PAT 464/564 (Winter 2026)

Generative AI for Music & Audio Creation

Lecture 8: Deep Generative Models

Instructor: Hao-Wen Dong

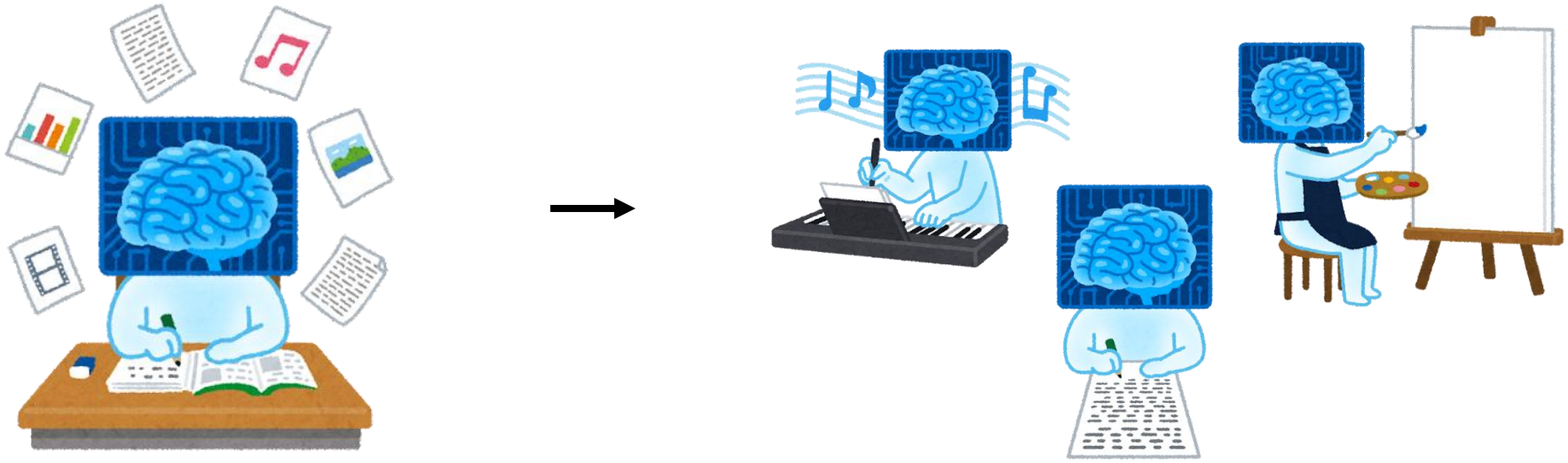
UNIVERSITY OF MICHIGAN

What I cannot create, I do not understand.

– Richard Feynman

Generative AI

What is Generative AI?

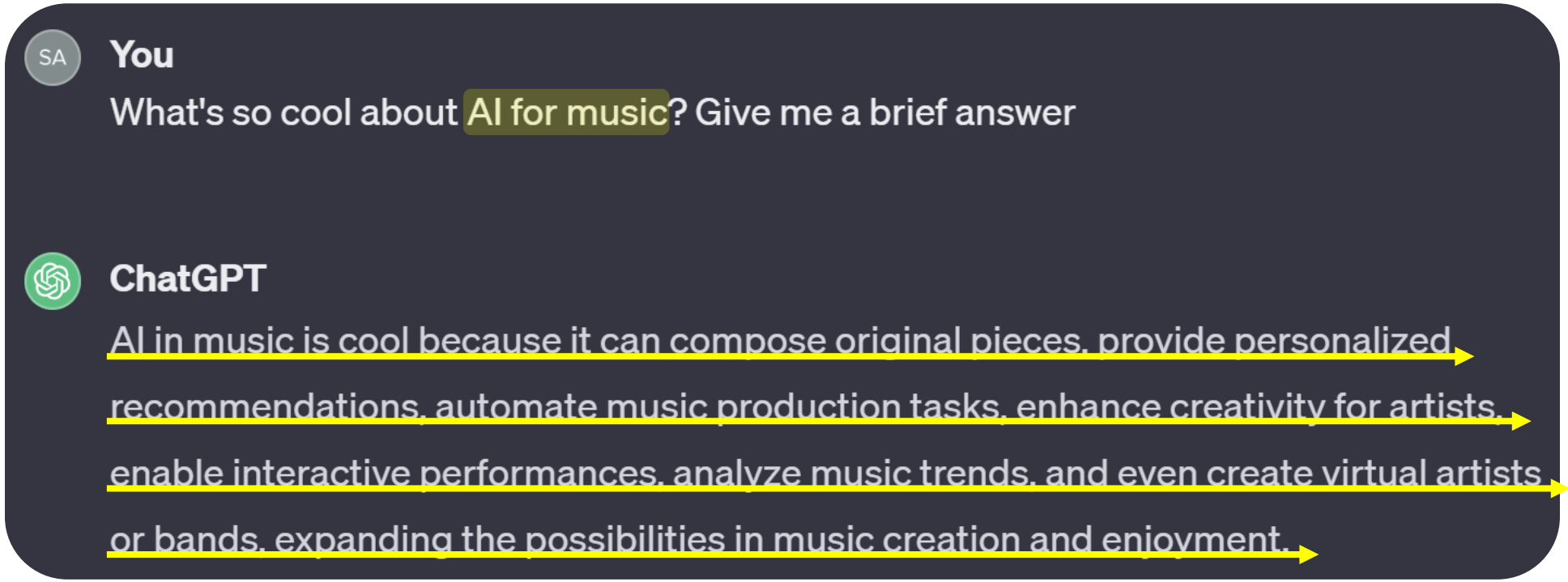


Generative AI is AI capable of generating text, images, music or other media.

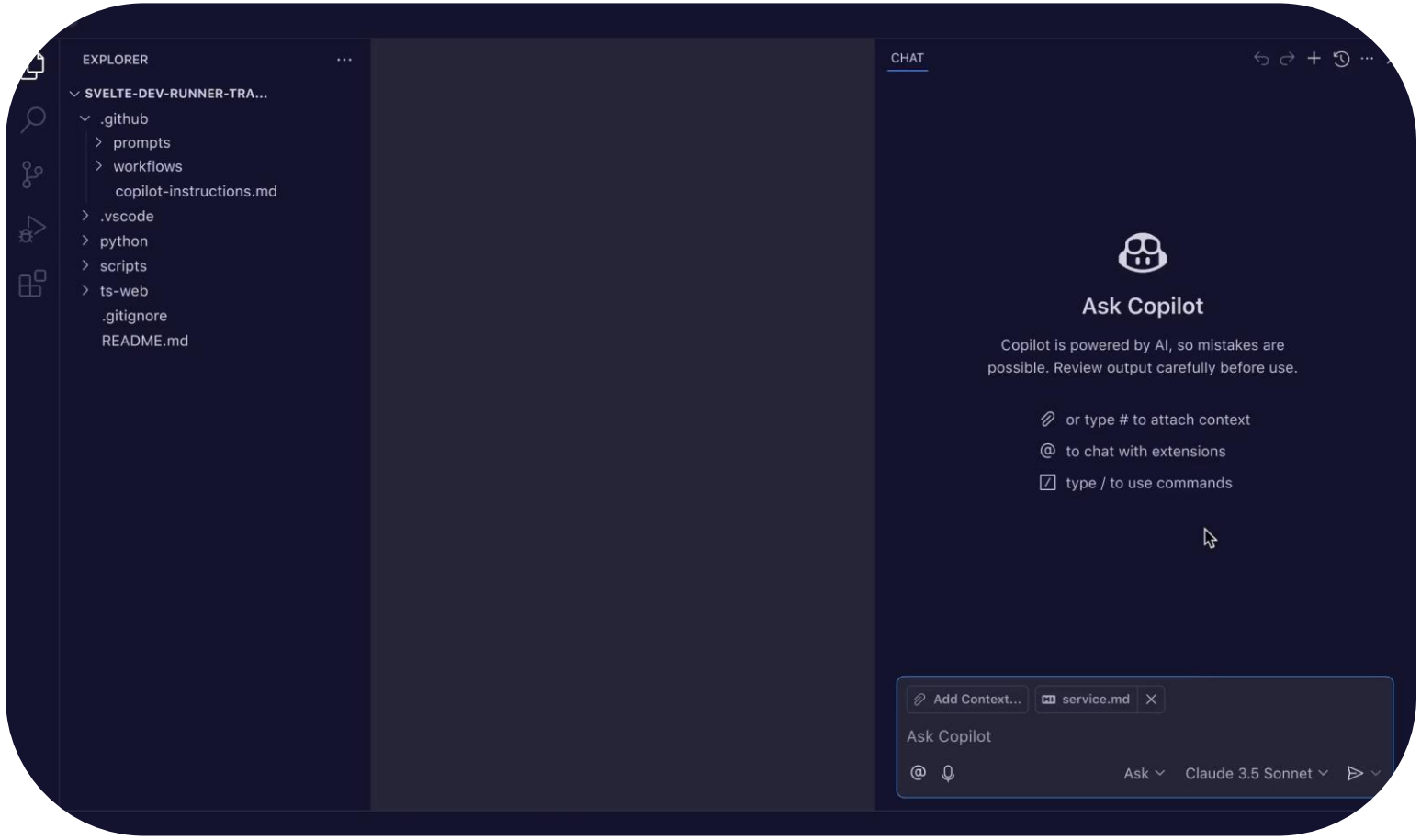
Analytic AI vs Generative AI

Created with Adobe Firefly

Generative AI for Text



Generative AI for Code



(Source: GitHub)

github.com/features/copilot

Generative AI for Visual Arts

(Source: Cosmopolitan)

First Prize in Digital arts at Colorado State Fair Fine Arts Competition

(Source: CNN Business)

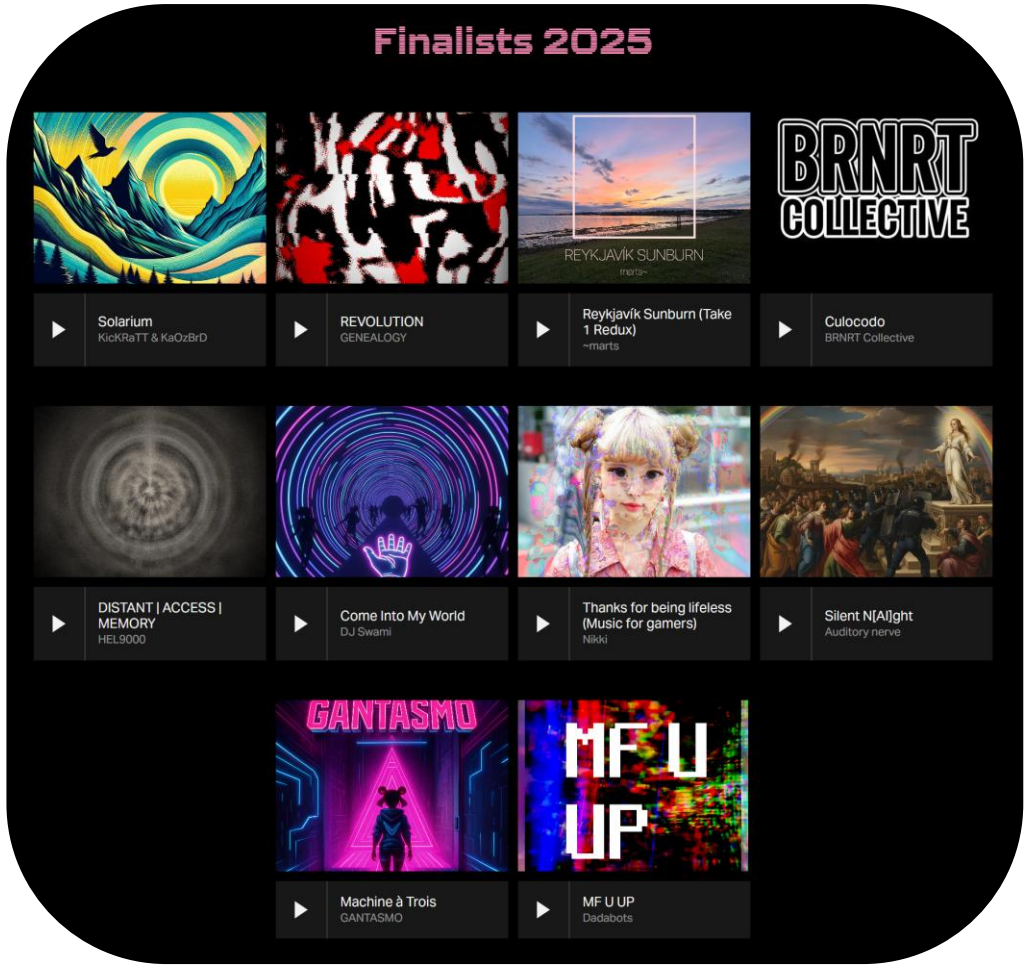
(Source: MoMA Magazine)

Sony World Photography Award in Creative Open Category

(Source: CNN)

Gloria Liu, "The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover," *Cosmopolitan*, June 21, 2022.
Rachel Metz, "AI won an art contest, and artists are furious," *CNN Business*, September 3, 2022.
Refik Anadol , "Refik Anadol on AI, Algorithms, and the Machine as Witness," *MoMA Magazine*, December 20, 2022.
Lianne Kolirin, "Artist rejects photo prize after AI-generated image wins award," *CNN*, April 18, 2023.

Generative AI for Music



(Source: AI Song Contest)

(Source: AI Song Contest)

aisongcontest.com

Entering Demons & Gods by Yaboi Hanoi (2022)

Winner of the
AI Song Contest 2022

youtu.be/PbrRoR3nEVw

soundcloud.com/yaboihanoi/enter-demons-and-gods

สวัสดี
ญาบอย ฮานอย นะครัชช

Generative AI for Videos

(Source: OpenAI)

(Source: OpenAI)

OpenAI, Sora: Creating video from text, openai.com/sora, Feb 15, 2024.

Generative AI for Videos with Music

Video **Runway Gen-2**

Music **MusicGen**

進捗共有チャンネル, "Runway Gen-2 Sample / Tiger Whiskey," *YouTube*, July 9, 2023.
Kaoru Naito, *Twitter*, twitter.com/ka0ru_1620/status/1678313226453520385, July 10, 2023.

This is NOT Morgan Freeman (2021)

youtu.be/oxXpB9pSETo

This is NOT Morgan Freeman (2021)

youtu.be/F4G6GNFz0O8

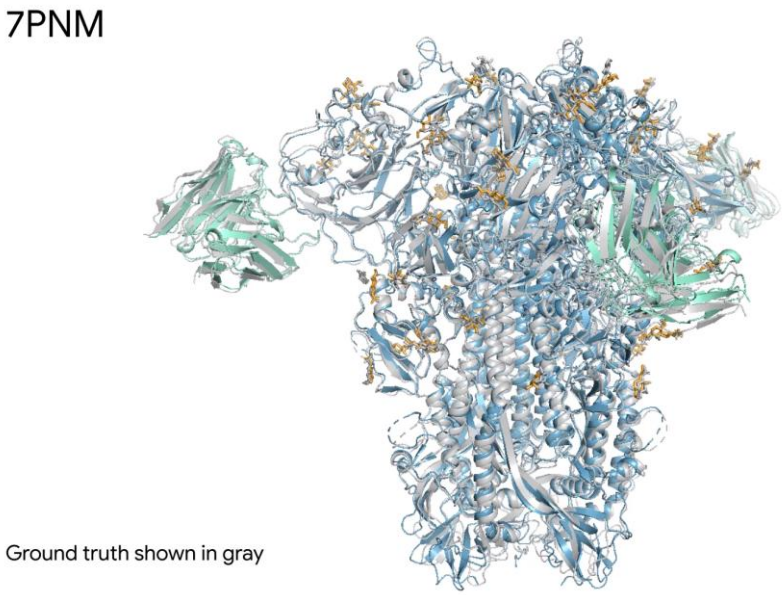
Generative AI for World Models

(Source: Google Deepmind)

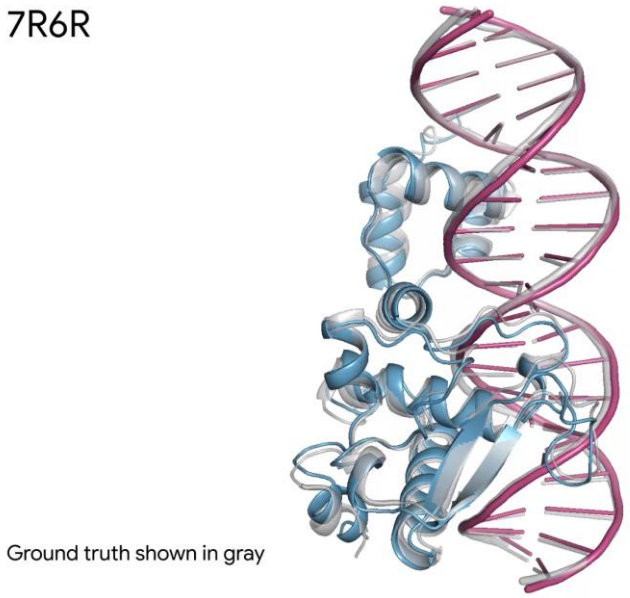
(Source: Google Deepmind)

deepmind.google/blog/genie-3-a-new-frontier-for-world-models/

Generative AI for Molecules



(Source: Google Deepmind)



(Source: Google Deepmind)

Google DeepMind AlphaFold Team and Isomorphic Labs, "AlphaFold 3 predicts the structure and interactions of all of life's molecules," *Google Blog - The Keyword*, 2024.

Generative AI for Rain

(Source: Ravuri et al., 2021)

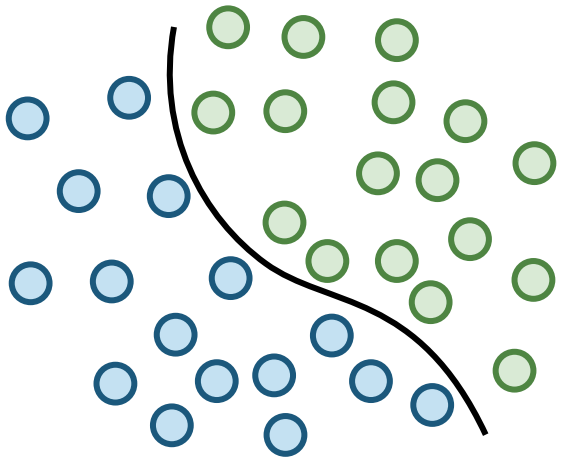
The Nowcasting Team, "Nowcasting the next hour of rain," *DeepMind Blog*, September 29, 2021.
Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed, "Skillful Precipitation Nowcasting using Deep Generative Models of Radar," *Nature*, 597:672–677, 2021.

Key Properties of Generative AI

- **One-to-many mapping**
  - There are many possible outputs given one input
- **The idea of "ground truth" becomes ambiguous**
  - Each training sample represents only one of the many possibilities
- **Some outputs are more likely than others**
  - Hence trainable through probabilistic modeling

Discriminative vs Generative Models

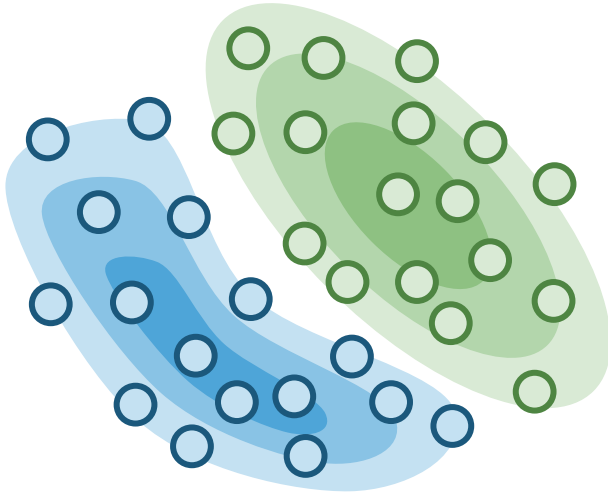
Discriminative



Discriminative models learn the decision boundary

$P(y|x)$

Generative

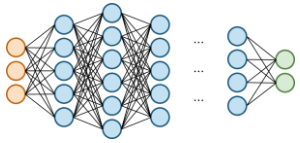


Generative models learn the underlying distribution

$P(x)$ or $P(x|y)$

Unconditional vs. Conditional Generative Models

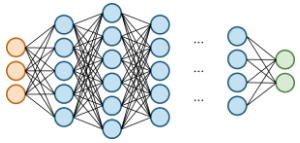
Unconditional



Sample

Conditional

Condition

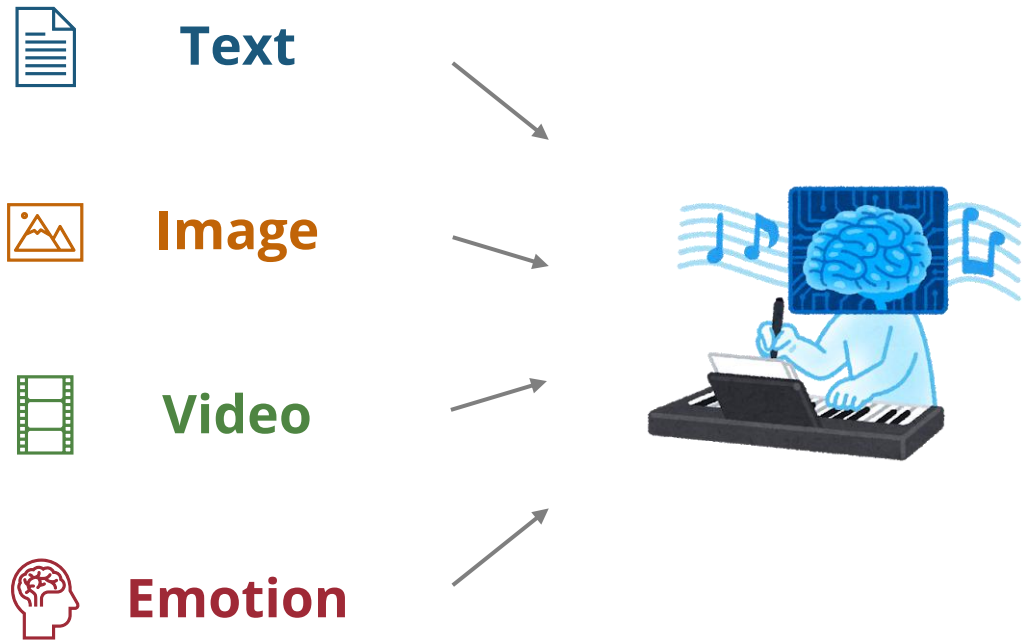


Sample

A Michigan space wolverine

nytimes.com/interactive/2024/01/19/technology/artificial-intelligence-image-generators-faces-quiz.html
Generated by ImageFX

Conditions during Training / Controls during Inference



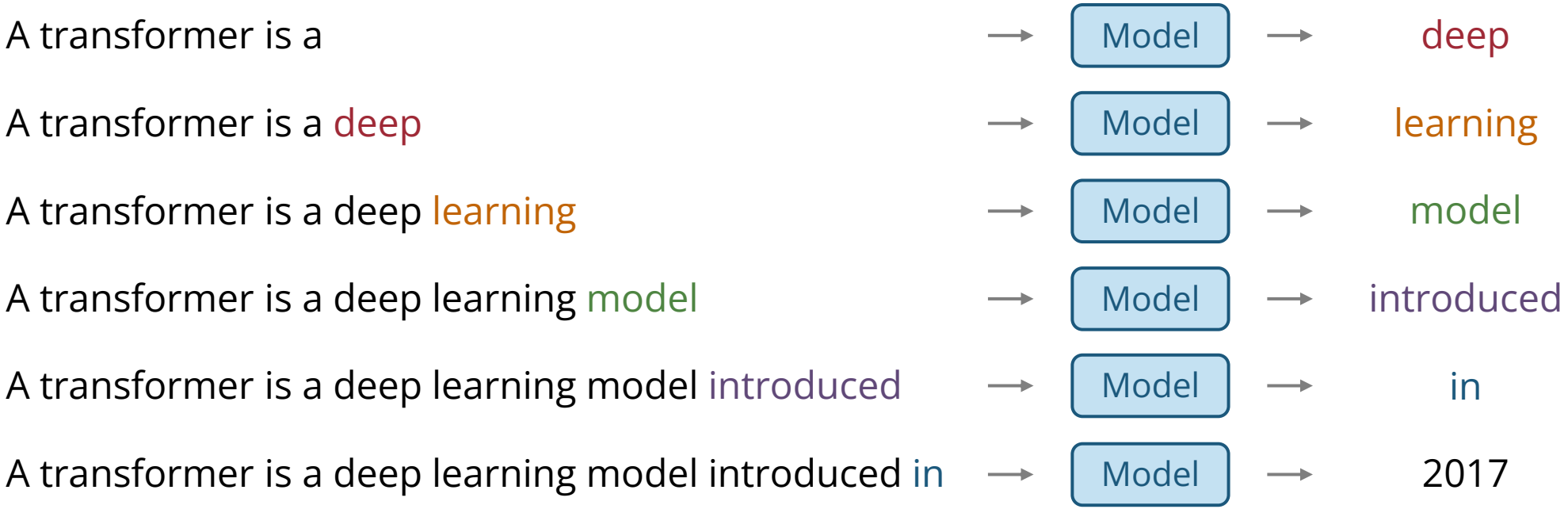
Deep Generative Models

Representative Types of Deep Generative Models

- **Deep autoregressive models**
  - Recurrent neural network (RNN)
  - Long short-term memory (LSTM)
  - Transformer model
- **Deep latent variable models**
  - Generative adversarial network (GAN)
  - Variational autoencoder (VAE)
  - Diffusion model
  - Flow-based model
- *And many others…*

Deep Autoregressive Models

- **Intuition**: Decompose the generation of a sequence into generating one item after another



Deep Autoregressive Models

- **Intuition**: Decompose the generation of a sequence into generating one item after another

$$P(\ x_i \mid \underbrace{x_1, x_2, \ldots, x_{i-1}}\ )$$

Next word — Previous words

$P(\text{ electrical} \mid \text{A transformer is a})$ ↑

$P(\text{ character} \mid \text{A transformer is a})$ ↑

$P(\quad \text{gene} \quad \mid \text{A transformer is a})$ ↑

$P(\quad \text{model} \quad \mid \text{A transformer is a})$ ↑

$P(\quad \text{food} \quad \mid \text{A transformer is a})$ ↓

$P(\quad \text{musical} \quad \mid \text{A transformer is a})$ ↓

Deep Autoregressive Models

- **Intuition**: Decompose the generation of a sequence into generating one item after another

$$P(\, x_i \mid \underbrace{x_1, x_2, \ldots, x_{i-1}}\,)$$

Next word — Previous words

The whole sentence

$$X = (x_0, x_1, \ldots, x_N)$$

2nd word given 1st word — Last word given all previous words

$$P(X) = P(x_0)\; P(x_1 \mid x_0)\; P(x_2 \mid x_0, x_1)\; \ldots\; P(\, x_N \mid x_1, x_2, \ldots, x_{N-1}\,)$$

1st word — 3rd word given 1st & 2nd words

Deep Autoregressive Models

- **Intuition**: Decompose the generation of a sequence into generating one item after another

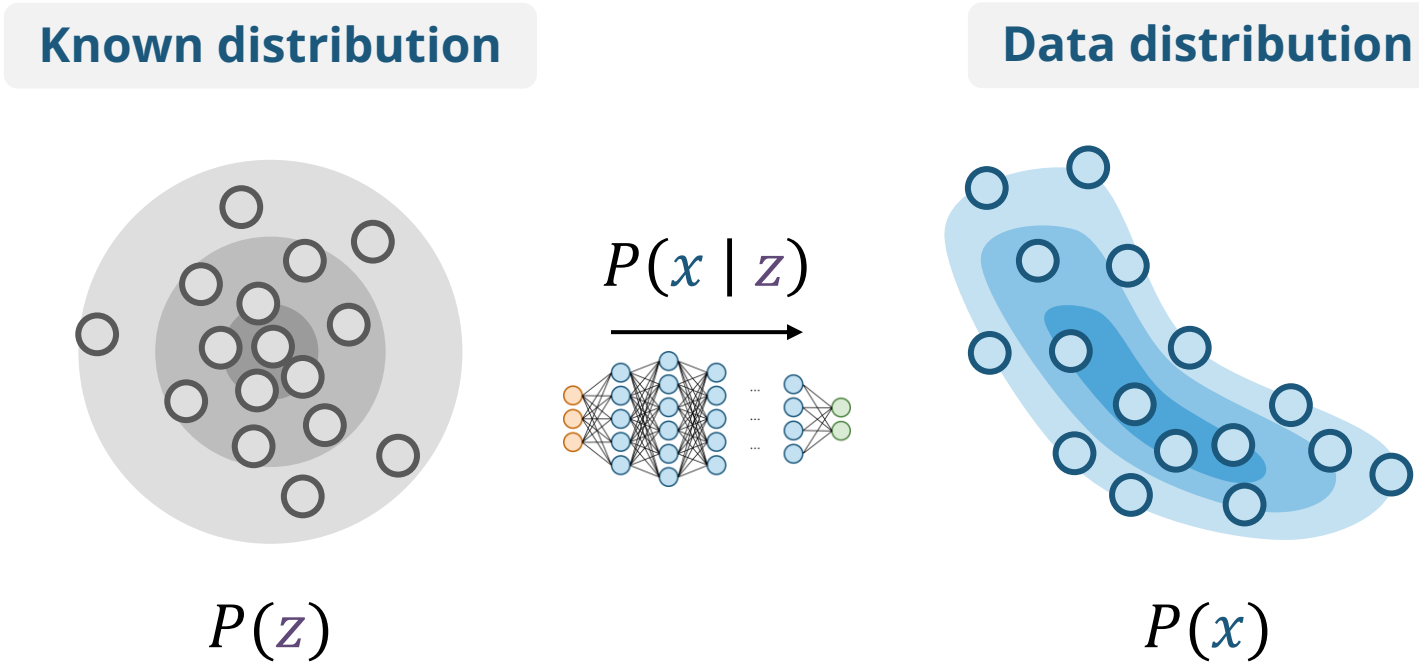
What we want the model to learn!

$$P(X) = P(x_0)\,P(x_1 \mid x_0)\,P(x_2 \mid x_0, x_1) \ldots P(x_N \mid x_1, x_2, \ldots, x_{N-1})$$

$$= P(x_0) \prod_{i=1}^{N} P(x_i \mid x_1, x_2, \ldots, x_{i-1})$$

Deep Latent Variable Models

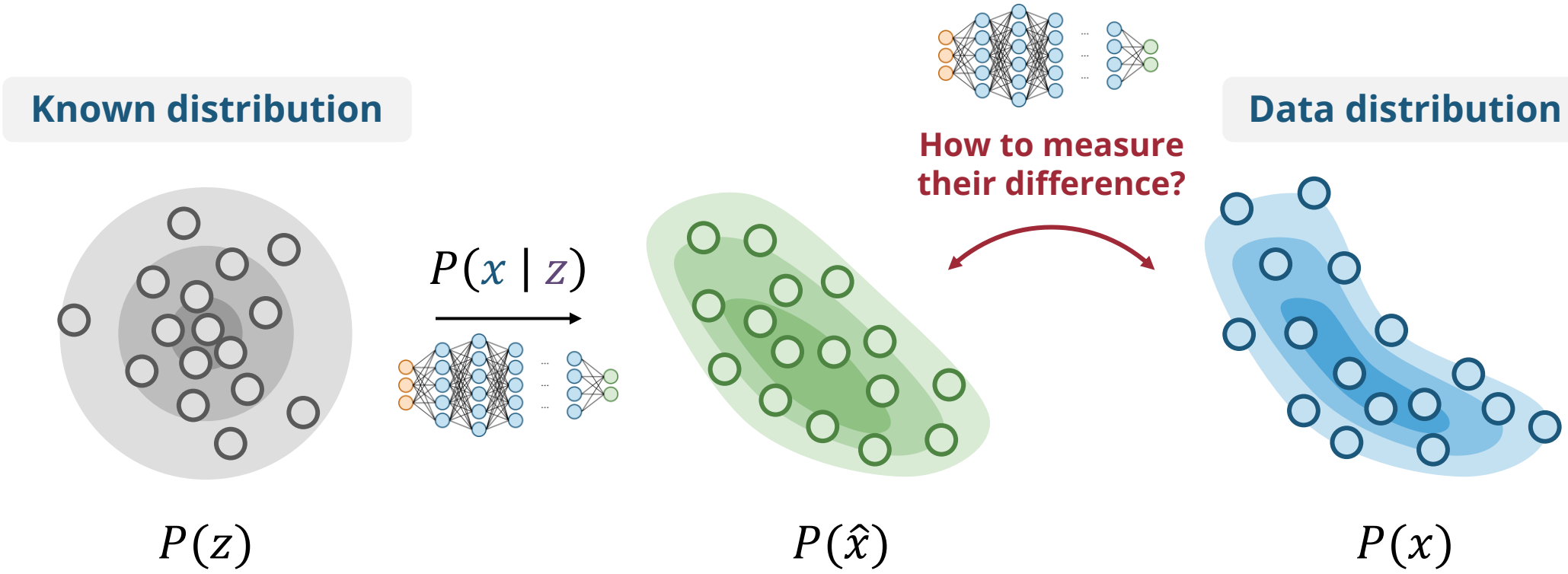
- **Intuition**: Learn to map a known distribution to the data distribution



$$P(x) = P(z)\ P(x \mid z)$$

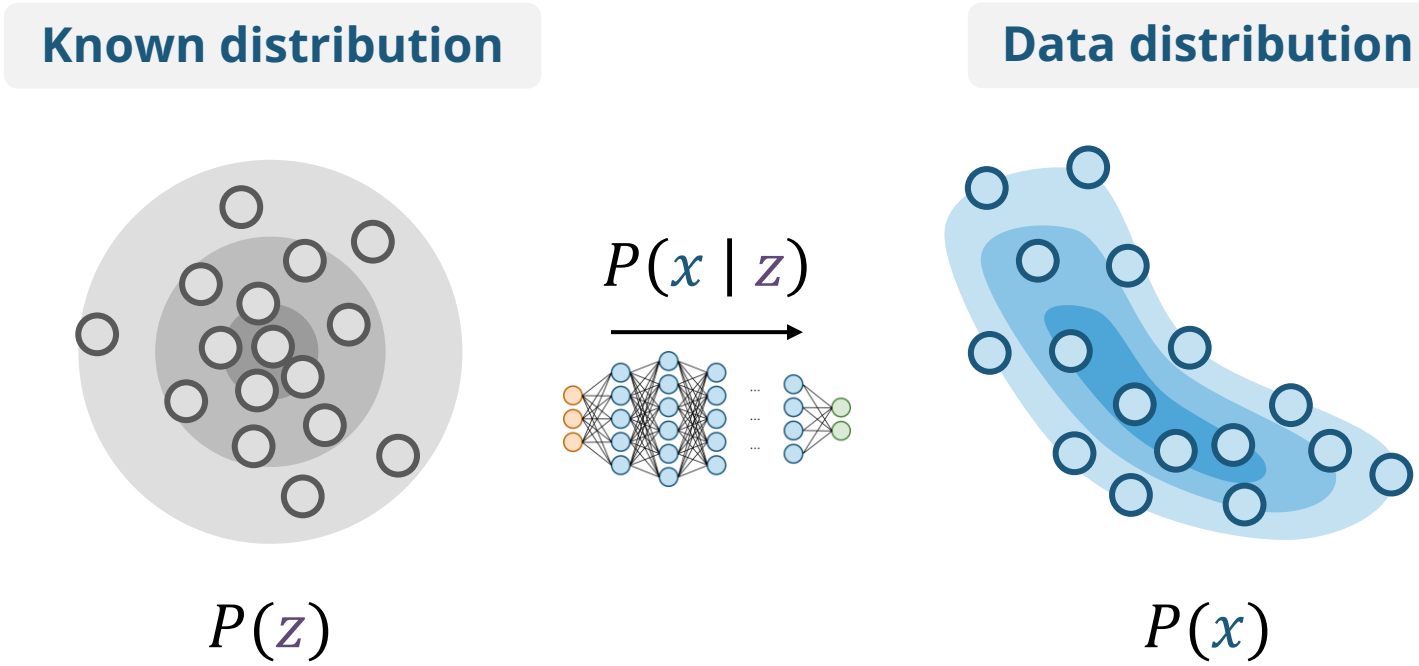
Deep Latent Variable Models

- **Intuition**: Learn to map a known distribution to the data distribution



Deep Latent Variable Models

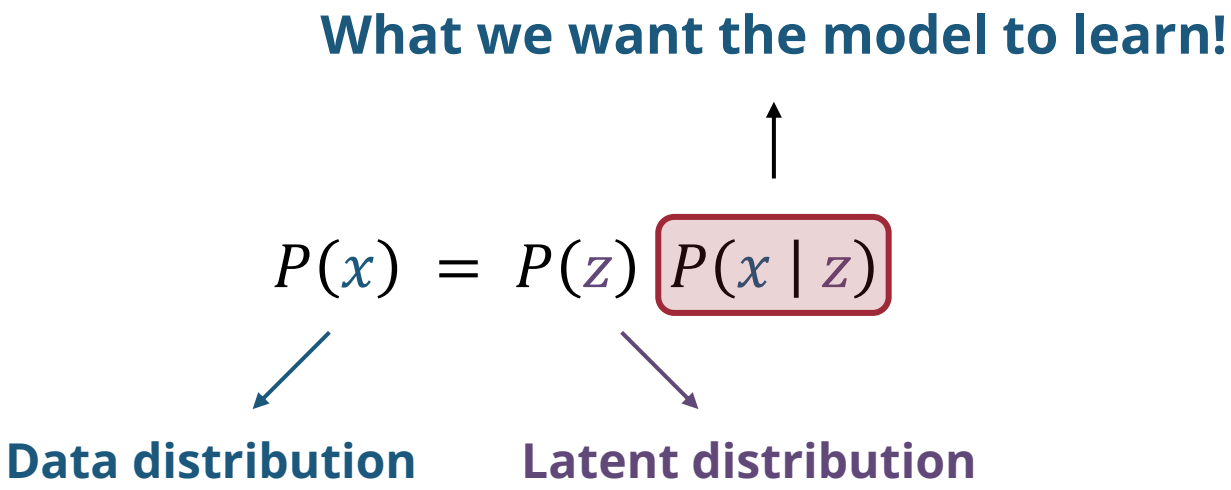
- **Intuition**: Learn to map a known distribution to the data distribution



$$P(x) = P(z)\ P(x \mid z)$$

Deep Latent Variable Models

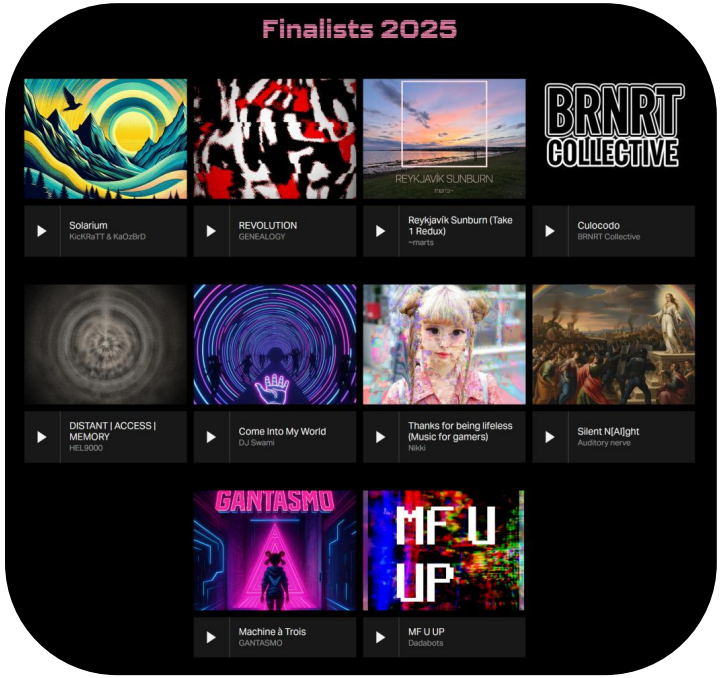
- **Intuition**: Learn to map a known distribution to the data distribution



AI Song Contest

AI Song Contest

- Annual international competition showcasing the **creative potential of human–AI co-creativity in the songwriting process**



(Source: AI Song Contest)

(Source: AI Song Contest)

aisongcontest.com

Entering Demons & Gods by Yaboi Hanoi (2022)

Winner of the
AI Song Contest 2022

youtu.be/PbrRoR3nEVw

soundcloud.com/yaboihanoi/enter-demons-and-gods

Reading: The Making of Entering Demons & Gods (2022)

"**It was like a saxophonist trained in classical Thai motifs, who played a special 'Thai Edition' saxophone with Phi Nai tunings, had joined the musical conversation.** The same was true with the trumpet model and the ขลุ่ย 'Khlui' - a flute from Thai, Laos and Cambodian repertoire. I could assemble a **transcultural ensemble** to expand the sonic palette of Thai motifs, whilst adhering to underlying tunings and idiomatic inflections like never before."

lamtharnhantrakul.github.io/enter-demons-and-gods/

How would you touch me? by Synthetic Beat Brigade (2023)

youtu.be/O4cJ3acEGDw

Winner of the AI Song Contest 2023

Reading: The Making of How would you touch me? (2023)

"This project is a collaboration between **Artificial Intelligence (AI) enthusiasts in four fields: artist management, music and post-production, tech, and creative**. In contrast, the majority of the music industry sees AI as a threat. Our team understands that these technological advances will have a significant impact on how we produce music. Because of this, we have decided to **use AI for every step of the production process**. From ideation to creating the lyrics to producing the music."

drive.google.com/file/d/1QTQ7P3iZI6I0anIwNQ3ewf8g3JjDjesl/view

Reading: The Making of How would you touch me? (2023)

- **Ideation**: Spotify API, ChatGPT, Facebook Llama, Google Bison
- **Lyrics**: ChatGPT 2, Genius API
- **Composition**: AI Drummachine, Mofi, Tonetrasnfer, This patch does not exist, Albeatz, BaiscPitch, Magenta, AIVA, MuseNet
- **Vocals**: Soundly Voice Designer, Vocal Remove, Voice characteristics
- **Mastering**: Landr
- **Cover art & bandart**: Midjourney
- **Clip**: ComfyUI for Stable Diffusion + ControlNet

drive.google.com/file/d/1QTQ7P3iZI6I0anIwNQ3ewf8g3JjDjesl/view

Sudamérica by Onda Corta (2024)

Winner of the AI Song Contest 2024

youtu.be/OQcpItDUuik

SUDAMÉRICA - ONDA CORTA

Reading: The Making of Sudamérica (2024)

“To create our song, **we integrated AI tools at every stage of the creative process**. We started with three different language models (**ChatGPT**, **Claude** and **Gemini**) to generate the initial concept of "Sudamérica," gathering a wide range of ideas. We chose these models because they are the most widely used, allowing us to **understand what a larger number of users are getting when they inquire about our region**.”

aisongcontest.com/participants-2024/onda-corta

Reading: The Making of Sudamérica (2024)

“Key recurring themes like tango, soccer, and biodiversity were selected and filtered. Using **ElevenLabs**, **we generated over 100 audio clips of 5-10 seconds based on these themes**, as well as South American country names and traditional musical styles. We chose Eleven Labs for its ethical transparency in using its database and because it allowed us to explore brief audio clips instead of complete musical creations. **This gave us greater control over the composition and allowed us to use varied sounds from prompts like "biodiversity" or "Bolivia" as sound textures in our song**.”

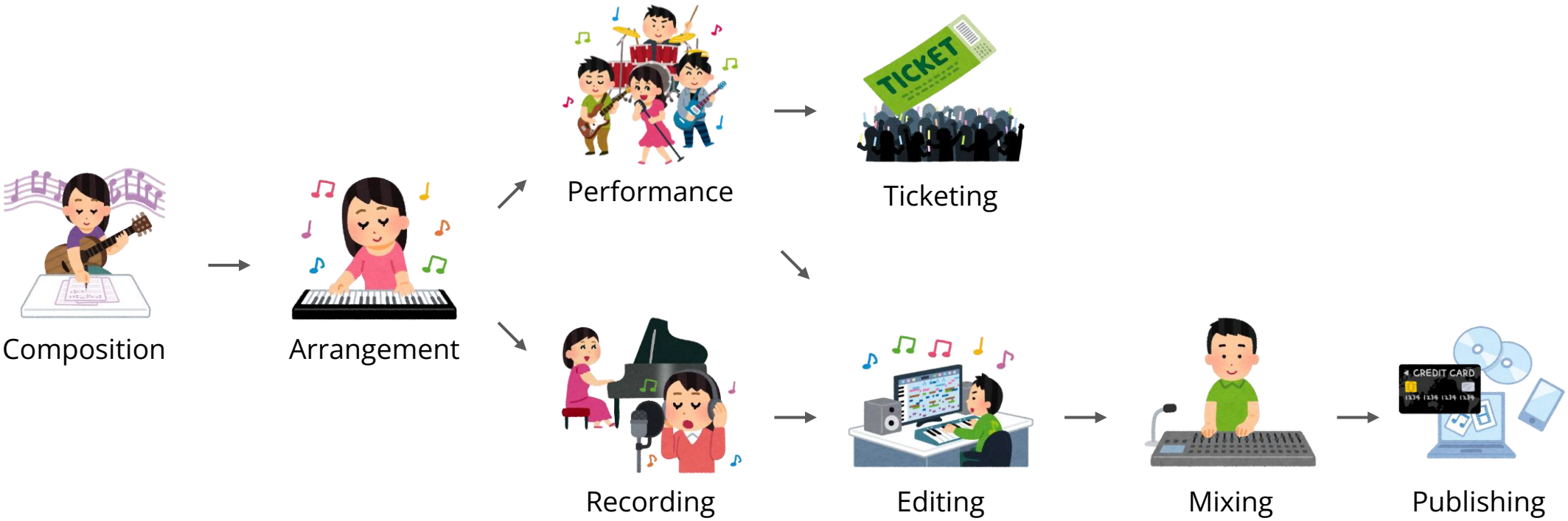
aisongcontest.com/participants-2024/onda-corta

Reading: The Making of Sudamérica (2024)

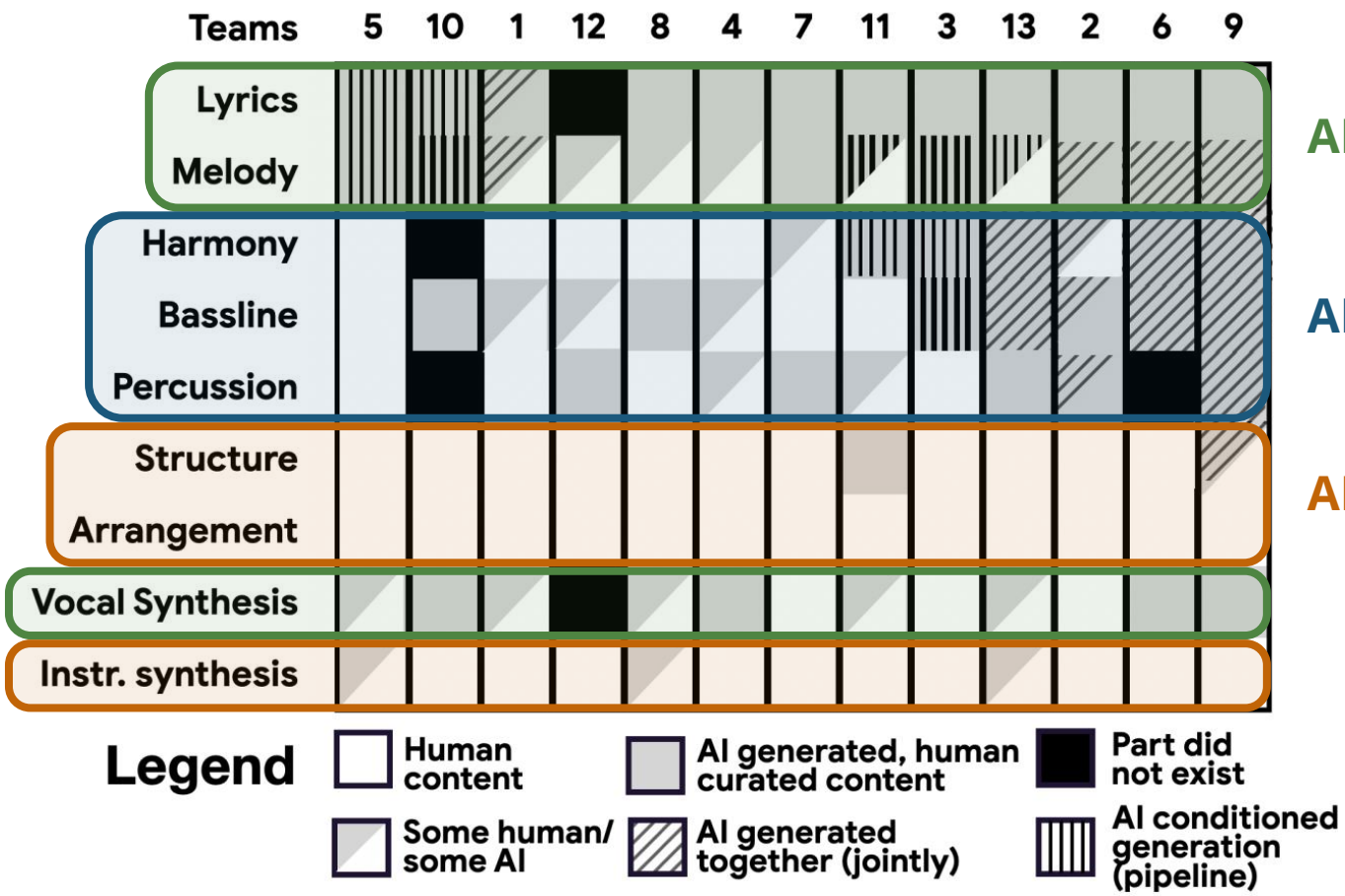
"For the lyrics, we used **Claude.ai** to generate several options based on selected concepts. We curated the best verses and choruses for thematic and stylistic coherence. Andrés selected 23 audio clips for the song, using **Moises.ai** to separate one key sample into bass, percussion, and piano tracks, which became the song's foundation. He composed the song in a DAW, looping and layering samples, and recorded vocals, further enhanced with **KITS.ai** for a unique touch."

aisongcontest.com/participants-2024/onda-corta

An Overly-Simplified Music Production Workflow



Analyzing Human-AI Music Co-creation (Huang et al., 2020)



AI is helpful

AI can be helpful

AI is not so helpful

First AI Song Contest! (in 2020)

(Source: Huang et al., 2020)

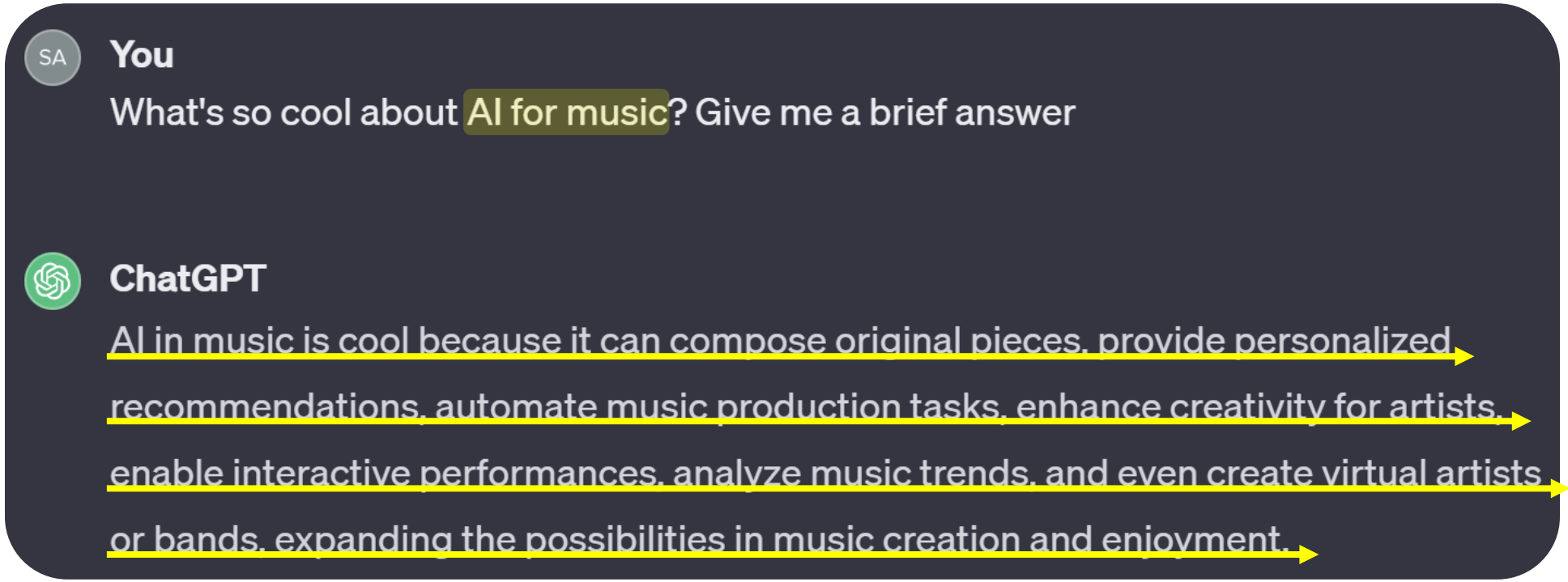
Cheng-Zhi Anna Huang, Hendrik Vincent Koops, Ed Newton-Rex, Monica Dinculescu, and Carrie J. Cai, "AI Song Contest: Human-AI Co-Creation in Songwriting," *ISMIR*, 2020.

Recap

Analytic AI vs Generative AI

Created with Adobe Firefly

Generative AI for Text



Generative AI for Visual Arts

(Source: Cosmopolitan)

First Prize in Digital arts at Colorado State Fair Fine Arts Competition

(Source: CNN Business)

(Source: MoMA Magazine)

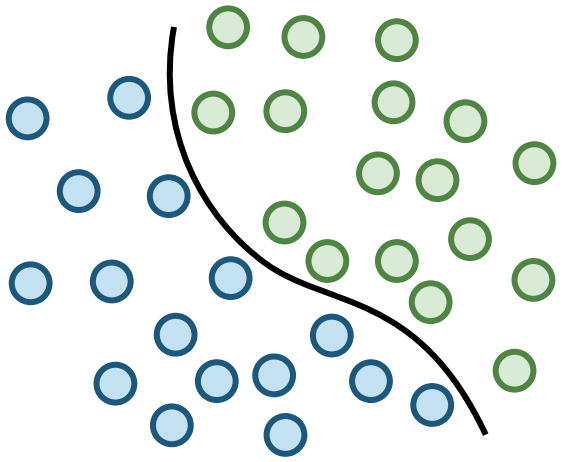
Sony World Photography Award in Creative Open Category

(Source: CNN)

Gloria Liu, "The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover," *Cosmopolitan*, June 21, 2022.
Rachel Metz, "AI won an art contest, and artists are furious," *CNN Business*, September 3, 2022.
Refik Anadol , "Refik Anadol on AI, Algorithms, and the Machine as Witness," *MoMA Magazine*, December 20, 2022.
Lianne Kolirin, "Artist rejects photo prize after AI-generated image wins award," *CNN*, April 18, 2023.

Discriminative vs Generative Models

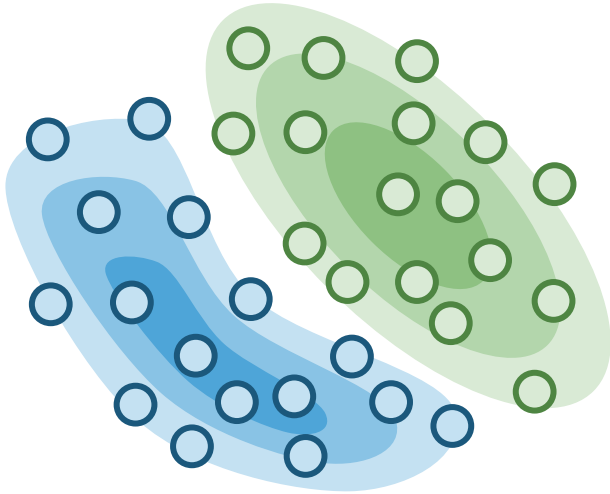
Discriminative



Discriminative models learn the decision boundary

$P(y|x)$

Generative

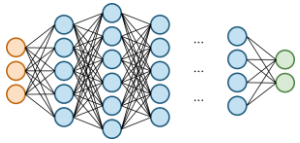


Generative models learn the underlying distribution

$P(x)$ or $P(x|y)$

Unconditional vs. Conditional Generative Models

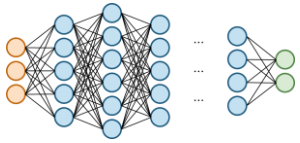
Unconditional



Sample

Conditional

Condition



Sample

A Michigan space wolverine

nytimes.com/interactive/2024/01/19/technology/artificial-intelligence-image-generators-faces-quiz.html
Generated by ImageFX

Representative Types of Deep Generative Models

- **Deep autoregressive models**
  - Recurrent neural network (RNN)
  - Long short-term memory (LSTM)
  - Transformer model
- **Deep latent variable models**
  - Generative adversarial network (GAN)
  - Variational autoencoder (VAE)
  - Diffusion model
  - Flow-based model
- *And many others…*

Deep Autoregressive Models

- **Intuition**: Decompose the generation of a sequence into generating one item after another

$$P(\ x_i \mid \underbrace{x_1, x_2, \ldots, x_{i-1}}\ )$$

Next word — Previous words

$P(\ \text{electrical} \mid \text{A transformer is a}\ )$ ↑

$P(\ \text{character} \mid \text{A transformer is a}\ )$ ↑

$P(\ \text{gene} \mid \text{A transformer is a}\ )$ ↑

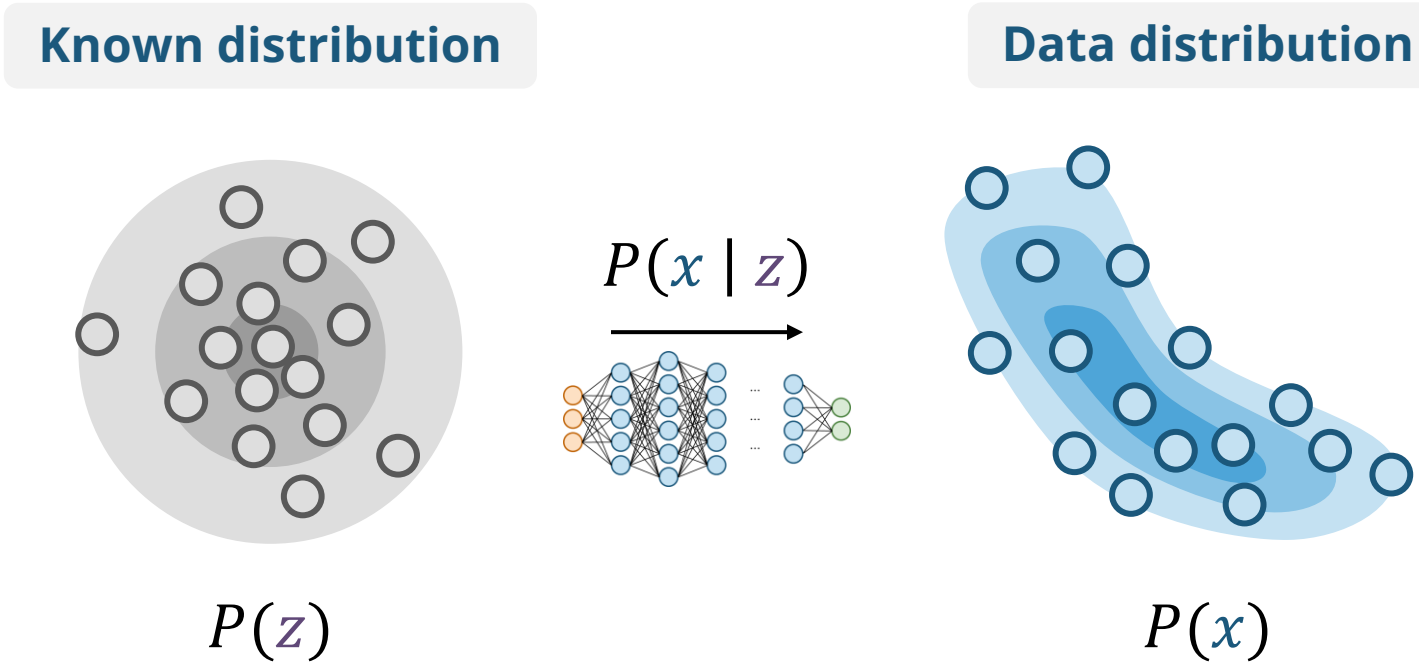
$P(\ \text{model} \mid \text{A transformer is a}\ )$ ↑

$P(\ \text{food} \mid \text{A transformer is a}\ )$ ↓

$P(\ \text{musical} \mid \text{A transformer is a}\ )$ ↓

Deep Latent Variable Models

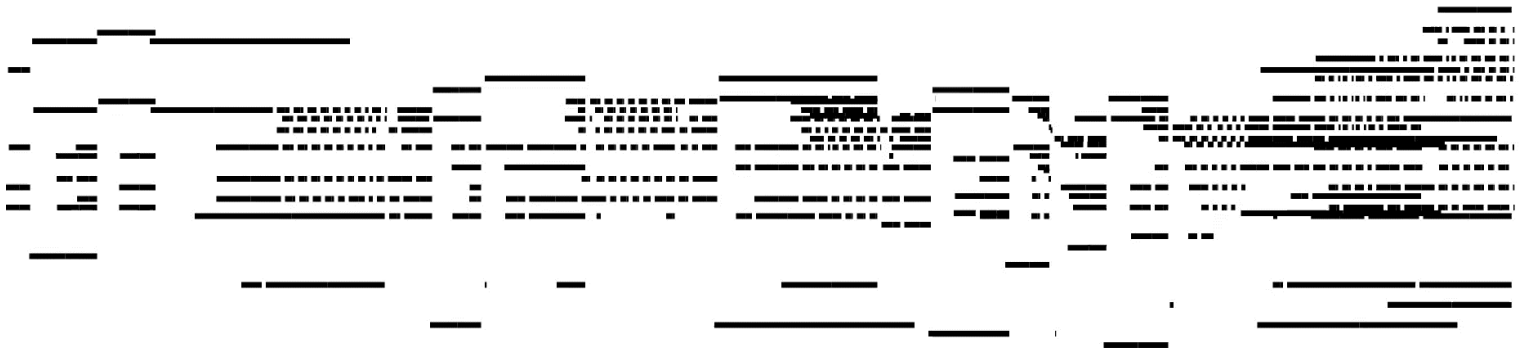
- **Intuition**: Learn to map a known distribution to the data distribution



$$P(x) = P(z)\ P(x \mid z)$$

Next Lecture

Language-based Music Generation



(Source: Huang et al., 2018)