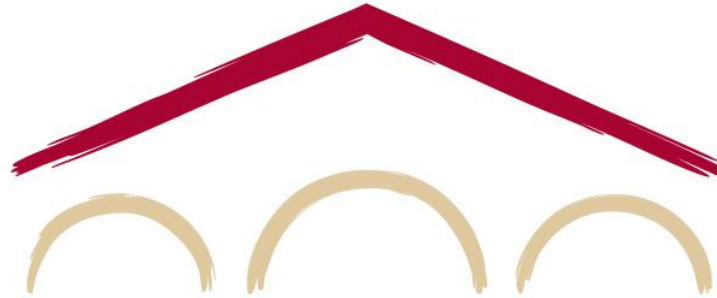


Natural Language Processing with Deep Learning

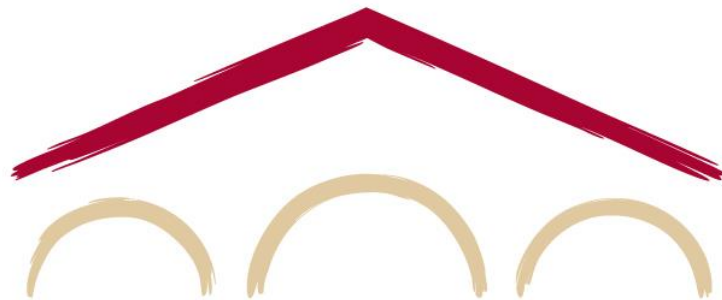
CS224N/Ling284



Diyi Yang & Yejin Choi

Lecture 1: Introduction and the History of NLP

Natural Language Processing 与 Deep Learning CS224N/Ling284



Diyi Yang & Yejin Choi

Lecture 1 : 导论与 NLP 发展史

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.**

课程计划

Lecture 1 : 导论与 NLP 发展史

1. 课程介绍 (10 分钟)
2. NLP 的历史 : 我们如何走到今天 (65 分钟)

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

课程后勤简介

- 讲师：D i y i Yang , Y e j i n C h o i
- 首席 TA：J u l i e K a l l i n i
- 课程经理：J o h n C h o
- 助教：很多优秀的人！详见官网
- 时间：周二 / 周四 4 : 3 0 - 5 : 5 0 太平洋时间, 在 Nvidia 报告厅 (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/>
 - 助教、教学大纲、答疑 / o f f i c e h o u r s、E d (用于所有课程提问 / 讨论)
 - Office hours start **Thursday!**
 - P y t h o n / n u m p y 和 P y T o r c h 教程：前两个周五
 - 每节课前上传幻灯片 P D F

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

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

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

1. 深度学习应用于 NLP 的现代有效方法的基础
 - 基础优先：word vectors、recurrent networks、attention
 - 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 以及通过计算机生成语言
3. An understanding of and **ability to build systems** for some of the major problems 涉及语言与计算：
 - 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

课程作业与评分政策

- 4 x 1.5-week Assignments: 6% + 3 x 14%: 48%
 - HW 1 今天发布！下周二截止！下午 4 : 3 0
 - 提交到 Canvas 中的 Gradescope (即使用 @stanford.edu 邮箱的 G
- 期末默认或自定义课程项目 (1 - 3 人) : 4 9 %
 - 项目提案 : 8 % , 里程碑 : 6 % , 海报或网页摘要 : 3 % , 报告 : 3 2 %
- 课堂参与 : 3 %
 - 客座讲座反馈、Ed 论坛、课程评价、karma — 详见官网！
- 迟交政策
 - 6 天免费延期 ; 之后每迟交一天扣总成绩 1 %
 - 每次作业迟交不得超过 3 天 , 除非提前获得许可

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)

课程作业与评分政策

- Collaboration policy:
 - Please read the website and the [Honor Code](#) (荣誉准则) to understand allowed collaboration and how to document it: Don't take code off the web; acknowledge working with other students; write your own 作业解答
 - 学生必须独立提交 [CS 224N](#) 作业解答
- AI 工具使用政策
 - Large language models are great, but we don't want ChatGPT's solutions to our assignment
 - 允许使用 AI 工具协作；严禁让 AI 直接回答作业问题
 - Employing AI tools to substantially complete assignments will be considered a violation of the Honor Code (参见 [Generative AI](#) 政策指引 for 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

作业整体规划（需独立完成！）

- HW 1 是一个轻松的入门作业 — Jupyter / IPython Notebook
- Hw2 covers neural network foundations and calculating tensor derivatives, and you 围绕依存句法分析构建一个小型网络来完成 NLP 任务
- HW 3 是从零编写 Transformer 并理解 attention 机制
- Hw4 focuses on LLMs evaluation and redteaming
- 期末项目的更多细节将在后续介绍，你可以选择：
 - 选择默认项目
 - 你将实现一个 GPT，然后对其进行微调和适配以完成下游任务
 - 开放式但起步较容易；对许多人是不错的选择
 - 提出自定义期末项目，经我们批准
 - You will receive feedback from a **mentor** (TA/prof/postdoc/PhD)
- 可组建 1 - 3 人团队；可使用任何语言 / 库

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

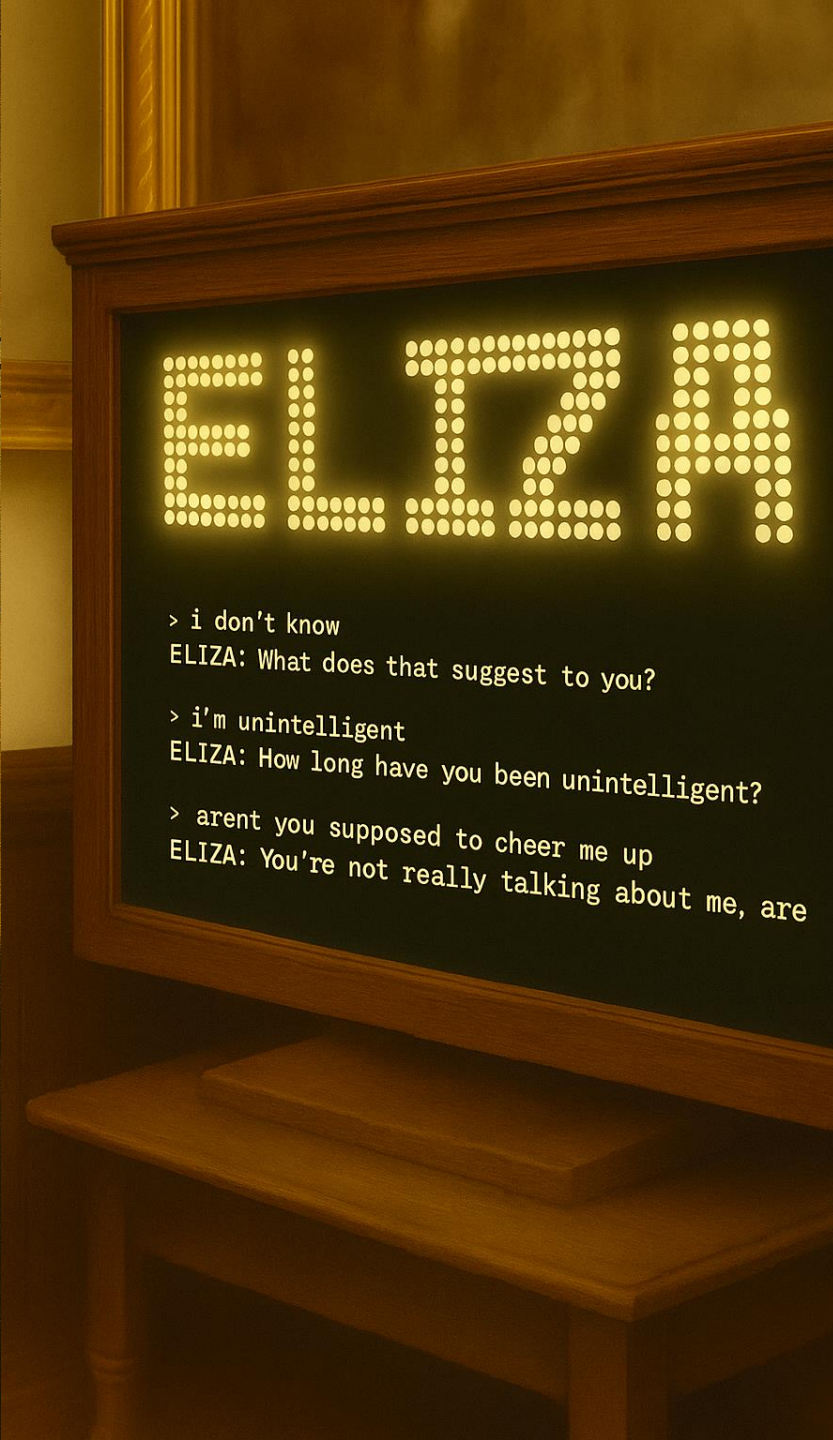
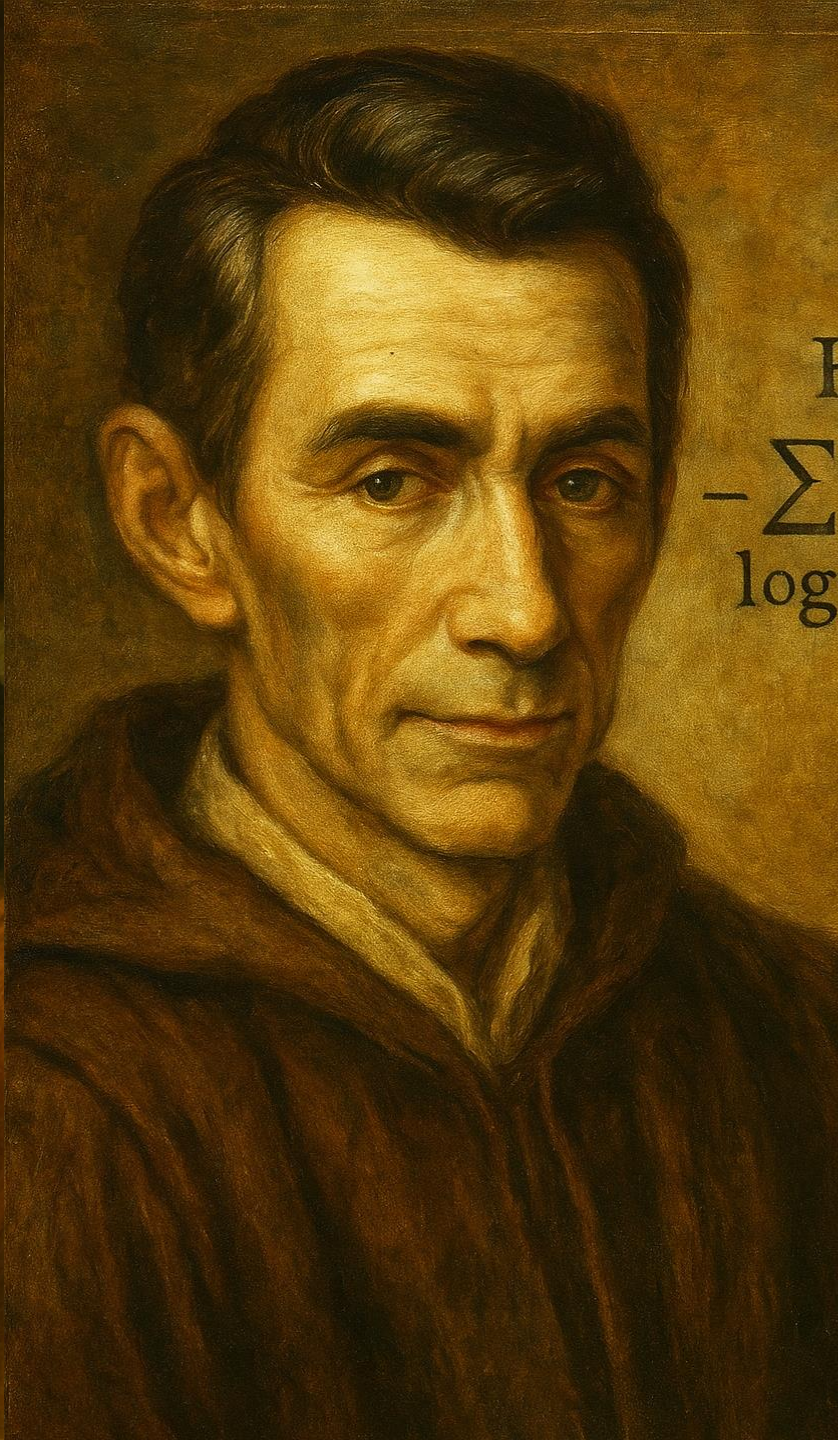
Key learning today: **understand the paradigm shifts in NLP and how our assumptions about language shape what is possible in each era.**

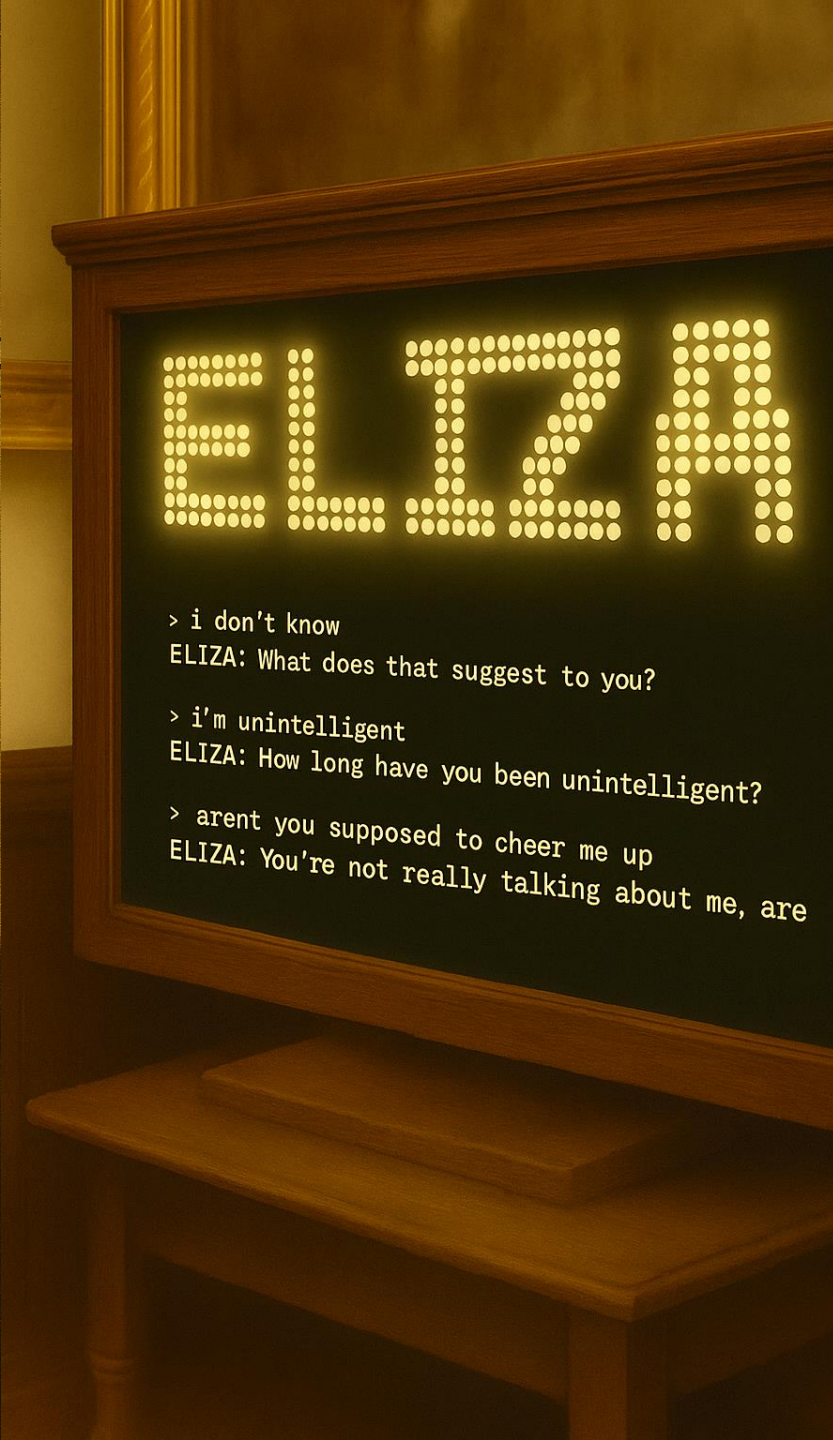
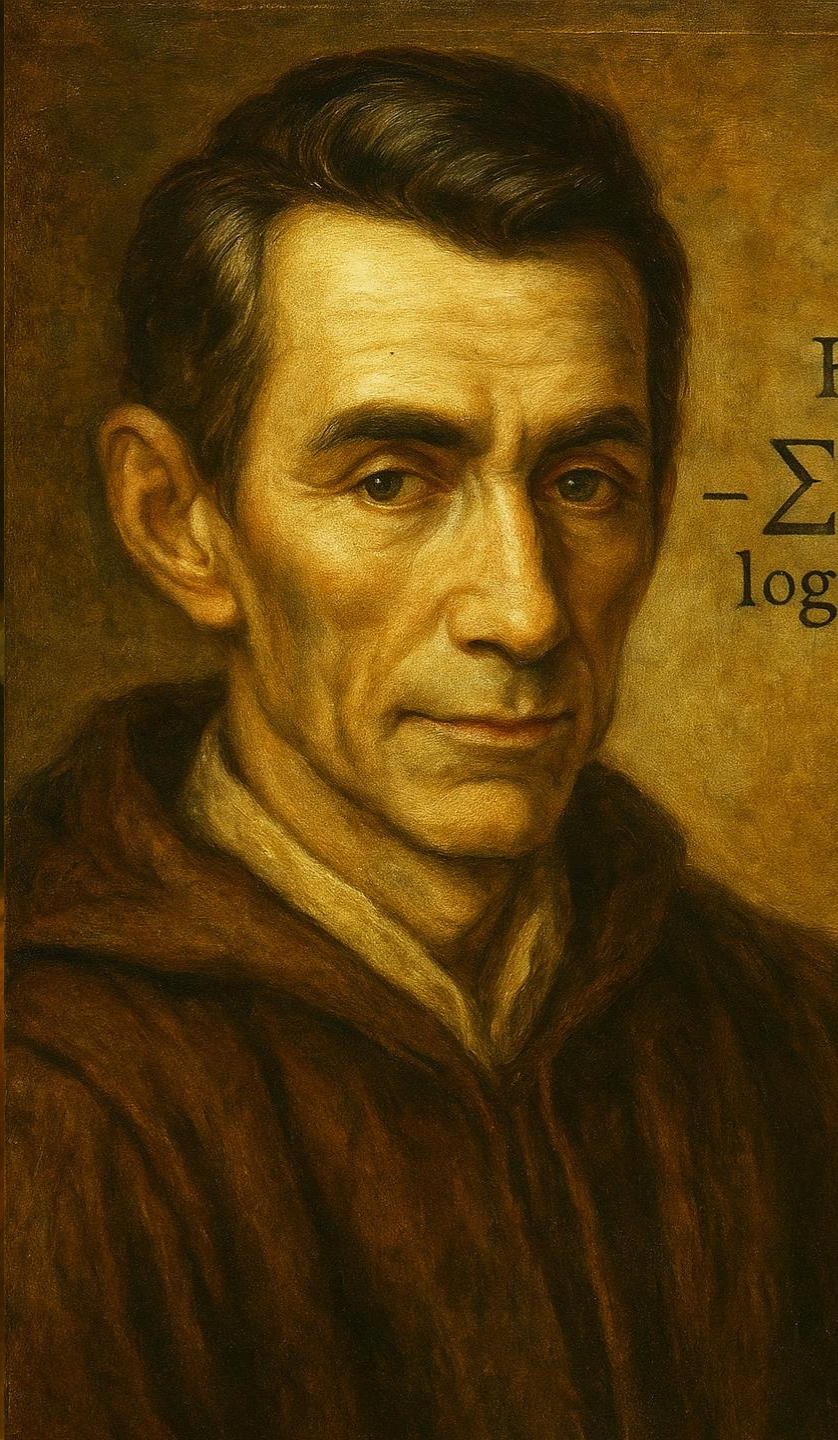
课程计划

Lecture 1 : 导论与 Word Vectors

1. 课程介绍 (10 分钟)
2. NLP 的历史 : 我们如何走到今天 (65 分钟)
 - A **creative** 回顾 NLP 历史的视角
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

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

...

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

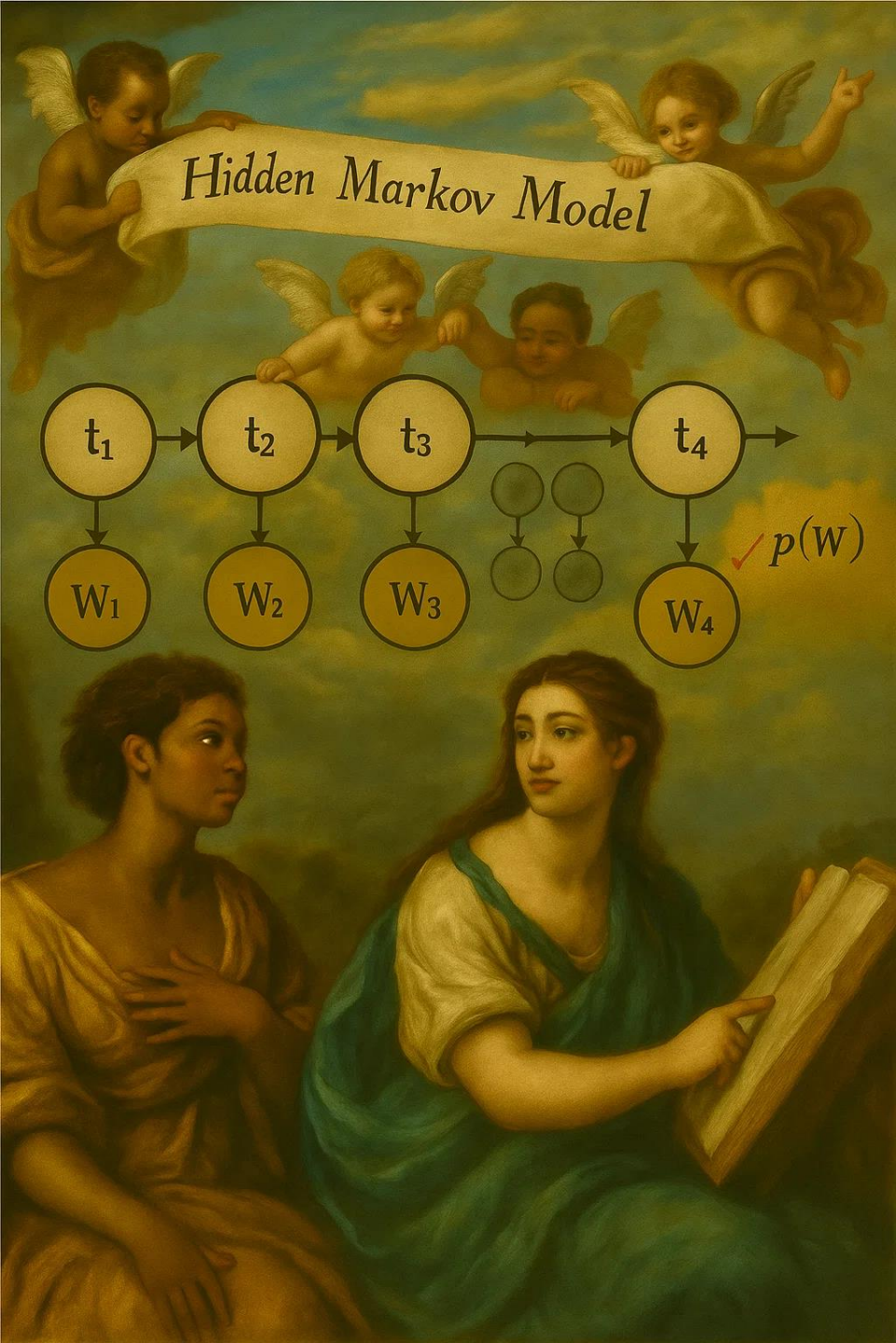
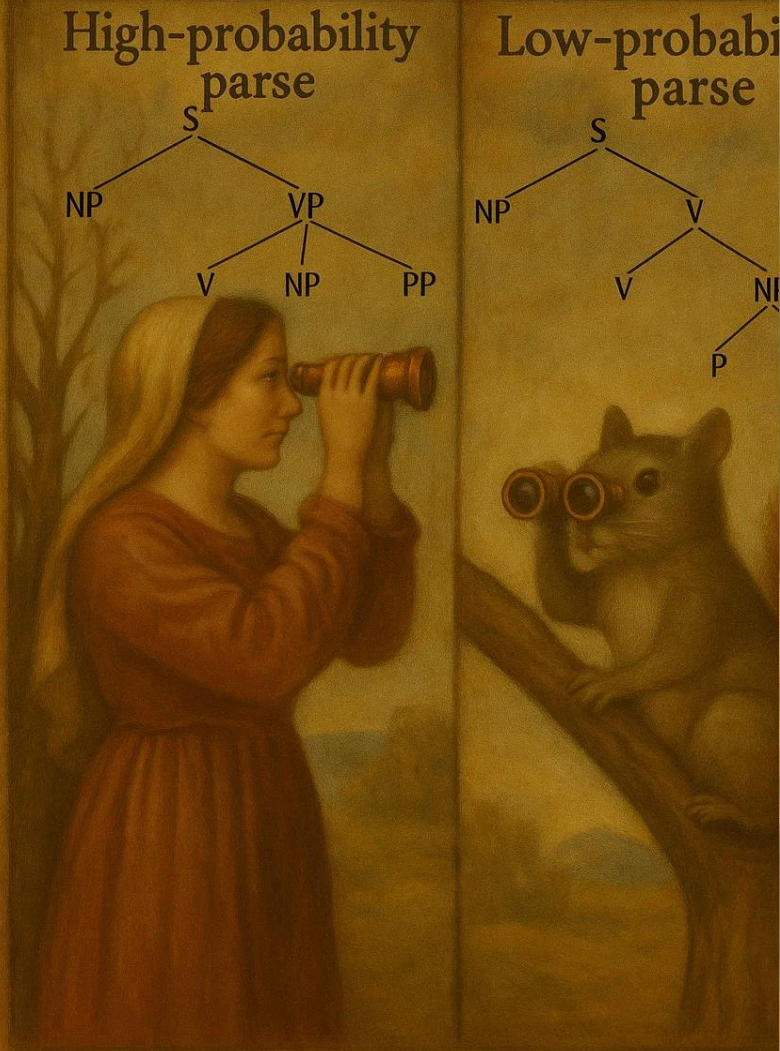
$$P(\text{this, is, a, sentence}) = P(\text{this})P(\text{is} | \text{this})P(\text{a} | \text{this, is})P(\text{sentence} | \text{is, a})$$

...

BAG OF WORDS



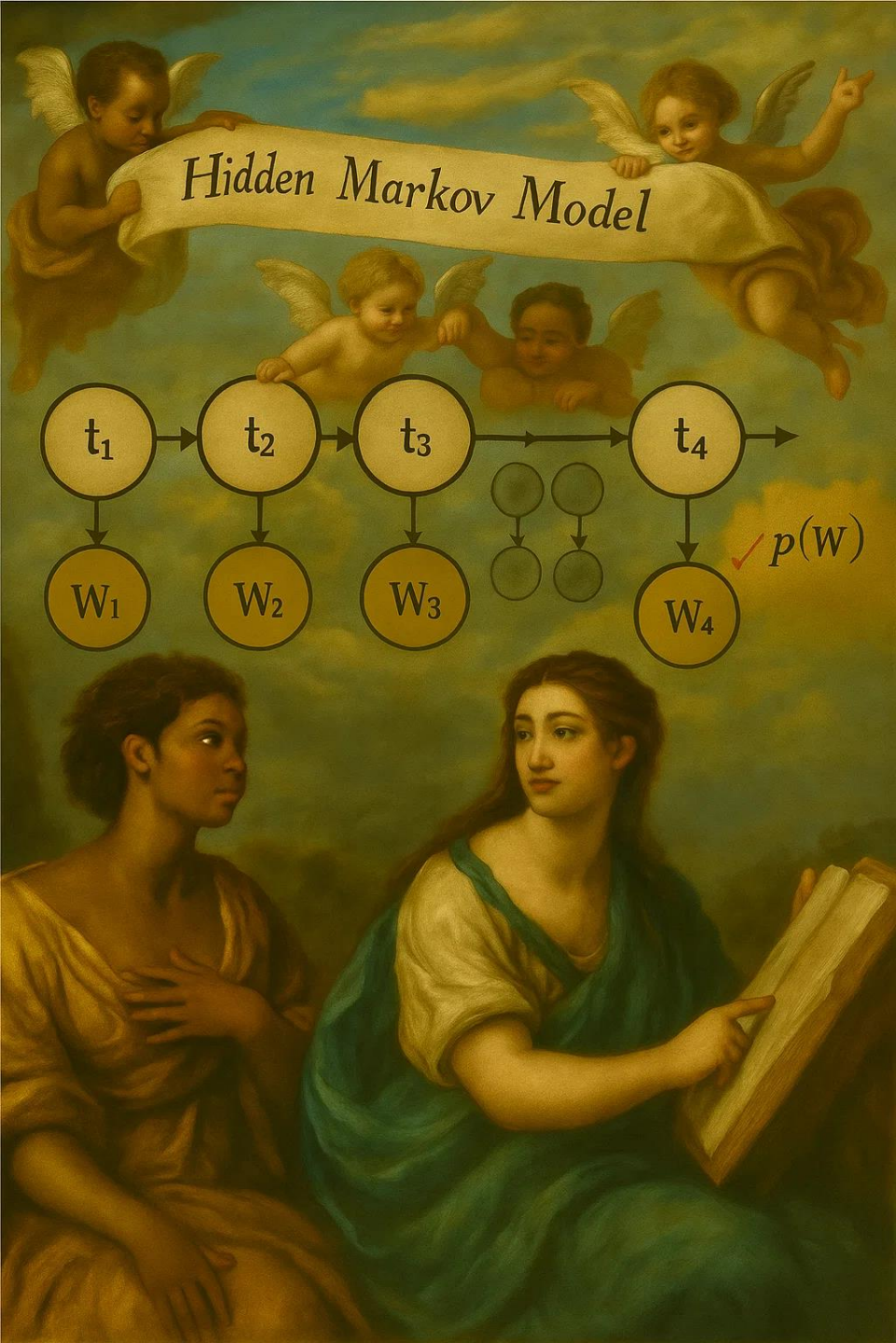
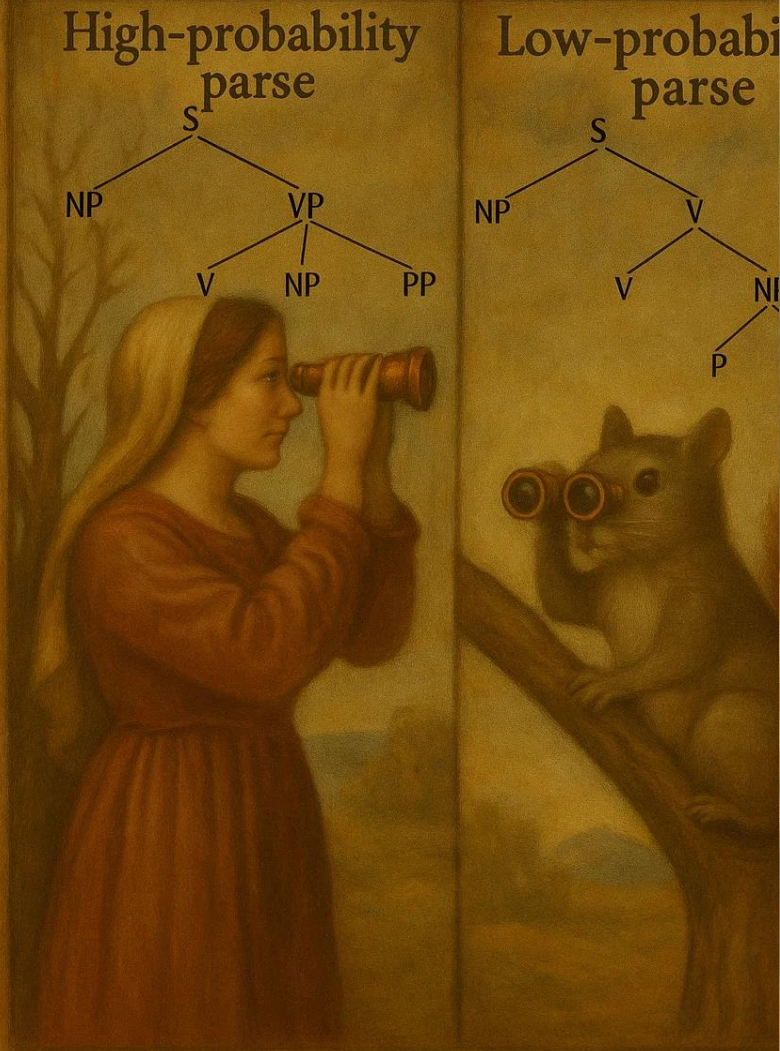
She spotted the squirrel with binoculars

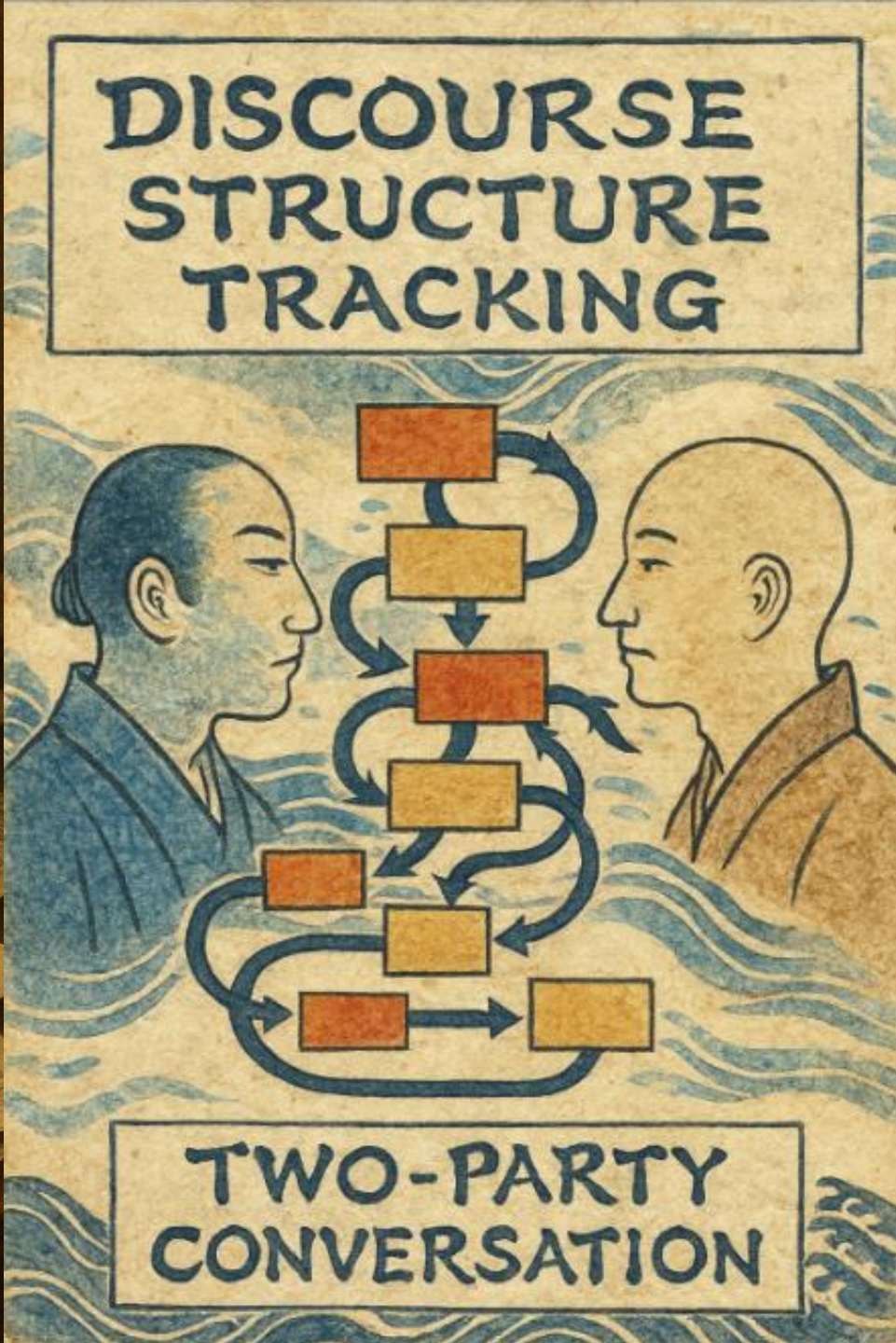
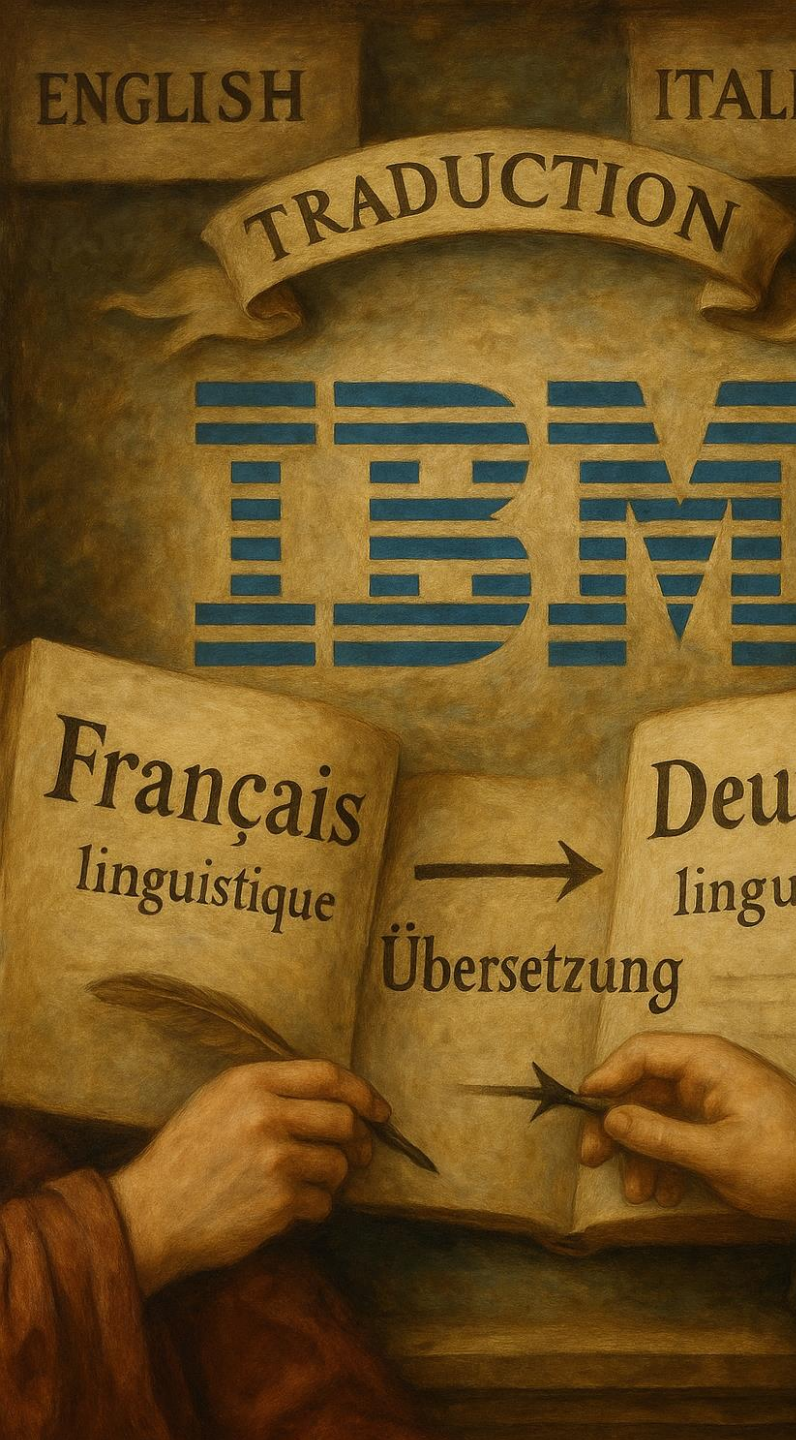


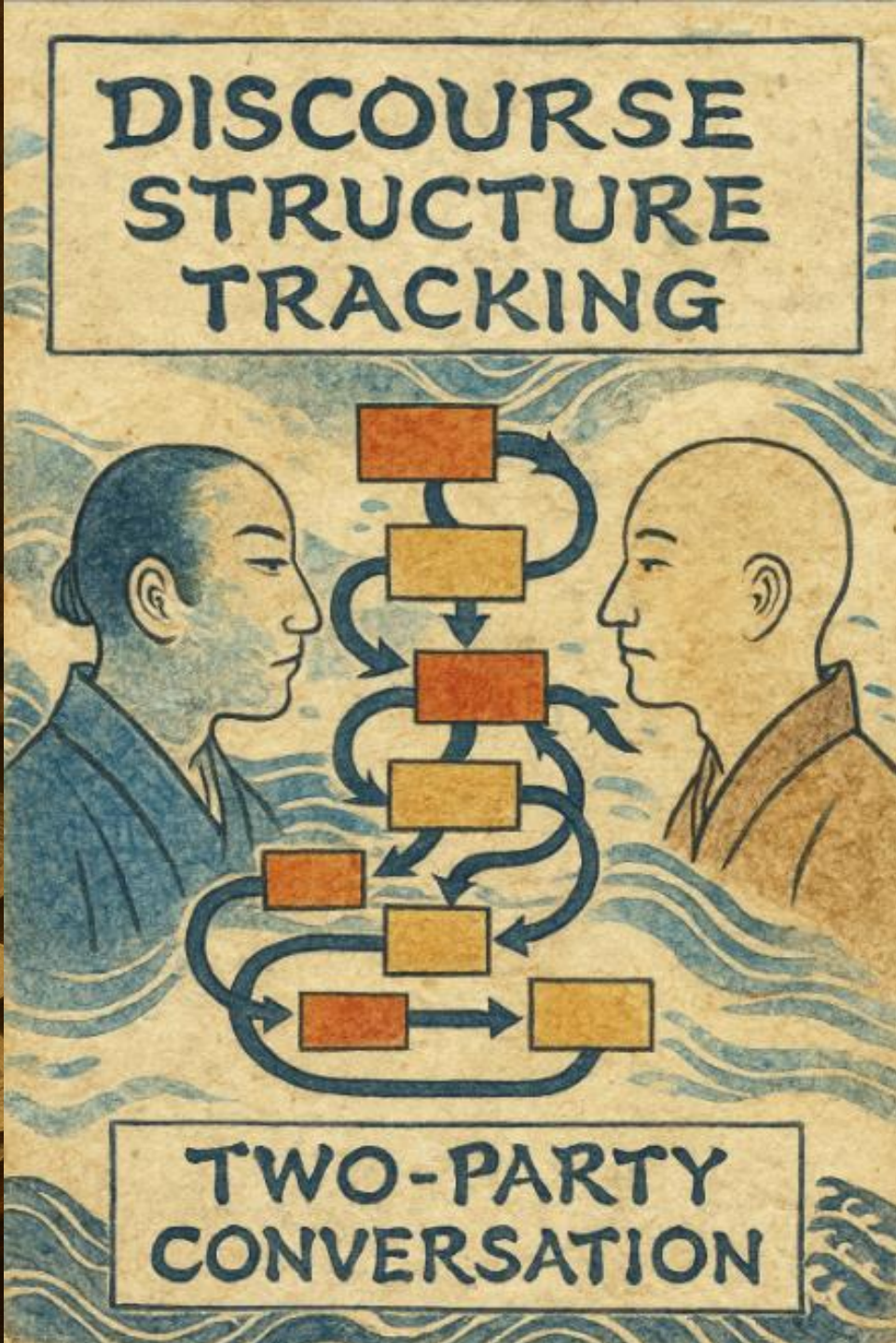
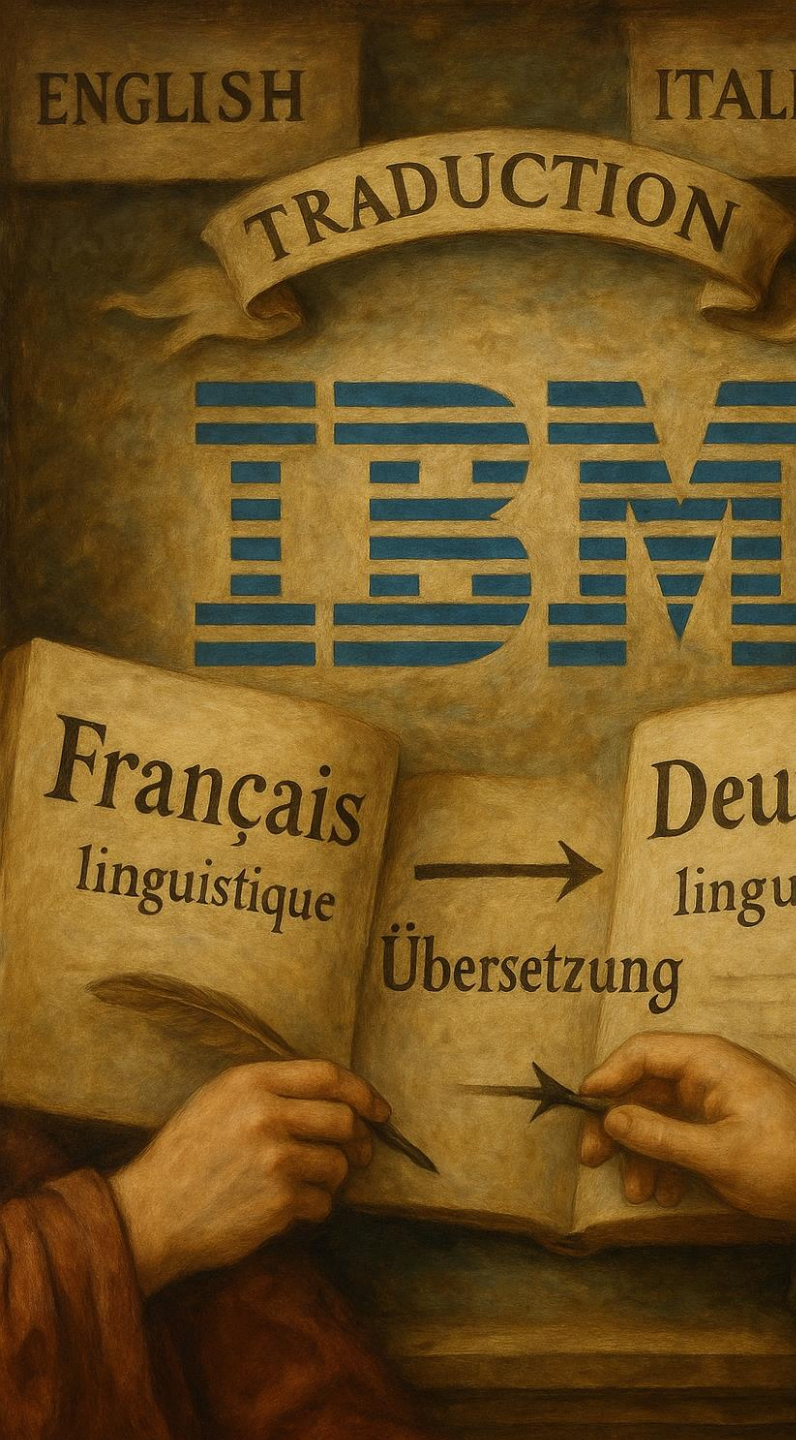
BAG OF WORDS

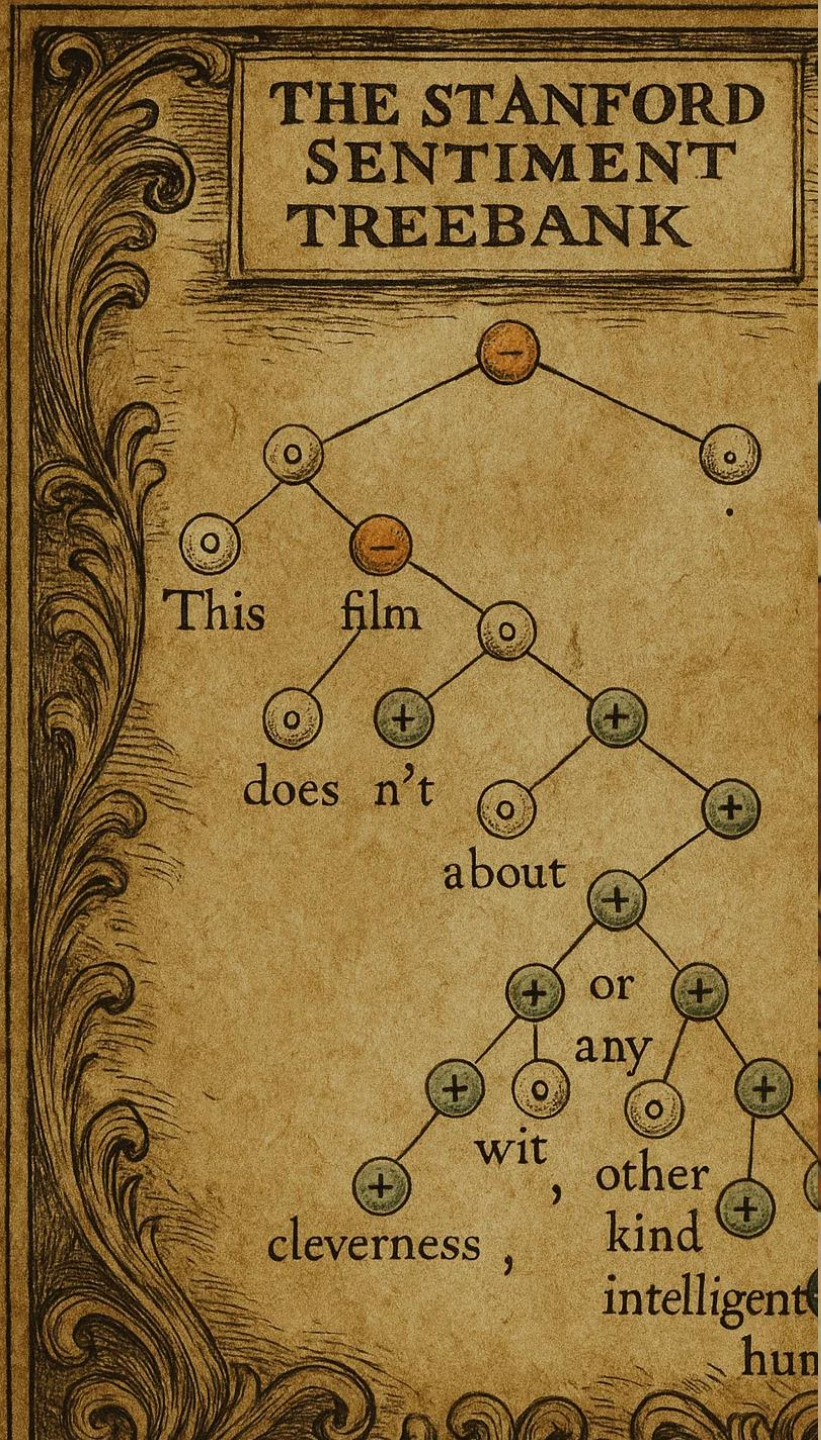
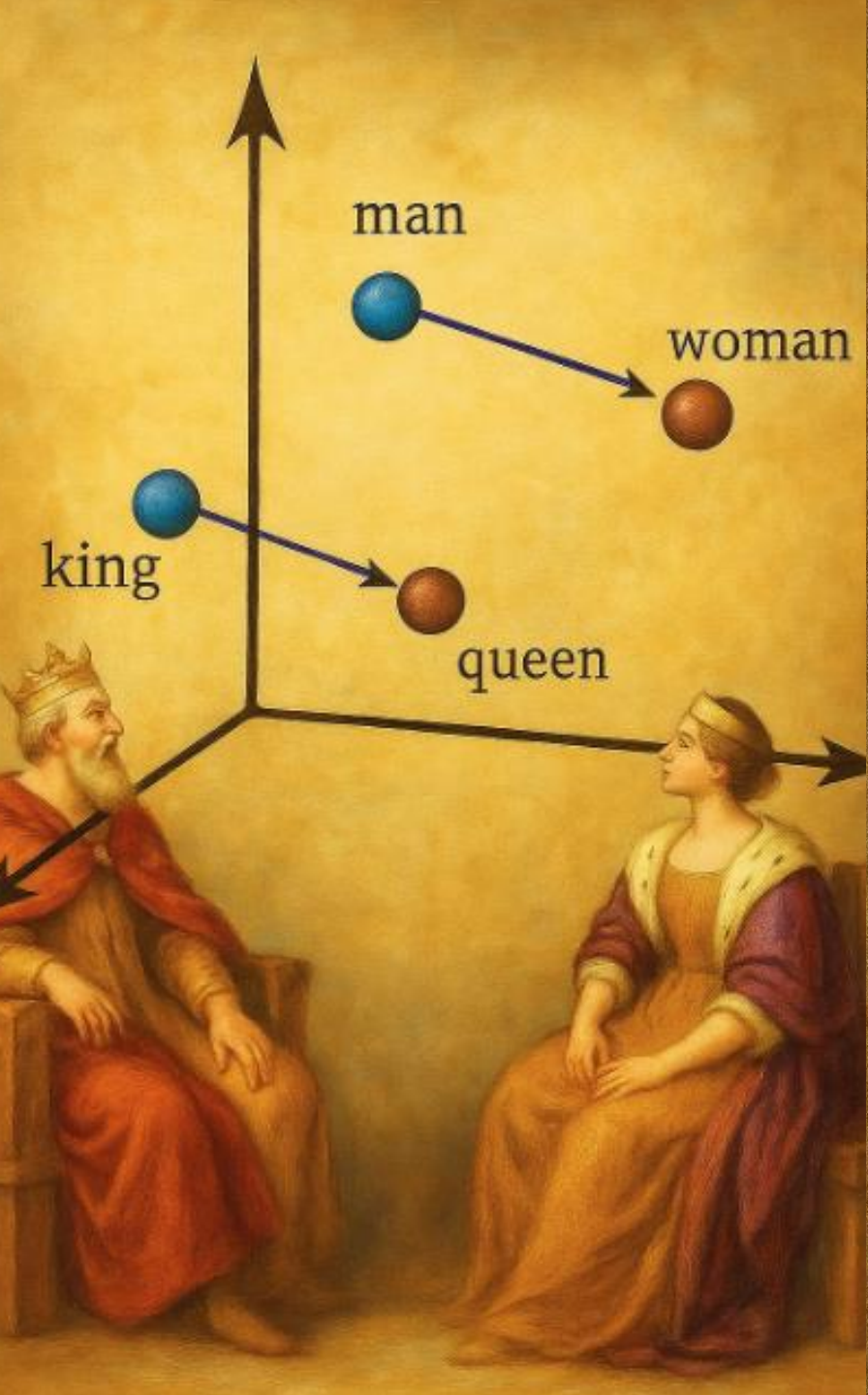


She spotted the squirrel with binoculars



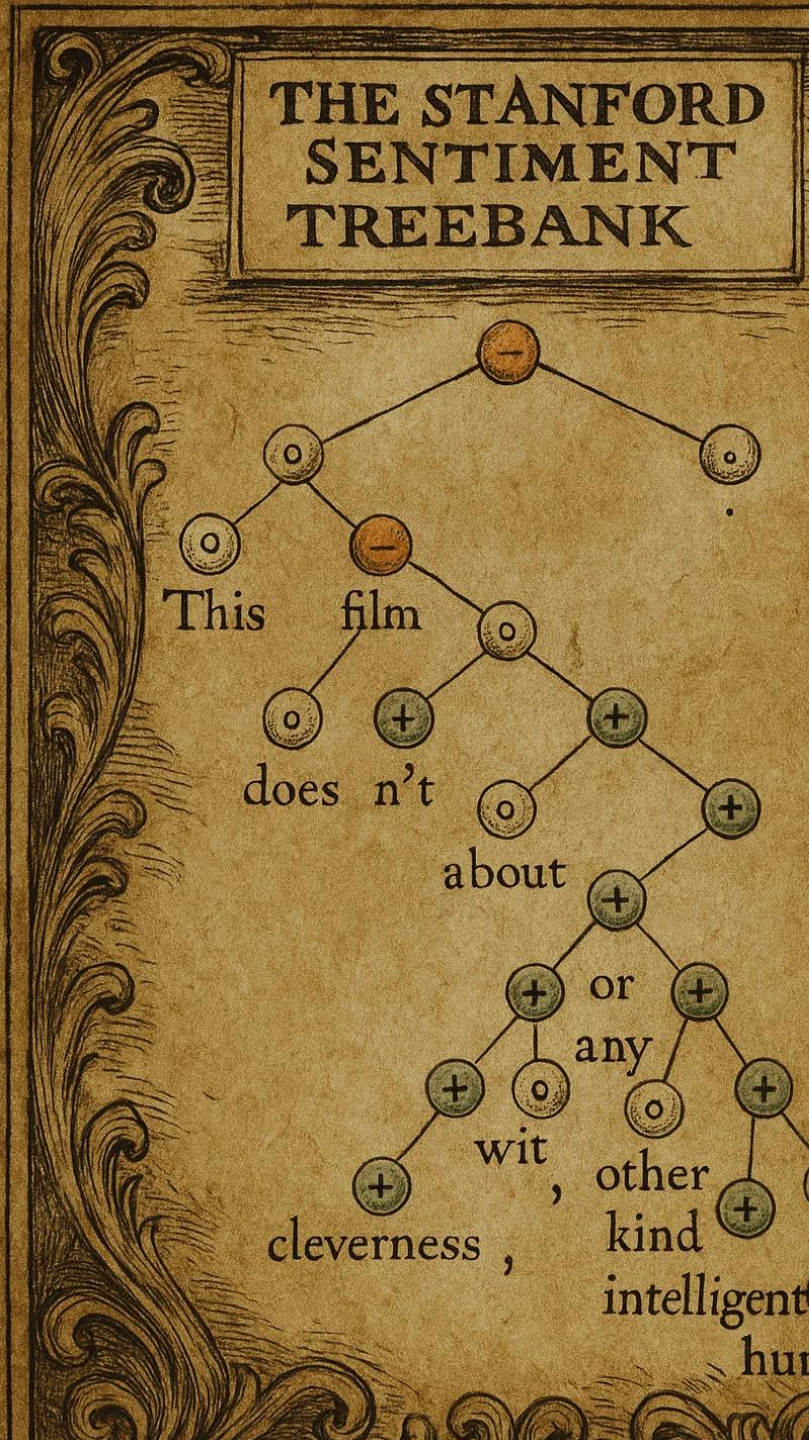
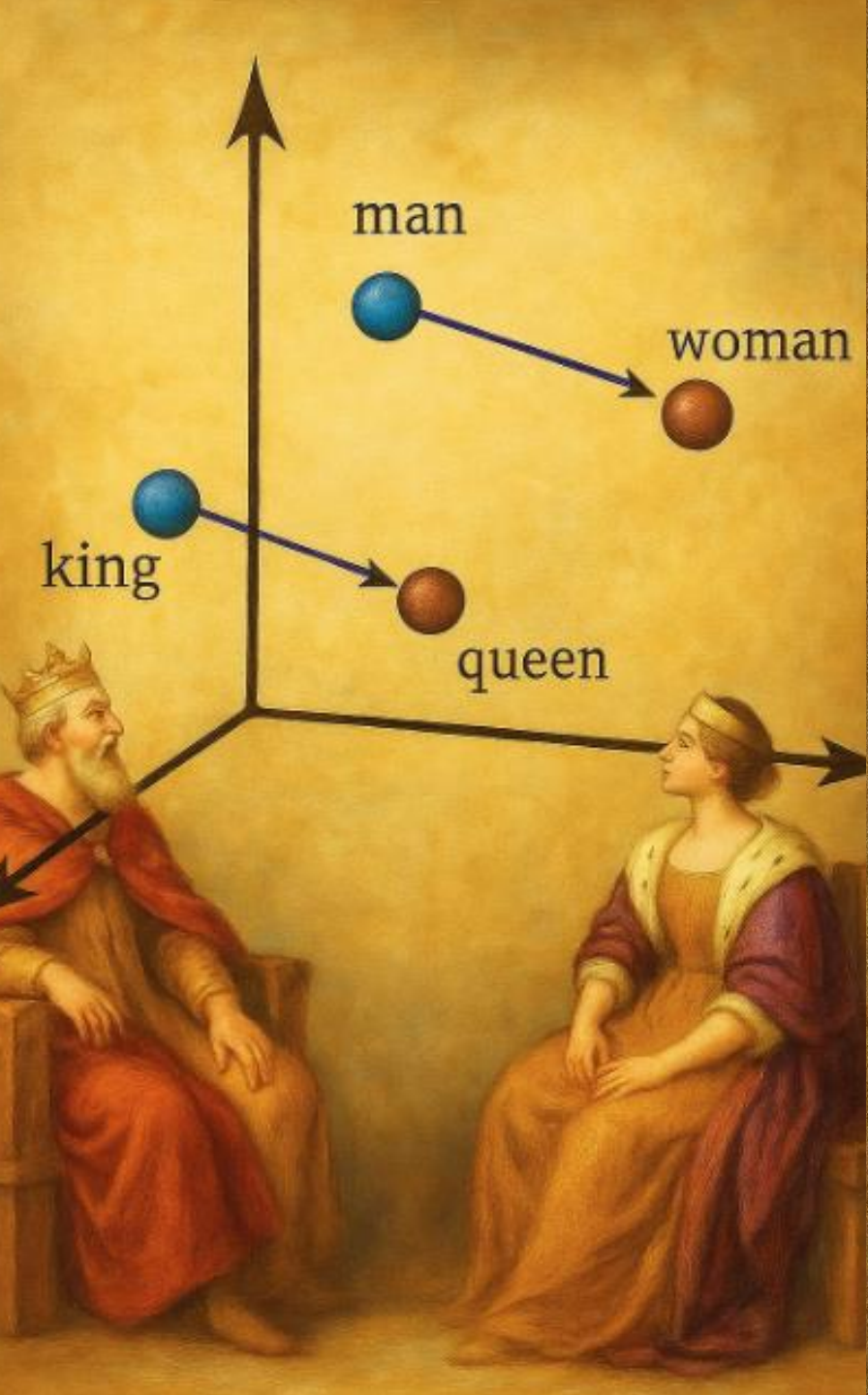






Word Embeddings with Co-Occurrence Matrices = GloVe

counts	like	enjoy	deep	learning	NLP	flying
like	0	1	0	0	0	0
enjoy	0	0	1	0	0	0
deep	0	0	0	0	0	0
learning	0	0	0	1	0	1
NLP	0	0	0	0	0	0
flying	0	0	1	0	0	1
	0	0	1	0	1	0



Word Embeddings with Co-Occurrence Matrices = GloVe

counts	like	enjoy	deep	learning	NLP	flying
like	0	1	0	0	0	0
enjoy	0	0	1	0	0	0
deep	0	0	0	0	0	0
learning	0	0	0	1	0	1
NLP	0	0	0	0	0	0
flying	0	0	1	0	0	1
	0	0	1	0	1	0

SEQUENCE-TO-SEQUENCE MODELS

Encoder

Context Vector

Decoder

Input Sequence

Output Sequence



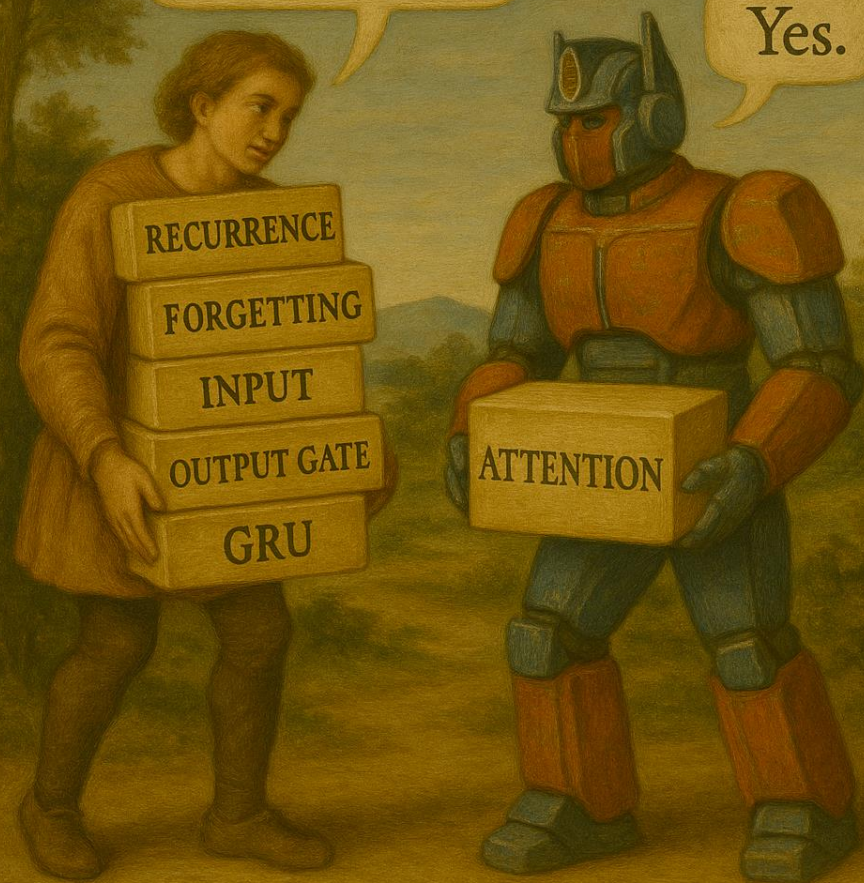
Extractive QA



SQuAD

Is that really all you need

Yes.



SEQUENCE-TO-SEQUENCE MODELS

Encoder

Context Vector

Decoder

Input Sequence

Output Sequence



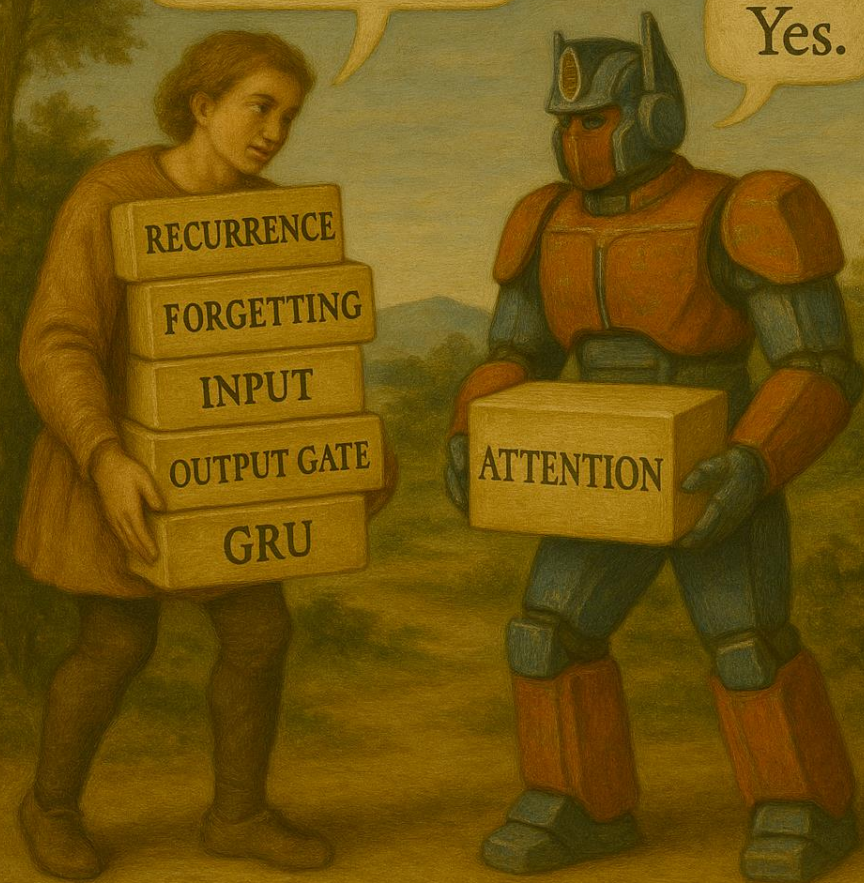
Extractive QA



SQuAD

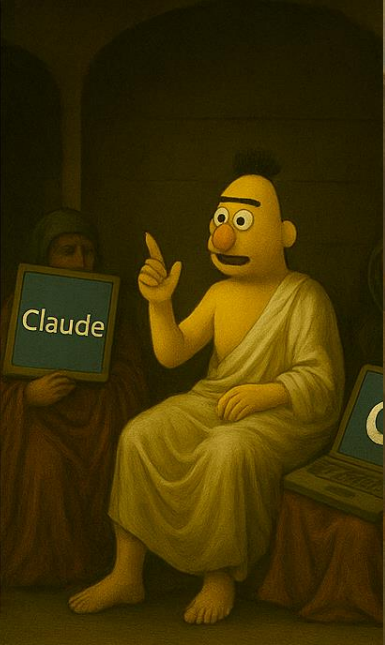
Is that really all you need

Yes.

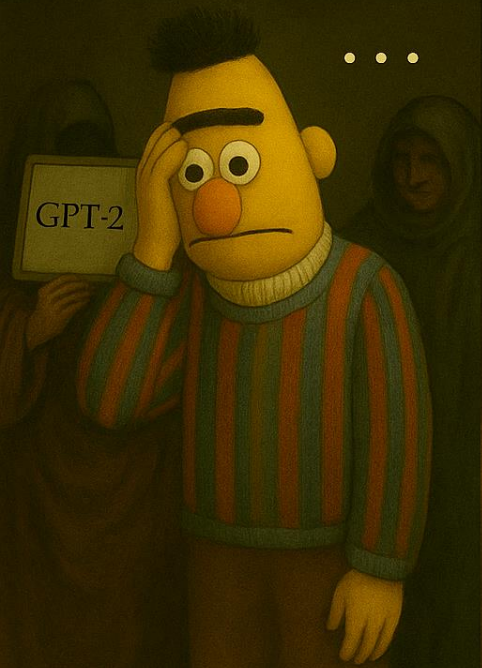




BIRTH

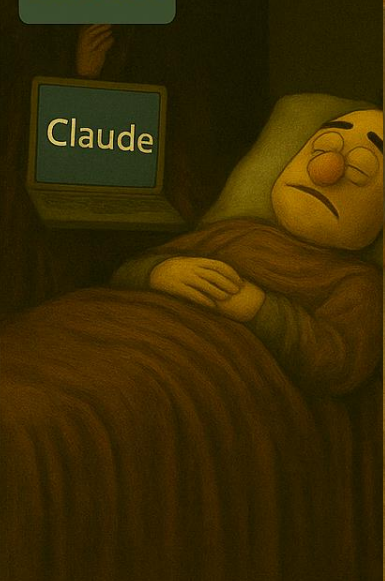


DECLINE



...

ChatGPT



Claude



GLUE

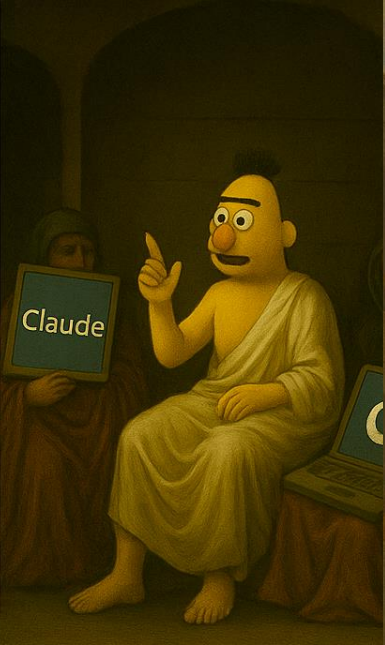
CoLA
SST-2
MRPC
STS-B
QQP
MNLI
QNLI
RTE
WNLI



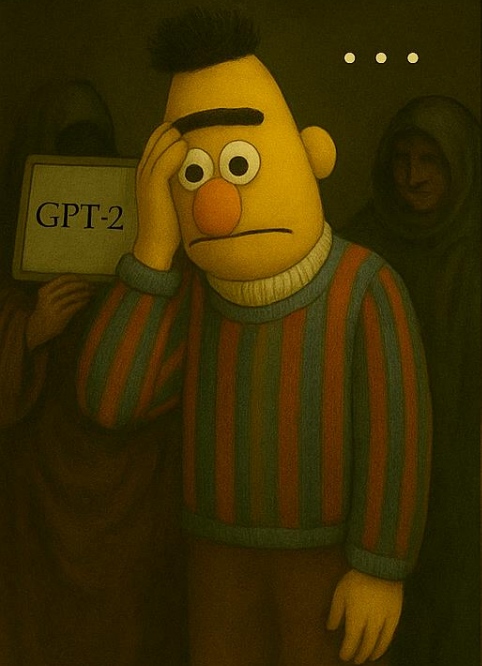
mBERT



BIRTH

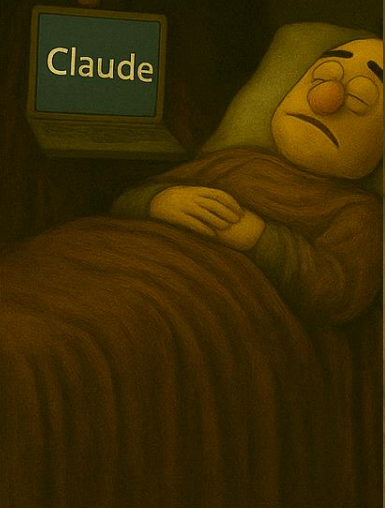


DECLINE



...

ChatGPT



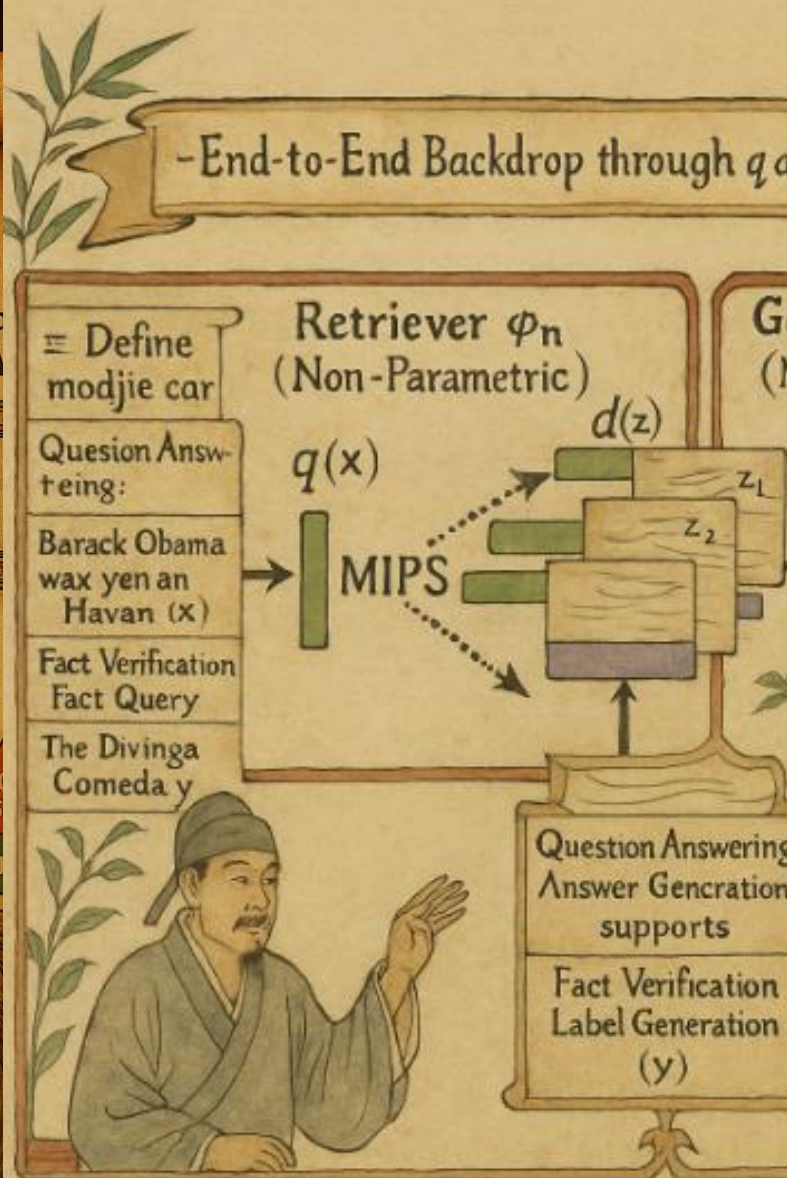
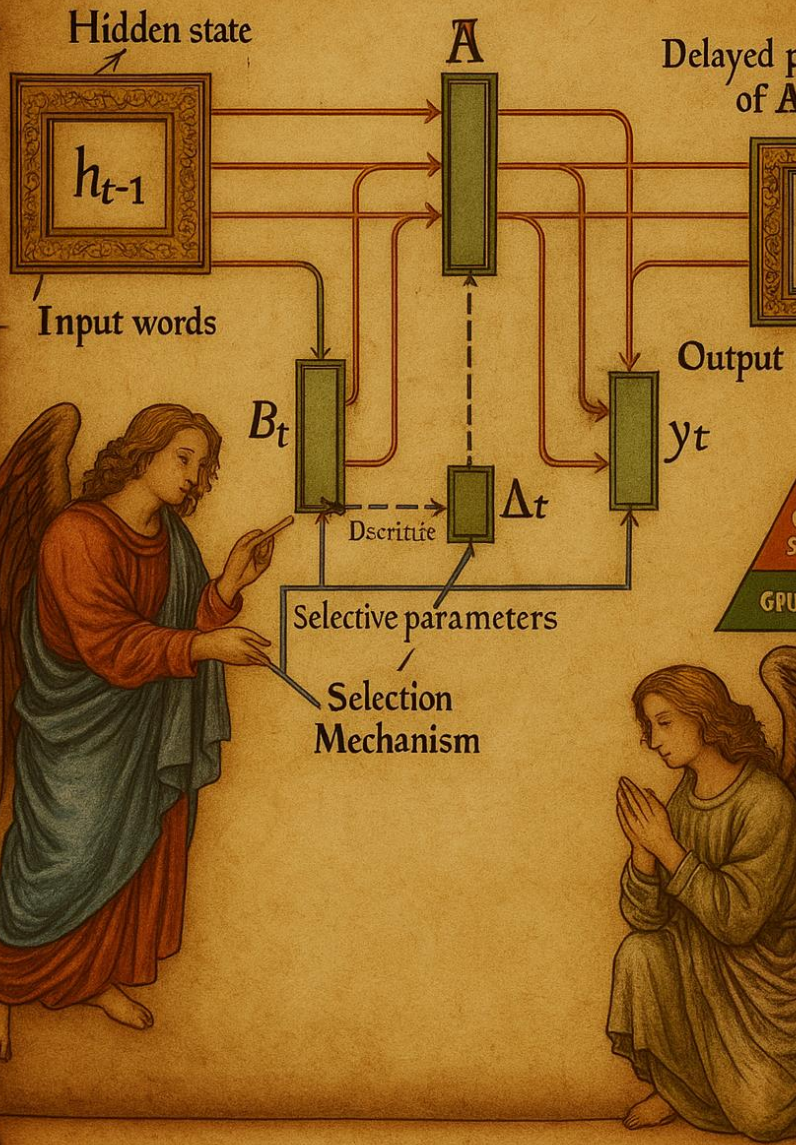


Figure 1: Overview of our approach, We co-trained retriever (Query Encoder + Document Encoder) end-to-end. For query x , we use-Maximum Inner Product Search (MIPS) to find the top-K documents.



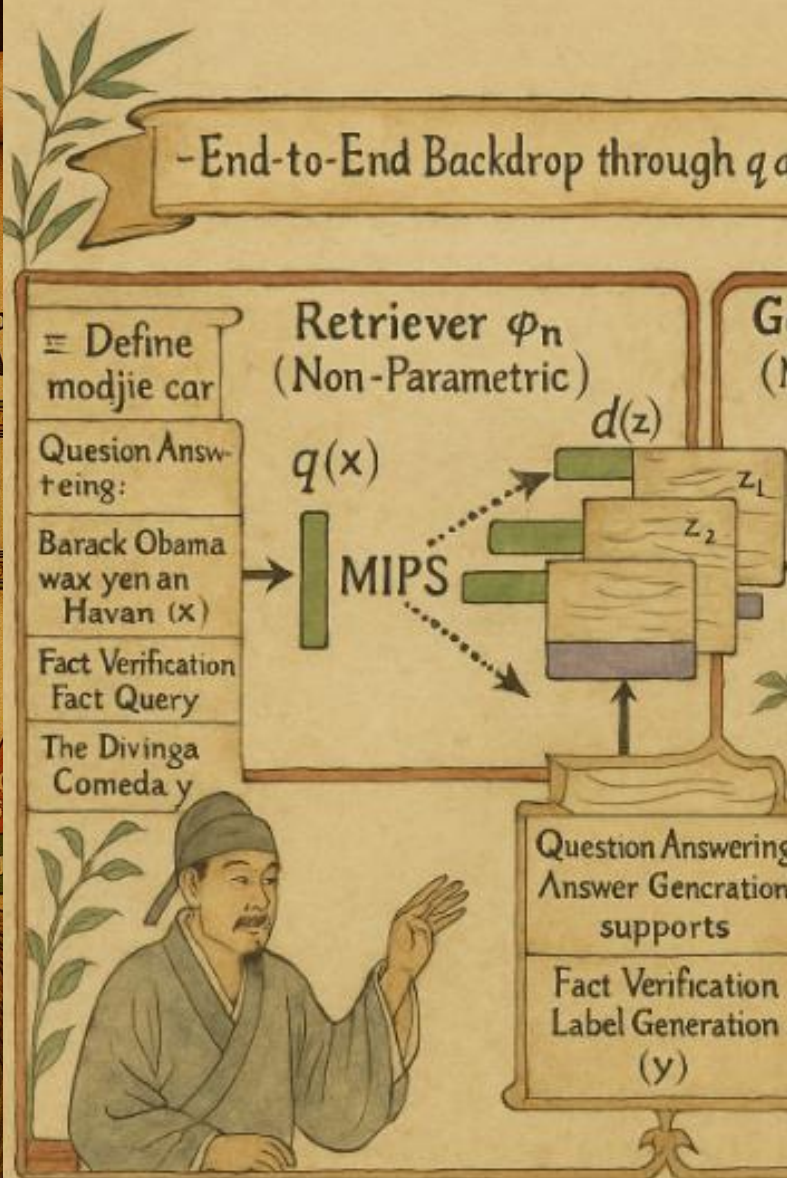
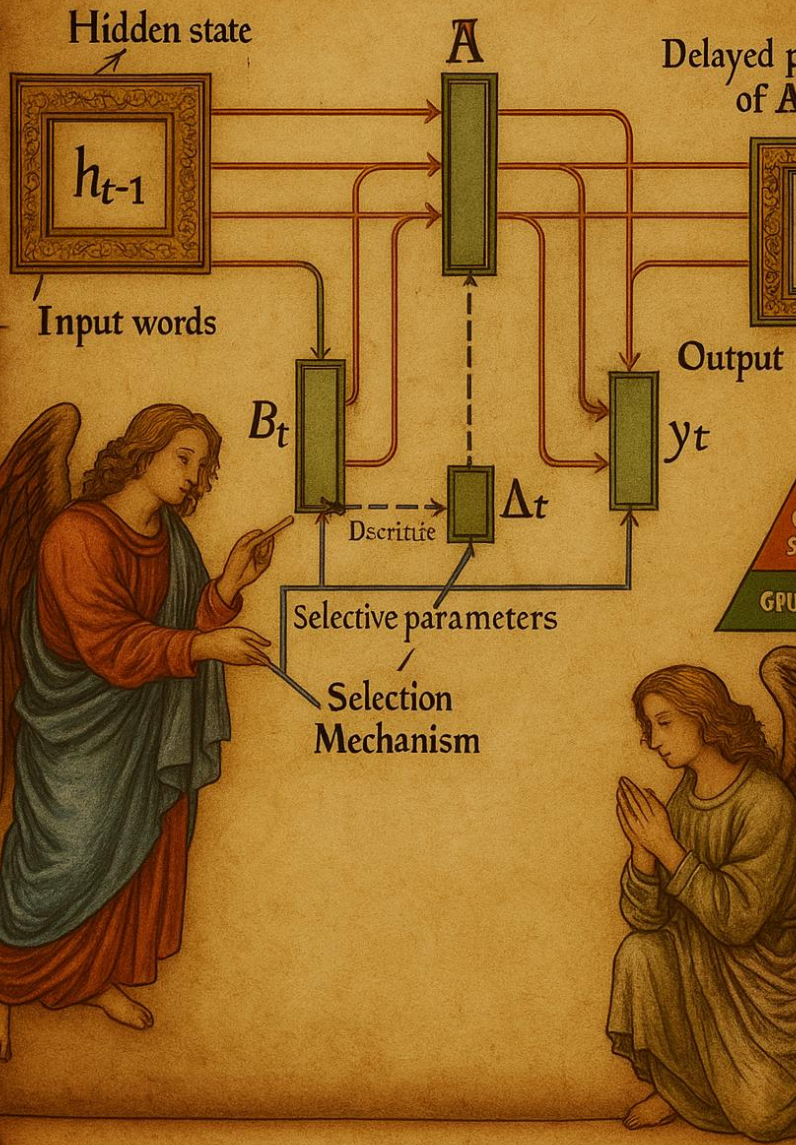
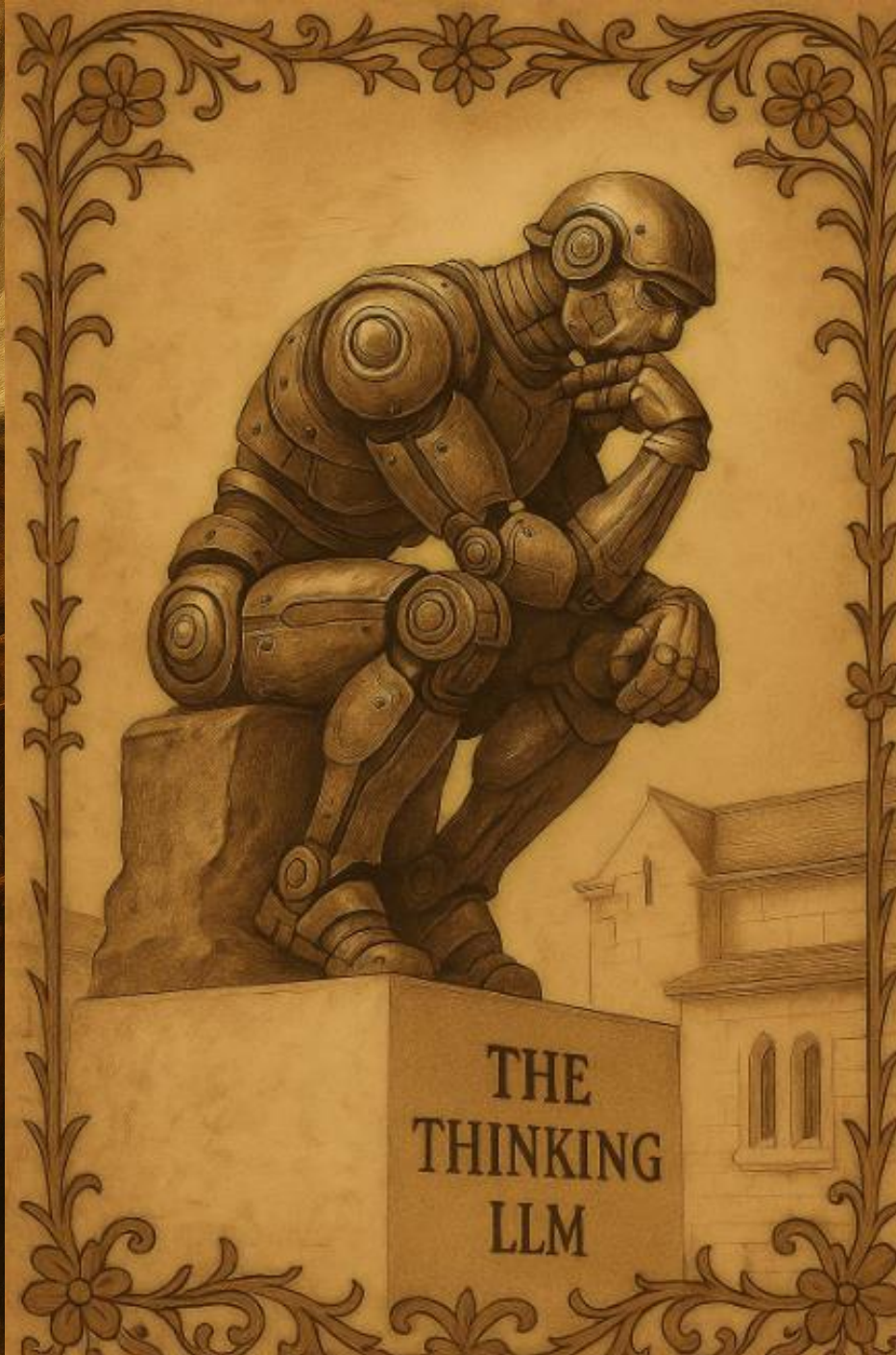
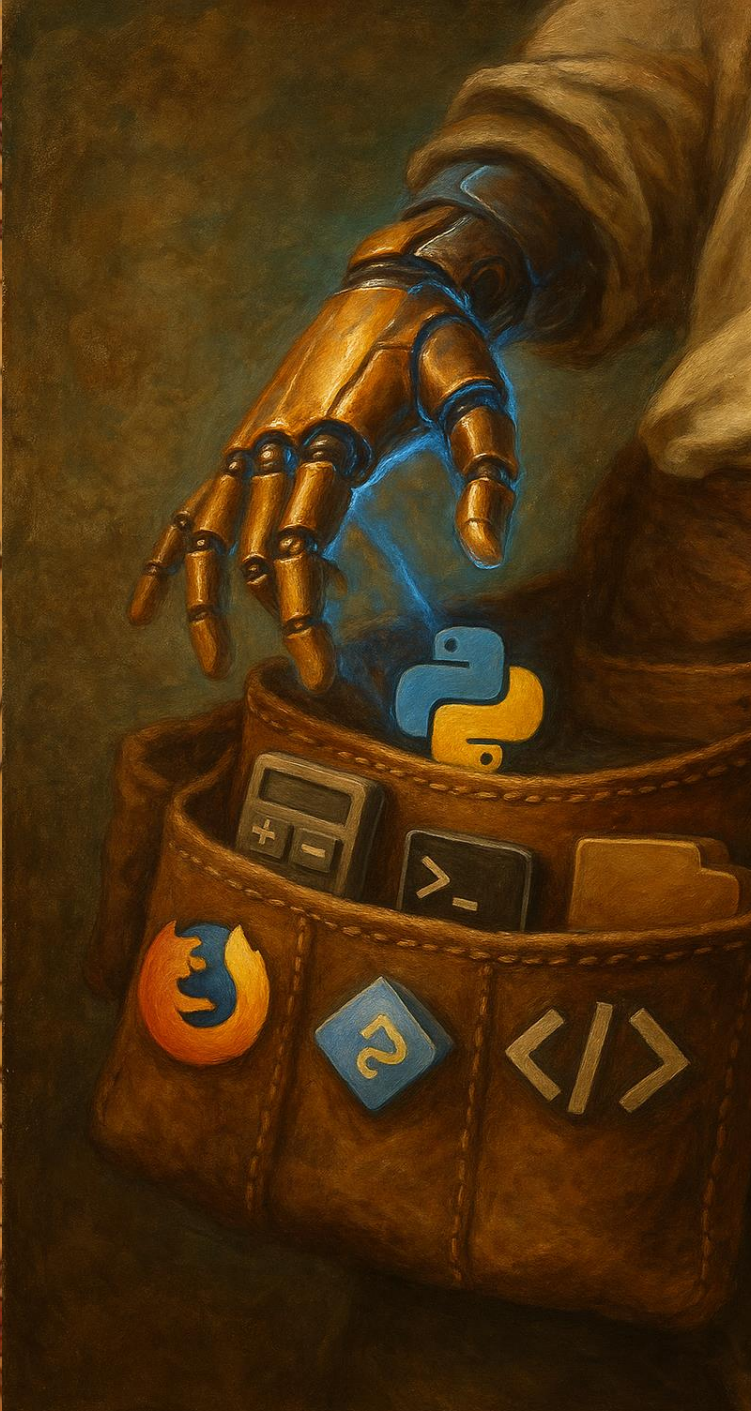
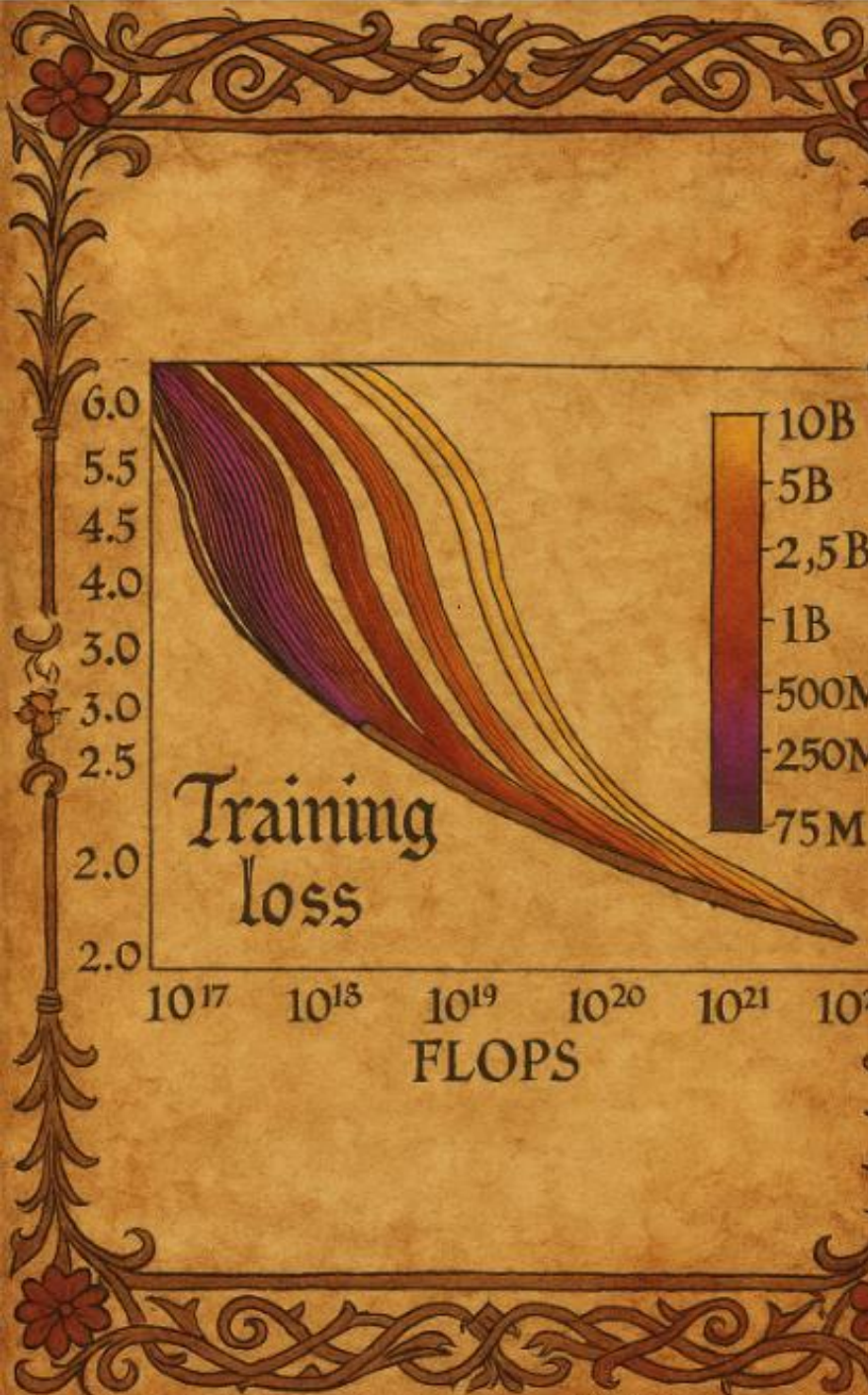
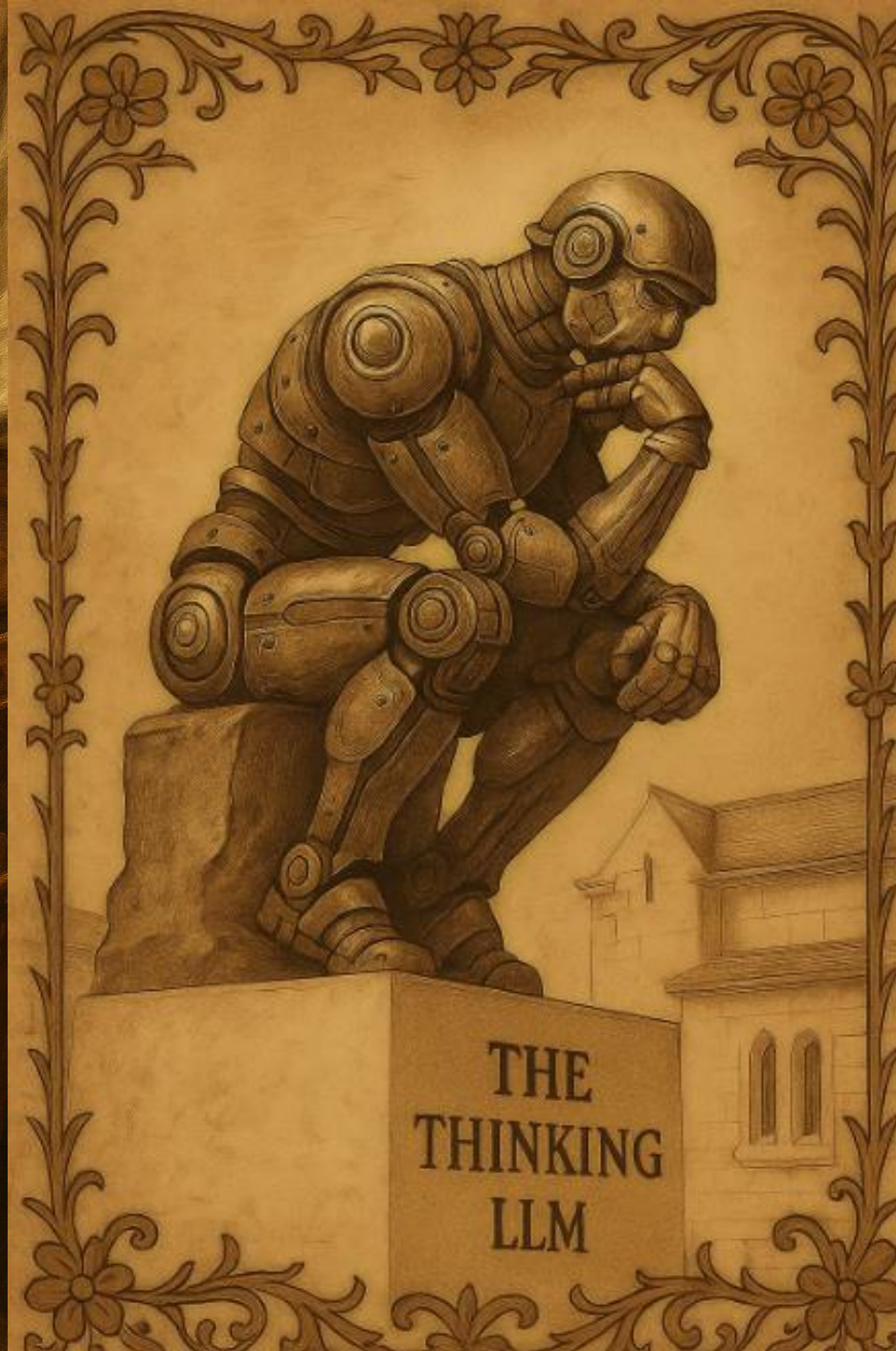
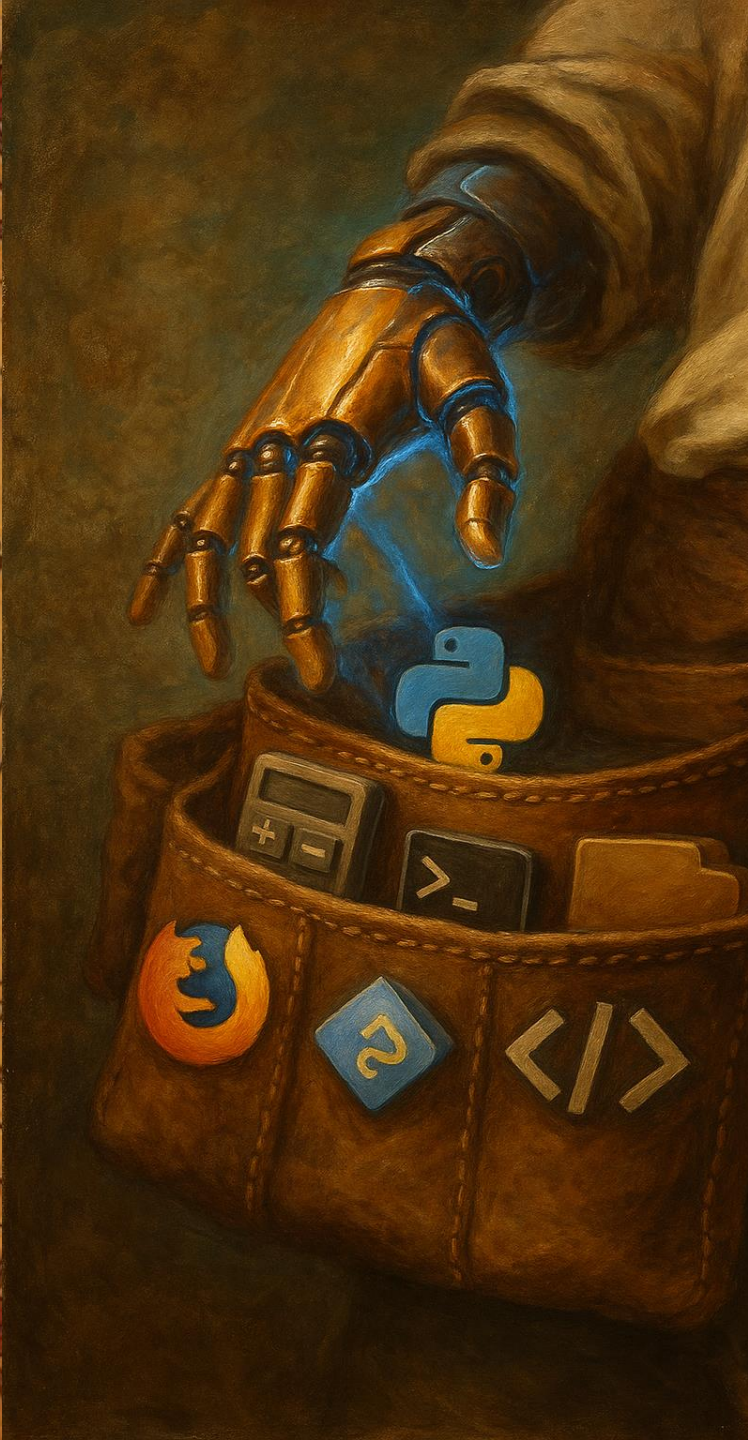
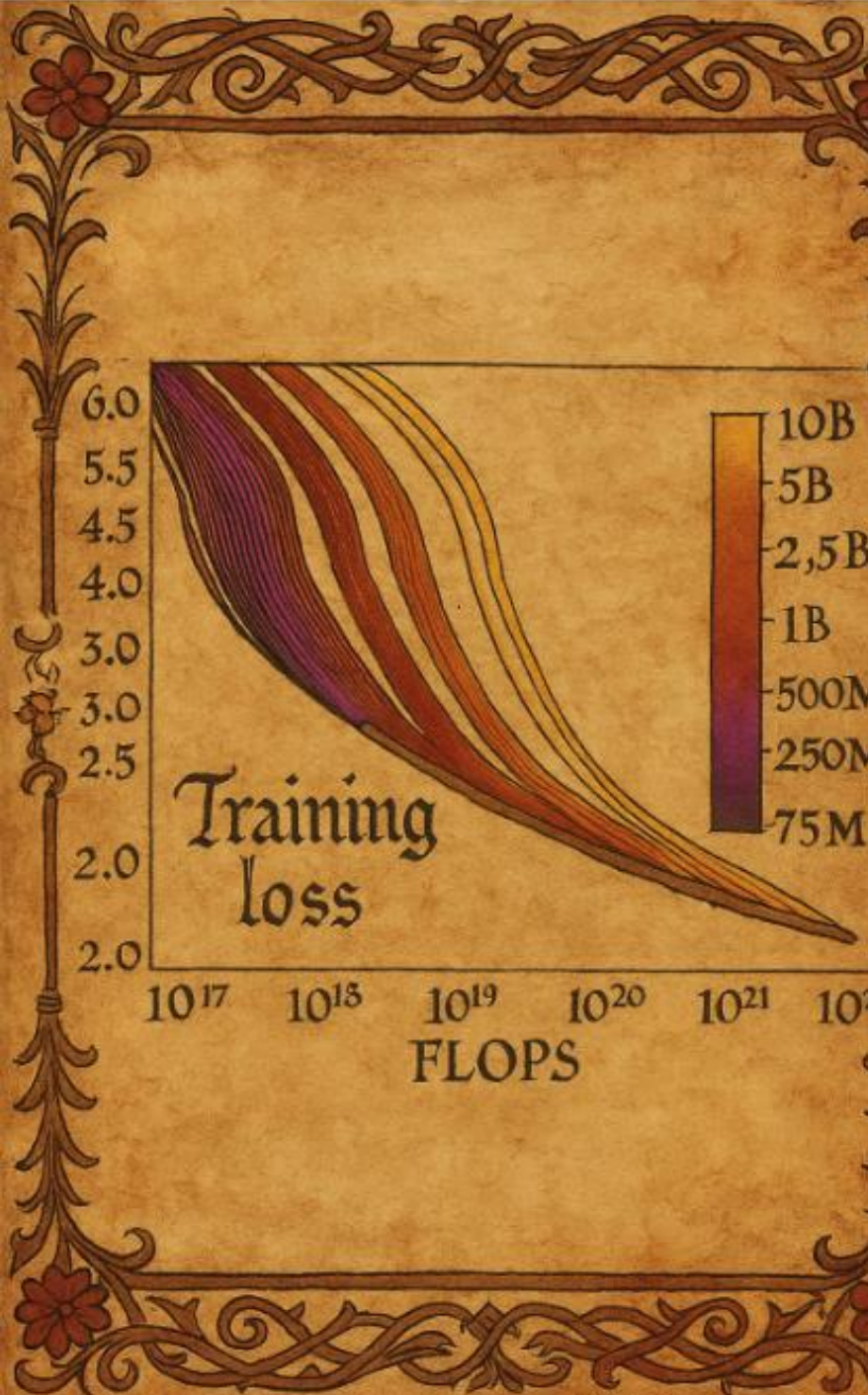


Figure 1: Overview of our approach, We co-train retriever (Query Encoder + Document Encoder) end-to-end. For query x , we use-Maximum Inner Product Search (MIPS) to find the top-K documents.







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.**

课程计划

Lecture 1 : 导论与 Word Vectors

1. 课程介绍 (10 分钟)

2. NLP 的历史 : 我们如何走到今天 (65 分钟)

- 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, who should we invite? (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* (the canonical text for a generation).
 - Key figure in Stanford NLP and the broader field.
- 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 **his 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

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, what is your 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 field.
- 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" linguistics 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 **his 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

客座讲座：Chris Manning 教授，CS224N 的创立者



NLP 发展史