STATS 100A: Probability

Instructor: Xiaowu Dai
Email: dai@stat.ucla.edu
Office: Math Science 8917

Plan

Part 1: basic concepts and rules through examples

Part 2: random variables --- one at a time

Part 3: two or more random variables

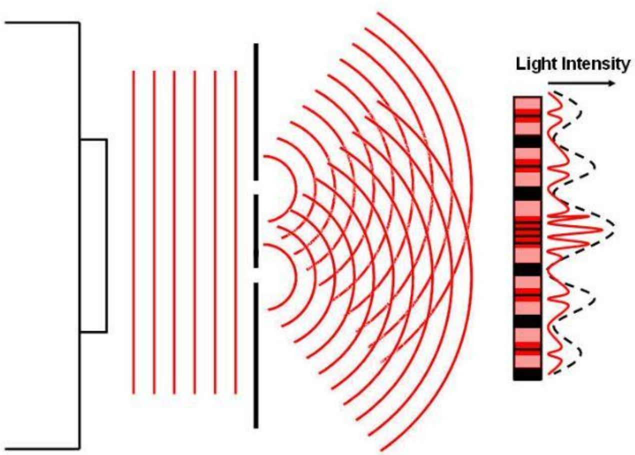
Emphases

- Concepts, intuitions
- Calculations in precise notation
- Scientific applications

Why do we need probability theory?

In the following, all the pictures are taken from internet and all the credits belong to the authors
The content of this lecture will not be in the homework assignments and final exam

Double Slit Diffraction



At the most fundamental level, the physical laws are probabilistic

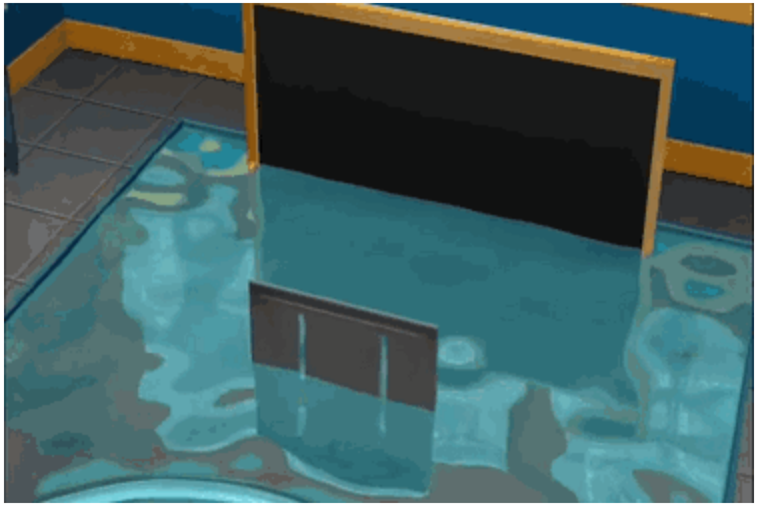
- Two-slit experiment: an electron goes through two slits simultaneously like water waves that can be superposed

1. The evolution of wave function is governed by Schrodinger equation
2. Born: |wave function|^2 = probability density function
3. Observer's measurement, wave function collapses

- Heisenberg uncertainty principle
- Bohr: wave-particle duality
- Einstein (against probabilistic interpretation): God does not play dice
- Schrodinger (also against) cat paradox: superposition of alive and dead

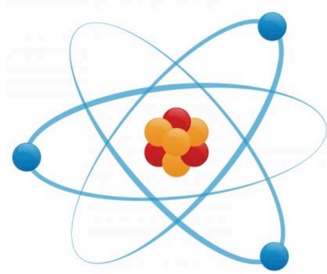
Two-slit experiment

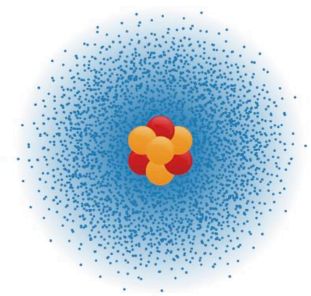
Wave function → probability density

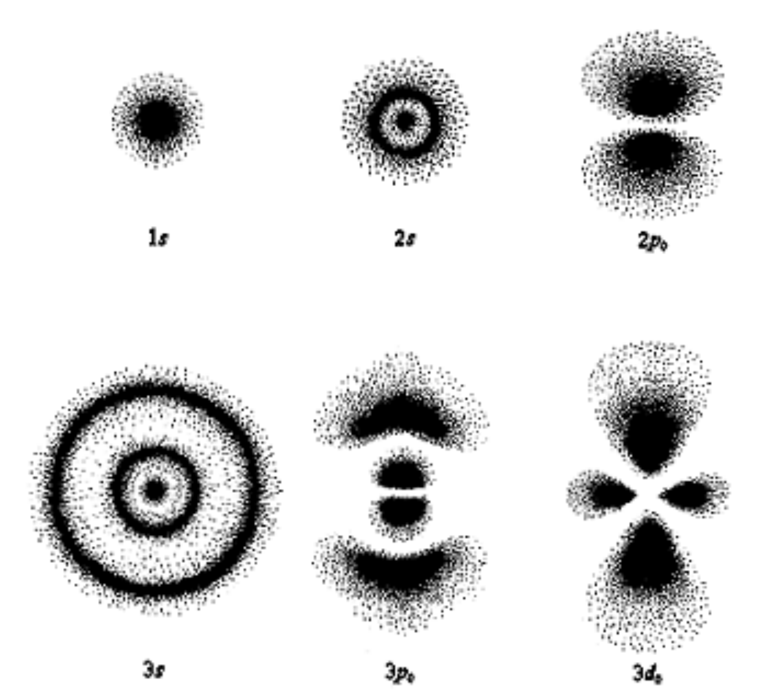


Electron cloud: population of possibilities
probability density

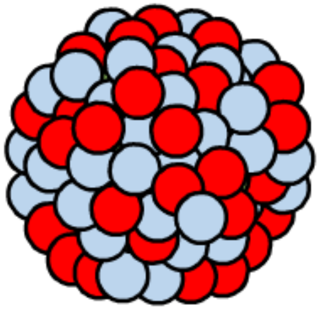
$$p(x) = |\Psi(x)|^2$$



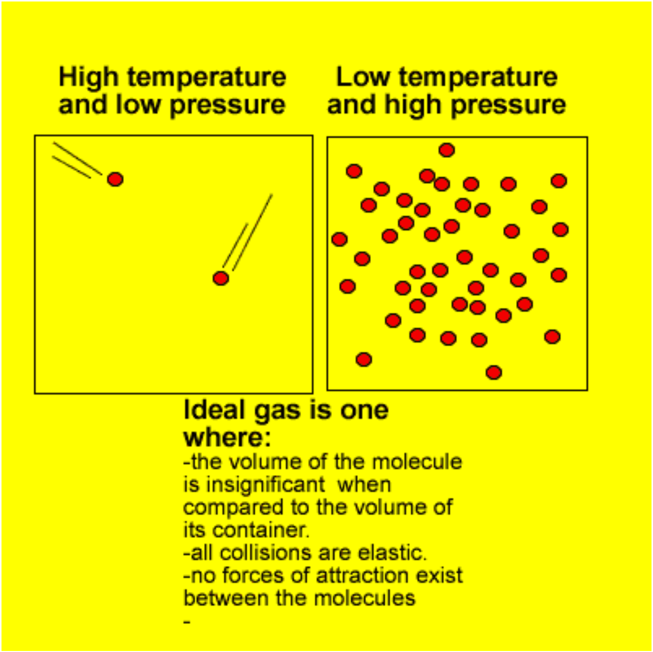


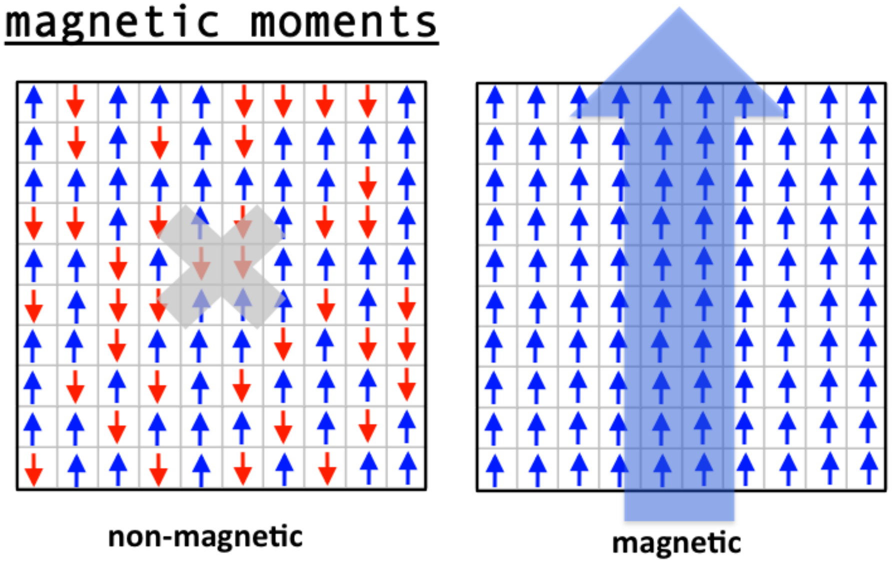


Particle decay



$^{226}_{88}\mathrm{Ra}$





Even if physical laws are deterministic

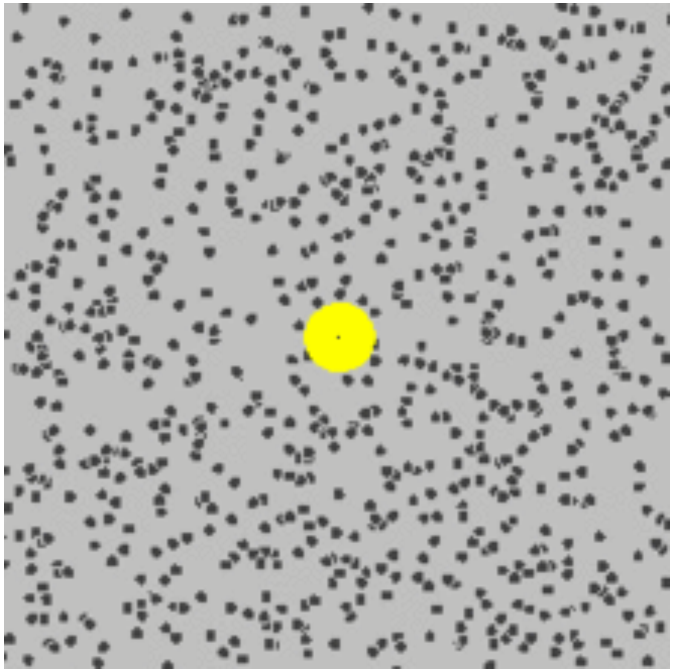
If we study the motions of 10^23 elements, we need statistical models

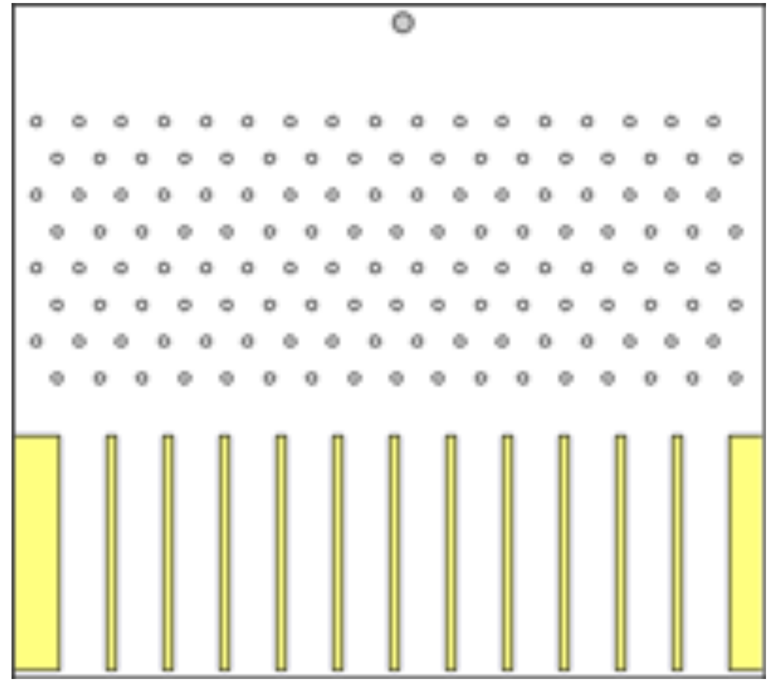
Configuration X in Omega = {X: energy(X) = fixed}

X(t) evolves in Omega and traverses every configuration with equal frequency

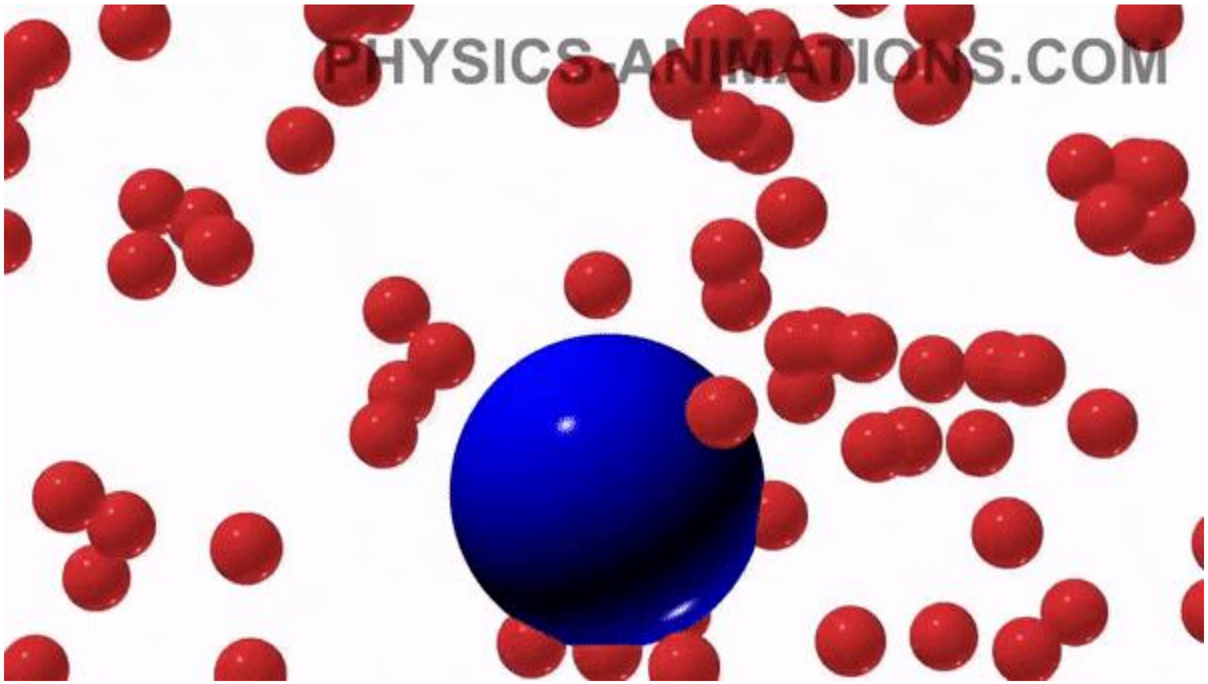
at a random time t, X(t) ~ uniform(Omega) → subsystem x(t) of X(t)

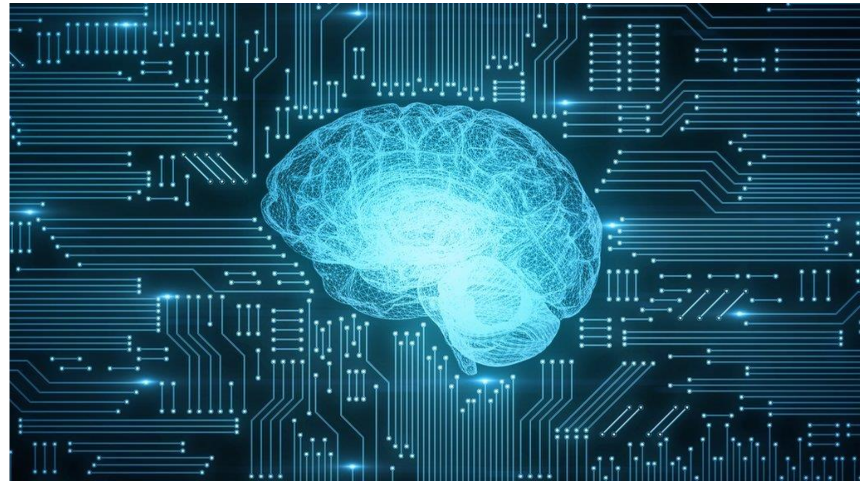
- Ideal gas: temperature, pressure etc. are statistical properties
- Ferromagnetism: Ising model, magnetism is also statistical property
- Phase transition: water/ice/gas, same model →different distributions
- Arrow of time: increase in entropy, again a statistical property





- **Brownian motion**: dust particles in water show zig-zag paths
- Einstein: caused by bombardments from invisible water molecules
- established existence of molecules and atoms
- Left: 2D; Right: 1D simplified, Galton board/quincunx, Gaussian, CLT
- Stochastic differential equation
- Finance, Black-Scholes equation
- Diffusion models for generative AI

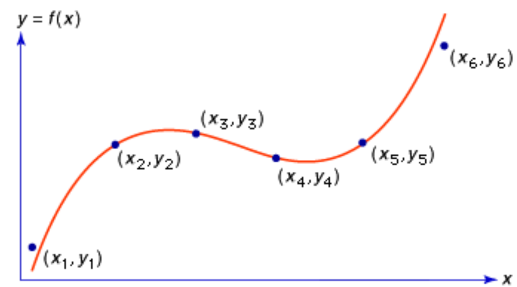

PHYSICS-ANIMATIONS.COM



Artificial intelligence, machine learning, deep learning, computer vision, natural language processing, robotics …

All based on probabilistic framework

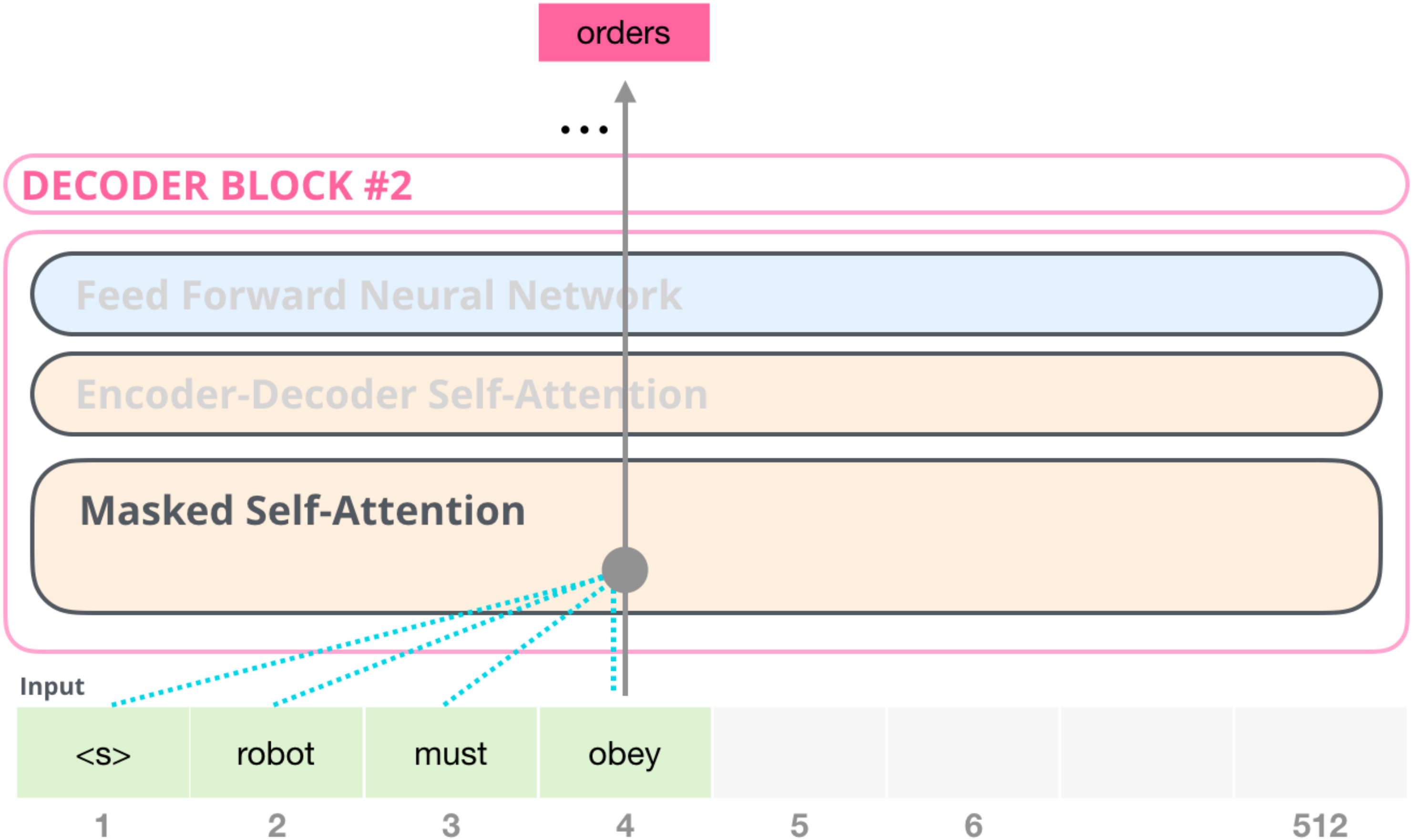
- Training data come from a probability distribution
  $$(x_i, y_i, i = 1, ..., n) \sim p(x, y)$$
- Learn about this probability distribution
  $$\max_\theta \frac{1}{n} \sum_{i=1}^{n} \log p_\theta(y_i|x_i) \rightarrow p_{\hat{\theta}}(y|x)$$



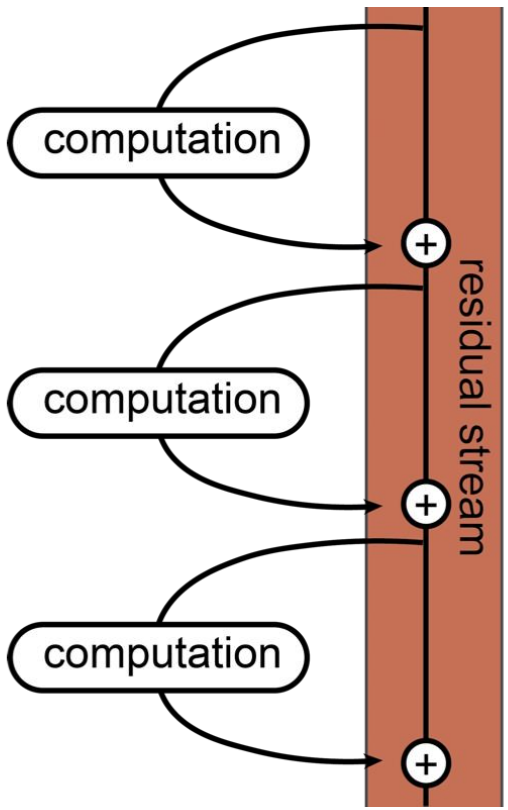
© 2003 Encyclopædia Britannica, Inc.

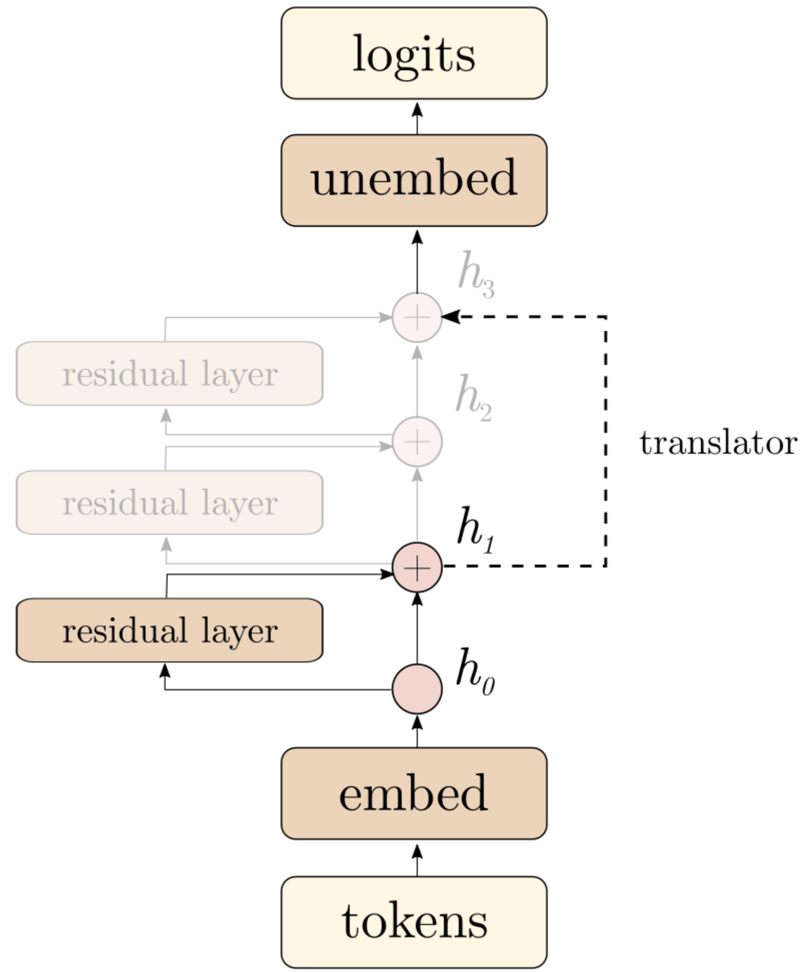
- Generalize to testing data: x: instruction, y: completion
- Memorization and generalization

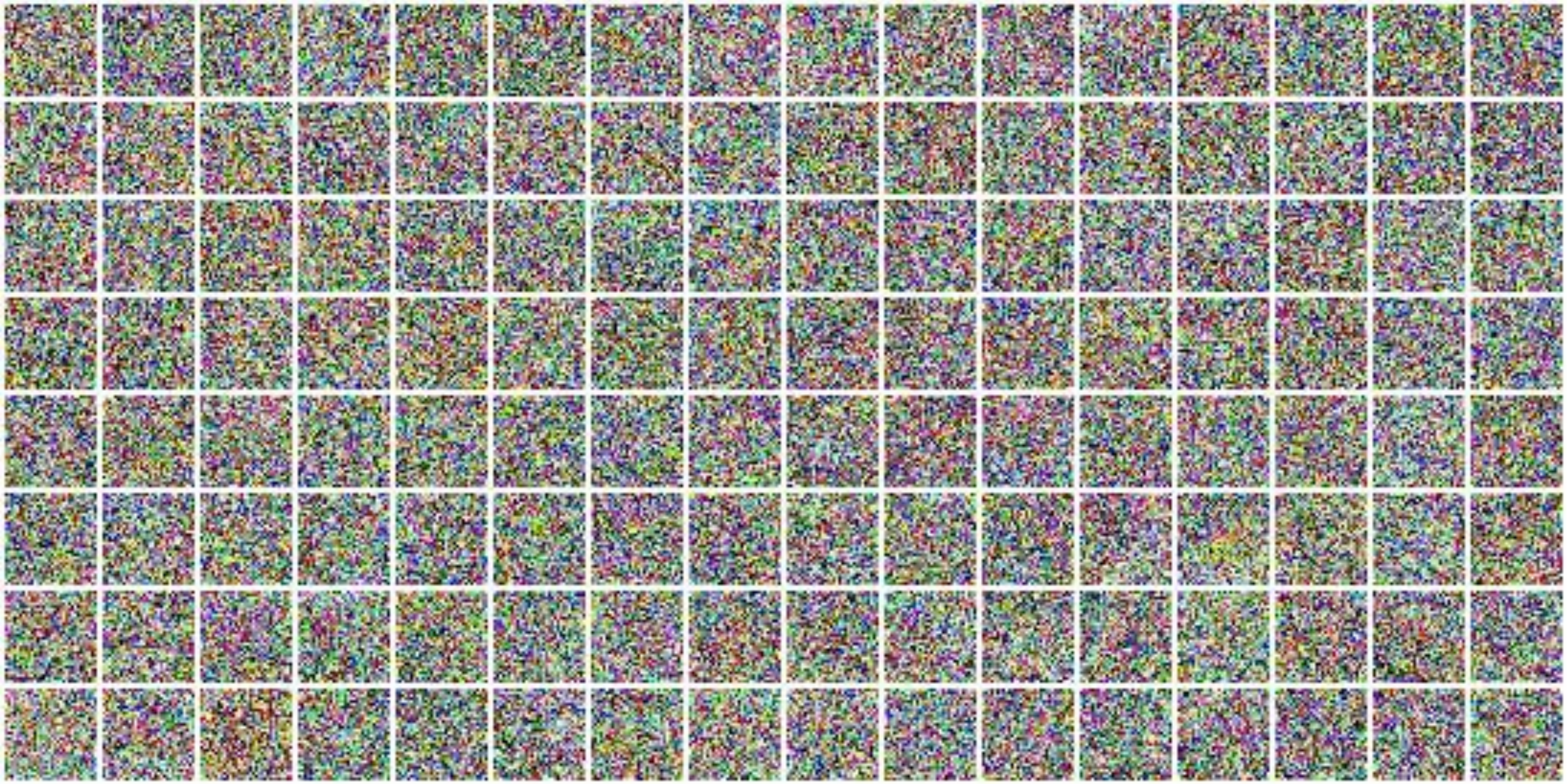
GPT-4



next word prediction $$\prod_t p_\theta(y_t | y_{<t}, x)$$

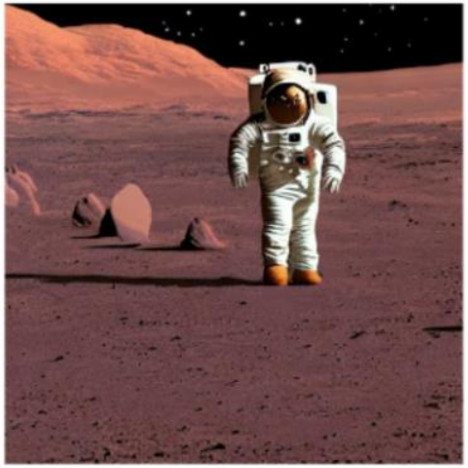

computation
computation
computation
residual stream


logits
unembed
h_3
residual layer
h_2
translator
residual layer
h_1
residual layer
h_0
embed
tokens



Diffusion denoising probability model $$\prod_t p_\theta(y_{t-1}|y_t, x)$$

Stable Diffusion



DALLE 2

Midjourney

Alone astronaut on Mars, mysterious, colorful, hyper realistic

Stable Diffusion

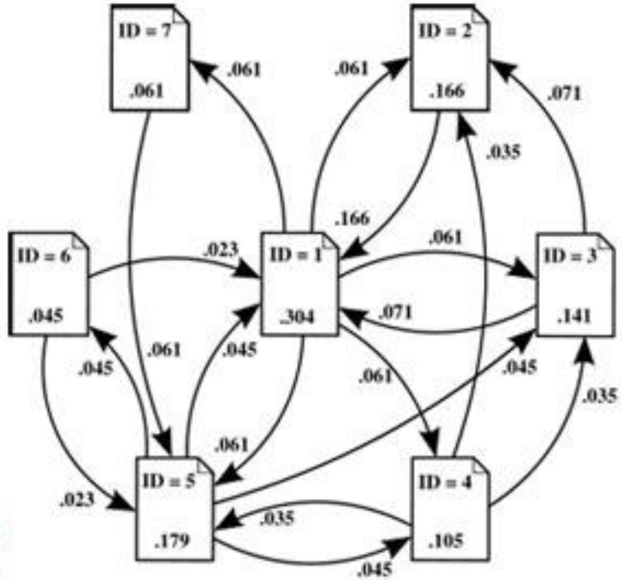
DALLE 2

Midjourney

Pyramid shaped mountain above a still lake, covered with snow

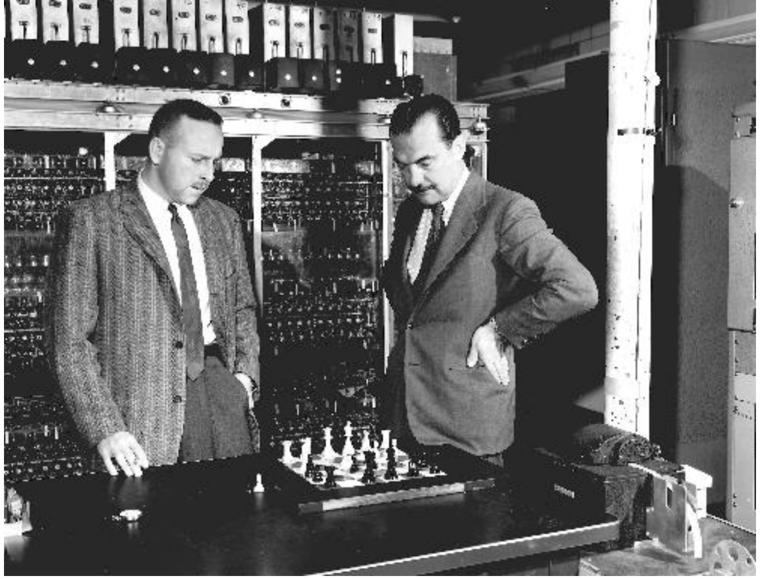

SORA

Prompt: A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.



Google page-rank, random surfing model

- 85% probability, randomly click a link to the next page
- 15% probability, randomly go to another page
- **Markov chain** → stationary distribution
- Imagine 1 billion people follow the same random walk
- Eventually they will be distributed over all the webpages
- Popularity of a page is based on its stationary distribution

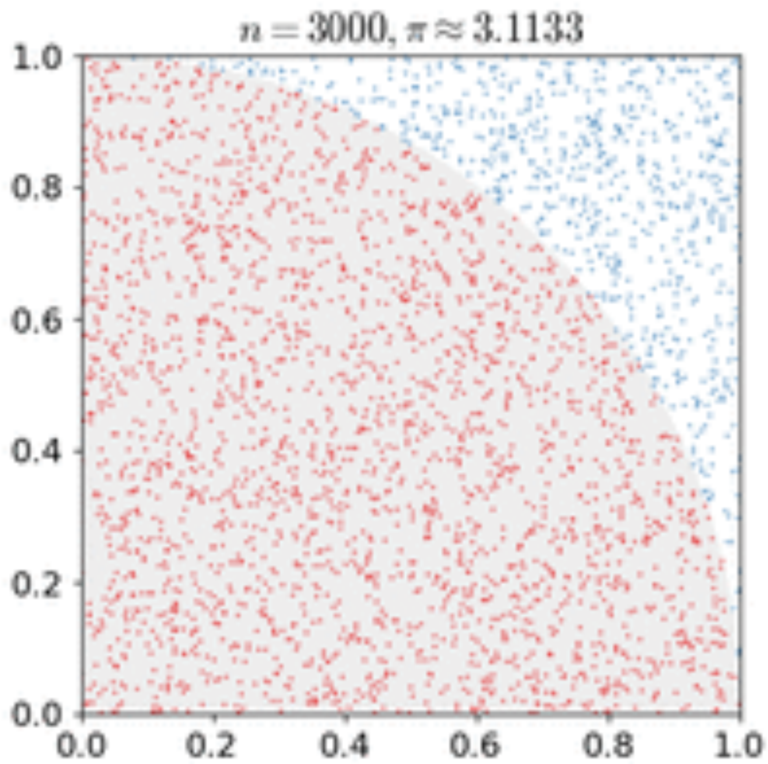


Monte Carlo algorithm

- Monte Carlo – European Las Vegas
- Stan Ullman: uncle going to Monte Carlo
- Metropolis: Monte Carlo algorithm
- High-dimensional integration or optimization
- Ironically, the most accurate algorithms may be based on randomness
- Avoid the curse of dimensionality
- population (e.g., 300 million) → sample (e.g., 1000 examples)

Metropolis algorithm: ranked #1 algorithm in scientific computing

- The first computers: MANIAC
- Manhattan project: atomic bomb, hydrogen bomb
- **Markov chain** Monte Carlo



$n = 3000, \pi \approx 3.1133$
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Foundation for statistics, machine learning, AI, etc.

Applications in many other areas and disciplines

Importance only comes after calculus and linear algebra