PAT 498/598 (Fall 2024)

Special Topics: Generative AI for Music and Audio Creation

Lecture 3: Intro to AI Music II

Instructor: Hao-Wen Dong

SCHOOL OF MUSIC, THEATRE & DANCE
PERFORMING ARTS TECHNOLOGY
UNIVERSITY OF MICHIGAN

Tentative Schedule

Generative AI Background

Week	Date	Lecture
1	Aug 26	Introduction
		Background
	Aug 28	├ AI & machine learning fundamentals
2	~~Sep 2~~	├ ~~No Class (Labor Day)~~
	Sep 4	├ Deep learning fundamentals I
3	Sep 9	├ Deep learning fundamentals II
	Sep 11	├ Language models – RNNs, LSTMs & transformers
4	Sep 16	├ Generative adversarial nets & diffusion models
	Sep 18	└ Music & audio processing fundamentals

Symbolic Music Generation

Week	Date	Lecture
		Symbolic Music Generation
5	Sep 23	├ Melody generation
	Sep 25	├ Harmony & chord progression generation
6	Sep 30	├ Polyphonic music generation
	Oct 2	├ Multitrack music generation
7	Oct 7	├ Multimodal music generation I
	Oct 9	└ Multimodal music generation II
8	~~Oct 14~~	~~No Class (Fall Study Break)~~

Audio Synthesis

Week	Date	Lecture
		Audio Synthesis
	Oct 16	├ Time-domain audio synthesis I
9	Oct 21	├ Time-domain audio synthesis II
	Oct 23	├ Frequency-domain audio synthesis I
10	Oct 28	├ Frequency-domain audio synthesis II
	Oct 30	├ Multimodal audio synthesis I
11	Nov 4	└ Multimodal audio synthesis II
	Nov 6	Project pitch & discussion
12	~~Nov 11~~	~~No Class (Travel)~~
	~~Nov 13~~	~~No Class (Travel)~~

Assistive Music Creation Tools

Week	Date	Lecture
		Assistive Music Creation Tools
13	Nov 18	├ Neural audio effects
	Nov 20	├ Auto-mixing
14	Nov 25	└ Live performance & interactive systems
	~~Nov 27~~	~~No Class (Thanksgiving)~~
15	Dec 2	Discussions — ethical concerns & copyright issues
	Dec 4	Review
16	Dec 9	**Project presentation**

Symbolic Music Generation

(Recap) Piano Genie (2018)

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

piano-genie.glitch.me/

Chris Donahue, Ian Simon, Sander Dieleman, "Piano Genie," IUI, 2019.

(Recap) Fruit Genie Live (2019)

youtu.be/L4wvXrPmIkU & magenta.tensorflow.org/fruitgenie

Audio Synthesis

(Recap) Tone Transfer (2020)

youtu.be/bXBliLjImio & magenta.tensorflow.org/tone-transfer

sites.research.google/tonetransfer

(Recap) Yaboi Hanoi – Entering Demons & Gods (2022)

https://youtu.be/PbrRoR3nEVw

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

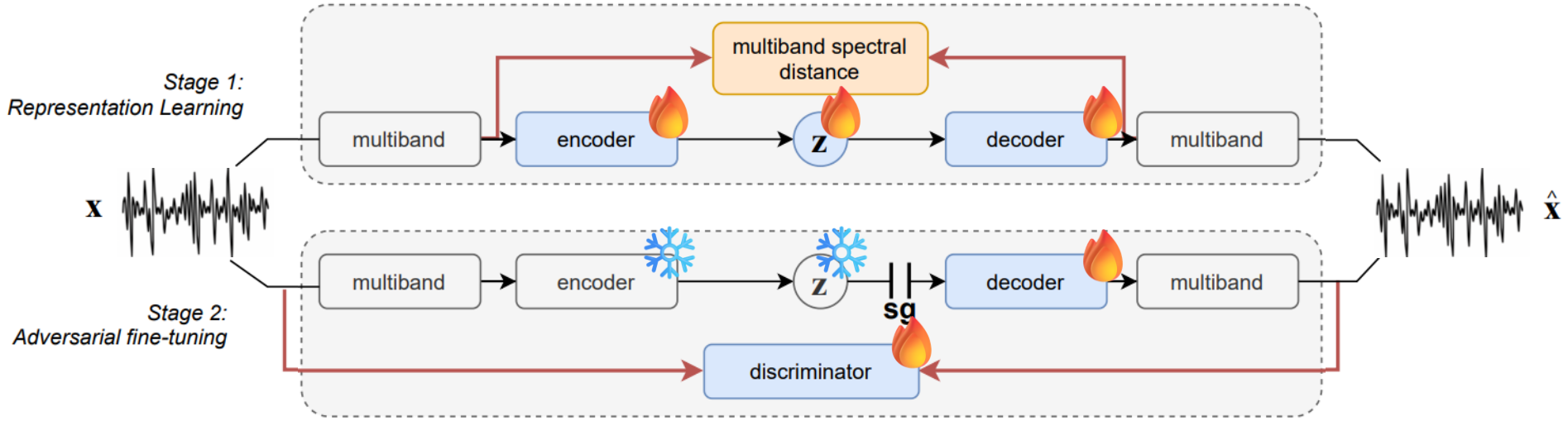
Realtime Neural Audio Synthesis – RAVE (2022)

youtu.be/dMZs04TzxUI

github.com/acids-ircam/RAVE

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

Realtime Neural Audio Synthesis – RAVE (2022)



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

Realtime Neural Audio Synthesis – RAVE (2022)

youtu.be/jAlRf4nGgYI

github.com/acids-ircam/RAVE

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

Assistive Music Creation Tools

Assistive Music Creation Tools

- Any tools used in the music creation pipeline
- Some tasks are well-developed (e.g., auto-tune)
- Counterintuitively, **assistive tools are often time more difficult to build**
  - For it requires **FIRST understanding the input** and **THEN perform the task**

Topics We’ll Cover

- Neural audio effects
- Auto-mixing
- Live performance & interactive systems

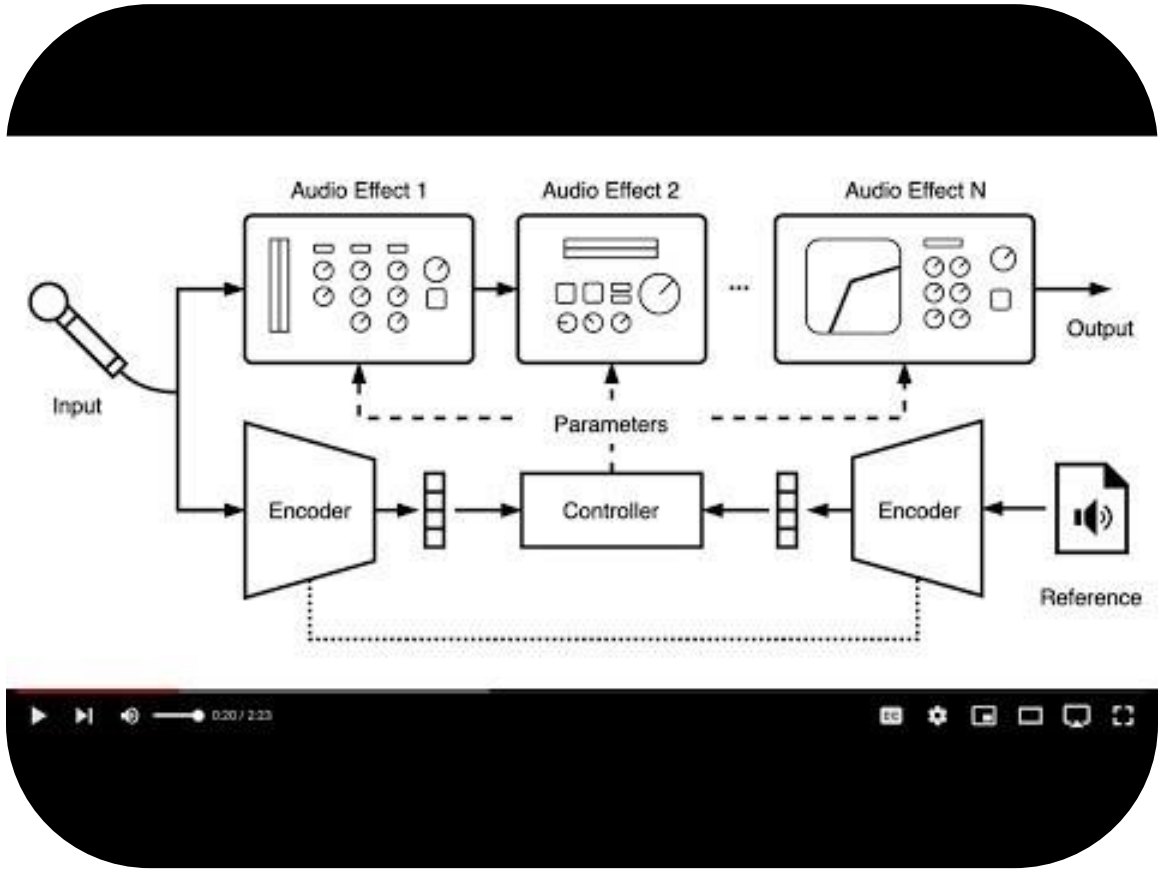
Neural Audio Effects (2021)

youtu.be/Zmo8kB-SfF4

colab.research.google.com/github/csteinmetz1/steerable-nafx/blob/master/steerable-nafx.ipynb

Christian J. Steinmetz and Joshua D. Reiss, "Steerable discovery of neural audio effects," *NeurIPS ML4CD Workshop*, 2021.

DeepAFx-ST: Style Transfer of Audio Effects (2022)



youtu.be/lZp455wiMk4?t=100

Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

DeepAFx-ST: Style Transfer of Audio Effects (2022)

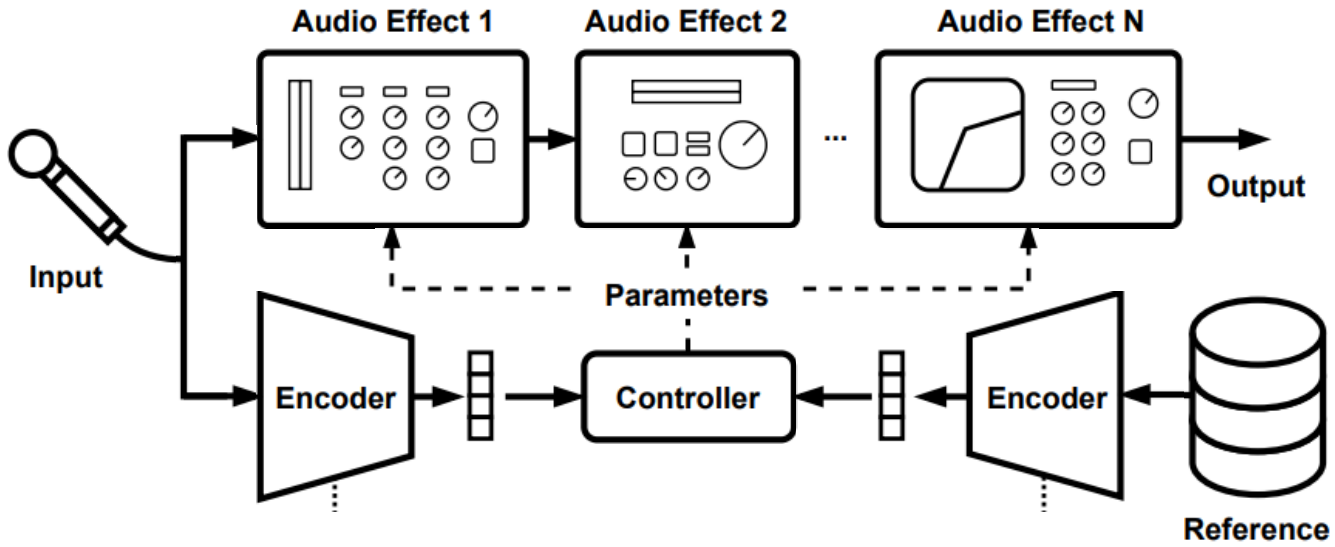


Fig. 1. Our *DeepAFx-ST* method imposes the audio effects and production style from one recording to another by example. We use a shared-weight encoder to analyze the input and a style reference signal, then compare each with a controller that outputs the parameters of effects that themselves perform style manipulation.

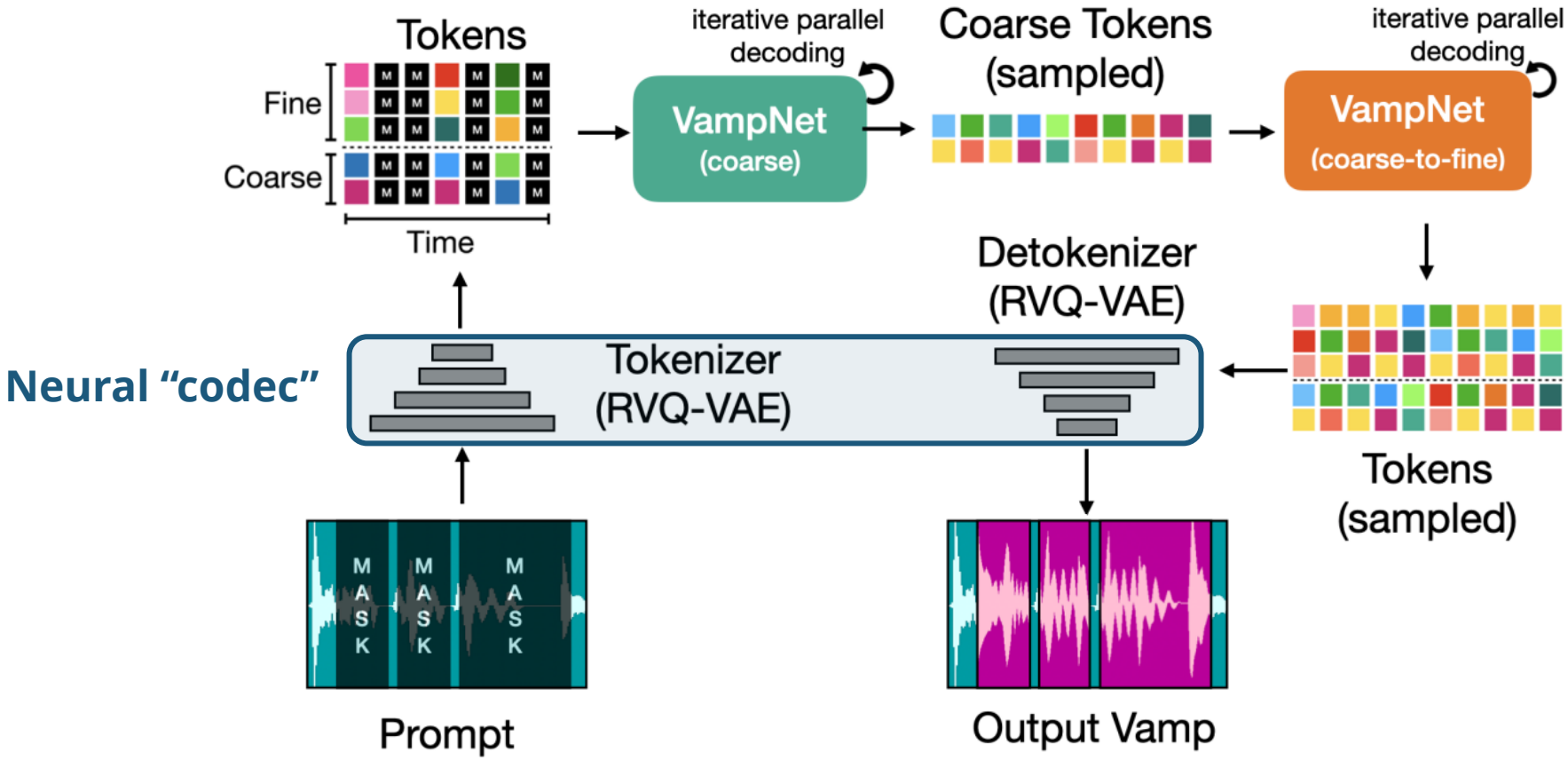
Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

unloop: a looper that doesn't repeat itself (2023)

youtu.be/yzBI8Vcjd2s

github.com/hugoflores garcia/unloop

VampNet (2023)



Hugo Flores Garcia, Prem Seetharaman, Rithesh Kumar, and Bryan Pardo, "VampNet: Music Generation via Masked Acoustic Token Modeling," *ISMIR*, 2023.

AI Creative Agents (2015)

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

Somax 2 (2019)

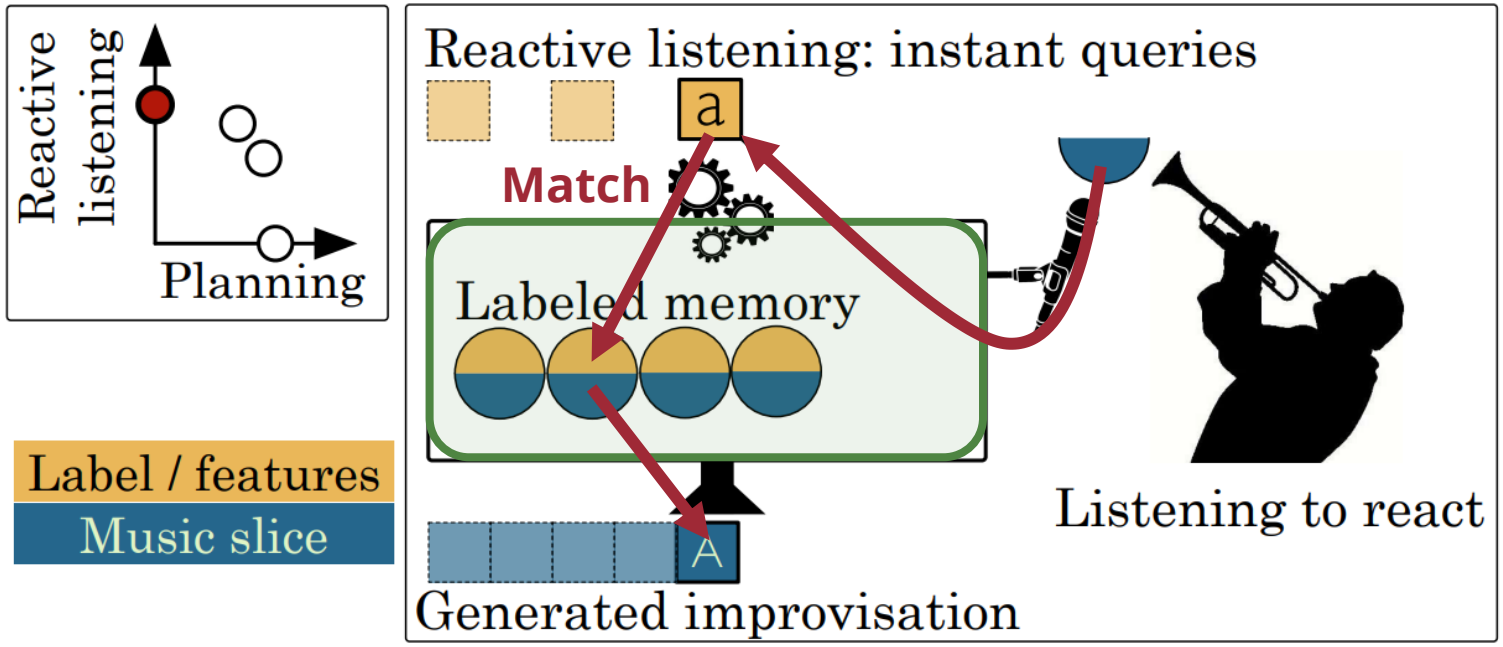


Figure 1. Somax: music generation guided by reactive listening.

Jérôme Nika, Ken Déguernel, Axel Chemla–Romeu-Santos, Emmanuel Vincent, and Gérard Assayag, "DYCI2 agents: merging the "free", "reactive", and "scenario-based" music generation paradigms," *ICMC*, 2017.

ImproteK (2017)

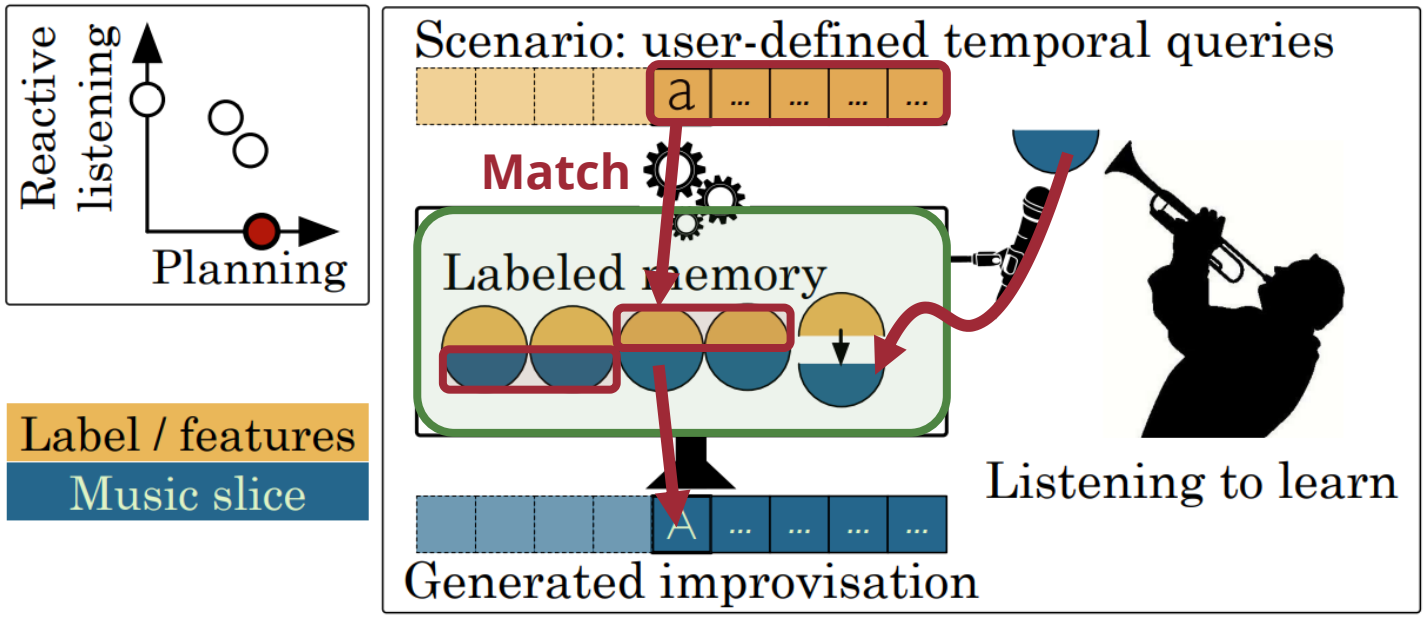


Figure 2. ImproteK: music generation guided by a temporal scenario.

Jérôme Nika, Ken Déguernel, Axel Chemla–Romeu-Santos, Emmanuel Vincent, and Gérard Assayag, "DYCI2 agents: merging the "free", "reactive", and "scenario-based" music generation paradigms," *ICMC*, 2017.

ImproteK (2017)

For the scenario C A B B C C B A:

Matching both the history of the memory and the future of the scenario

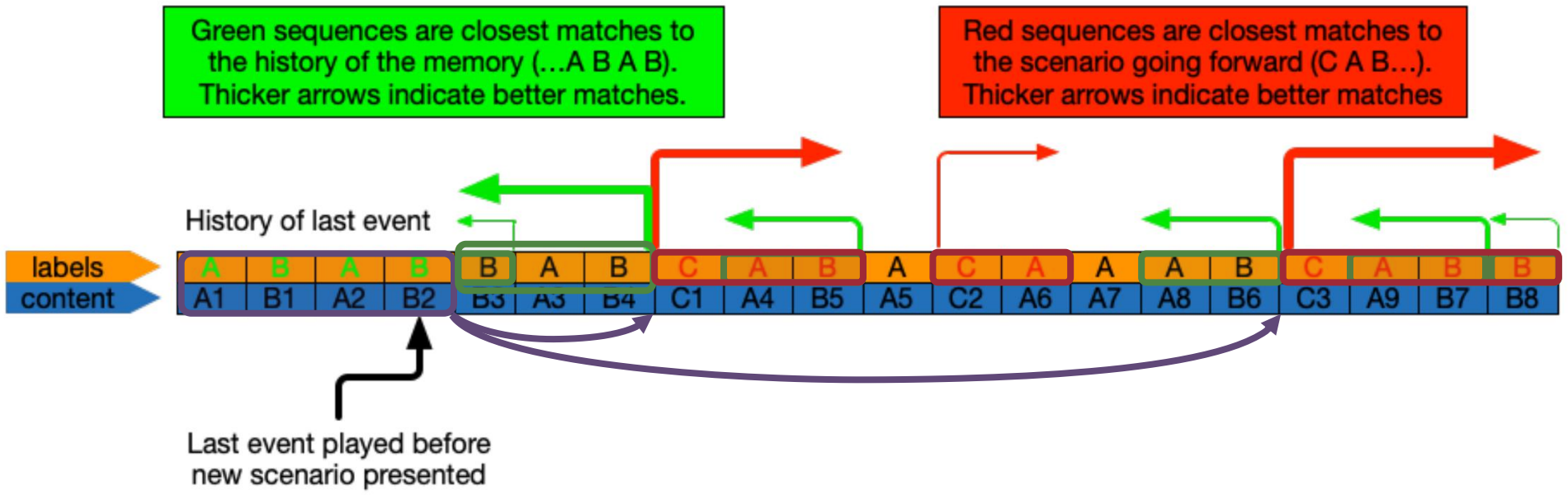


Figure 4: How an Agent responds to a Scenario

Jérôme Nika, Augustin Muller, Joakim Borg, Gérard Assayag, and Matthew Ostrowski, "Dicy2 for Live," *Ircam*, 2022.

Human-AI Music Co-creation

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

youtu.be/O4cJ3acEGDw & bit.ly/45vlmuT

Synthetic Beat Brigade - 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

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

AI Song Contest

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

aisongcontest.com

(Part of) Assignment 1

- Please listen to the **ten finalists of AI Song Contest 2024** and **read the about pages** by clicking the cover arts
- **Vote for your favorites**
- Answer the following questions briefly (5-10 sentences each):
  - Which is your favorite song? What did they do well? What can be improved?
  - What is one dimension that most finalists didn't look into or didn't do well on?
  - What tasks are easy for current AI? What are difficult?

Intro to Machine Learning

(Recap) What is Artificial Intelligence?

AI is the study of how to make computers **do things at which, at the moment, people are better**.

– Elaine Rich and Kevin Knight, 1991

1997

(Source: Britannica)

2016

(Source: The Guardian)

20??

(Source: SC2HL)

Elaine Rich and Kevin Knight, *Artificial Intelligence.* United Kingdom: McGraw-Hill, 1991.
https://www.britannica.com/topic/Deep-Blue
https://www.theguardian.com/technology/2016/mar/15/alphago-what-does-google-advanced-software-go-next
https://www.youtube.com/watch?v=PFMRDm_H9Sg

What is Machine Learning?

Machine Learning is a field of study that gives computers the ability to **learn without being explicitly programmed**.

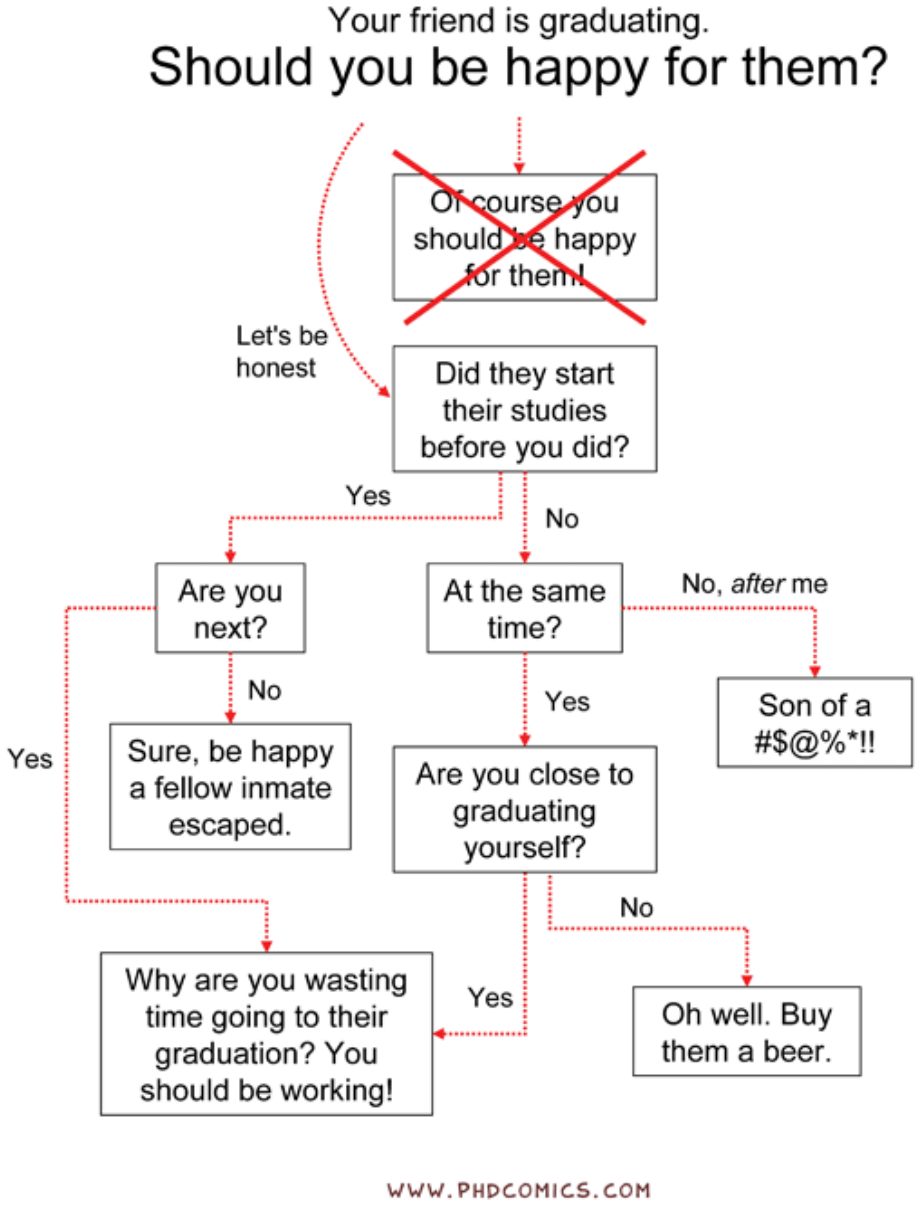
– Attributed to Arthur Samuel

A computer program is said to learn from **experience E** with respect to some **class of tasks T** and **performance measure P** if ***its performance at tasks in T, as measured by P, improves with experience E***.

– Tom M. Mitchell, 1997

Example: Decision Trees

- One of the simplest machine learning models
- Human-like decision process
- More interpretable



Toy Example: Animal Classification

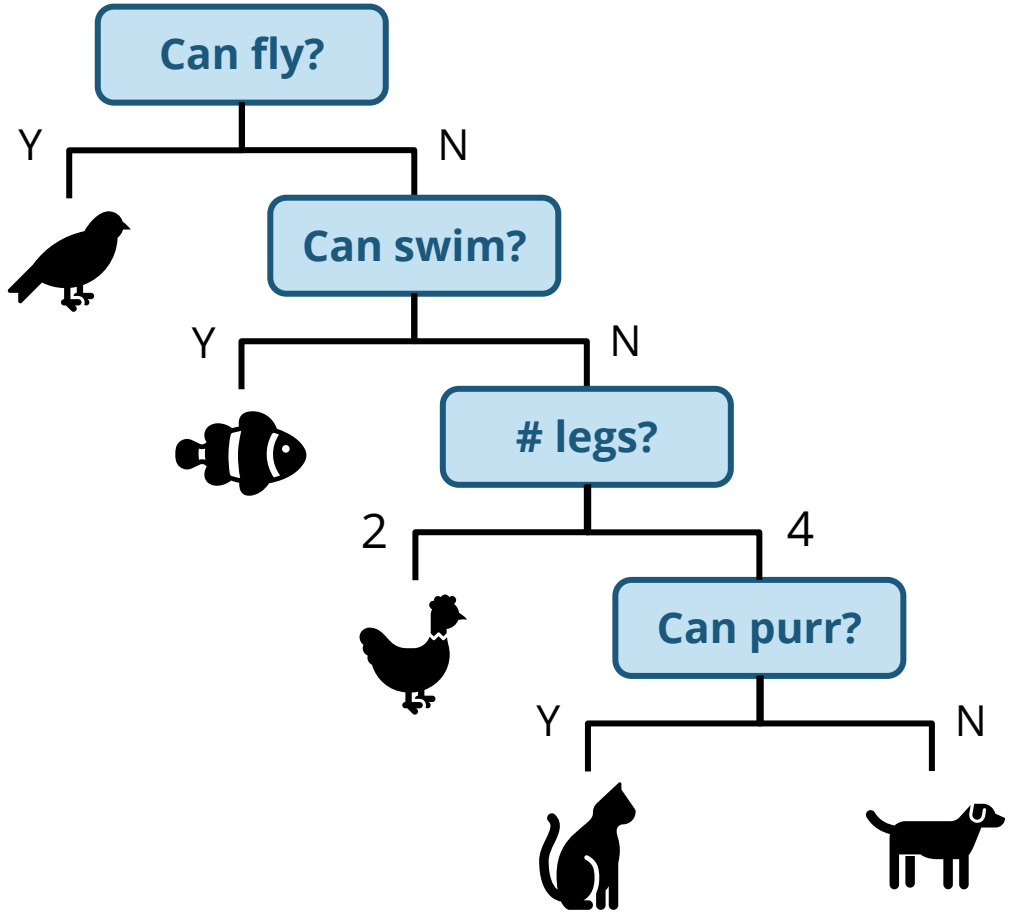
- Suppose we have the following dataset

	Can fly?	Can swim?	# of legs	Can purr?
	N	N	2	N
	N	N	4	Y
	N	Y	0	N
	Y	N	2	N
	N	N	4	N

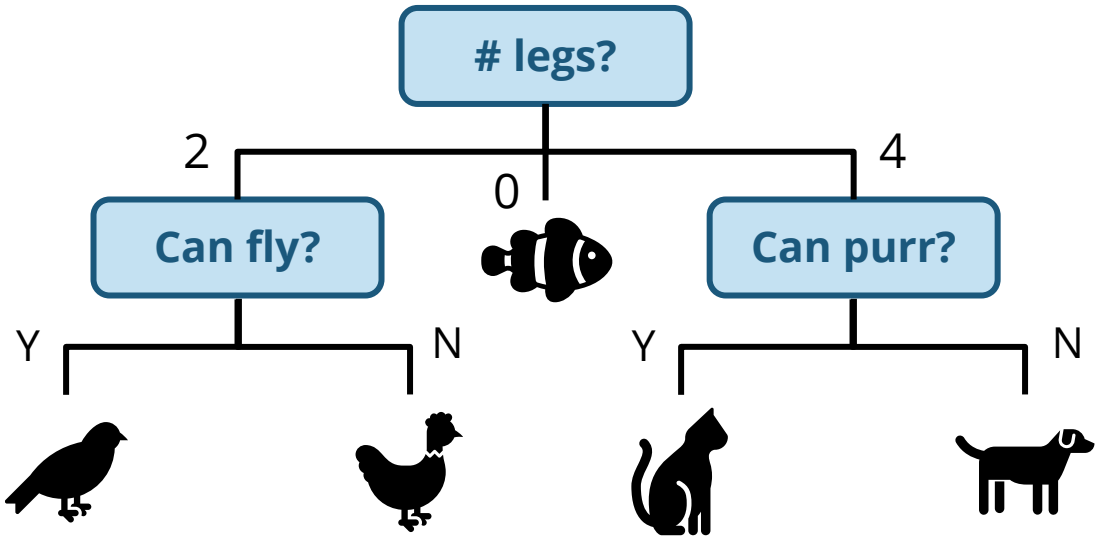
Features

Building a Decision Tree

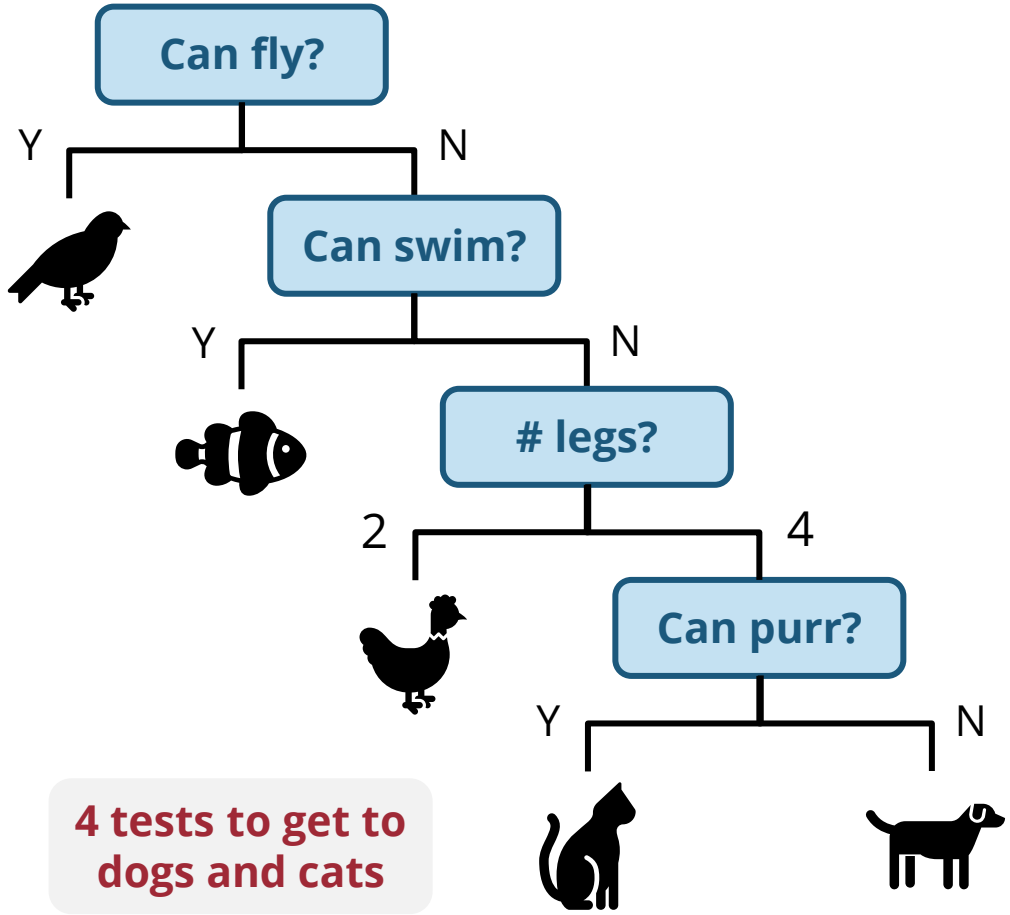
	Can fly?	Can swim?	# of legs	Can purr?
	N	N	2	N
	N	N	4	Y
	N	Y	0	N
	Y	N	2	N
	N	N	4	N



How to Determine the Order of Features to Test?



2 tests maximum to get to any animal type



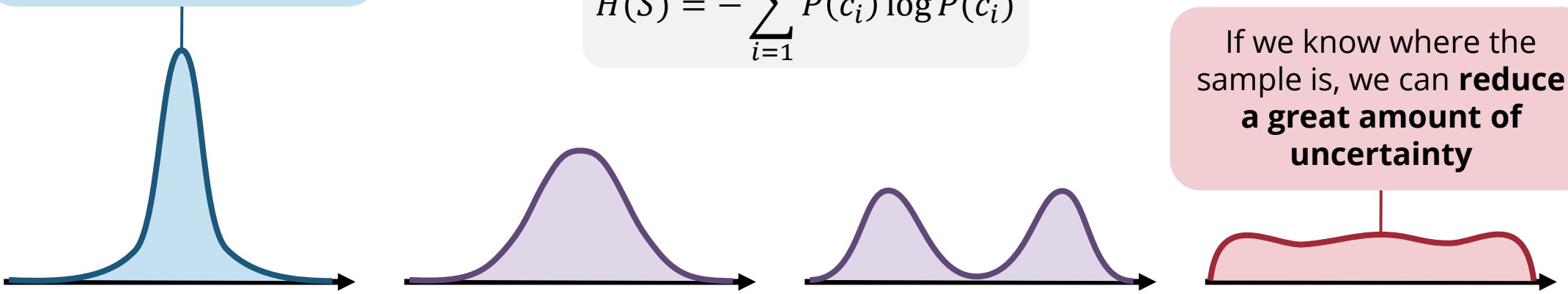
4 tests to get to dogs and cats

Entropy of a Distribution

Even without knowing anything, we know **the sample is very likely to be around the peak**

$$H(S) = -\sum_{i=1}^{C} P(c_i) \log P(c_i)$$

If we know where the sample is, we can **reduce a great amount of uncertainty**



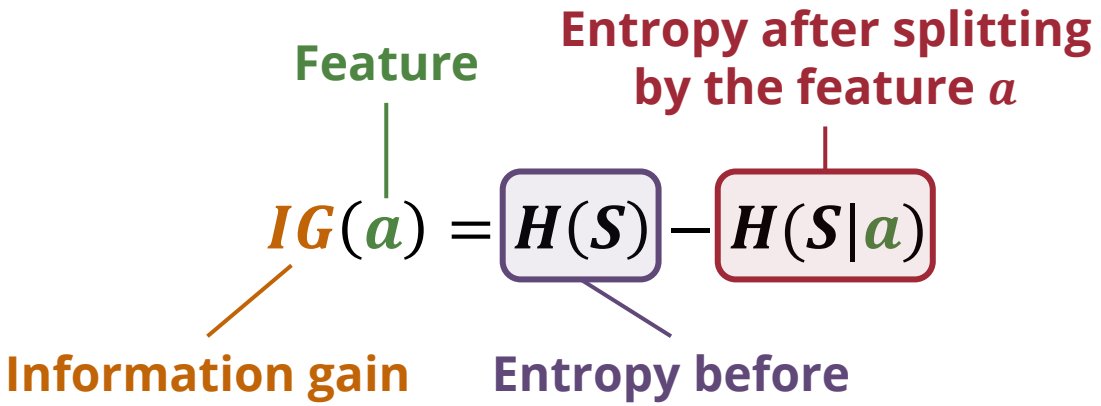
Low entropy ⟷ **High entropy**

Provide less information

Provide more information

How to Determine the Order of Features to Test?

- Occam's razor: *The simplest explanation is usually the best one*
- In each iteration, pick the feature that gives the highest **information gain**



The idea behind the ID3 and C4.5 algorithms

Toy Example: Animal Classification

- Suppose we have the following dataset

Largest information gain if selected!

	Can fly?	Can swim?	# of legs	Can purr?
	N	N	2	N
	N	N	4	Y
	N	Y	0	N
	Y	N	2	N
	N	N	4	N

Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

• **Task T**	**Animal classification**
• **Performance metric P**	Percentage of correct predictions
• **Experience E**	Examples of animals with their features

Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

• **Task T**	**Stock price prediction**
• **Performance metric P**	Average difference from actual stock price
• **Experience E**	History stock price

Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

• **Task T**	**Piano transcription**
• **Performance metric P**	Percentage of correctly predicted notes
• **Experience E**	Recordings with sheet music

Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

• **Task T**	**Beat tracking**
• **Performance metric P**	Average difference from actual timings
• **Experience E**	Recordings with beat timestamps

Types of Machine Learning

- **Supervised learning** — Given **pairs of example inputs and outputs**
  - **Classification**: *discrete* outputs
  - **Regression**: *continuous* outputs
- **Unsupervised learning** — Given ***only* example inputs**
  - **Self-supervised learning**
- **Semi-supervised learning** — Given **example inputs** and **a few example outputs**
- **Reinforcement learning** — Given **scalar rewards** for **a sequence of actions**

Many generative AI models based on self-supervised learning!