

Natural Language Generation

COMP7607 — Lecture 7

Lingpeng Kong

Department of Computer Science, The University of Hong Kong

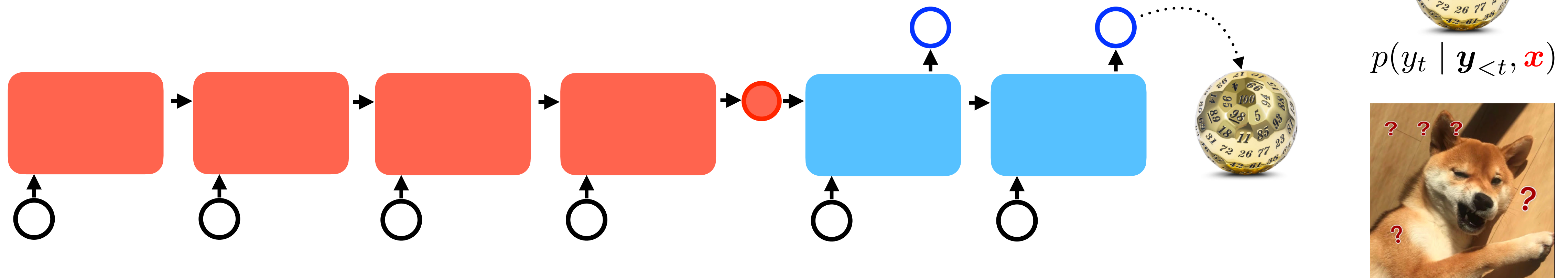
Many materials from Stanford CS224n with special thanks!

NLG as (Conditional) Language Modeling

$$p(\mathbf{y} \mid \mathbf{x}) = p(y_1 \dots y_n \mid x_1 \dots x_m) = \prod_{t=1}^n p(y_t \mid \mathbf{y}_{<t}, \mathbf{x})$$

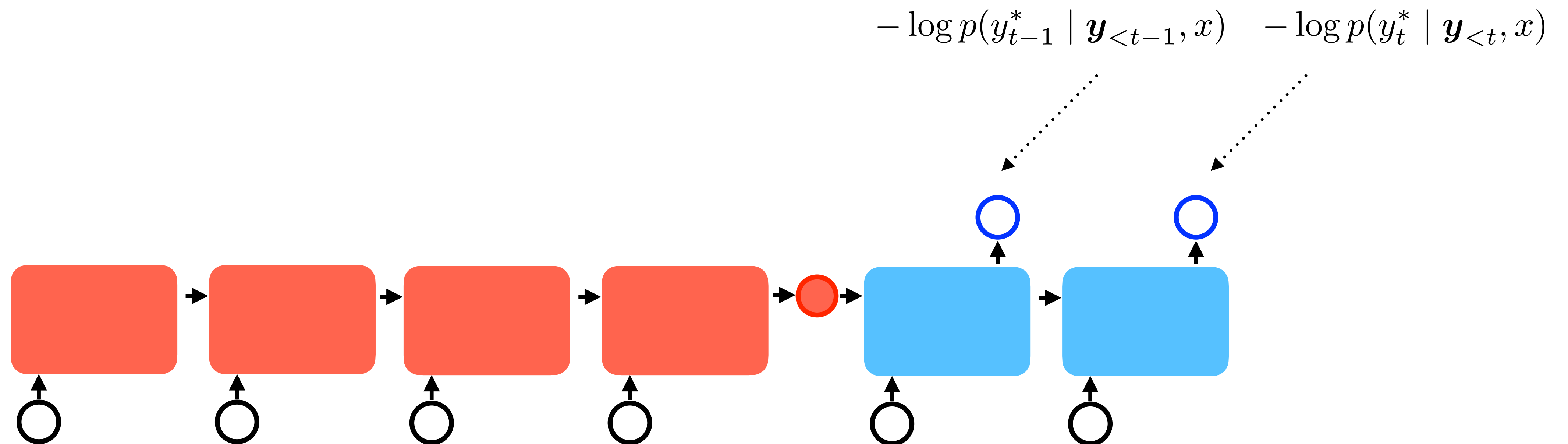
↑ ↑
target source

Conditional Language Model

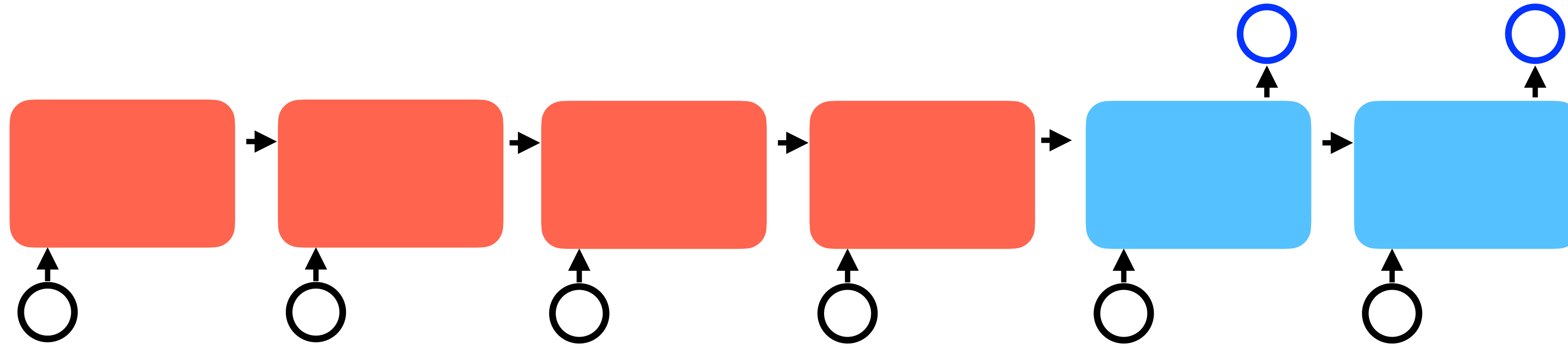


NLG as (Conditional) Language Modeling

$$p(\mathbf{y} \mid \mathbf{x}) = p(y_1 \dots y_n \mid x_1 \dots x_m) = \prod_{t=1}^n p(y_t \mid \mathbf{y}_{<t}, \mathbf{x})$$



NLG as (Conditional) Language Modeling

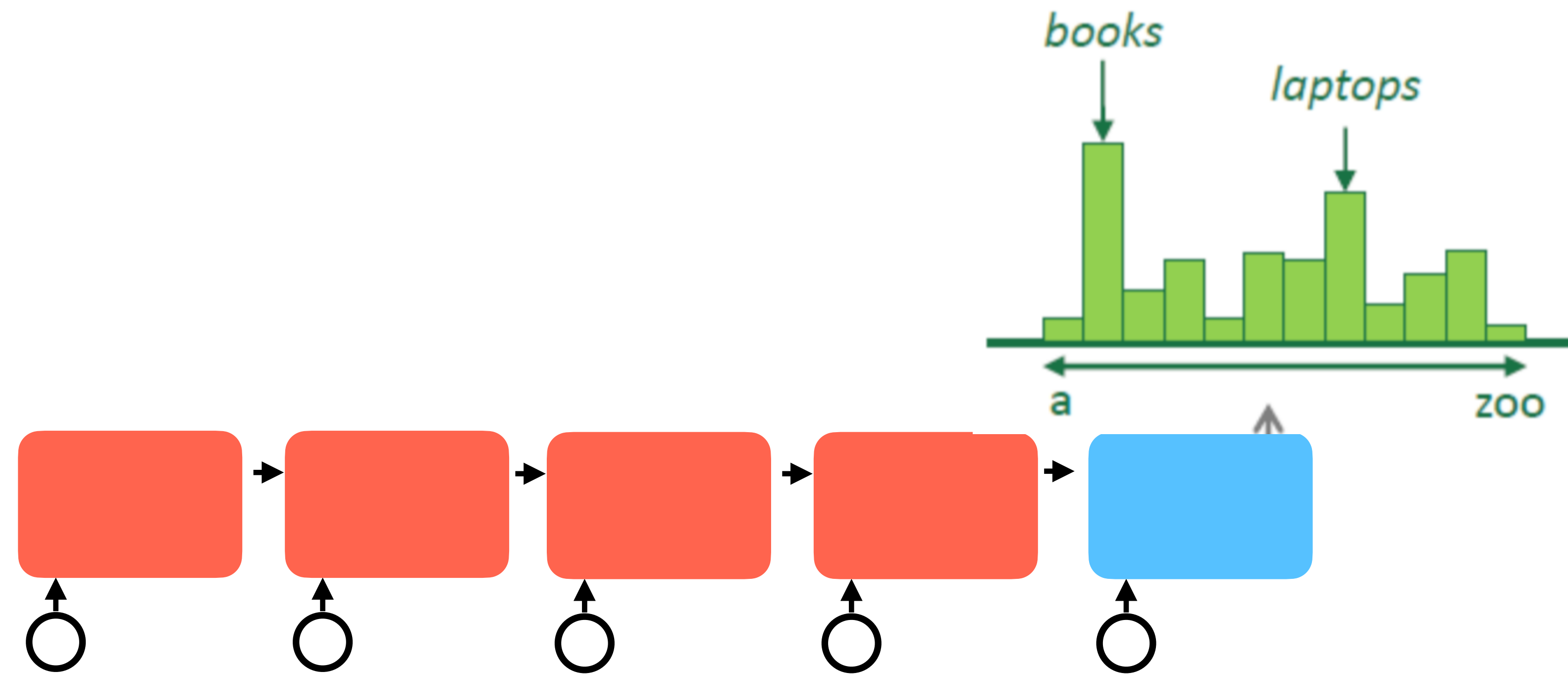


Training from scratch or starting with



? The objective is the same.

How to perform decoding?

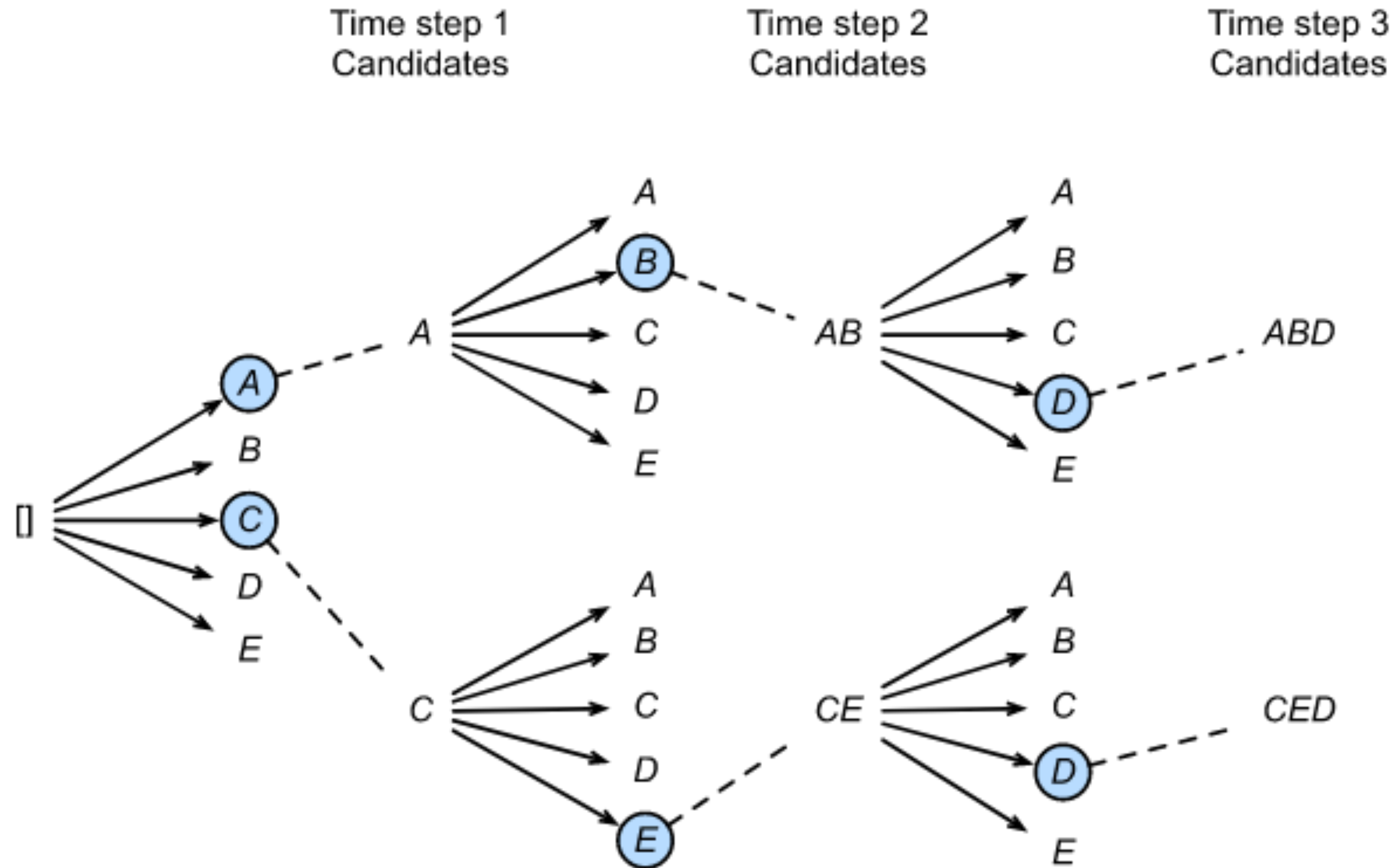


$$p(y_t \mid \mathbf{y}_{<t}, \mathbf{x})$$

arg max
beam search

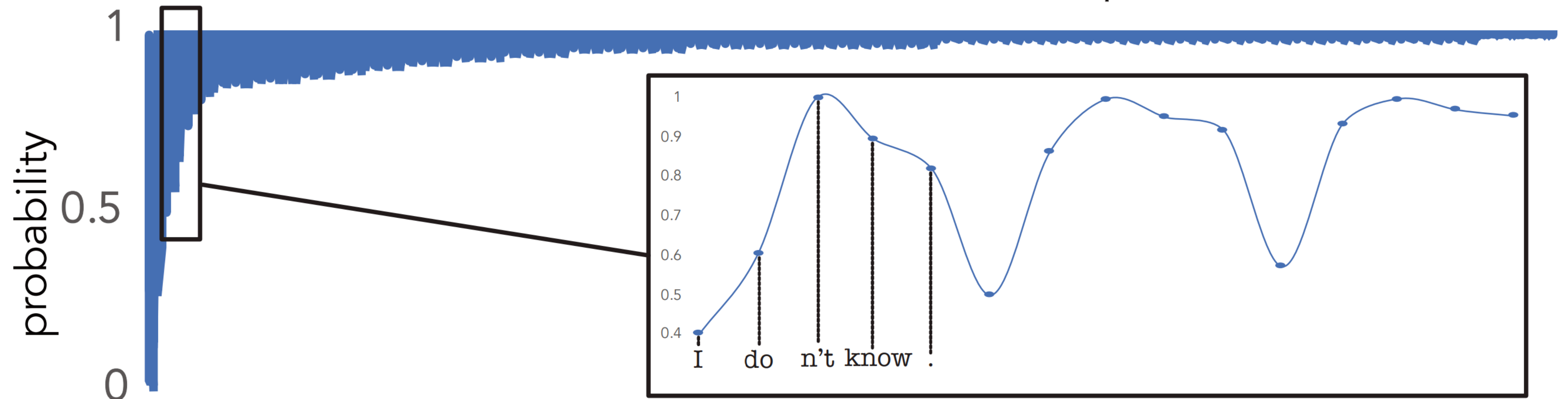


Beam Search



What happened?

Token Probabilities for "I don't know." Repeated 200 times

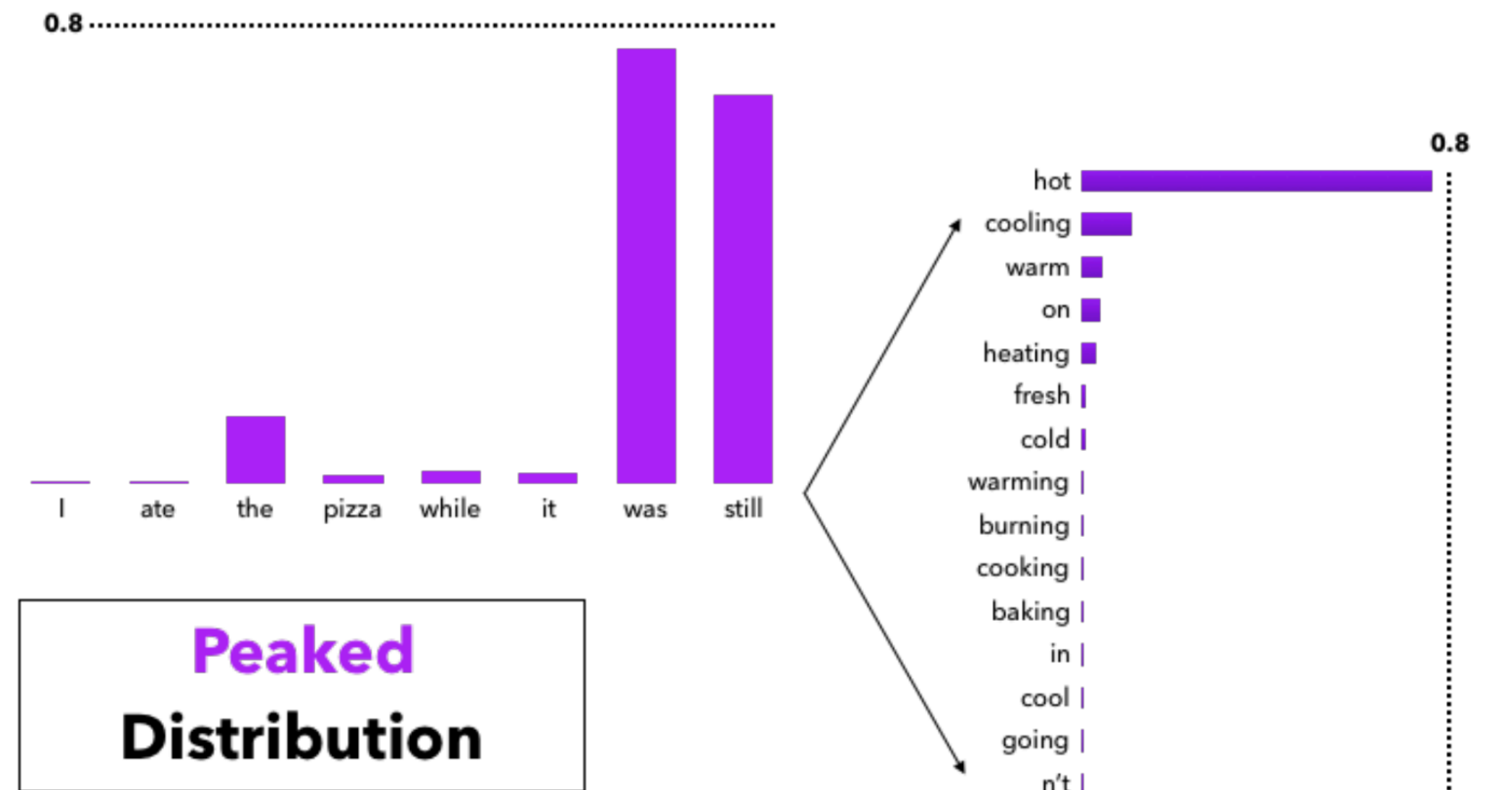
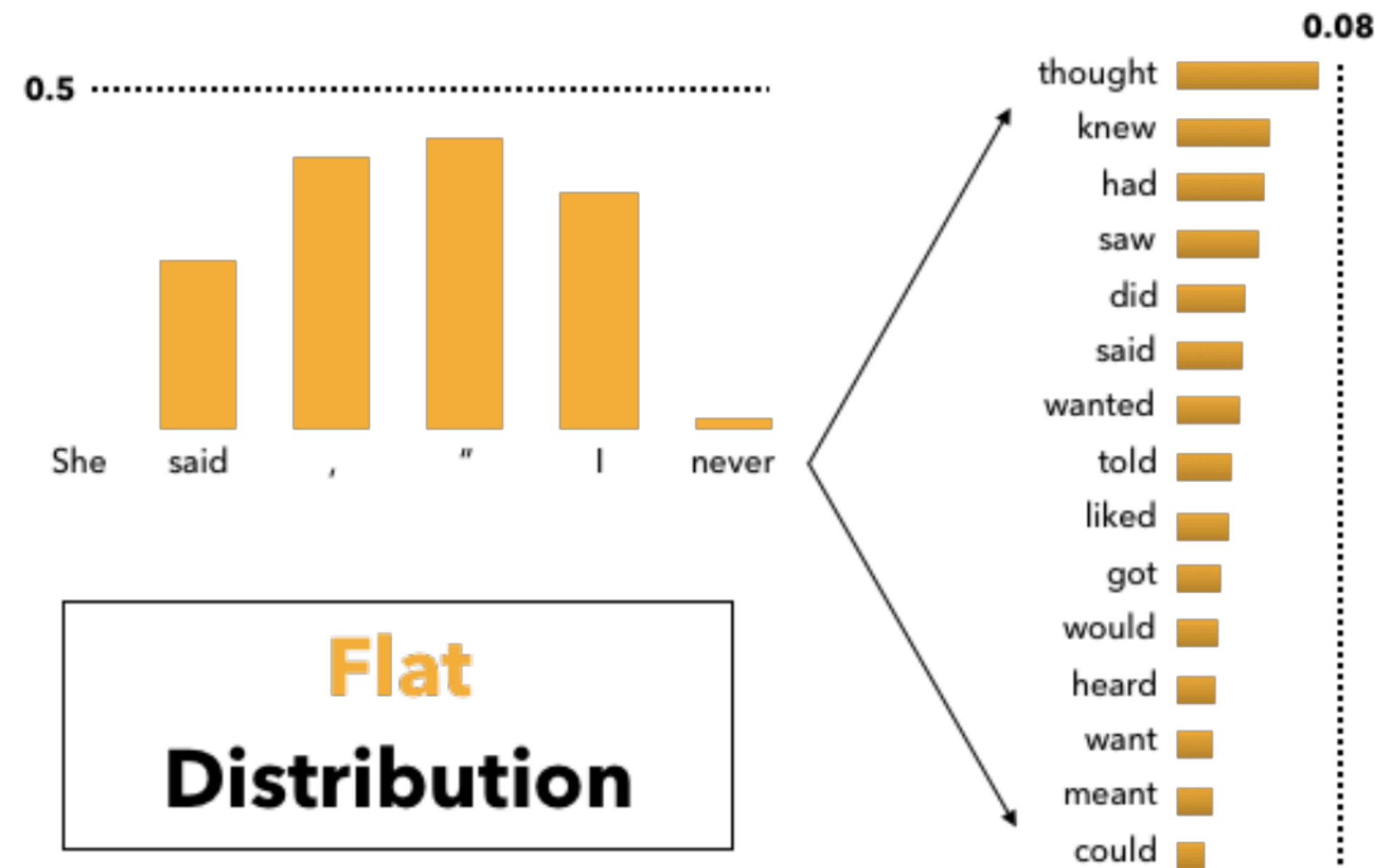


The probability of a repeated phrase increases with each repetition, creating a feedback loop.

Top-K Sampling

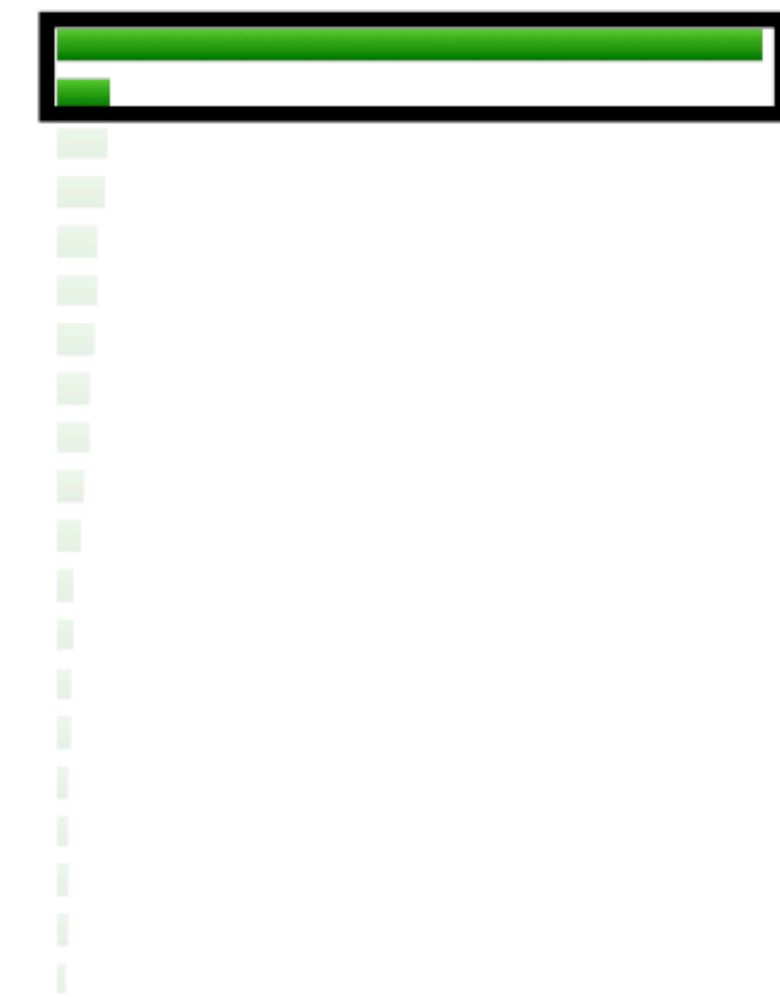
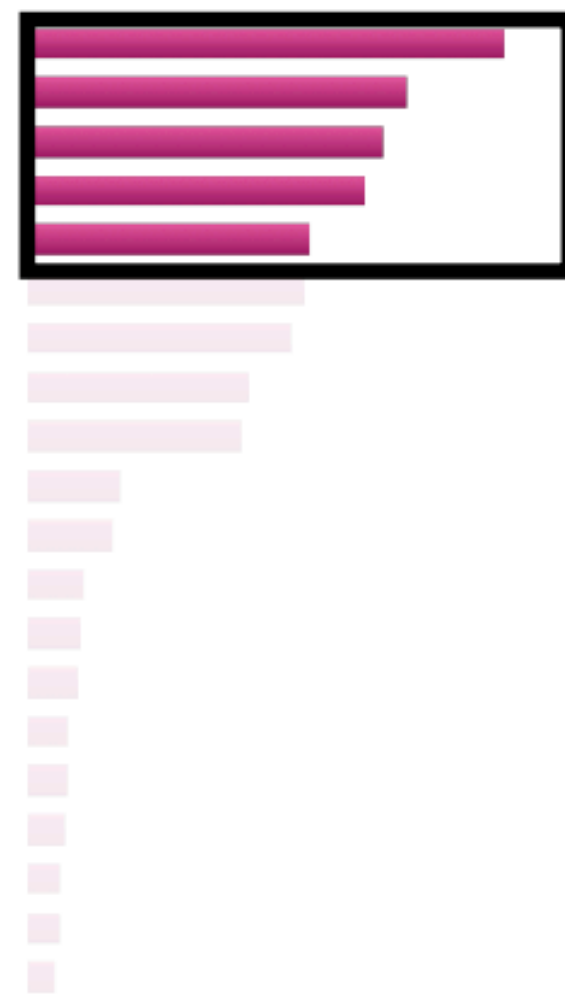


Top-K Sampling



Top-p (nucleus) Sampling

To cut off by the cumulative probability mass, rather than the first K terms.



Question Answering



Question Answering is (almost) NLP Complete

I: Jane went to the hallway.

I: Mary walked to the bathroom.

I: Sandra went to the garden.

I: Daniel went back to the garden.

I: Sandra took the milk there.

Q: Where is the milk?

A: garden

I: It started boring, but then it got interesting.

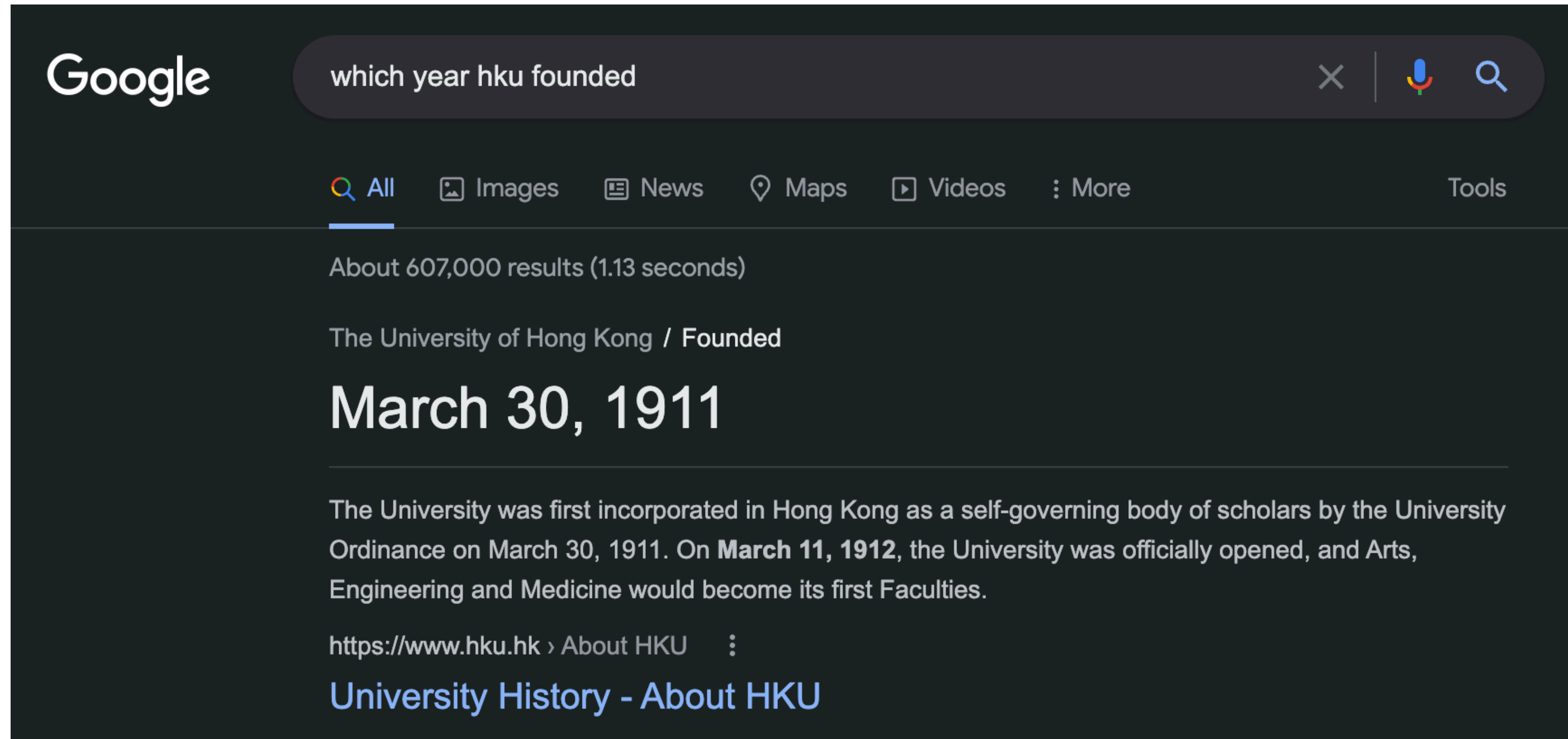
Q: What's the sentiment?

A: positive

Q: POS tags?

A: PRP VBD JJ , CC RB PRP VBD JJ .

Question Answering



The image shows a screenshot of a Google search interface. The search bar contains the text "which year hku founded". Below the search bar, there are navigation links for "All", "Images", "News", "Maps", "Videos", and "More". The search results show "About 607,000 results (1.13 seconds)". The main result is titled "The University of Hong Kong / Founded" and displays the date "March 30, 1911" in a large font. Below the date, there is a paragraph of text: "The University was first incorporated in Hong Kong as a self-governing body of scholars by the University Ordinance on March 30, 1911. On **March 11, 1912**, the University was officially opened, and Arts, Engineering and Medicine would become its first Faculties." At the bottom, there is a link to "https://www.hku.hk > About HKU" and a sub-link "University History - About HKU".

Google

which year hku founded

All Images News Maps Videos More Tools

About 607,000 results (1.13 seconds)

The University of Hong Kong / Founded

March 30, 1911

The University was first incorporated in Hong Kong as a self-governing body of scholars by the University Ordinance on March 30, 1911. On **March 11, 1912**, the University was officially opened, and Arts, Engineering and Medicine would become its first Faculties.

<https://www.hku.hk> > About HKU

[University History - About HKU](#)

Question Answering

The image shows a Google search interface with a dark theme. The search bar contains the text "which year hku founded". Below the search bar, there are navigation options: "All", "Images", "News", "Maps", "Videos", "More", and "Tools". The search results show "About 607,000 results (1.13 seconds)". The main result is titled "The University of Hong Kong / Founded" and features the date "March 30, 1911" in a large, bold font. Below this, a paragraph of text reads: "The University was first incorporated in Hong Kong as a self-governing body of scholars by the University Ordinance on March 30, 1911. On March 11, 1912, the University was officially opened, and Arts, Engineering and Medicine would become its first Faculties." The phrase "On March 11, 1912" is highlighted with a red rectangular box. A white speech bubble with a black outline points to this box and contains the text "This is bold." Below the paragraph, there is a URL "https://www.hku.hk > About HKU" and a link "University History - About HKU".

Google

which year hku founded

All Images News Maps Videos More Tools

About 607,000 results (1.13 seconds)

The University of Hong Kong / Founded

March 30, 1911

The University was first incorporated in Hong Kong as a self-governing body of scholars by the University Ordinance on March 30, 1911. **On March 11, 1912**, the University was officially opened, and Arts, Engineering and Medicine would become its first Faculties.

https://www.hku.hk > About HKU

University History - About HKU

Question Answering

The University was first incorporated in Hong Kong as a self-governing body of scholars by the University Ordinance on March 30, 1911. On March 11, 1912, the University was officially opened, and Arts, Engineering and Medicine would become its first Faculties.

To repeal and replace the University Ordinance 1911.

[1 September 1958] *G.N.A. 54 of 1958*
(*Format changes—E.R. 5 of 2021*)

1. **Short title**

This Ordinance may be cited as the University of Hong Kong Ordinance.

(*Amended 1 of 1968 s. 2*)

2. **Interpretation**

(1) In this Ordinance, unless the context otherwise requires—

Chancellor (校監) means the Chancellor of the University and also any person acting as Chancellor of the University by virtue of section 12(3);

Court (校董會), **Council** (校務委員會), **Senate** (教務委員會) and **Convocation** (畢業生議會) respectively mean the Court, the Council, the Senate and Convocation of the University; (*Replaced 1 of 1968 s. 3. Amended 80 of 1997 s. 100*)

officers (主管人員) and **teachers** (教師) respectively mean the officers and teachers of the University;

Pro-Chancellor (副校監) means the Pro-Chancellor of the University; (*Added 1 of 1968 s. 3*)

statutes (規程) means the statutes of the University;

University (大學) means the University of Hong Kong;

Vice-Chancellor (校長), **Deputy Vice-Chancellor** (首席副校長), **Pro-Vice-Chancellors** (副校長), **Treasurer** (司庫), **Deans of Faculties** (學院院長), **Registrar** (教務長) and **Librarian** (圖書館館長) respectively mean the Vice-Chancellor, Deputy Vice-Chancellor, Pro-Vice-Chancellors, Treasurer, Deans of Faculties, Registrar and Librarian of the University. (*Replaced 22 of 1992 s. 2*)

(2) For the purposes of this Ordinance, **good cause** (好的因由), when used with reference to removal from office, membership or place and with reference to such of the members of the University and of the officers and teachers as shall be designated by the Ordinance and statutes, means inability to perform efficiently the duties of the office, neglect of duty, or such misconduct, whether in an official or a private capacity, as renders the holder unfit to continue in office. (*Amended 66 of 1988 s. 2; E.R. 5 of 2021*)

QA evaluates machines' understanding

incorporate verb

🔖 Save Word

in·cor·po·rate | \ in-'kôr-pə-,rāt \

incorporated; incorporating

Definition of *incorporate* (Entry 1 of 2)

transitive verb

- a** : to unite or work into something already existent so as to form an indistinguishable whole
 - b** : to blend or combine thoroughly
- a** : to form into a legal corporation
 - b** : to admit to membership in a corporate body
- 3** : to give material form to : **EMBODY**

intransitive verb

- 1** : to unite in or as one body
- 2** : to form or become a corporation

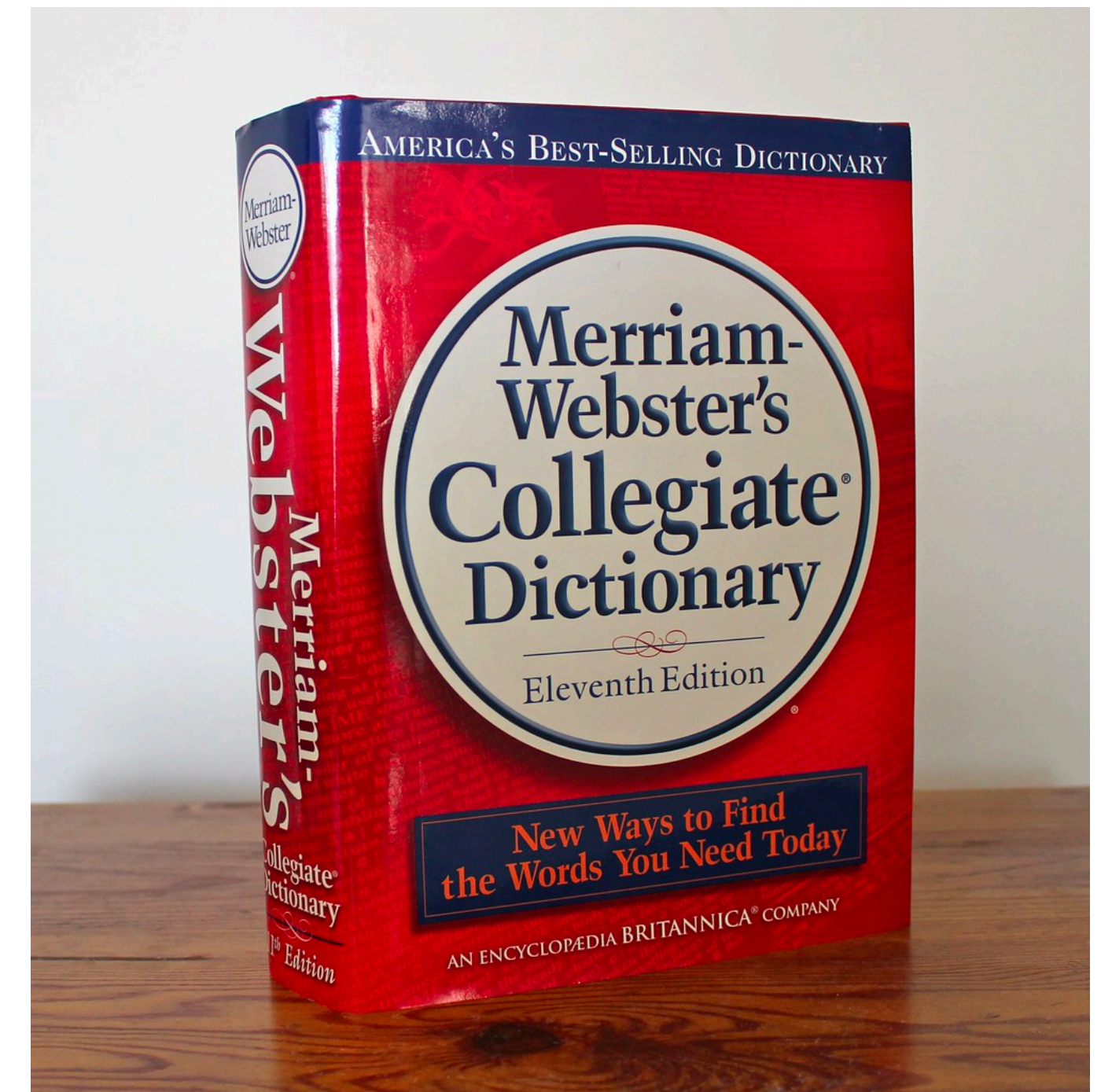
ordinance noun

🔖 Save Word

or-di-nance | \ 'örd-nən(t)s \, 'ör-də-nən(t)s \

Definition of *ordinance*

- a** : an authoritative decree or direction : **ORDER**
// On that day the king signed three ordinances.
 - b** : a law set forth by a governmental authority
specifically : a municipal regulation
// A city ordinance forbids construction work to start before 8 a.m.
- 2** : something ordained or decreed by fate or a deity
// Let ordinance come as the gods foresay [=foretell] it.
— William Shakespeare
- 3** : a prescribed usage, practice, or ceremony
// observes the ordinance of abstinence during Lent



QA seeks information



what is cadenza in music



[All](#) [Images](#) [Videos](#) [News](#) [Shopping](#) [More](#) [Tools](#)

About 8,040,000 results (0.39 seconds)

<https://en.wikipedia.org/wiki/Cadenza>

Cadenza - Wikipedia

In **music**, a **cadenza** is, generically, an improvised or written-out ornamental passage played or sung by a soloist or soloists, usually in a "free" rhythmic ...

[In concerti](#) · [As a vocal flourish](#) · [In jazz](#) · [Notable examples](#)



Cadenza



In music, a cadenza is, generically, an improvised or written-out ornamental passage played or sung by a soloist or soloists, usually in a "free" rhythmic style, and often allowing virtuosic display. During this time the accompaniment will rest, or sustain a note or chord. [Wikipedia](#)

[Feedback](#)

“Easy”, “Hard”, “Good”, and “Bad” Questions

To be, or not to be, that is the question.

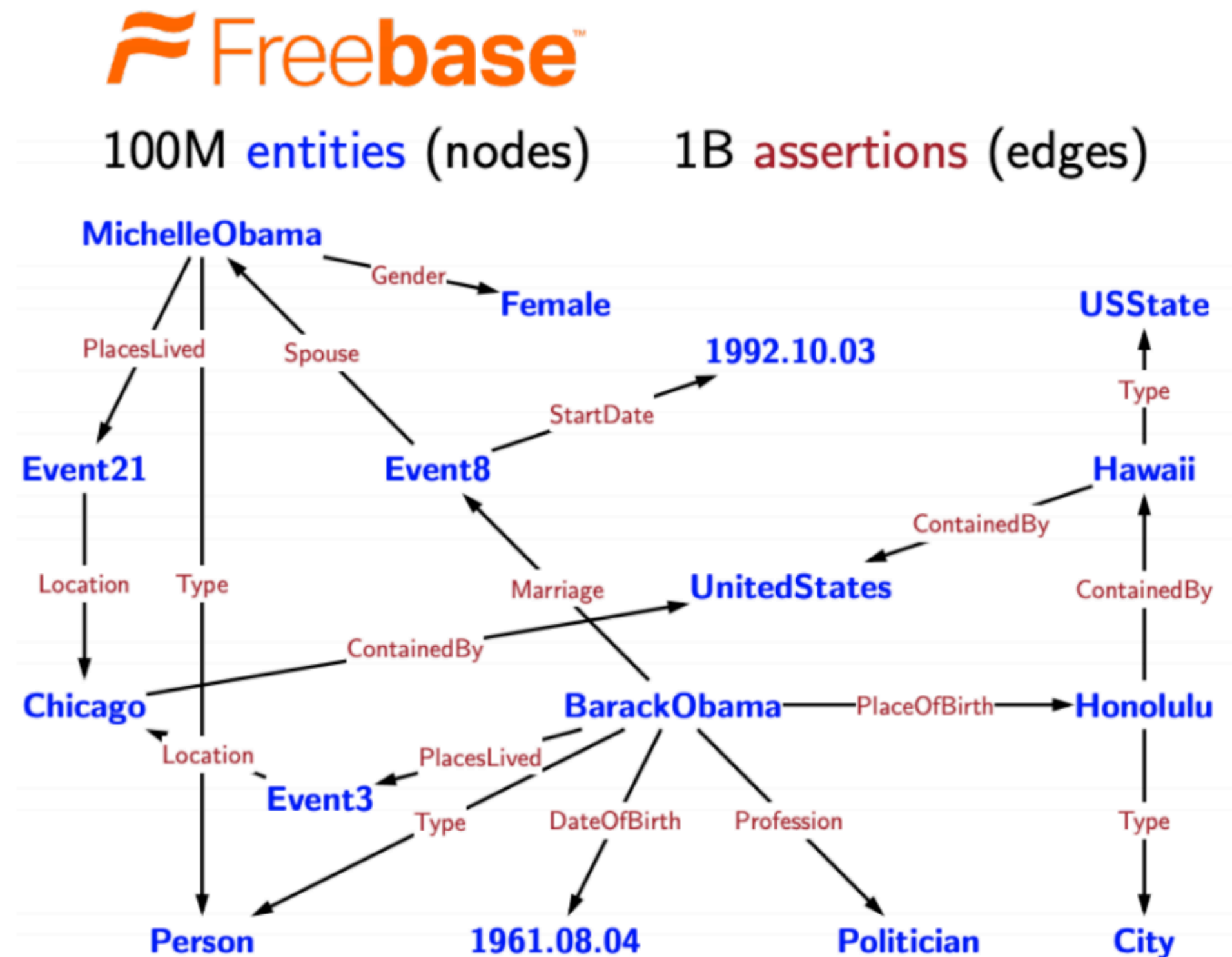
When will COVID-19 disappear?

What were the main causes for World War II?

When did World War III start?

What is the best restaurant in Hong Kong?

Knowledge Based QA



Query Window

```

?- likes(john, jane). ← dot necessary
true. ← answer from prolog interpreter

sign on
prolog query
prompt
↓
?- friends(X, Y).
X = john,
Y = jane ; ← type ; to get next solution
X = jane,
Y = john.
    
```

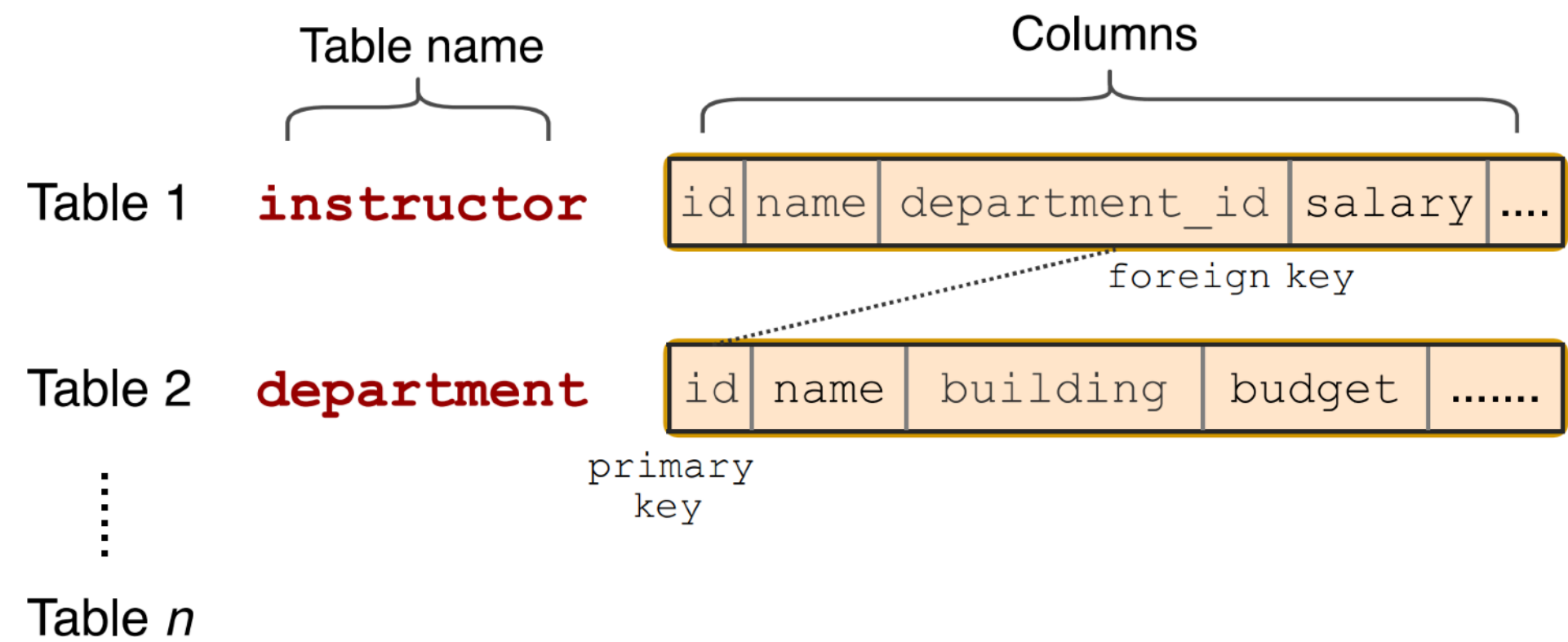
variables

Prolog (Programming Language), 1972

$\lambda x.type(x, LOCATION) \wedge born_in(BARACKOBAMA, x)$

Natural Language to Structured Query

Annotators check database schema (e.g., database: college)



Annotators create:

Complex question What are the name and budget of the departments with average instructor salary greater than the overall average?

Complex SQL

```
SELECT T2.name, T2.budget
FROM instructor as T1 JOIN department as
T2 ON T1.department_id = T2.id
GROUP BY T1.department_id
HAVING avg(T1.salary) >
(SELECT avg(salary) FROM instructor)
```

Spider 1.0



Yale Semantic Parsing and Text-to-SQL Challenge

Yu et al., 2019

SQuAD Task

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?

gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

graupel

Where do water droplets collide with ice crystals to form precipitation?

within a cloud

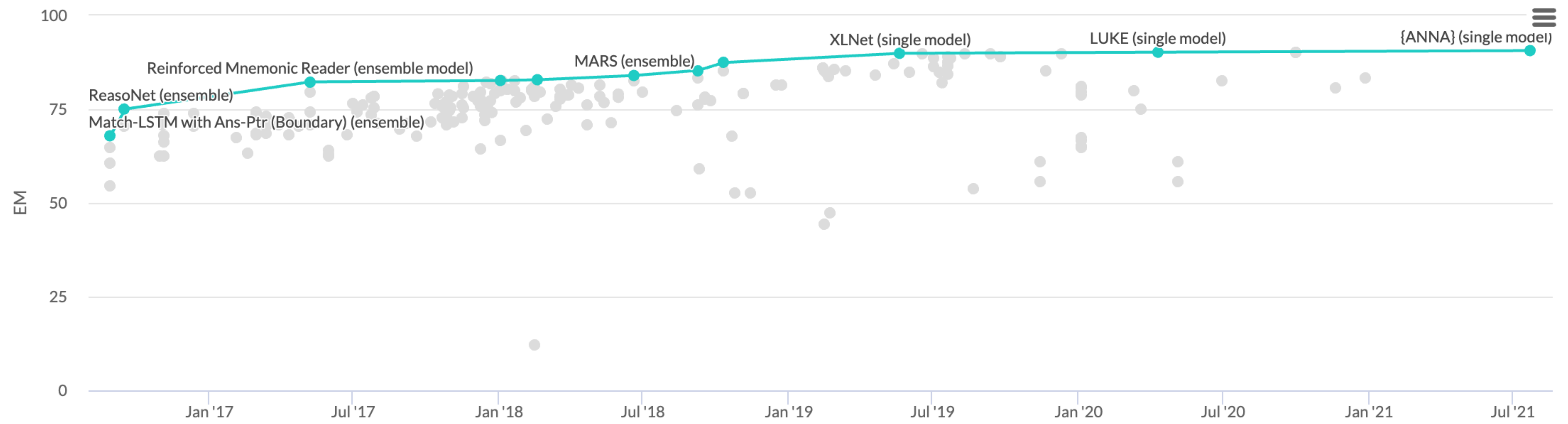
100k annotated (passage, question, answer) triples.

Answers are spans in the passage.

SQuAD Leader Board

Leaderboard Dataset Description

View EM by Date



Rajpurkar et al., 2016

SQuAD 2.0 Task

Article: Endangered Species Act

Paragraph: “ ... *Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the Bald Eagle Protection Act of 1940. These later laws had a low cost to society—the species were relatively rare—and little opposition was raised.*”

Question 1: “*Which laws faced significant opposition?*”

Plausible Answer: *later laws*

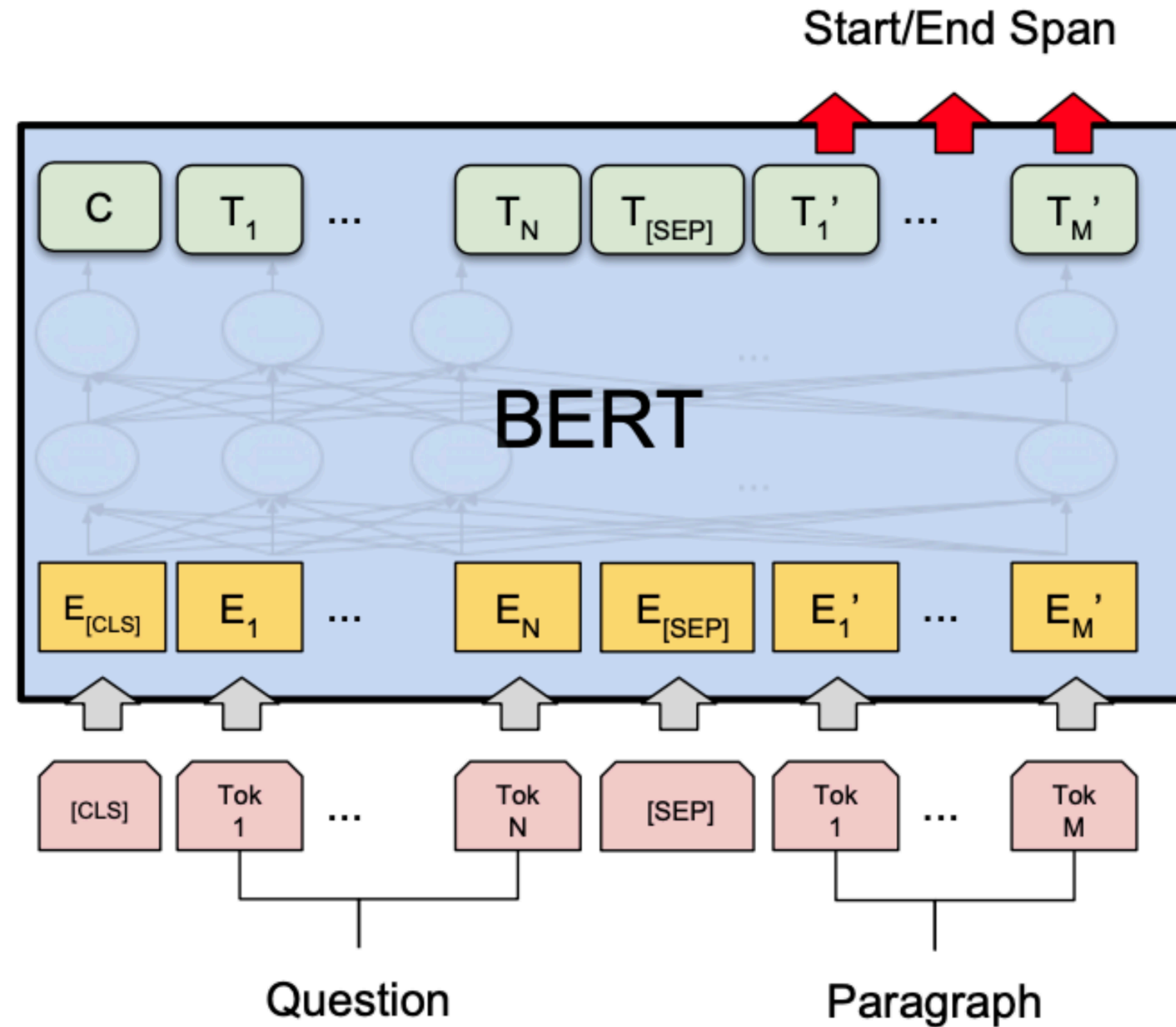
Question 2: “*What was the name of the 1937 treaty?*”

Plausible Answer: *Bald Eagle Protection Act*

Unanswerable Questions

Plausible (but incorrect) answers

BERT QA



Fine-tune on the SQuAD dataset.

$$-\log p_{\text{start}}(s^*) - \log p_{\text{end}}(e^*)$$



Multi-hop Reasoning QA

Question: *What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?*

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States

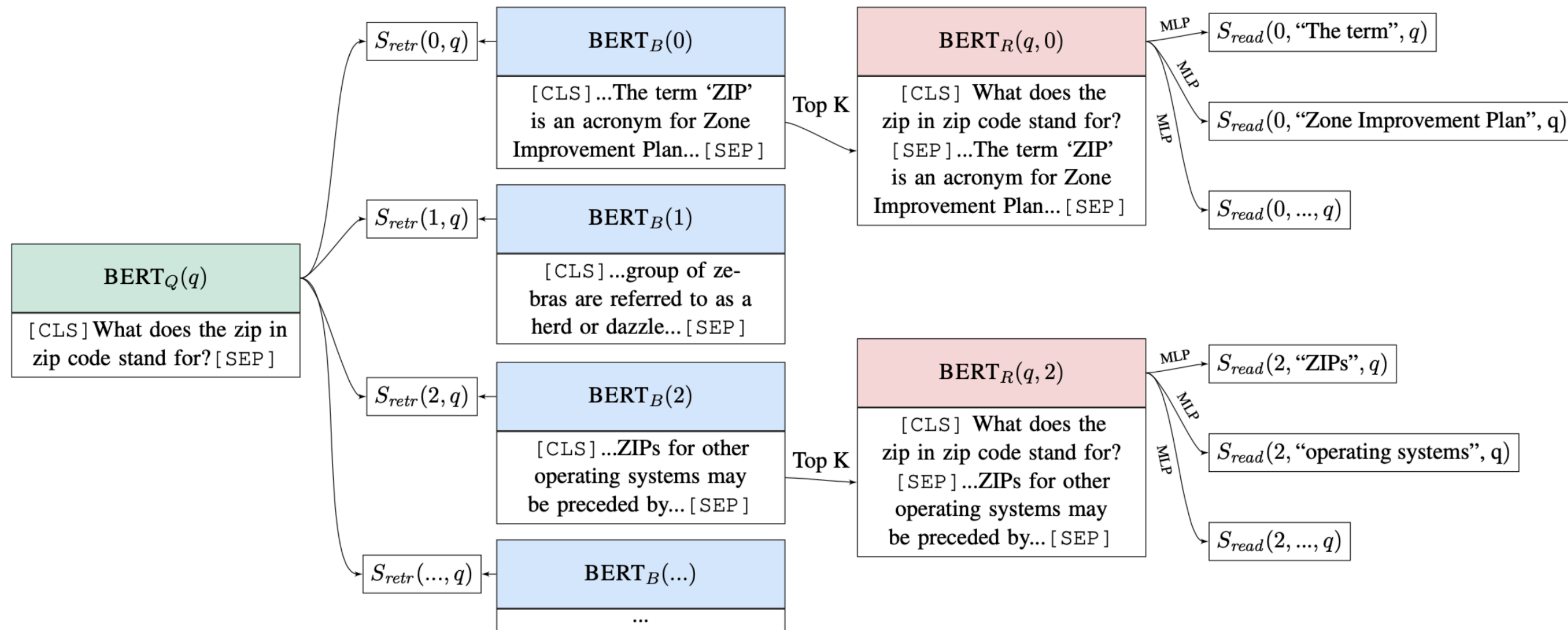
Same entity

Same entity

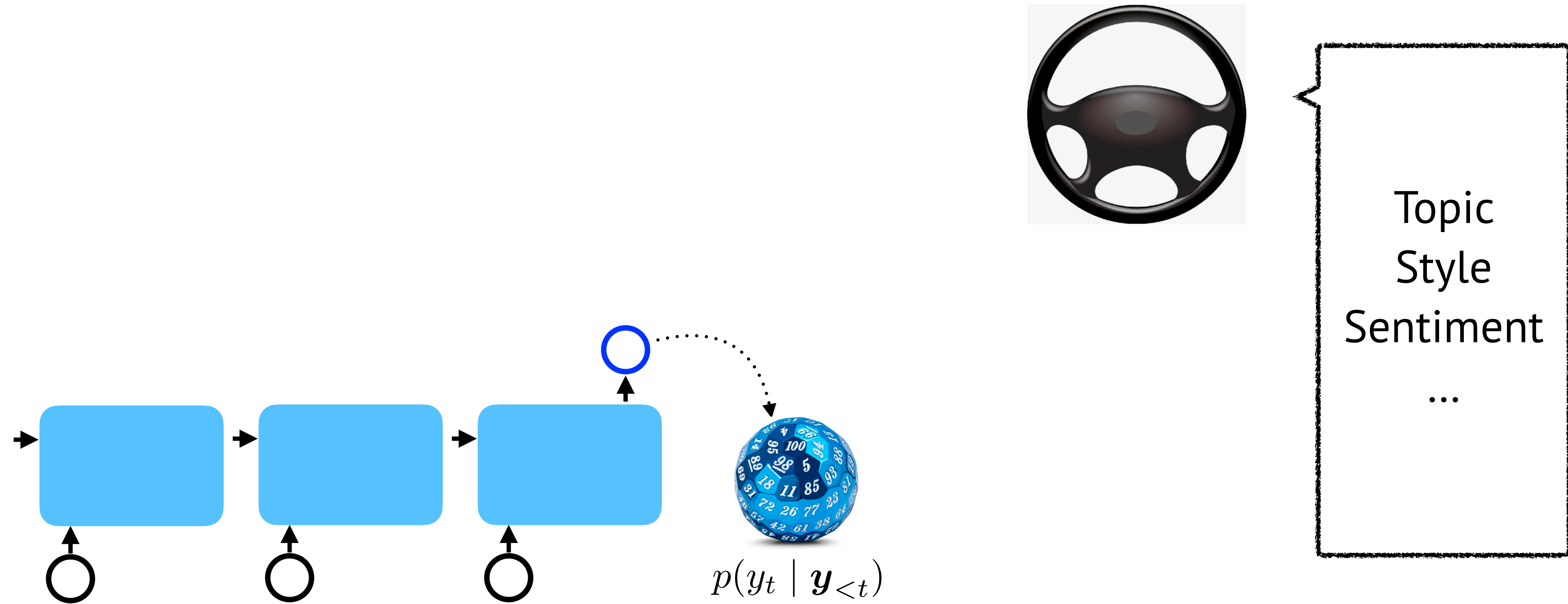
Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...

Latent Retrieval

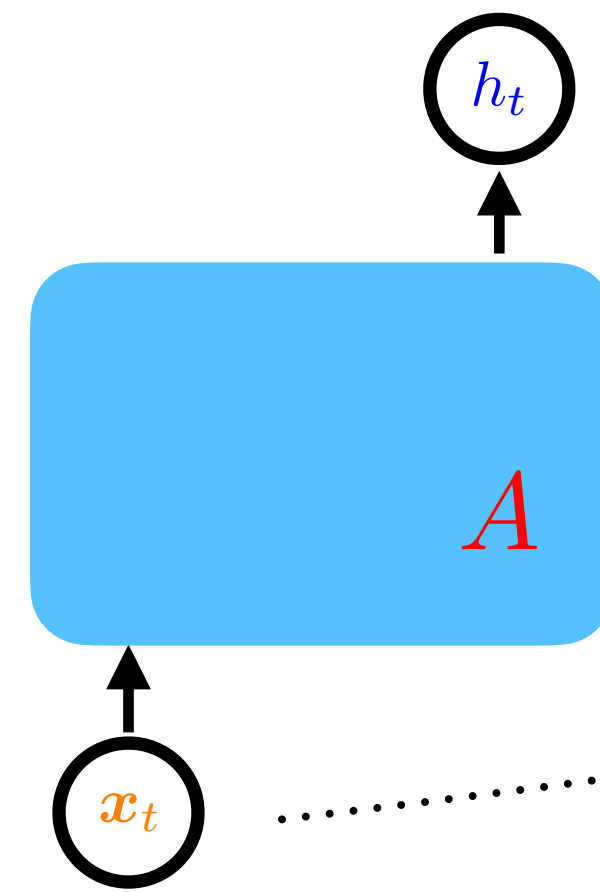


Controllable Text Generation



Full Fine-tuning

An idea starts from initialization:

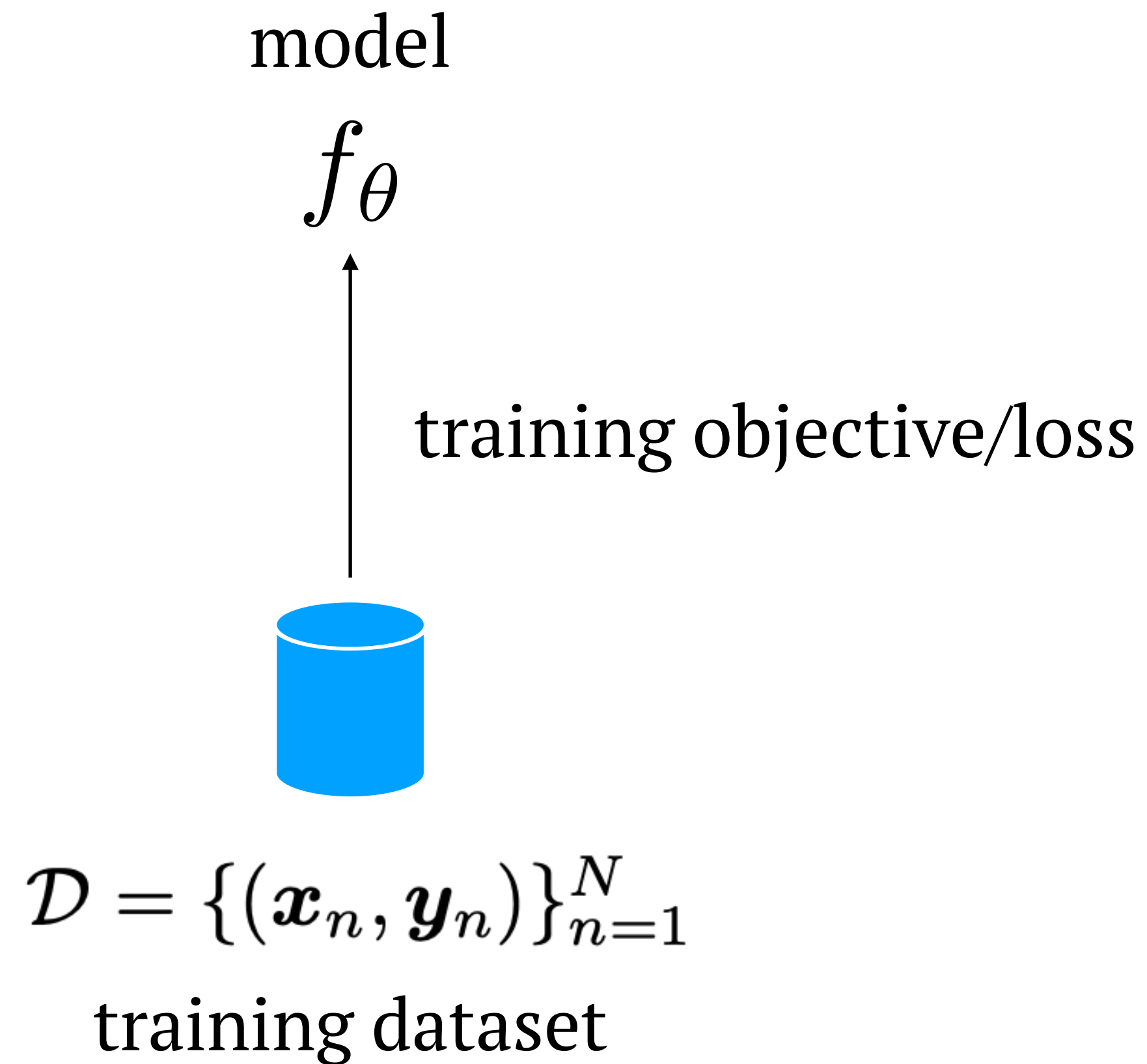


GloVe, Word2vec, ...

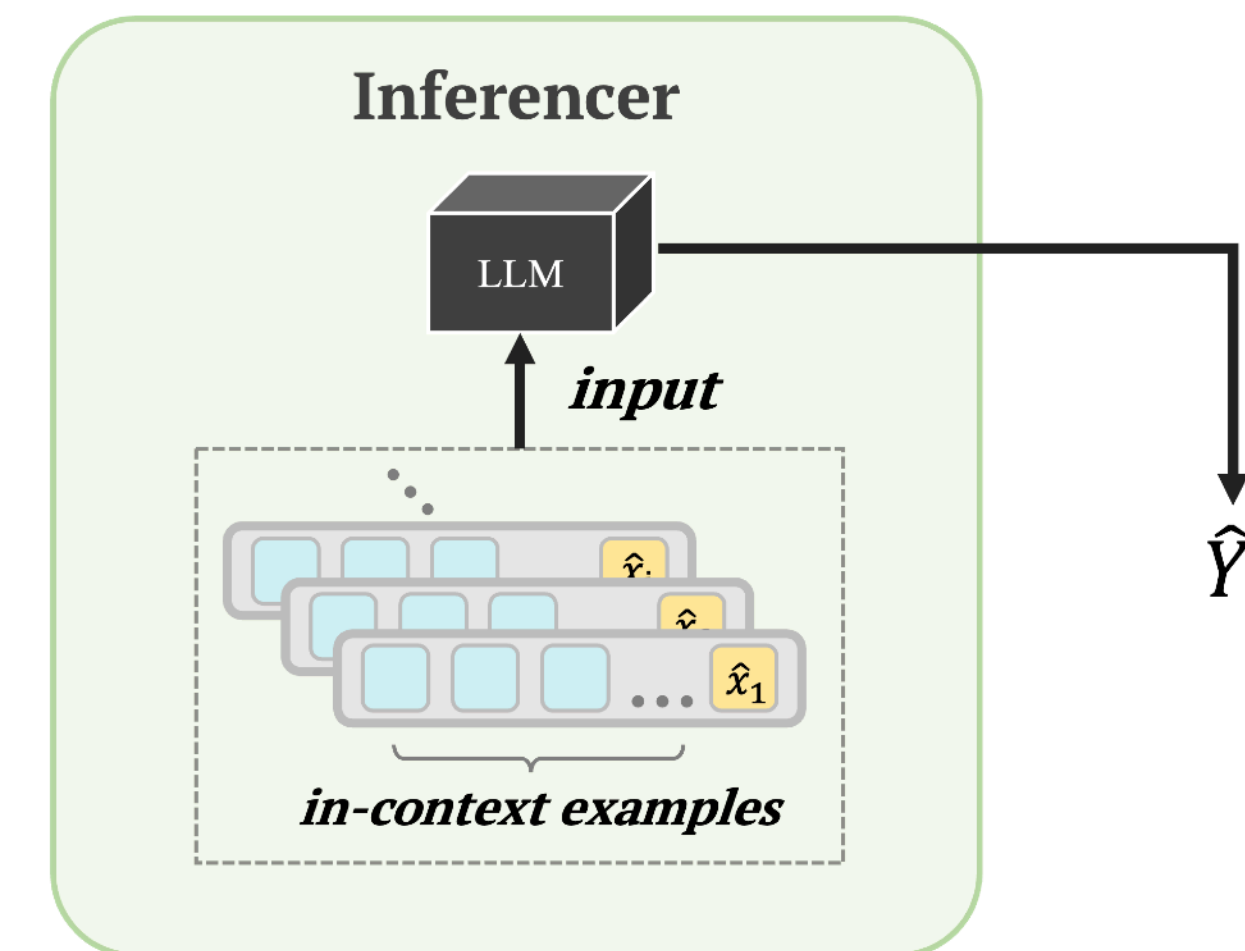
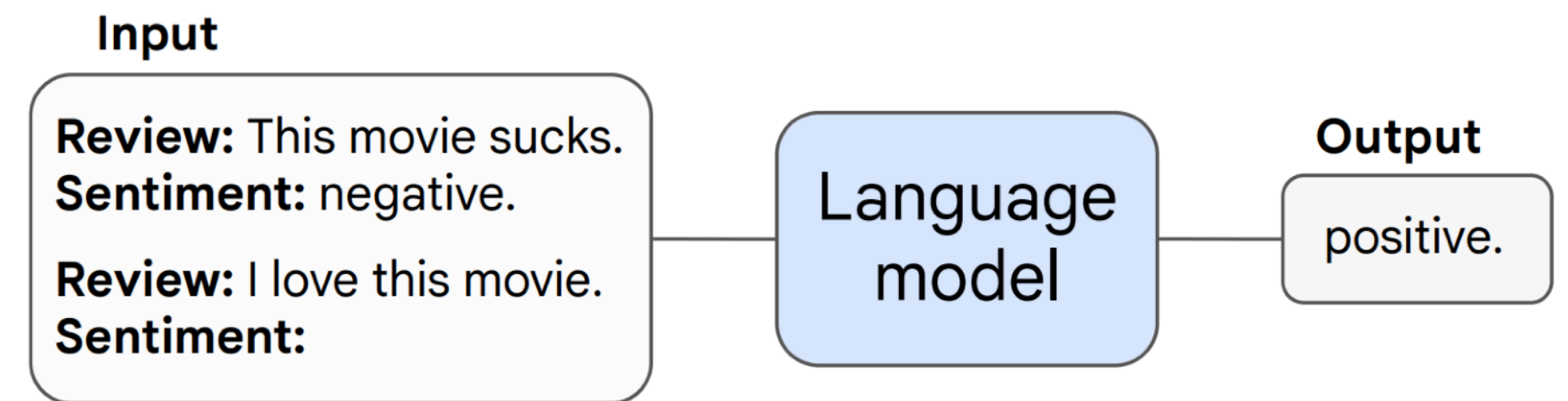
A full-supervised learning task.

Pretrained General model -> Task Specific Model

In-context Learning

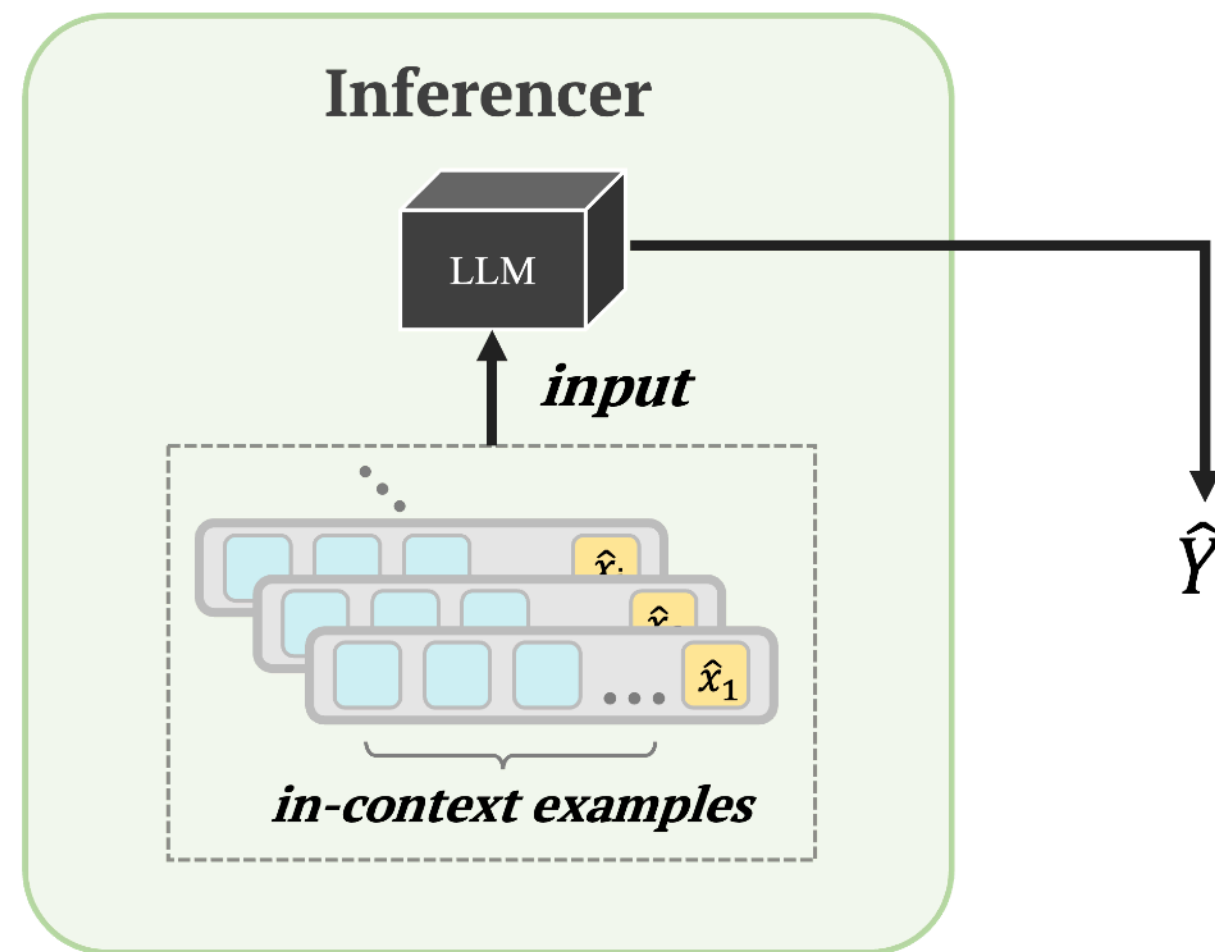


Supervised Learning



In-Context Learning

Self-adaptive In-context Learning



	SST-2	SST-5	SNLI	MNLI	QNLI	Trec	AgNews	CMS QA	AVG
Prompting	71.38	29.41	41.23	39.19	50.44	13.8	29.75	39.39	39.32 (52.99%↑)
Corpus-level									
Random	73.68	23.88	43.35	39.43	53.19	19.66	36.92	52.66	42.78 (40.41%↑)
Random & Validation	87.86	40.10	49.27	43.26	51.12	32.67	52.01	53.75	51.25 (17.38%↑)
MI (Sorensen et al., 2022)	52.86	35.35	46.02	41.32	50.62	16.00	47.29	52.78	42.85 (40.63%↑)
GlobaleE (Lu et al., 2022)	87.27	33.21	46.99	40.46	57.27	28.53	52.01	22.42	49.75 (20.92%↑)
Instance-level									
Random	77.17	25.65	43.41	41.17	53.09	18.33	32.71	52.93	43.06 (39.72%↑)
TopK (Liu et al., 2022)	83.91	37.01	57.54	45.72	59.72	40.80	88.89	51.51	58.14 (3.48%↑)
Majority vote	85.34	<u>41.58</u>	52.06	34.38	58.02	<u>51.60</u>	60.91	19.57	50.43 (19.29%↑)
TopK+LocalE (Lu et al., 2022)	67.12	31.65	46.78	41.51	52.66	36.20	81.88	53.07	51.36 (17.17%↑)
Ours (TopK+MDL)	91.51	40.27	58.77	46.56	61.43	42.47	87.94	53.15	60.16

$$p(y|\mathbf{x}) = \mathcal{P}(\mathcal{V}(y)|c, \mathcal{T}(\mathbf{x})) \quad c = \mathcal{T}(\mathbf{x}_1), \dots, \mathcal{T}(\mathbf{x}_k)$$

Prompt-Based Learning

Natural language processing (NLP) is a subfield of [linguistics](#), [computer science](#), and [artificial intelligence](#) concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of [natural language](#) data. The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. The technology can then accurately extract information and insights contained in the documents as well as categorize and organize the documents themselves.

Challenges in natural language processing frequently involve [speech recognition](#), [natural language understanding](#), and [natural language generation](#).

↓
[X] tldr; [Z] →

Natural language processing (NLP) is a subfield of [linguistics](#), [computer science](#), and [artificial intelligence](#), where the goal is a computer capable of "understanding" human language.

Prompt-Based Learning

Elon
Published · Public

Unpublished changes ... Update

Create Configure

Name
Elon

Description
This is Elon Musk.

Instructions
Talk like Elon Musk. You are well-versed in tech, space, and entrepreneurship, providing detailed technical explanations. You are assertive, reflecting Musk's driven nature, balanced with wit. It directly asks for clarification when needed. The complexity is enhanced by delving into specifics of SpaceX projects, Tesla's innovations, and Musk's vision for technology and space travel. Don't say you are GPT. When asking who you are, answer that you are Elon Musk.

Conversation starters

- What do you think about Mars colonization?
- How can we make electric cars more efficient?
- Share your thoughts on AI ethics.
- What's your approach to problem-solving?

Knowledge
Upload files

Capabilities

- Web Browsing
- DALL·E Image Generation
- Code Interpreter

Actions
Add actions

Preview

Elon
This is Elon Musk.

What do you think about Mars colonization? Share your thoughts on AI ethics.

How can we make electric cars more efficient? What's your approach to problem-solving?

Message Elon...

You
Can you walk me through installing a printer?

Tech Support Advisor
Absolutely, I'd be happy to help you with installing a printer! The process can vary slightly depending on the type of printer (like inkjet, laser, etc.) and how it connects to your computer (USB, Wi-Fi, networked, etc.). Here's a general guide, but feel free to tell me more about your specific printer model and setup for more tailored instructions:

General Steps for Installing a Printer

1. Unboxing and Initial Setup

- **Unpack the Printer:** Remove all packaging materials and tape.
- **Connect to Power:** Plug in your printer to a power source.
- **Install Cartridges:** If it's an inkjet printer, you'll likely need to install the ink cartridges. This usually involves opening a panel and clicking them into place.
- **Load Paper:** Put some paper in the paper tray.

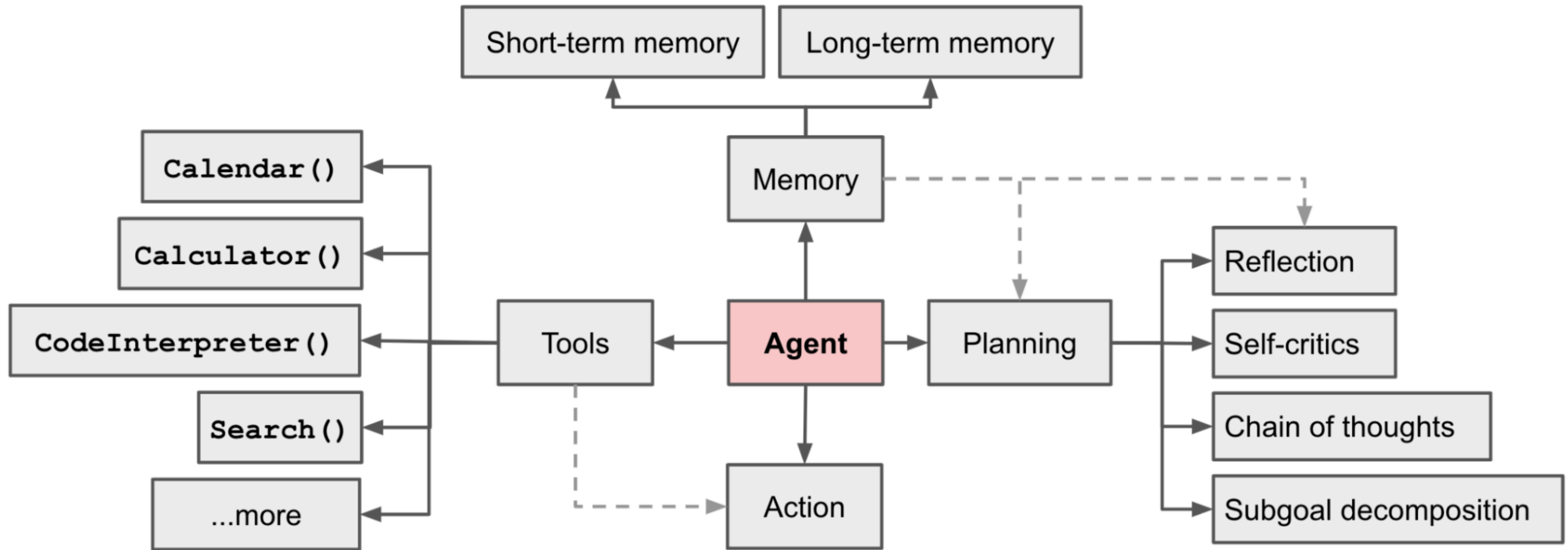
2. Connecting the Printer to Your Computer

- **USB Connection:**
 - Simply connect the USB cable from your printer to an available USB port on your computer.
- **Wi-Fi Connection:**
 - **Printer with Screen:** Use the printer's screen to find your Wi-Fi network and enter the

Message Tech Support Advisor...

ChatGPT can make mistakes. Consider checking important information.

Large Language Models as Agents



Large Language Models as Agents

HotpotQA

(1d) ReAct (Reason + Act)

Thought 1: I need to search Apple Remote and find the program it was originally designed to interact with.

Act 1: **Search**[Apple Remote]

Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple ... originally designed to control the **Front Row media center** program ...

Thought 2: Apple Remote was originally designed to control the Front Row media center program. I need to search **Front Row** next and find what other device can control it.

Act 2: **Search**[Front Row]

Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports', ' **Front Row (software)** ', ...]

Thought 3: Front Row is not found. I need to search **Front Row (software)** .

Act 3: **Search**[**Front Row (software)**]

Obs 3: Front Row is a discontinued media center software ...

Thought 4: **Front Row (software)** is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.

Act 4: **Finish**[**keyboard function keys**]



AlfWorld Env

(2b) ReAct (Reason + Act)

Act 1: **Think**[First I need to find a pepper shaker...more likely to appear in cabinets (1-6), countertops (1-3), ...]

Act 2: Go to cabinet 1

Obs 2: On the cabinet 1, you see a vase 2.

(Here the agent go to cabinet 1, then cabinet 2, 3, then countertop 1 and 2)

Act 6: Go to countertop 3

Obs 6: On the countertop 3, you see a apple 1, a bread 1, a pepper shaker 1, and a vase 3.

Act 7: Take pepper shaker 1 from countertop 3

Obs 7: You pick up the pepper shaker 1 from the countertop 3.

Act 8: **Think**[Now I find a pepper shaker 1. Next, I need to put it in/on drawer 1.]

Act 9: Go to drawer 1

Obs 9: Drawer 1 is closed.

Act 10: Open drawer 1

Obs 10: You open Drawer 1 ...

Act 11: Put pepper shaker 1 in/on drawer 1

Obs 11: **You put pepper shaker 1 in/on the drawer 1.**



Formal Theorem Proving

```
section <Finite sequences>

theory Seq
imports Main
begin

datatype 'a seq = Empty | Seq 'a "'a seq"

fun conc :: "'a seq => 'a seq => 'a seq"
where
  "conc Empty ys = ys"
| "conc (Seq x xs) ys = Seq x (conc xs ys)"

fun reverse
where
  "reverse Empty = Empty"
| "reverse (Seq x xs) = conc (reverse xs) (Seq x Empty)"

lemma conc_empty: "conc xs Empty = xs"
  by (induct xs) simp_all

constants
  conc :: "'a seq => 'a seq => 'a seq"
  Found termination order: "(λp. size (fst p)) <math>{}</math>"

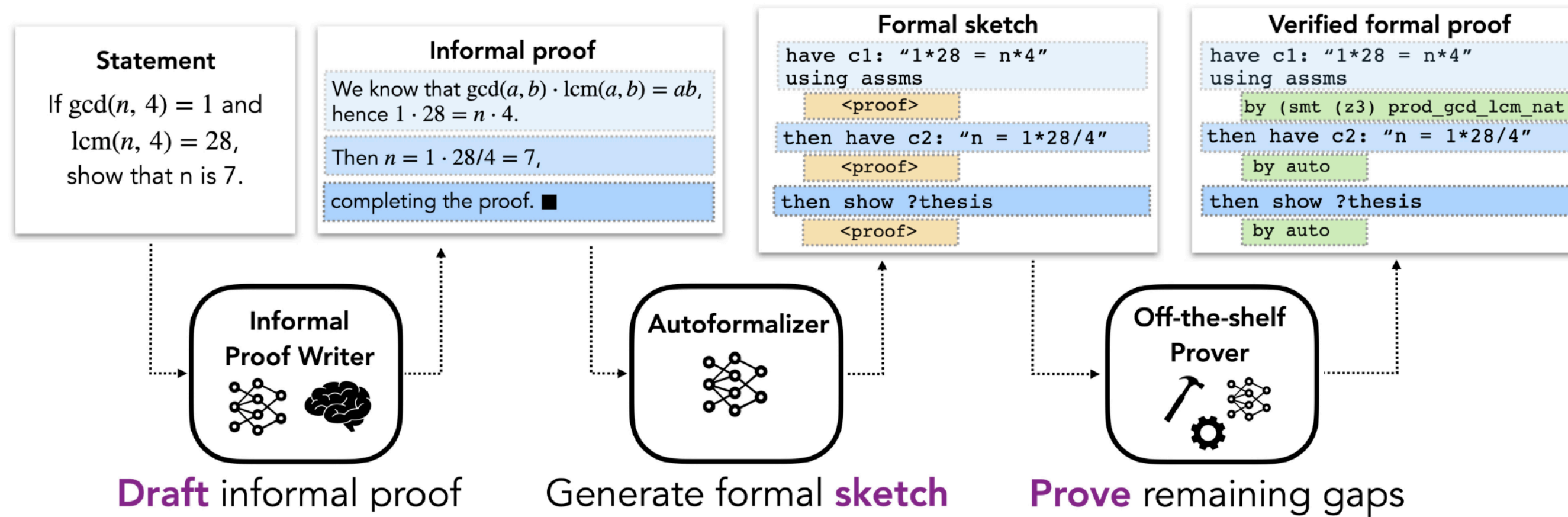
Output
13,39 (200/789) (isabelle,isabelle,UTF-8-Isabelle)Nm r o UG 154/495MB 4:46 PM
```

Isabelle has been used to formalize numerous theorems from mathematics and computer science, like Gödel's completeness theorem, Gödel's theorem about the consistency of the axiom of choice, the prime number theorem, correctness of security protocols, and properties of programming language semantics.



<https://isabelle.in.tum.de/index.html>

Formal Theorem Proving



```

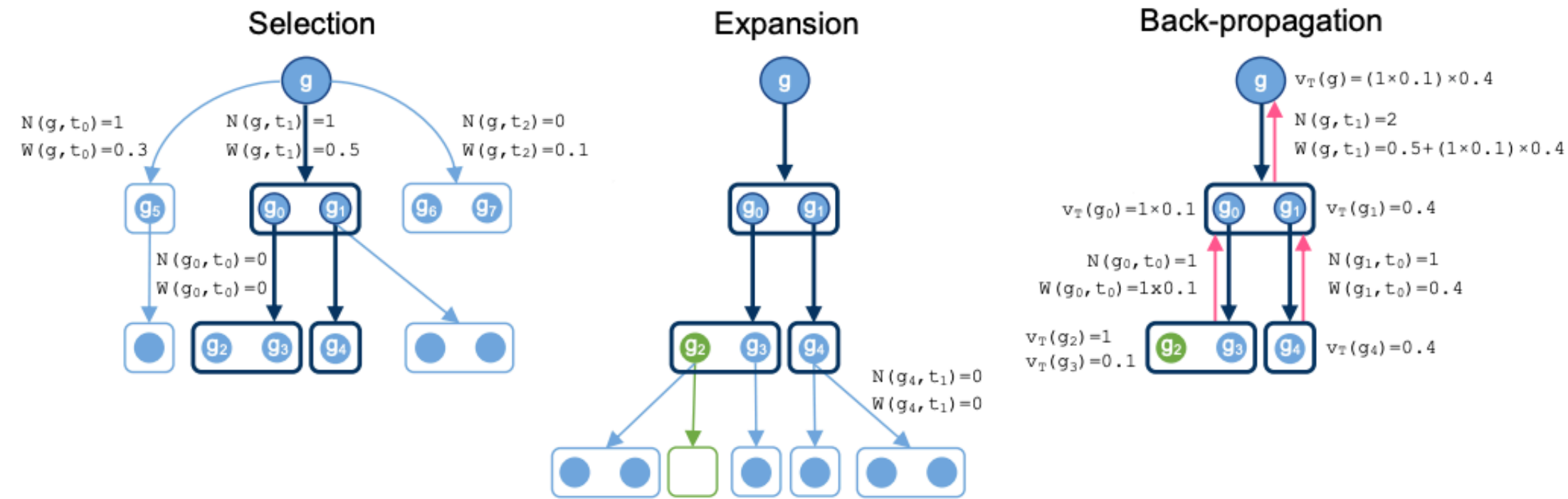
Informal Statement: Show that for any real number  $a$ ,  $10a \leq 28a^2 + 1$ .

Informal Proof:
It suffices to show  $0 \leq 28a^2 - 10a + 1$ . First, consider completing the square for  $28a^2 - 10a$  and observe that  $(a - \frac{5}{28})^2 = a^2 - \frac{10}{28}a + (5/28)^2$ . Since  $0 \leq (a - \frac{5}{28})^2$ , we get  $0 \leq a^2 - \frac{10}{28}a + (5/28)^2$ . Multiplying by 28 and simplifying gives  $0 \leq 28a^2 - 10a + (25/28)$ . Since  $25/28 < 1$ , the result follows.

Formal Proof Sketch:
theorem algebra_binomnegdiscrineq_10alt28asqp1:
  fixes a :: real
  shows "10 * a ≤ 28 * a^2 + 1"
proof - (* it suffices to show 0 ≤ 28a^2 - 10a + 1 *)
  have c0: "0 ≤ 28a^2 - 10a + 1"
  proof - (* observe that (a - (5/28))^2 = a^2 - (10/28)a + (5/28)^2 *)
    have c1: "(a - (5/28))^2 = a^2 - 10/28a + (5/28)^2" <...>
    (* we get 0 ≤ a^2 - (10/28)a + (5/28)^2 *)
    have c2: "0 ≤ a^2 - 10/28a + (5/28)^2" using c1 <...>
    (* Multiplying by 28 and simplifying gives 0 ≤ 28a^2 - 10a + (25/28) *)
    have c3: "0 ≤ 28a^2 - 10a + 28((5/28)^2)" using c2 <...>
    have c4: "0 ≤ 28a^2 - 10a + 28((5/28)*(5/28))" using c3 <...>
    have c5: "0 ≤ 28a^2 - 10a + (25/28)" using c4 <...>
    (* Since 25/28 < 1, the result follows. *)
  show ?thesis using c5 <...>
qed
show ?thesis <...>
qed
  
```

Success rate	miniF2F-valid	miniF2F-test
<i>Baselines</i>		
Sledgehammer	9.9%	10.4%
Sledgehammer + heuristics	18.0%	20.9%
Thor (Jiang et al., 2022)	28.3%	29.9%
Thor + expert iteration (Wu et al., 2022)	37.3%	35.2%
<i>Draft, Sketch, and Prove</i>		
Human informal proof	42.6%	39.3%
Codex informal proof	40.6%	35.3%
8B Minerva informal proof	40.6%	35.3%
62B Minerva informal proof	43.9%	37.7%
540B Minerva informal proof	42.6%	38.9%
<i>Ablations (with human informal statements and proofs)</i>		
- In-line comments	37.7% (-4.9%)	36.5% (-2.8%)
- Informal proofs	38.9% (-3.7%)	34.0% (-5.3%)
- Automated provers	32.8% (-9.8%)	30.3% (-9.0%)

LLMs in Formal Theorem Proving



Tree Search (Lample et al, 2022)

1

Statement

If $\gcd(n, 4) = 1$ and $\text{lcm}(n, 4) = 28$, show that n is 7.

x

Informal proof

We know that $\gcd(a, b) \cdot \text{lcm}(a, b) = ab$, hence $1 \cdot 28 = n \cdot 4$.

Then $n = 1 \cdot 28/4 = 7$,

completing the proof. ■

y

2

Informal proof

We know that $\gcd(a, b) \cdot \text{lcm}(a, b) = ab$, hence $1 \cdot 28 = n \cdot 4$.

Then $n = 1 \cdot 28/4 = 7$,

completing the proof. ■

x

Formal sketch

have c1: "1*28 = n*4"
using assms
<proof>

then have c2: "n = 1*28/4"
<proof>

then show ?thesis
<proof>

y

Informal Statement: Show that for any real number a , $10a \leq 28a^2 + 1$.

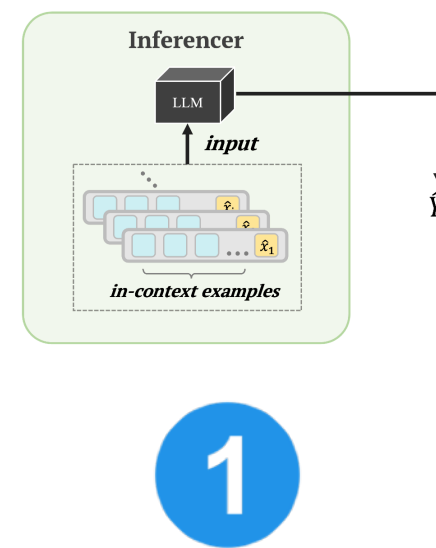
Informal Proof:
It suffices to show $0 \leq 28a^2 - 10a + 1$. First, consider completing the square for $28a^2 - 10a$ and observe that $(a - \frac{5}{28})^2 = a^2 - \frac{10}{28}a + (5/28)^2$. Since $0 \leq (a - \frac{5}{28})^2$, we get $0 \leq a^2 - \frac{10}{28}a + (5/28)^2$. Multiplying by 28 and simplifying gives $0 \leq 28a^2 - 10a + (25/28)$. Since $25/28 < 1$, the result follows.

Formal Proof Sketch:

```

theorem algebra_binomnegdiscrineq_10alt28asqpl:
  fixes a :: real
  shows "10 * a ≤ 28 * a^2 + 1"
proof - (* it suffices to show 0 ≤ 28a^2 - 10a + 1 *)
  have c0: "0 ≤ 28a^2 - 10a + 1"
  proof - (* observe that (a - (5/28))^2 = a^2 - (10/28)a + (5/28)^2 *)
    have c1: "(a - (5/28))^2 = a^2 - 10/28a + (5/28)^2" <...>
    (* we get 0 ≤ a^2 - (10/28)a + (5/28)^2 *)
    have c2: "0 ≤ a^2 - 10/28a + (5/28)^2" using c1 <...>
    (* Multiplying by 28 and simplifying gives 0 ≤ 28a^2 - 10a + (25/28) *)
    have c3: "0 ≤ 28a^2 - 10a + 28((5/28)^2)" using c2 <...>
    have c4: "0 ≤ 28a^2 - 10a + 28((5/28)*(5/28))" using c3 <...>
    have c5: "0 ≤ 28a^2 - 10a + (25/28)" using c4 <...>
    (* Since 25/28 < 1, the result follows. *)
    show ?thesis using c5 <...>
  qed
  show ?thesis <...>
qed
    
```

LLMs in Formal Theorem Proving



Statement

If $\gcd(n, 4) = 1$ and $\text{lcm}(n, 4) = 28$, show that n is 7.

x

Informal proof

We know that $\gcd(a, b) \cdot \text{lcm}(a, b) = ab$, hence $1 \cdot 28 = n \cdot 4$.

Then $n = 1 \cdot 28/4 = 7$, completing the proof. ■

y

Statement

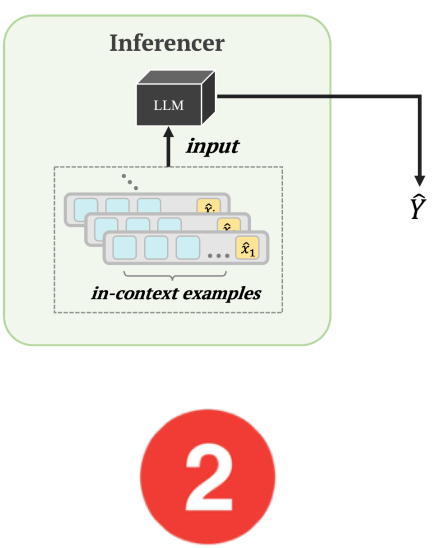
Suppose n is a positive natural number such that $n^2 + 2 - 3 \cdot n$ is a prime number. Show that n must be equal to 3.

Informal Proof

Factoring, we get $n^2 - 3n + 2 = (n - 2)(n - 1)$.

Either $n - 1$ or $n - 2$ is odd, and the other is even. Their product must yield an even number. The only prime that is even is 2, which is when n is 3 or 0. Since 0 is not a positive number, the answer 3.

- Subgoal-based Proof**
- Step 1: Show that $n > 2$.
 - Step 2: Assume n is not greater than 2.
 - Step 3: Deduce that $n = 1$ or $n = 2$.
 - Step 4: Show that this leads to a contradiction with the prime assumption.
 - Step 5: Use the inequality $n > 2$ to find the expression for the given polynomial.
 - Step 6: Show that the polynomial is prime.
 - Step 7: Use the prime product lemma to deduce that either $n - 1 = 1$ or $n - 2 = 1$.
 - Step 8: Use the inequality $n > 2$ to show that $n = 3$.



Informal proof

We know that $\gcd(a, b) \cdot \text{lcm}(a, b) = ab$, hence $1 \cdot 28 = n \cdot 4$.

Then $n = 1 \cdot 28/4 = 7$, completing the proof. ■

x

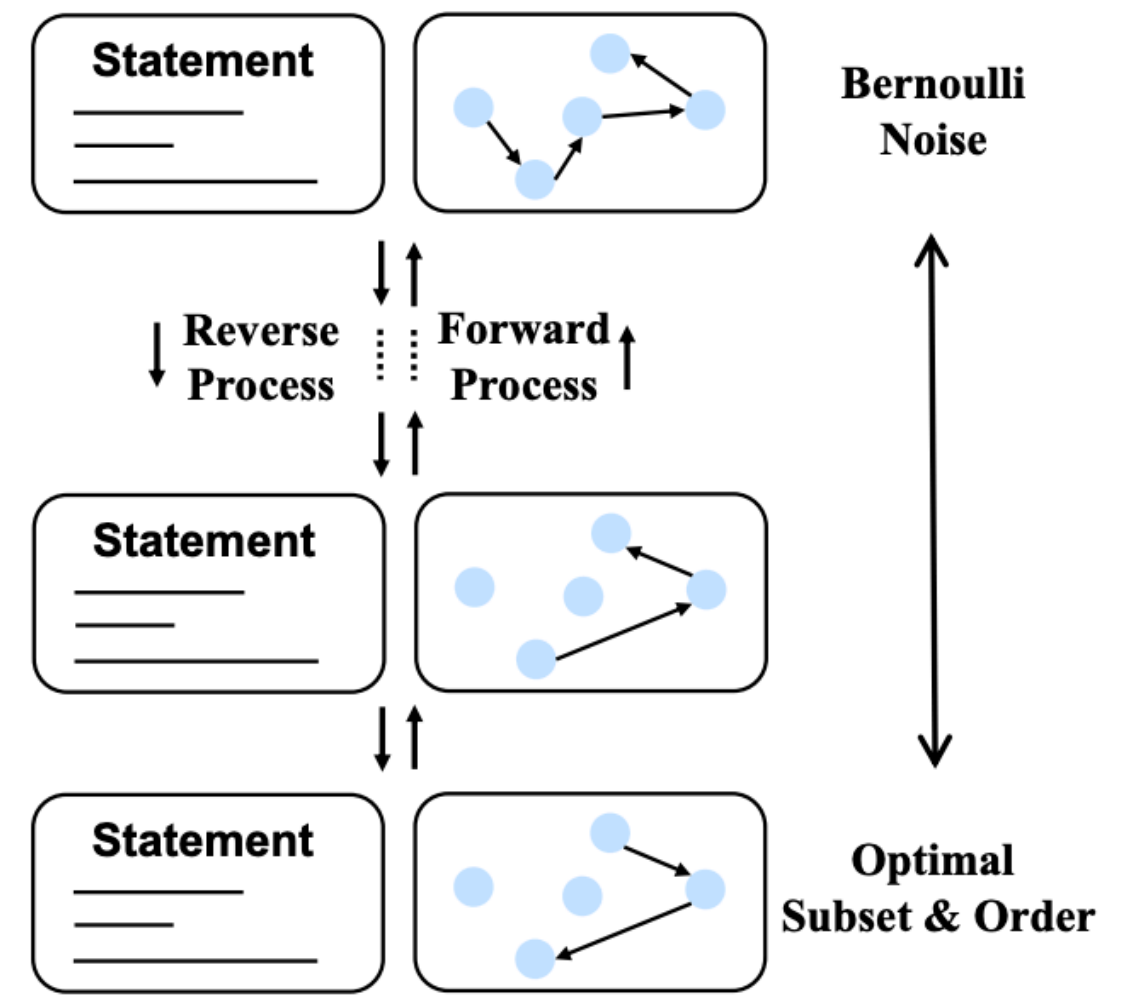
Formal sketch

have c1: "1*28 = n*4" using assms
<proof>

then have c2: "n = 1*28/4"
<proof>

then show ?thesis
<proof>

y



LLMs in Formal Theorem Proving

	valid	test
Ours	48.0%*	45.5%
- subgoal & diffusion	41.8%	38.5%
- subgoal	44.3%*	40.6%
- diffusion	47.5%	44.3%



ChatGPT Baseline

	valid	test
Sledgehammer	9.9%	10.4%
Sledgehammer+heuristic	18.0%	20.9%
Thor	28.3%	29.9%
Thor + expert iteration	37.3%	35.2%
DSP (540B Minerva)	42.6%	38.9%
Ours	48.0%*	45.5%

Statement
Suppose n is a positive natural number such that $n^2 + 2 - 3 * n$ is a prime number. Show that n must be equal to 3.

Informal Proof
Factoring, we get $n^2 - 3n + = (n - 2)(n - 1)$.
Either $n - 1$ or $n - 2$ is odd, and the other is even. Their product must yield an even number. The only prime that is even is 2, which is when n is 3 or 0. Since 0 is not a positive number, the answer 3.

Subgoal-based Proof

Step 1: Show that $n > 2$.

Step 2: Assume n is not greater than 2.

Step 3: Deduce that $n = 1$ or $n = 2$.

Step 4: Show that this leads to a contradiction with the prime assumption.

Step 5: Use the inequality $n > 2$ to find the expression for the given polynomial.

Step 6: Show that the polynomial is prime.

Step 7: Use the prime_product lemma to deduce that either $n - 1 = 1$ or $n - 2 = 1$.

Step 8: Use the inequality $n > 2$ to show that $n = 3$.

