PAT 463/563 (Fall 2025)

# Music & AI

## Lecture 18: Discussions, Challenges & Opportunities

Instructor: Hao-Wen Dong

UNIVERSITY OF MICHIGAN

# 🚀 Project: Presentation & Report

- **Presentation** in class on **Dec 1**
  - **10-min** presentation that summarizes your motivations, methods, results, analysis and discussions
  - You may follow any structure that best suits your narrative
- **Report** due at **11:59pm ET** on **Dec 15**
  - A **2 to 3-page** (excluding references) report that summarizes your motivations, methods, results, analysis and discussions
  - You may use any template
- **No late submissions!** Submit your work early and update it later.

# 🚀 Project: Rubrics

- **Presentation** (**20pt**)
  - Attendance (10pt)
  - Clarity (5pt)
  - Organization & presentation (5pt)
- **Report** (**20pt**)
  - Writing clarity (5pt)
  - Organization & presentation (5pt)
  - Results (5pt)
  - Discussion (5pt)

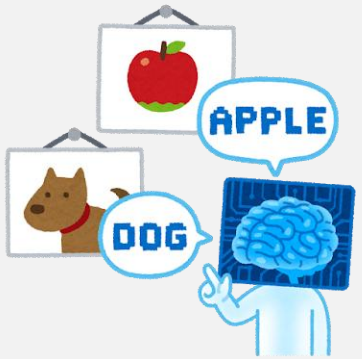
# Course Evaluation

- Your feedback is highly appreciated!
- Enter at umich.bluera.com/umich

# What is AI?

# What is Artificial Intelligence?

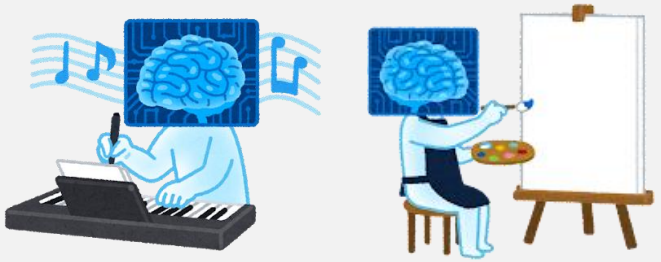
**Systems that think like humans**




**Systems that think rationally**

**Systems that act like humans**

**Systems that act rationally**



# 🤔 What is your expectation of a Real AI?

- **Thinking vs Acting**
  - An AI **needs to know how to think**
  - An AI **doesn't need to know how to think**

- **Human vs AI**
  - An AI **needs to behave like a human**
  - An AI **doesn't need to behave like a human**

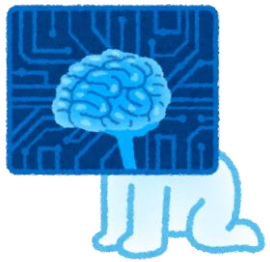
# 🤔 What is your expectation of a Real AI?

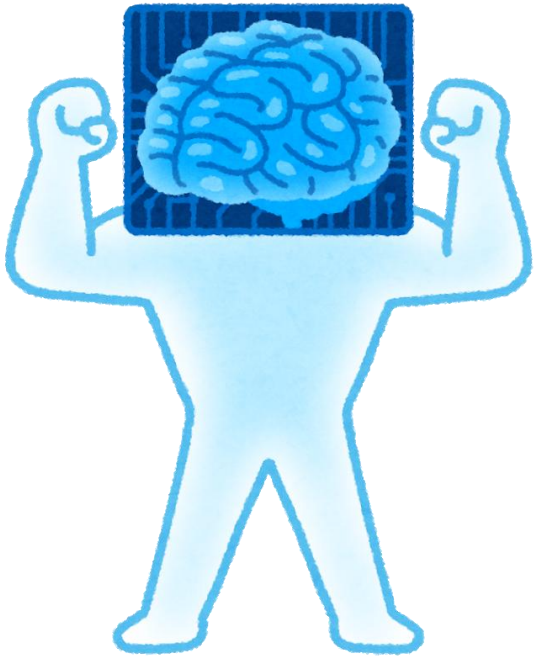
- **Thinking vs Acting**
  - An AI **needs to know how to think** **(Votes: 0)**
  - An AI **doesn't need to know how to think** **(Votes: 3)**

- **Human vs AI**
  - An AI **needs to behave like a human** **(Votes: 0)**
  - An AI **doesn't need to behave like a human** **(Votes: 3)**

# 🤔 Is current music AI more capable than you thought?

- **Yes** or **No**?
- **Yes**, in what aspects?
- **No**, in what aspects?





# 🤔 Discussions

- **To what extent of human involvements** can a song still be called AI music?
- **Shall we intervene** if AI-generated material doesn't sound polished?
- What is the **goal of AI music**?

> "Whatever you now find weird, ugly, uncomfortable and nasty about a new medium will surely **become its signature**."
>
> – Brian Eno, 1996

# Copyright

# Purpose of Copyright

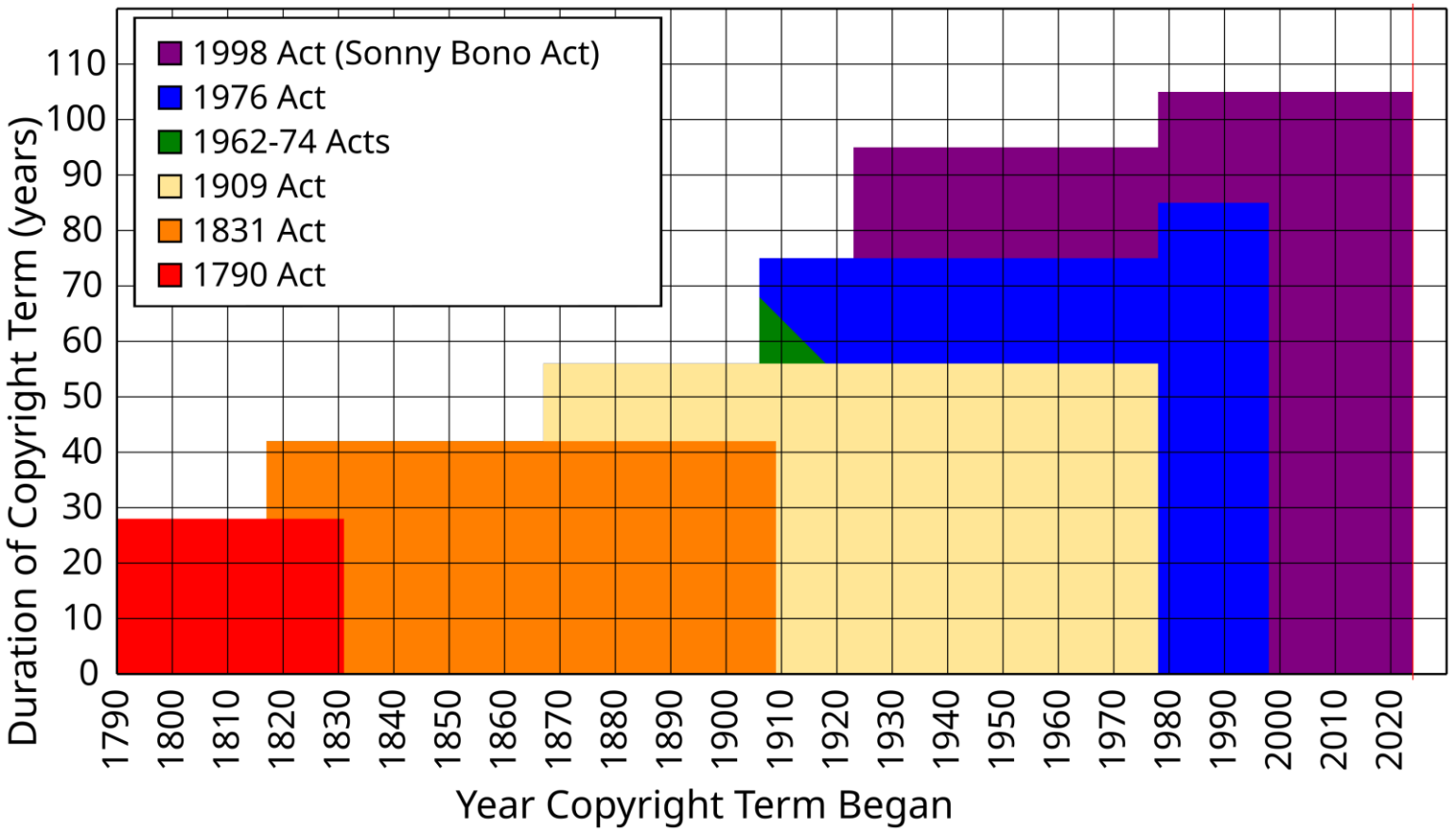
- According to Copyright Alliance:
  - "The primary purpose of copyright is to **induce and reward authors**, through the provision of property rights, to create new works and to make those works available to the public to enjoy."
  - "The theory is that by granting certain exclusive rights to creators that allow these creators to protect their creative works against theft, **creators receive the benefit of economic rewards** and **the public receives the benefit of the creative works** that might not otherwise be created or disseminated."

copyrightalliance.org/education/copyright-law-explained/copyright-basics/purpose-of-copyright/

# Paul McCartney on AI & Copyright


youtu.be/DpebXMTlYNo

# Expansion of Copyright Law




(Source: Tom Bell via Wikimedia Commons)

# 🤔 Purpose of Copyright

- What about **meme culture**?
- Think about **portrait vs photography**
  - While photography displace realistic portraits, it also **frees portrait painting into new art forms** like impressionism and abstract arts
  - Will AI music **raise the collective standards of music** for the public?

# Fairly Trained: L Certification

- All of the **training data used for the model(s)** being certified must fall into one of the following categories:
  - Be **explicitly provided to the model developer for the purposes of being used as training data**, according to a contractual agreement with a party that has the rights required to enter such an agreement
  - Be available under an **open license** appropriate to the use-case
  - Be in the **public domain globally**
  - Be **fully owned by the model developer**


**Fairly Trained Certified**

fairlytrained.org

# Fairly Trained Certified Companies













fairlytrained.org/certified-models

# Fairly Trained Certified Products & Models








fairlytrained.org/certified-models

# 🤔 Music Gen AI & Copyright

- How does AI-generated music compare to other industry-disrupting technologies like **recording** and **music sampling**?
- Does model training justify **fair use**?
- Is a music Gen AI model **a derivative of its training dataset**?
- Can a music Gen AI model **reproduce a song** in its training dataset?

# The Many Lawsuits about Generative AI

The New York Times

**The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work (Published 2023)**

The New York Times sued OpenAI and Microsoft for copyright infringement on Wednesday, opening a new front in the increasingly intense legal battle.

Dec 27, 2023

Reuters

**Music publishers ask court to halt AI company Anthropic's use of lyrics**

Three music publishers are asking a federal court judge to issue a preliminary injunction that would prevent artificial intelligence company...

Nov 17, 2023


AP News

**Reddit sues AI company Anthropic for allegedly 'scraping' user comments to train chatbot Claude**

Social media platform Reddit sued the artificial intelligence company Anthropic on Wednesday, alleging that it is illegally "scraping" the...

Jun 4, 2025


AP News

**Disney and Universal sue AI firm Midjourney for copyright infringement**

Disney and Universal sued popular artificial intelligence image-generator Midjourney on Wednesday, marking the first time major Hollywood companies have taken...

Jun 11, 2025

AP News

**Warner Bros. sues Midjourney for AI-generated images of Superman, Bugs Bunny and other characters**

The lawsuit alleges Midjourney trained its AI system on "illegal copies" of Warner Bros. works and encourages its users to pick iconic...

Sep 5, 2025

AP News

**Reddit sues AI company Perplexity and others for 'industrial-scale' scraping of user comments**

Social media platform Reddit sued the artificial intelligence company Perplexity AI and three other entities on Wednesday, alleging their...

1 month ago

nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html
reuters.com/legal/music-publishers-ask-court-halt-ai-company-anthropics-use-lyrics-2023-11-17/
apnews.com/article/reddit-sues-ai-company-anthropic-claude-chatbot-f5ea042beb253a3f05a091e70531692d
apnews.com/article/disney-universal-midjourney-copyright-lawsuit-722b1b892192e7e1628f7ae5da8cc427
apnews.com/article/warner-bros-midjourney-ai-copyright-lawsuit-dc-studios-b87d80d7b4a4dfdcf0ee149d30830551
apnews.com/article/reddit-sues-ai-company-anthropic-claude-chatbot-f5ea042beb253a3f05a091e70531692d

# UMG, Sony & Warner v. Udio & Suno (June 2024)

AP News

## Music record labels sue AI song-generators Suno and Udio for copyright infringement

The Recording Industry Association of America announced the lawsuits Monday brought by labels including Sony Music Entertainment, Universal...

Jun 24, 2024

Reuters

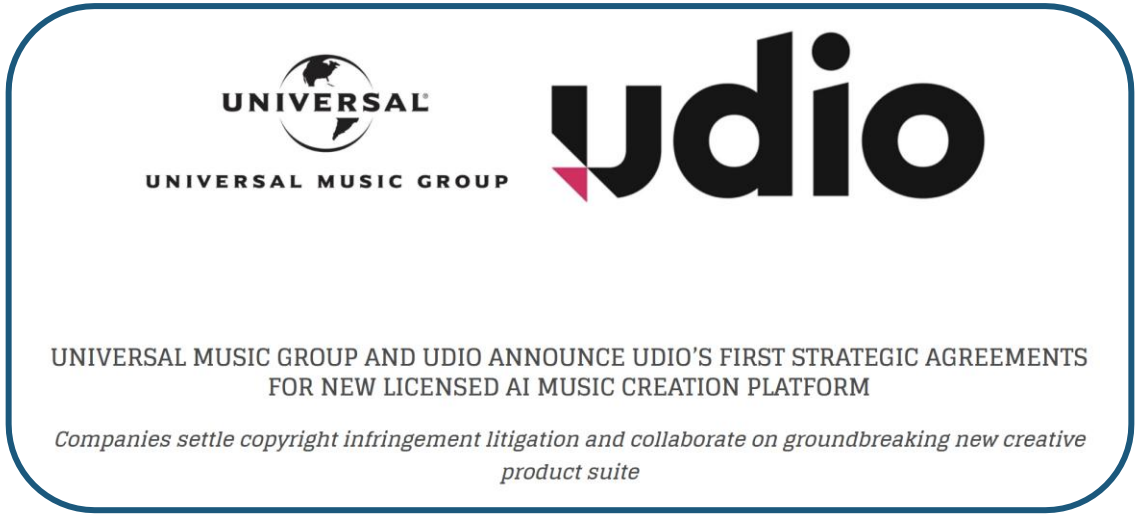
## Music labels sue AI companies Suno, Udio for US copyright infringement


Major record labels Sony Music , Universal Music Group and Warner Records sued artificial intelligence companies Suno and Udio on Monday,...

Jun 24, 2024

apnews.com/article/ai-music-generators-sued-suno-udio-riaa-37a398d326ebb53105538f0d1088233e
reuters.com/technology/artificial-intelligence/music-labels-sue-ai-companies-suno-udio-us-copyright-infringement-2024-06-24/

# UMG-Udio Settlement (October 2025)




"The new platform, which will be launched in 2026, will be powered by new cutting-edge generative AI technology that will be **trained on authorized and licensed music**. The new subscription service will transform the user engagement experience, creating a licensed and protected environment to **customize, stream and share music responsibly**, on the Udio platform."

"Udio's existing product will remain available to users during the transition period with creations controlled within a **walled garden** and the service amended in multiple ways—including **fingerprinting, filtering, and other measures**—before the launch of the updated service."

universalmusic.com/universal-music-group-and-udio-announce-udios-first-strategic-agreements-for-new-licensed-ai-music-creation-platform/

# Reading: How Will the Law Handle Generative AI?

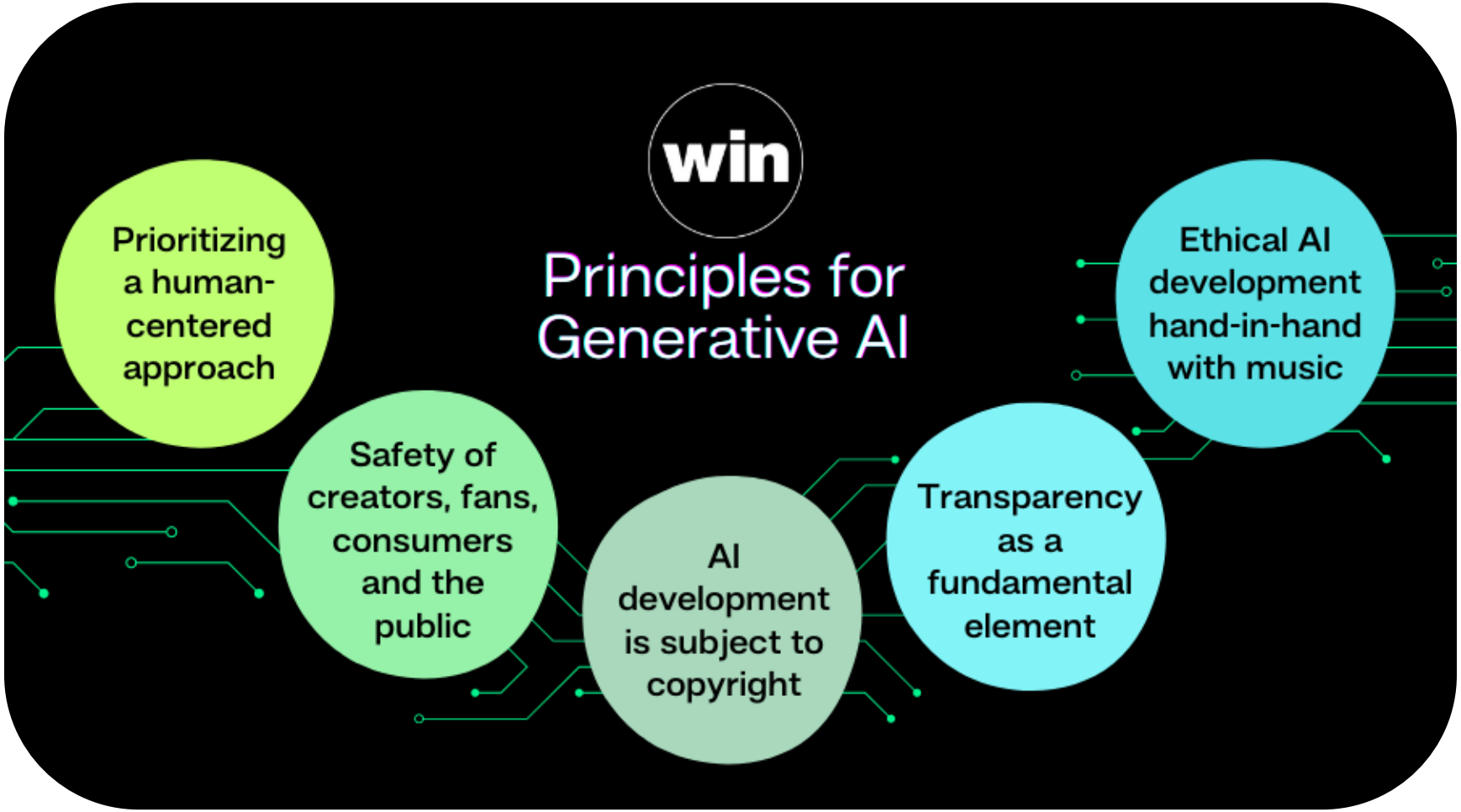

youtu.be/VxJXPWFoYoc

# Ethical Considerations

# Ethics vs Laws

> Ethics is doing **more than the law requires** and **less than the law allows**.
>
> – Michael Josephson

# Principles for Generative AI (Worldwide Independent Network)




winformusic.org/ai-principles/

# Ethical Implications of Music GenAI (Barnett, 2023)

- **Loss of agency and authorship**
- **Creativity stifling**
  - The repetitive nature of the music generation and that by limiting the creative output to possibilities of the model may result in a similar bound on human creativity
- **Predominance of western bias**
- **Copyright infringement**
- **Cultural appropriation**
  - Generative models make it easier to use content from marginalized cultures without any accompanying investment in or engagement from the community

Julia Barnett, "The Ethical Implications of Generative Audio Models: A Systematic Literature Review," *AIES*, 2023.

# Ethical Implications of Speech GenAI (Barnett, 2023)

- **Phishing and fraud**
- **Misinformation and deepfakes**
- **Security and privacy**
  - The potential for risk to security and privacy of individuals as a result of speech generative models, especially when they only require small segments of training data to produce a realistic voice of a targeted speaker
- **Non-consensual use of biometric data**

Julia Barnett, "The Ethical Implications of Generative Audio Models: A Systematic Literature Review," *AIES*, 2023.

# 🔥 **HW 1**: Real or Fake!?

# AI Drake: “Heart on My Sleeve” (Ghostwriter977, 2023)

youtu.be/S2qxvg9NNPM

# This is NOT Morgan Freeman (2021)


youtu.be/oxXpB9pSETo

# This is NOT Morgan Freeman (2021)


youtu.be/F4G6GNFz0O8

# Resources on Ethical Implications of Music AI Research

- Andre Holzapfel, Bob L. Sturm, and Mark Coeckelbergh "Ethical Dimensions of Music Information Retrieval Technology," *TISMIR*, 1(1):44–55, 2018.
- Rujing Huang, Bob L. T. Sturm, and Andre Holzapfel, "De-centering the West: East Asian Philosophies and the Ethics of Applying Artificial Intelligence to Music," *ISMIR*, 2021.
- Rujing Stacy Huang, Andre Holzapfel, Bob L. T. Sturm, and Anna-Kaisa Kaila, "Beyond Diverse Datasets: Responsible MIR, Interdisciplinarity, and the Fractured Worlds of Music," *TISMIR*, 6(1):43–59, 2023.
- Julia Barnett, "The Ethical Implications of Generative Audio Models: A Systematic Literature Review," *AIES*, 2023.

# Resources on Diversity in Music AI Research

- Xavier Serra, Martin Clayton, and Barış Bozkurt, "Computational Approaches for Analysis of Non-Western Music Traditions," *ISMIR Tutorials*, 2018.
- Georgina Born, "Diversifying MIR: Knowledge and Real-World Challenges, and New Interdisciplinary Futures," *TISMIR*, 3(1):193–204 2020.
- Lorenzo Porcaro, Carlos Castillo, and Emilia Gómez "Diversity by Design in Music Recommender Systems," *TISMIR*, 4(1):114–126, 2021.

# Environmental Concerns

# GPU Energy Consumption

V100

**250W**

A100

**400W**

RTX 5090

**575W**

Multiple GPUs

(Source: Supermicro)

**2-6kW**

**~1kW**

**~1.5kW**

**~2kW**

# Training BLOOM (A 176B Parameter LLM) (Luccioni et al., 2022)

| | |
|---|---|
| **Total training time** | 118 days, 5 hours, 41 min |
| **Total number of GPU hours** | 1,082,990 hours |
| **Total energy used** | 433,196 kWh |
| **GPU models used** | Nvidia A100 80GB |
| **Carbon intensity of the energy grid** | 57 $gCO_2eq/kWh$ |

**Average household power usage:**
**10.5 MWh / year**

**A household for 40+ years**
(or **40+ households for a year**)

(Source: Luccioni et al., 2023)

| Model name | Number of parameters | Datacenter PUE | Carbon intensity of grid used | Energy consumption | $CO_2eq$ emissions | $CO_2eq$ emissions × PUE |
|---|---|---|---|---|---|---|
| GPT-3 | 175B | 1.1 | 429 $gCO_2eq/kWh$ | 1,287 MWh | *502 tonnes* | 552 tonnes |
| Gopher | 280B | 1.08 | 330 $gCO_2eq/kWh$ | *1,066 MWh* | *352 tonnes* | 380 tonnes |
| OPT | 175B | *1.09* [2] | *231$gCO_2eq/kWh$* | *324 MWh* | 70 tonnes | *76.3 tonnes* [3] |
| BLOOM | 176B | 1.2 | 57 $gCO_2eq/kWh$ | 433 MWh | 25 tonnes | 30 tonnes |

(Source: Luccioni et al., 2023)

Alexandra Sasha Luccioni, Sylvain Viguier, and Anne-Laure Ligozat, "Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model," *JMLR*, 24(253):1–15, 2023.
eia.gov/energyexplained/use-of-energy/electricity-use-in-homes.php

# The Rapidly-Growing GPU Needs

Tom's Hardware

**First in-depth look at Elon Musk's 100,000 GPU AI cluster — xAI Colossus reveals its secrets**

YouTuber ServeTheHome was granted access to the Supermicro servers within the 100,000 GPU beast, showing off several facets of the supercomputer...

Oct 28, 2024

Tom's Hardware

**Meta is using more than 100,000 Nvidia H100 AI GPUs to train Llama-4 — Mark Zuckerberg says that Llama 4 is being trained on a cluster "bigger than anything that I've seen"**

Meta is using more than 100,000 Nvidia H100 AI GPUs to train Llama-4 — Mark Zuckerberg says that Llama 4 is being trained on a cluster "bigger...

Oct 31, 2024

tomshardware.com/desktops/servers/first-in-depth-look-at-elon-musks-100-000-gpu-ai-cluster-xai-colossus-reveals-its-secrets
tomshardware.com/tech-industry/artificial-intelligence/meta-is-using-more-than-100-000-nvidia-h100-ai-gpus-to-train-llama-4-mark-zuckerberg-says-that-llama-4-is-being-trained-on-a-cluster-bigger-than-anything-that-ive-seen

# The Rapidly-Growing GPU & Energy Needs in Michigan

Michigan Engineering News

## U-Michigan announces most advanced AI research complex with historic Los Alamos alliance

A state-of-the-art AI research facility under development by the University of Michigan and Los Alamos National Laboratory will bring one of the world's most...

Feb 3, 2025

Reuters

## OpenAI, Oracle plan 1 gigawatt Stargate data center in Michigan with Related Digital


OpenAI, Oracle plan 1 gigawatt Stargate data center in Michigan with Related Digital ... Oct 30 (Reuters) - OpenAI, Oracle and Related Digital...

3 weeks ago

Reuters

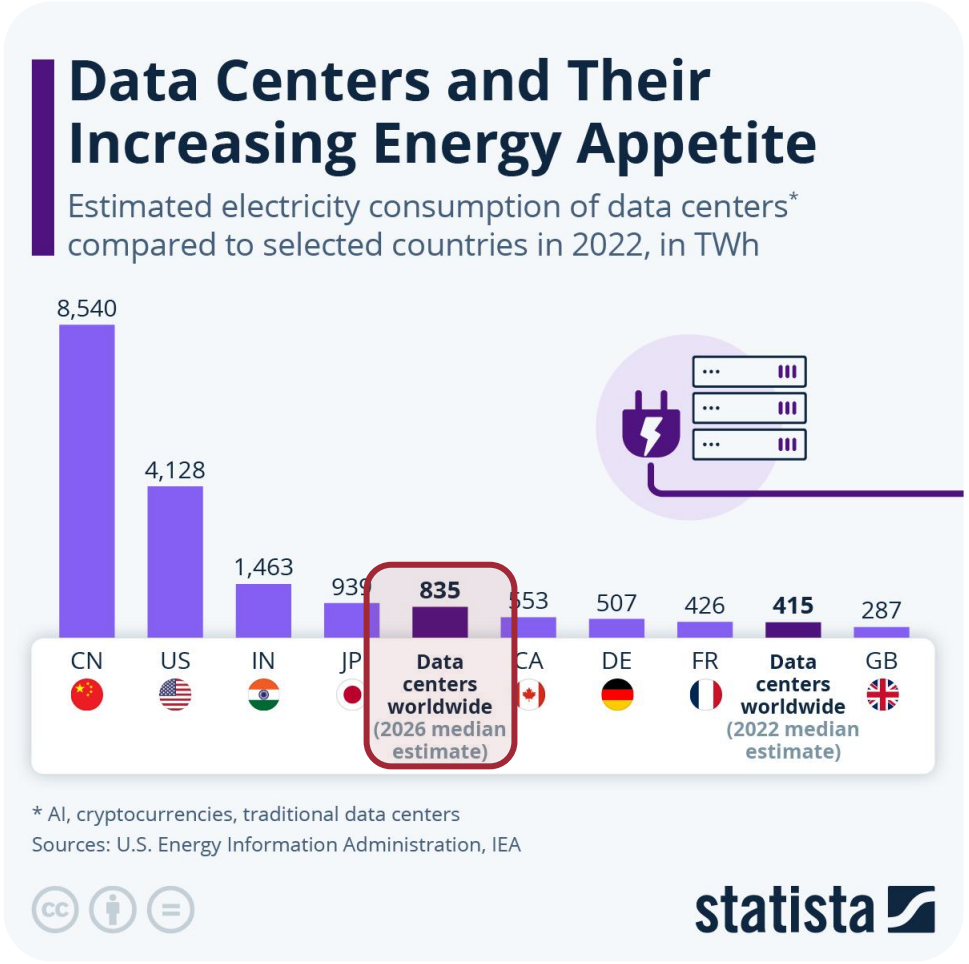
## DTE Energy raises five-year spending plan by 22% on data center power demand

U.S. Midwest electric utility DTE Energy on Thursday raised its five-year capital investment plan by $6.5 billion to $36.5 billion,...

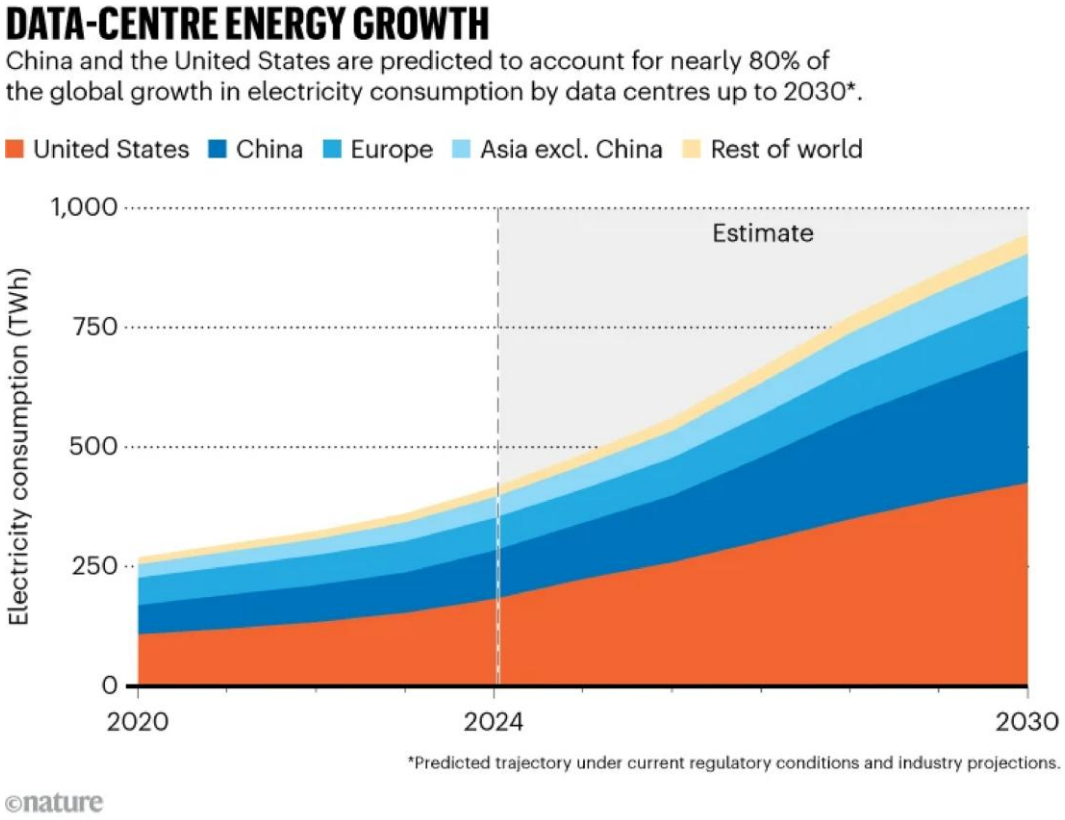
3 weeks ago

news.engin.umich.edu/2025/02/u-michigan-announces-most-advanced-ai-research-complex-with-historic-los-alamos-alliance/
reuters.com/technology/openai-oracle-related-digital-announce-new-stargate-data-center-michigan-2025-10-30/
reuters.com/business/energy/dte-energy-adds-65-billion-five-year-spending-plan-data-center-power-demand-2025-10-30/

# Increasing Energy Consumption of Data Centers




(Source: Statista)

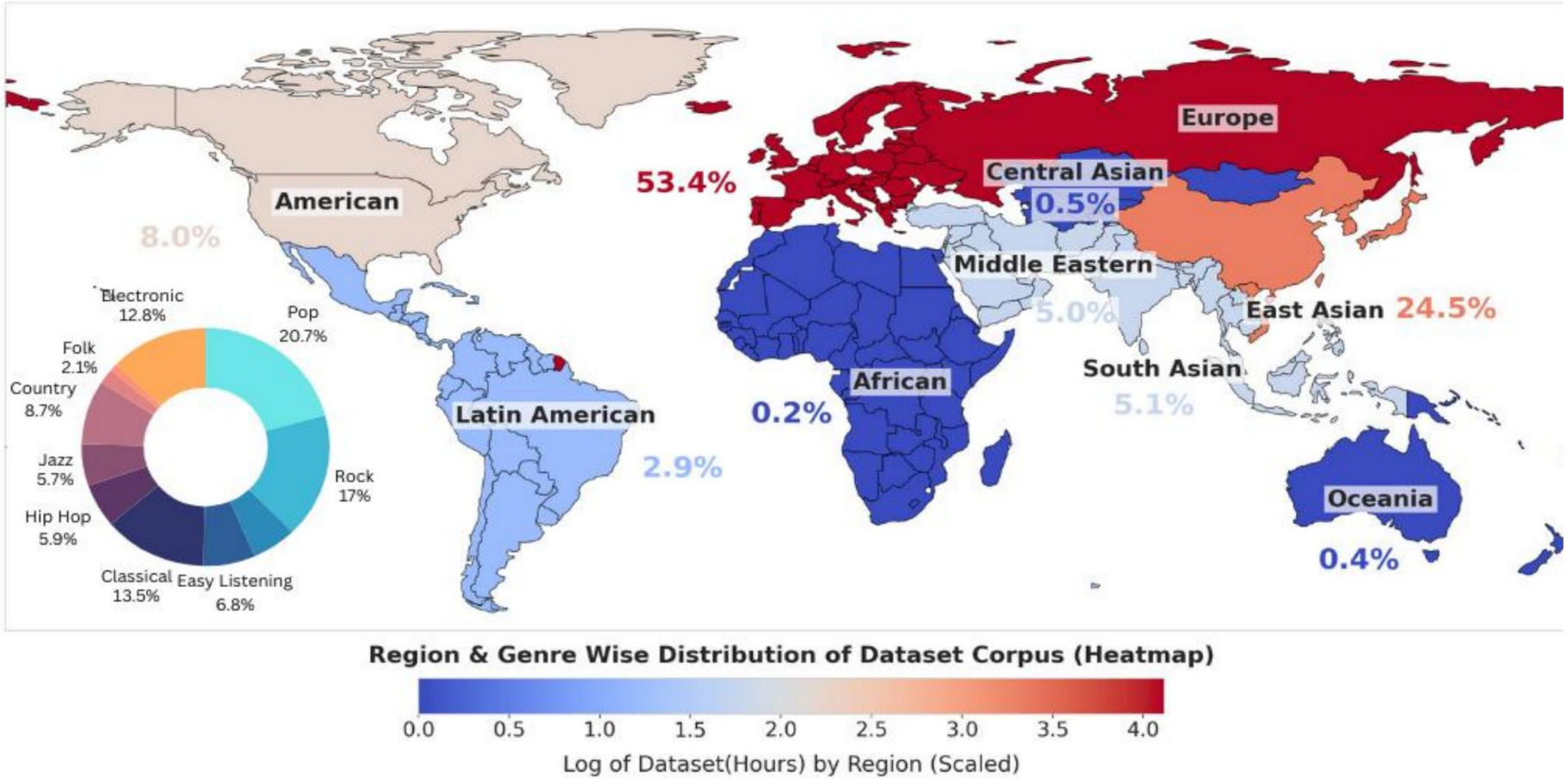



(Source: Nature/Scientific American)

statista.com/chart/32689/estimated-electricity-consumption-of-data-centers-compared-to-selected-countries/
scientificamerican.com/article/ai-will-drive-doubling-of-data-center-energy-demand-by-2030/

# Cultural Concerns

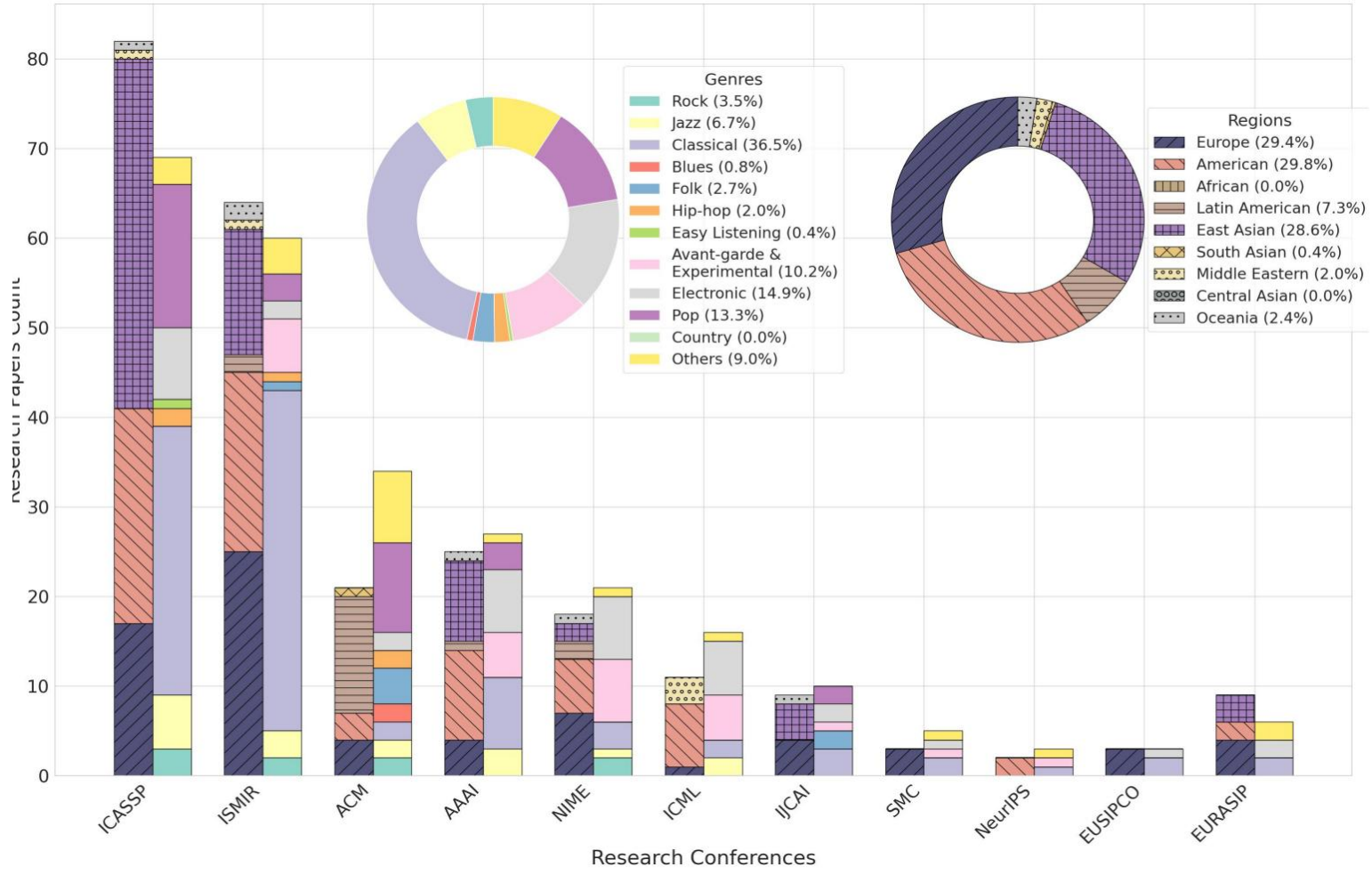
# Global Divide in AI Music Datasets (Mehta et al., 2024)




(Source: Mehta et al., 2024)

Atharva Mehta, Shivam Chauhan, and Monojit Choudhury, "Missing Melodies: AI Music Generation and its "Nearly" Complete Omission of the Global South," *arXiv preprint arXiv:2412.04100*, 2024.

# Global Divide in AI Music Research (Mehta et al., 2024)




(Source: Mehta et al., 2024)

Atharva Mehta, Shivam Chauhan, and Monojit Choudhury, "Missing Melodies: AI Music Generation and its "Nearly" Complete Omission of the Global South," *arXiv preprint arXiv:2412.04100*, 2024.

# Implications

- **Limiting Global South creativity**
  - Limits the potential for Global South music genres to evolve and adapt in the digital age
- **Economic disparities**
  - Worsens the economic disparities within the music industry
- **Reinforcement of existing biases**
  - Reinforces existing cultural biases, perpetuating a cycle where Global South music is viewed as less important or less valuable
- **Cultural erosion**
  - Genres such as Hindustani folk and traditional Arabic Maqam that represent centuries-old traditions, philosophies, and artistic expressions could fade from the public mind

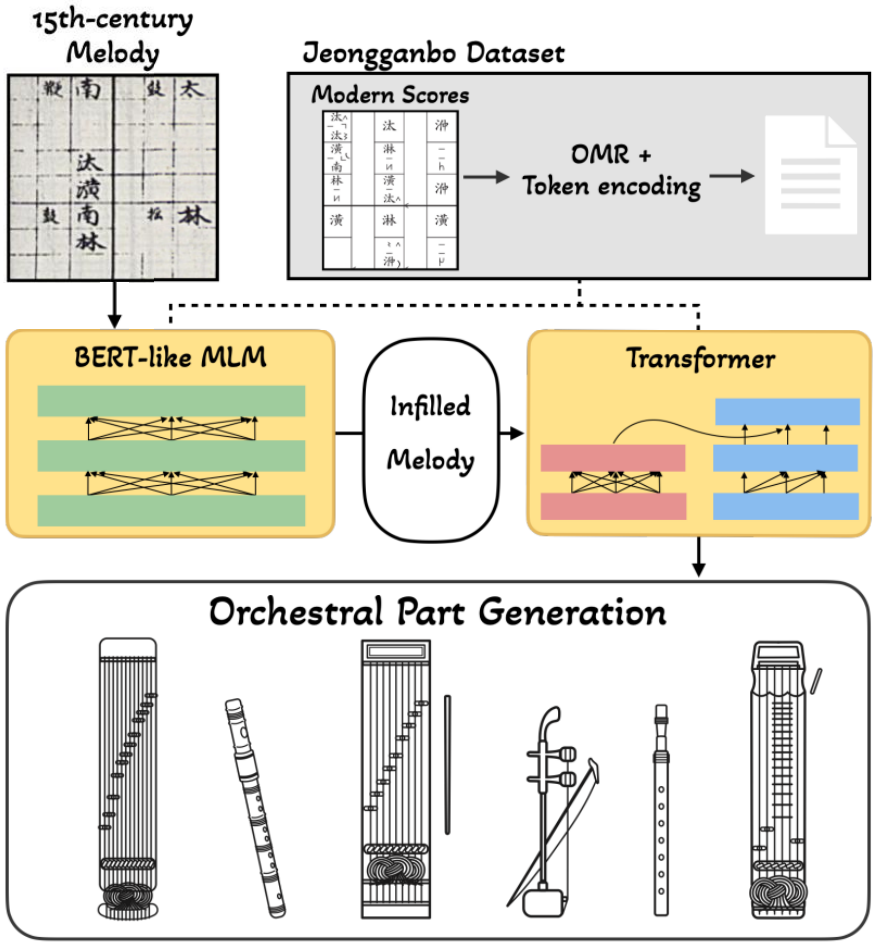
Atharva Mehta, Shivam Chauhan, and Monojit Choudhury, “Missing Melodies: AI Music Generation and its "Nearly" Complete Omission of the Global South,” *arXiv preprint arXiv:2412.04100*, 2024.

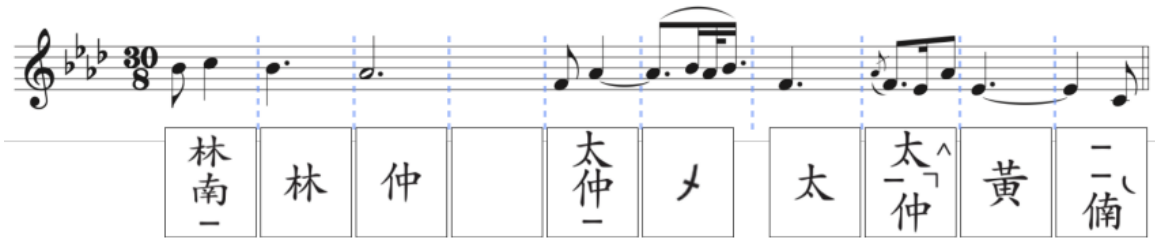
# Recommendations (Mehta et al., 2024)

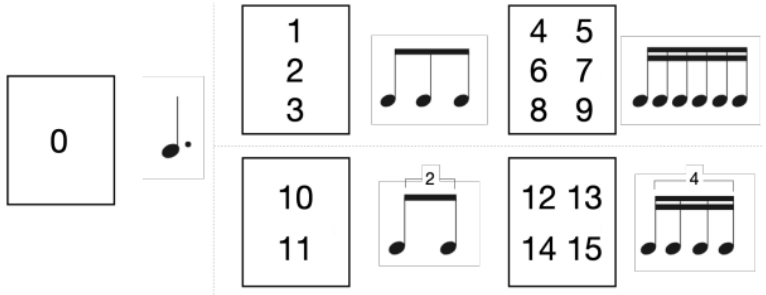
- **Explicit mention** of genres and model limitations
- **Avoid generation** when uncertainty exists
- Investing on **inclusive datasets**
- **Transfer learning** for underrepresented styles
- **Inclusive evaluation**

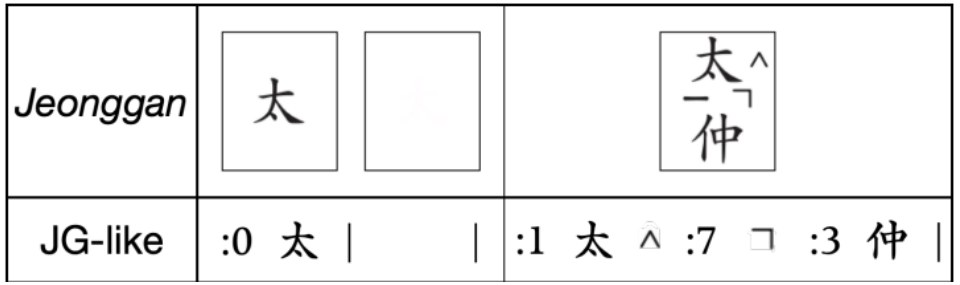
Atharva Mehta, Shivam Chauhan, and Monojit Choudhury, “Missing Melodies: AI Music Generation and its "Nearly" Complete Omission of the Global South,” *arXiv preprint arXiv:2412.04100*, 2024.

# Reviving Korean Court Music with AI (Han et al., 2024)








(Source: Han et al., 2024)

Danbinaerin Han, Mark Gotham, Dongmin Kim, Hannah Park, Sihun Lee, and Dasaem Jeong, "Six Dragons Fly Again: Reviving 15th-Century Korean Court Music with Transformers and Novel Encoding," *ISMIR*, 2024.

# Six Dragons Fly Again (2024)


youtu.be/7zS1FSG7dcg

# Challenges of AI Music

# The Five Challenges

**Representations**

**Usability**

**Creativity**

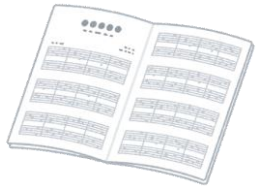
**Multimodality**

**Personalization**

# Challenge 1: Representations

**How can we best represent music for machine learning?**

# Music Generation: Four Paradigms



**Symbolic music generation**

**Text-based**

```
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

**Image-based**

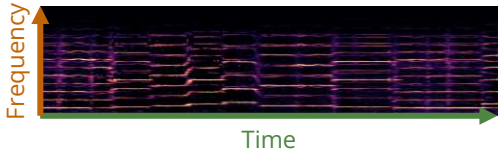

Piano roll

**Audio-domain music generation**

**Time series-based**

Waveform

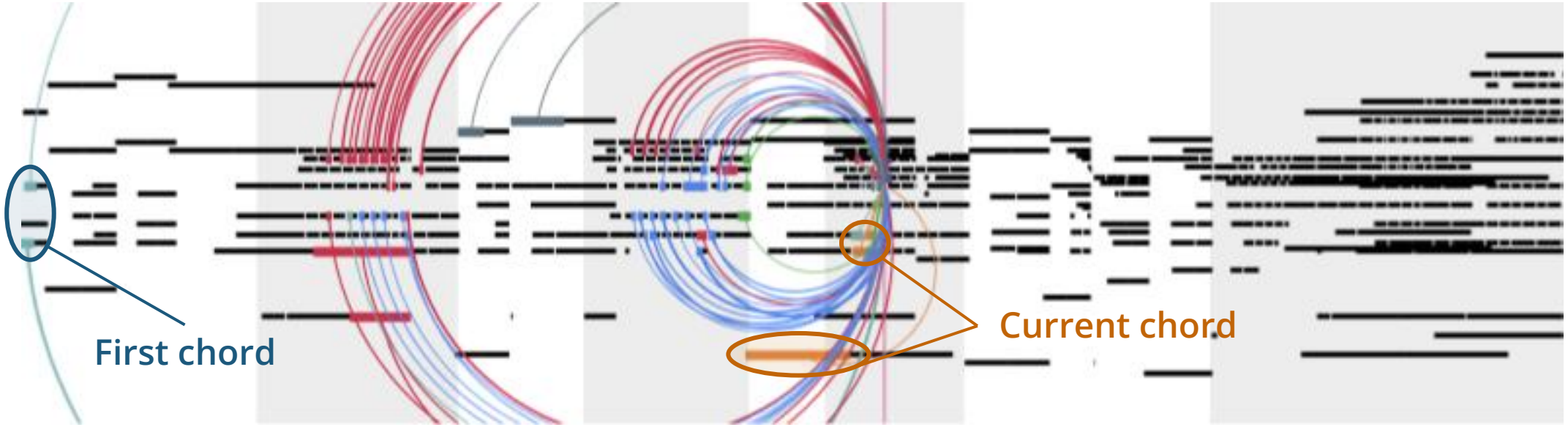
**Image-based**




Spectrogram

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

(Each color represents an attention head)




(Source: Huang et al., 2018)

Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, and Douglas Eck, "Music Transformer: Generating Music with Long-Term Structure," *ICLR*, 2019.
Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, and Douglas Eck, "Music Transformer: Generating Music with Long-Term Structure," *Magenta Blog*, December 13, 2018.

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

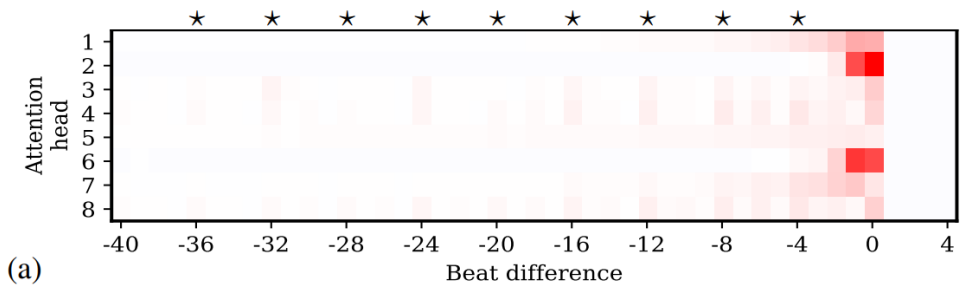
- We proposed two new quantities for measuring **mean relative attention**

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

- The MMT model attends more to notes

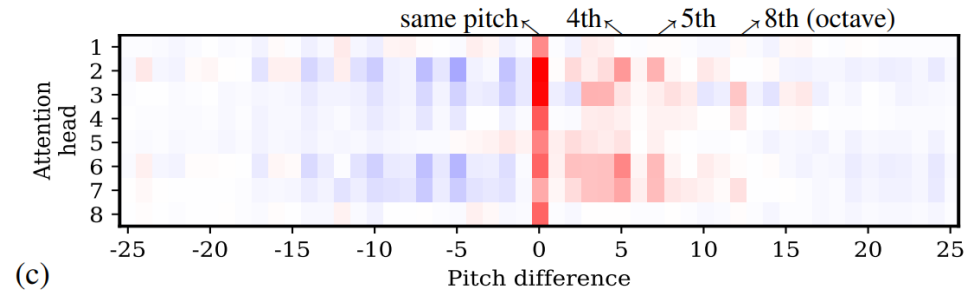
that are $4N$ beats away in the past

Positive and negative mean relative attention gain

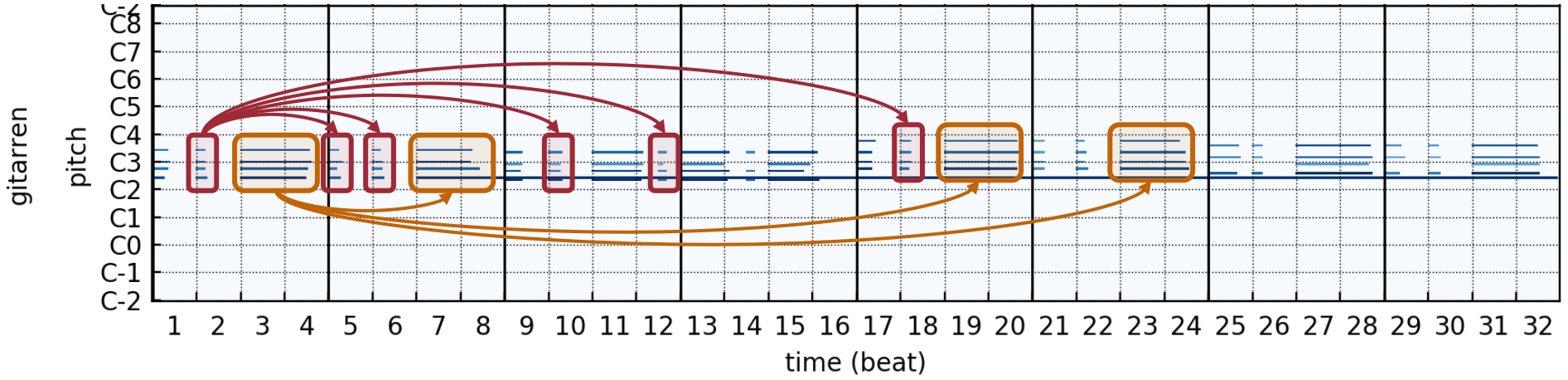



that has a pitch in an octave above which forms a consonant interval

Positive and negative mean relative attention gain



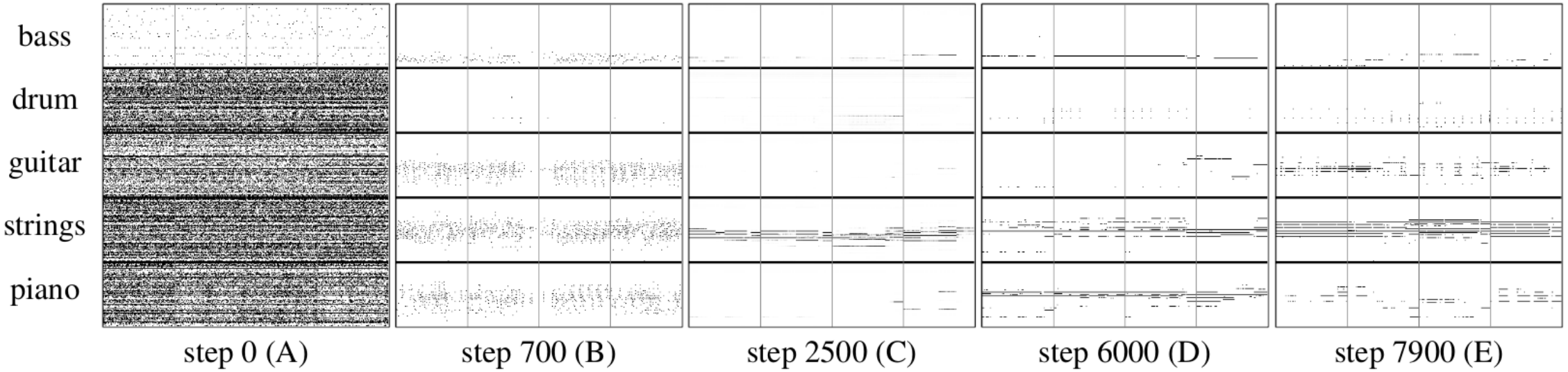
# Why Piano Rolls?




**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)

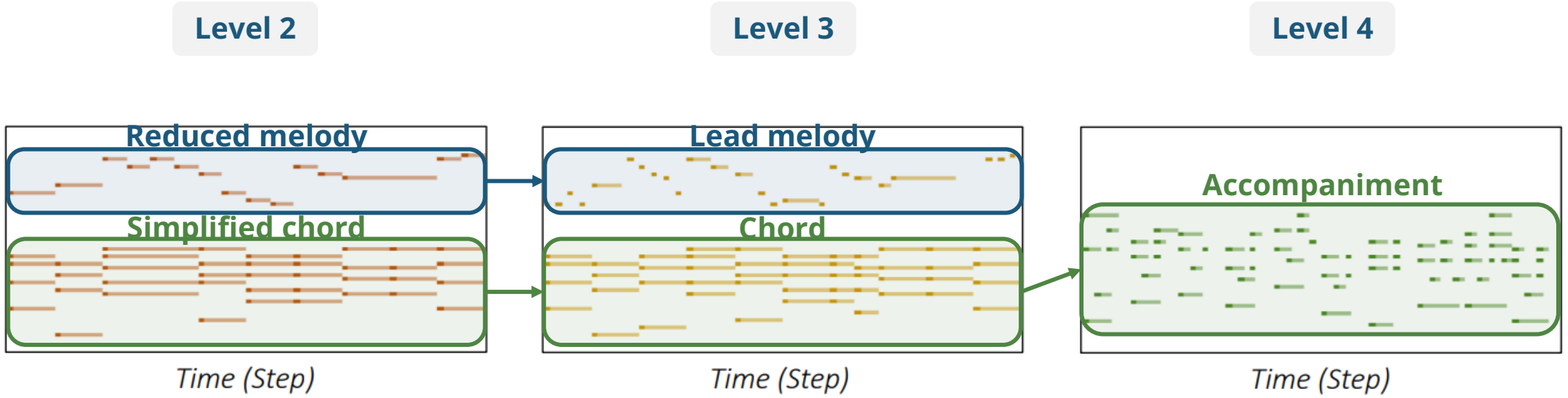
**Examples of generated music**




(Source: Dong et al., 2018)

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.

# Cascaded Diffusion Models (Wang et al., 2024)




(Source: Wang et al., 2024)

wholesonggen.github.io

Ziyu Wang, Lejun Min, and Gus Xia, "Whole-Song Hierarchical Generation of Symbolic Music Using Cascaded Diffusion Models," *ICLR*, 2024.

# Challenge 1: Representations

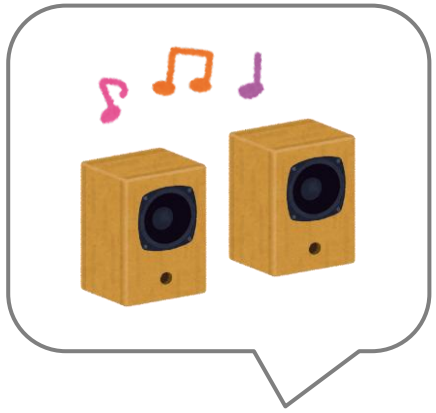
**How can we best represent music for machine learning?**

# Challenge 2: Multimodality

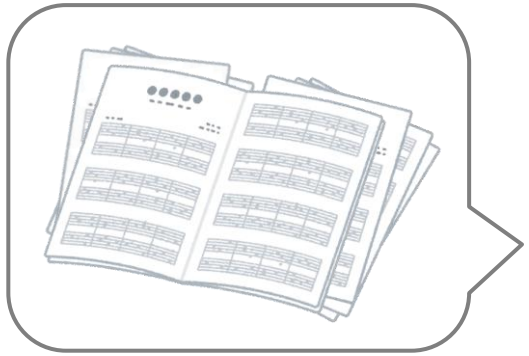
**Can AI learn to create music by “listening to” music rather than “reading” music?**

# Human-inspired Machine Learning for Music & Audio

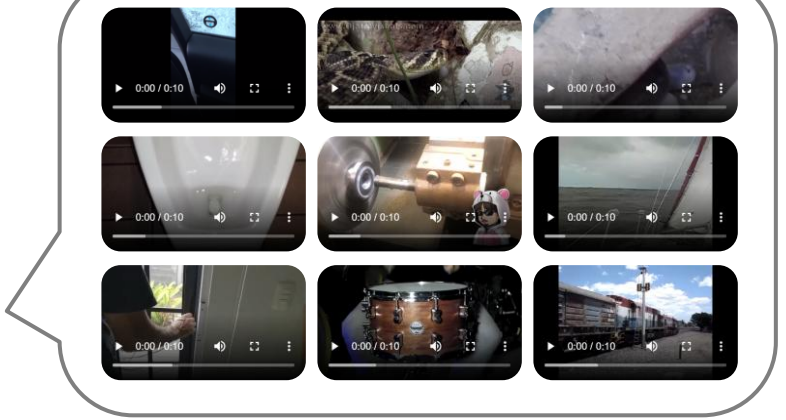
Learning from listening



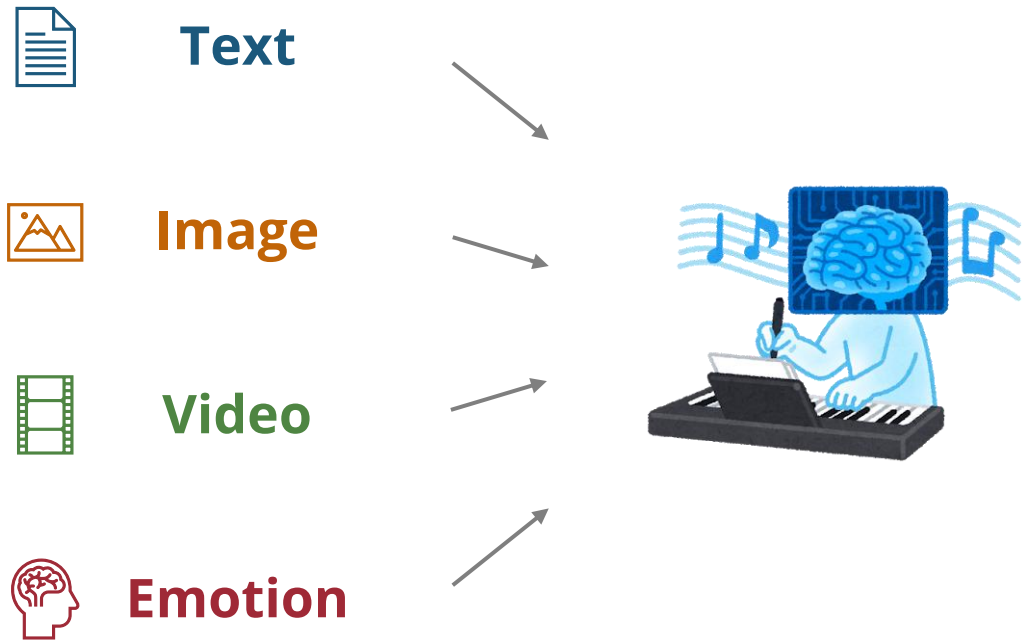
Learning from reading



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 3: Usability

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

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

**Large language models**
(GPT-4)

**Pretrained generative audio models**
(MusicGen, AudioLDM, Bark)

**Instructions** → **Audio Script** → **Audio**

| 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 | Auto |
| Speech | Foreground | N/A | Host | -15 | N/A | Now let's connect with our on-site reporter ... | Auto |
| Sound effect | Foreground | N/A | N/A | -35 | N/A | Transition swoosh. | 1 |
| Sound effect | Background | 2 | N/A | -30 | Begin | Background noise of busy engineering office. | Auto |
| Speech | Foreground | N/A | Reporter | -15 | N/A | We're here at the headquarters of … | Auto |
| Speech | Foreground | N/A | Director | -15 | N/A | Thank you, so it's a fantastic … | Auto |
| Speech | Foreground | N/A | Reporter | -15 | N/A | This is truly an impressive feat … | Auto |

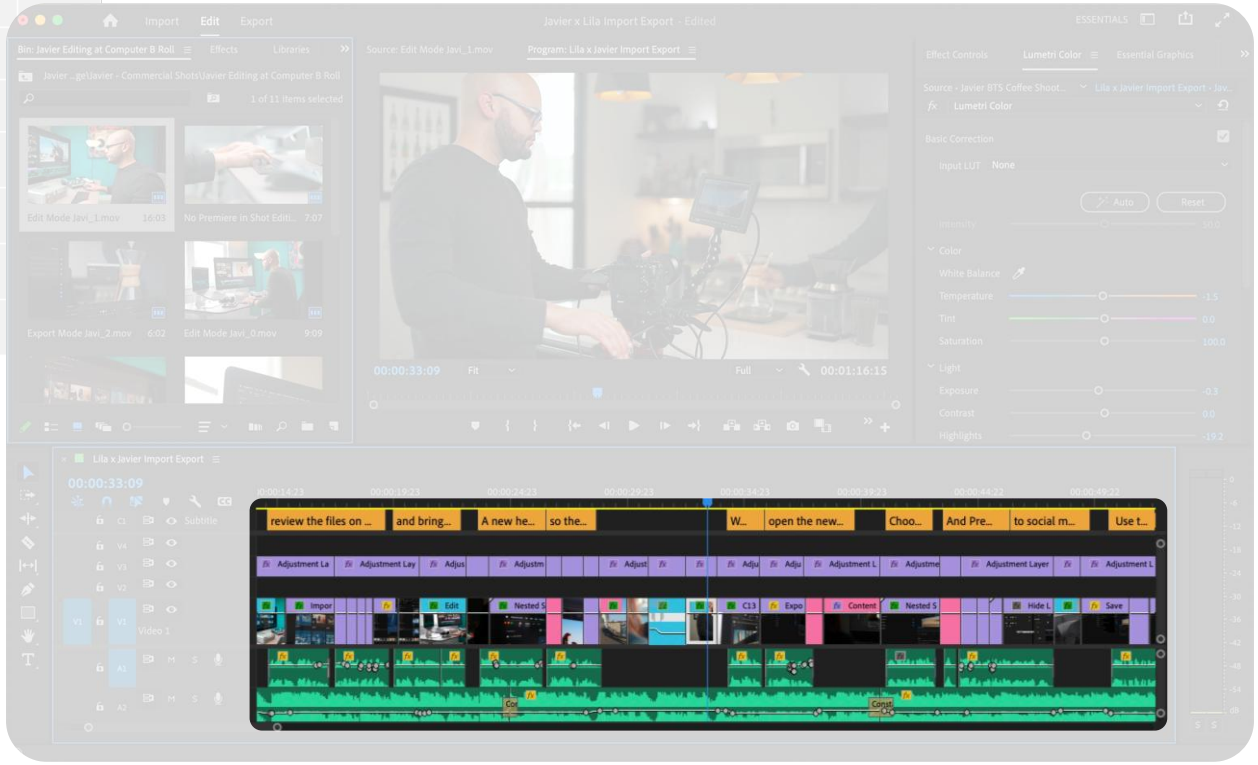
**Interactable intermediate outputs**

(Source: 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, and Wenwu Wang, "WavJourney: Compositional Audio Creation with Large Language Models," *arXiv preprint arXiv:2307.14335*, 2023.

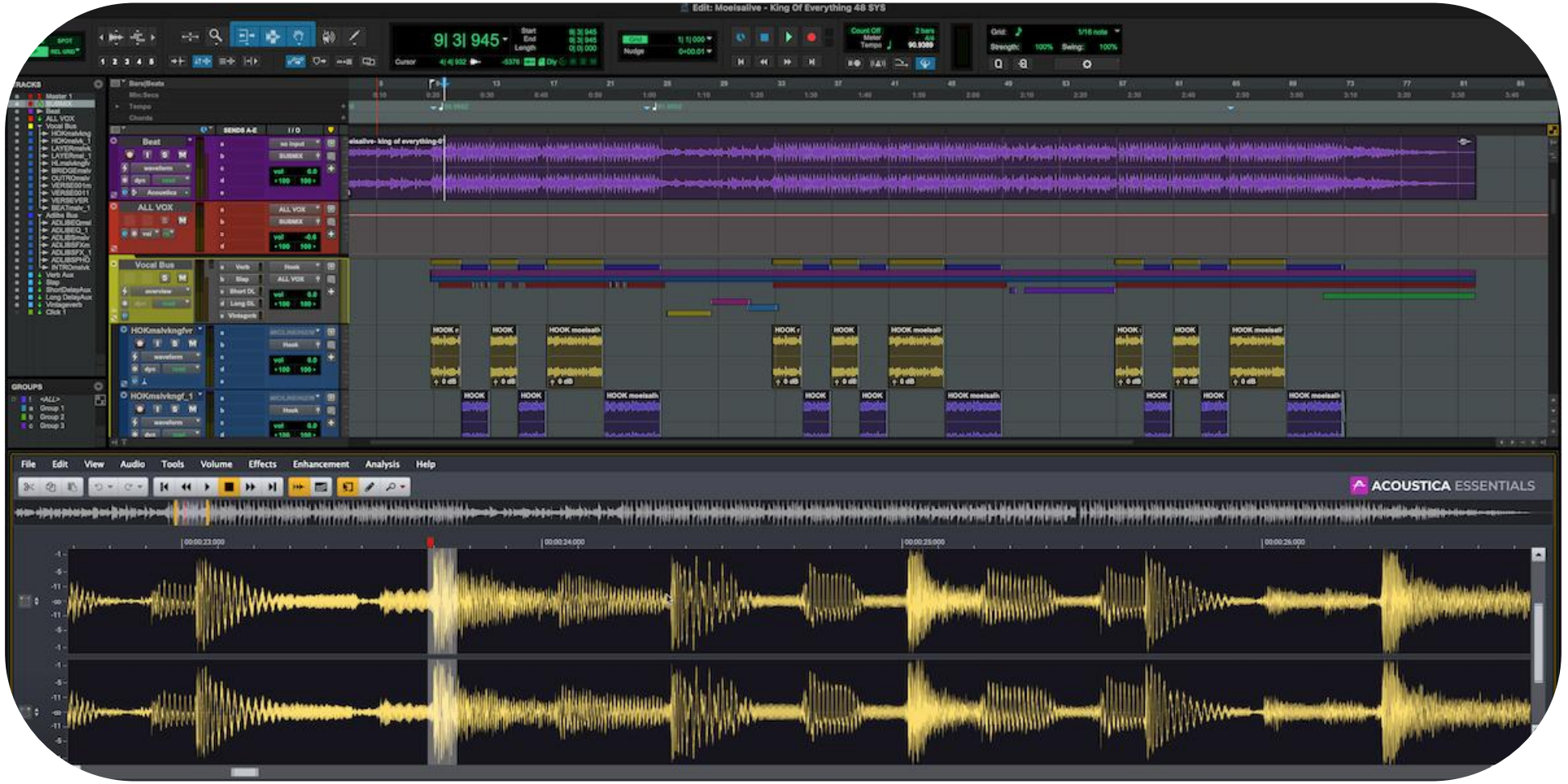
# Integrating GenAI into the Creative Workflow

| 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 … | |



**Integration into professional creative workflow**

# Integrating GenAI into the Music Creative Workflow




(Source: Avid)

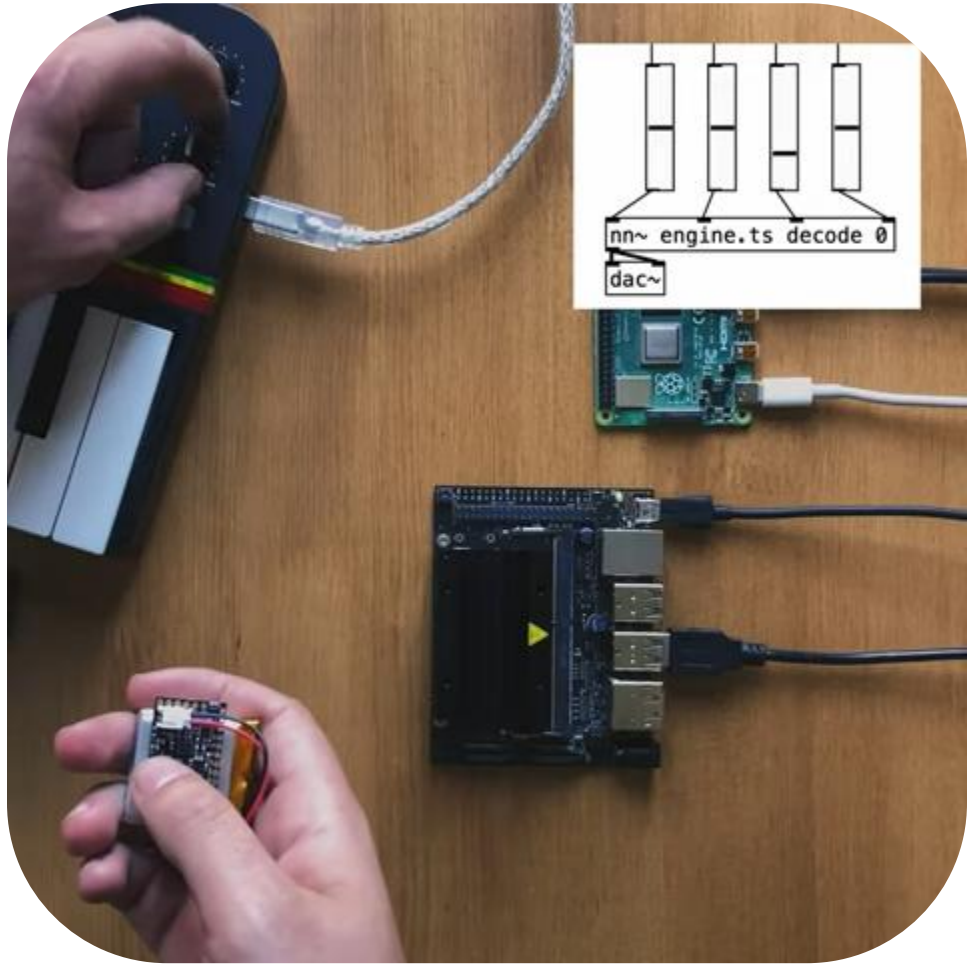
avid.com/pro-tools/whats-new

# Integrating GenAI into the Music Creative Workflow


(Source: Avid)

avid.com/pro-tools

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




youtu.be/jAIRf4nGgYI

Antoine Caillon and Philippe Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," *arXiv preprint arXiv:2111.05011*, 2021.

# 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.

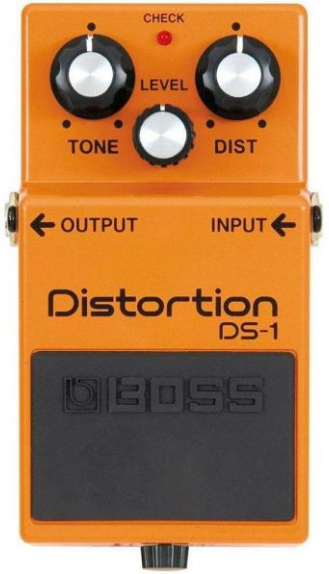
# Ease of Personalization for Artists

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

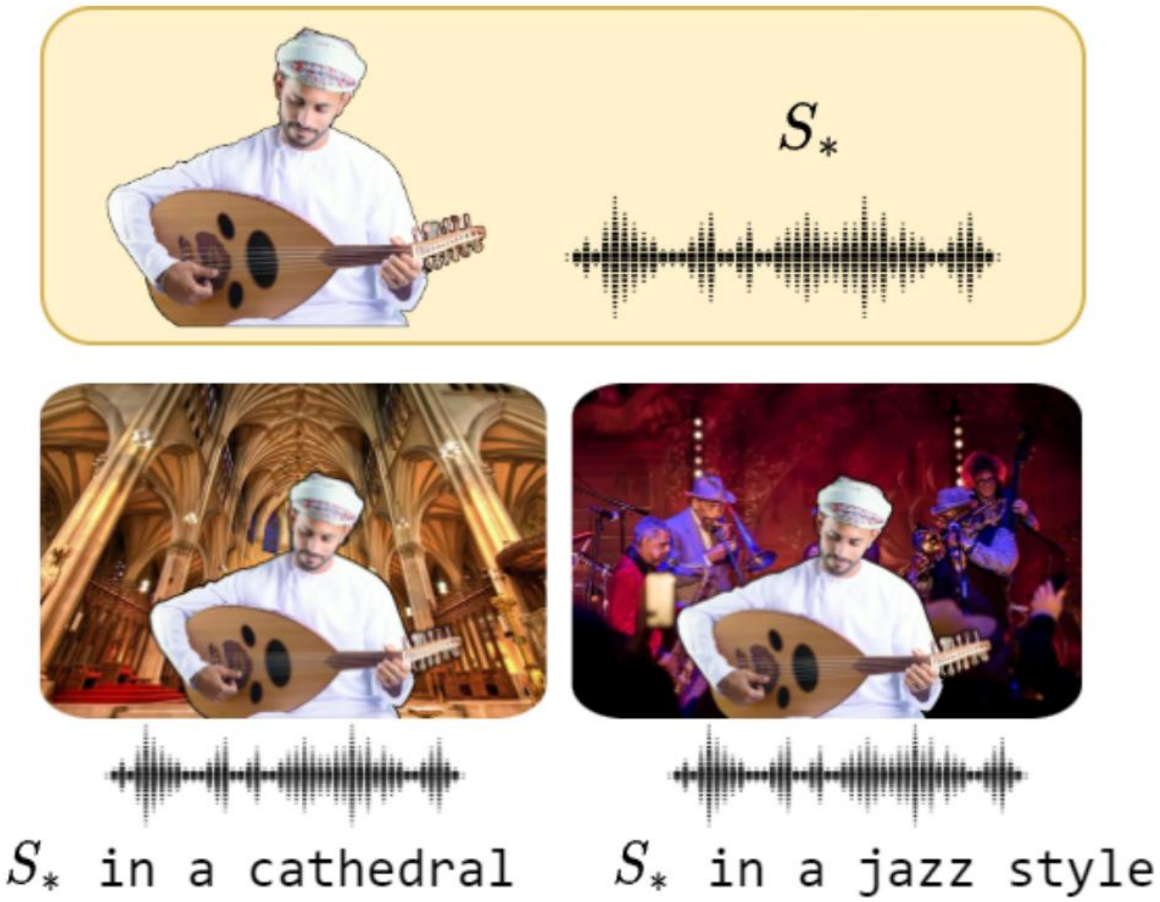
- **Can we do better?**

# Overfitting vs Distortion

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



# Personalized Text-to-Music Generation (Plitsis et al., 2024)




**S***

A disco song with a **S***

**S***

A recording of a **S*** song with a hip hop drum beat accompaniment

(Source: Plitsis et al., 2024)

Manos Plitsis, Theodoros Kouzelis, Georgios Paraskevopoulos, Vassilis Katsouros, and Yannis Panagakis, "Investigating Personalization Methods in Text to Music Generation," *ICASSP*, 2024.

# Challenge 4: Personalization

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

# Challenge 5: Creativity

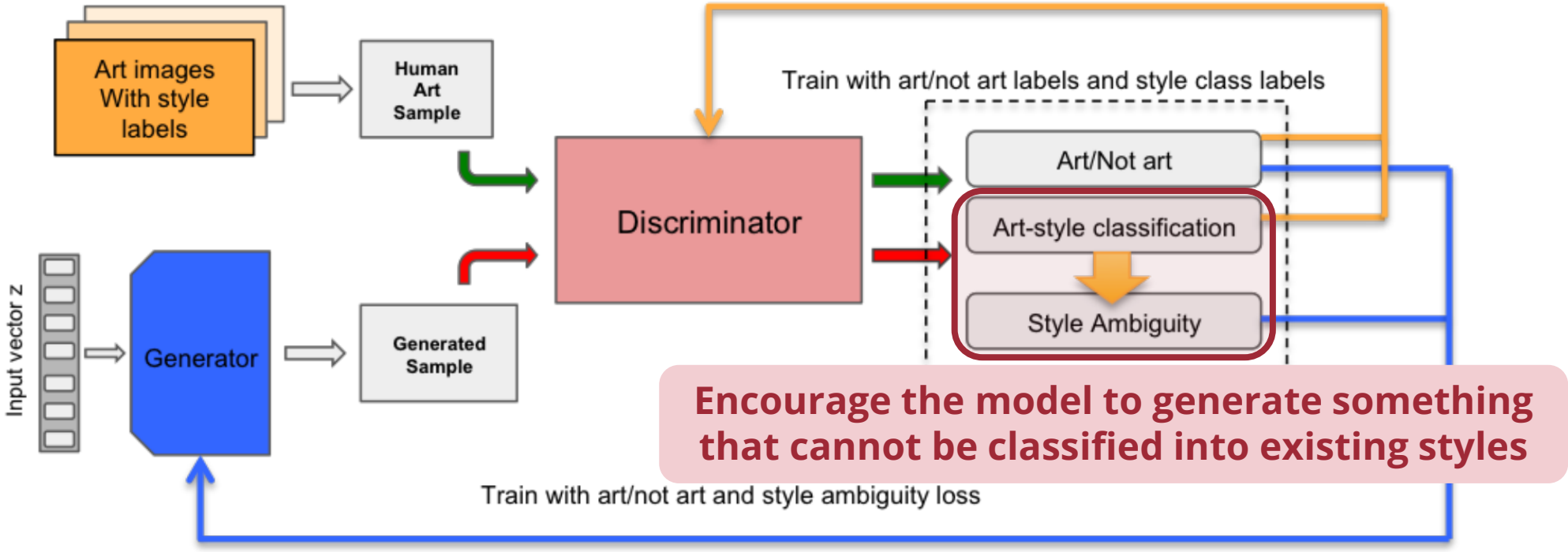
**Can AI ever be creative? How can AI augment human creativity?**

# 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)

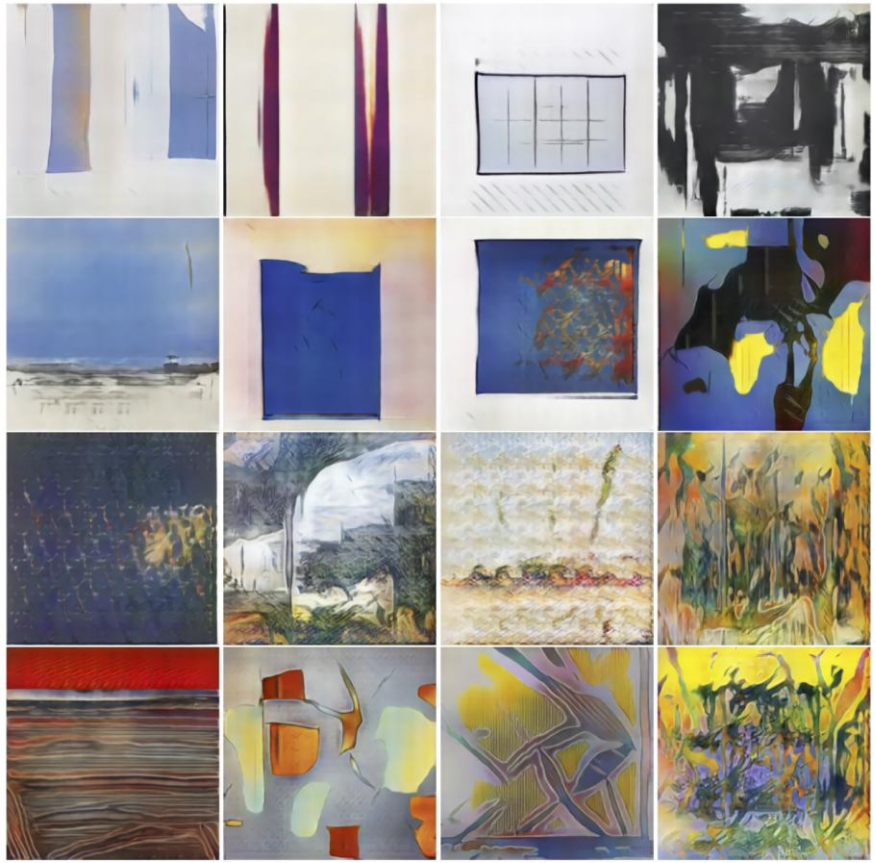



(Source: Elgammal et al., 2017)

Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone, "CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms," *ICCC*, 2017.

# Creative Adversarial Network (Elgammal et al., 2017)

**Example generated images**

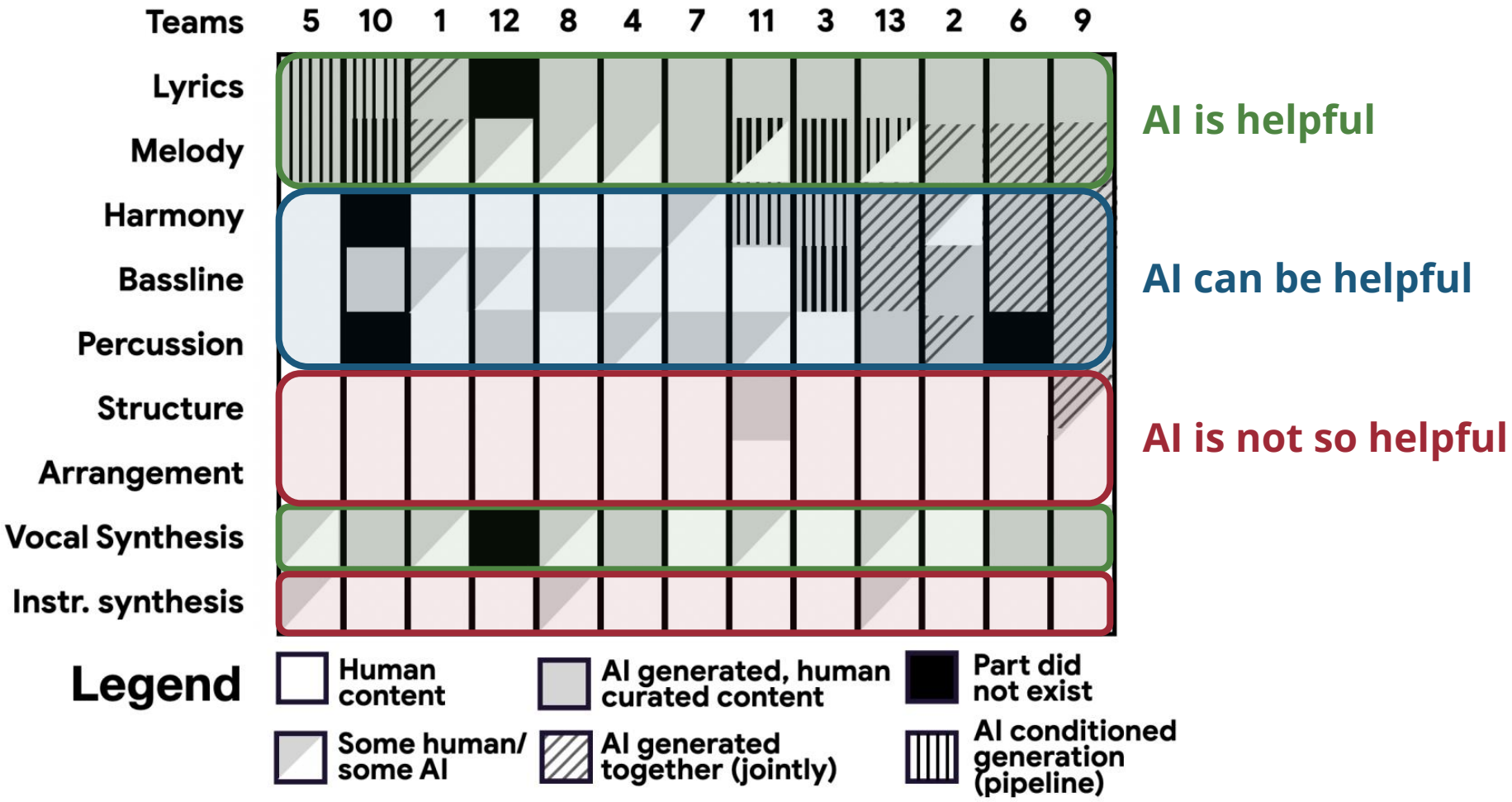


**Best samples**

(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," *ICCC*, 2017.

# How can AI Augment Human Creativity?




(Source: Huang et al., 2020)

Cheng-Zhi Anna Huang, Hendrik Vincent Koops, Ed Newton-Rex, Monica Dinculescu, and Carrie J. Cai, "AI Song Contest: Human-AI Co-Creation in Songwriting," *ISMIR*, 2020.

# What is Creativity?

> Creativity is the ability to come up with ideas or artefacts that are **new**, **surprising** and **valuable**.
>
> – Margaret Boden, 2007

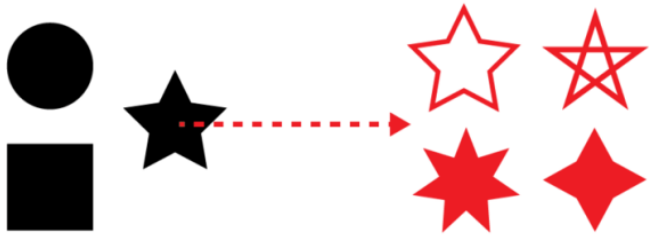
# Three Types of Creativity (Boden, 2007)

## Combinatorial Creativity

Combining existing ideas and things into something new

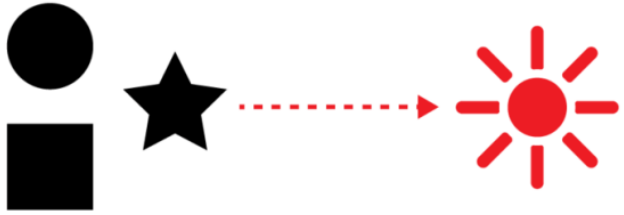
## Exploratory Creativity

Exploring possibilities within a domain



## Transformative Creativity

Radically new ideas that redefine domain and applicable rules



(Source: van Kuijk, 2023)

linkedin.com/pulse/creativity-ungraspable-mystic-proces-thus-can-ai-jasper-van-kuijk/
Margaret A. Boden, "Creativity in a nutshell," *Think*, 5(15):83–96, 2007.

# AI is Good at Combinatory Creativity

**Prompt**: A Michigan space wolverine


Generated by ImageFX

# AI is Good at Combinatory Creativity

**Prompt**: An MIT space beaver

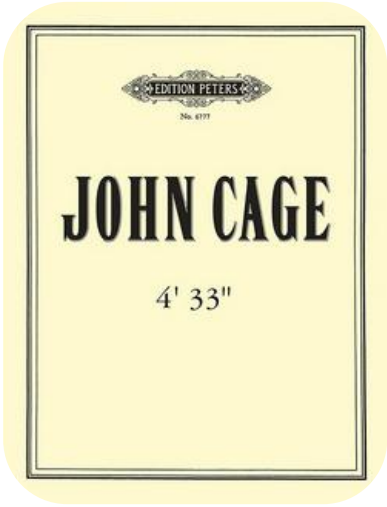

Generated by ImageFX

# Can AI learn Transformative Creativity?

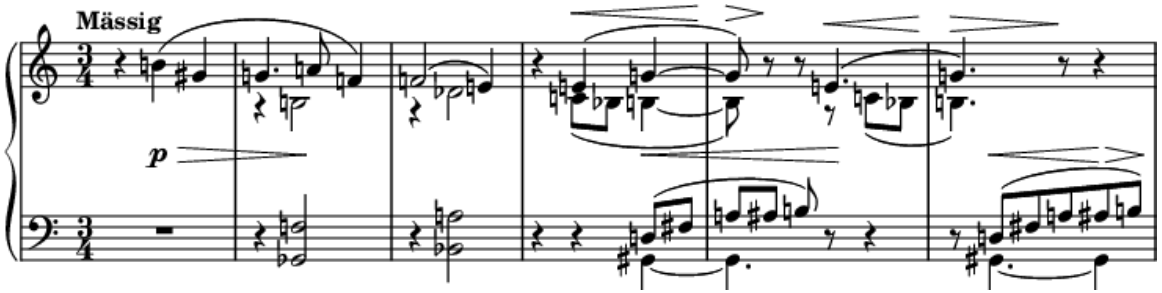
**Duchamp & conceptual art**


**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

# Reading: Can Computer Create Arts?

youtu.be/HPMCWtoC_rM

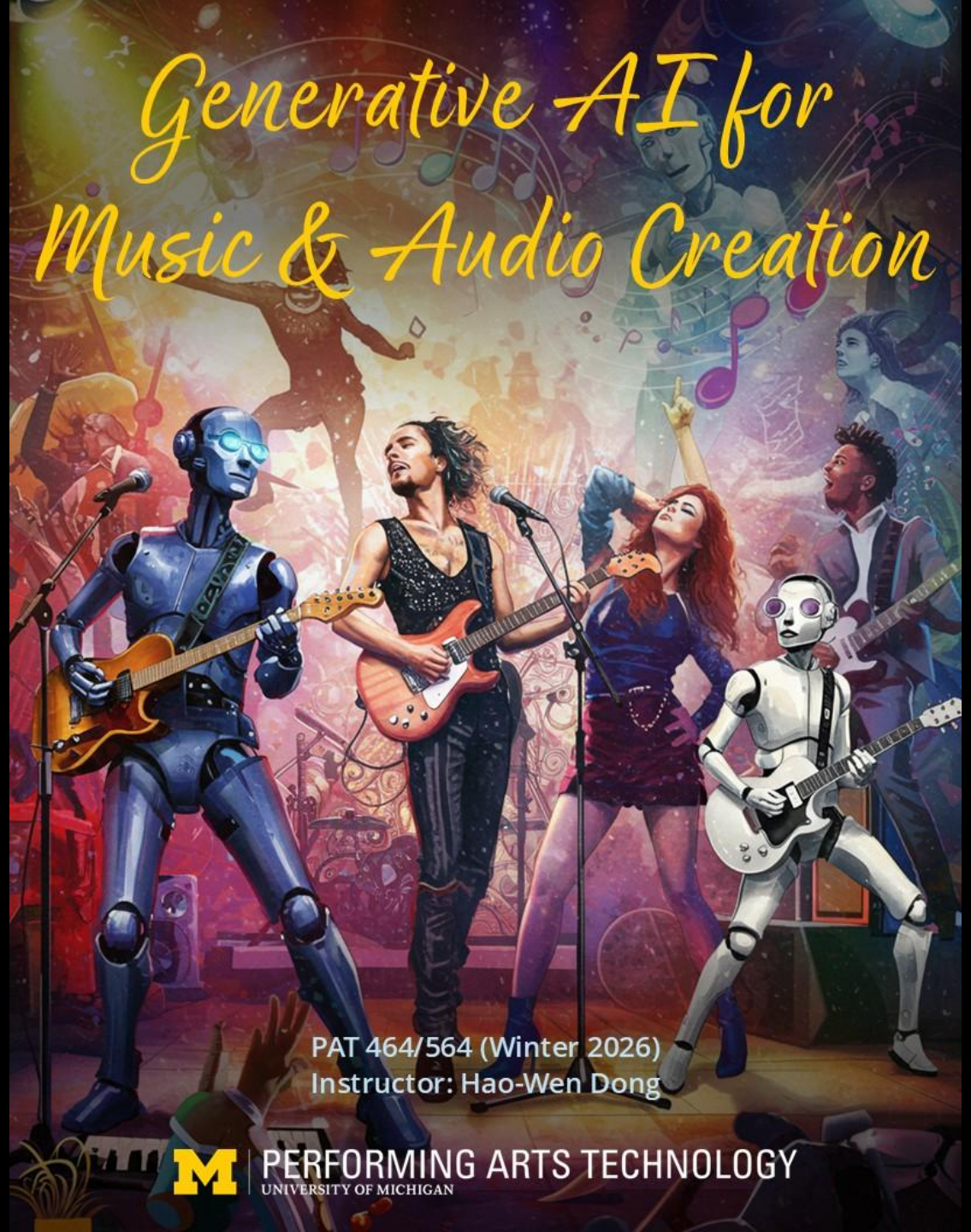
# Challenge 5: Creativity

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

# The Five ~~Challenges~~ Opportunities!


- **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?

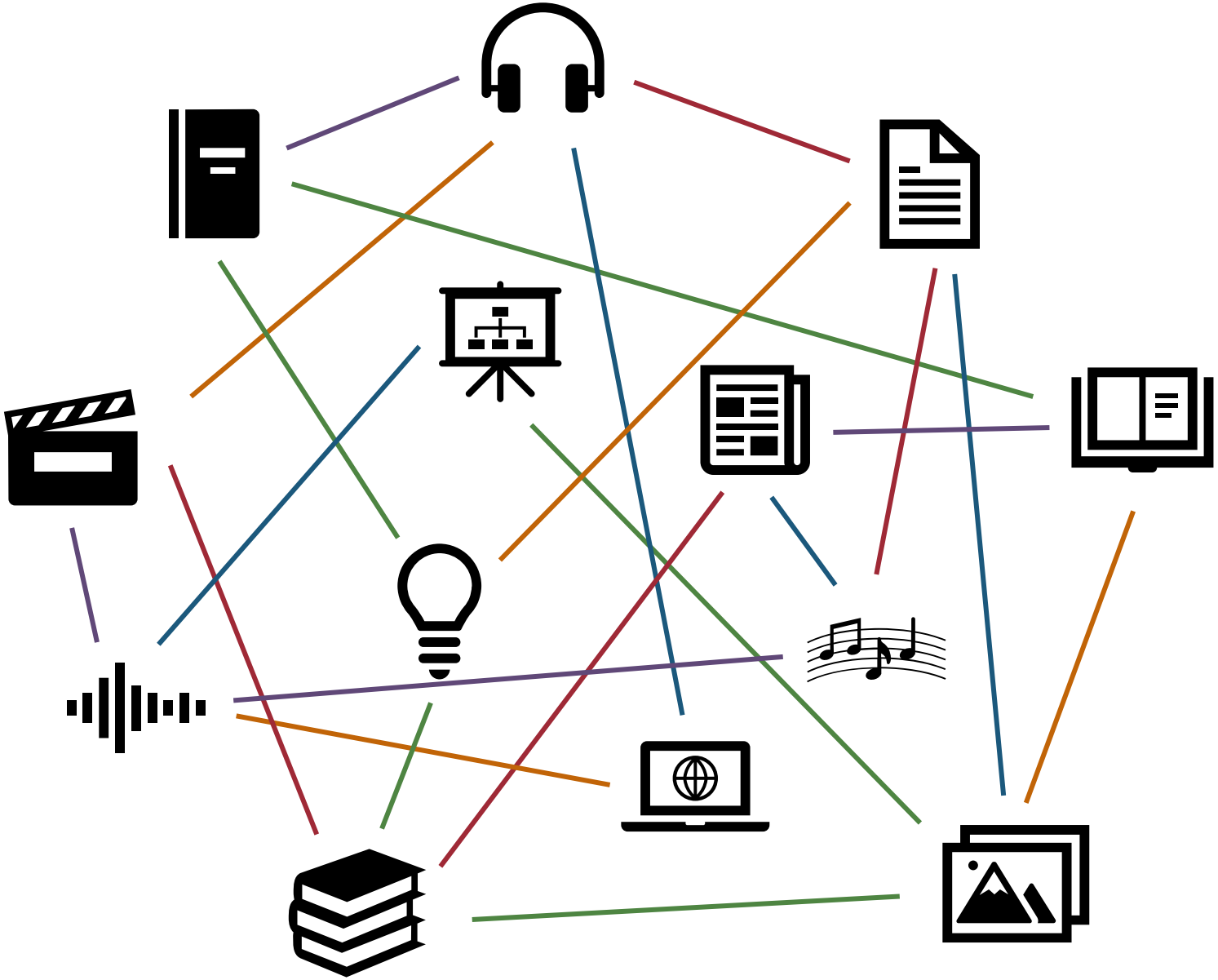
# Final Thoughts




**Winter 2026**
**Generative AI for Music and Audio Creation**
**(PAT 464/564)**

## Dive deeper into GenAI for music & audio

If you want to learn more about all the latest music and audio generation models



**Teachers**

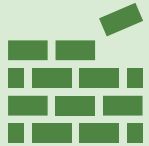
**organize knowledge**

**Researchers**

**create knowledge**

**Engineers**

**apply knowledge**



**Artists**

**challenge knowledge**