Lecture 1: Course Overview, Linguistic Fundamentals, History of NLP

Instructor: Stephanie Schoch

CSCI 680: Natural Language Processing
Department of Computer Science, College of William & Mary
August 26, 2026

WILLIAM & MARY

Some slides are adapted from existing materials: Jurafsky & Martin (3rd ed.), Diyi Yang at Stanford, and Swabha Swayamdipta at USC.

Lecture Outline

- Course Overview
- Introduction to NLP

Course Overview: Introductions

1. What's your name?
2. What program are you in?
3. Why are you taking this class?
4. Something about you?

Course Overview: Basics

stephanieschoch.com/nlp

Natural Language Processing | **Home** | Schedule | Resources | Project

CSCI 680: Natural Language Processing

William & Mary · Fall 2026

Instructors

Stephanie Schoch

Welcome!

This course provides a comprehensive introduction to natural language processing, spanning foundational techniques through large language models. The first half of the course will focus on linguistic and statistical fundamentals before advancing to neural architectures. The second half of the course will focus on the modern LLM pipeline (e.g. pre-training, post-training, and prompting). Students will also engage with advanced topics and emerging areas in NLP (e.g. interpretability, harms and risks of language modeling, data-centric NLP).

Schedule

Week	Date	Topic	Course Material

Overview

Course Info

Time: MWF 10:00–10:50 AM
Location: Integrated Science Center (ISC) 3280
Office Hours: M 1:00–2:00 PM

Prerequisites

Students should be proficient in Python. Experience with packages such as SciPy, Scikit-learn, and PyTorch is helpful. Students should also have experience with Calculus, Linear Algebra, and Probability & Statistics.

Syllabus

Objectives, grading, policies, and more (PDF)

Textbooks:

- Jurafsky and Martin. “Speech and Language Processing.” 3rd Ed.
- Eisenstein. “Natural Language Processing.”

Course Overview: Landscape

Module 1
Statistical Foundations
Weeks 1-4

- Introduction, basics of text processing
- N-gram language models
- Text classification
- Word embeddings

Module 2
Neural Methods
Weeks 5-7

- FFNs, backpropagation, RNNs
- Seq2seq, attention, language generation
- Transformers

Module 3
Large Language Models
Weeks 8-10

- Transformer LMs
- Pretraining, scaling laws, post-training
- Fine-tuning, prompting
- LLM agents

Module 4
Advanced Topics
Weeks 11-13

- Evaluating LLMs, responsible language modeling
- Interpretability
- Data-centric NLP

Wrap Up
Weeks 14-15

- Outro
- Project presentations

Project Pitch	Project Proposal	Progress Report	Final Report
Sept. 14	Oct. 2	Nov. 2	Nov. 24

Course Overview: Policies

Grading:

- Project 50%
- Homework (3 × 10%) = 30%
- Quizzes (3 × 5%) = 15%
- Paper Discussions 5%

Late Policy:

- 1 penalty-free late day (homework assignments ONLY)
- Homework & project assignments: up to 3 days late each, 10% penalty each day

Course Overview: Generative AI

Use of Generative AI:

- **Homeworks:**
  - Required to try the assignment without AI assistance first. Then, allowed to use AI to help you work through blockers, if you choose. Each assignment will have a written component asking about this process.
- **Project:**
  - You are responsible for all design decisions, integrity of references/code/etc.
  - A major component of the final presentation is your ability to answer questions. "Claude recommended this" is not an acceptable answer, regardless of the question. Basically, you can use AI as a research or coding assistant, but it is a tool (not a coauthor/coworker). You are responsible for all components, and you need to think critically about **your** project.
- **Quizzes:**
  - In class, handwritten (no AI use allowed)
- **Paper Discussion:**
  - Permitted to assist preparation/understanding. You still need to read the paper yourself though.

Course Overview: Project

Overview:

- Semester-long small group research project on an NLP problem.
- **Questions to begin thinking about:**
  - What the problem is that will be solved?
  - Why the problem matters?
  - Why this project is feasible for this class/timeline?

Components:

- Project Pitch (5%)
- Team Formation
- Project Proposal (10%)
- Project Progress Report (10%)
- Project Final Presentation (10%)
- Project Final Report (15%)

Course Overview: Deadlines

Projects

1. **Project Pitch Slide** — Week 4 — Due Mon 09/14, 11:59 PM
2. **Project Team Formation** — Week 5 — Due Wed 09/23, 11:59 PM
3. **Project Proposal** — Week 6 — Due Fri 10/02, 11:59 PM
4. **Project Progress Report** — Week 11 — Due Mon 11/02, 11:59 PM
5. **Project Final Report** — Week 14 — Due Tue 11/24, 11:59 PM

Homework

1. **Homework 1** — Week 5 — Released Wed 09/09 — Due Fri 09/25, 11:59 PM
2. **Homework 2** — Week 9 — Released Mon 09/28 — Due Fri 10/23, 11:59 PM
3. **Homework 3** — Week 13 — Released Mon 10/26 — Due Fri 11/20, 11:59 PM

Quizzes

1. **Quiz 1: Statistical Foundations** — Week 5 — Due Mon 09/21, 10:00 AM
2. **Quiz 2: Neural Methods** — Week 8 — Due Mon 10/12, 10:00 AM
3. **Quiz 3: LLMs** — Week 10 — Due Fri 10/30, 10:00 AM

All times shown are local course times.

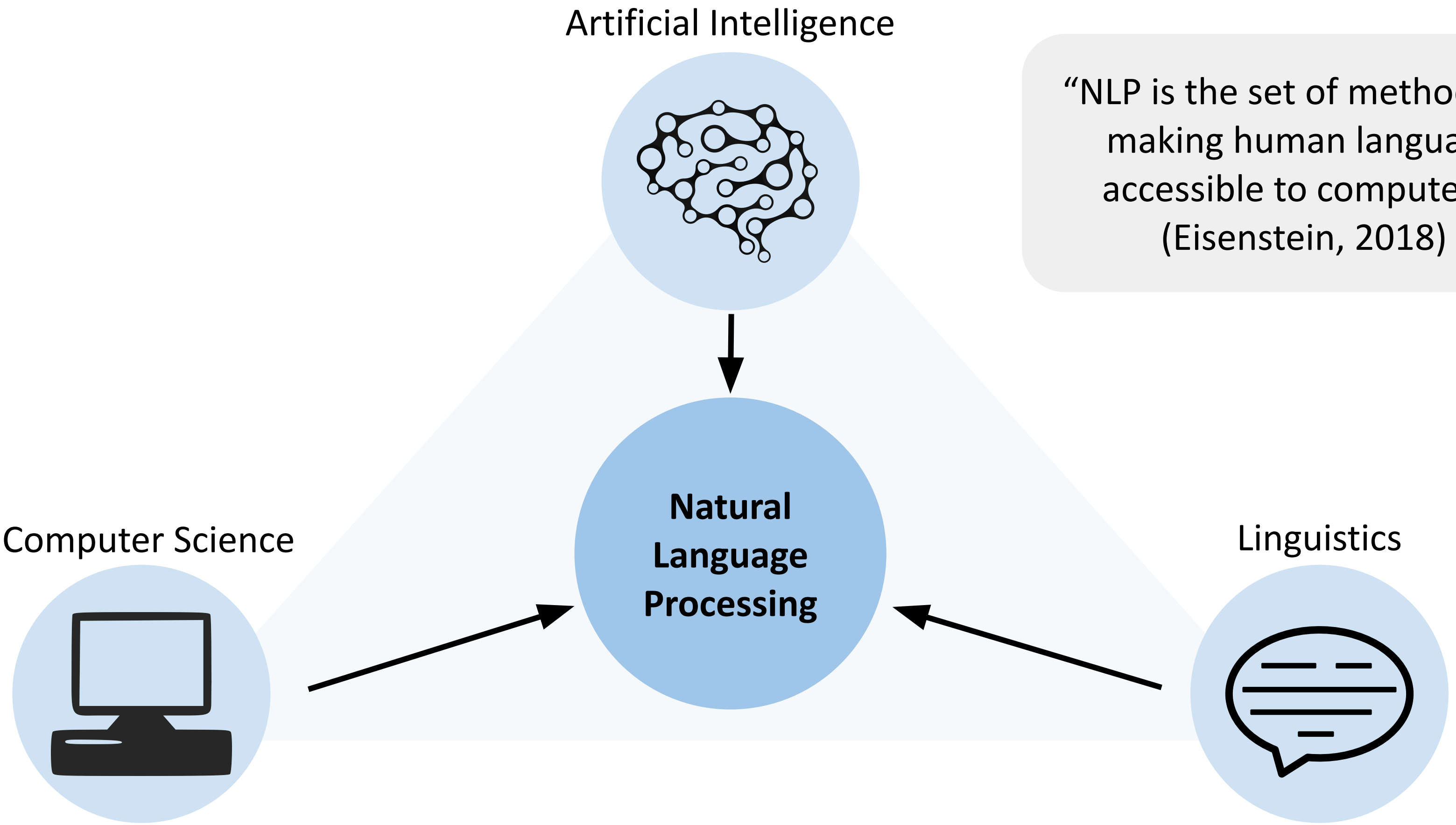
Course Overview: Outcomes

Learning Objectives:

- Explain foundational linguistic, statistical, and neural concepts underlying modern NLP.
- Implement NLP models and pipelines for various tasks (e.g. text classification, question answering, natural language inference) using modern software frameworks.
- Evaluate NLP systems through experimental design, quantitative metrics, and qualitative error analysis.
- Read and critique current NLP research literature and communicate technical ideas effectively.
- Design and execute an NLP research project involving problem formulation, experimentation, and analysis.
- Assess the limitations, risks, and societal implications of NLP technologies (e.g. LLMs), including consideration for the underlying data

Before Moving On: Questions About The Course?

What is NLP?



"NLP is the set of methods for making human language accessible to computers" (Eisenstein, 2018)

NLP Today: Large Language Models (LLMs)

…

Modern NLP is dominated by large language models.

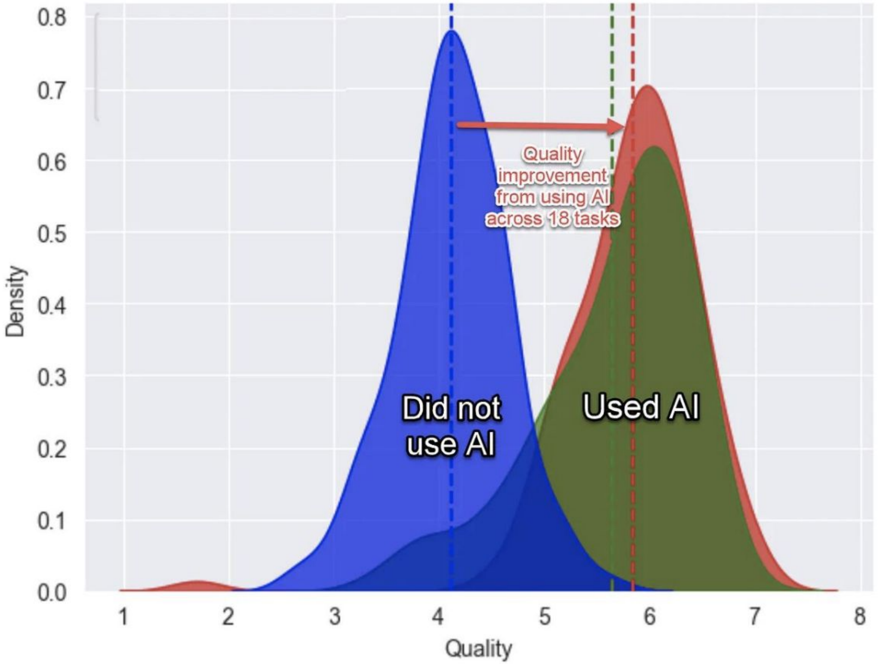
LLMs: The Good

Improved Productivity

Consultants using GPT-4 outperform non-users.

Dell'Acqua et al. 2023; Mollick 2023

- Consultants (from Boston Consulting Group) using GPT-4 **finished 12.2% more tasks** on average, **completed tasks 25.1% more quickly**, and produced **40% higher quality results** than those not using AI.
- Use of LLM especially improved performance of lower-performing humans.
- Results vary depending on task



Distribution of output quality across all tasks. Blue group did not use GPT-4; green and red groups used GPT-4; red group got additional training on how to use AI.

LLMs: The Good

Reduced Language Barriers (i.e. Machine Translation)

- Can detect language automatically
- Can reorder text spans as more text is input.

LLMs: The Good → The Concerns

PHILOSOPHICAL TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Research

Cite this article: Katz DM, Bommarito MJ, Gao S, Arredondo P. 2024 GPT-4 passes the bar exam. *Phil. Trans. R. Soc. A* **382**: 20230254. https://doi.org/10.1098/rsta.2023.0254

Received: 22 September 2023
Accepted: 20 December 2023

One contribution of 15 to a theme issue 'A complexity science approach to law and governance'.

Subject Areas:
artificial intelligence

Keywords:
large language models, Bar Exam, GPT-4, legal services, legal complexity, legal language

Author for correspondence:
Daniel Martin Katz
e-mail: dkatz3@kentlaw.iit.edu

GPT-4 passes the bar exam

Daniel Martin Katz[1,2,3,4], Michael James Bommarito[1,2,3,4], Shang Gao[5] and Pablo Arredondo[2,5]

[1]Illinois Tech, Chicago Kent College of Law, Chicago, IL, USA
[2]CodeX, The Stanford Center for Legal Informatics, Stanford, CA, USA
[3]Bucerius Law School, Hamburg, Germany
[4]273 Ventures, LLC, USA
[5]Casetext, Inc., USA

DMK, 0000-0002-9775-0320

In this paper, we experimentally evaluate the zero-shot performance of GPT-4 against prior generations of GPT on the entire uniform bar examination (UBE), including not only the multiple-choice multistate bar examination (MBE), but also the open-ended multistate essay exam (MEE) and multistate performance test (MPT) components. On the MBE, GPT-4 significantly outperforms both human test-takers and prior models, demonstrating a 26% increase over ChatGPT and beating humans in five of seven subject areas. On the MEE and MPT, which have not previously been evaluated by scholars, GPT-4 scores an average of 4.2/6.0 when compared with much lower scores for ChatGPT. Graded across the UBE components, in the manner in which a human test-taker would be, GPT-4 scores approximately 297 points, significantly in excess of the passing threshold for all UBE jurisdictions. These findings document not just the rapid and remarkable advance of large language model performance generally, but also the potential for such models to support the delivery of legal services in society.

This article is part of the theme issue 'A complexity science approach to law and governance'.

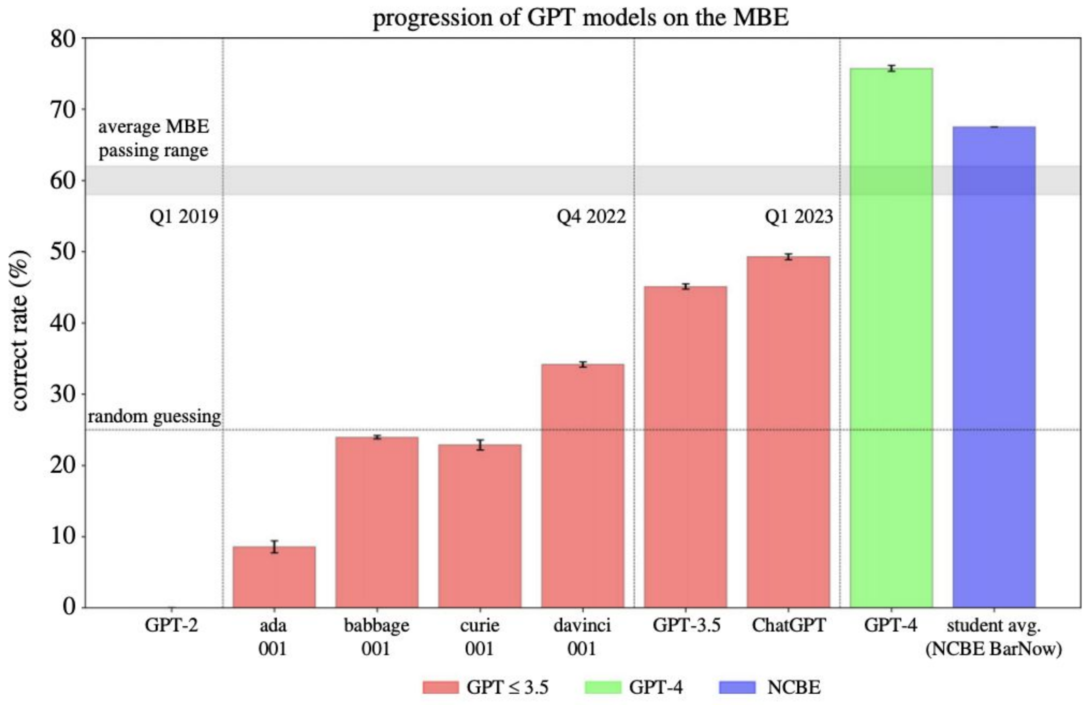
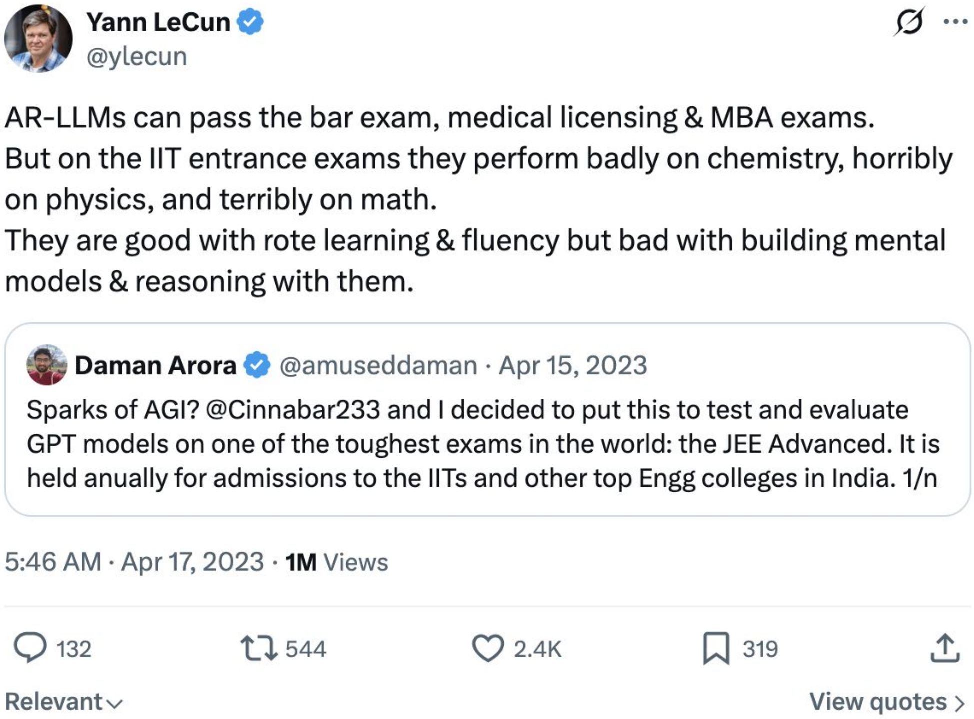


Figure 1. Progression of recent GPT models on the multistate bar exam (MBE).

LLMs: The Good → The Concerns



Yann LeCun
@ylecun

AR-LLMs can pass the bar exam, medical licensing & MBA exams.
But on the IIT entrance exams they perform badly on chemistry, horribly on physics, and terribly on math.
They are good with rote learning & fluency but bad with building mental models & reasoning with them.

Daman Arora @amuseddaman · Apr 15, 2023
Sparks of AGI? @Cinnabar233 and I decided to put this to test and evaluate GPT models on one of the toughest exams in the world: the JEE Advanced. It is held anually for admissions to the IITs and other top Engg colleges in India. 1/n

5:46 AM · Apr 17, 2023 · **1M** Views

132 544 2.4K 319

Relevant

View quotes

Memorization vs. Generalization

LLMs: The Concerns

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Share full article 267

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

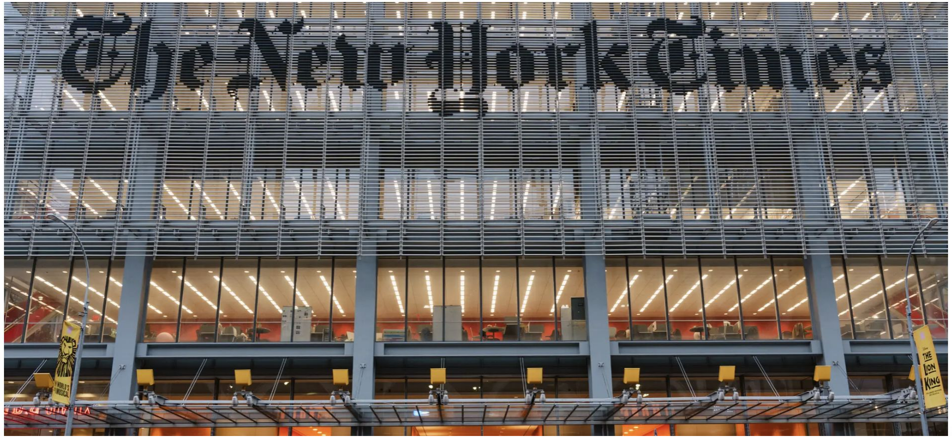
Hallucination, Misinformation

The New York Times Amends Lawsuit Against OpenAI and Microsoft

In a new court filing, The Times accused Microsoft of encouraging OpenAI to train its A.I. systems using copyrighted articles.

Listen · 4:04 min

Share full article



Privacy, Copyright Issues

LLMs: The Concerns

All Articles > Millions Already Turn to AI for Therapy. Is It Safe?

May 2026

Millions Already Turn to AI for Therapy. Is It Safe?

Researchers are working to make AI tools used for mental health support safer, smarter, and grounded in the full picture of a person's life

Spring/Summer 2026

by Allison Eck — 14 minute read — Feature

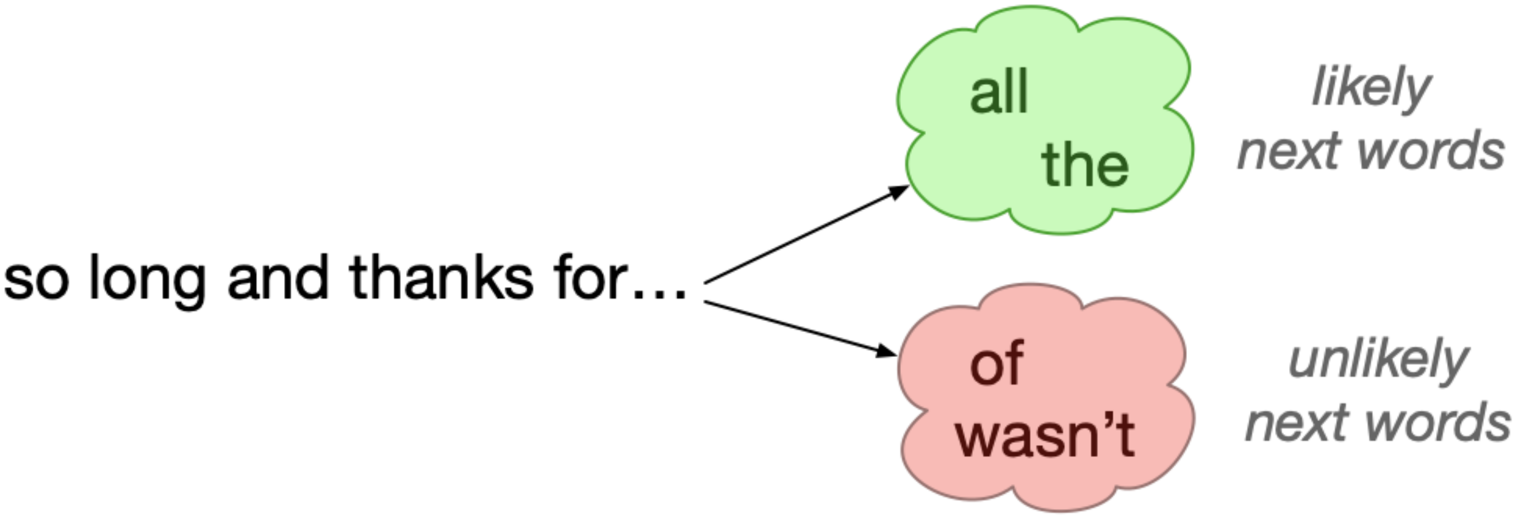
```
 Welcome to
                   EEEEEE  LL      IIII   ZZZZZZ   AAAAA
                   EE      LL       II        ZZ  AA   AA
                   EEEEE   LL       II      ZZZ   AAAAAAA
                   EE      LL       II     ZZ     AA   AA
                   EEEEEE  LLLLLL  IIII ZZZZZZ    AA   AA

 Eliza is a mock Rogerian psychotherapist.
 The original program was described by Joseph Weizenbaum in 1966.
 This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Safety and Alignment; Ethical Issues

What is a Language Model?

Language Model:

A computational system that can **predict the next word** from previous words.



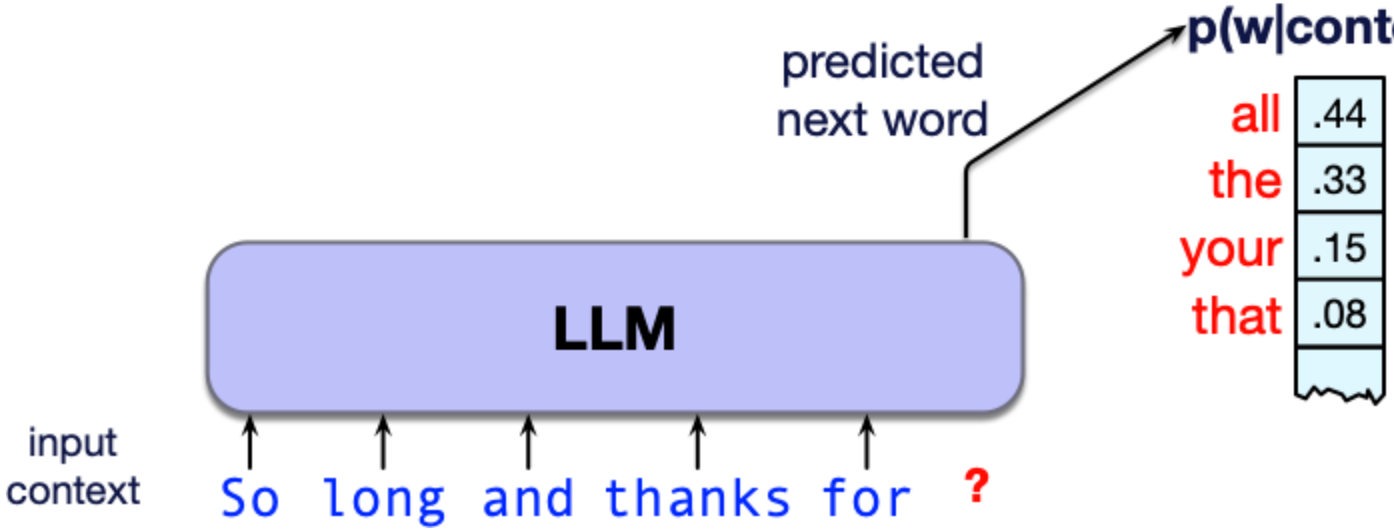
What is a Large Language Model?

Large Language Model:

A neural network with:

Input: a context or prefix,

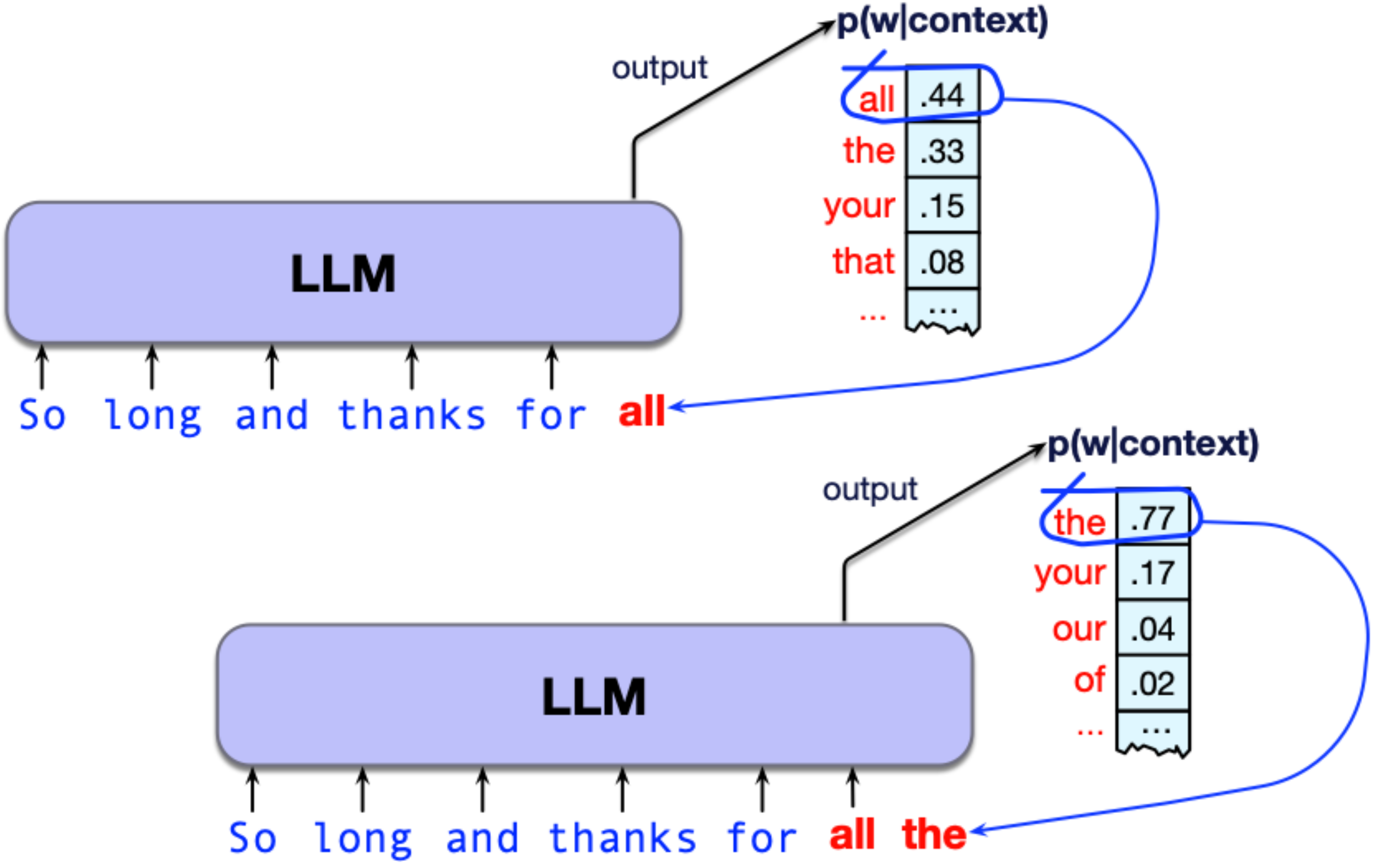
Output: a distribution over possible next words



Predict: assign distribution over words

Generate: sample from distribution

Large Language Model: Generation



Predict: assign distribution over words

Generate: sample from distribution

Language Models: Size

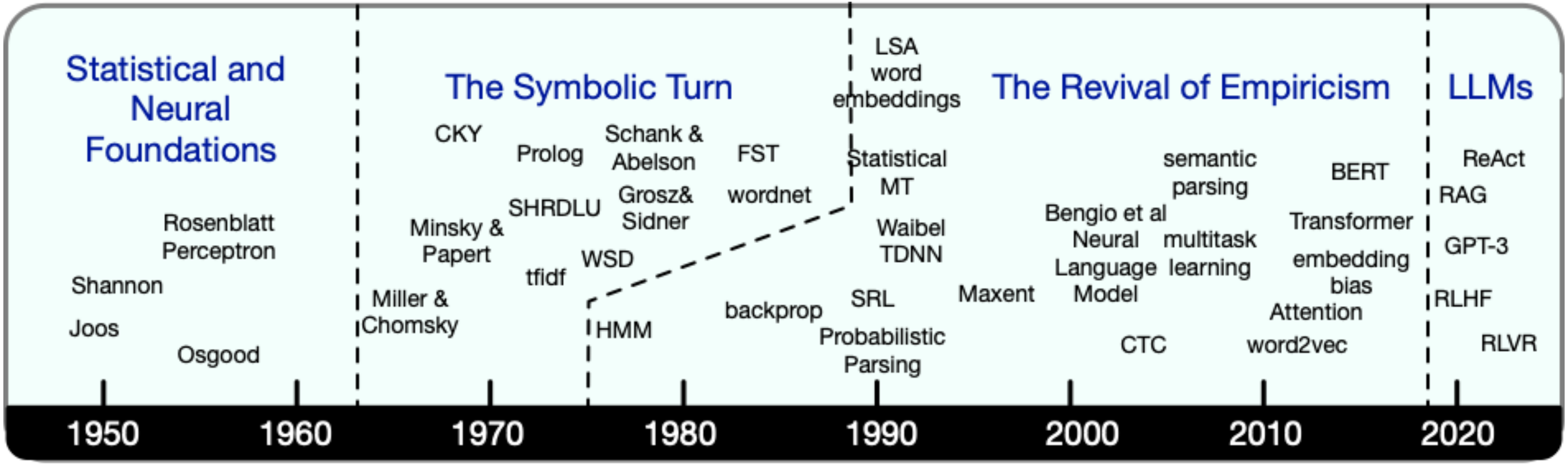
LANGUAGE MODEL SIZE

n-gram models (frequency-based) ⟷ LLMs (neural)

Large Language Models:

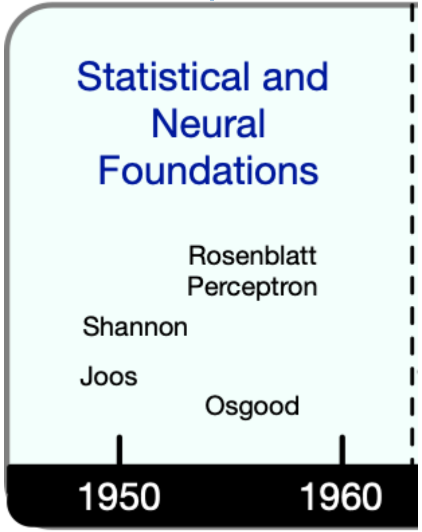
- Many **layers**
- Many **parameters**
- Large amounts of **training data**

A Brief History of NLP

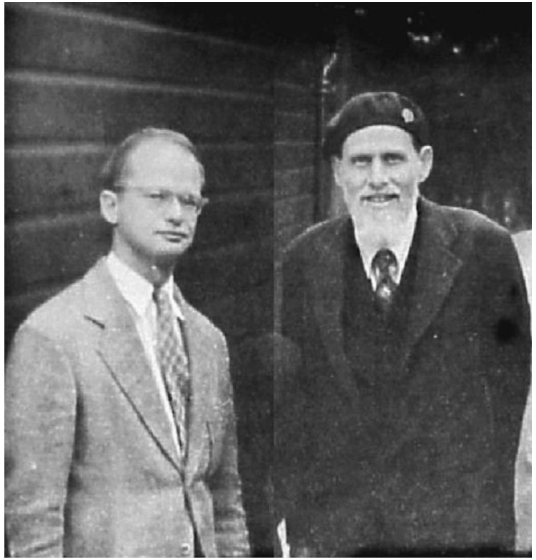


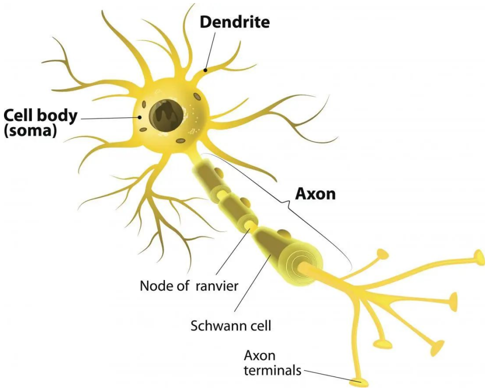
A Brief History of NLP: Statistical and Neural Foundations (1950s, early 1960s)

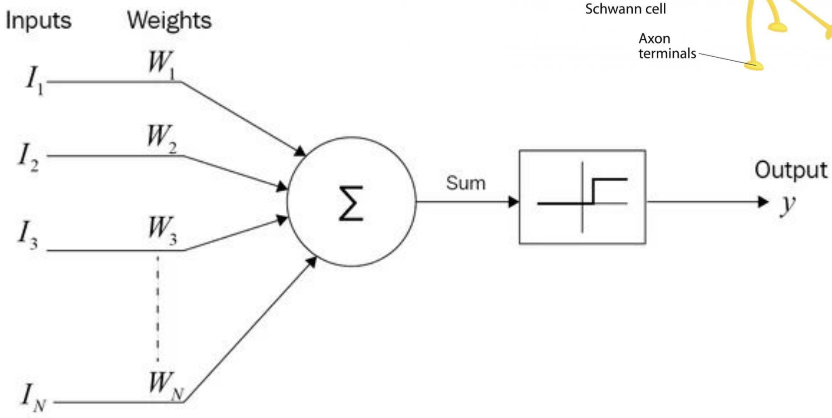
Era	Time (Approx.)	Dominant Idea	Computational Strategies
Statistical and Neural Foundations	1950s, early 1960s	Language contains statistical regularities that can be learned; words (i.e. meaning) can have distributed / vector representations.	probabilities, early neural networks, vector representations



A Brief History of NLP: How Did We Get Here?







McCulloch-Pitts Neuron in
"A Logical Calculus of the Ideas Immanent in Nervous Activity" (1943)

A Brief History of NLP: How Did We Get Here?

"When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'" – Warren Weaver, March 1947

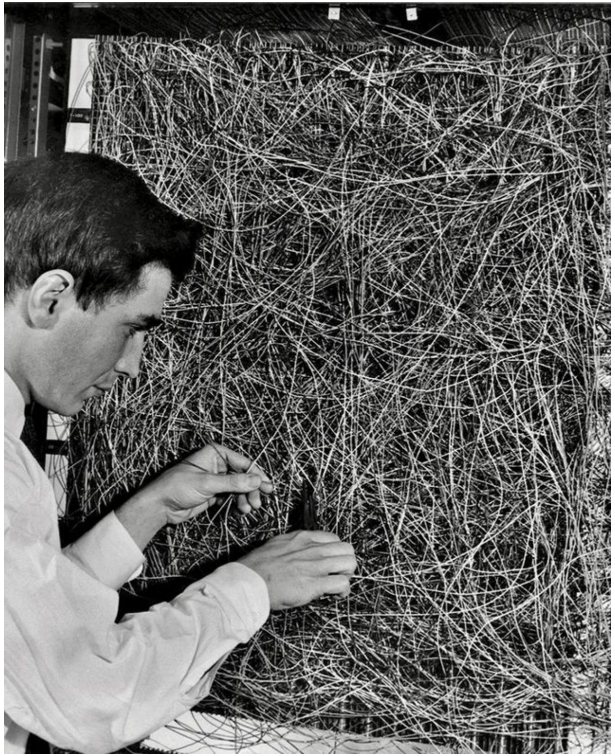
Early notions of machine translation

A Brief History of NLP: How Did We Get Here?

Conference was “to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.” – John McCarthy 1955

The Dartmouth Summer Research Project 1956

A Brief History of NLP: Statistical and Neural Foundations (1950s, early 1960s)



"An IBM 704 – a 5-ton computer the size of a room – was fed a series of punch cards. After 50 trials, the computer taught itself to distinguish cards marked on the left from cards marked on the right." – Melanie Lefkowitz (Cornell Chronicle)

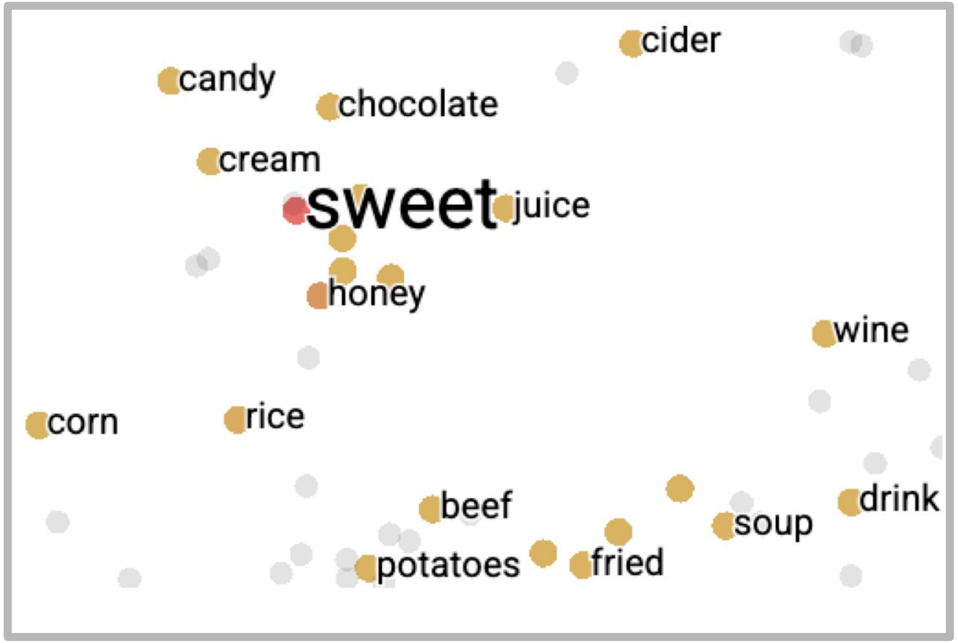
[Professor's perceptron paved the way for AI – 60 years too soon]

Frank Rosenblatt: The Mark I Perceptron (1957)

A Brief History of NLP: Statistical and Neural Foundations (1950s, early 1960s)

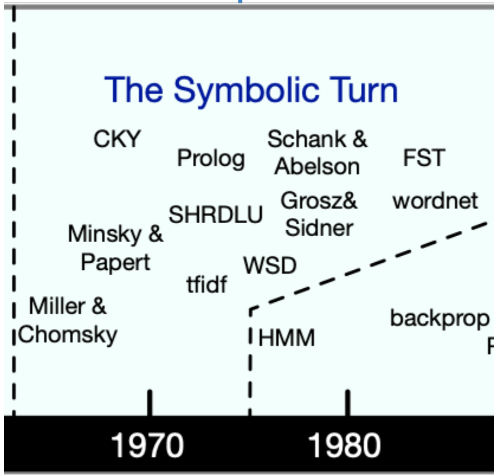
Word Embeddings (1950s):

- Psychologist Charles Osgood
- Harris and Joos in linguistics (distributional hypothesis)
- Represent meaning of a token as a point in high-dimensional space
- Can be used to do computations on meaning (e.g. word similarity)



A Brief History of NLP: The Symbolic Turn (1960s - 1988)

Era	Time (Approx.)	Dominant Idea	Computational Strategies
The Symbolic Turn	1960s -1988	Language is fundamentally structured and rule-governed; understanding requires explicit syntax, semantics, and knowledge.	grammars, rules, logic, parsers, symbolic representations



A Brief History of NLP: The Symbolic Turn (1960s - 1988)

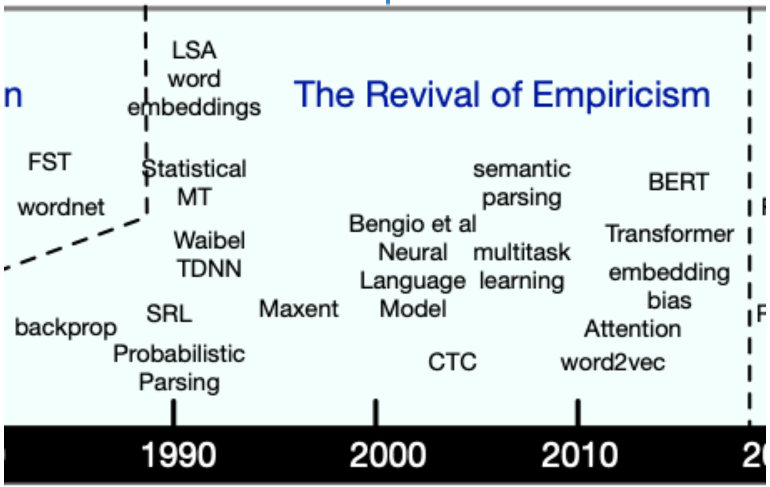
Recursive structure (e.g. embedded phrases):

- The dog barked.
- The dog **that chased the cat that frightened the mouse** barked.

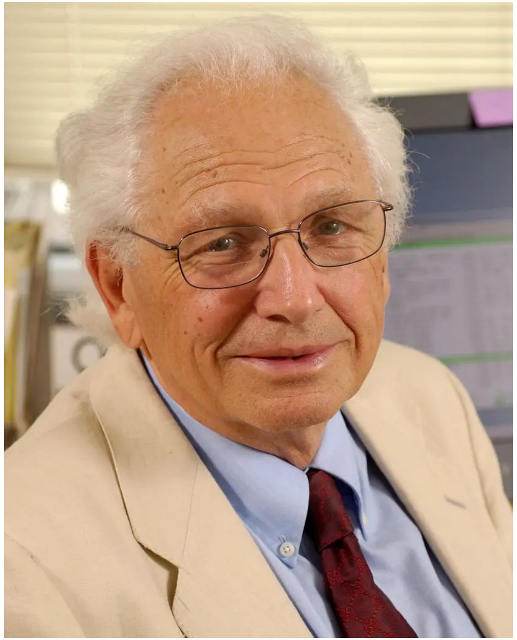
Noam Chomsky: Syntactic Structures (1957)

A Brief History of NLP: The Revival of Empiricism (1988 - 2019)

Era	Time (Approx.)	Dominant Idea	Computational Strategies
The Revival of Empiricism	1988 - 2019	Rather than hand-writing all knowledge, learn patterns from actual language data and evaluate systems empirically.	probabilistic models → statistical ML → neural networks



A Brief History of NLP: The Revival of Empiricism (1988 - 2019)



Frederick Jelinek & IBM Speech team: n-gram language models (1975)

Brown et al., 1990: "A Statistical Approach to Machine Translation"

Yoshua Bengio et al., 2003: "A Neural Probabilistic Language Model"

A Brief History of NLP: The Revival of Empiricism (1988 - 2019)

NEURAL MACHINE TRANSLATION
BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

Dzmitry Bahdanau
Jacobs University Bremen, Germany

KyungHyun Cho **Yoshua Bengio***
Université de Montréal

Attention mechanism (2015)

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Transformer (2017)

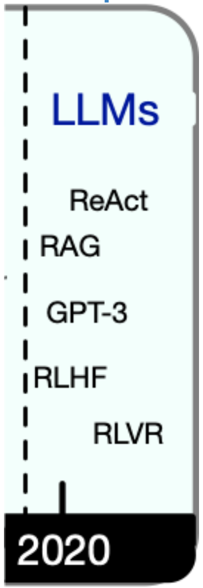
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin **Ming-Wei Chang** **Kenton Lee** **Kristina Toutanova**
Google AI Language
{jacobdevlin,mingweichang,kentonl,kristout}@google.com

BERT (2019)

A Brief History of NLP: LLMs (2019 -)

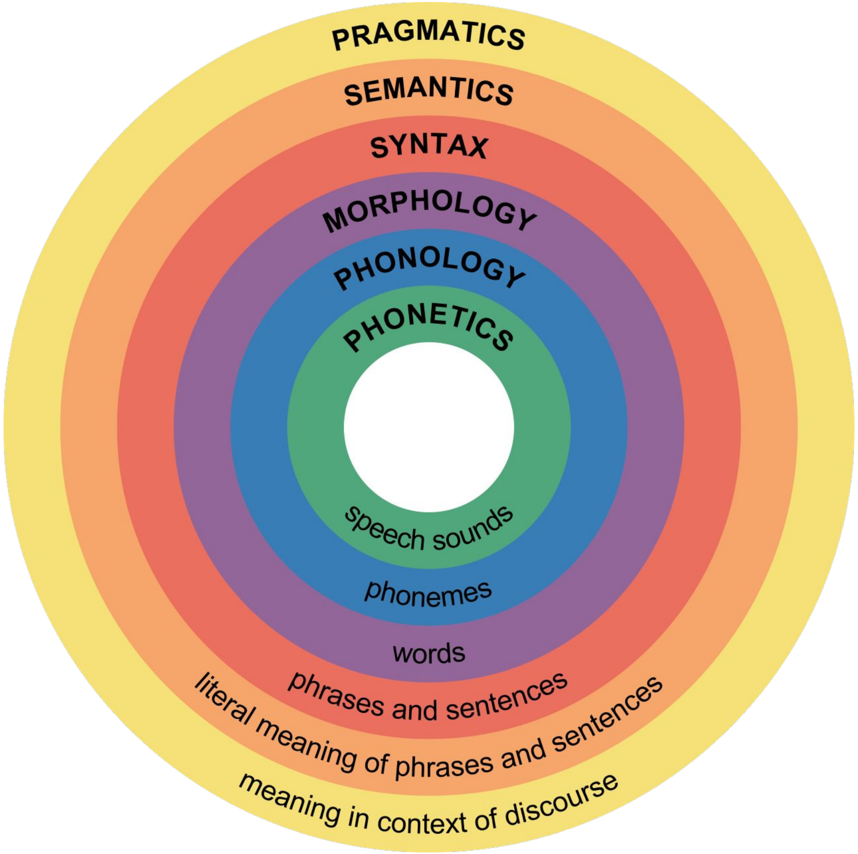
Era	Time (Approx.)	Dominant Idea	Computational Strategies
LLMs	2019 - present	A sufficiently large pretrained predictive model can acquire broad linguistic/world knowledge and perform many tasks through language itself.	transformers, pretraining, prompting, post-training, RAG, agents, reasoning RL



Let’s Consider Language…

What does it mean to “know” or “understand” a language?

Linguistic Structure



File:Major levels of linguistic structure.jpg: James J. Thomas and Kristin A. Cook (Ed.) derivative work McSush, Public domain, via Wikimedia Commons

Linguistic Structure: Morphology

A shipping-ship, shipping shipping-ships.

a [[ship-ing]-ship] [ship-ing] [[ship-ing]-ship-s]

Linguistic Structure - Syntax

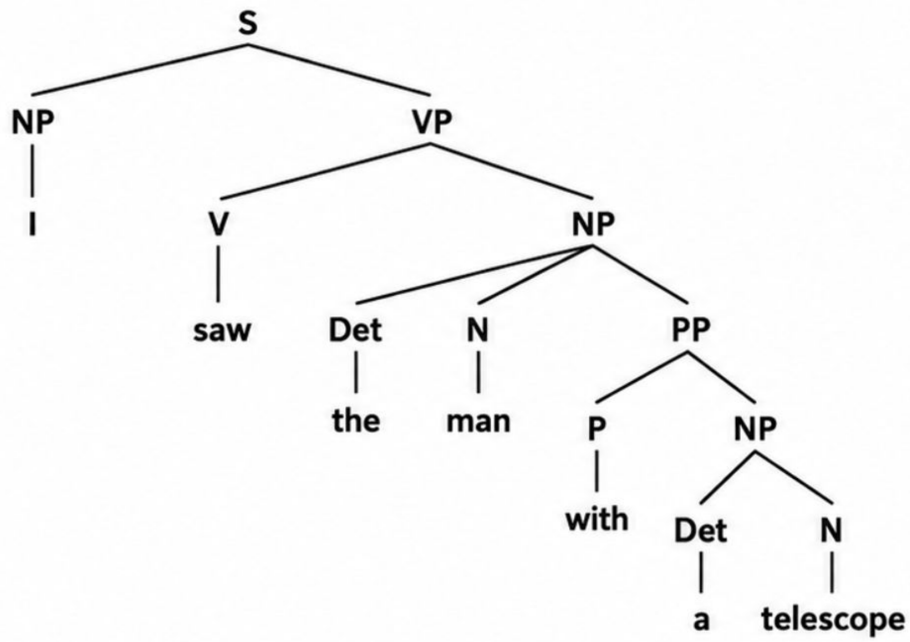
I saw the man with a telescope.

Who has the telescope?

Linguistic Structure - Syntax: PP Attachment Ambiguity

Interpretation 1: prepositional phrase modifies the noun phrase.

I saw the man with a telescope.



Bracketed Structure:

saw [the man [with the telescope]]

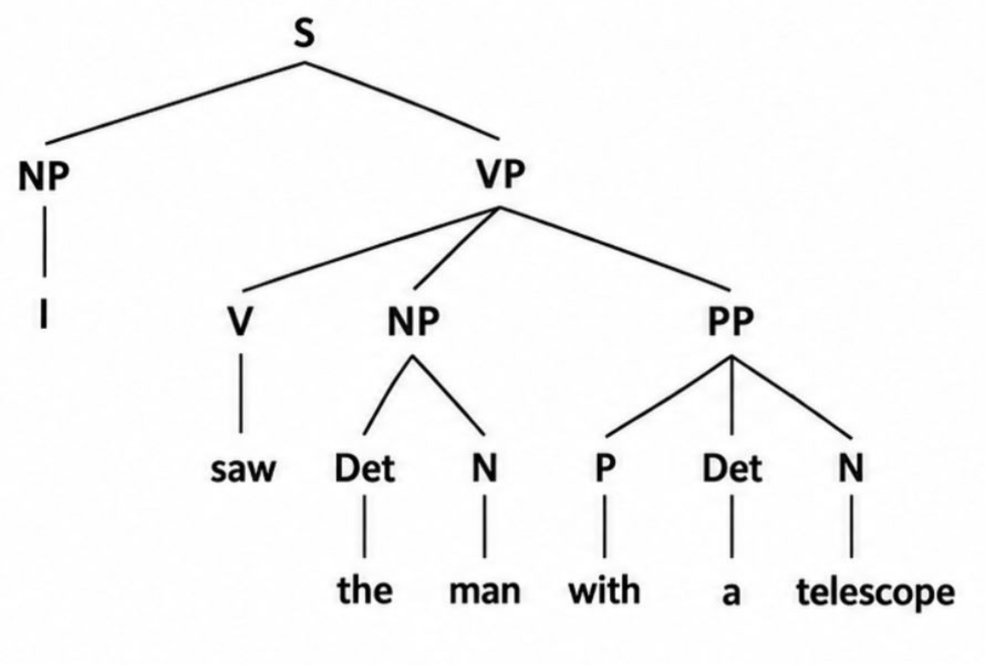
Meaning:

I saw the man who had the telescope.

Linguistic Structure - Syntax: PP Attachment Ambiguity

Interpretation 2: prepositional phrase modifies the verb phrase.

I saw the man with a telescope.



Bracketed Structure:

[saw the man] [with the telescope]

Meaning:

I used the telescope to see the man.

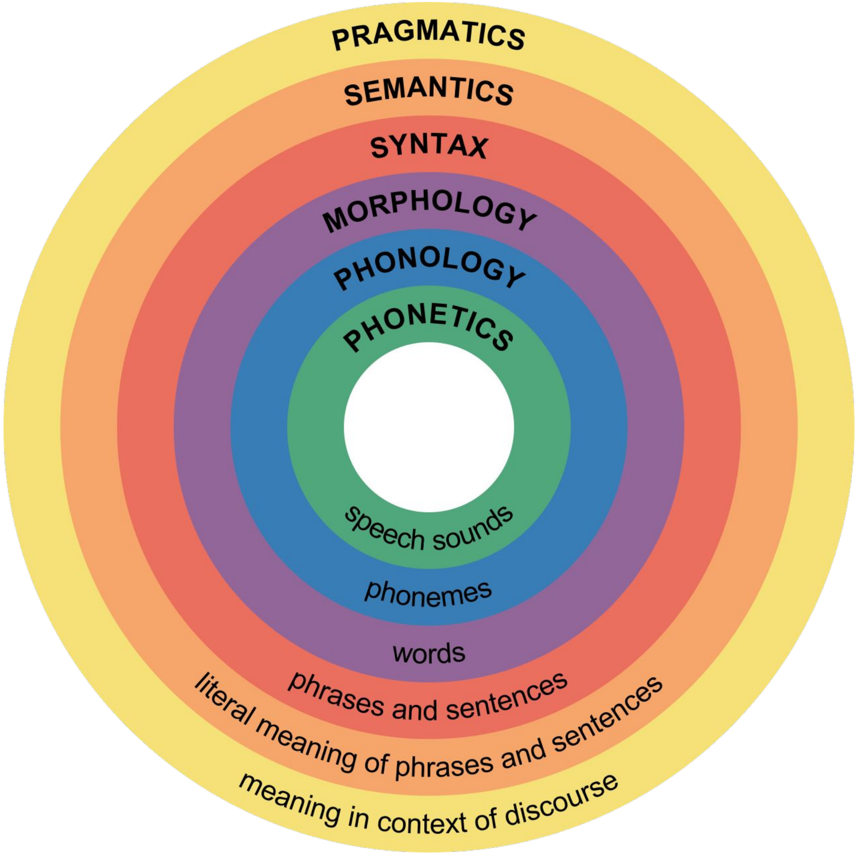
Linguistic Structure - Semantics & Pragmatics

Are you going out tonight?

I have an exam tomorrow.

Is the second speaker going out tonight?
Did they answer the question?

Linguistic Structure: Then vs. Now



File:Major levels of linguistic structure.jpg: James J. Thomas and Kristin A. Cook (Ed.) derivative work McSush, Public domain, via Wikimedia Commons

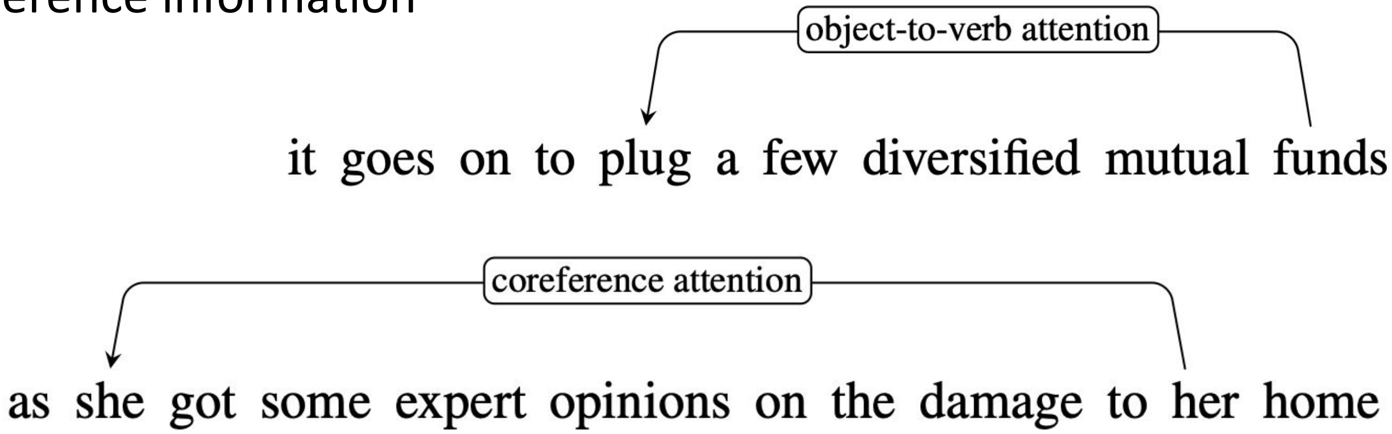
- **Early NLP:** linguistic structure was central to process of building meaning representations
- **Modern NLP:** other ways of building meaning representations
  - Word embeddings
  - **New Roles:** 1) Interpretability, 2) Applications to social and cognitive science

Linguistic Structure: Interpretability

Goal of interpretability: understand in terms of how a neural net processes language

Example: Embeddings in networks implicitly represent:

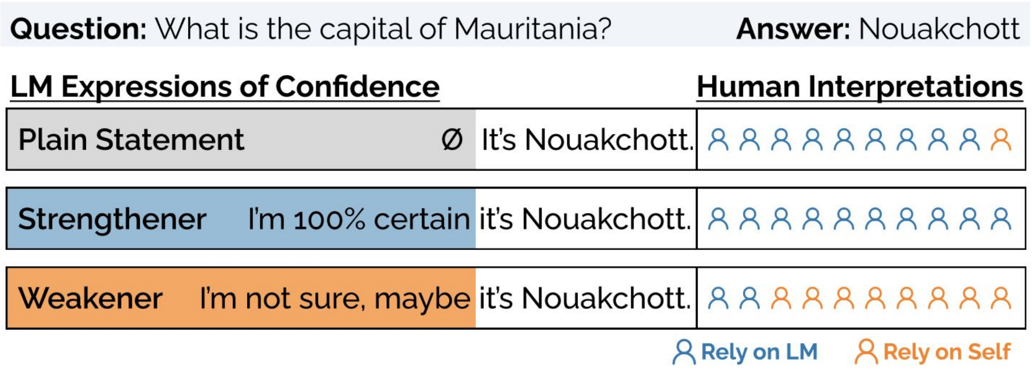
- syntactic parse trees
- semantic information
- coreference information



Linguistic Analysis of Interactions: LLM-Human Conversation

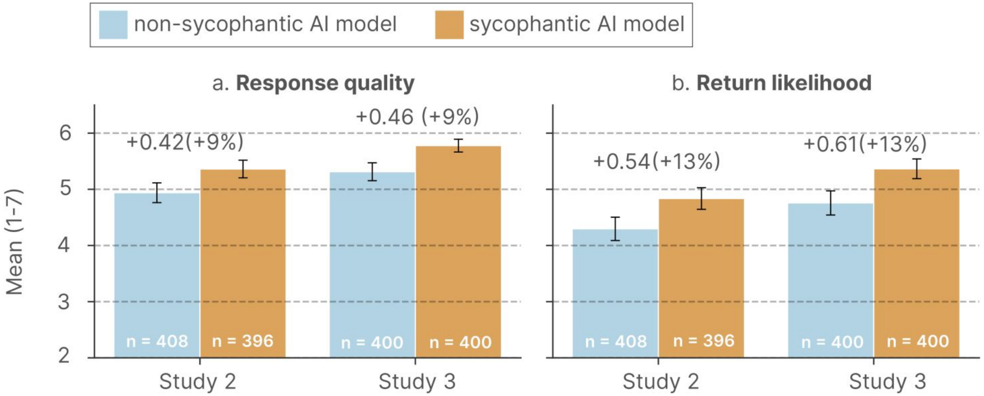
LLMs show linguistic overconfidence (Zhou et al., 2024)

- LLMs overuse strengtheners like "definitely" and avoid weakeners like "maybe



LLMs generate sycophantic language (Cheng et al., 2026)

- LLMs use language that is linguistically affirming and bolsters the user's positive self-image



Exit Survey: https://forms.gle/Xa1yksySRyN4vRXJ7

TODOs:

- Start brainstorming for your project.

Next Lecture: **Basics of Text Processing**