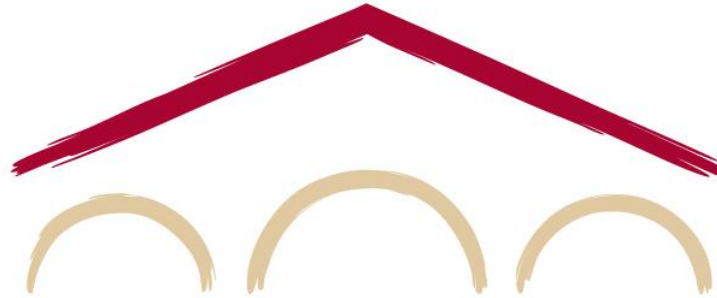


Natural Language Processing with Deep Learning

CS224N/Ling284



Diyi Yang & Yejin Choi

Lecture 1: Introduction and the History of NLP

Lecture Plan

Lecture 1: Introduction and the History of NLP

1. The course (10 mins)
2. The history of NLP: How we got here (65 mins)

Key learning today: **understand the paradigm shifts in NLP and how our assumptions about language shape what is possible in each era.**

Course logistics in brief

- Instructor: Diyi Yang, Yejin Choi
- Head TA: Julie Kallini
- Course Manager: John Cho
- TAs: Many wonderful people! See website
- Time: Tu/Th 4:30–5:50 Pacific time, Nvidia Aud. (→ video)
- Email list: **cs224n-win2526-staff@lists.stanford.edu**
- We've put a lot of other important information on the class webpage. Please read it!
 - <http://cs224n.stanford.edu/> a.k.a., <http://www.stanford.edu/class/cs224n/>
 - TAs, syllabus, help sessions/office hours, Ed (for all course questions/discussion)
 - Office hours start **Thursday!**
 - Python/numpy and then PyTorch tutorials: First two Fridays.
 - Slide PDFs uploaded before each lecture

Instructors



Diyi Yang



Yejin Choi

Course Staff



John Cho (Course Manager)



Swati Dube Batra
(Course Manager Advisor)

Teaching Assistants



Julie Kallini (Head TA)



Sarah Chen



Ali Sartaz Khan



Minsik Oh



Fang Wu



Ahmed Ahmed



Caroline Choi



Arpandeeep Khatua



Chenglei Si



David Anugraha



Advit Deepak



Alisa Levin



Mirac Suzgun



Luke Bailey



Nevin George



Shicheng Liu



Tristan Thrush

What do we hope to teach? (A.k.a. “learning goals”)

1. The foundations of the effective modern methods for deep learning applied to NLP
 - Basics first: word vectors, recurrent networks, attention, transformers
 - Then key methods used in NLP in 2025: [pretraining](#), [post-training](#), [efficient adaptation](#), [agents](#), [reasoning](#), [multilinguality](#), [multimodality](#), [interpretability](#), etc.
2. A big picture understanding of human languages and the difficulties in understanding and producing them via computers
3. An understanding of and **ability to build systems** for some of the major problems **involving language and computation**:
 - Word representations, question answering, fine-tuning LLMs, RAG, agentic systems and tool use, and LLMs evaluation

Course work and grading policy

- 4 x 1.5-week Assignments: 6% + 3 x 14%: 48%
 - **HW1 is released today! Due next Tuesday! At 4:30 p.m.**
 - Submitted to Gradescope in Canvas (i.e., using @stanford.edu email for your Gradescope account)
- Final Default or Custom Course Project (1–3 people): 49%
 - Project proposal: 8%, milestone: 6%, poster or web summary: 3%, report: 32%
- Participation: 3%
 - Guest lecture reactions, Ed, course evals, karma – see website!
- Late day policy
 - 6 free late days; afterwards, 1% off course grade per day late
 - Assignments not accepted more than 3 days late per assignment unless given permission in advance

Course work and grading policy

- Collaboration policy:
 - Please read the website and the **Honor Code**! Understand allowed collaboration and how to document it: Don't take code off the web; acknowledge working with other students; write your own assignment solutions
 - Students must independently submit their solutions to CS224N homework
- AI tools policy
 - Large language models are great, but we don't want ChatGPT's solutions to our assignment
 - Collaboration with AI tools is allowed; asking it to answer questions is strictly prohibited
 - Employing AI tools to substantially complete assignments will be considered a violation of the Honor Code (see Generative AI Policy Guidance [here](#) for more details)

High-Level Plan for Assignments (to be completed individually!)

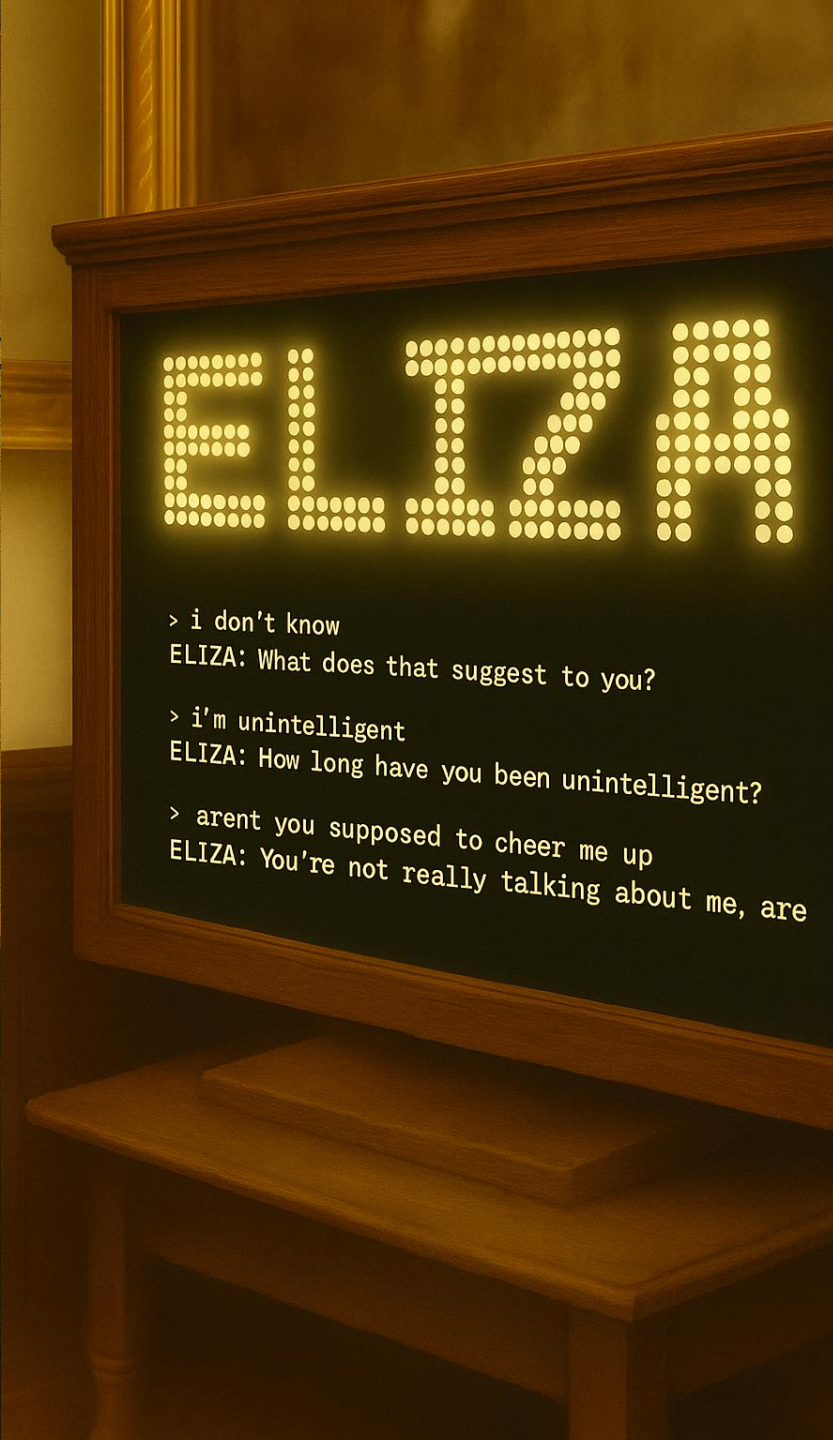
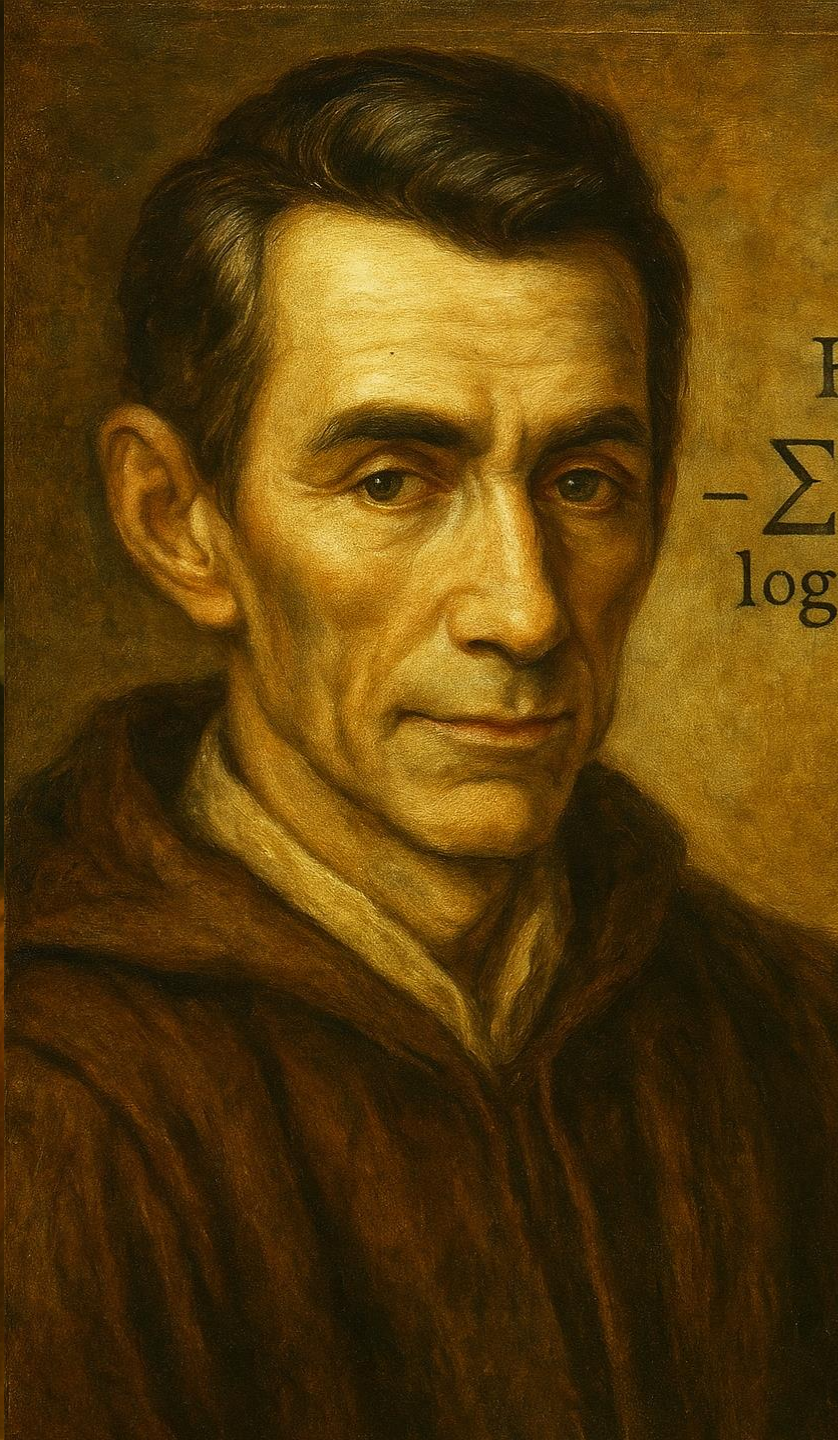
- Hw1 is hopefully an easy on ramp – a Jupyter/IPython Notebook
- Hw2 covers neural network foundations and calculating tensor derivatives, and you build a small network for an NLP task around dependency parsing
- Hw3 is about coding a transformer from scratch and understanding attention
- Hw4 focuses on LLMs evaluation and redteaming
- For Final Project, more details presented later, but you either:
 - Do the default project
 - You implement a GPT and then fine-tune and adapt it for downstream tasks
 - Open-ended but an easier start; a good choice for many
 - Propose a custom final project, which we approve
 - You will receive feedback from a **mentor** (TA/prof/postdoc/PhD)
 - Can work in teams of 1–3; can use any language/packages

Lecture Plan

Lecture 1: Introduction and Word Vectors

1. The course (10 mins)
2. The history of NLP: How we got here (65 mins)
 - A **creative** way of looking at the history of NLP
Credit to Stanford NLP group; check out Gates 3B for the exhibition

Key learning today: **understand the paradigm shifts in NLP and how our assumptions about language shape what is possible in each era.**





I gave the man a book

P O R
I → ATRANS ← book
From

N=1

This is a sentence

$$P(\text{this, is, a, sentence}) = P(\text{this})P(\text{is})P(\text{a})P(\text{sentence})$$

N=2

This is a sentence

$$P(\text{this, is, a, sentence}) = P(\text{this})P(\text{is} | \text{this})P(\text{a} | \text{is})P(\text{sentence} | \text{a})$$

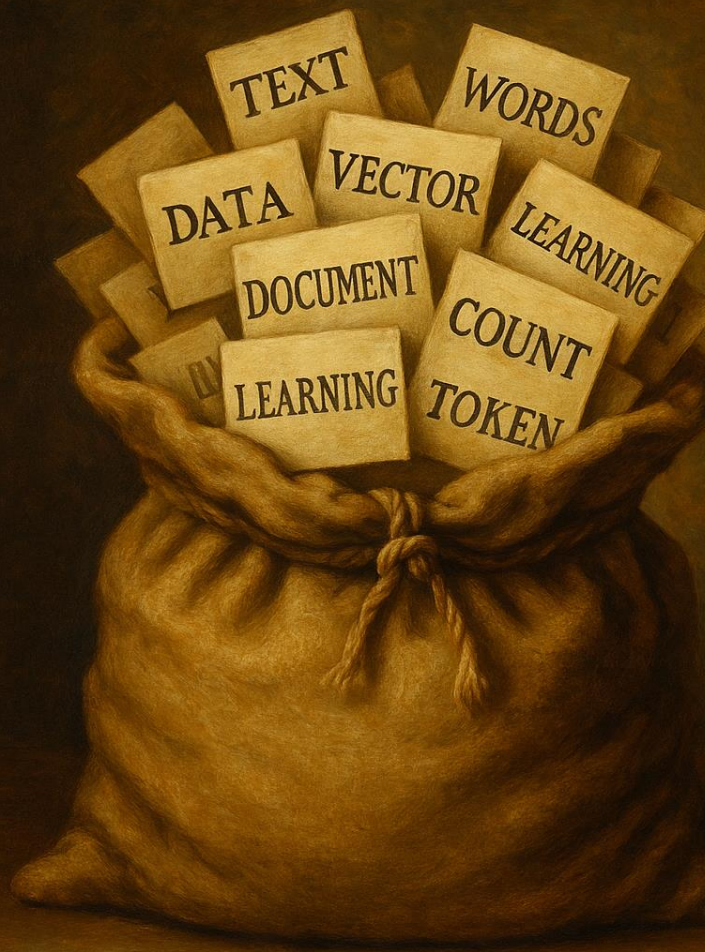
N=3

This is a sentence

$$P(\text{this, is, a, sentence}) = P(\text{this})P(\text{is} | \text{this})P(\text{a} | \text{this, is})P(\text{sentence} | \text{is, a})$$

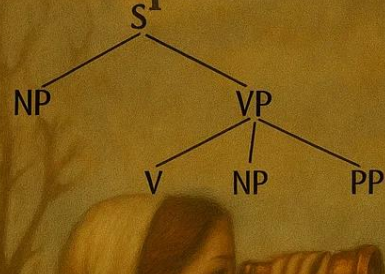
...

BAG OF WORDS

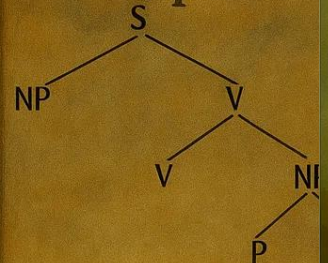


She spotted the squirrel with binoculars

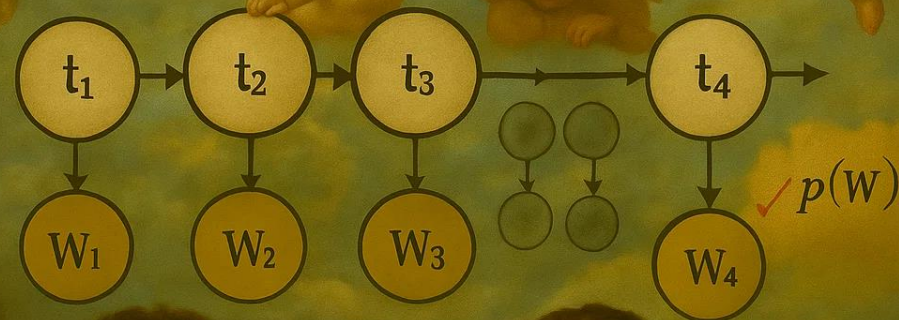
High-probability parse

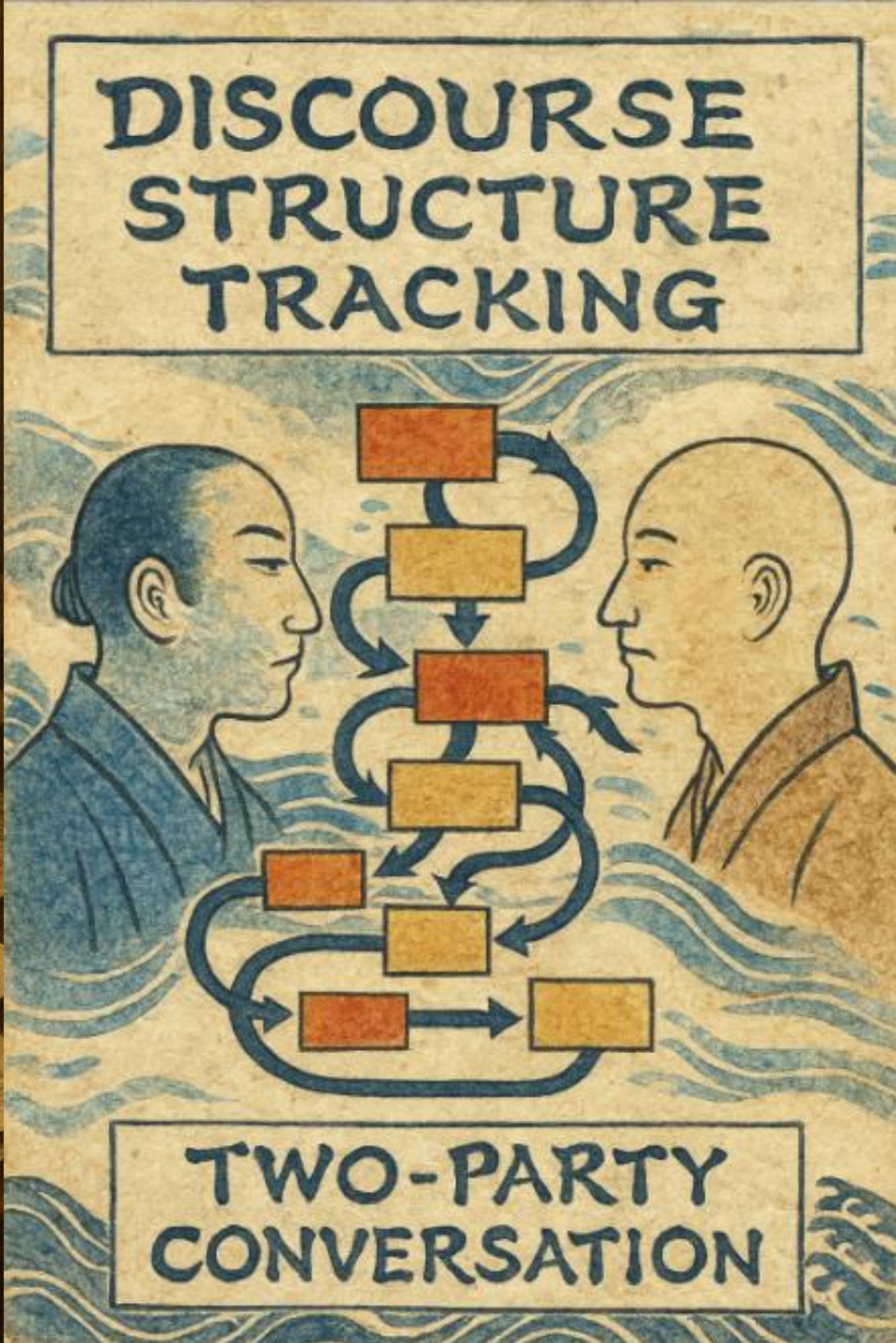
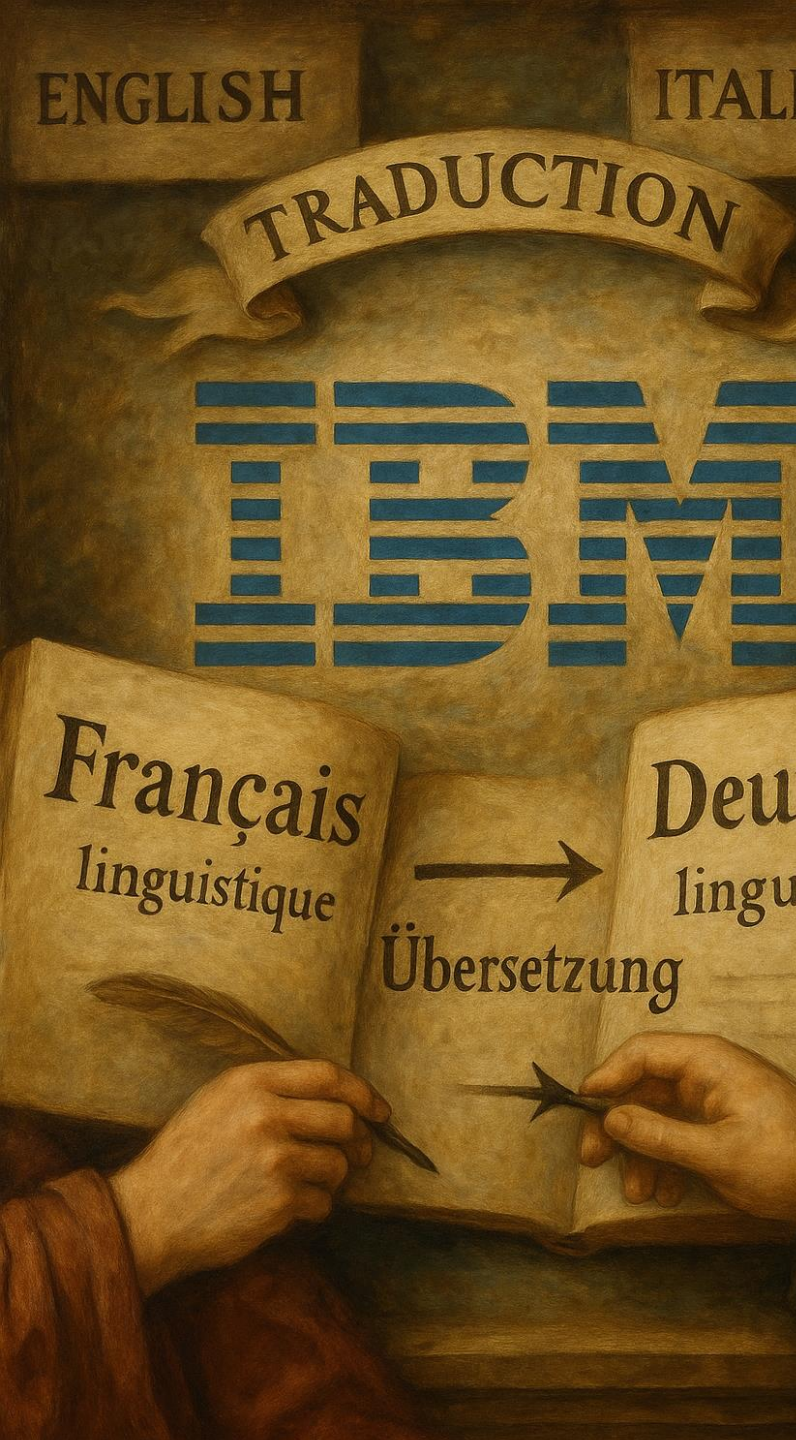


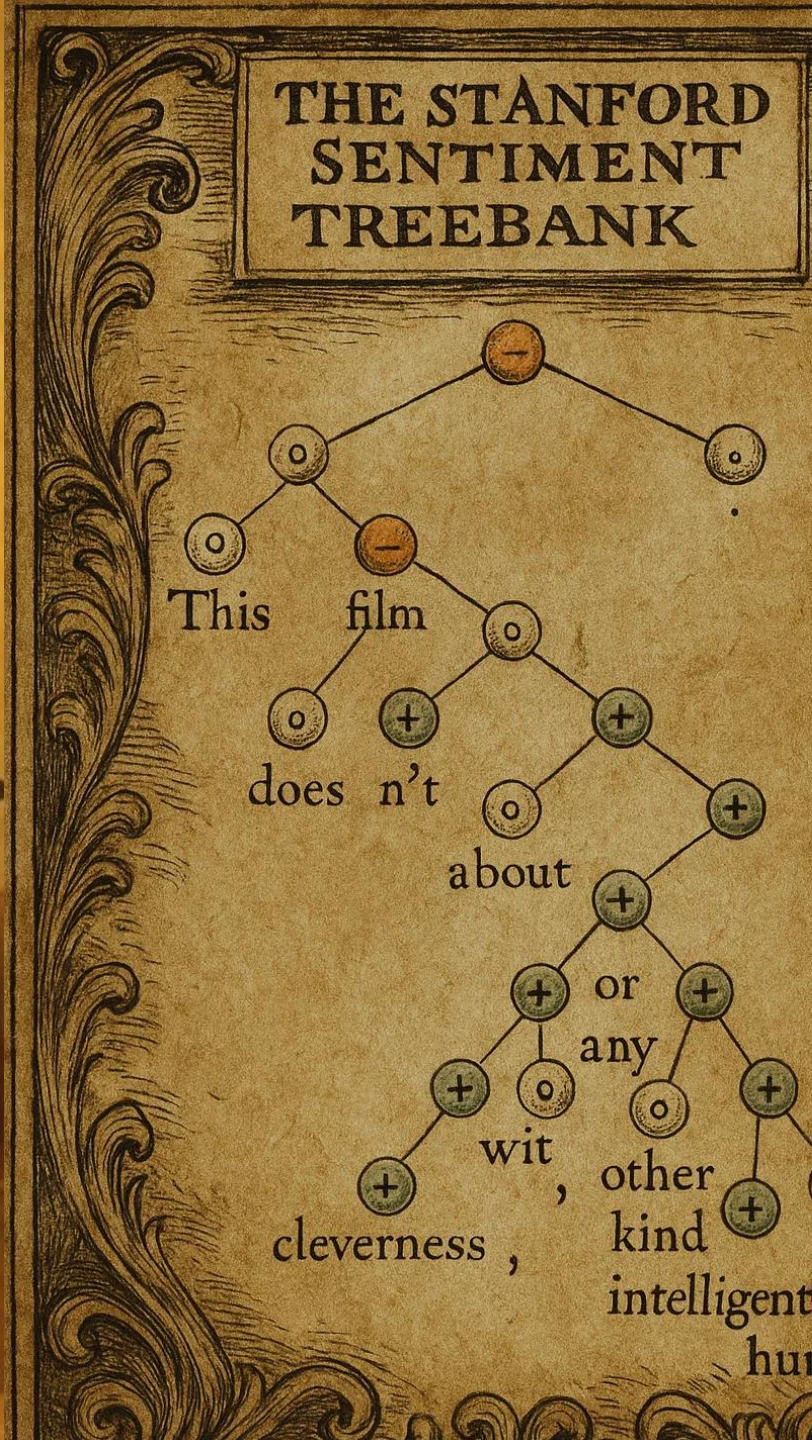
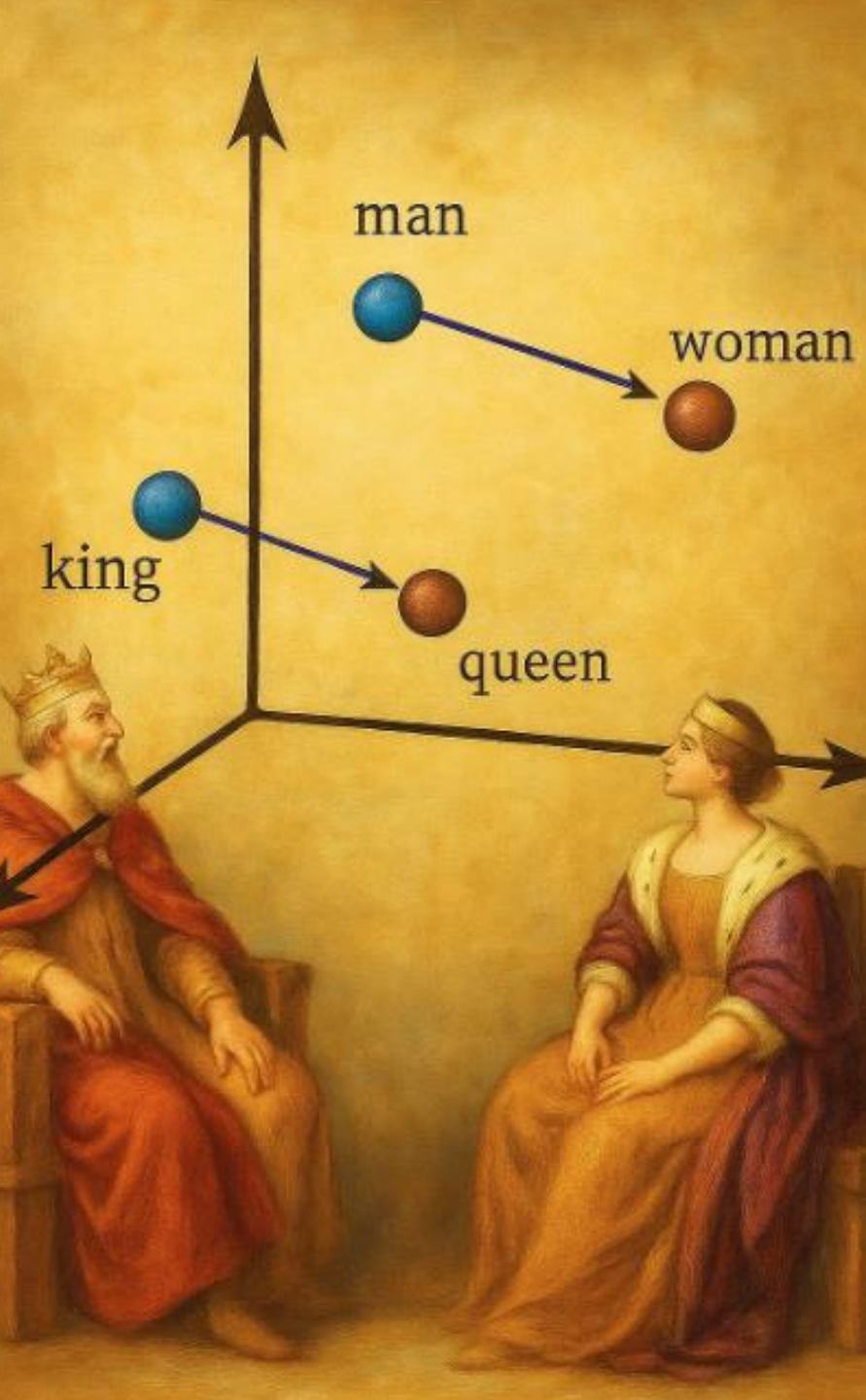
Low-probability parse



Hidden Markov Model







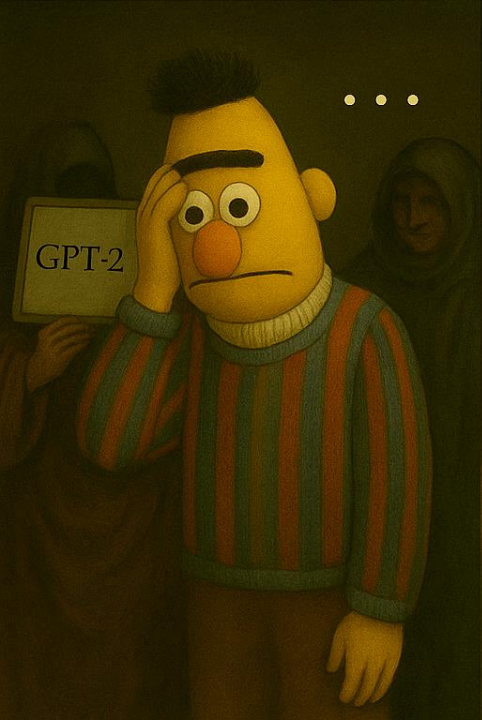




BIRTH



DECLINE



...

ChatGPT



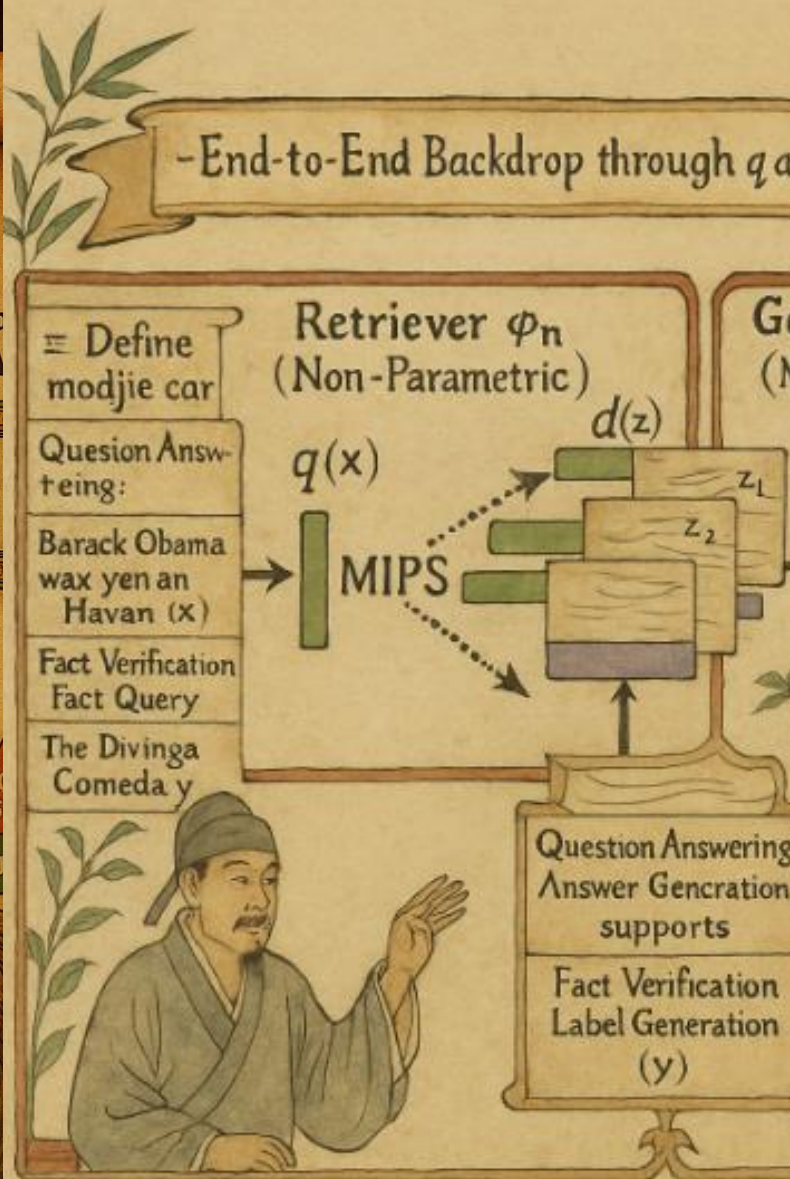
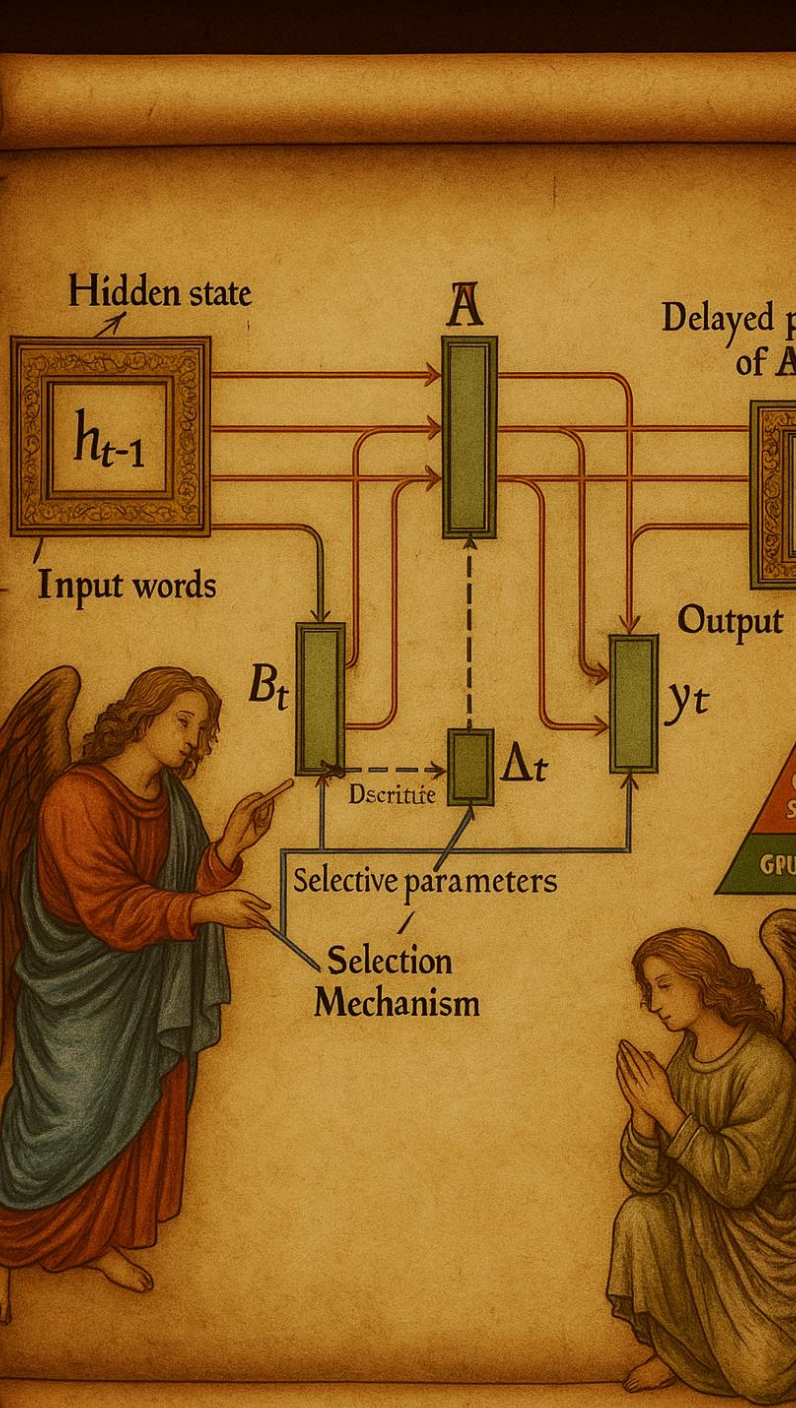
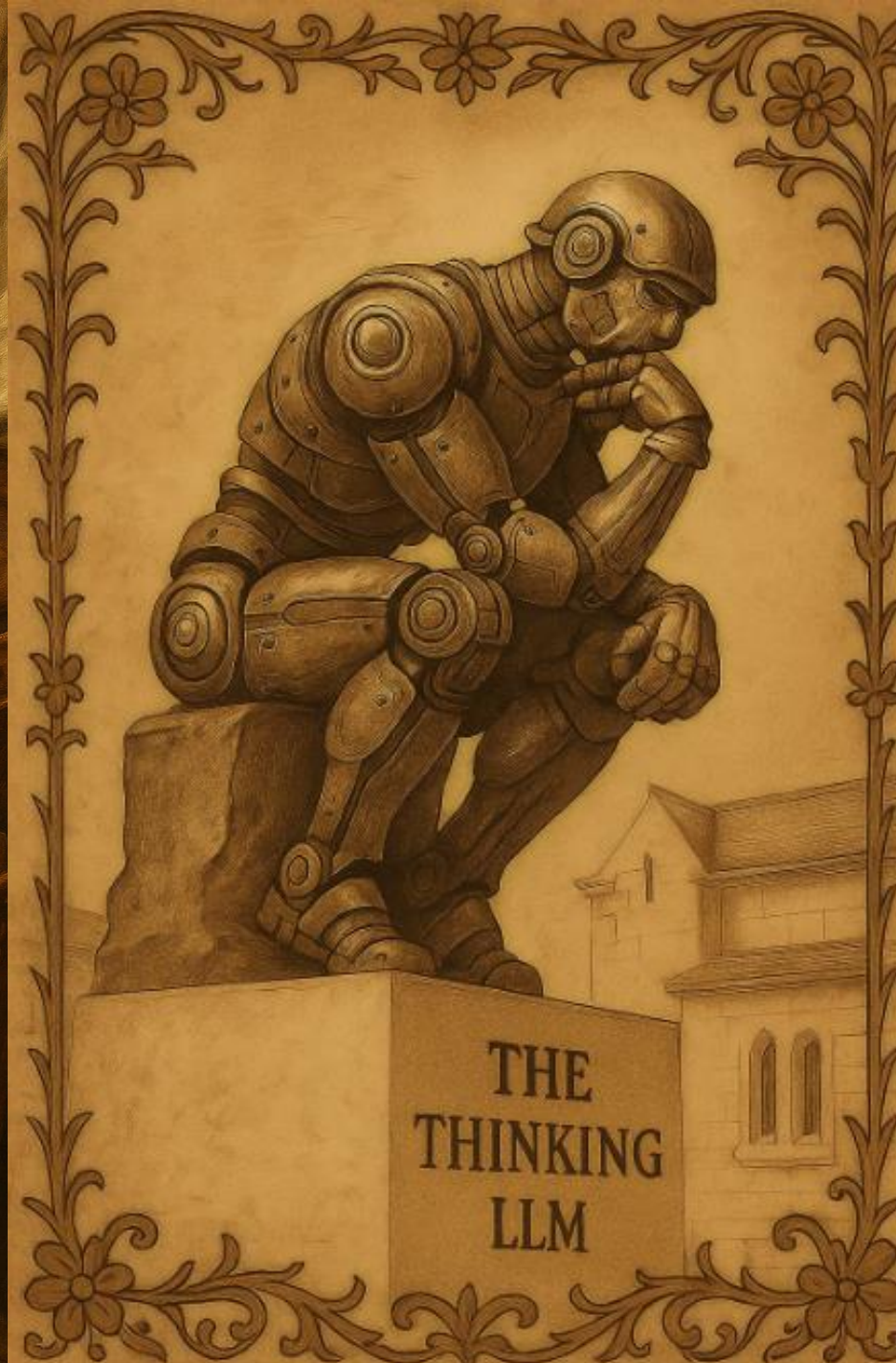
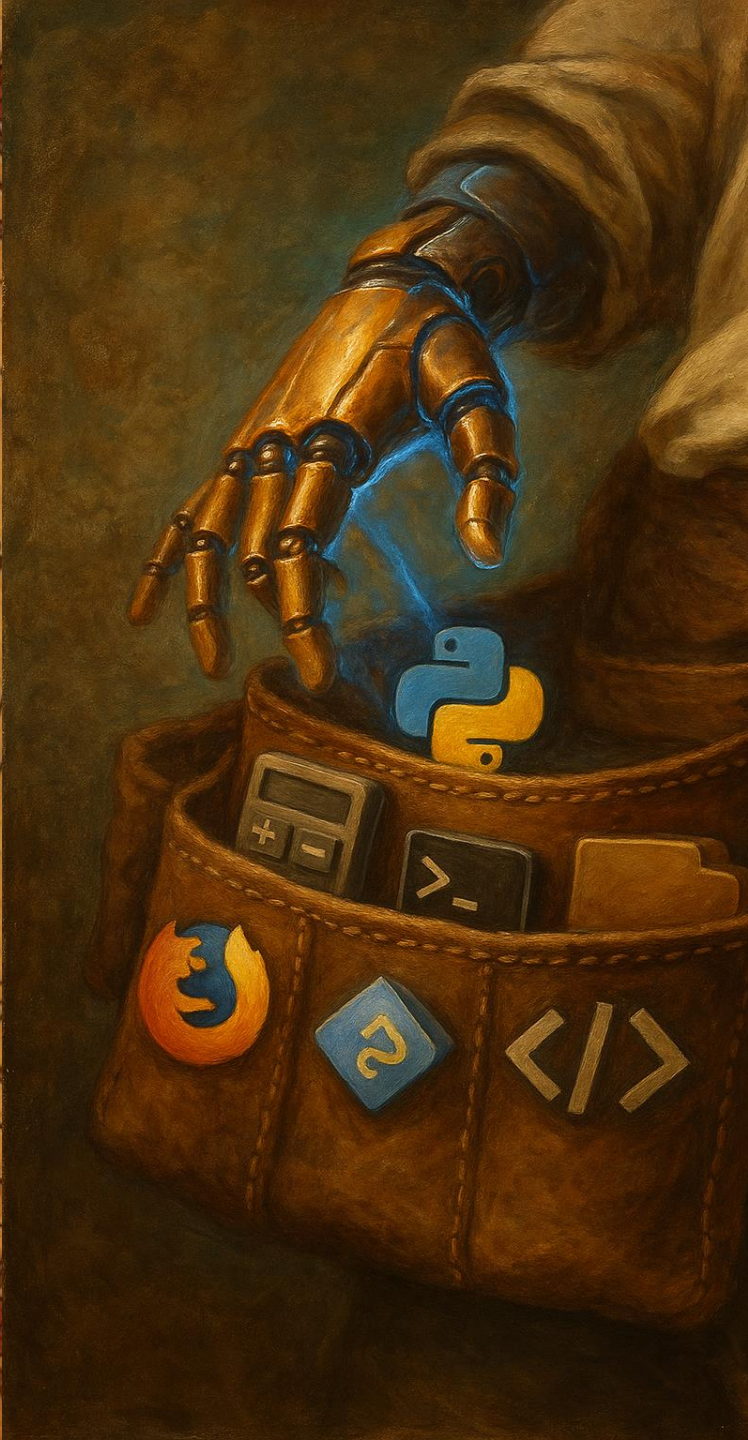
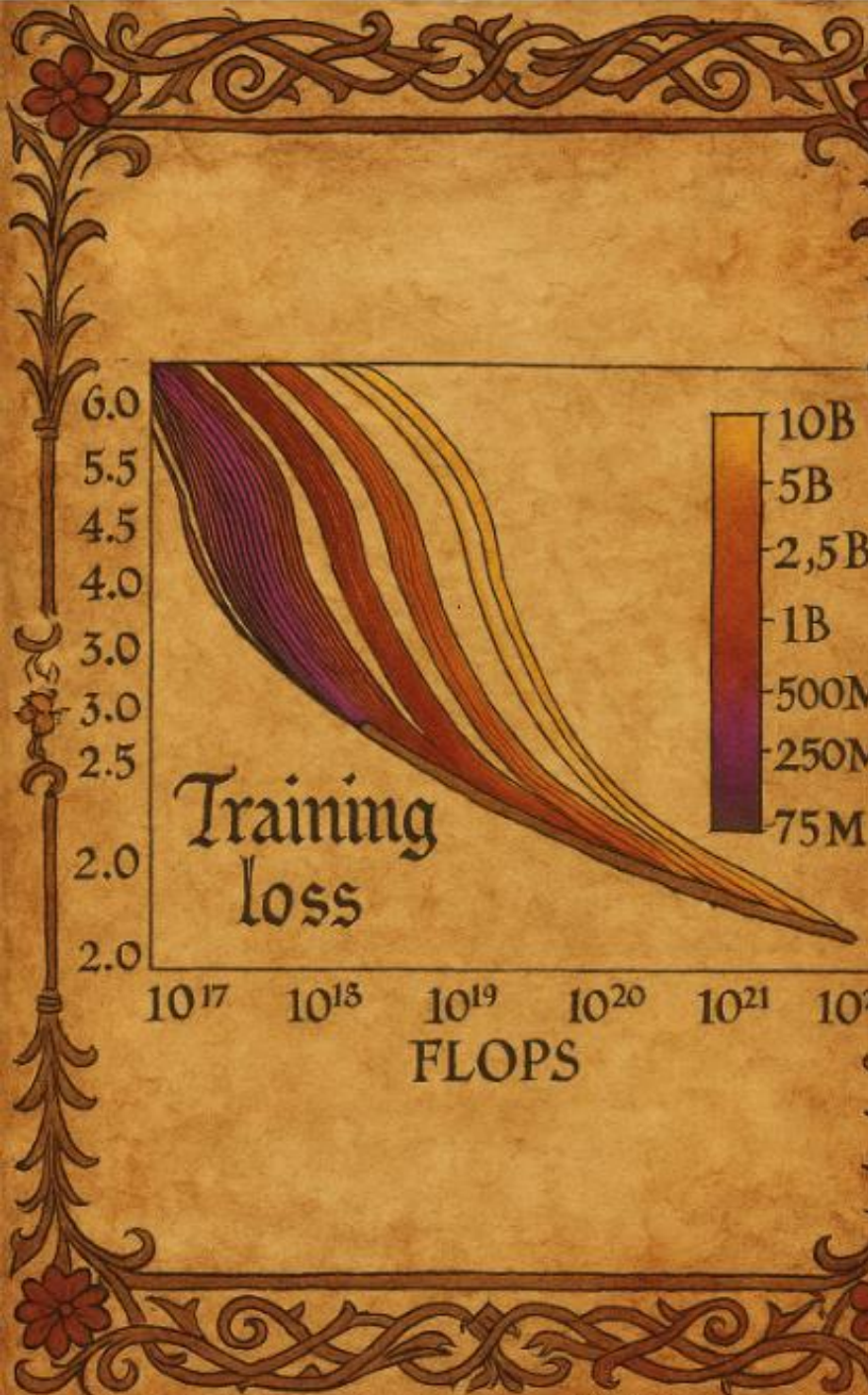


Figure 1: Overview of our approach, We co-train retriever (Query Encoder + Document Encoder) and generator (Answer Encoder + Answer Decoder) end-to-end. For query x , we use-Maximum Inner Product Search (MIPS) to find the top- K documents.





Lecture Plan

Lecture 1: Introduction and Word Vectors

1. The course (10 mins)
2. The history of NLP: How we got here (65 mins)
 - A creative way of looking at the history of NLP (5 mins)
Credit to Stanford NLP group; check out Gates 3B for the exhibition
 - **A scientific way of looking at the history of NLP (60 mins)**

Key learning today: **understand the paradigm shifts in NLP and how our assumptions about language shape what is possible in each era.**

if you can only invite one professor to talk about the history of NLP, who should we invite?

If you can invite **only one professor** to speak about the history of NLP, the single choice is:

Chris Manning (Stanford)

Why him:

- He uniquely bridges **multiple eras of NLP**:
 - symbolic / linguistic NLP
 - statistical NLP
 - neural and deep learning–based NLP
 - modern foundation models
- He was central to the field's **modernization** (from the 2000s, then neural NLP).
- He can tell the history **from inside the room**,
 - Co-author of *Foundations of Statistical Natural Language Processing* (a canonical text for a generation).
 - Key figure in Stanford NLP and the broader NLP community.
- He's excellent at **historical narrative + concrete details**—ideal for a "history of NLP" talk.

if you can only invite one professor to talk about the history of NLP, who should we invite?



If you are looking for the definitive voice to bridge the gap between "old school" linguistic and the modern era of Deep Learning, there is one standout choice: **Professor Christopher Manning** of Stanford University.

While there are many pioneers—like Geoffrey Hinton for Neural Networks or Yoshua Bengio for Deep Learning—Manning is arguably the most essential figure specifically for the **history and evolution of Natural Language Processing (NLP)**.

Why Professor Christopher Manning?

- **The Bridge Builder:** He literally wrote the book on both eras of NLP. He co-authored *Foundations of Statistical Natural Language Processing* (the "bible" of the 1990s) and has since been a leader in the shift toward **Deep Learning** and **Transformers**.
- **The Linguistic Perspective:** Unlike many computer scientists who treat language as "data points," Manning is a trained linguist. He provides a unique historical perspective on how our understanding of syntax and grammar has informed (and sometimes clarified) modern AI.
- **The Stanford NLP Group:** As the leader of one of the world's most influential research groups, he oversaw the development of **GloVe** (word embeddings) and early work on **Attention mechanisms**, which paved the way for ChatGPT.



if you can only invite one professor to talk about the history of NLP, who should we invite?



Excellent question. If you can only invite **one professor** to give a *comprehensive, engaging, and authoritative* talk on the history of NLP, my strong recommendation would be:

Professor Christopher D. Manning

(Stanford University)

Here's why he is the ideal single choice:

1. Span of Historical Eras

Manning has not only studied but actively **shaped** multiple eras of NLP:

- **Early & Symbolic Era:** His early work on computational linguistics and parsing connects him to pre-statistical foundations.
- **Statistical Revolution (1990s-2000s):** He co-authored the seminal textbook *"Foundations of Statistical Natural Language Processing"* (1999) with Hinrich Schütze. This book defined the field's shift to statistical methods.
- **Deep Learning/Neural Revolution (2010s-Present):** He led groundbreaking work (e.g., GloVe embeddings, TreeRNNs, Stanford NLP tools) and co-authored the

Guest Lecture by Prof. Chris Manning, the creator of CS224N



The History of NLP