PAT 498/598 (Winter 2025)

Music & Al

Lecture 1: Introduction

Instructor: Hao-Wen Dong



Welcome! Tell Us about Yourself!

- Name
- Pronouns
- Program/year
- What is your **most familiar instrument** (if any)?
- What is your **most familiar programming language** (if any)?
- Have you ever used any **AI Music tools**? Which tools?

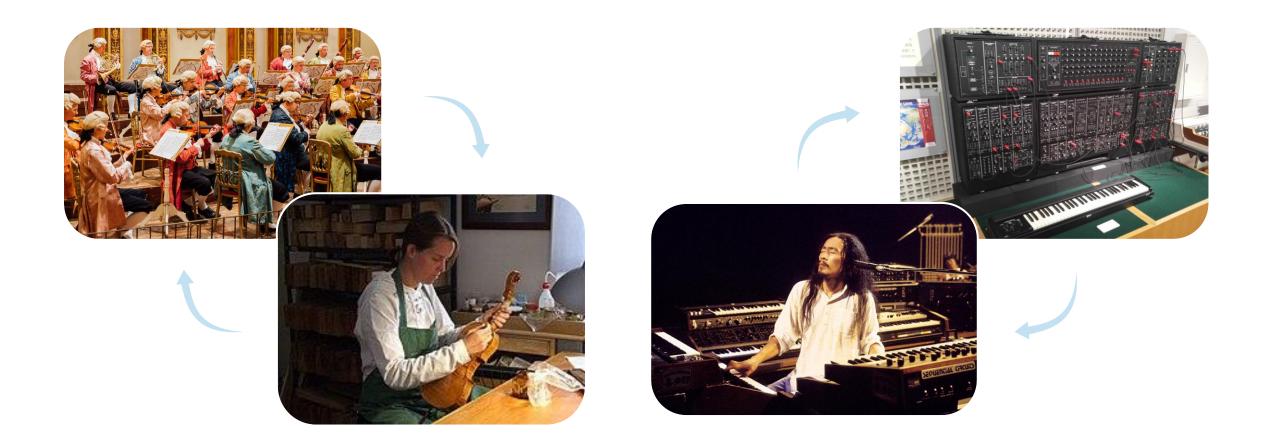
About Me

- Hao-Wen (Herman) Dong
- Pronouns: he/him
- Email: hwdong@umich.edu
- Office: Stearns 131 (15 min walk to the north from Moore)
- <u>Office hours</u>: By appointments
- <u>Research areas</u>: Generative AI for music and audio creation



Music & AI

Music & Technology



(Source: Yamaha)





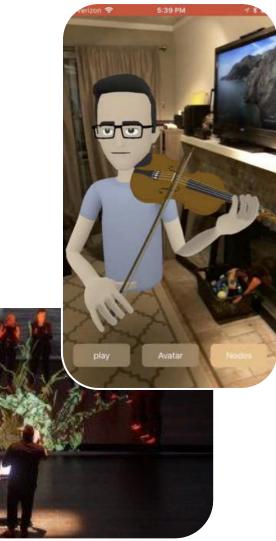
(Source: Robot Gizmos)

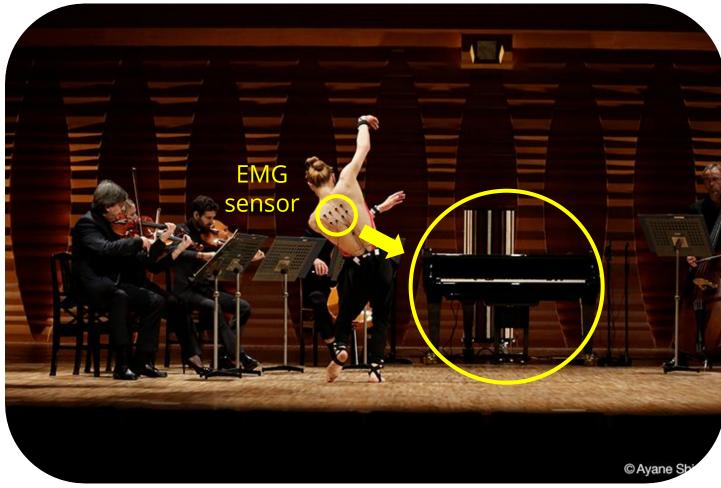
Shlizerman et al., "Audio to Body Dynamics," Proc. CVPR, 2018. yamaha.com/en/news_release/2018/18013101/ sankei.com/article/20240113-CQCOSQHJWFIYPJJKZDCITRTRVI/ roboticgizmos.com/shimon-musical-robot-deep-learning/ nbcdfw.com/entertainment/the-scene/how-verdigris-ensemble-is-using-ai-to-create-a-new-concert-experience/3366031/

(Source: NBC DFW)

(Source: Sankei Shimbun)

(Shlizerman et al., 2019)





(Source: Yamaha)

Yamaha Global, "Yamaha Artificial Intelligence (AI) Transforms a Dancer into a Pianist - Short Version," YouTube, youtu.be/21injmy1wsU, 2018.



(Source: Yamaha)

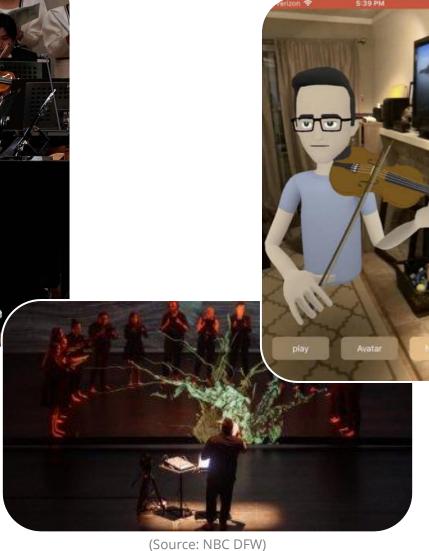




(Source: Robot Gizmos)

Shlizerman et al., "<u>Audio to Body Dynamics</u>," *Proc. CVPR*, 2018. <u>yamaha.com/en/news_release/2018/18013101/</u> <u>sankei.com/article/20240113-CQCOSQHJWFIYPJJKZDCITRTRVI/</u> <u>roboticgizmos.com/shimon-musical-robot-deep-learning/</u> <u>nbcdfw.com/entertainment/the-scene/how-verdigris-ensemble-is-using-ai-to-create-a-new-concert-experience/3366031/</u>

(Source: Sankei Shimbun)



)

(Shlizerman et al., 2019)

11



(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.



(Source: Yamaha)





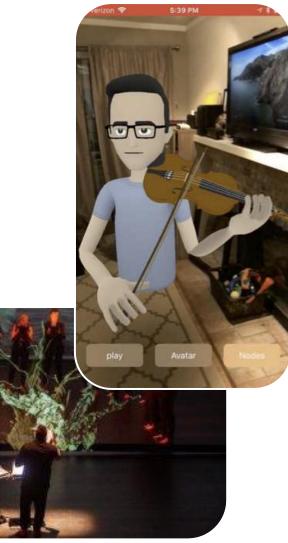
(Source: Robot Gizmos)

Shlizerman et al., "<u>Audio to Body Dynamics</u>," *Proc. CVPR*, 2018. <u>yamaha.com/en/news_release/2018/18013101/</u> <u>sankei.com/article/20240113-CQCOSQHJWFIYPJJKZDCITRTRVI/</u> <u>roboticgizmos.com/shimon-musical-robot-deep-learning/</u> <u>nbcdfw.com/entertainment/the-scene/how-verdigris-ensemble-is-using-ai-to-create-a-new-concert-experience/3366031/</u>

(Source: NBC DFW)

(Source: Sankei Shimbun)

(Shlizerman et al., 2019)





Shlizerman et al., "<u>Audio to Body Dynamics</u>," *Proc. CVPR*, 2018. yamaha.com/en/news_release/2018/18013101/ sankei.com/article/20240113-CQCOSQHJWFIYPJJKZDCITRTRVI/ roboticgizmos.com/shimon-musical-robot-deep-learning/

(Source: Sankei Shimbun)



(Source: Yamaha)





(Source: Robot Gizmos)

Shlizerman et al., "<u>Audio to Body Dynamics</u>," *Proc. CVPR*, 2018. <u>yamaha.com/en/news_release/2018/18013101/</u> <u>sankei.com/article/20240113-CQCOSQHJWFIYPJJKZDCITRTRVI/</u> <u>roboticgizmos.com/shimon-musical-robot-deep-learning/</u> <u>nbcdfw.com/entertainment/the-scene/how-verdigris-ensemble-is-using-ai-to-create-a-new-concert-experience/3366031/</u>

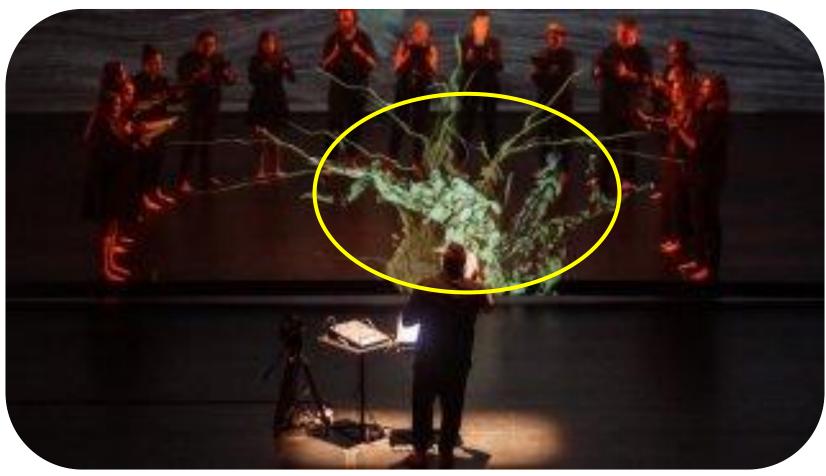
(Source: NBC DFW)

(Source: Sankei Shimbun)

(Shlizerman et al., 2019)



17



(Source: NBC DFW)

(Source: Yamaha)





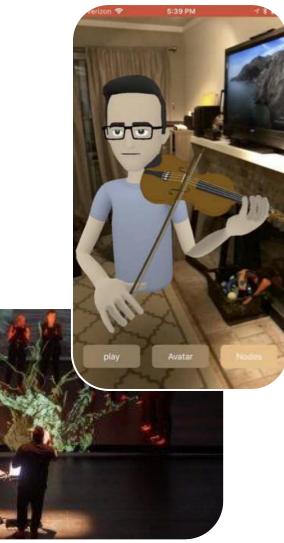
(Source: Sankei Shimbun)

(Source: Robot Gizmos)

Shlizerman et al., "<u>Audio to Body Dynamics</u>," *Proc. CVPR*, 2018. <u>yamaha.com/en/news_release/2018/18013101/</u> <u>sankei.com/article/20240113-CQCOSQHJWFIYPJJKZDCITRTRVI/</u> <u>roboticgizmos.com/shimon-musical-robot-deep-learning/</u> <u>nbcdfw.com/entertainment/the-scene/how-verdigris-ensemble-is-using-ai-to-create-a-new-concert-experience/3366031/</u>

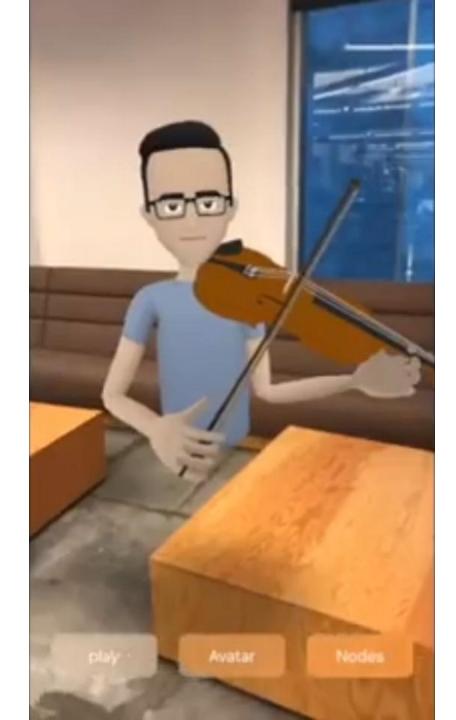
(Source: NBC DFW)

(Shlizerman et al., 2019)





(Shlizerman et al., 2019)



Course Logistics

Communications

- **Course website**: Syllabus, schedule, readings, recordings, etc.
- Email: Announcements
- Google Chat: Q&A



hermandong.com/teaching/ pat498_598_winter2025

Topics

Analysis

 Genre/mood classification, source separation, transcription, optical music recognition, beat-tracking, structural analysis ...

Creation

• Automatic music composition, arrangement, performance synthesis, improvisation ...

Retrieval & Processing

• Search, recommendation, enhancement, production, editing, mixing ...

Advanced Topics

Interactive tools, singing voice synthesis, audiovisual tools ...

Prerequisites

• **Prior coding experience** is recommended!

- If you've taken any programming course, you should be fine
- You should be comfortable reading code written by others

Assignments

- All assignments must be completed on your own
 - You are welcome to exchange ideas with your peers, but this should be in the form of concepts and discussion, not in the form of writing and code.
- Please turn in your work even if it's incomplete!
- Due at 11:59pm ET on the date specified
- Late submissions: 1 point deducted per day

Paper Presentation

- Read and present a research paper
- Group size: 1–2
- A pool of papers will be provided
 - A paper relevant to your final project would be ideal
 - You may propose a paper that is not in the pool
- We will discuss how to read and present a research paper in class
- Milestones (tentative)
 - Paper decision: Feb 10
 - Paper presentation: Feb 24 & 26

Project

- **Open-ended group project** (group size: 1–2)
 - Building a new AI music tool or Exploring creative & artistic use of AI tools
- Milestones (tentative)
 - Pitch: Mar 24
 - Proposal: Mar 31
 - Presentation: Apr 21
 - Final report: Apr 28
- Due at **11:59pm ET** on the date specified
- No late submissions! Submit your work early and update it later.

Grading

- Homework (40%)
- Paper-reading presentation (15%)
- **Project** (45%)
 - Presentation 15%
 - Results 15%
 - Final report 15%
- All grading and regrade requests will be handled on Gradescope

Optional Reading

- Course slides for "<u>Deep Learning for Music Analysis and Generation</u>" by Yi-Hsuan Yang
- "Intelligent Music Production" by Brecht De Man, Ryan Stables, and Joshua D. Reiss
- "Fundamentals of Music Processing" by Meinard Müller (notebooks)

Policies: Attendance

- In-person attendance is strongly encouraged
- Zoom recordings will be available
- Please attend in-person
 - For **project pitch** (March 24)
 - For **project presentation** (Apr 21)

Policies: Generative Al Usage

- Feel free to use GenAl tools (U-M GPT, ChatGPT, Stable Diffusion, DALLE, etc.) in your workflow. However, you must disclose your usage of GenAl services in your write-ups.
- You take full responsibility for AI-generated materials as if you had produced them yourself: ideas should be attributed and facts should be true.

Policies: Academic Integrity

- Plagiarism and cheating violate SMTD's Academic Code of Conduct. All plagiarism, cheating and other academic misconduct cases will be reported to SMTD's Office of Academic and Student Affairs.
- All assignments must be completed on your own. You are welcome to exchange ideas with your peers, but this should be in the form of concepts and discussion, not in the form of writing and code.
- You must provide proper citations/references for any external resources you use in your writing and code.

Any Questions on the Syllabus?





Not so long ago in June 2023

Prompt: relaxing and smooth jazz played in a stylish cafe





Prompt: delightful country music with acoustic guitars





Prompt: cinematic and suspenseful orchestral music





huggingface.co/spaces/facebook/MusicGen



Not so long ago in July 2023



Video Runway Gen-2 Music MusicGen



The State of the Art

Until when...

Suno Al

suno.com



Music FX (2024)

saxophone - solo. T	a smooth - he sound is both	Ģ	TRACK 1/2	\rightarrow	?
sophisticated - and slow tempo	playful - with a	SMOOTH THE SOU SOPHIST	PIECE WITH A SAXOPHONE ND IS BOTH ICATED AND P LOW TEMPO.		A JAZ SMOO THE S SOPH WITH
F	♦ Start over	0.	0:02 / 0:30		• -
C More chill medium tem bass her		¦† Settings	🛃 Download	Copy share link	

aitestkitchen.withgoogle.com/tools/music-fx

Music FX DJ (2024)

MusicFX DJ 🔹					? ;
	chill ∢				
	swirling phasers	i			
	polka				
	Add a prompt		间		
	C more doo wop		ebm salsa snappy snare	mbira	
		funk dru	mmer		
· · ···	•••	Θ	+ :: ;;;		
Density	1	Brightness	Chaos	Drums Bass	Other
			0-0- - - - 0-0 0-0- - - - - 0-0	BPM KEY	~

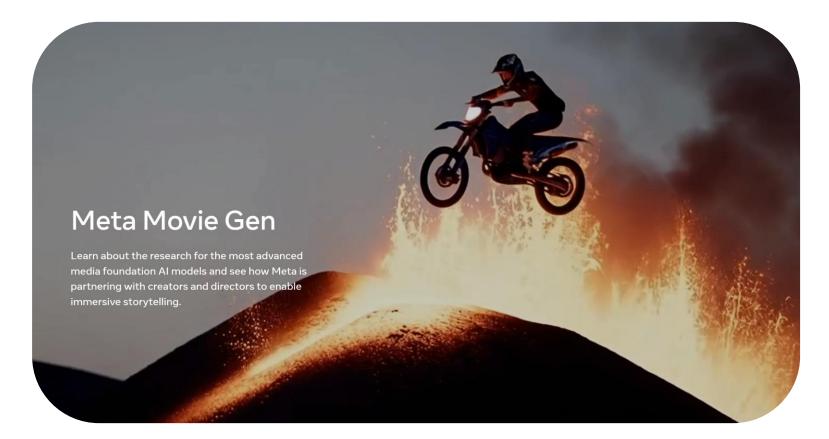
aitestkitchen.withgoogle.com/tools/music-fx-dj





youtube.com/live/IUQW5LgBZvQ

MovieGen



ai.meta.com/research/ movie-gen/

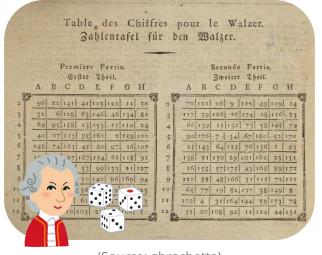


Source: YouTube)

Al Music in the Early Days

The Early Days

Musical Dice Game (1792)



(Source: gbrachetta)

gbrachetta.github.io/Musical-Dice/

ILLIAC Suite (1957)



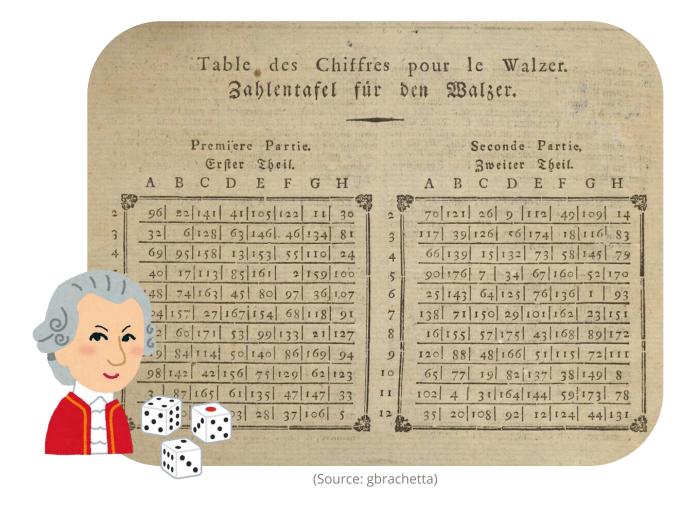
(Source: Illinois Distributed Museum)





(Source: The Guardian)

Musical Dice Game (1792)



gbrachetta.github.io/Musical-Dice/



Lejaren Hiller – ILLIAC Suite for String Quartet (1957)

ILLIAC SUITE FOR STRING QUARTET

I. EXPERIMENT NO. I



youtu.be/n0njBFLQSk8 & music.arts.uci.edu/abauer/3.1/scores/Hiller_Illiac_Suite.pdf 3

10

Emily Howell (by David Cope) – Prelude (2007)



Reading: David Cope on Emily Howell

- <u>David Cope: 'You pushed the button and out came hundreds and</u> <u>thousands of sonatas'</u>
 - Interview by Tim Adams published on *The Observer*, *The Guardians*, July 10, 2010

"People tell me they don't hear soul in the music," he says. "When they do that, I pull out a page of notes and ask them to show me where the soul is. We like to think that what we hear is soul, but I think audience members put themselves down a lot in that respect. The feelings that we get from listening to music are something we produce, it's not there in the notes. It comes from emotional insight in each of us, the music is just the trigger."



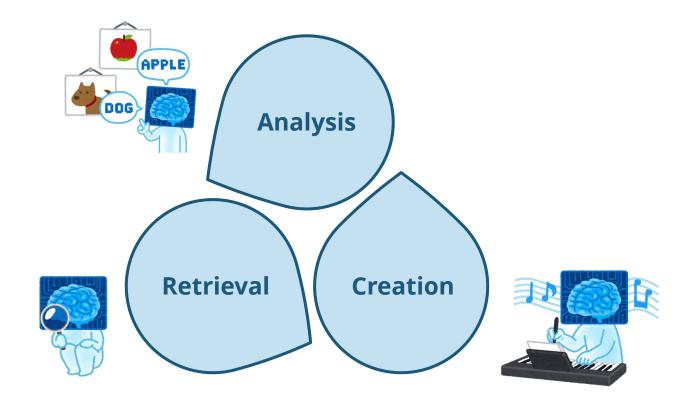
Analytic Al vs Generative Al





Music Information Research (MIR)

• "Intelligent ways to analyze, retrieve and create music" (Yang 2018)



Music Information Research (MIR)

EE



a female cat engineer making an electric chip in a classroom

Music



a cat playing heavy metal

CS



a cat engineer debugging on laptop

Topics

Analysis

 Genre/mood classification, source separation, transcription, optical music recognition, beat-tracking, structural analysis ...

Creation

• Automatic music composition, arrangement, performance synthesis, improvisation ...

Retrieval & Processing

- Search, recommendation, enhancement, production, editing, mixing ...

Advanced Topics

Interactive tools, singing voice synthesis, audiovisual tools ...

What about ChatGPT?

Large Language Models (LLMs)

• The models behind ChatGPT!

🗚 You

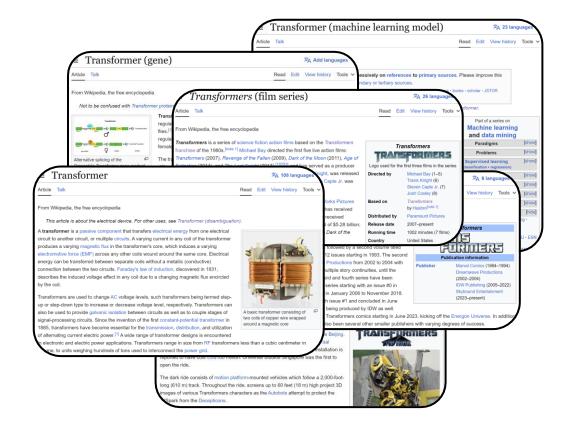
What's so cool about AI for music? Give me a brief answer

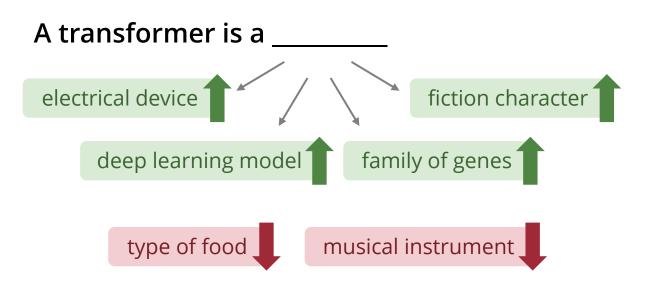
ChatGPT Word-by-word generation

Al in music is cool because it can compose original pieces. provide personalized recommendations, automate music production tasks, enhance creativity for artists, enable interactive performances, analyze music trends, and even create virtual artists, or bands, expanding the possibilities in music creation and enjoyment.

Language Models

• Predicting the next word given the past sequence of words

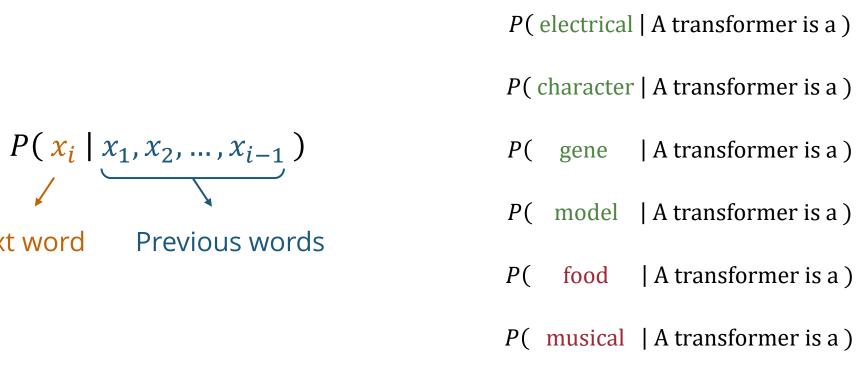




Language Models (Mathematically)

Next word

• A class of machine learning models that learn the next word probability



Language Models – Generation

• How do we generate a new sentence using a trained language model?

A transformer is a	\rightarrow	Model -	 deep
A transformer is a <mark>deep</mark>	\rightarrow	Model -	 learning
A transformer is a deep learning		Model -	 model
A transformer is a deep learning model	\rightarrow	Model -	 introduced
A transformer is a deep learning model introduced	\rightarrow	Model -	 in
A transformer is a deep learning model introduced in	\rightarrow	Model -	 2017

Designing a Machine-readable Music Language

- How can we "represent" music in a way that machines understand?
- Why not using sheet music "images" directly?
 - Machines still have a hard time reading sheet music
 - A challenging task known as "optical music recognition" (OMR)
- Examples:
 - ABC notation
 - MIDI



ABC Notation

- A simple text-based notation
- Use letters to denote pitches
 - Lower octave (A–G), higher octave (a–g)
- Use prefix to denote accidentals
 - Sharp (^), flat (_), natural (=)





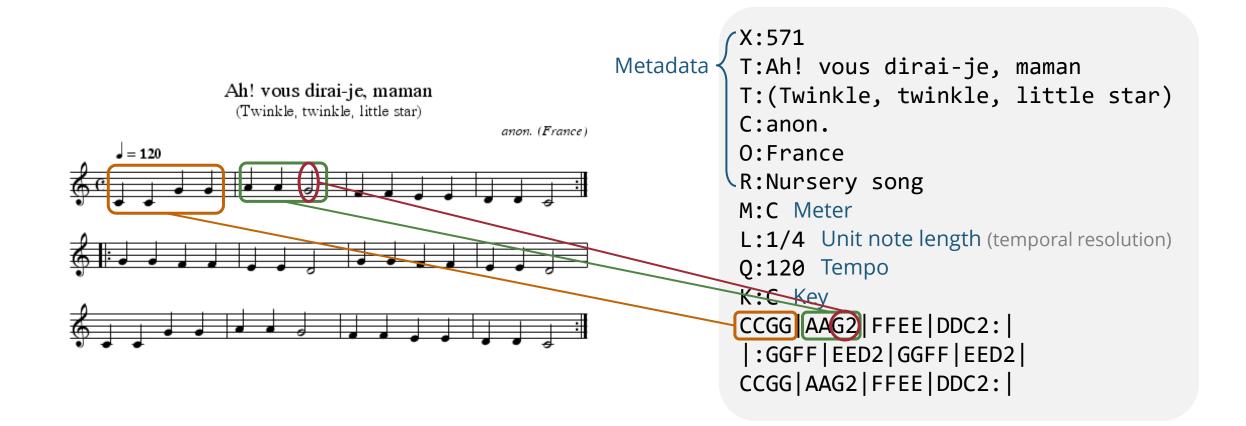
C, D, E, F, |G, A, B, C | D E F G | A B c d | e f g a | b c' d' e' | f' g' a' b'

What is this song in ABC notation?

CCGG AAG2 FFEE DDC2: |:GGFF EED2 GGFF EED2 CCGG AAG2 FFEE DDC2:

Twinkle, twinkle, little star!

An Example of ABC Notation

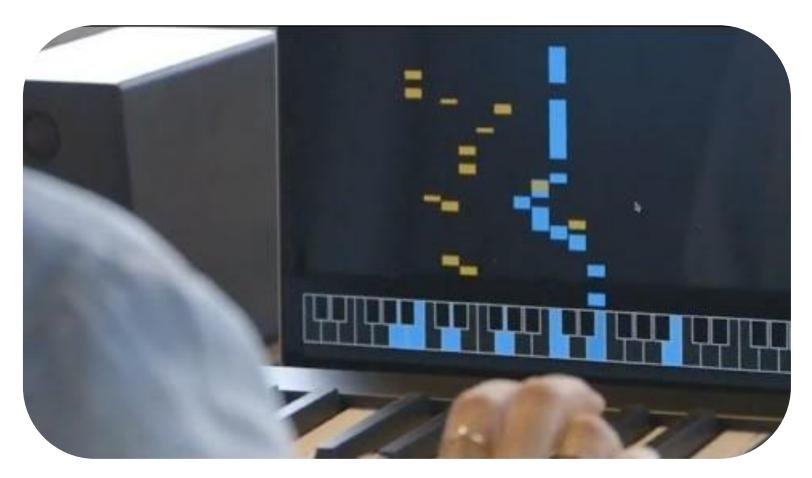


ChatGPT can also Compose Songs in ABC Notation!

SA	You Write a cheerful folk song using ABC notation		\$4
6	ChatGPT Certainly! Here's a cheerful folk song in ABC notation:		8
	vbnet	பி Copy code	0
	X:1		6
	T:Sunshine on the Meadow M:4/4		0
	K:C		.
	G2 G2 E2 E2 F2 F2 D4 E2 E2 G2 G2 A4 G4 G2 G2 E2 E2 F2 F2 D4 E2 E2 G2 G2 C4 E4		
	G2 G2 E2 E2 F2 F2 D4 E2 E2 G2 G2 A4 G4		
	G2 G2 E2 E2 F2 F2 D4 G4 F4 E4 D4 C4 z4		
	G2 G2 E2 E2 F2 F2 D4 E2 E2 G2 G2 A4 G4		^
	G2 G2 E2 E2 F2 F2 D4 E2 E2 G2 G2 C4 E4 G2 G2 E2 E2 F2 F2 D4 E2 E2 G2 G2 A4 G4		
	G2 G2 E2 E2 F2 F2 D4 G4 F4 E4 D4 C4 z4		J
	Feel free to play around with the tempo and instrumentation to give it you	r own unique	8
	style!		
	<u>chat.openai.com</u>		}



Example: A.I. Duet (Mann et al, 2016)



<u>youtu.be/0ZE1bfPtvZo</u> <u>experiments.withgoogle.com/ai/ai-duet/view</u>

More Examples

Example: Piano Genie (Donahue et al., 2018)



youtu.be/YRb0XAnUpIk & magenta.tensorflow.org/pianogenie

piano-genie.glitch.me/

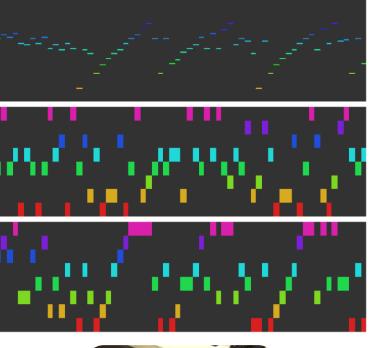


Example: Piano Genie (Donahue et al., 2018)

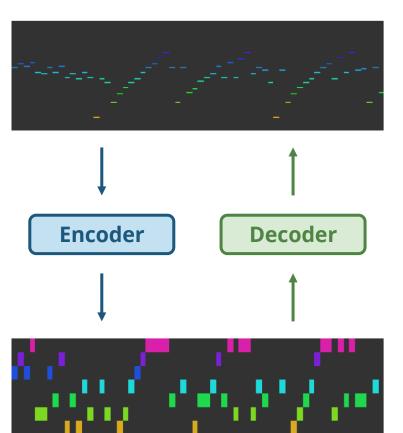
Input melody

Baseline

Proposed

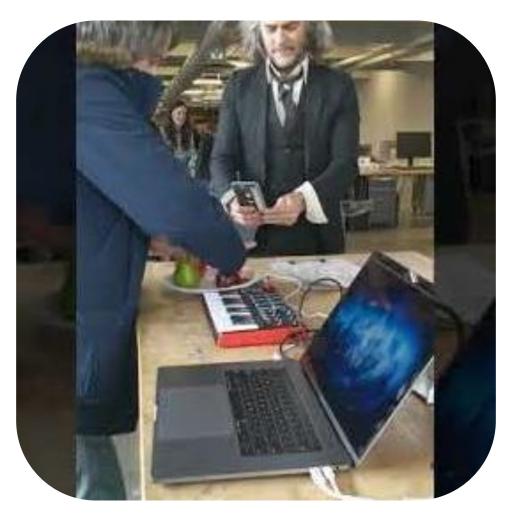






(Source: Donahue et al., 2019)

Example: Fruit Genie (2019)



youtu.be/HoVs4kC68no

Example: Fruit Genie Live (2019)



youtu.be/L4wvXrPmIkU

Example: JS Bach Doodle (2019)

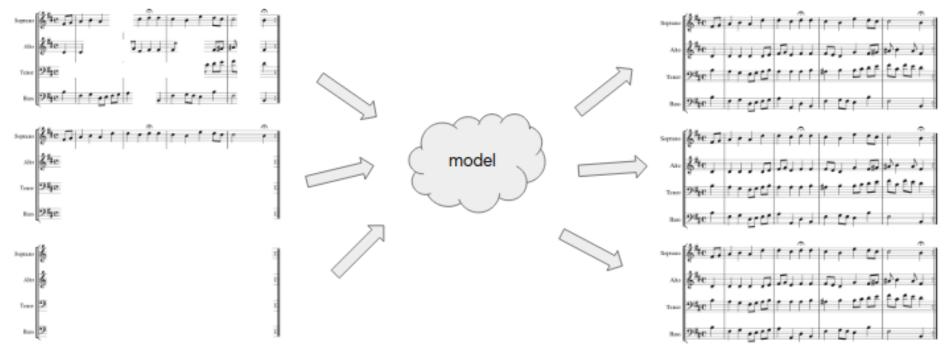


youtu.be/XBfYPp6KF2g & magenta.tensorflow.org/coconet

doodles.google/doodle/ celebrating-johannsebastian-bach/

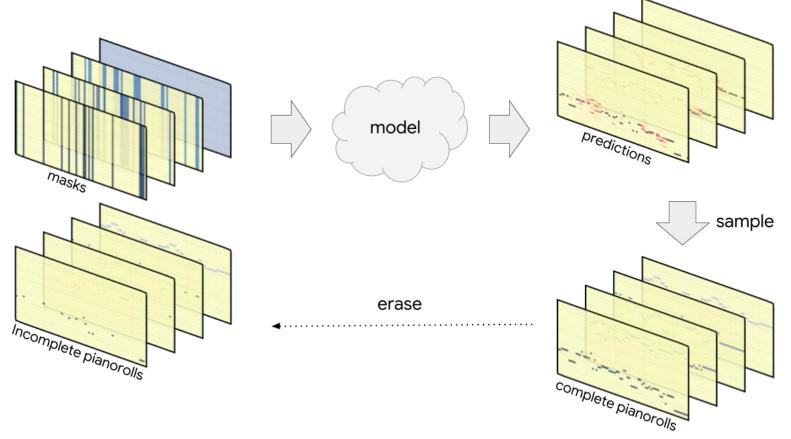


• Based on Orderless NADE (Uria et al, 2014)



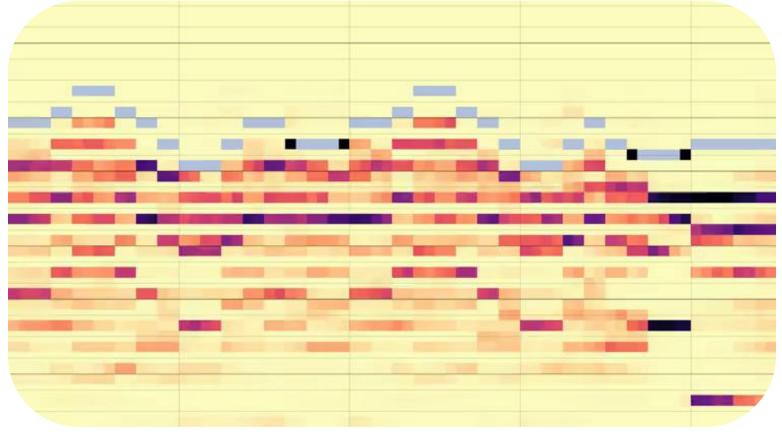
(Source: Huang et al., 2019)

Benigno Uria, Iain Murray, and Hugo Larochelle, "<u>A Deep and Tractable Density Estimator</u>," *ICML*, 2014. Cheng-Zhi Anna Huang, Tim Cooijmans, Adam Roberts, Aaron Courville, and Douglas Eck, "<u>Counterpoint by Convolution</u>," *ISMIR*, 2017. Cheng-Zhi Anna Huang, Tim Cooijmans, Monica Dinculescu, Adam Roberts, and Curtis Hawthorne, "<u>Coconet: the ML model behind today's Bach Doodle</u>," *Magenta Blog*, 2019.

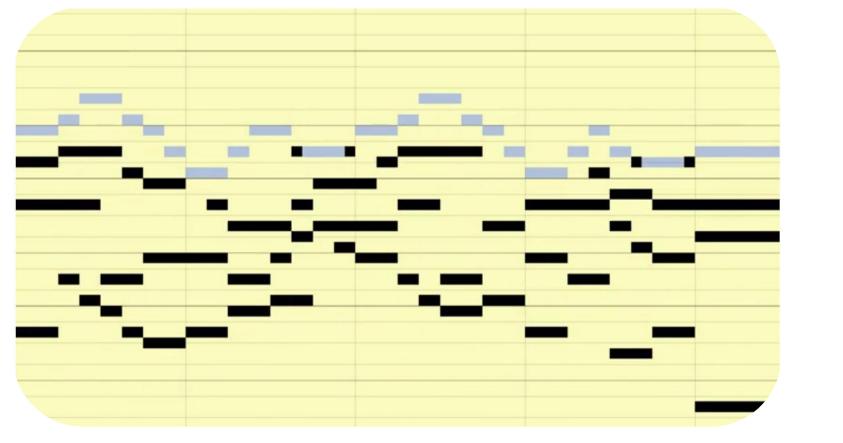


(Source: Huang et al., 2019)

Cheng-Zhi Anna Huang, Tim Cooijmans, Adam Roberts, Aaron Courville, and Douglas Eck, "<u>Counterpoint by Convolution</u>," *ISMIR*, 2017. Cheng-Zhi Anna Huang, Tim Cooijmans, Monica Dinculescu, Adam Roberts, and Curtis Hawthorne, "<u>Coconet: the ML model behind today's Bach Doodle</u>," *Magenta Blog*, 2019.

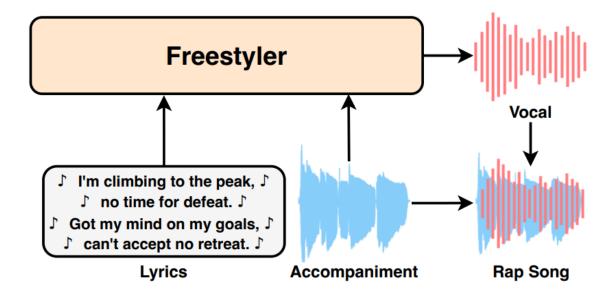


(Source: Huang et al., 2017)



(Source: Huang et al., 2017)

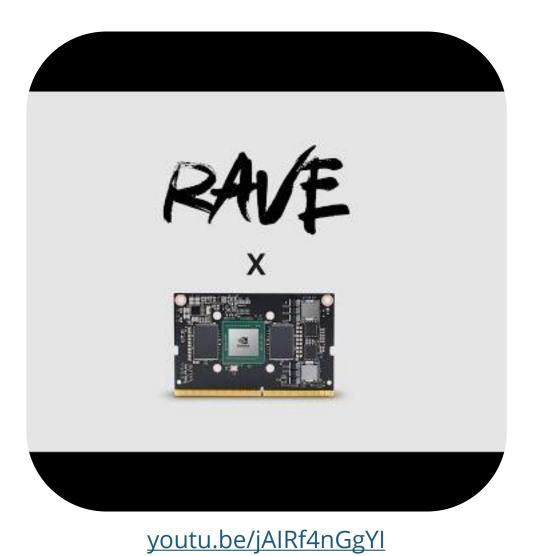
Example: Freestyler (Ning et al., 2024)



nzqian.github.io/Freestyler

Ziqian Ning, Shuai Wang, Yuepeng Jiang, Jixun Yao, Lei He, Shifeng Pan, Jie Ding, and Lei Xie, "Drop the beat! Freestyler for Accompaniment Conditioned Rapping Voice Generation," arXiv preprint arXiv:2408.15474, 2024.

Example: RAVE (Caillon & Esling, 2021)



github.com/acidsircam/RAVE



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

Example: MusicLDM (Chen et al., 2023)



youtu.be/DALv7ea6cv0