AIR5101 / CIE6021

Generative AI

Lecture 1: Introduction

Instructor: Zhen Liu

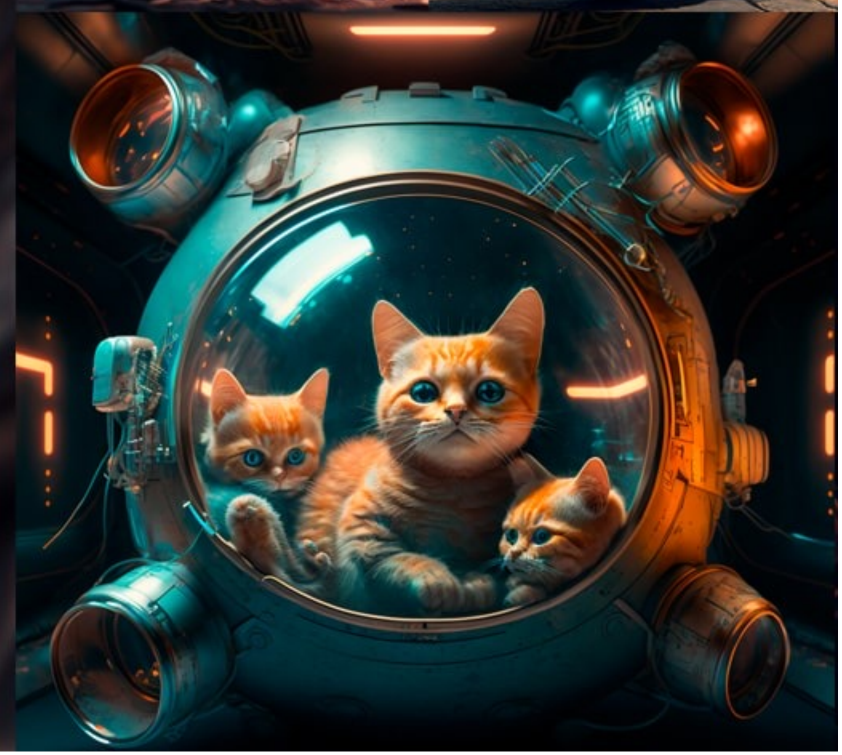
Spring 2025, CUHK-Shenzhen

What is GenAI?

I do not want to spend a semester learning GenAI. Just tell me what GenAI is.

Generative AI (GenAI) refers to artificial intelligence models designed to create new content such as text, images, audio, video, or even code. Instead of just analyzing or recognizing data, these models generate new data that resembles the examples they were trained on.



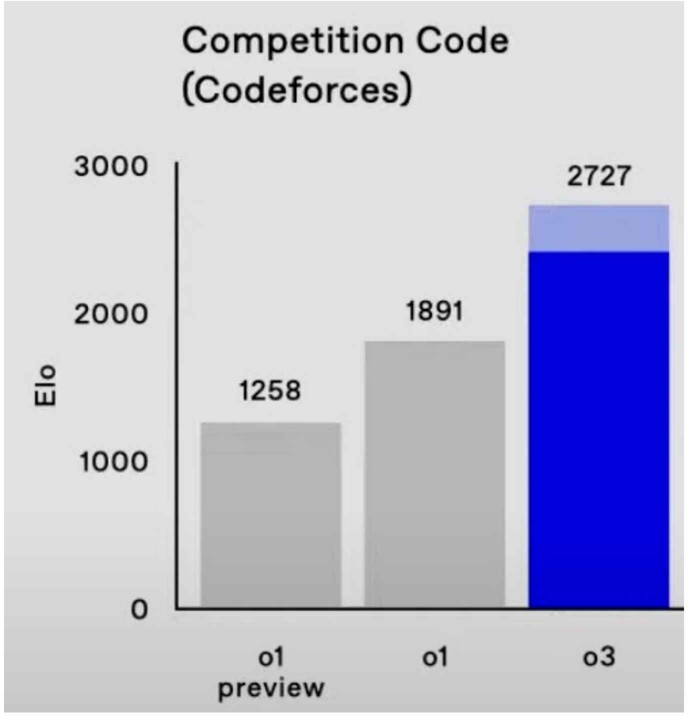
Code Writing

Write a mandelbrot fractal algorithm in Python

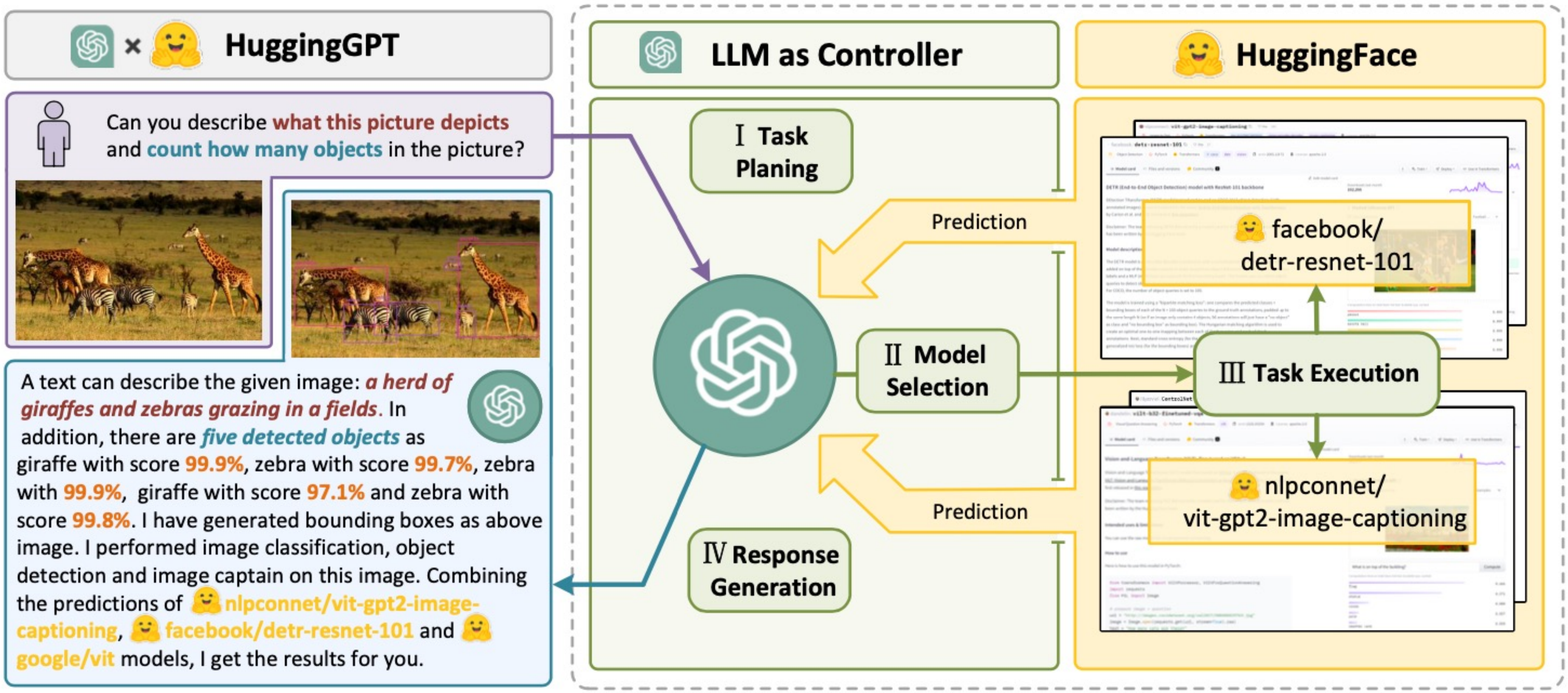
Here's a simple implementation of the Mandelbrot fractal algorithm in Python:

python Copy code

```python
import numpy as np
import matplotlib.pyplot as plt

def mandelbrot(c, maxiter):
    z = c
    for n in range(maxiter):
        if abs(z) > 2:
            return n
        z = z*z + c
    return maxiter

def mandelbrot_set(xmin, xmax, ymin, ymax, width, height, maxiter):
    r1 = np.linspace(xmin, xmax, width)
    r2 = np.linspace(ymin, ymax, height)
    n3 = np.empty((width, height))
    for i in range(width):
        for j in range(height):
            n3[i,j] = mandelbrot(r1[i] + 1j*r2[j], maxiter)
    return (r1, r2, n3)

def plot_mandelbrot(xmin, xmax, ymin, ymax, width=10, height=10, maxiter=100):
    x, y, z = mandelbrot_set(xmin, xmax, ymin, ymax, width, height, maxiter)
    plt.imshow(z, extent=(xmin, xmax, ymin, ymax), cmap='bone', origin='lower')
    plt.show()

plot_mandelbrot(-2, 0.5, -1.25, 1.25, width=500, height=500, maxiter=80)
```



LLM Agents



HuggingGPT
Can you describe what this picture depicts and count how many objects in the picture?
A text can describe the given image: a herd of giraffes and zebras grazing in a fields. In addition, there are five detected objects as giraffe with score 99.9%, zebra with score 99.7%, zebra with 99.9%, giraffe with score 97.1% and zebra with score 99.8%. I have generated bounding boxes as above image. I performed image classification, object detection and image captain on this image. Combining the predictions of nlpconnet/vit-gpt2-image-captioning, facebook/detr-resnet-101 and google/vit models, I get the results for you.
LLM as Controller
HuggingFace
I Task Planing
Prediction
II Model Selection
III Task Execution
facebook/ detr-resnet-101
nlpconnet/ vit-gpt2-image-captioning
Prediction
IV Response Generation

Key components of AI

- **Learning**: capable of learning from experience
- **Reasoning & decision making**: capable of making long-term decisions, especially in stochastic environments with partial information
- **Interaction**: capable of understanding and communicating with humans

Features of Generative AI

- **Creation**: can generate never-seen examples in a semantically meaningful way
- **Scalability**: can be stably trained on huge datasets with lots of parameters
- **Scaling Law**: performance gets better with more compute, data & parameters
- **Cross-modality**: cross-modality generalization once trained on multi-modal datasets

Course Information

This is an advanced-level course on Generative AI

Upon completion, you are expected to

- Understand how the state-of-the-arts methods work
- Be able to read and think critically of recent papers

Course Structure

A mix of lectures and seminars

- Including ~6 guest lectures (subject to change)
- Student presentations on recent papers

Assignments + final projects (will discuss later)

What we will cover

A brief overview on generative models

- VAE, EBM, GAN, normalizing flow, diffusion model, flow matching, autoregressive models

Basics of modern GenAI models

- Architectures, training, inference, finetuning, quantization, alignment, test-time scaling, etc.

Applications and research frontiers

- 3D, video, planning/robotics, etc.

What this course is not for

Intro to math/ML/AI

Comprehensive intro to Bayesian methods (we touch a bit on those for GenAI)

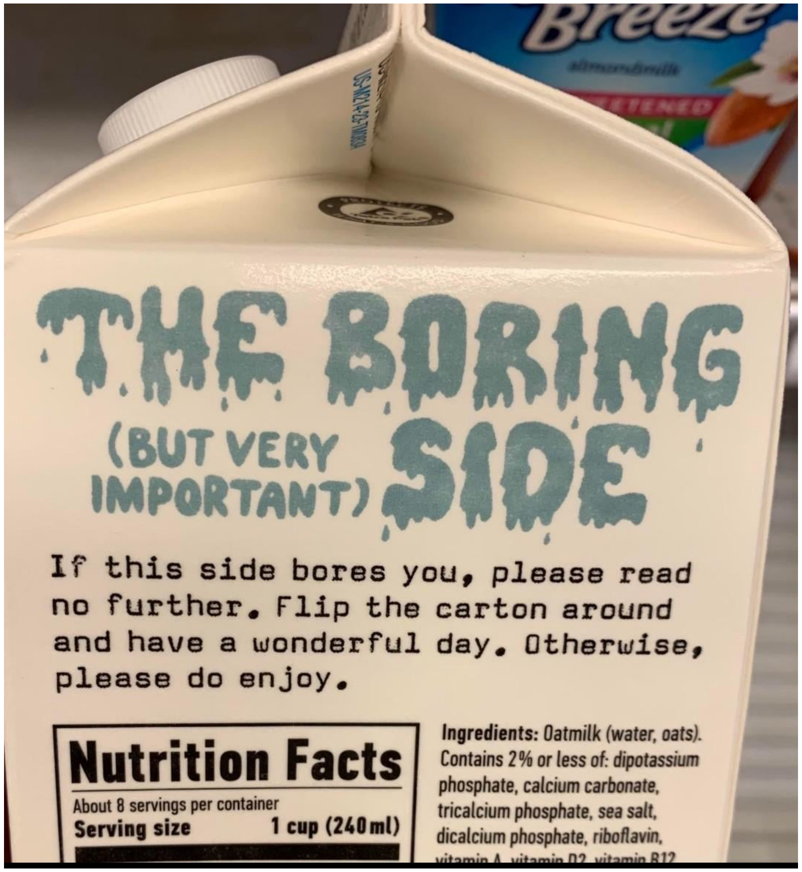
Highly domain-specific CV/NLP/robotics tasks

"How to design better prompts for generating lovely cat images?"*

Amazing tricks that work only for GPT-4 but not for LLaMA

*Disclaimer: we do cover prompt design, but as a general ML problem

Course Logistics



Teaching Team

Instructor: Zhen Liu

- zhenliu@cuhk.edu.cn
- Office Hour: 3pm-4pm TR

TAs:

Yinghong Liao

- 220019033@link.cuhk.edu.cn
- Office hour: 10-11am Friday

Haoyang Li

- 222041054@link.cuhk.edu.cn
- Office hour: 10-11am Wednesday

Course Logistics

- Time: 1:30 – 3pm TR
- Course website: https://air5101.github.io/
- Slides and course updates will be posted on the website
- Email policy: ALWAYS have the subject line starting with [AIR5101/CIE6021]
    - Otherwise it will be ignored

Prerequisites

Linear algebra, calculus, probability theory / statistics

Basics of machine learning / computer vision / natural language processing

Proficiency in Python and ideally libraries like PyTorch

Questions to ask yourself:

- What is the difference between forward and reverse KL? What is a validation set?
- What is the rank of a matrix? How about the singular value?
- If p(x) and p(y|x) are both Gaussians with known mean and variances, what is the formula for p(y)?

Grading

Coding Assignments – 30%

Paper Presentations & Reviews – 30%

Final (Group) Project – 40%

Coding Assignments

Around 4 in total

Tasks:

- Implementation of toy models with PyTorch
- Simple experiments with modern GenAI libraries
- Slightly open-ended questions

Paper Presentations & Reviews

Presentations – Week 3 (excluding holidays) to Week 12/13

- 1 to 4 paper presentations (group of 1 or 2 people) per class
- More presentations per class towards the end of this semester

Reviews

- Each student is required to submit a paper review per week
- Unless they have their paper presentation for that week

Paper Reviews

Imagine that you are reviewing a paper for a conference

- What is the background of this paper? What did the authors do?
- Major contributions? What experiments?
- Strengths and weaknesses?
- Critics – What should have been done? What was done incorrectly?
- Future work? Questions for the authors?

Plenty of examples from ICLR openreview. e.g., https://openreview.net/forum?id=PxTIG12RRHS

Final Project

Groups of 1-4 people

Final week for presentations (details TBD)

Deadlines:

- Feb 13 – Form a team
- Feb 20 – Project proposal
- May 10 – Project report
- Final week – Project presentation

Project Topics

Novel techniques / methods

Empirical observations (e.g., https://openreview.net/forum?id=Sy8gdB9xx)

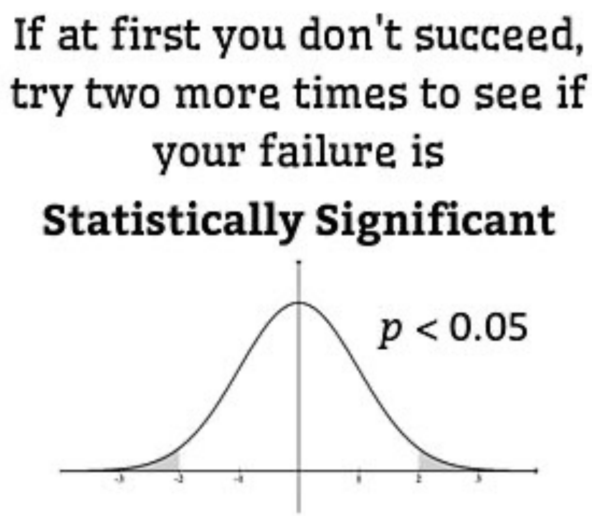
Applications

Ablation studies / engineering recipes (e.g., https://arxiv.org/abs/2102.09672)

Theoretical analysis (not very recommended)

Plagiarism

Assignments too hard?



Ask your classmates / TA / me to help after you try hard

LLM Policy

Allowed only for paper polishing / understanding basics concepts

Always remember to write your first draft in your own words

Penalized if significant cues of heavy LLM use are found repeatedly

Generative AI

Generative + AI

Generative + AI

Artificial Intelligence

Dartmouth workshop, Summer 1956

A grand (and failed) proposal

“ We propose that a 2-month, 10-man study of artificial intelligence... ”

A grand (and failed) proposal

“ We propose that a 2-month, 10-man study of artificial intelligence... ”

“ the conjecture that every … feature of intelligence… a machine can be made to simulate it. ”

A grand (and failed) proposal

“ We propose that a 2-month, 10-man study of artificial intelligence... ”

“ the conjecture that every ... feature of intelligence... a machine can be made to simulate it. ”

“ ...to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. ”

A grand (and failed) proposal

" We propose that a <u>2-month, 10-man study</u> of artificial intelligence... "

" the conjecture that every … feature of intelligence… a machine can be made to simulate it. "

" …to make machines <u>use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves</u>. "

" …a significant advance can be made in… <u>a summer</u>. "

Symbolic AI

Examples:

- First-order logics
- Constraint satisfaction problems
- Syntax and grammar rules

Issues:

- Exponential growth of search space
- Ambiguity in problem definitions

Failure of Symbolic AI

The spirit is willing but the flesh is weak.

↓

(Russian)

↓

The vodka is good but the meat is rotten.

Connectionism

Intelligence == connections of neurons

1958 Rosenblatt: Perceptron as a linear classifier

- Cannot even approximate XOR

Multi-layer perceptrons, convolutional nets:

- Too many parameters in the 70s-80s
- Speed? Generalization?

Statistical Machine Learning

“Every time I fire a linguist,
the performance of the speech recognizer goes up”

Statistical ML

Methods that learn from data:

- Support vector machine
- Graphical models
- Kernel methods

with worst-case statistical guarantees:

- VC dimension, Rademacher complexity, ...

Deep Learning

- AlexNet (2012): a large convolutional network on ImageNet of >1m images.
- AlphaGo (2016): Reinforcement learning with thousands of GPU hours to search and explore
- BERT (2018): generalizable text embeddings from learning on large-scale corpus

Foundation Models

- CLIP (2021): zero-shot cross-modality generalization from training on Internet-scale noisy image-text pairs
- StableDiffusion (2022): a large convolutional network on ImageNet of >1m images.
- ChatGPT (2022): Huge decoder-only transformer trained to achieve human-like conversation capability trained with thousands of GPUs for months

What in common: scaling

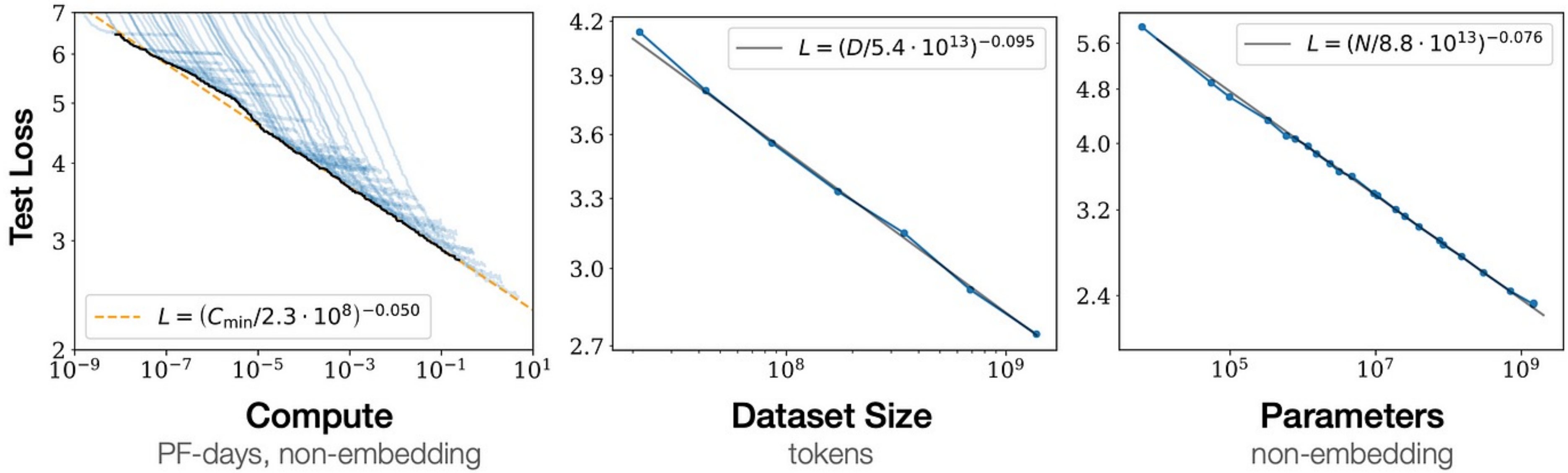
“The Bitter Lesson”

In short,

Major AI breakthrough == More compute

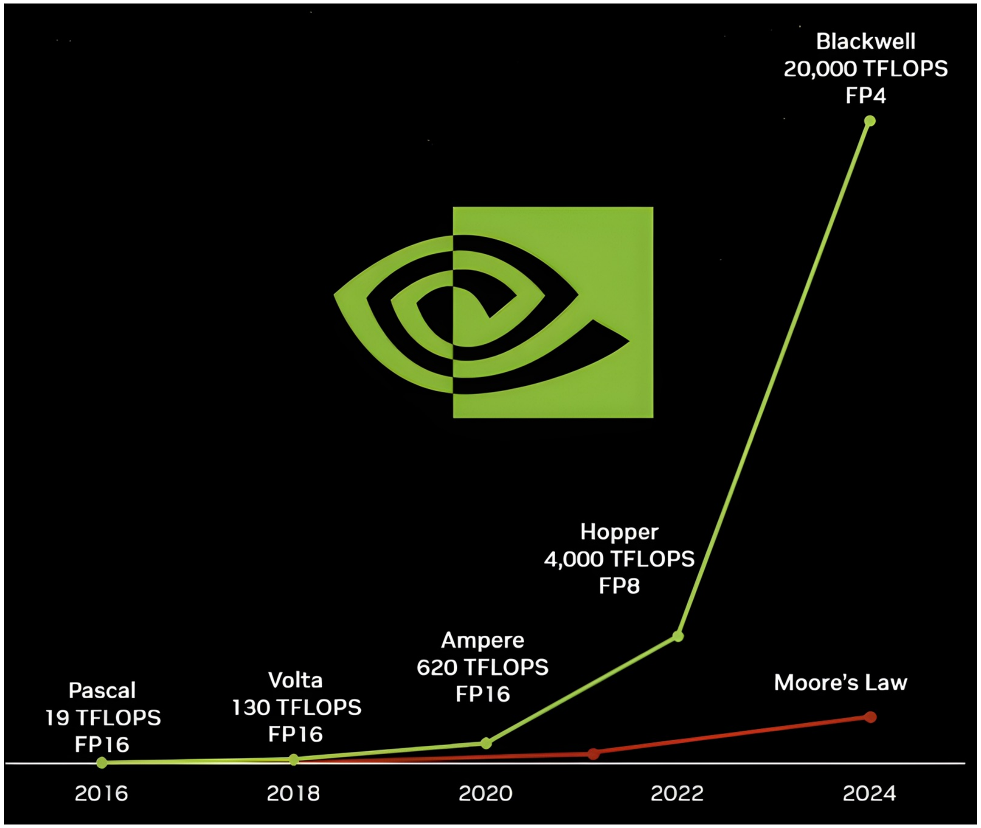
Quote from Richard Sutton. http://www.incompleteideas.net/IncIdeas/BitterLesson.html

Scaling Law



Test Loss
7
6
5
4
3
2
$L = (C_{min}/2.3 \cdot 10^{8})^{-0.050}$
10^{-9} 10^{-7} 10^{-5} 10^{-3} 10^{-1} 10^{1}
Compute
PF-days, non-embedding
4.2
3.9
3.6
3.3
3.0
2.7
$L = (D/5.4 \cdot 10^{13})^{-0.095}$
10^{8} 10^{9}
Dataset Size
tokens
5.6
4.8
4.0
3.2
2.4
$L = (N/8.8 \cdot 10^{13})^{-0.076}$
10^{5} 10^{7} 10^{9}
Parameters
non-embedding

"Huang's Law"


Blackwell
20,000 TFLOPS
FP4
Hopper
4,000 TFLOPS
FP8
Ampere
620 TFLOPS
FP16
Volta
130 TFLOPS
FP16
Pascal
19 TFLOPS
FP16
Moore's Law
2016
2018
2020
2022
2024

Source of scaling?

Massive Unlabeled Datasets
+ Huge Compute

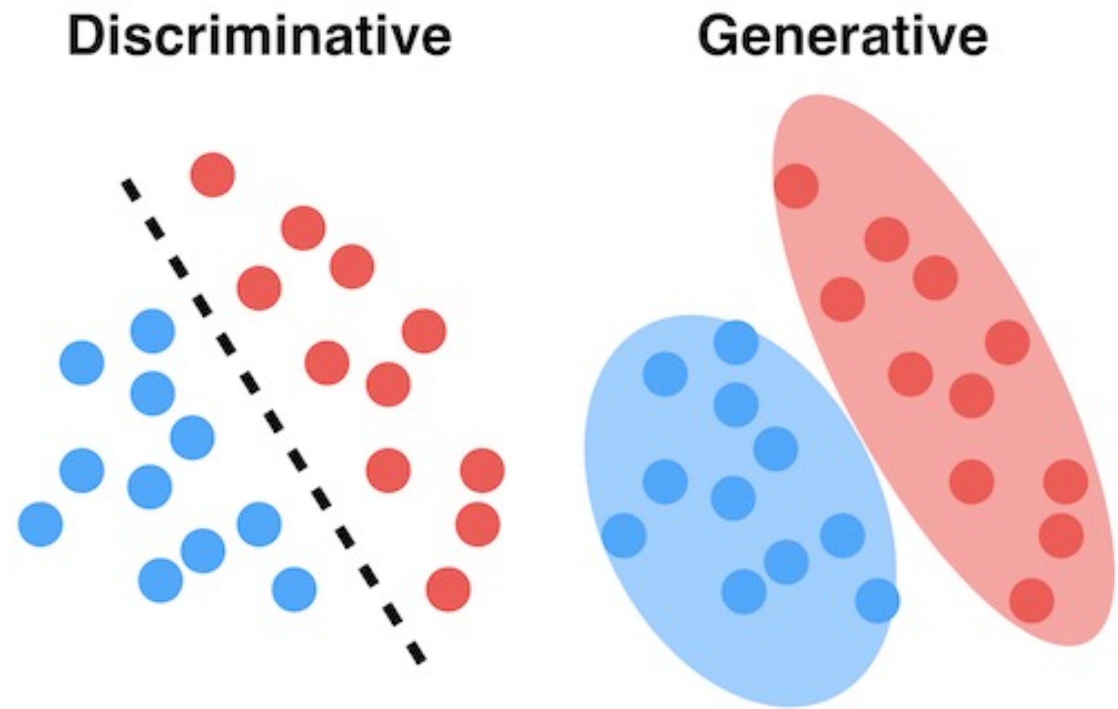
Huge compute with unlabeled datasets?

Generative Models

Generative + AI

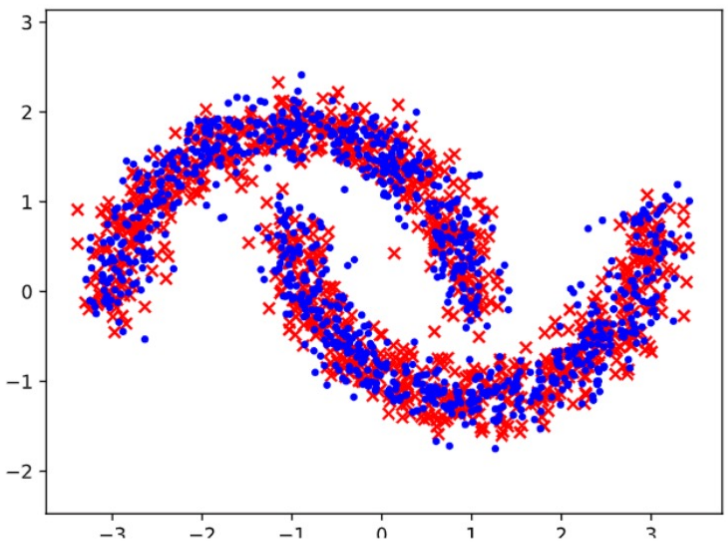
Generative vs. Discriminative

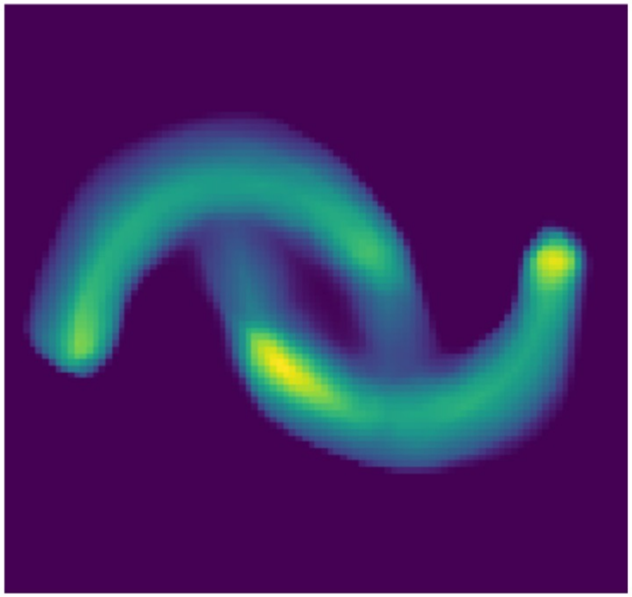
$$p(y|x)$$



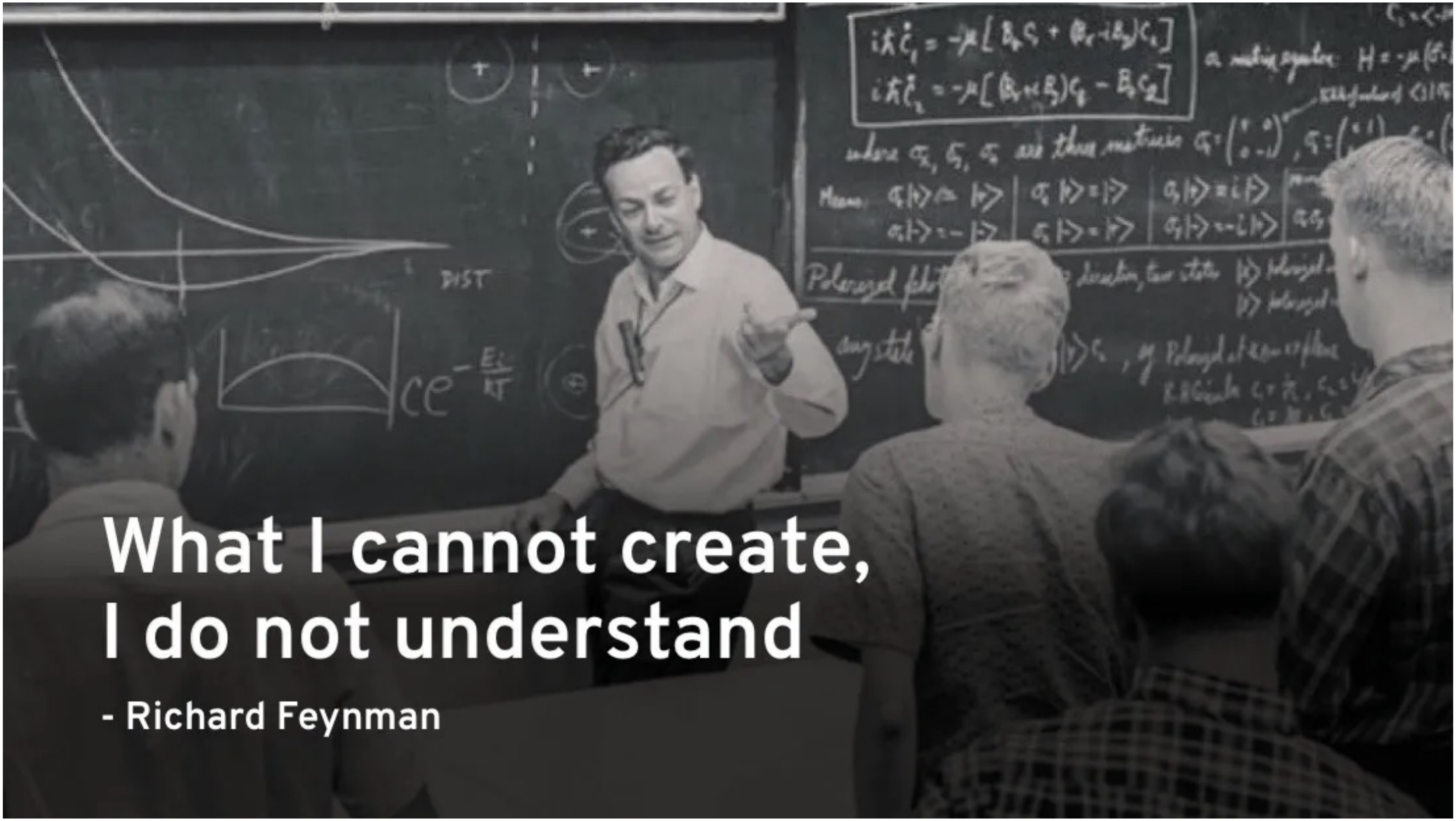
$$p(x)$$

Generative Models as Extrapolation

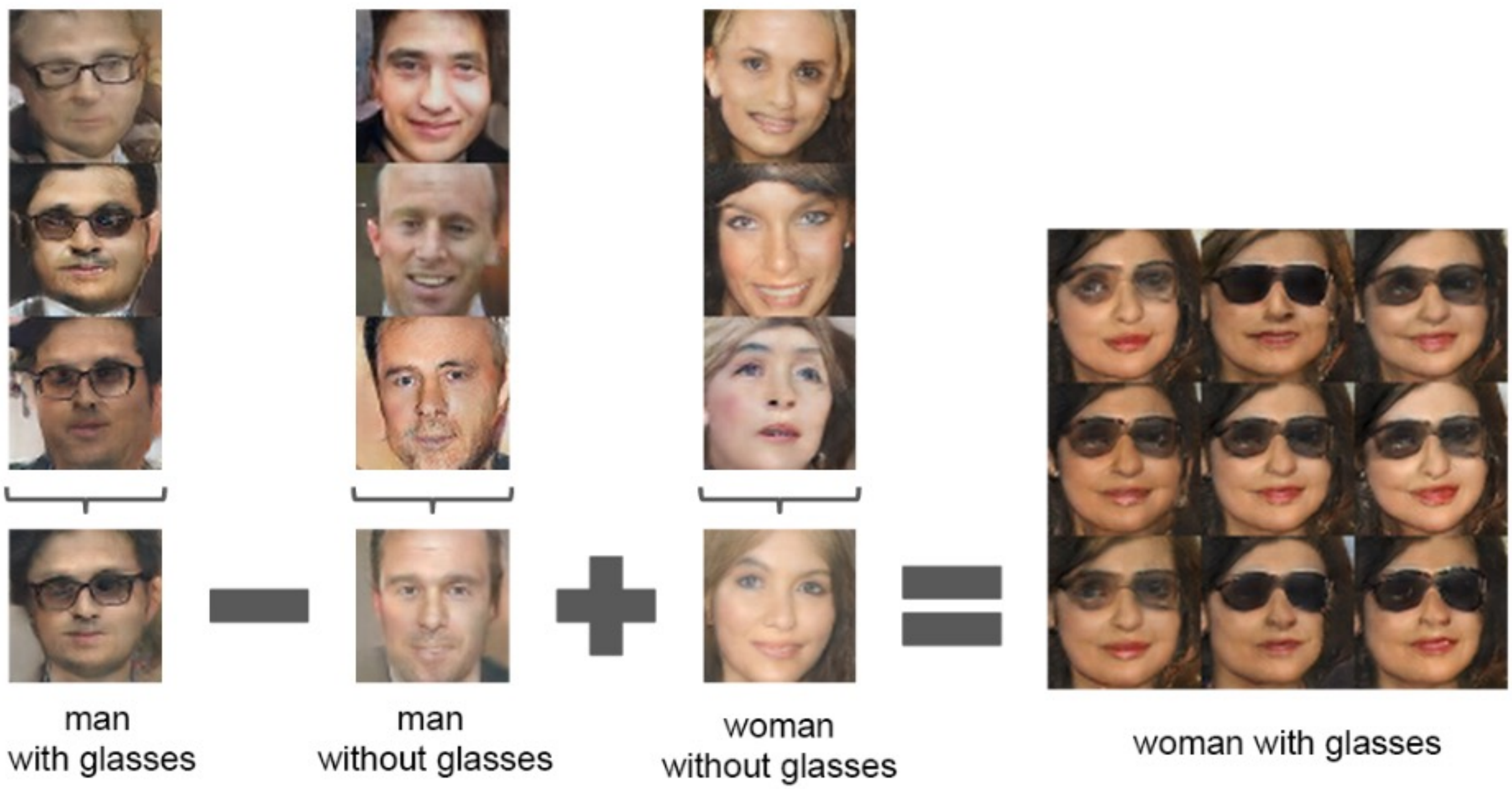




Red: Training data
Blue: Samples from learned distribution


What I cannot create,
I do not understand
- Richard Feynman

Compositional Extrapolation of Semantics



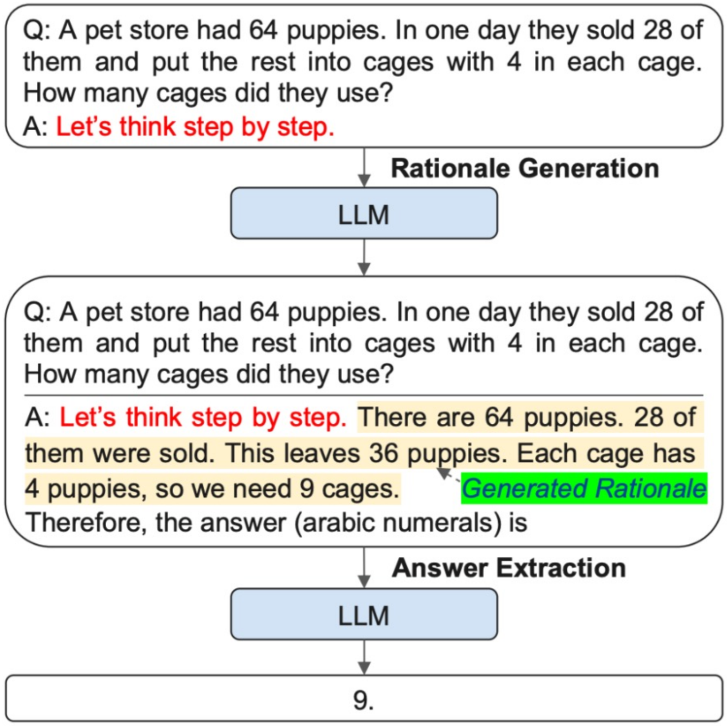
Semantics Composition – Text to Image

Input images

A [V] clock on the beach

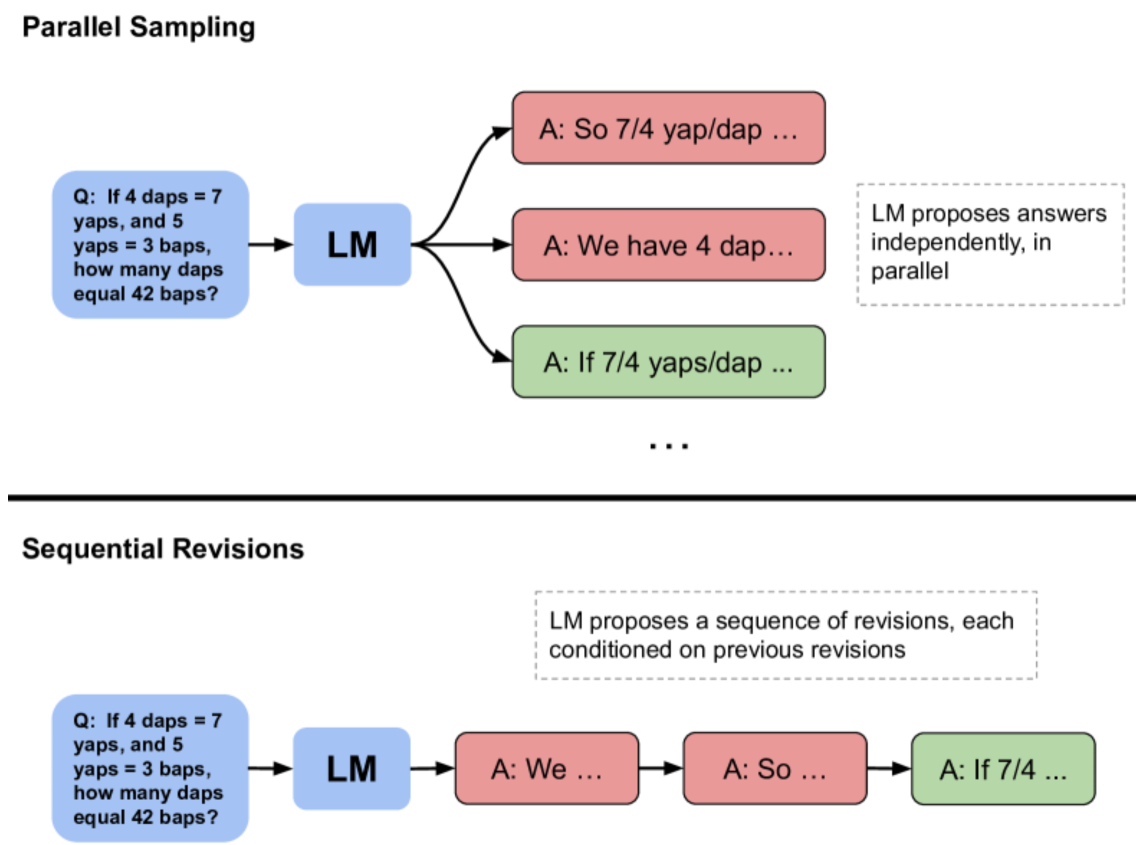
A [V] clock on top of green gras with sunflower

Compositionality and Reasoning



Composition of elementary reasoning steps

Test-time Scaling



GenAI Scaling Accelerated

Open-sourced Libraries

- Diffusers
- DeepSpeed
- vLLM

Open-sourced Models

- LLaMA
- Qwen
- StableDiffusion

Affordable Proprietary Models

- ChatGPT
- Claude

New Paradigms of GenAI

- General-task pretraining vs. Task-specific training
- Prompting vs. Finetuning
- Test-time scaling vs. Training-time scaling
- Synthetic data vs. Collected data
- ...

Back to Generative Models

How to train your generative model?

A Naïve Generative Model – Gaussian

Multivariate Gaussian $\mathcal{N}(\mu, \sigma^2 I)$

Maximum likelihood estimation (MLE) on a dataset $\mathcal{D}$

$$
\begin{aligned}
&\max_{\mu,\sigma} \mathop{\mathbb{E}}_{x\sim\mathcal{D}} \log p_{\mu,\sigma}(x) \\
=& \max_{\mu,\sigma} \mathop{\mathbb{E}}_{x\sim\mathcal{D}} \log \frac{\exp(-\|x-\mu\|^2/2\sigma^2)}{Z} \\
=& \min_{\mu,\sigma} \mathbb{E}_{x\sim\mathcal{D}} \frac{\|x-\mu\|^2}{2\sigma^2}
\end{aligned}
$$

A Naïve Generative Model – Gaussian

Optimization via gradient descent

$$\frac{\partial \mathcal{L}}{\partial \mu} = \mathbb{E}_{x \sim \mathcal{D}} \frac{\mu - x}{2\sigma^2}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\mathbb{E}_{x \sim \mathcal{D}} \frac{\|x - \mu\|^2}{\sigma^3}$$

A Naïve Generative Model – Gaussian

Alternative explanation of MLE estimation

$$\min_\theta D_{\mathrm{KL}}(p_{\mathcal{D}} \| p_\theta)$$ (Minimize KL divergence b/w data and model)

$$= \min_\theta \mathop{\mathbb{E}}_{x \sim p_{\mathcal{D}}} [\log p_{\mathcal{D}}(x) - \log p_\theta(x)]$$ (Definition of KL divergence)

$$= \max_\theta \mathop{\mathbb{E}}_{x \sim p_{\mathcal{D}}} \log p_\theta(x)$$

MLE objective = minimal (reverse) KL divergence

Challenges – Parametrization

MLE for general distributions

$$\max_{\theta} \mathop{\mathbb{E}}_{x \sim \mathcal{D}} \log p_\theta(x)$$

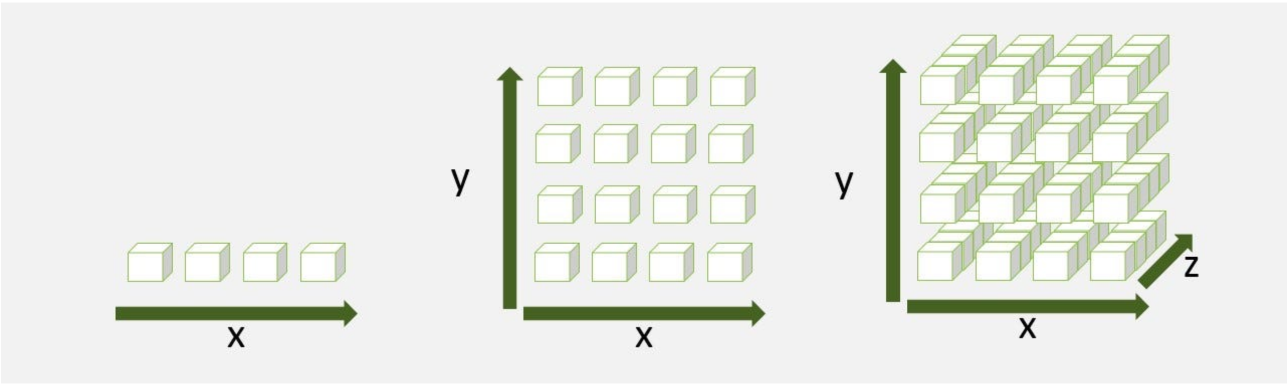
How to (or not to) parametrize $p_\theta(x)$ that satisfies $\int_x p_\theta(x)dx = 1$?

Curse of Dimensionality

Naïve solution: to discretize an n-dim space into K bins per dimension

$$\# \text{ of bins} = O(K^n)$$

Intractable in high-dimensional cases!



Challenges – Training

Minimize $D(p_{\mathcal{D}}, p_\theta)$ without explicit likelihood?

- Lower bounds of likelihood
- Efficient approximation of statistical distances / divergences

Challenges – Sampling

We know how to sample from simple distributions:

- Categorical distribution
- Gaussian distribution
- Beta distribution
- Gamma distribution
- …

What about a general and high-dimensional $p_\theta(x)$?

Curse of Dimensionality

Rejection sampling:

1. Sample from a simple proposal distribution $q(x)$, like $\mathcal{N}(0, I)$
2. Find some M such that $p_\theta(x)/[Mq(x)] \leq 1 \quad \forall x$
3. Sample $u \sim \mathrm{Uniform}(0, 1)$ and accept the sample x if $u \leq \dfrac{p_\theta(x)}{Mq(x)}$

In high-dimensional spaces:

The mass of $\mathcal{N}(0, I)$ is concentrated in a thin spherical shell

$\rightarrow$ $p_\theta(x)/[Mq(x)]$ is generally tiny $\rightarrow$ High rejection rate

Trade-off between Desired Properties

- Tractable Density / Likelihood
- Efficient Training
- Efficient Sampling
- Model Capacity

Next lecture: Variational Autoencoders

Start early and find your teammates