A PRACTICAL GUIDE TO GENERATIVE MUSIC AI FOR DEVELOPERS

Session 1: An Introduction to AI Music

ABOUT ME

- PhD student & Presidential Fellow at the MIT Media Lab
- MEng and MRes in *Artificial Intelligence & Machine Learning* at Imperial College London
- Classically-trained Jazz pianist

MY WORK

- Working with multiple GRAMMY-winning artists to develop Generative AI-powered musical instruments for live performances.

- Generally interested in exploring the field of real-time Artist-AI musical interaction.

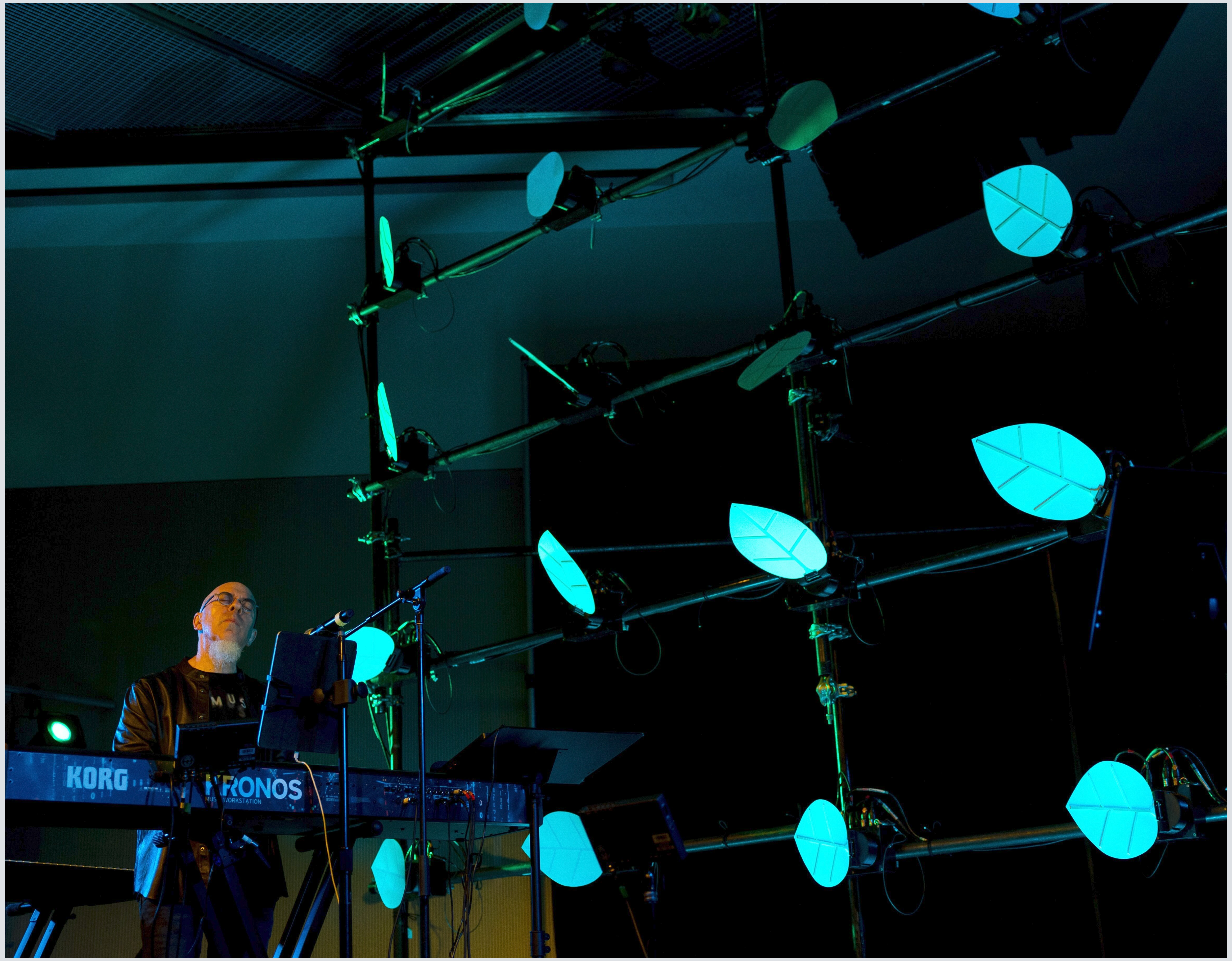


Photo Credit: Caroline Alden

A BRIEF HISTORY OF **AI** AND **MUSIC**

Before AI, "*generative music*" (a term popularized by Brian Eno) referred to music that was created by a **system** and that **changed over time**.

Diverse systems have been used to create music.

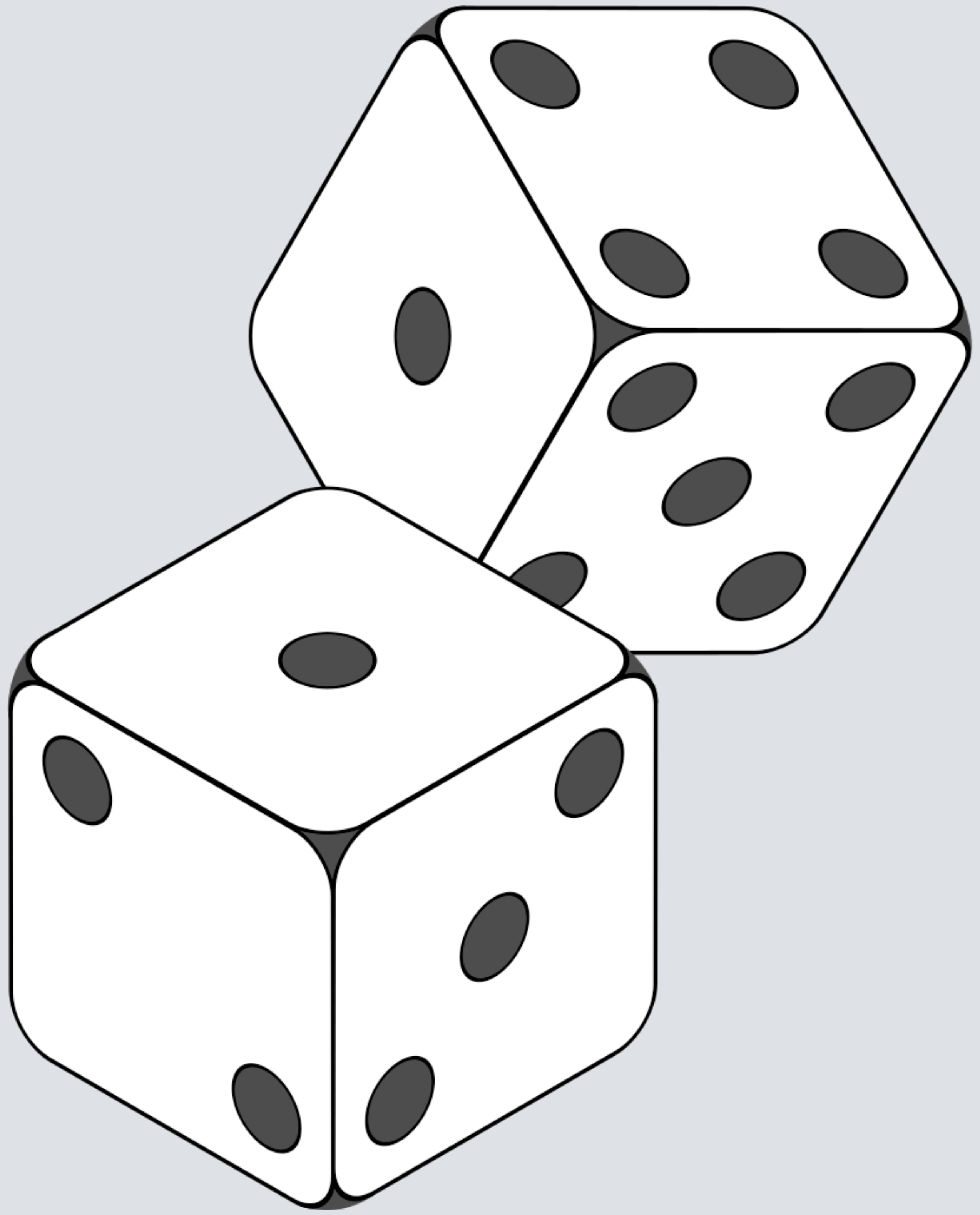
- Randomness and chance
- Algorithms and computers
- Deep Learning methods

ALEATORIC MUSIC

THROUGH RANDOMNESS AND CHANCE

Musikalisches Würfelspiel

- Attributed to Mozart, Haydn, and Bach.
- A number of pre-computed bars, randomly selected by dice roll.



Mozart's *Anleitung zum Componieren von Walzern so viele man will vermittelst zweier Würfel, ohne etwas von der Musik oder Composition zu verstehen*

("Instructions for the composition of as many waltzes as one desires with two dice, without understanding anything about music or composition")

ALGORITHMIC MUSIC

THROUGH ALGORITHMS AND COMPUTERS

Illiac Suite (1957)

Composed by Lejaren Hiller & Leonard Isaacson at the *University of Illinois, Urbana-Champaign*.

- First computer-generated piece
- Four mouvements

ALGORITHMIC MUSIC

THROUGH ALGORITHMS AND COMPUTERS

Project One & Project Two (1964-1966)

Composed by Gottfried Michael Koenig at the *Institute of Sonology* in Utrecht.

- Probabilistic rule-based system that defined musical parameters such as pitch, duration, dynamics, and timbre.

ALGORITHMIC MUSIC

THROUGH ALGORITHMS AND COMPUTERS

Jukedeck (2012)

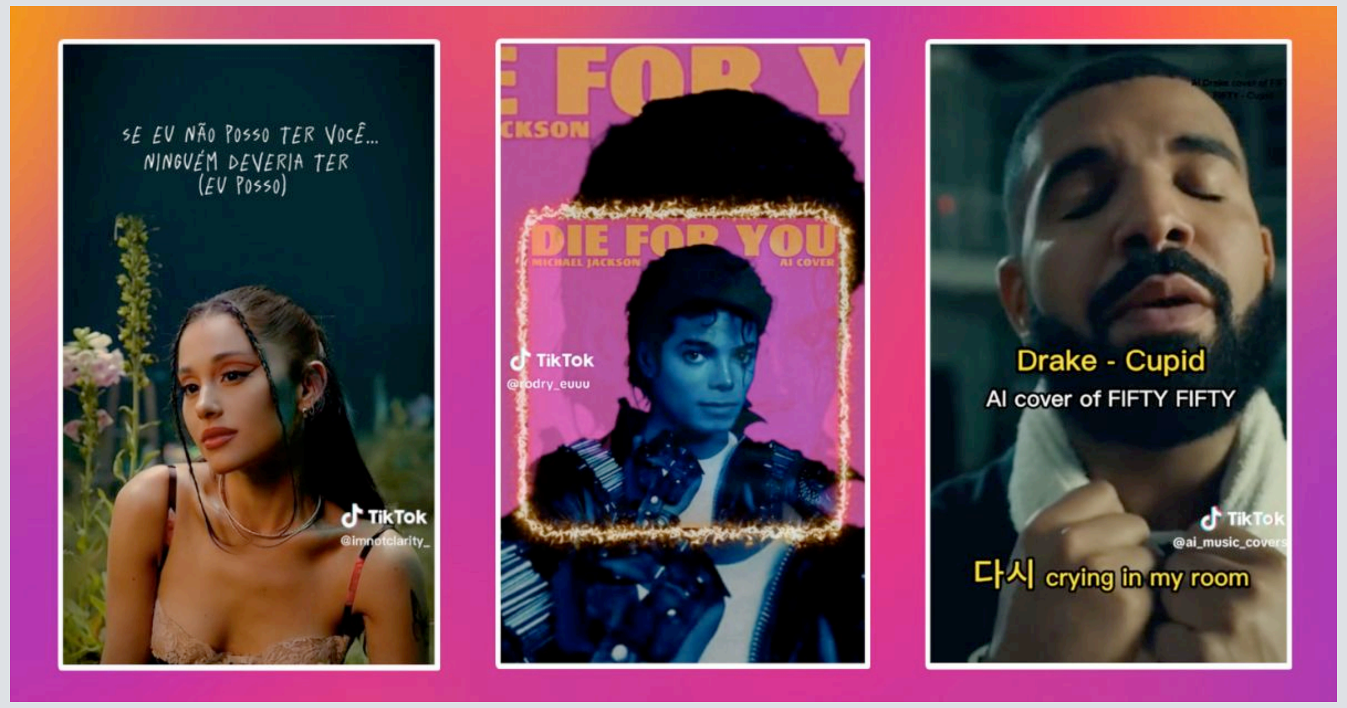
Founded by Ed-Newton Rex and Patrick Stobbs

- Royalty-free music generation through retrieval and samples and loops.
- Acquired by ByteDance.

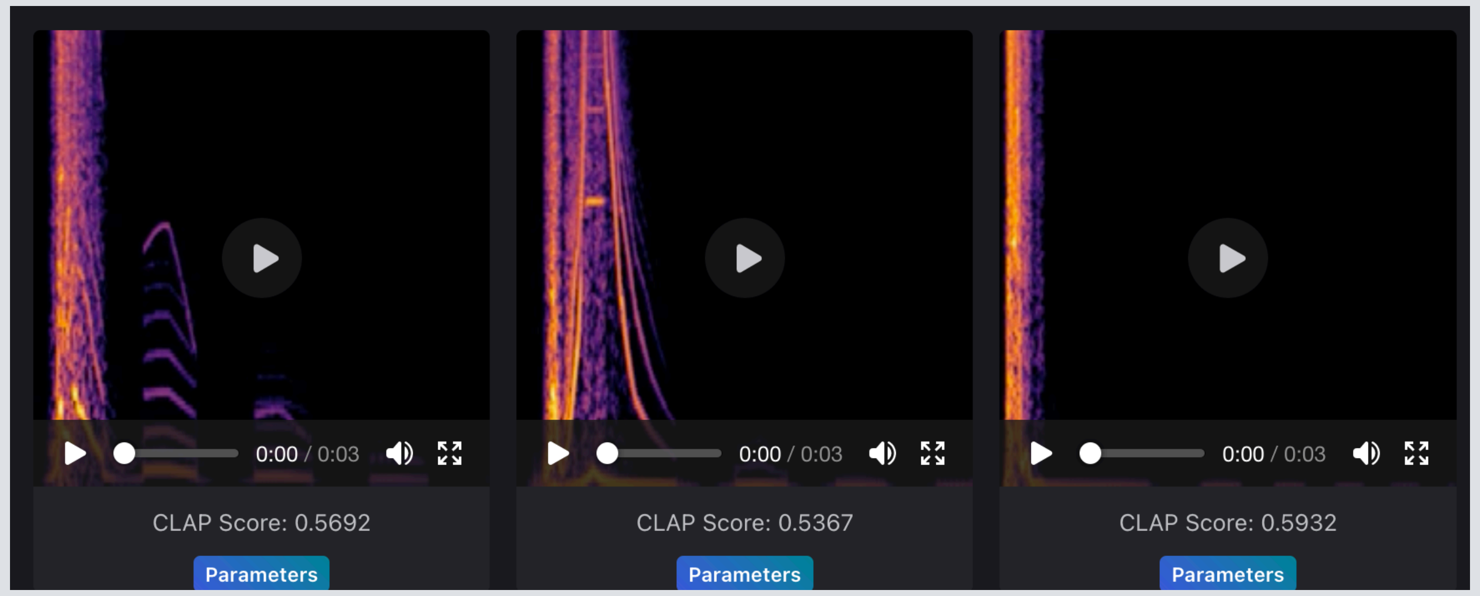
Jukedeck

CURRENT STATE OF **AI** AND **MUSIC**

MusicGen



SUNO



udio

SYMBOLIC VS AUDIO

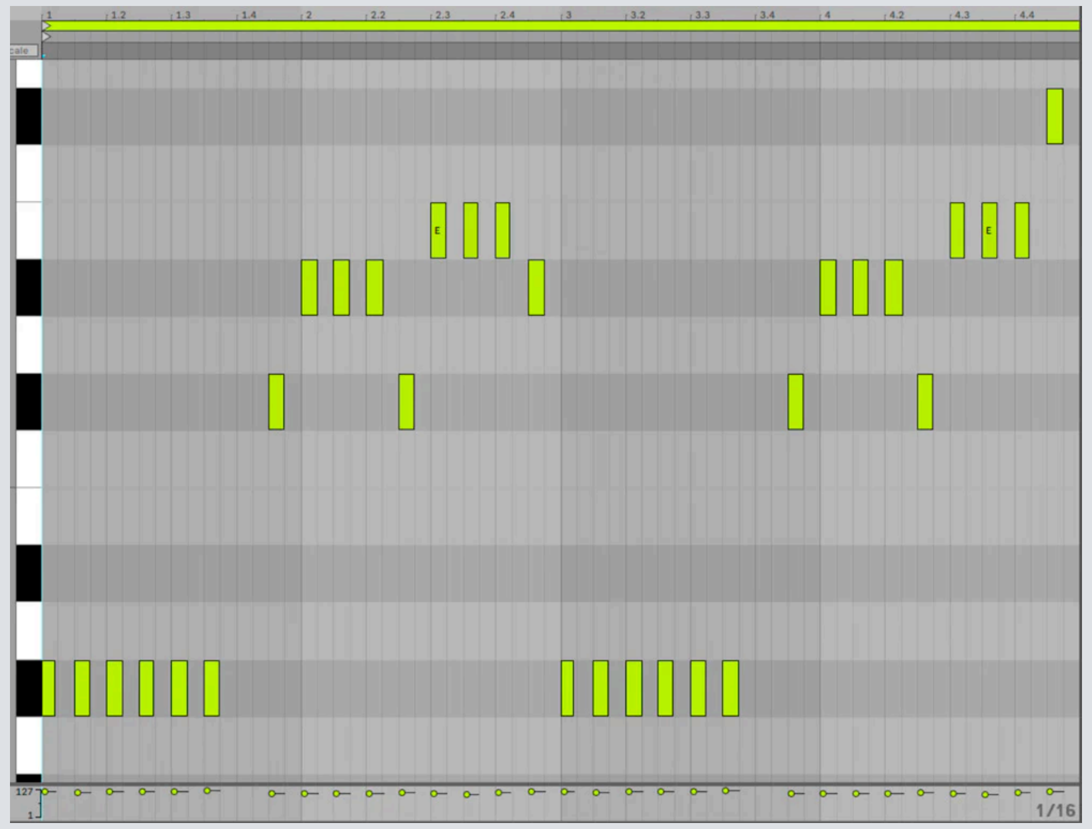
SYMBOLIC MUSIC

Control over

- Harmony
- Melody
- Rhythms
- Expression & Articulation
- Orchestration & Arrangement

SYMBOLIC MUSIC

Can be represented by




```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>4</octave>
  </pitch>
  <duration>2</duration>
  <type>half</type>
</note>
```

```
X: 0
T: My Tune
M: 4/4
L: 1/4
K: C
C,D,E,F | G,A,B,C | DEFG | ABcd | efga | bc’d’e’ | f’g’a’b’ |
```

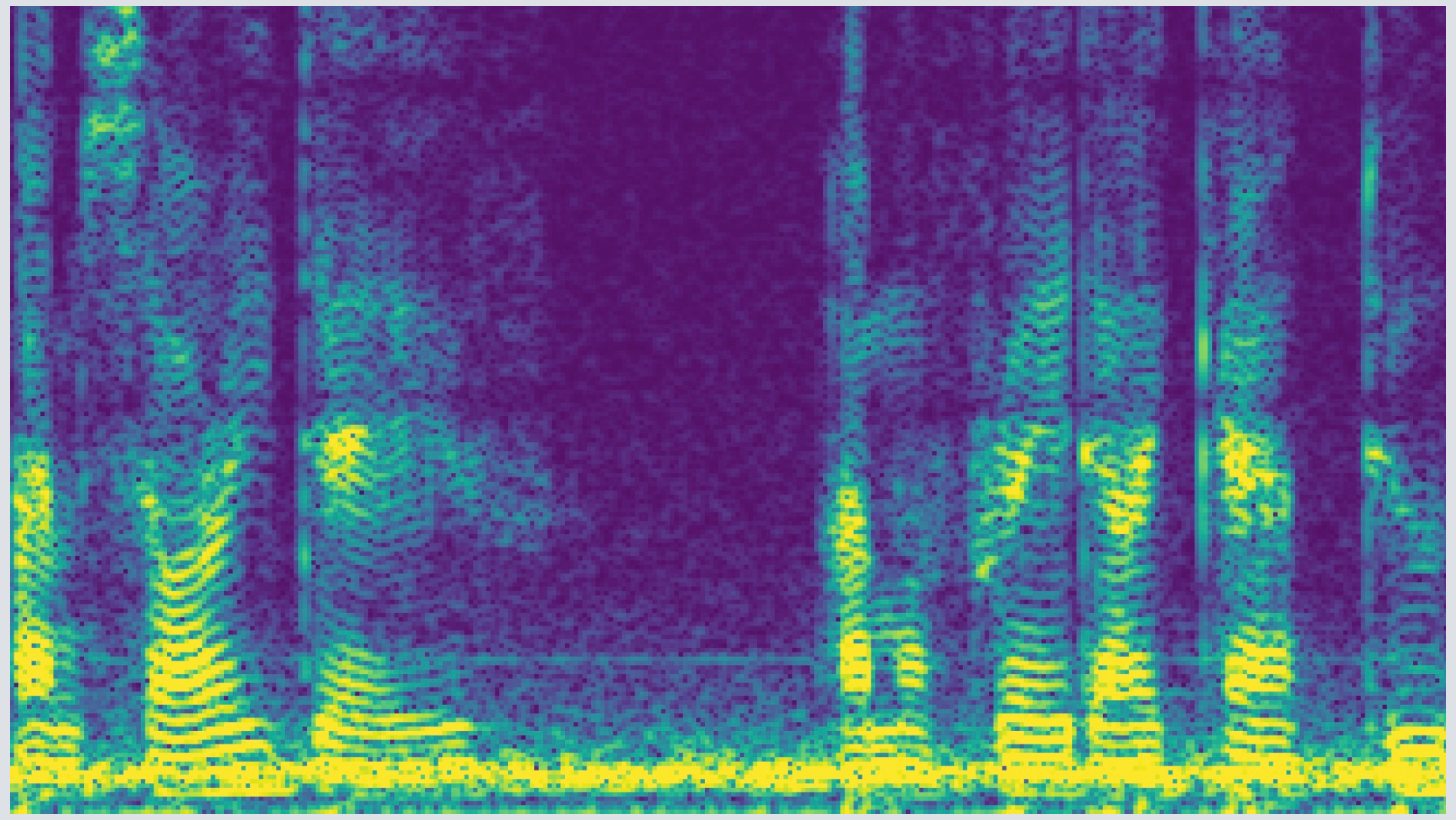
AUDIO

Control over

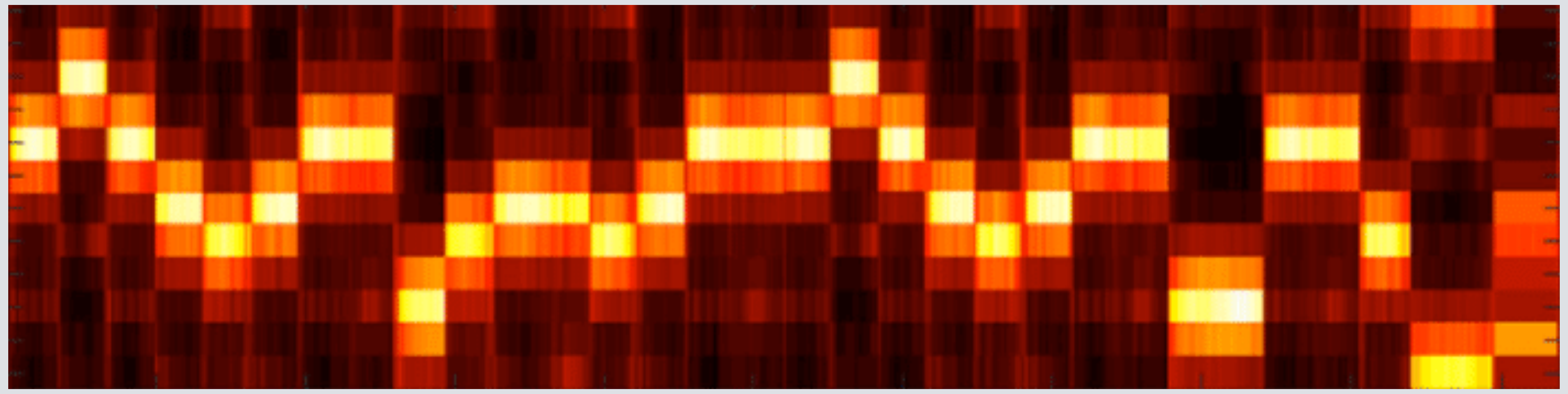
- Timbre
- Sonic details
- Phrasing
- Mixing & Mastering

AUDIO

Can be represented by



```
[[[[0.1561, ...,  0.2555],
   ...
 [[[0.8834, ..., 0.5366],
   ...
 [0.9060, ..., 0.2069]]]]
```



AVAILABLE TECHNOLOGIES

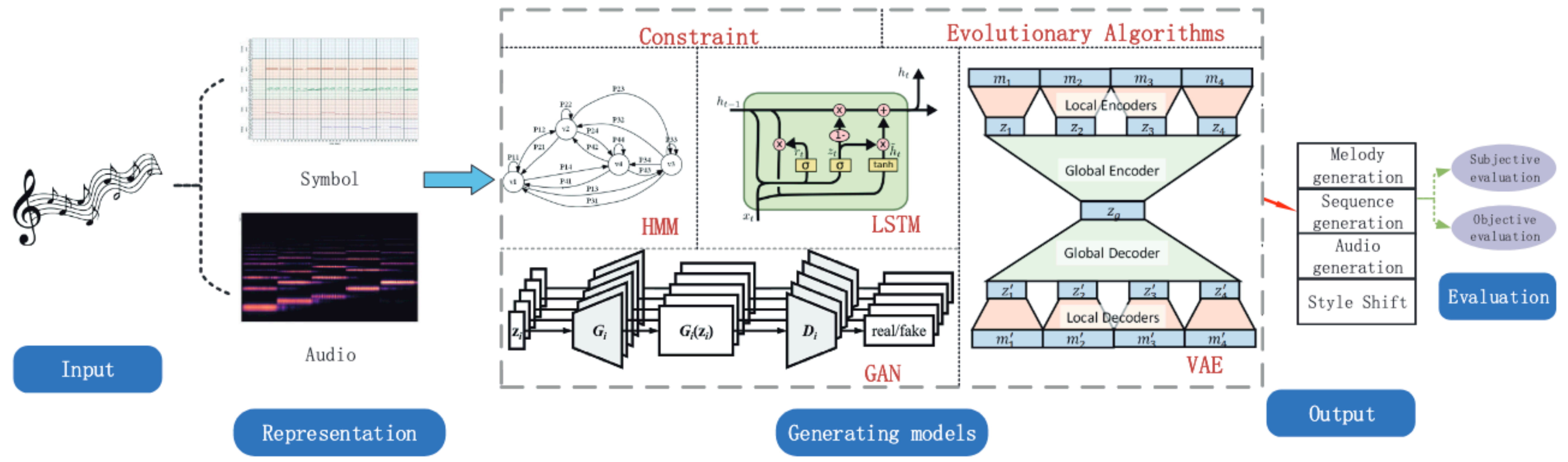


Fig. 1: Overview of music generation system.

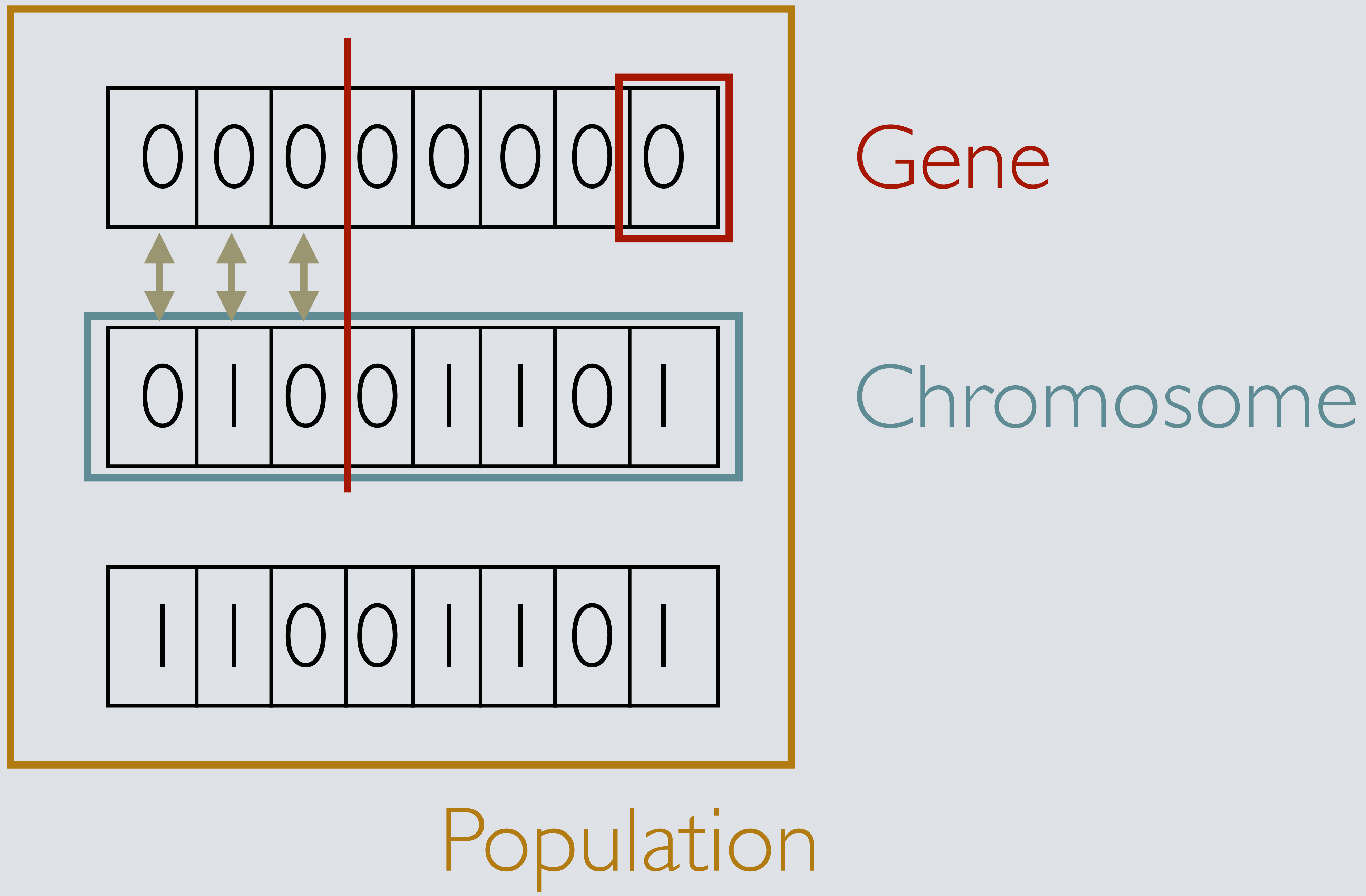
from (Wang et al., 2022)

AVAILABLE TECHNOLOGIES

PRE DEEP LEARNING

Genetic Algorithms

- Evolutionary search over populations of musical sequences
- Mutation, crossover, and selection optimize fitness (e.g., harmony, melody)
- Good at exploration but slow convergence
- e.g., John A. Bile's *GenJam*

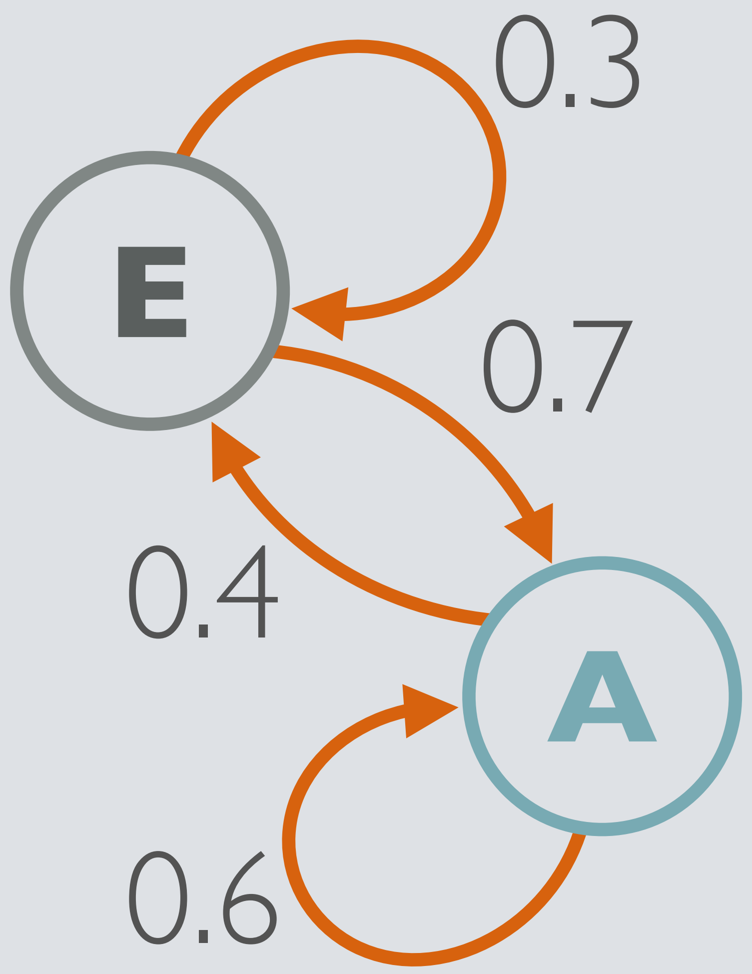


AVAILABLE TECHNOLOGIES

PRE DEEP LEARNING

Hidden Markov Models (HMMs)

- State-based system with transition probabilities.
- e.g. François Pachet's *Continuator*

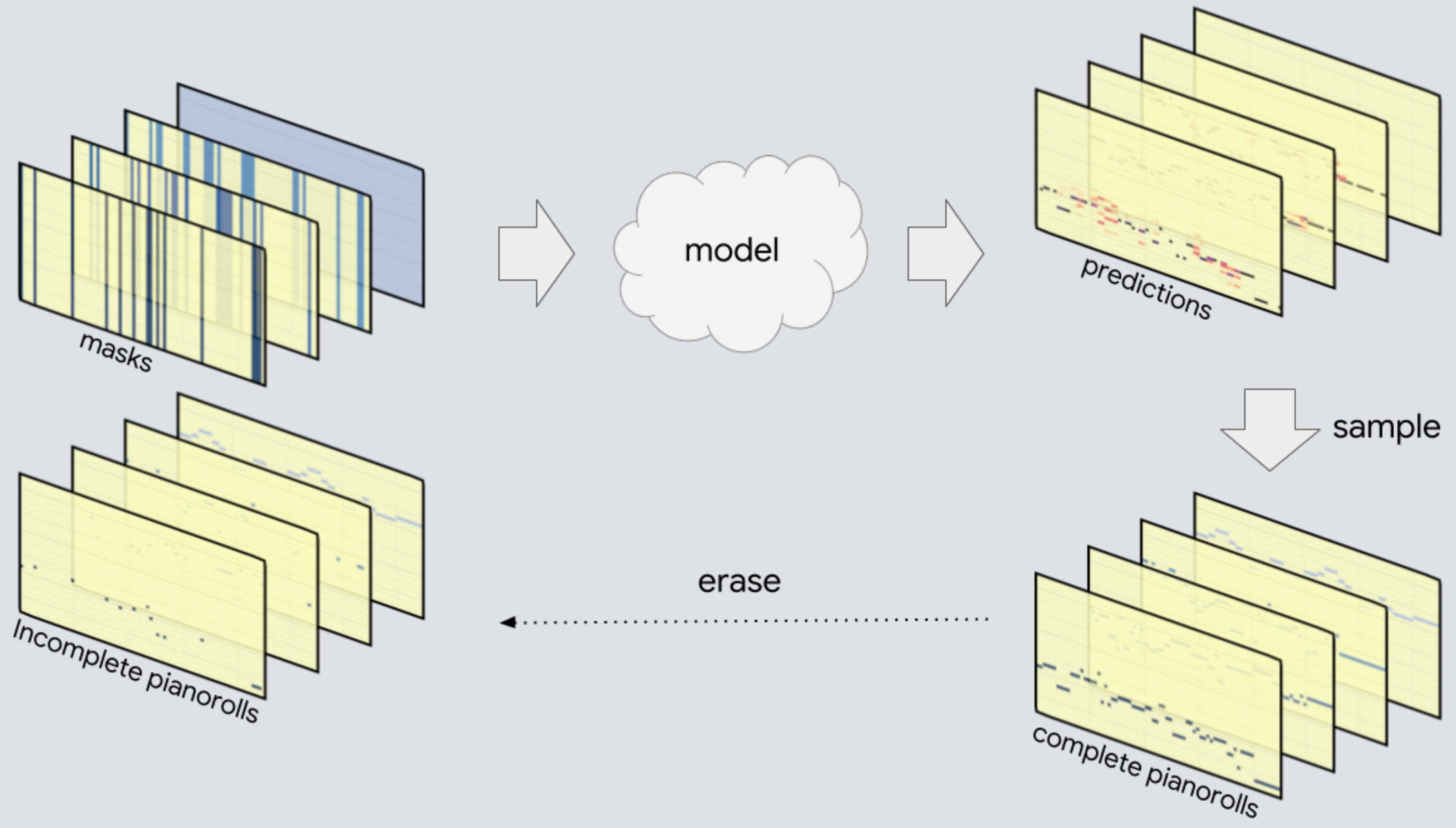


AVAILABLE TECHNOLOGIES

DEEP LEARNING - EARLY ERA

Convolutional Neural Networks (CNNs)

- Learn local patterns (e.g., rhythms, timbres) in audio or pianoroll formats
- Limited at modeling long-term structure
- Strong at capturing spatial or short-time structure
- e.g. Google Magenta's *Coconet*

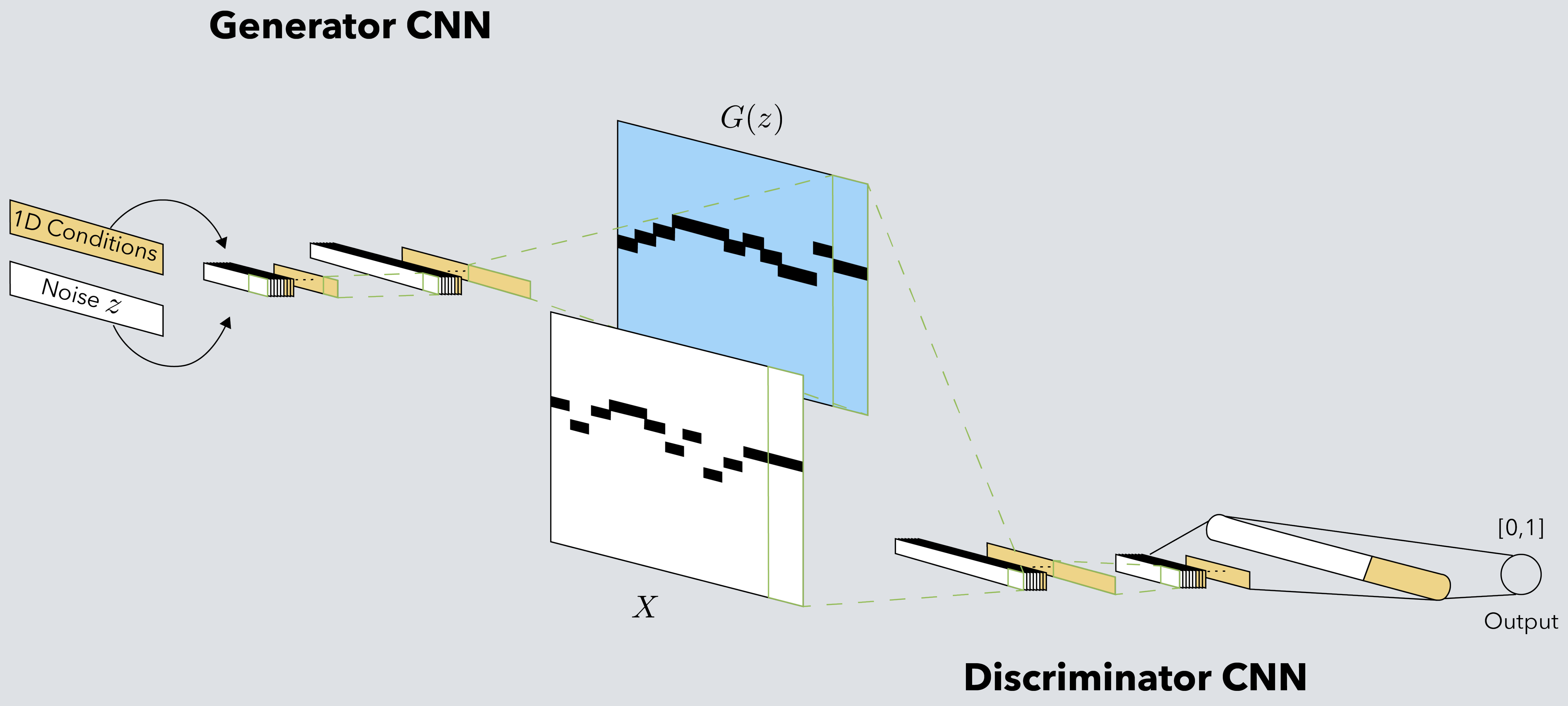


AVAILABLE TECHNOLOGIES

DEEP LEARNING - EARLY ERA

Generative Adversarial Networks (GANs)

- Generator creates music; discriminator evaluates realism.
- Adversarial training improves quality and diversity.
- Often used for symbolic generation (piano rolls) or spectrograms.
- e.g. Yang et al.'s *MidiNET*

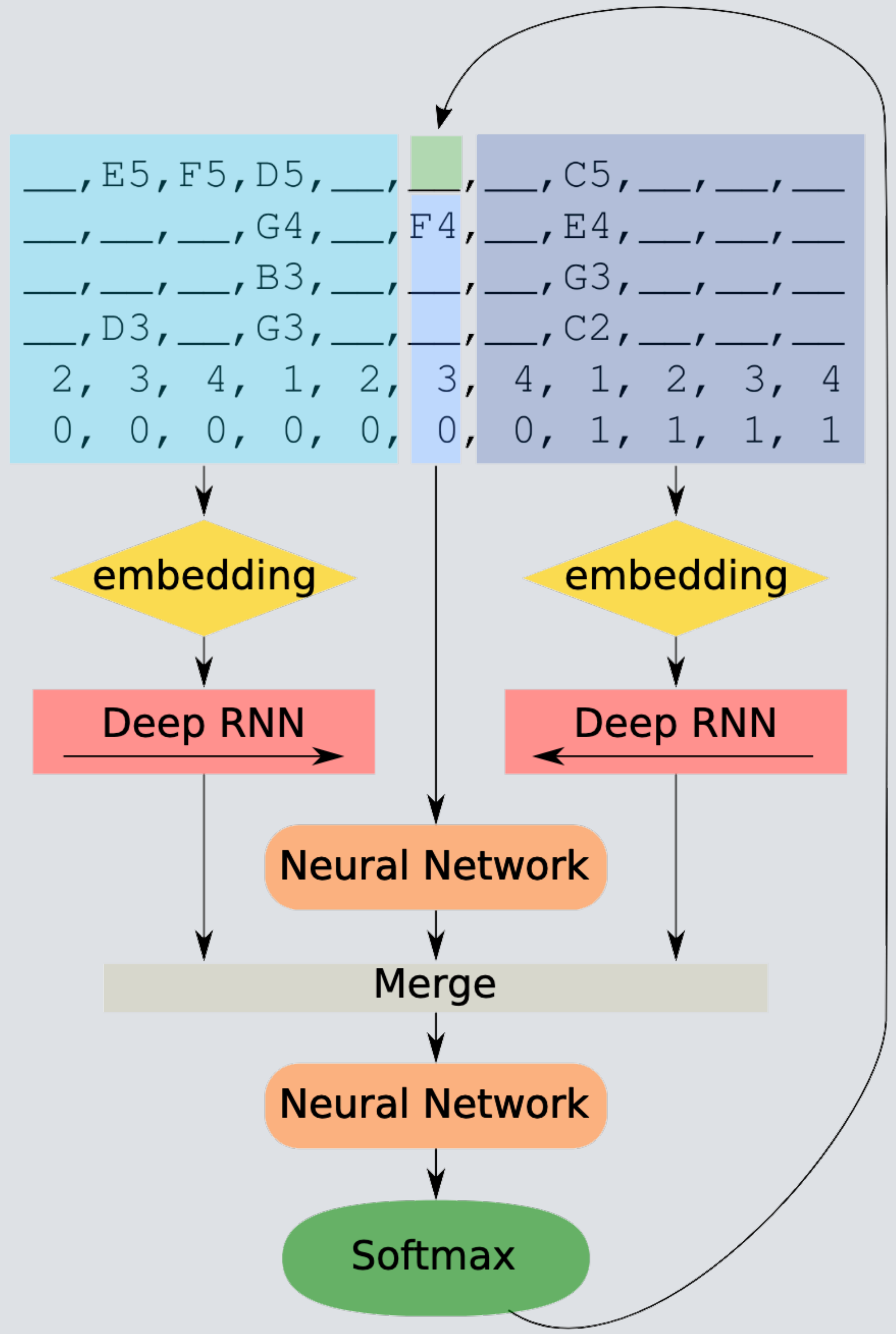


AVAILABLE TECHNOLOGIES

DEEP LEARNING - EARLY ERA

Recurrent Neural Networks (RNNs)

- Model sequences step-by-step with internal memory (hidden states)
- Good at short-term dependencies; struggle with very long-term coherence
- Variants like LSTM and GRU mitigate vanishing gradients
- e.g. Sony CSL's *DeepBach*

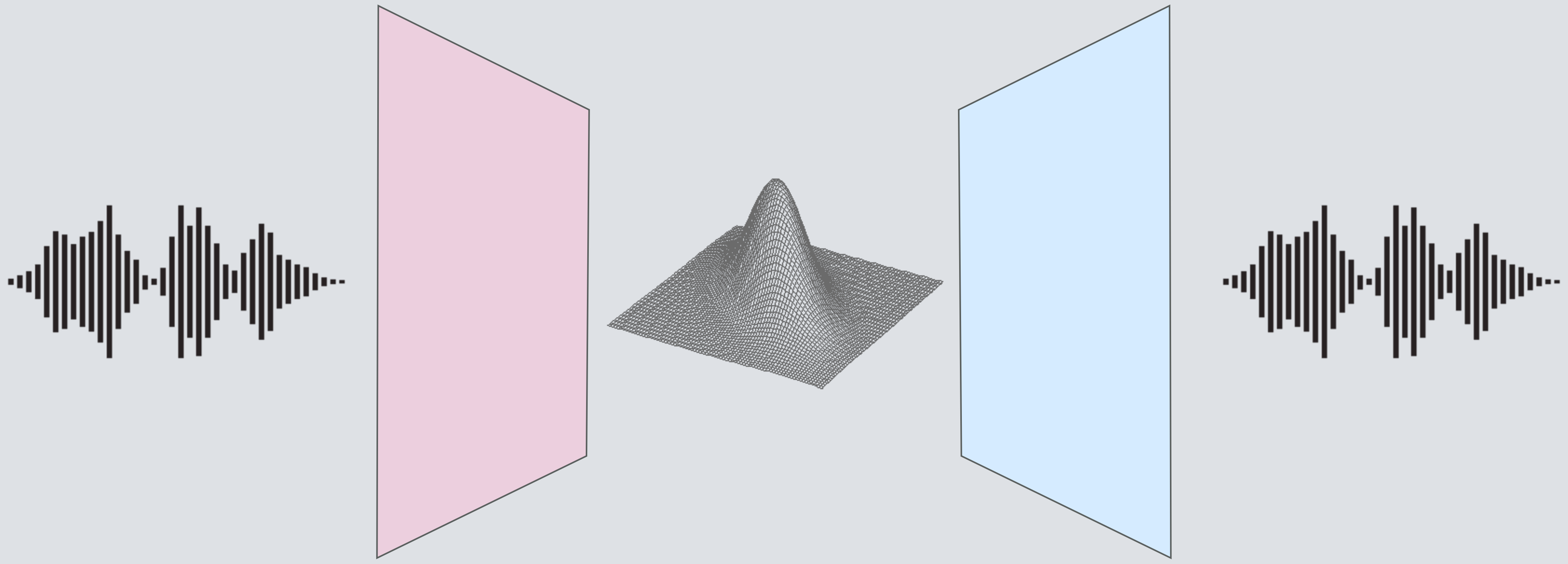


AVAILABLE TECHNOLOGIES

DEEP LEARNING - CURRENT ERA

Variational AutoEncoders (VAEs)

- Learn to create a latent representation that allows accurate reconstruction.
- Varying encoding levels
- Limited long-term dependencies
- e.g. OpenAI's *Jukebox*

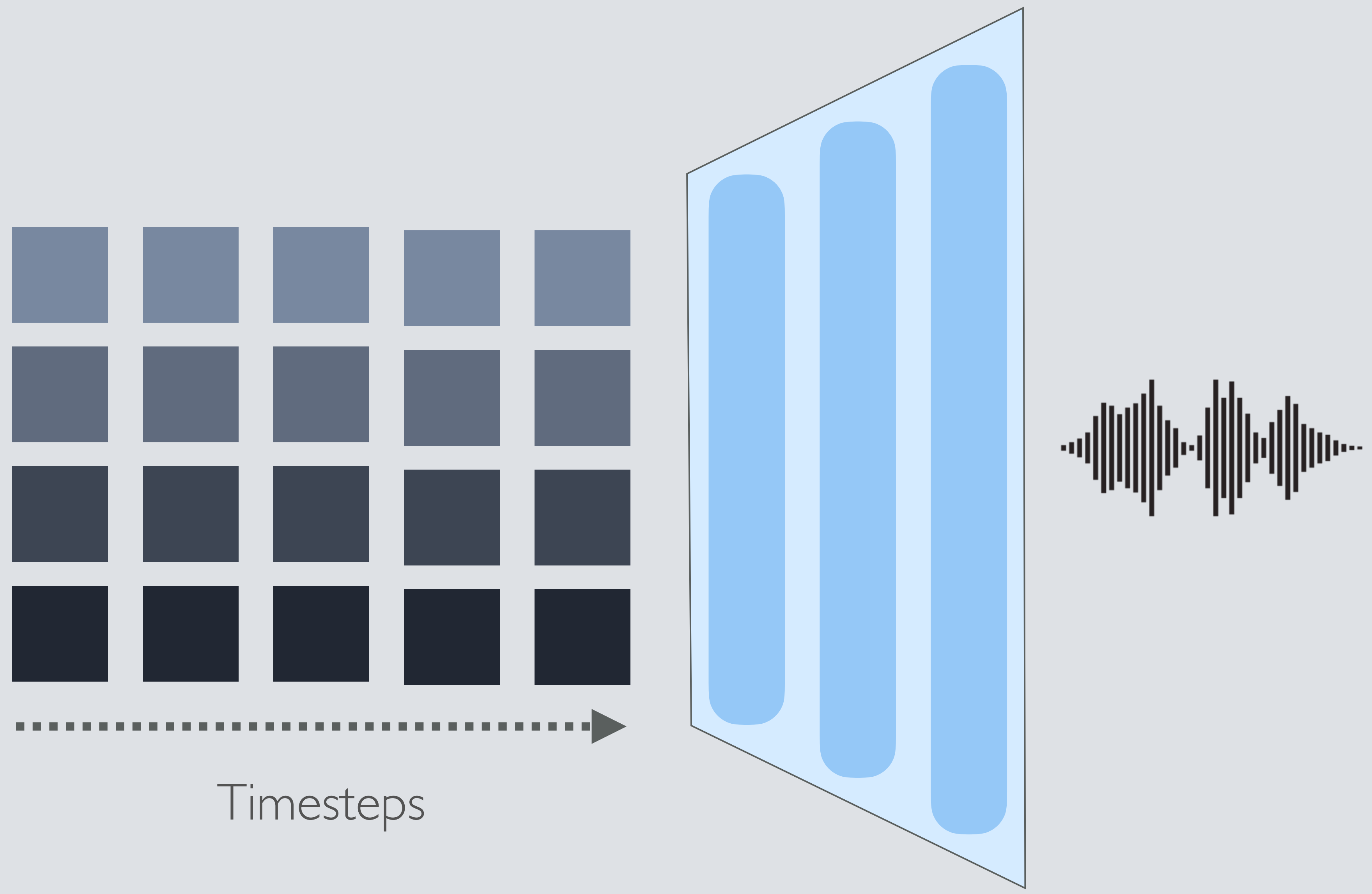


AVAILABLE TECHNOLOGIES

DEEP LEARNING - CURRENT ERA

Transformers

- Learn to model and auto-regressively complete a sequence of music tokens
- Better long-term dependencies
- e.g. Meta's *MusicGen*

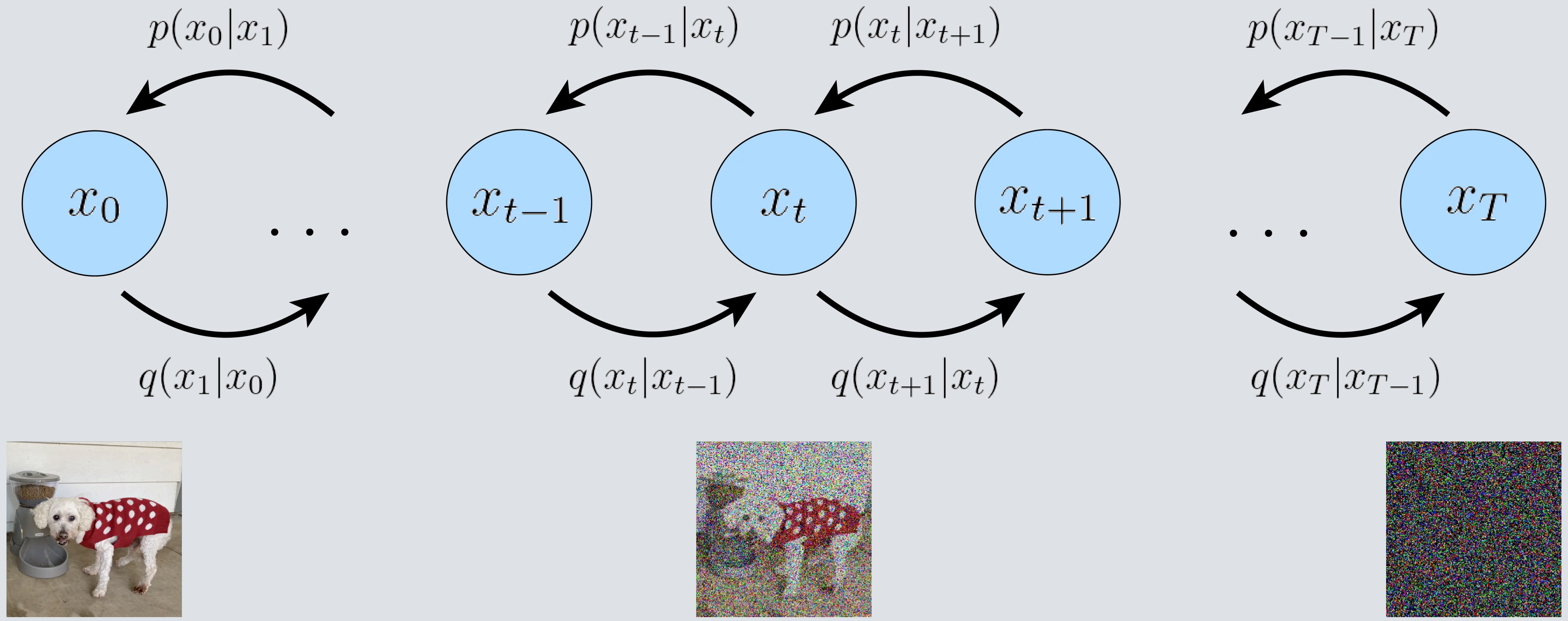


AVAILABLE TECHNOLOGIES

DEEP LEARNING - CURRENT ERA

Diffusion Models

- Learn a denoising process over a (latent) audio representation
- Great overall cohesion
- Limited long-term dependencies
- e.g. ByteDance's *AudioLDM2*



IN THIS COURSE

- Week 1: Introduction & Foundational Skills
- Week 2: ML Fundamentals applied to Music
- Week 3: (Re)using State-Of-The-Art Generative AI Models for Musical Creation
- Week 4: Real-Time Music Generation
- Week 5: Commercial Applications & Final Project Preparation
- Week 6: Final Projects

THANK YOU!

REFERENCES

- Hiller, L. 1981. Composing with computers: A progress report. Computer Music Journal, 5(4):7-21.
- Laske, O. 1981. Composition theory in Koenig's Project One and Project Two. Computer Music Journal, 5(4):54-65.
- Wang, L. et al. 2022. A Review of Intelligent Music Generation Systems.