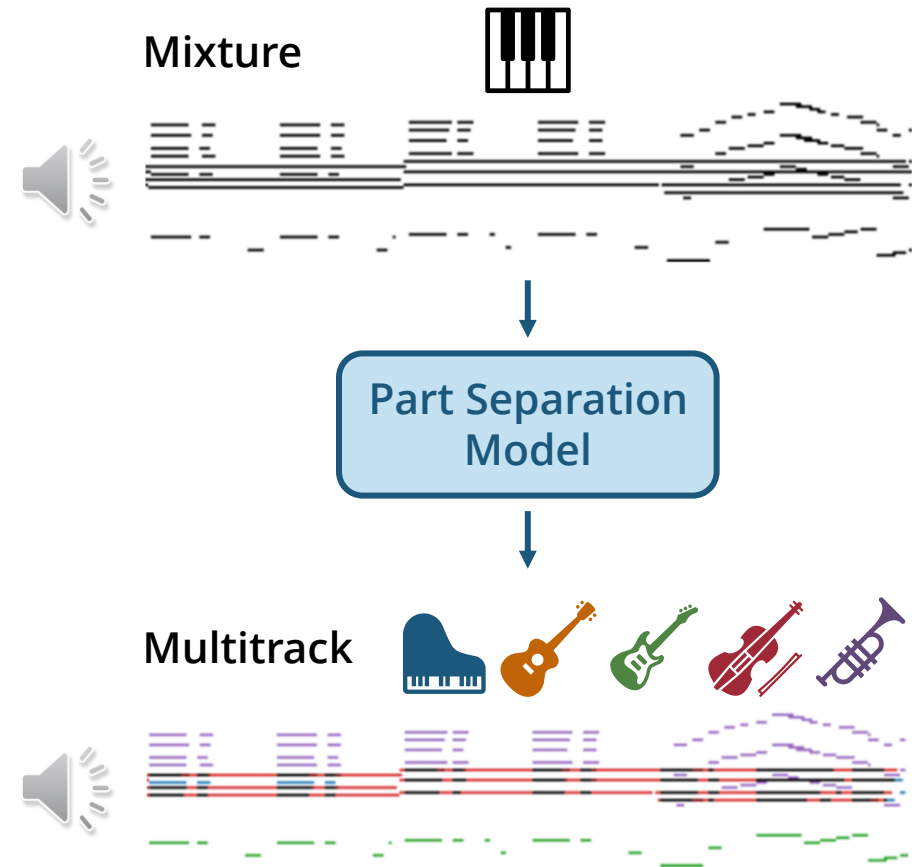
Stanford

UC San Diego

Training



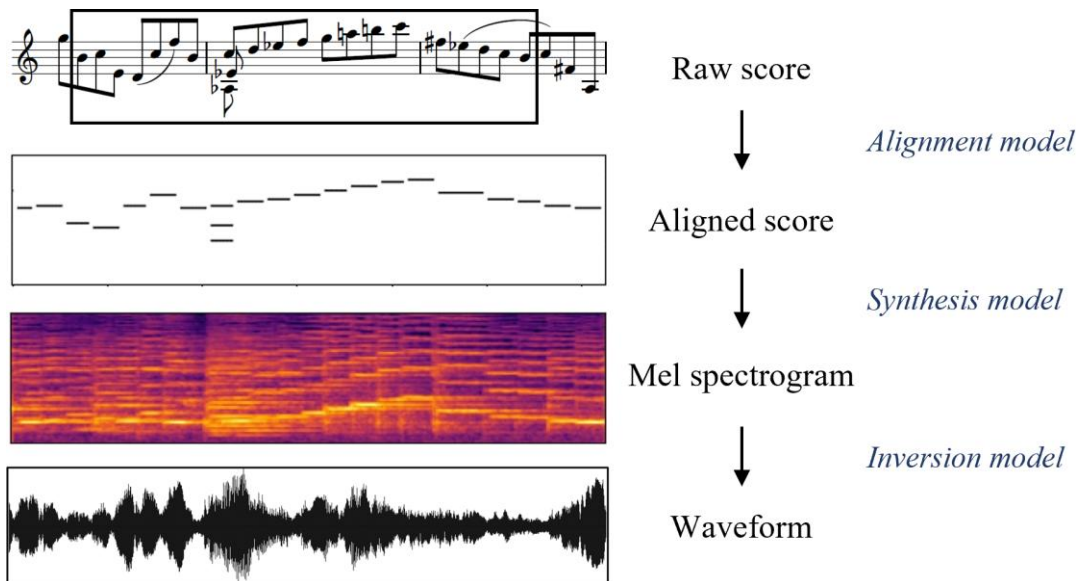
Inference



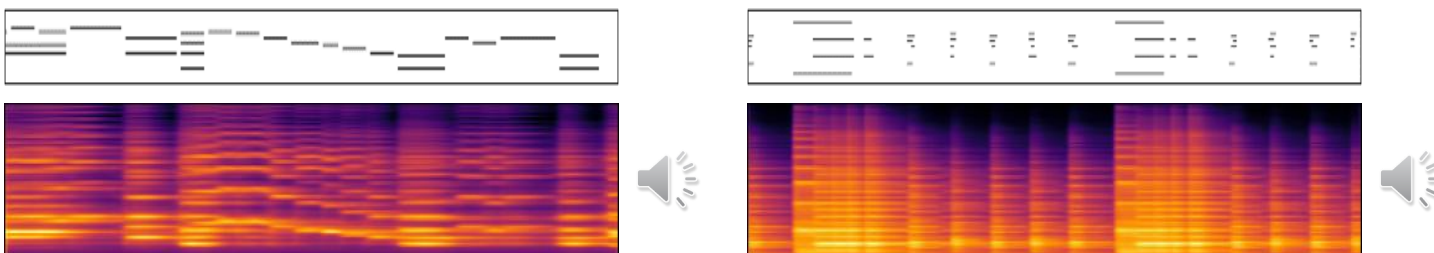
Hao-Wen Dong, Chris Donahue, Taylor Berg-Kirkpatrick and Julian McAuley, "Towards Automatic Instrumentation by Learning to Separate Parts in Symbolic Multitrack Music," *ISMIR*, 2021.

Synthesizing Expressive Violin Performance (ICASSP 2022)

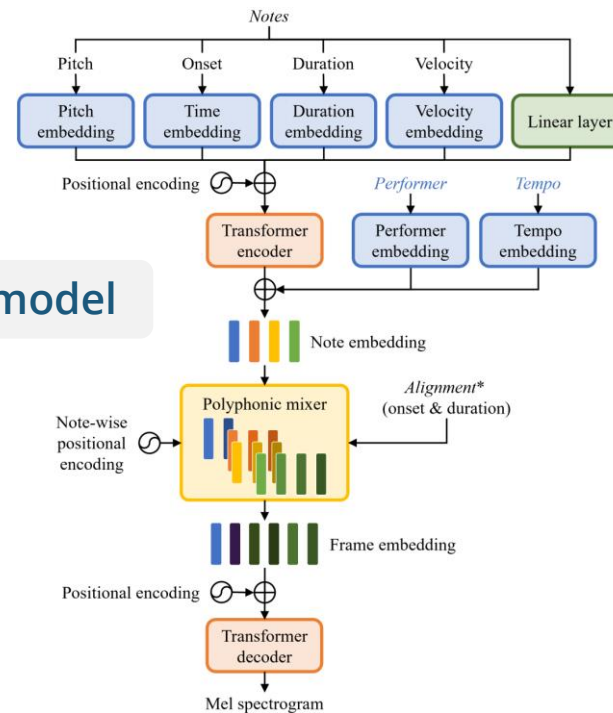
Performance synthesis



Example results



TTS-based model

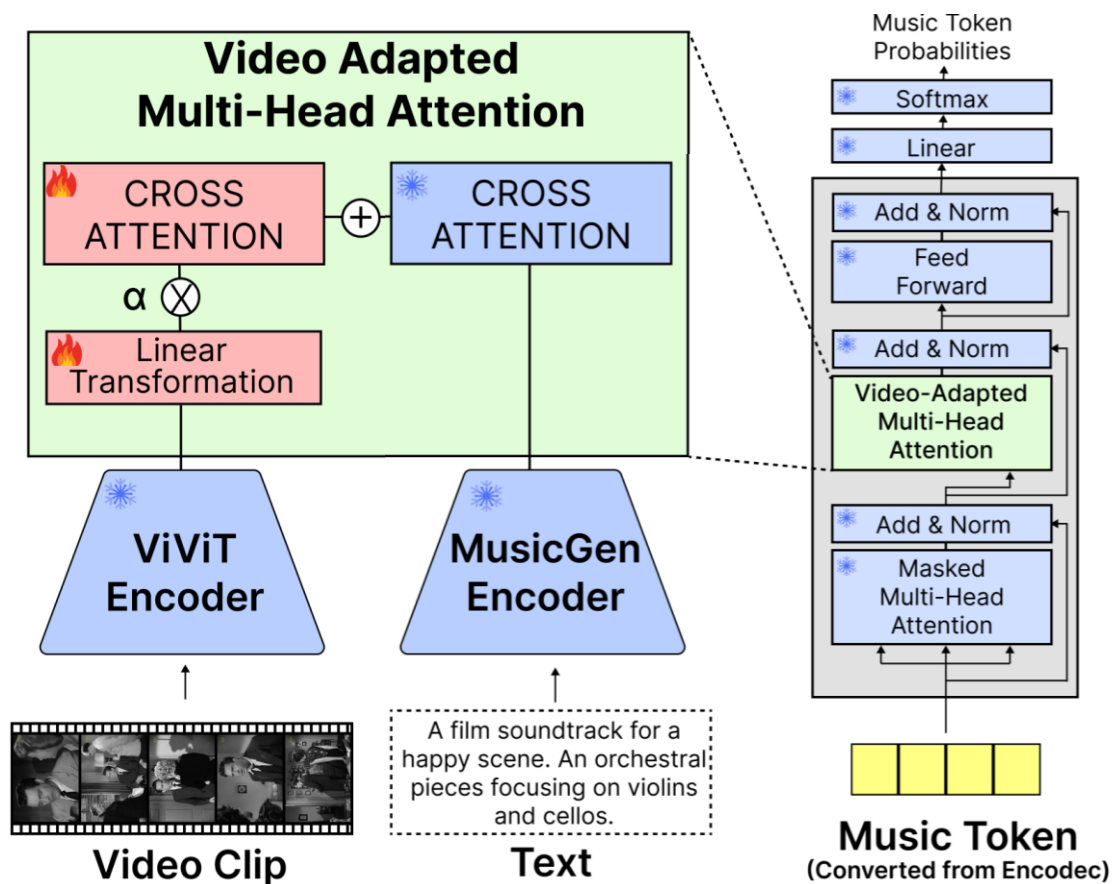


Dolby

UC San Diego

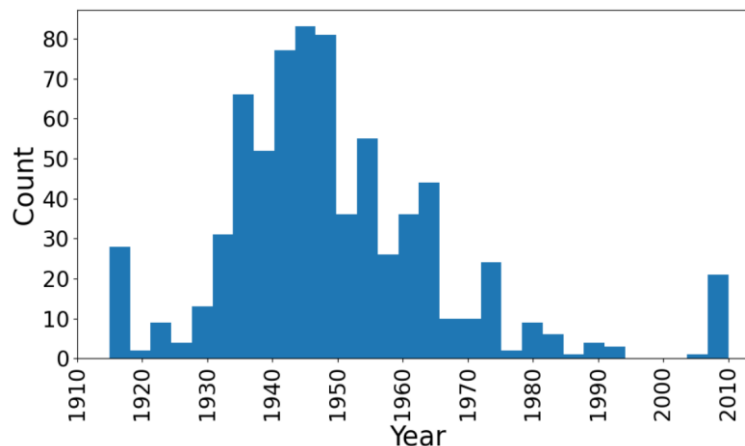
Video-Guided Text-to-Music Generation (ISMIR 2025)

Model



🔥 TRAINABLE ⚙️ FROZEN

Open Screen Soundtrack Library (OSSL)



736 video clips from **299 films**
in **public domain** or **CC-licensed**



UC San Diego Duke University M UNIVERSITY OF MICHIGAN

Video-Guided Text-to-Music Generation Using Public Domain Movie Collections

(ISMIR 2025)

Haven Kim, Zachary Novack, Weihan Xu,
Julian McAuley, Hao-Wen Dong

Challenges & Opportunities

| The Five Challenges

Representations

Usability

Creativity

Multimodality

Personalization

Challenge 1: Representations

How can we best represent music for machine learning?

Music Generation: Four Paradigms



Symbolic music generation

Text-based

Image-based

```
Program_change_0,  
Note_on_60, Time_shift_2, Note_off_60,  
Note_on_60, Time_shift_2, Note_off_60,  
Note_on_76, Time_shift_2, Note_off_67,  
Note_on_67, Time_shift_2, Note_off_67,  
...
```

MIDI



Piano roll



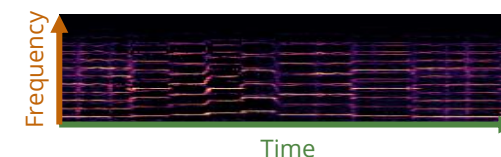
Audio-domain music generation

Time series-based

Image-based



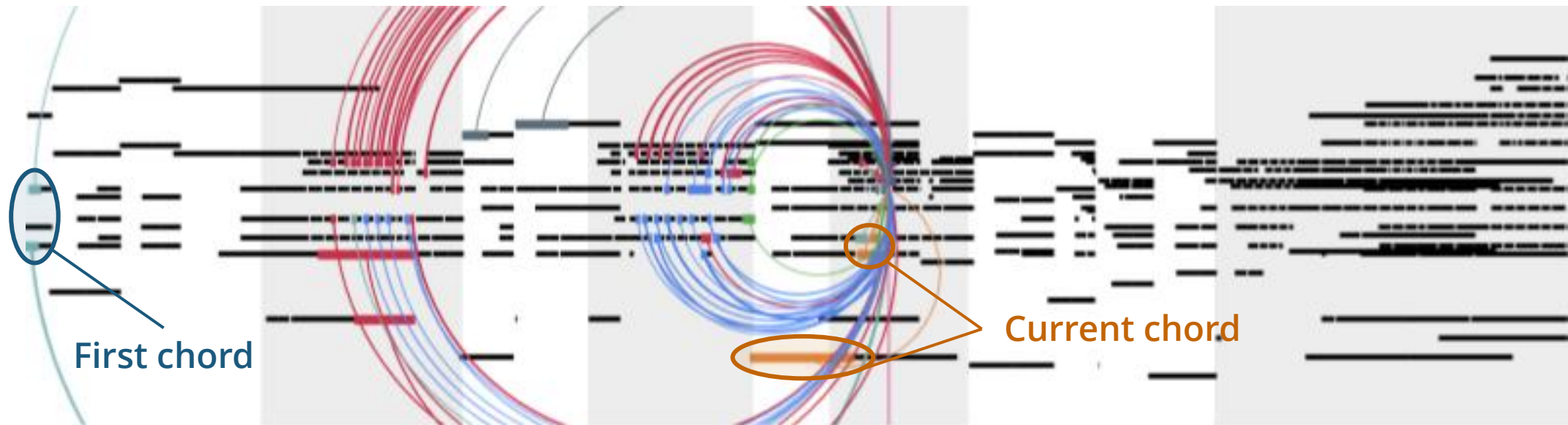
Waveform



Spectrogram

Visualizing Musical Self-attention (Huang et al., 2018)

(Each color represents an attention head)



(Source: Huang et al., 2018)

Analyzing Musical Self-attention (Dong et al., 2023)

- We proposed two new quantities for measuring **mean relative attention**

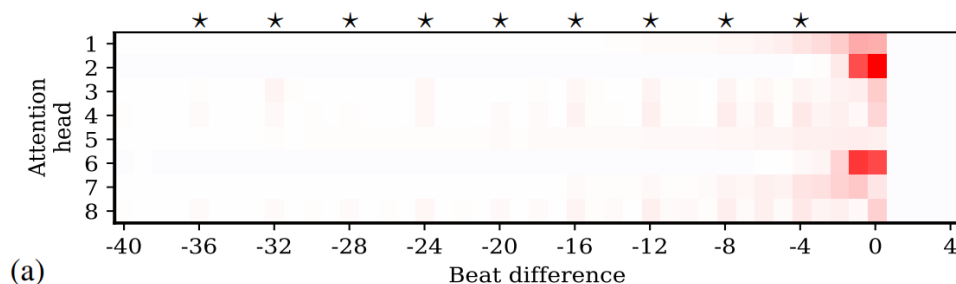
$$\gamma_k^{(d)} = \frac{\sum_{\mathbf{x} \in \mathcal{D}} \sum_{s > t} a_{s,t}(\mathbf{x}) \mathbb{1}_{x_t^{(d)} - x_s^{(d)} = k}}{\sum_{\mathbf{x} \in \mathcal{D}} \sum_{s > t} a_{s,t}(\mathbf{x})} \quad \tilde{\gamma}_k^{(d)} = \gamma_k^{(d)} - \frac{\sum_{\mathbf{x} \in \mathcal{D}} \sum_{s > t} \mathbb{1}_{x_t^{(d)} - x_s^{(d)} = k}}{\sum_{\mathbf{x} \in \mathcal{D}} \sum_{s > t} 1}$$

- The MMT model attends more to notes

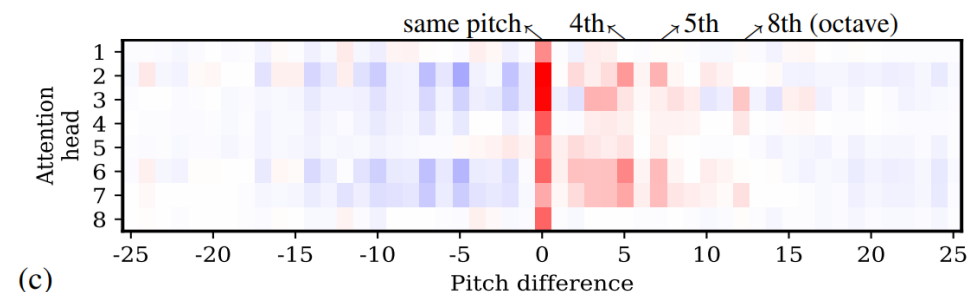
that are $4N$ beats away in the past

that has a pitch in an octave above which forms a consonant interval

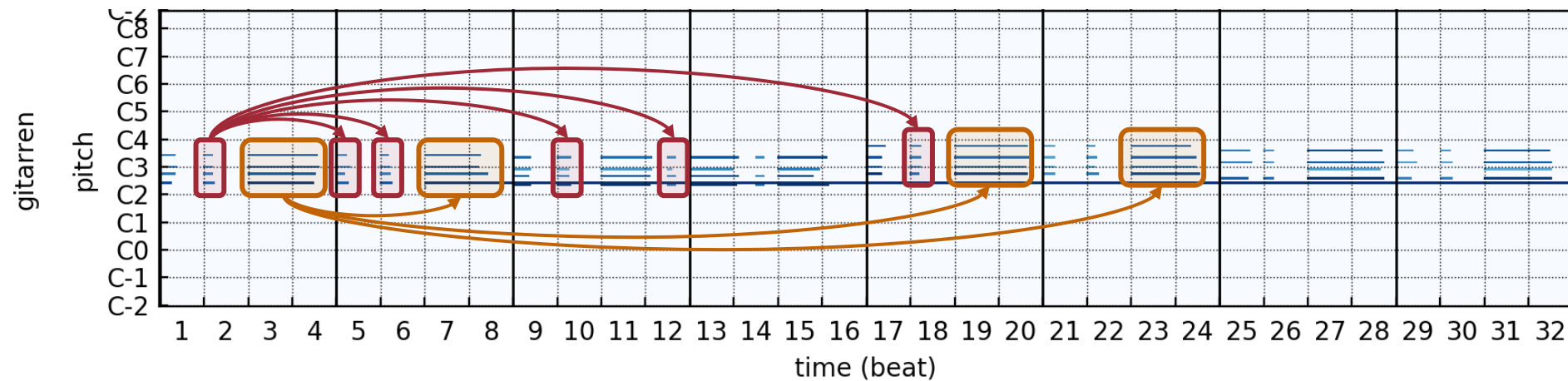
Positive and negative mean relative attention gain



Positive and negative mean relative attention gain



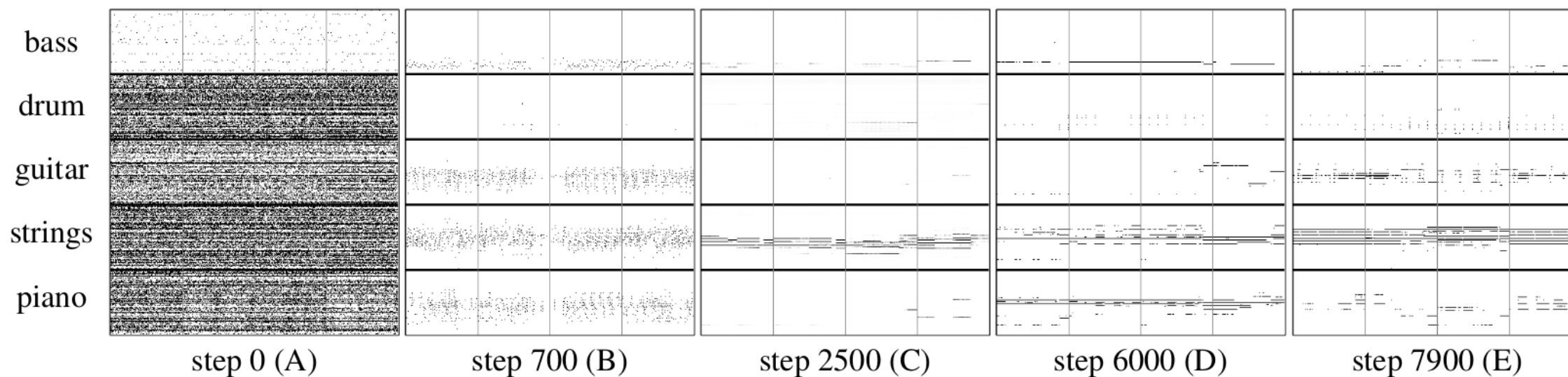
| Why Piano Rolls?



Many musical patterns like melodies, chords, scales and arpeggios
are **translational invariant** in the temporal and pitch axes

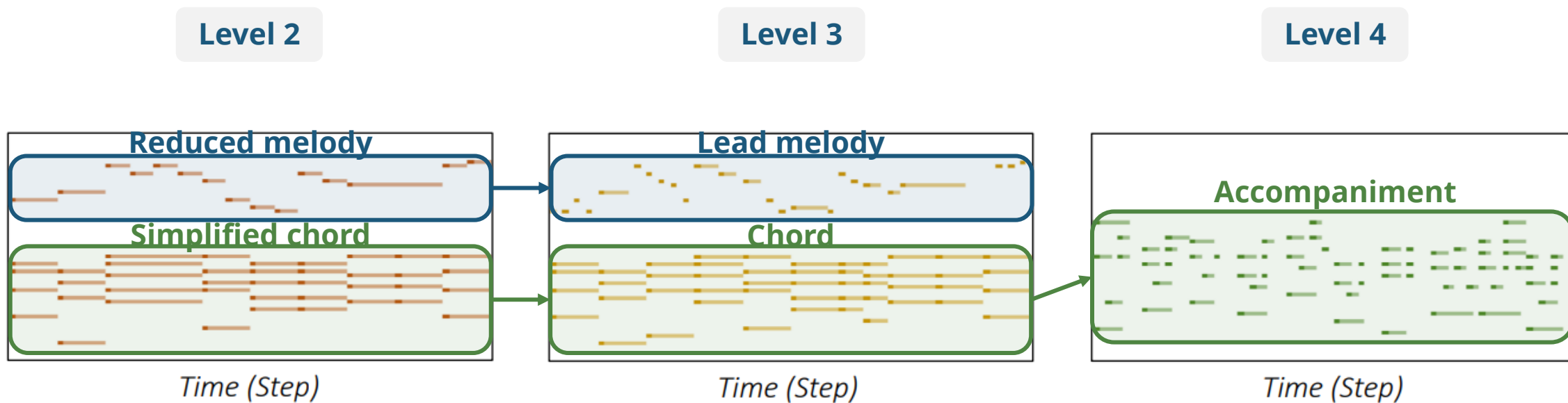
MuseGAN: Generating Multitrack Pianorolls (Dong et al., 2018)

Examples of
generated music



(Source: Dong et al., 2018)

Cascaded Diffusion Models (Wang et al., 2024)



(Source: Wang et al., 2024)

wholesonggen.github.io

Challenge 1: Representations

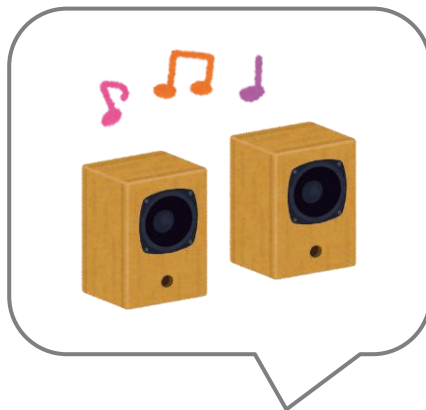
How can we best represent music for machine learning?

Challenge 2: Multimodality

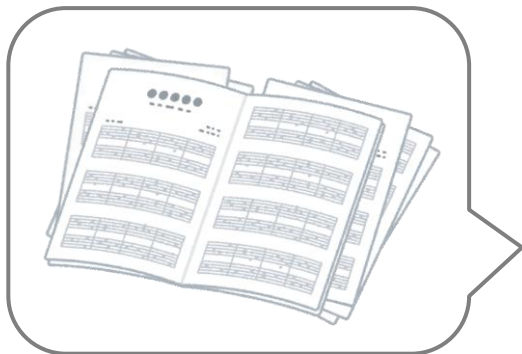
Can AI learn to create music by “listening to” music rather than “reading” music?

Human-inspired Machine Learning for Music & Audio

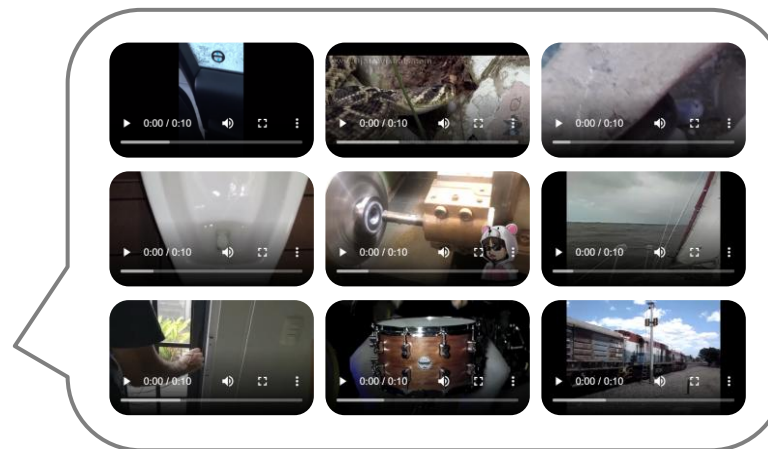
Learning from listening



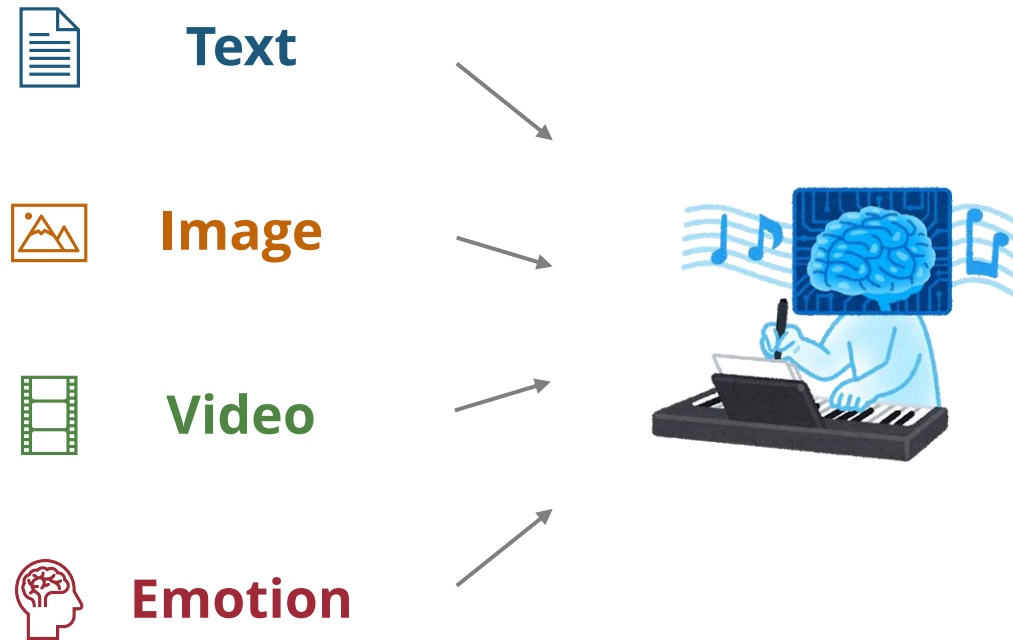
Learning from reading



Learning from watching



Multimodal Inputs for Generative Music AI



| AI Creative Agents (2015)



youtu.be/DggF9m9xqik & github.com/DYCI2/Dicy2

Shimon: An Improvising Robotic Marimba Player (2021)



(Source: Robot Gizmos)

roboticgizmos.com/shimon-musical-robot-deep-learning/
Georgia Tech Center for Music Technology, "Shimon with the Aarhus Jazz Orchestra," *YouTube*, youtu.be/ZpTV1-acSU8, 2021.



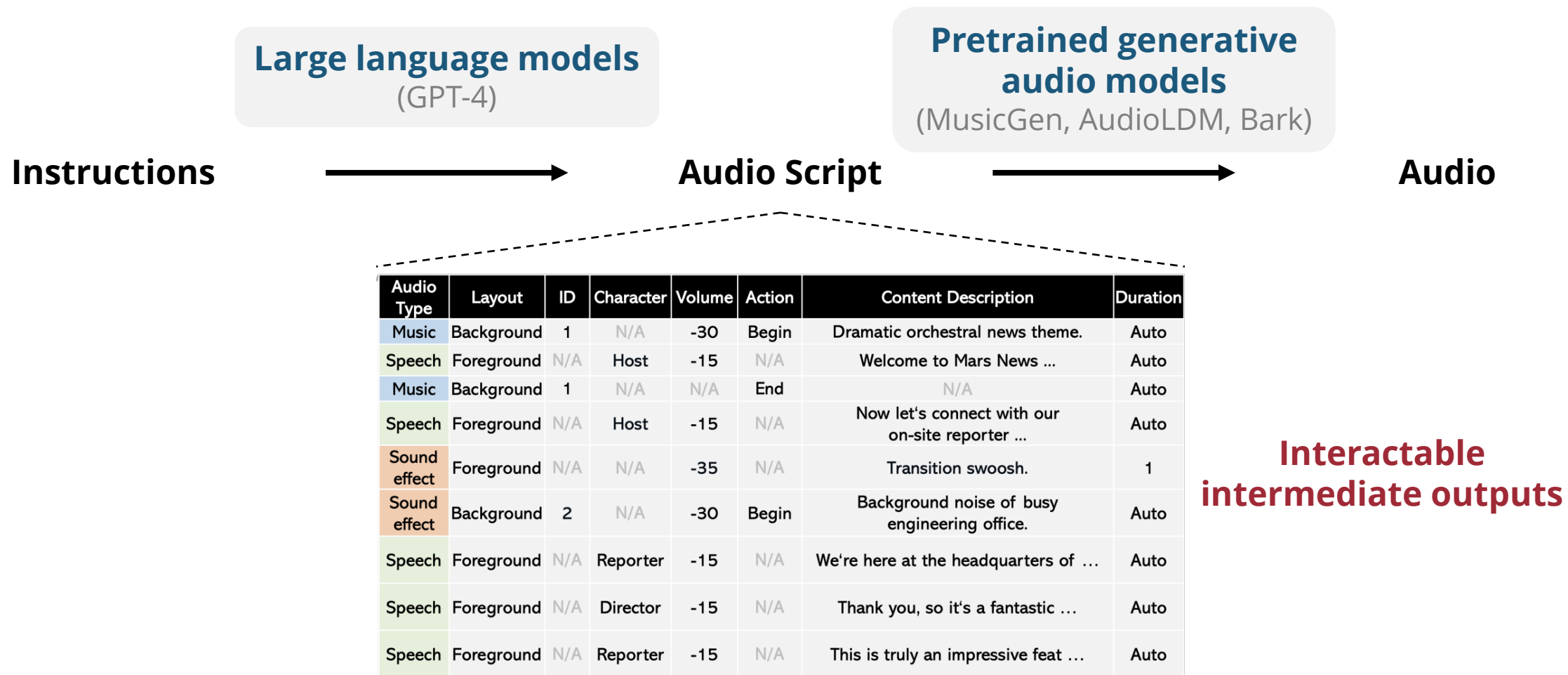
Challenge 2: Multimodality

Can AI learn to create music by “listening to” music rather than “reading” music?

Challenge 3: Usability

How can AI music tools be integrated into an artist's creative workflow?

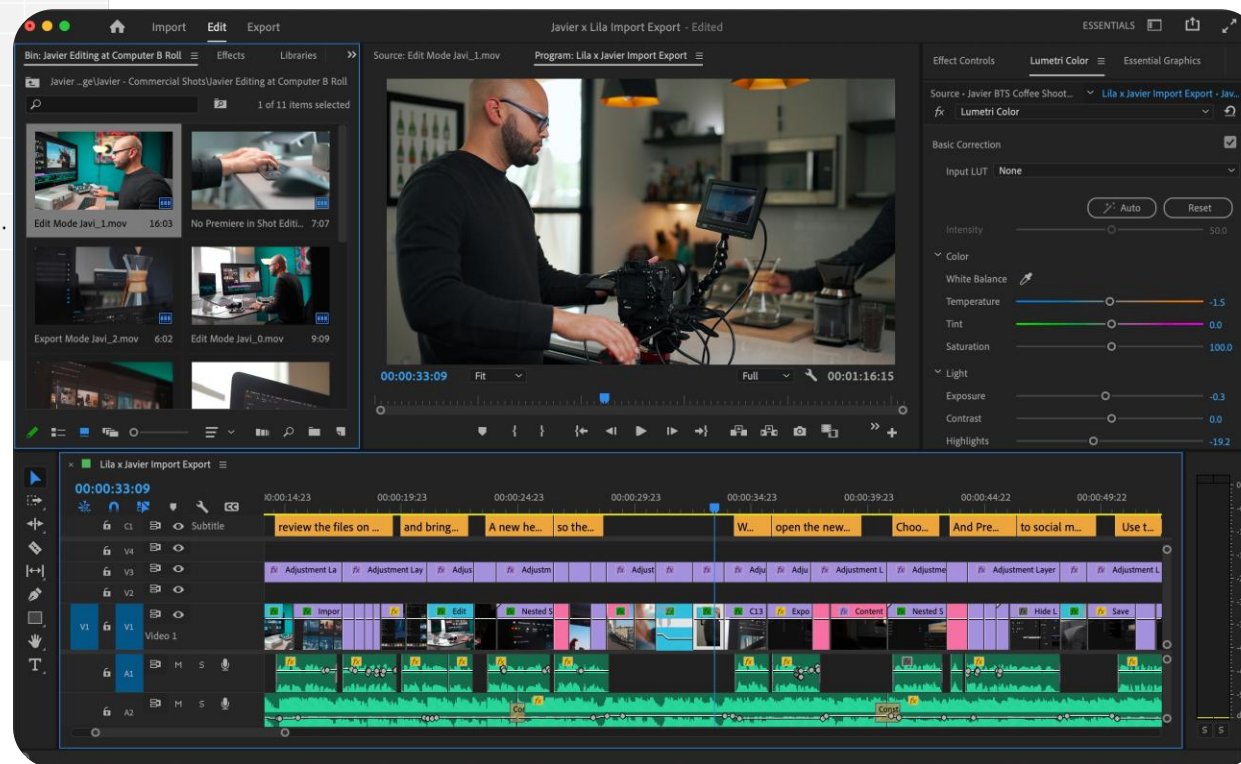
WavJourney: Compositional Audio Creation (Liu et al., 2023)



Xubo Liu, Zhongkai Zhu, Haohe Liu, Yi Yuan, Meng Cui, Qiushi Huang, Jinhua Liang, Yin Cao, Qiuqiang Kong, Mark D. Plumbley, Wenwu Wanga, "WavJourney: Compositional Audio Creation with Large Language Models," *arXiv preprint arXiv:2307.14335*, 2023.

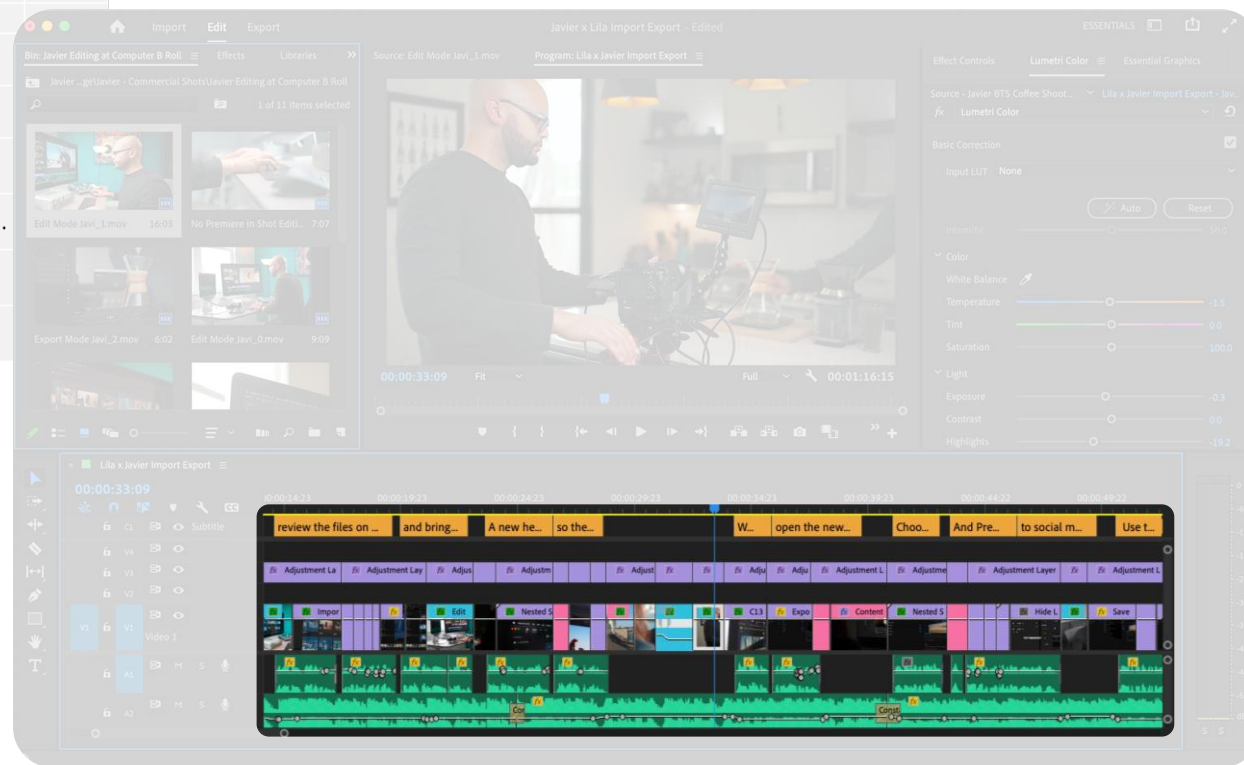
Integrating GenAI into the Creative Workflow

Audio Type	Layout	ID	Character	Volume	Action	Content Description	Duration
Music	Background	1	N/A	-30	Begin	Dramatic orchestral news theme.	Auto
Speech	Foreground	N/A	Host	-15	N/A	Welcome to Mars News ...	Auto
Music	Background	1	N/A	N/A	End	N/A	
Speech	Foreground	N/A	Host	-15	N/A	Now let's connect with our on-site reporter ...	
Sound effect	Foreground	N/A	N/A	-35	N/A	Transition swoosh.	
Sound effect	Background	2	N/A	-30	Begin	Background noise of busy engineering office.	
Speech	Foreground	N/A	Reporter	-15	N/A	We're here at the headquarters of ...	
Speech	Foreground	N/A	Director	-15	N/A	Thank you, so it's a fantastic ...	
Speech	Foreground	N/A	Reporter	-15	N/A	This is truly an impressive feat ...	



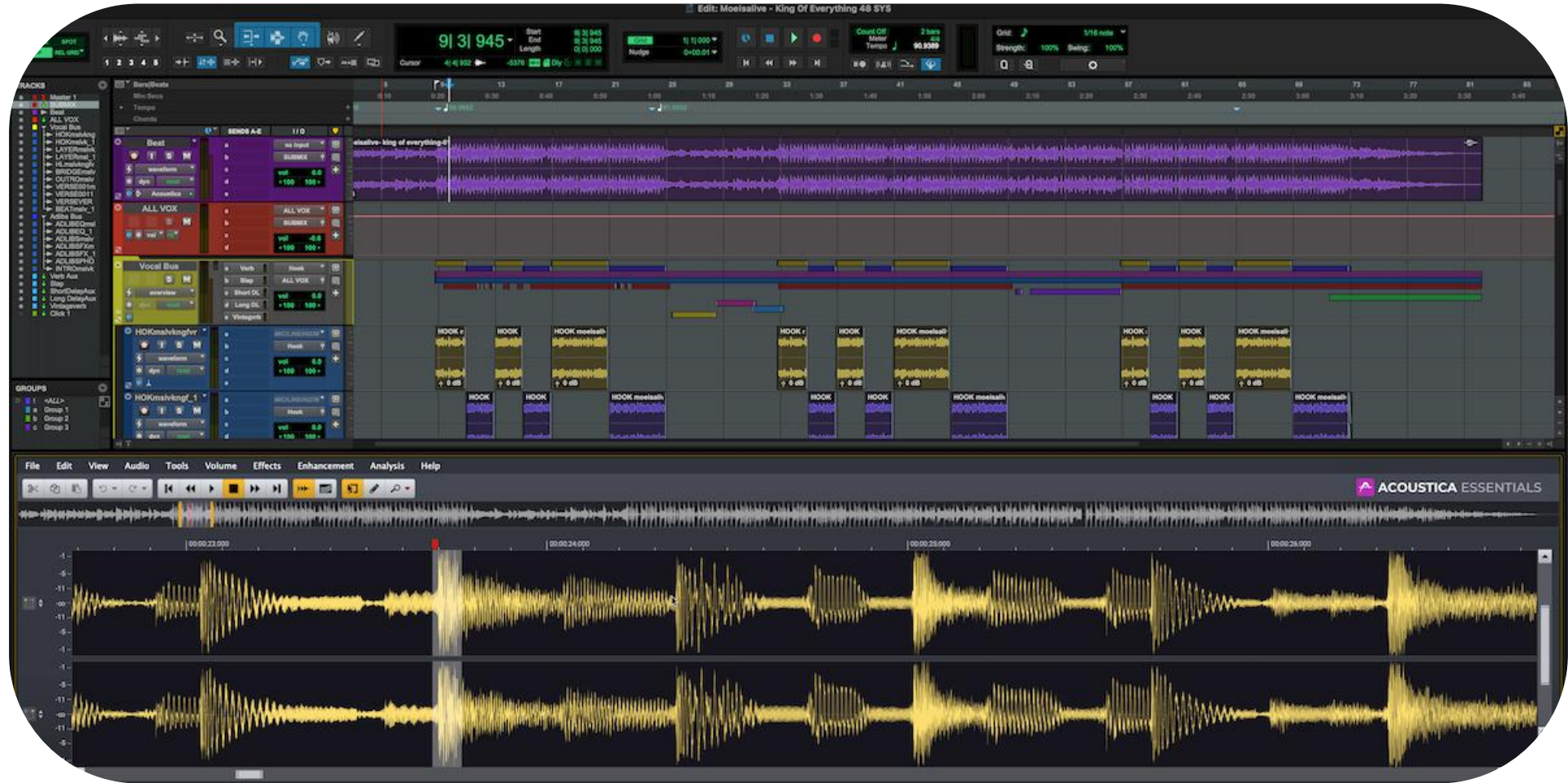
Integrating GenAI into the Creative Workflow

Audio Type	Layout	ID	Character	Volume	Action	Content Description	Duration
Music	Background	1	N/A	-30	Begin	Dramatic orchestral news theme.	Auto
Speech	Foreground	N/A	Host	-15	N/A	Welcome to Mars News ...	Auto
Music	Background	1	N/A	N/A	End	N/A	
Speech	Foreground	N/A	Host	-15	N/A	Now let's connect with our on-site reporter ...	
Sound effect	Foreground	N/A	N/A	-35	N/A	Transition swoosh.	
Sound effect	Background	2	N/A	-30	Begin	Background noise of busy engineering office.	
Speech	Foreground	N/A	Reporter	-15	N/A	We're here at the headquarters of ...	
Speech	Foreground	N/A	Director	-15	N/A	Thank you, so it's a fantastic ...	
Speech	Foreground	N/A	Reporter	-15	N/A	This is truly an impressive feat ...	



Integration into professional creative workflow

Integrating GenAI into the Music Creative Workflow



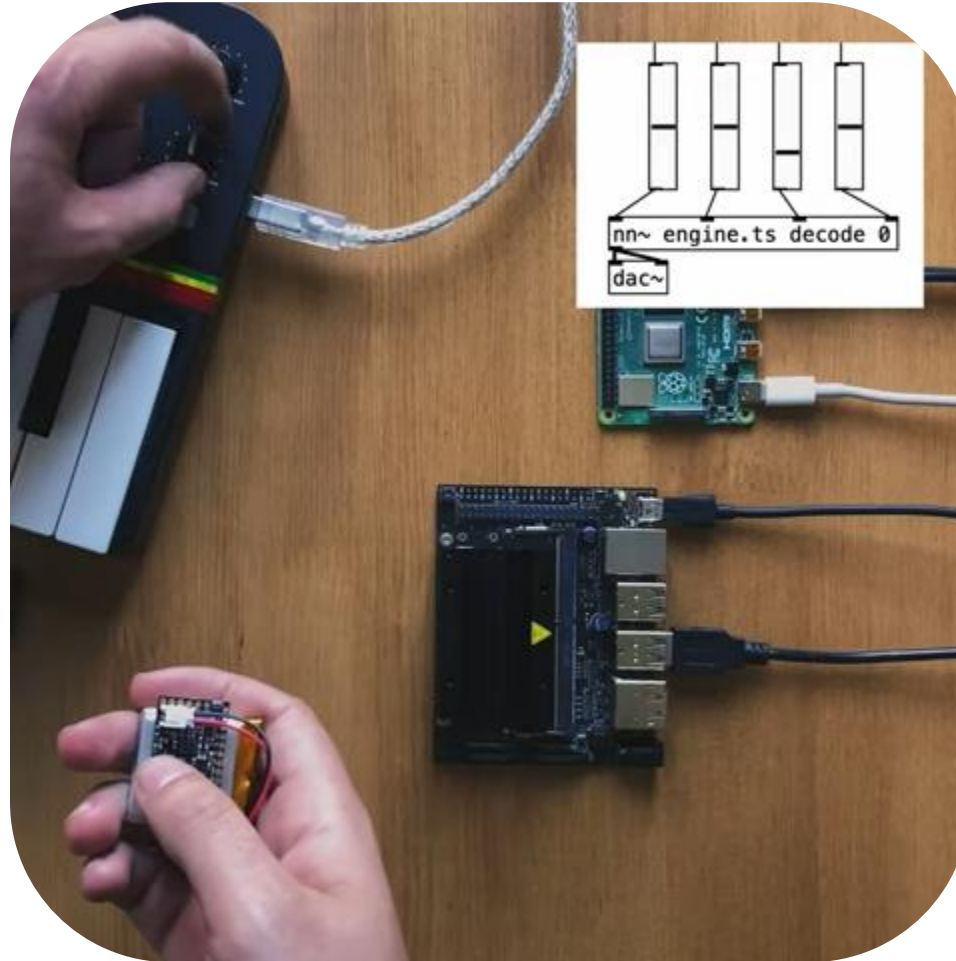
(Source: Avid)

| Integrating GenAI into the Music Creative Workflow



(Source: Avid)

RAVE: Real-time Audio Synthesis (Caillon & Esling, 2022)



youtu.be/jAlRf4nGgYI

Misusable Music Tools (Nao Tokui, 2024)

Throughout history, music and technology have often intertwined, with **new technologies being misused by artists** (turntables, etc).

– Nao Tokui, 2024

AI is more challenging to misuse because **it lacks a physical entity and operates as a black box.**

– Nao Tokui, 2024



(Source: Flintmi via [Wikimedia Commons](#))

x.com/naotokui_en/status/1756180102344499338

Nao Tokui, "Surfing human creativity with AI — A user's guide," *Lulu Press*, 2023.

Flintmi, [CC BY 3.0](#), via [Wikimedia Commons](#)

Without **deviation from the norm**,
progress is not possible.

– Frank Zappa



Challenge 3: Usability

How can AI music tools be integrated into an artist's creative workflow?

Challenge 4: Personalization

How can we make “my personal AI music tools”?

YACHT & Google Magenta

“The band first took all 82 songs from their back catalog and isolated each part, from bass lines to vocal melodies to drum rhythms; they then took those isolated parts and broke them up into four-bar loops. Then, **they put those loops into the machine learning model,** which **put out new melodies based on their old work. They did a similar process with lyrics, using their old songs plus other material they considered inspiring.** The final task was to pick lyrics and melodies that made sense, and pair them together to make a song.”



youtu.be/_yz8QYzcfxI

YACHT, “YACHT — SCATTERHEAD (4K Lyric Video)”, *YouTube*, July 26, 2019.

Megan Friedman, “Behind Magenta, the tech that rocked I/O,” *The Keyword*, May 20, 2019.

Adam Roberts, “YACHT’s new album is powered by ML + Artists”, *Magenta Blog*, September 13, 2019.

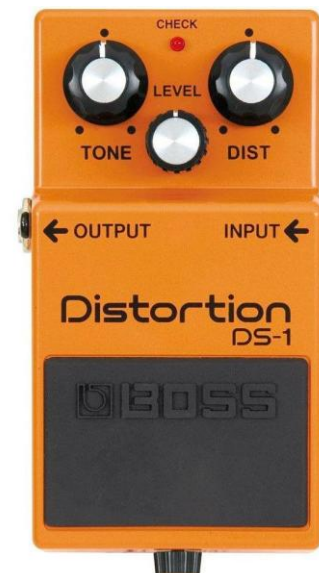
Ease of Personalization for Artists

- Through **finetuning our own models**
- Through **finetuning with live inputs**
- Python scripting vs friendly user interface

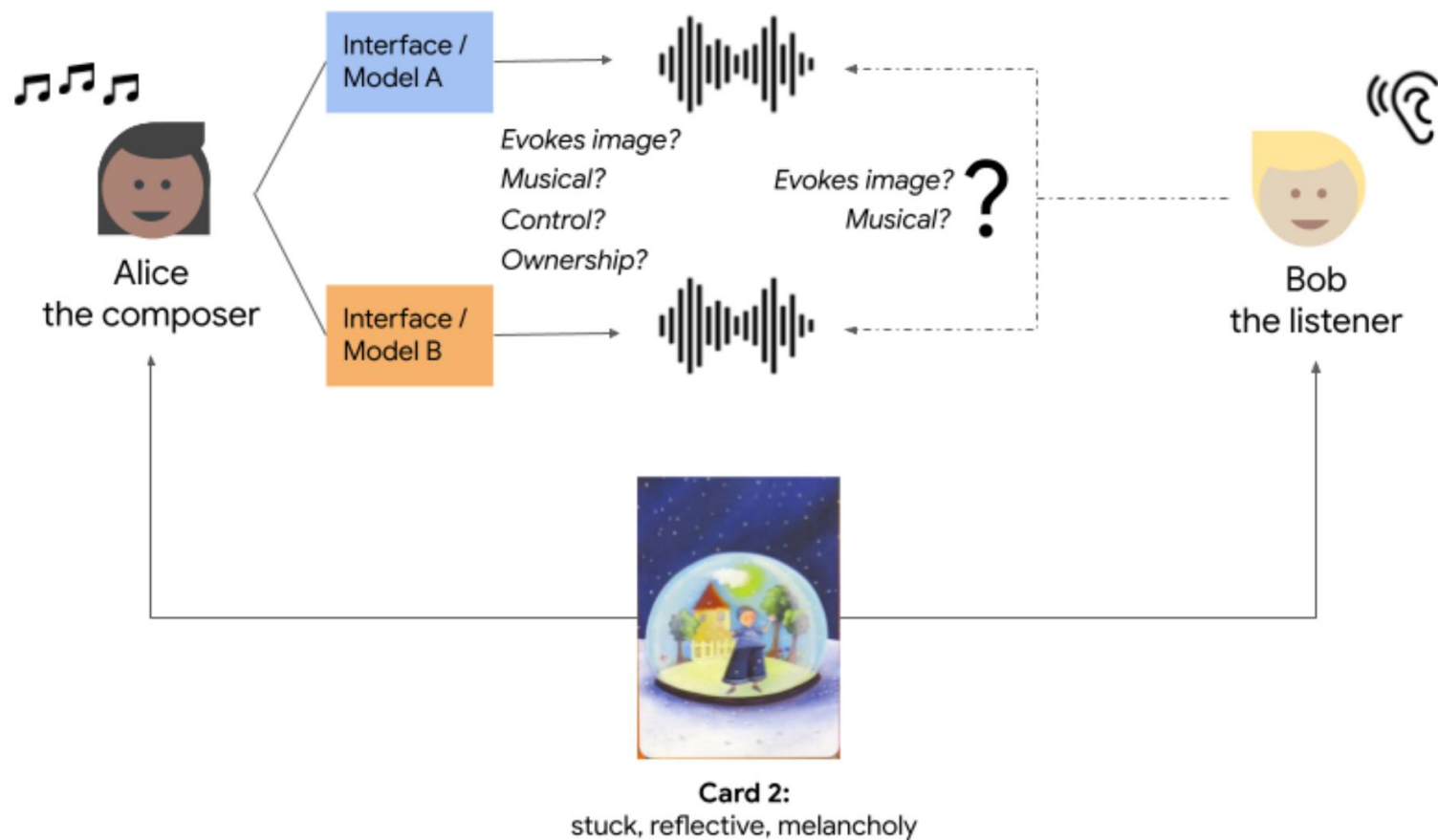
- **Can we do better?**

Overfitting vs Distortion

- Will **overfitting** be a new music expression, the “**distortion**” for AI music?



Expressive Communication: Alice & Bob (Louie, 2022)



(Source: Louie et al., 2022)

Challenge 4: Personalization

How can we make “my personal AI music tools”?

Challenge 5: Creativity

Can AI ever be creative? How can AI augment human creativity?



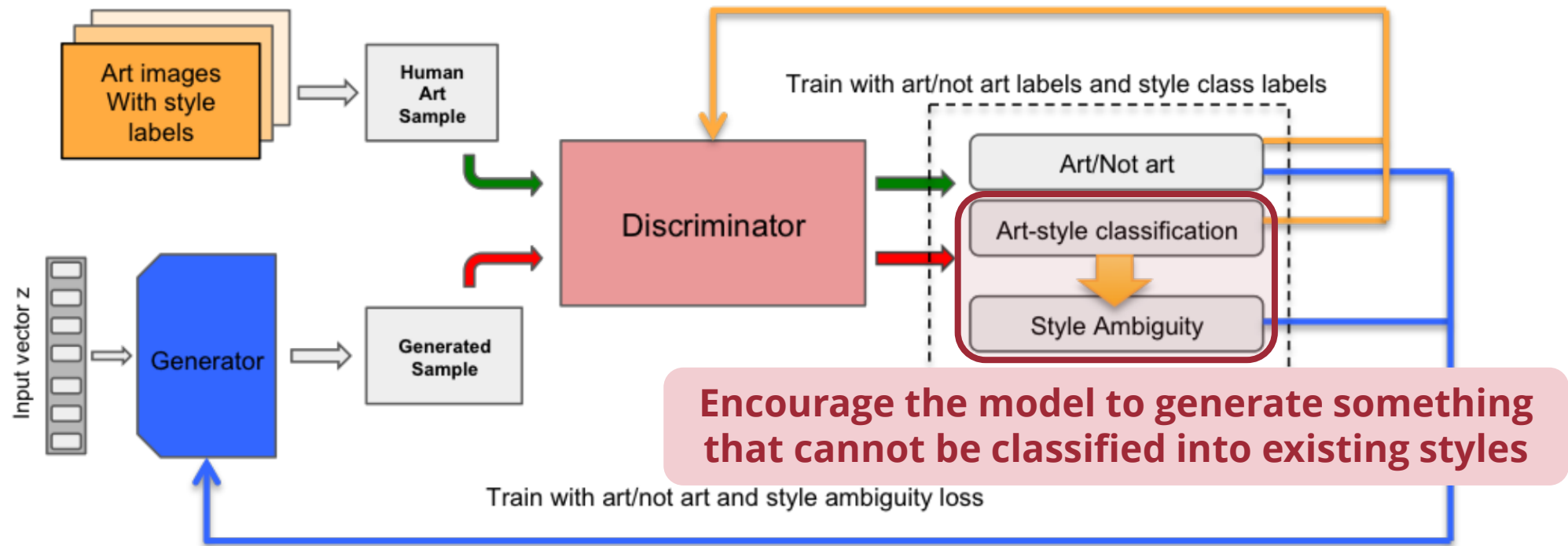
**Works of art make rules;
rules do not make works of art.**

– Claude Debussy

| The Curse of Machine Learning

- As the old saying goes, “**AI is only as good as the data it learns from.**”
- Machine learning models are trained to approximate some distribution in its formal formulation.
- This seems to contradict the idea of creativity that requires **the ability to extrapolate** and **think out of the box**.
- **Can AI ever be creative?**

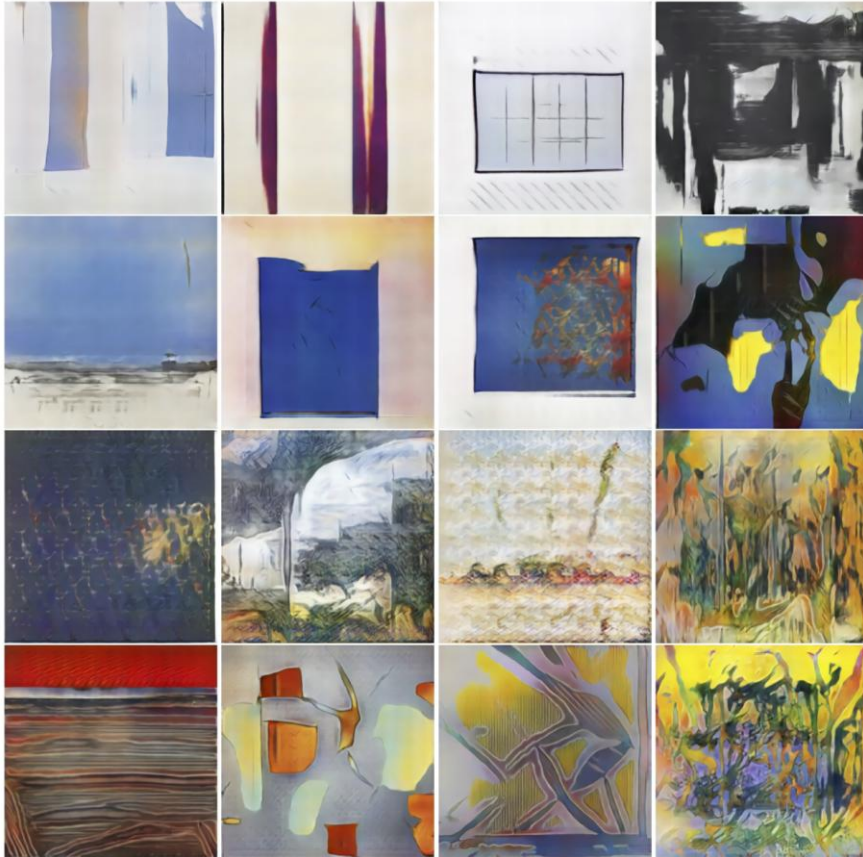
Creative Adversarial Network (Elgammal et al., 2017)



(Source: Elgammal et al., 2017)

Creative Adversarial Network (Elgammal et al., 2017)

Example generated images



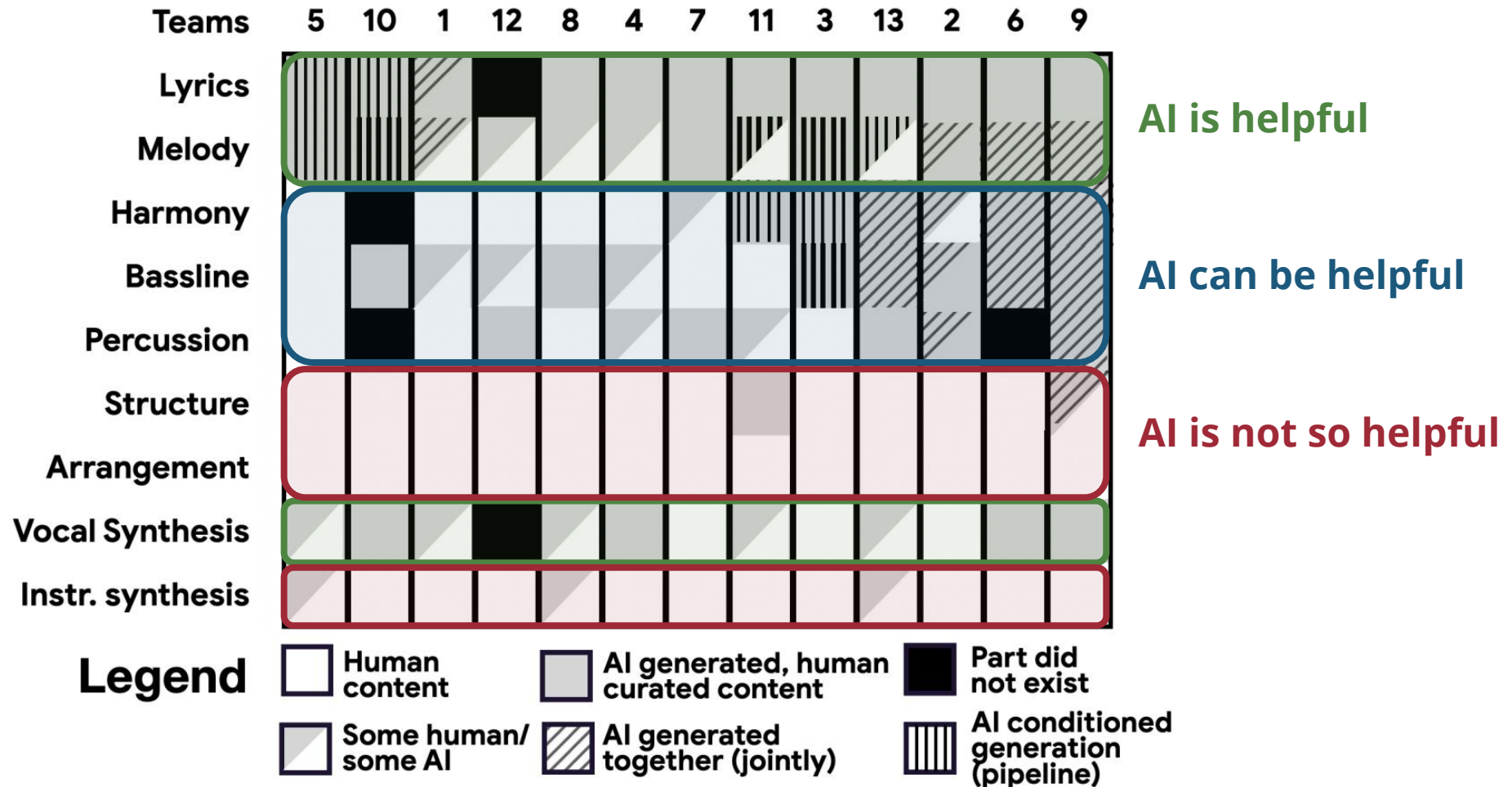
Best samples



(Source: Elgammal et al., 2017)

Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone, "[CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms](#)," ICCV, 2017.

How can AI Augment Human Creativity?



(Source: Huang et al., 2020)

| What is Creativity?

Creativity is the ability to come up with ideas or artefacts that are **new**, **surprising** and **valuable**.

– Margaret Boden, 2007

Three Types of Creativity (Boden, 2007)

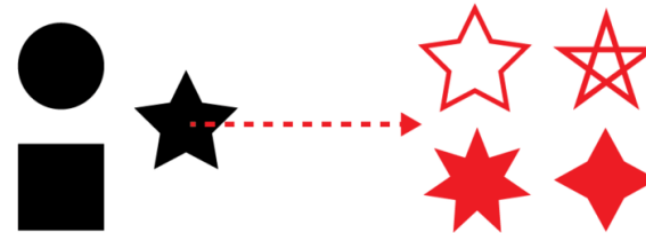
Combinatorial Creativity

Combining existing ideas and things into something new



Exploratory Creativity

Exploring possibilities within a domain



Transformative Creativity

Radically new ideas that redefine domain and applicable rules



(Source: van Kuijk, 2023)

| AI is Good at Combinatory Creativity

Prompt: A Michigan space wolverine



Can AI learn Transformative Creativity?

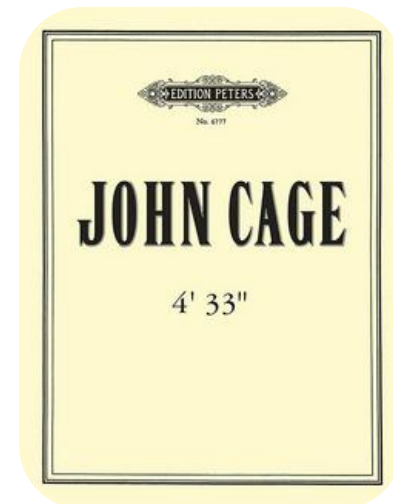
Duchamp & conceptual art



Picasso & cubism



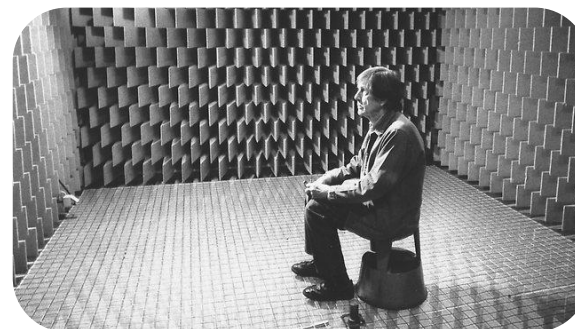
Cage & modernist music



Schoenberg & atonality

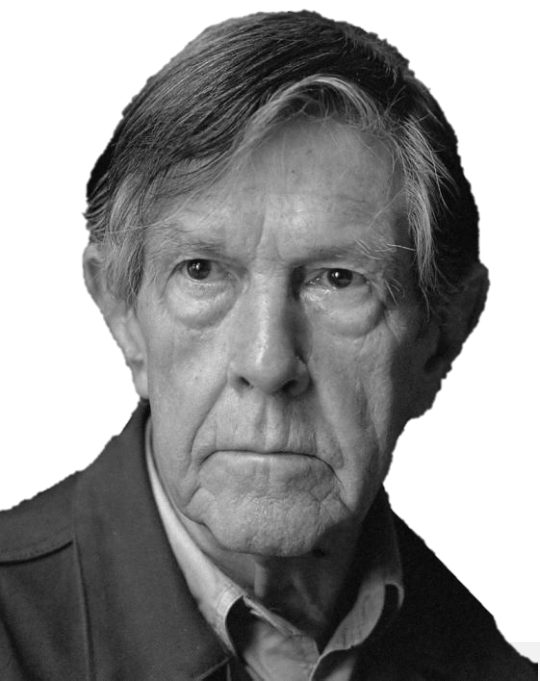


| John Cage's 4'33" (1952)



They missed the point. **There's no such thing as silence.** What they thought was silence, because they didn't know how to listen, was full of accidental sounds. You could hear the wind stirring outside during the first movement. During the second, raindrops began pattering the roof, and during the third the people themselves made all kinds of interesting sounds as they talked or walked out.

– John Cage, on the premiere of 4'33", 1952



| Creativity vs Art



Creativity is allowing yourself to **make mistakes**.
Art is knowing **which ones to keep**.

– Scott Adams

Challenge 5: Creativity

Can AI ever be creative? How can AI augment human creativity?

| The Five ~~Challenges~~ Opportunities!

Representations

Multimodality

Usability

Personalization

Creativity

- **Representations**: How can we best represent music for machine learning?
- **Multimodality**: Can AI learn to create music by “listening to” music rather than “reading” music?
- **Usability**: How can AI music tools be integrated into an artist’s creative workflow?
- **Personalization**: How can we make “my personal AI music tools”?
- **Creativity**: Can AI ever be creative? How can AI augment human creativity?

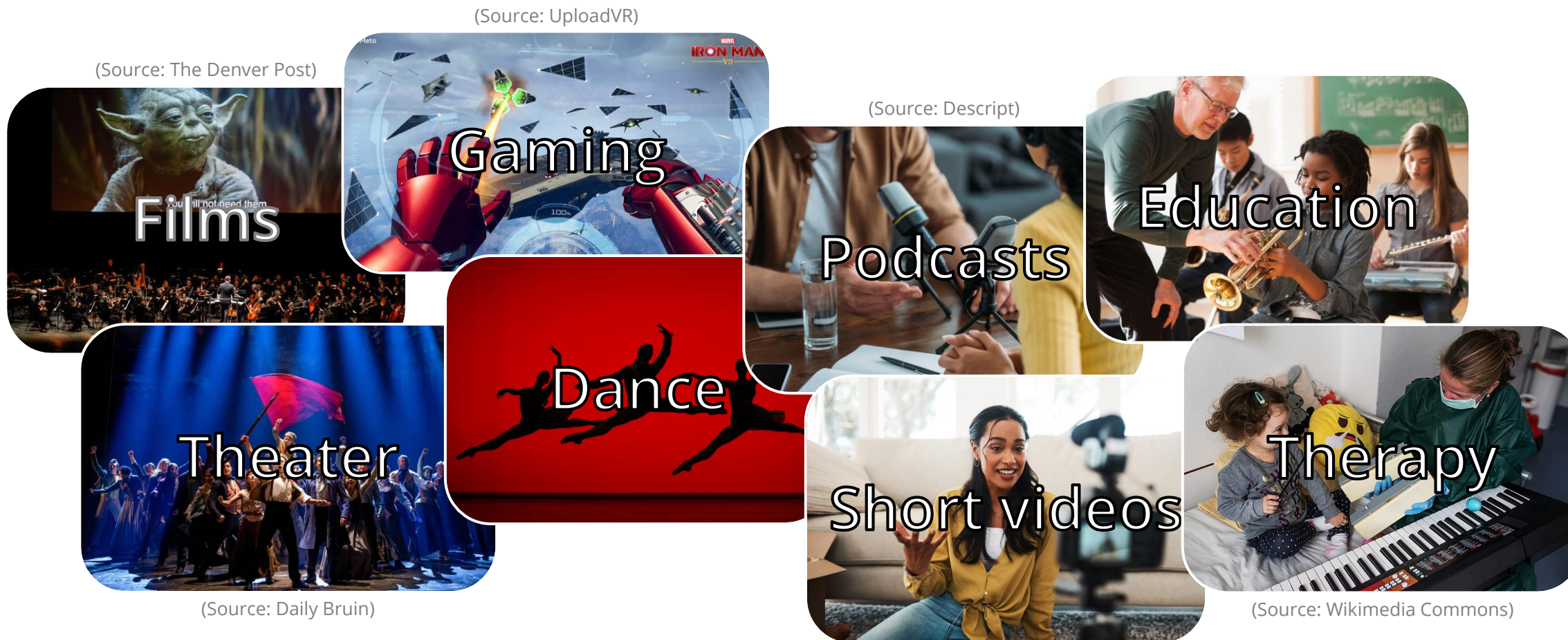
Conclusion

Music & Technology Co-evolves



Hildegard Dodel, Public domain, via Wikimedia Commons.
Taken at Hamamatsu Museum of Musical Instruments, August 2019.
yan, [CC BY-SA 4.0](#), via Wikimedia Commons.

Generative AI for Content Creation



Universitaetsmedizin, [CC BY-SA 4.0](#), via Wikimedia Commons
[uploadvr.com/iron-man-vr-breaks-free-from-cords-load-screens-on-quest-2/](#)
[descript.com/blog/article/what-is-the-best-audio-interface-for-recording-a-podcast](#)
[denverpost.com/2019/08/02/colorado-symphony-movie-scores-harry-potter-star-wars/](#)
[dailybruin.com/2023/08/04/theater-review-the-musical-les-misrables-offers-stellar-displays-and-impassioned-vocals](#)

Art challenges Technology




Music

**Augmenting Human Creativity
with AI**


AI

Technology inspires the Art



| The Five ~~Challenges~~ Opportunities!

Representations

Usability

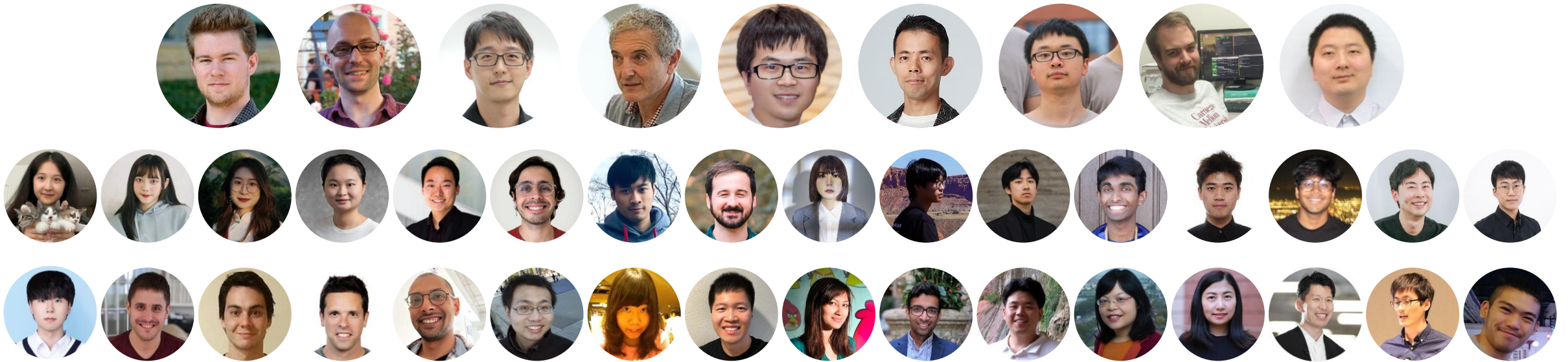
Creativity

Multimodality

Personalization

Generative AI for Music: Challenges & Opportunities

Nothing would have been possible without all my fantastic collaborators!



UC San Diego



SONY



hermandong.com / hwdong@umich.edu

