Question Answering and Machine Translation

CSE538 - Spring 2024

Guest Lecturer: Yash Kumar Lal

NLP, The Course

Overall NLP Concept

I. Syntax

Introduction to NLP; Tokenization; Words Corpora

One-hot, and Multi-hot encoding.
Parts-of-Speech; Named Entities;

Parsing; Verbal Predicates;Dependency Parsing

II. Semantics

Dependency Parsing; Word Sense Disambiguation

Vector Semantics (Embeddings), Word2vec

Probabilistic Language Models
Ngram Classifier, Topic Modeling

Overall NLP Concept

III. Language Modeling

Ethical Considerations

Masked Language Modeling (autoencoding)

Generative Language Modeling (autoregressive)

Applying LMs

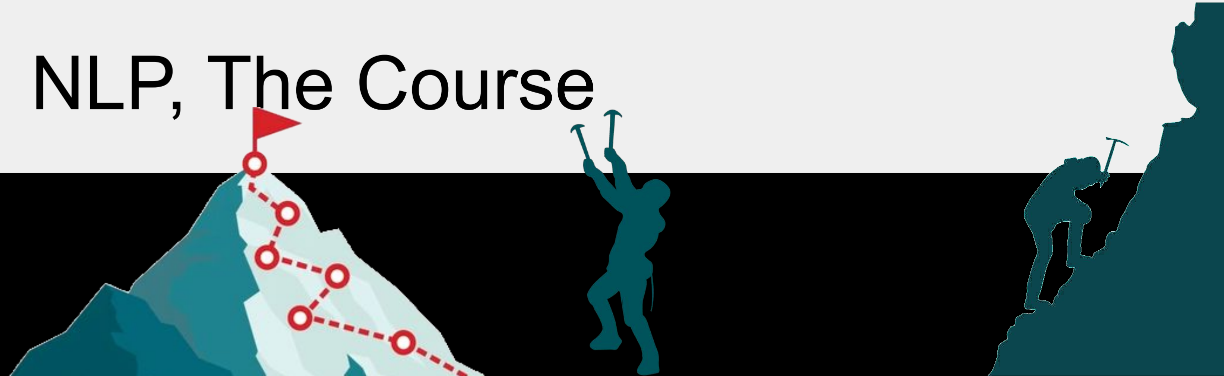
IV. Applications

Language and Psychology
(advanced sentiment)

Speech and Audio Processing, Dialog (chatbots)

Question Answering, Translation

NLP, The Course



Overall NLP Concept

I. Syntax

Introduction to NLP; Tokenization; Words Corpora

One-hot, and Multi-hot encoding.
Parts-of-Speech; Named Entities;

Parsing; Verbal Predicates;Dependency Parsing

II. Semantics

Dependency Parsing; Word Sense Disambiguation

Vector Semantics (Embeddings), Word2vec

Probabilistic Language Models
Ngram Classifier, Topic Modeling

Overall NLP Concept

III. Language Modeling

Ethical Considerations

Masked Language Modeling (autoencoding)

Generative Language Modeling (autoregressive)

Applying LMs

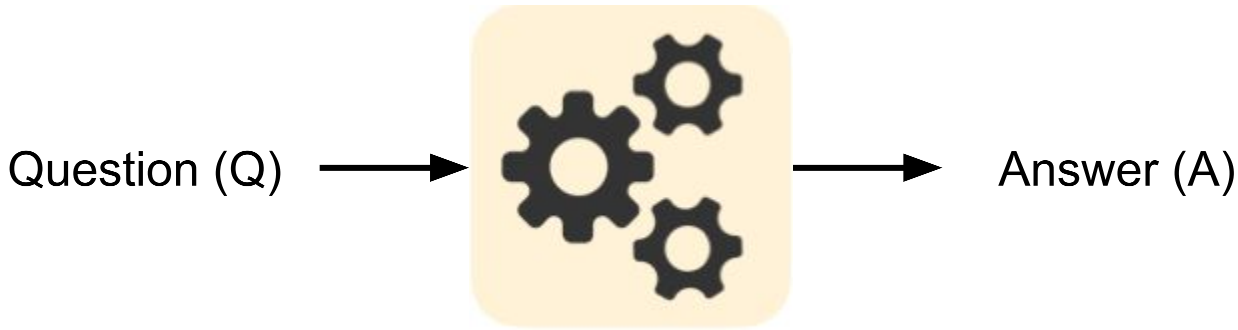
IV. Applications

Language and Psychology
(advanced sentiment)

Speech and Audio Processing, Dialog (chatbots)

Question Answering, Translation

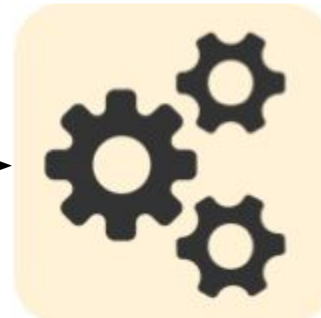
What is Question Answering?



What is Question Answering?

The goal of question answering is to build systems that automatically answer questions posed by humans in a natural language

What is Question Answering?

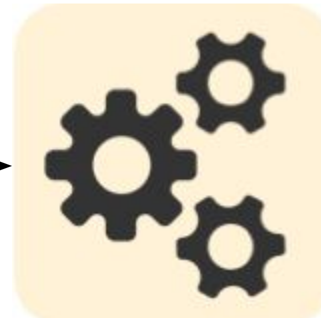
Question (Q) →  → Answer (A)

The goal of question answering is to build systems that automatically answer questions posed by humans in a natural language

Who is the first person to go to Mariana Trench?

The first person to go to the Mariana Trench was the American oceanographer and adventurer Don Walsh, who descended to its deepest point, the Challenger Deep, in 1960.

What is Question Answering?

Question (Q) →  → Answer (A)

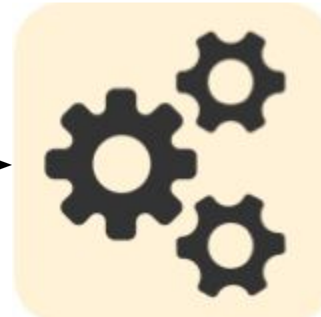
The goal of question answering is to build systems that automatically answer questions posed by humans in a natural language

Who is the first person to go to Mariana Trench?

The first person to go to the Mariana Trench was the American oceanographer and adventurer Don Walsh, who descended to its deepest point, the Challenger Deep, in 1960.

Q: From a user's perspective, are you happy with the answer?

What is Question Answering?

Question (Q) →  → Answer (A)

The goal of question answering is to build systems that automatically answer questions posed by humans in a natural language

GPT-4 visual input example, Extreme Ironing:

User What is unusual about this image?

Source: https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg

GPT-4 The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi.

What is Question Answering?

The goal of question answering is to build systems that automatically answer questions posed by humans in a natural language

GPT-4 visual input example, Extreme Ironing:

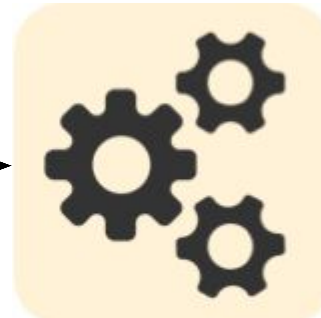
User: What is unusual about this image?

Source: https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg

GPT-4: The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi.

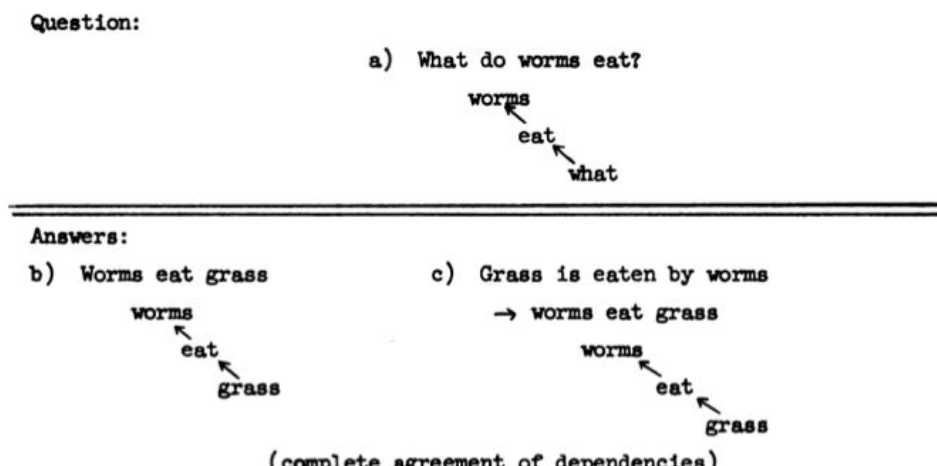
Q: From a user's perspective, are you happy with the answer?

What is Question Answering?

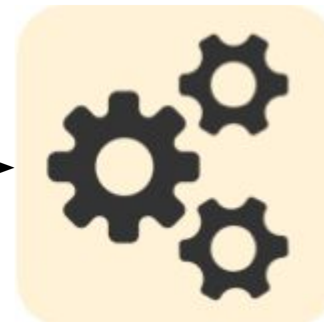
Question (Q) →  → Answer (A)

The goal of question answering is to build systems that automatically answer questions posed by humans in a natural language

The earliest QA systems were built in the 1960s!
(Simmons et al., 1964)



Question Answering: A Taxonomy

Question (Q) →  → Answer (A)

Information	Question	Answer
Documents (corpus)	Factoid vs non-factoid	Single fact
Document	Open vs closed domain	Explanation
Knowledge Base	Simple vs multi-step	Document
Other modalities of data (image, video...)	Narrative	Extracted span
		Image or other object

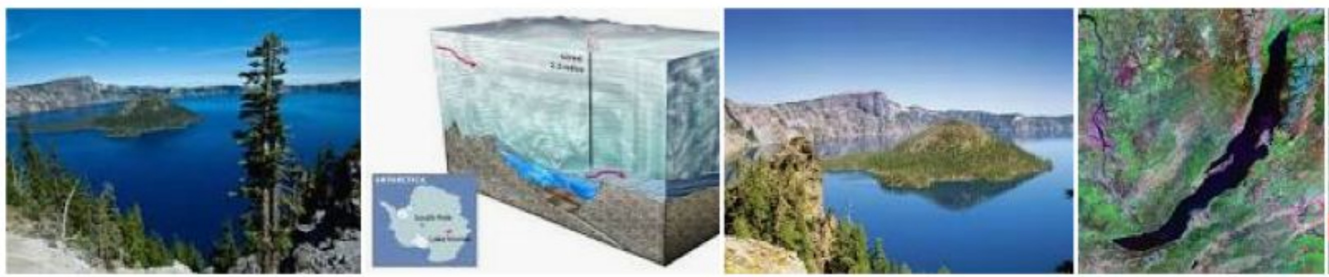
Real-World Applications Everywhere!

Google

Where is the deepest lake in the world?

All Maps Images News Videos More Settings Tools

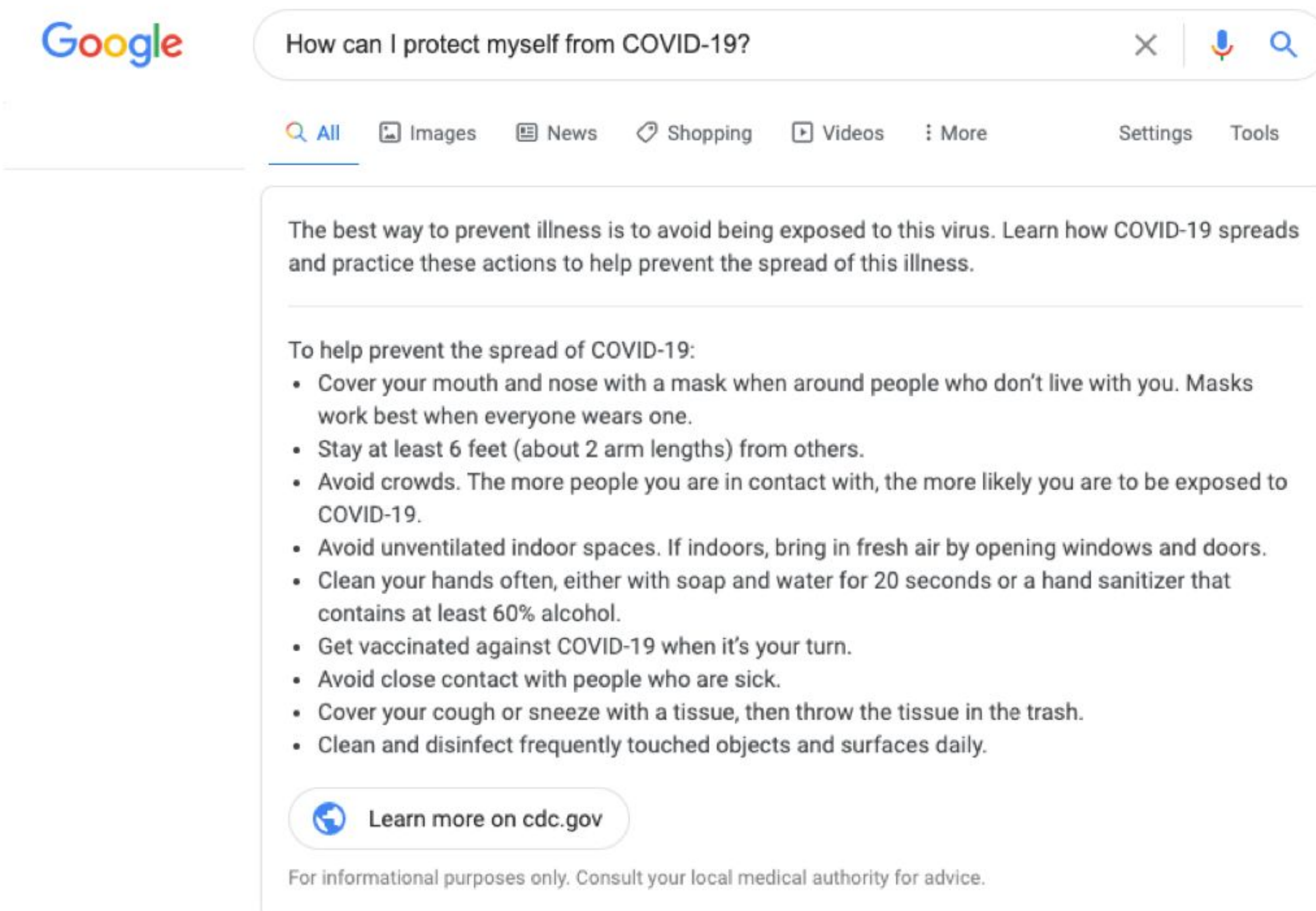
About 21,100,000 results (0.71 seconds)



Siberia

Lake **Baikal**, in Siberia, holds the distinction of being both the deepest lake in the world and the largest freshwater lake, holding more than 20% of the unfrozen fresh water on the surface of Earth.

Real-World Applications Everywhere!



The best way to prevent illness is to avoid being exposed to this virus. Learn how COVID-19 spreads and practice these actions to help prevent the spread of this illness.

To help prevent the spread of COVID-19:

- Cover your mouth and nose with a mask when around people who don't live with you. Masks work best when everyone wears one.
- Stay at least 6 feet (about 2 arm lengths) from others.
- Avoid crowds. The more people you are in contact with, the more likely you are to be exposed to COVID-19.
- Avoid unventilated indoor spaces. If indoors, bring in fresh air by opening windows and doors.
- Clean your hands often, either with soap and water for 20 seconds or a hand sanitizer that contains at least 60% alcohol.
- Get vaccinated against COVID-19 when it's your turn.
- Avoid close contact with people who are sick.
- Cover your cough or sneeze with a tissue, then throw the tissue in the trash.
- Clean and disinfect frequently touched objects and surfaces daily.

Learn more on cdc.gov

For informational purposes only. Consult your local medical authority for advice.

Areas in Question Answering

Reading Comprehension	● Answer based on a document ● Context is a one (or more) document(s)
Open-Domain QA	● Answer based on encyclopedic knowledge ● Context is the Internet (all knowledge)
Visual QA	● Answer is simple and factual ● Context is one/multiple image(s)

Areas in Question Answering

Reading Comprehension	● Answer based on a document ● Context is a one (or more) document(s)
Open-Domain QA	● Answer based on encyclopedic knowledge ● Context is the Internet (all knowledge)
Visual QA	● Answer is simple and factual ● Context is one/multiple image(s)

Reading Comprehension

Comprehend a passage of text and answer questions about its content

Passage (*P*) + Question (*Q*) → Answer (*A*)

Reading Comprehension (MCTest)

Comprehend a passage of text and answer questions about its content

Passage (*P*) + Question (*Q*) → Answer (*A*)

P

Alyssa got to the beach after a long trip. She's from Charlotte. She traveled from Atlanta. She's now in Miami. She went to Miami to visit some friends. But she wanted some time to herself at the beach, so she went there first. After going swimming and laying out, she went to her friend Ellen's house. Ellen greeted Alyssa and they both had some lemonade to drink. Alyssa called her friends Kristin and Rachel to meet at Ellen's house.

Q Why did Alyssa go to Miami?

A To visit some friends

Richardson et al, 2013.

Reading Comprehension (MCTest)

Comprehend a passage of text and answer questions about its content

Passage (*P*) + Question (*Q*) → Answer (*A*)

P

Alyssa got to the beach after a long trip. She's from Charlotte. She traveled from Atlanta. She's now in Miami. She went to Miami to visit some friends. But she wanted some time to herself at the beach, so she went there first. After going swimming and laying out, she went to her friend Ellen's house. Ellen greeted Alyssa and they both had some lemonade to drink. Alyssa called her friends Kristin and Rachel to meet at Ellen's house.

Q Why did Alyssa go to Miami?

A To visit some friends

- ~3k questions from ~1k articles
- Multiple-choice questions
- Need for paraphrase, coreference resolution and dealing with many distractors

Richardson et al, 2013.

Reading Comprehension (SQuAD)

Comprehend a passage of text and answer questions about its content

Passage (*P*) + Question (*Q*) → Answer (*A*)

P

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”.

Q

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

A

Within a cloud

Rajpurkar et al, 2016.

Reading Comprehension (SQuAD)

Comprehend a passage of text and answer questions about its content

Passage (*P*) + Question (*Q*) → Answer (*A*)

P

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”.

Q

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

A

Within a cloud

- 100k annotated (passage, question, answer) triples
- Answer is a short segment of text (or span) in passage
- Questions are crowd-sourced, passages are from English Wikipedia, usually 100-150 words long

Rajpurkar et al, 2016.

Evaluating Reading Comprehension

Passage (P) + Question (Q) → Answer (A)

P

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”.

Q

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

A

Within a cloud

Inside clouds

Clouds

M

Collide inside clouds

- Exact match (EM): 0 or 1
- max{0, 0, 0} = 0

Rajpurkar et al, 2016.

Evaluating Reading Comprehension

Passage (P) + Question (Q) → Answer (A)

P

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”.

Q

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

A

Within a cloud

Inside clouds

Clouds

M

Collide inside clouds

- F1: Partial credit
- max{0.33, 0.67, 0.33} = 0.67

Rajpurkar et al, 2016.

Models for Reading Comprehension

Passage (P) + Question (Q)   Answer (A)

Input: $P = (p_1, p_2, \ldots, p_N)$, $Q = (q_1, q_2, \ldots, q_M)$ N~100, M~15

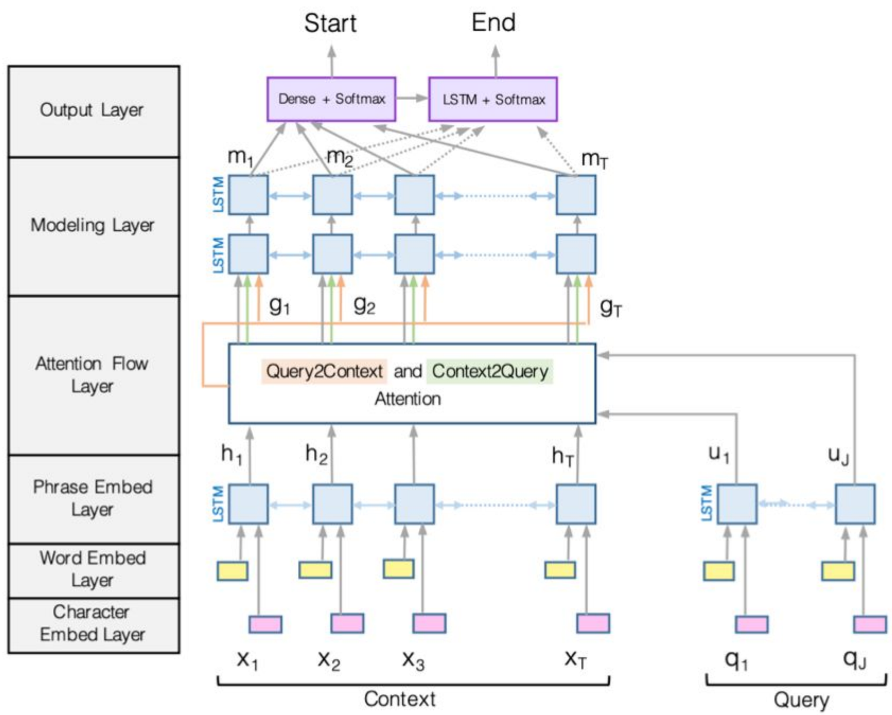
Output: $0 < \text{start} < \text{end} < N+1$ answer is a span in the passage

LSTM-Based Models for Reading Comprehension

Passage (P) + Question (Q) Answer (A)

Input: $P = (p_1, p_2, \ldots, p_N)$, $Q = (q_1, q_2, \ldots, q_M)$ $N \sim 100$, $M \sim 15$

Output: $0 < \text{start} < \text{end} < N+1$ answer is a span in the passage



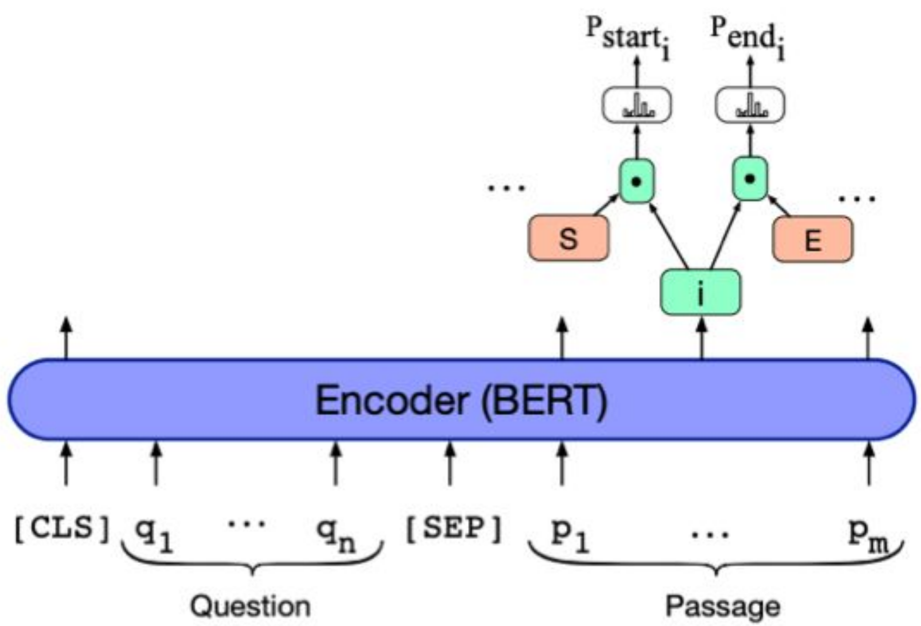
Seo et al, 2017.

BERT-Based Models for Reading Comprehension

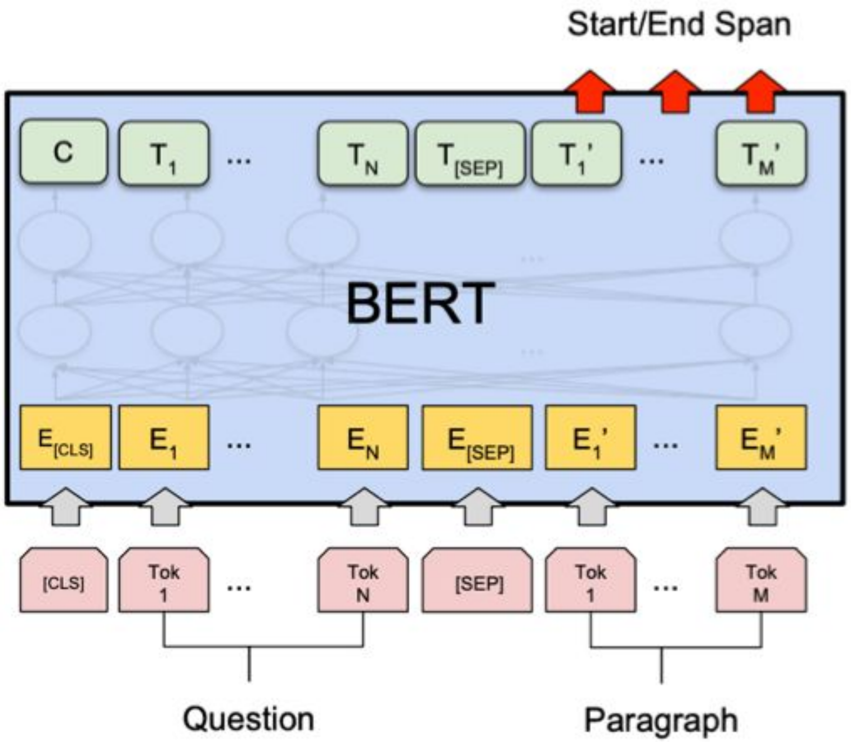
Passage (P) + Question (Q) → Answer (A)

Input: $P = (p_1, p_2, \ldots, p_N)$, $Q = (q_1, q_2, \ldots, q_M)$ — N~100, M~15

Output: $0 < \text{start} < \text{end} < N+1$ — answer is a span in the passage

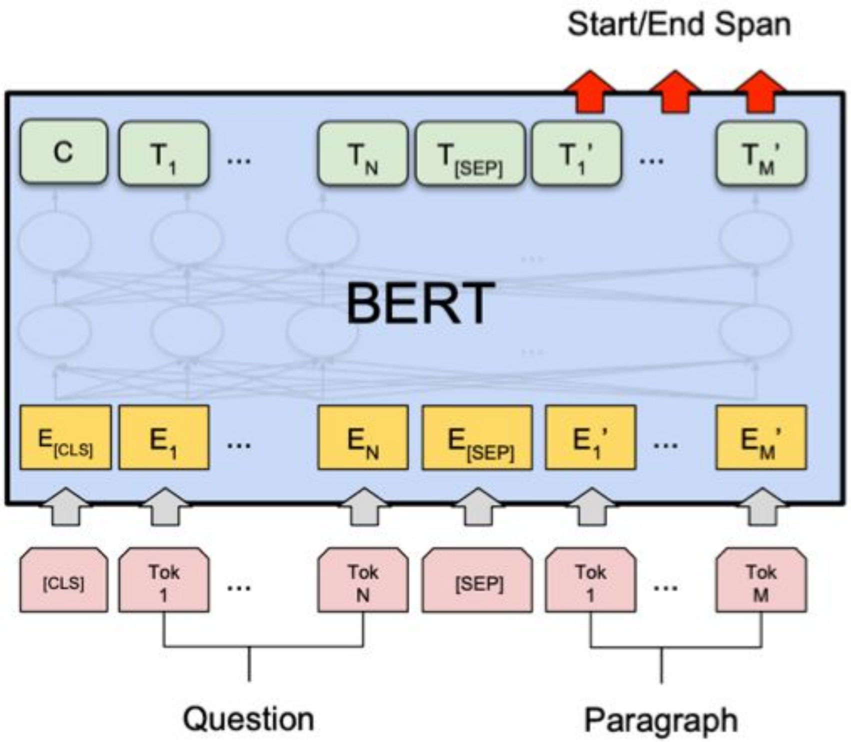


BERT-Based Models for Reading Comprehension

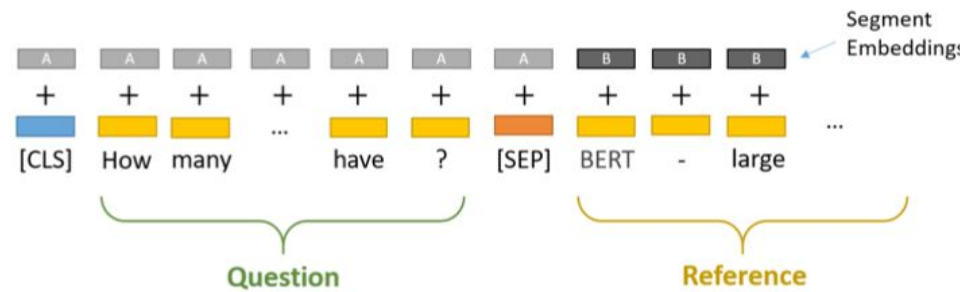


Question = Segment A
Passage = Segment B
Answer = predicting two endpoints in segment B

BERT-Based Models for Reading Comprehension



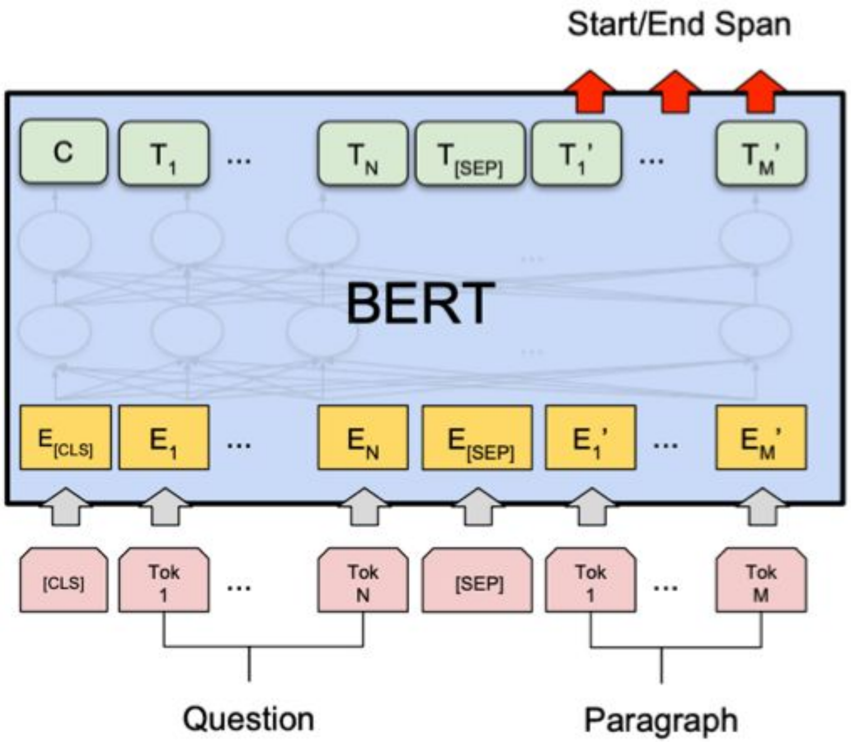
Question = Segment A
Passage = Segment B
Answer = predicting two endpoints in segment B



Question: How many parameters does BERT-large have?

Reference Text: BERT-large is really big... it has 24 layers and an embedding size of 1,024, for a total of 340M parameters! Altogether it is 1.34GB, so expect it to take a couple minutes to download to your Colab instance.

BERT-Based Models for Reading Comprehension



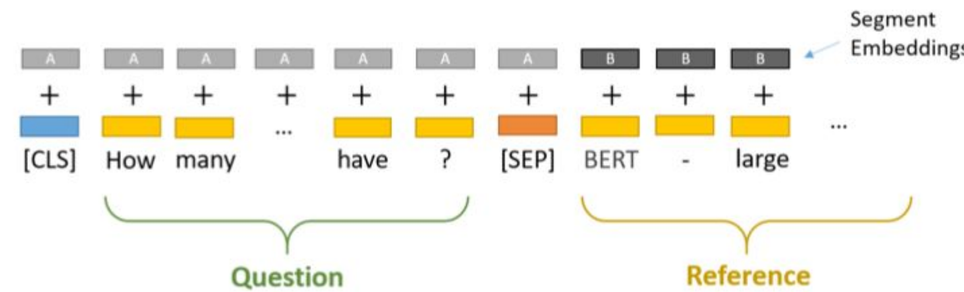
$$\mathcal{L} = -\log p_{\text{start}}(s^*) - \log p_{\text{end}}(e^*)$$

$$p_{\text{start}}(i) = \text{softmax}_i(\mathbf{w}_{\text{start}}^{\intercal}\mathbf{h}_i)$$

$$p_{\text{end}}(i) = \text{softmax}_i(\mathbf{w}_{\text{end}}^{\intercal}\mathbf{h}_i)$$

where $\mathbf{h}_i$ is the hidden vector of c_i, returned by BERT

Question = Segment A
Passage = Segment B
Answer = predicting two endpoints in segment B

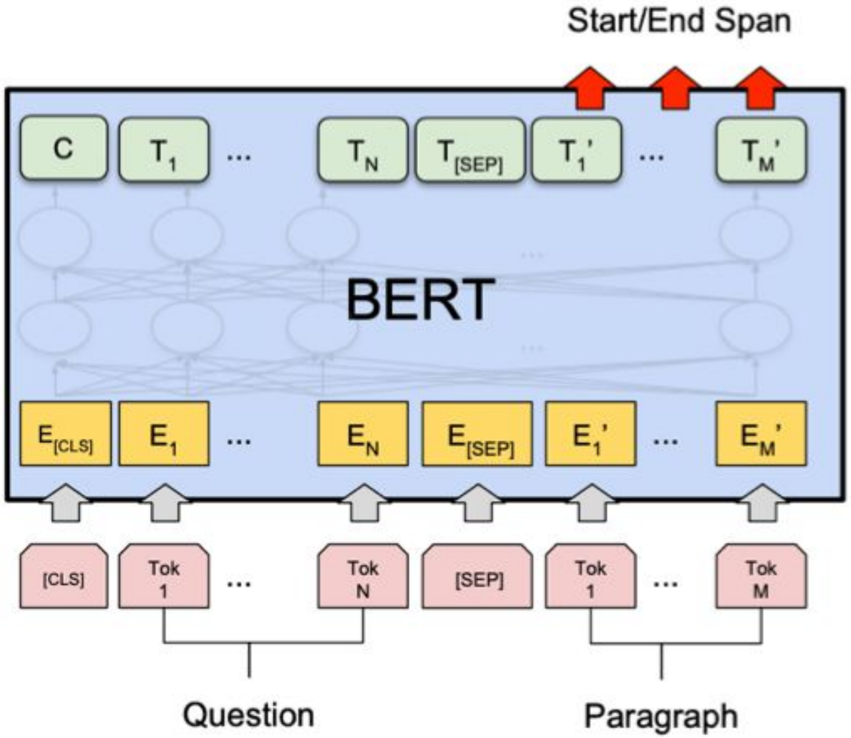


Question: How many parameters does BERT-large have?

Reference Text: BERT-large is really big... it has 24 layers and an embedding size of 1,024, for a total of 340M parameters! Altogether it is 1.34GB, so expect it to take a couple minutes to download to your Colab instance.

mccormickml.com

Is Reading Comprehension Solved?



	F1	EM
Human performance	91.2*	82.3*
BiDAF	77.3	67.7
BERT-base	88.5	80.8
BERT-large	90.9	84.1
XLNet	94.5	89.0
RoBERTa	94.6	88.9
ALBERT	94.8	89.3

dev set, except for human performance

Is Reading Comprehension Solved?

Questions that require long answers

Question: How do Jellyfish function without brains or nervous systems? [...] (60 words)

Answer: Jellyfish may not have a brain, but they have a rough nervous system and innate behaviours. However, they are very simple creatures. They're invertebrate: creatures without a backbone. Most jellyfish have really short life spans. Sometimes just a couple of hours. [...] As their name implies, they are largely composed of basically jelly inside a thin membrane. They're over 95% water. (327 words)

Documents: [...] Jellyfish do not have brains, and most barely have nervous systems. They have primitive nerve cells that help them orient themselves in the water and sense light and touch. [...] While they dont possess brains, the animals still have neurons that send all sorts of signals throughout their body. [...] They may accomplish this through the assistance of their nerve rings. Jellyfish don't have brains, and that's just where things begin. They don't have many of the body parts that are typical in other animals. [...] (1070 words)

ELI5 (Fan et al., 2019)

Is Reading Comprehension Solved?

Questions that require long answers

Question: How do Jellyfish function without brains or nervous systems? [...] (60 words)

Answer: Jellyfish may not have a brain, but they have a rough nervous system and innate behaviours. However, they are very simple creatures. They're invertebrate: creatures without a backbone. Most jellyfish have really short life spans. Sometimes just a couple of hours. [...] As their name implies, they are largely composed of basically jelly inside a thin membrane. They're over 95% water. (327 words)

Documents: [...] Jellyfish do not have brains, and most barely have nervous systems. They have primitive nerve cells that help them orient themselves in the water and sense light and touch. [...] While they dont possess brains, the animals still have neurons that send all sorts of signals throughout their body. [...] They may accomplish this through the assistance of their nerve rings. Jellyfish don't have brains, and that's just where things begin. They don't have many of the body parts that are typical in other animals. [...] (1070 words)

ELI5 (Fan et al., 2019)

Questions that require discrete reasoning

Q: Where did Charles travel to first, Castile or Barcelona?

In 1517, the seventeen-year-old King sailed to Castile, where he was formally recognised as King of Castile. There, his Flemish court provoked much scandal, … In May 1518, Charles traveled to Barcelona in Aragon, where he would remain for nearly two years.

DROP (Dua et al., 2019)

Is Reading Comprehension Solved?

Setting	CoQA	DROP	QuAC	SQuADv2	RACE-h	RACE-m
Fine-tuned SOTA	**90.7**[a]	**89.1**[b]	**74.4**[c]	**93.0**[d]	**90.0**[e]	**93.1**[e]
GPT-3 Zero-Shot	81.5	23.6	41.5	59.5	45.5	58.4
GPT-3 One-Shot	84.0	34.3	43.3	65.4	45.9	57.4
GPT-3 Few-Shot	85.0	36.5	44.3	69.8	46.8	58.1

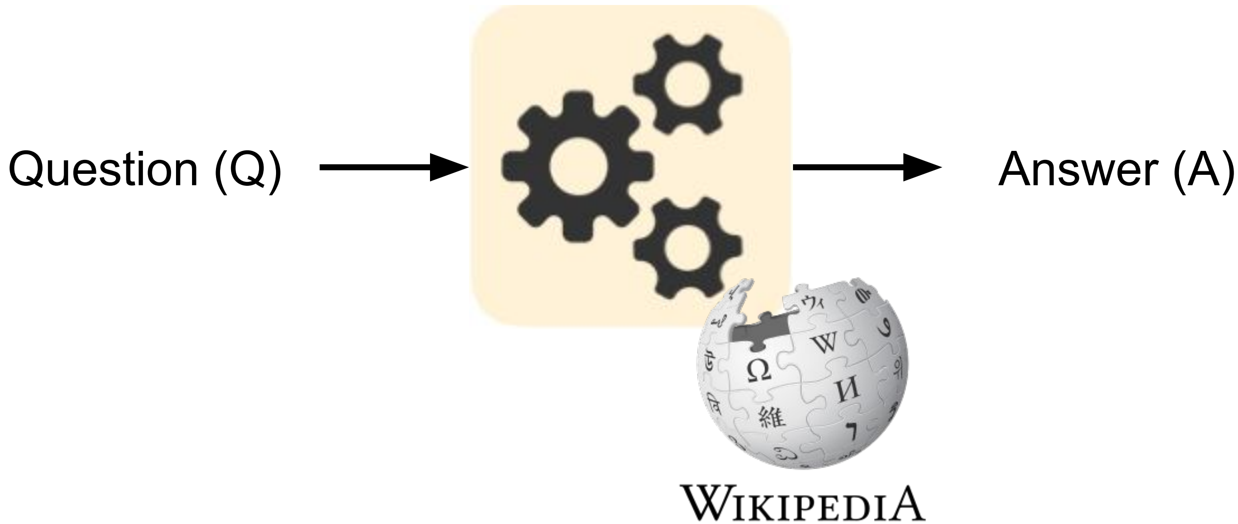
Why few-shot GPT-3 is still not enough?

- Context length is long - cannot use enough examples
- No fine-tuning

Areas in Question Answering

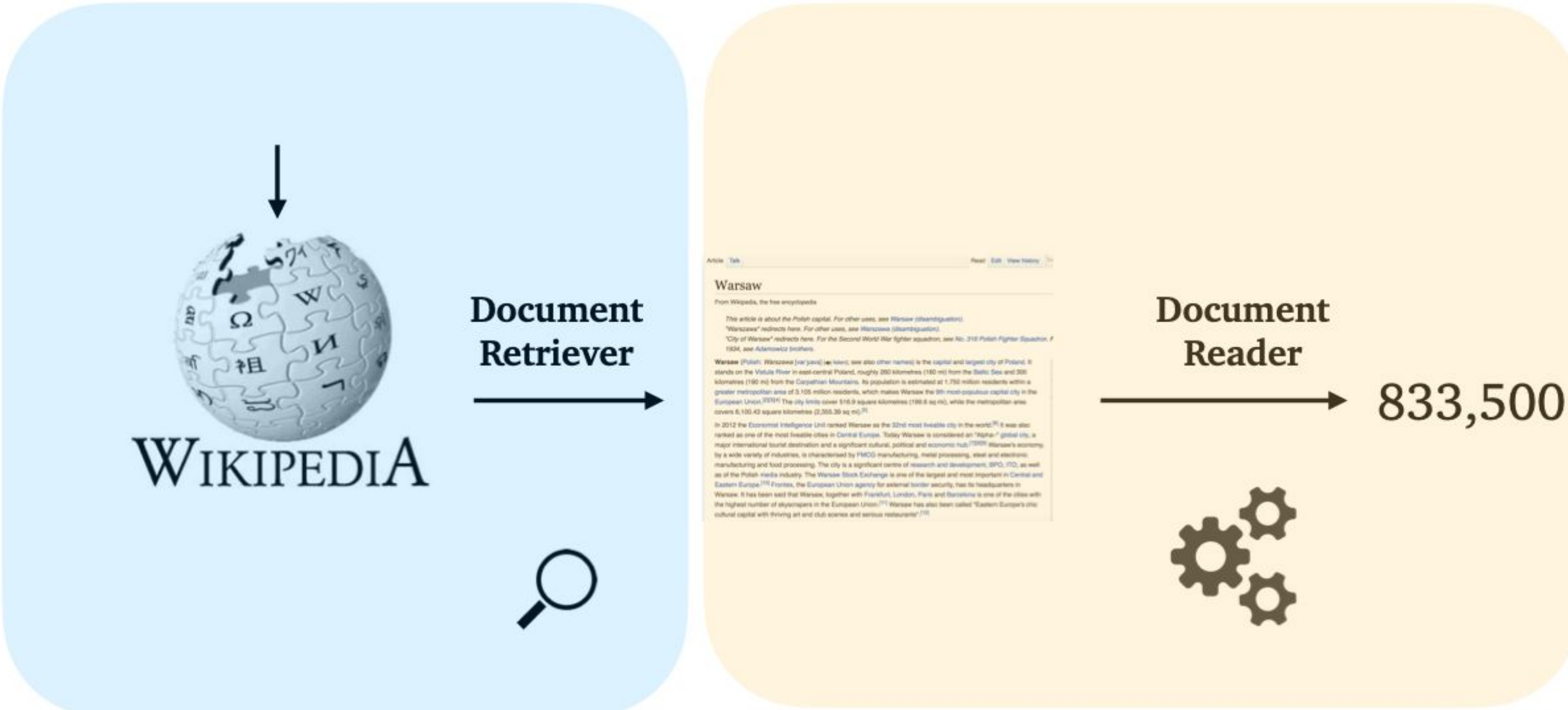
Reading Comprehension	• Answer based on a document • Context is a one (or more) document(s)
Open-Domain QA	• Answer based on encyclopedic knowledge • Context is the Internet (all knowledge)
Visual QA	• Answer is simple and factual • Context is one/multiple image(s)

Open-Domain Question Answering



- No given passage, just a large collection of documents (e.g., Wikipedia)
- No idea where answer is located
- Have to answer any open-domain questions
- Very challenging, but more practical

Open-Domain Question Answering



Chen et al, 2017.

Retriever-Reader Framework

Input: $D = (D_1, D_2, \ldots, D_N)$, Q

D: large collection of documents

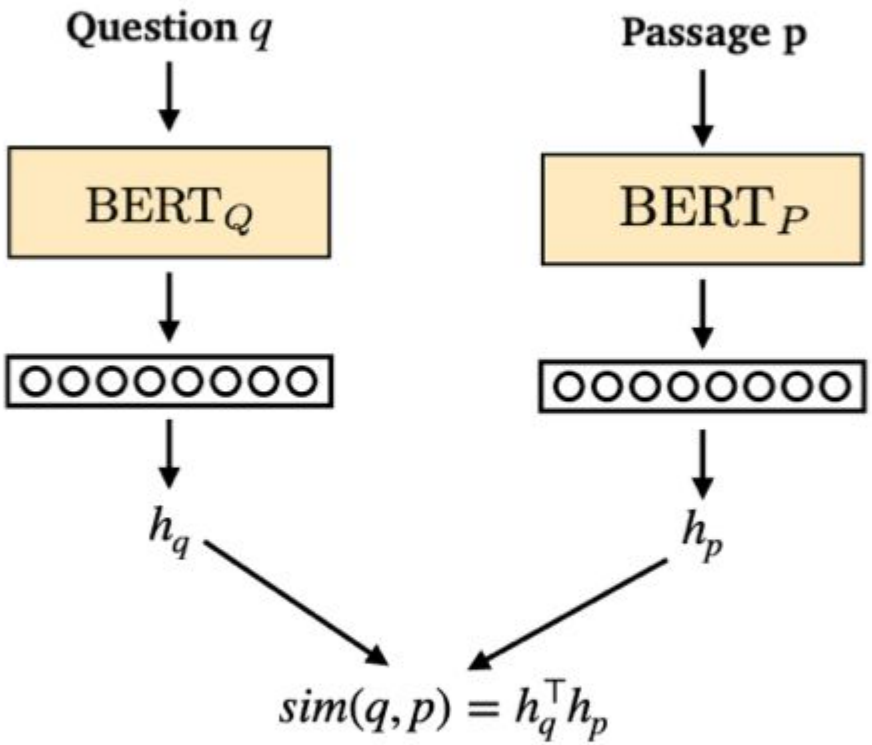
Output: an answer string A

Retriever: $f(D, Q) \rightarrow (P_1, P_2, \ldots, P_K)$

K is pre-defined

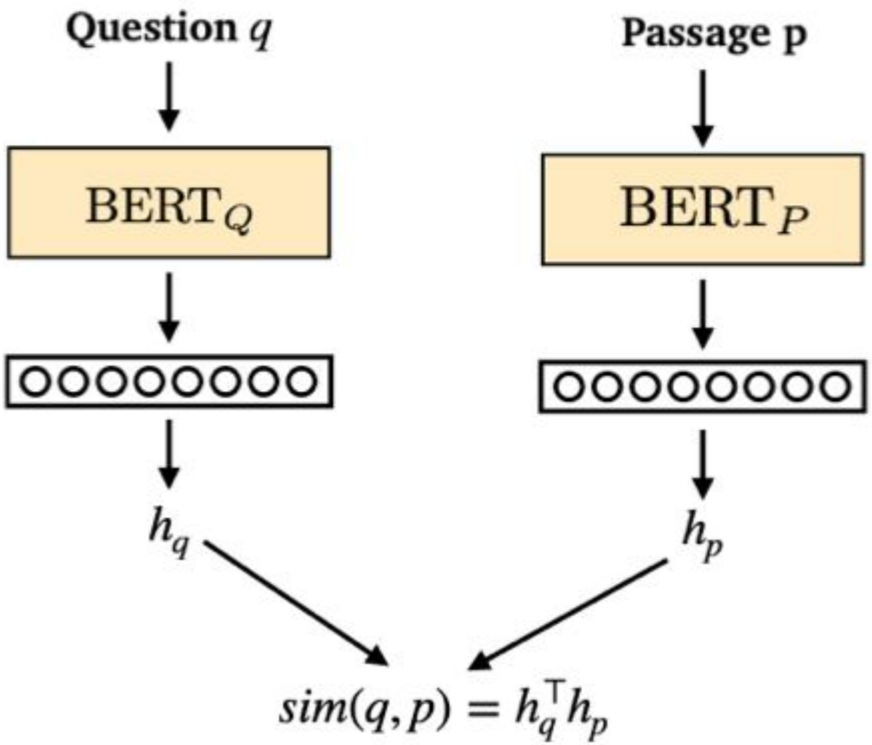
Reader: $g(Q, \{P_1, P_2, \ldots, P_K\}) \rightarrow A$

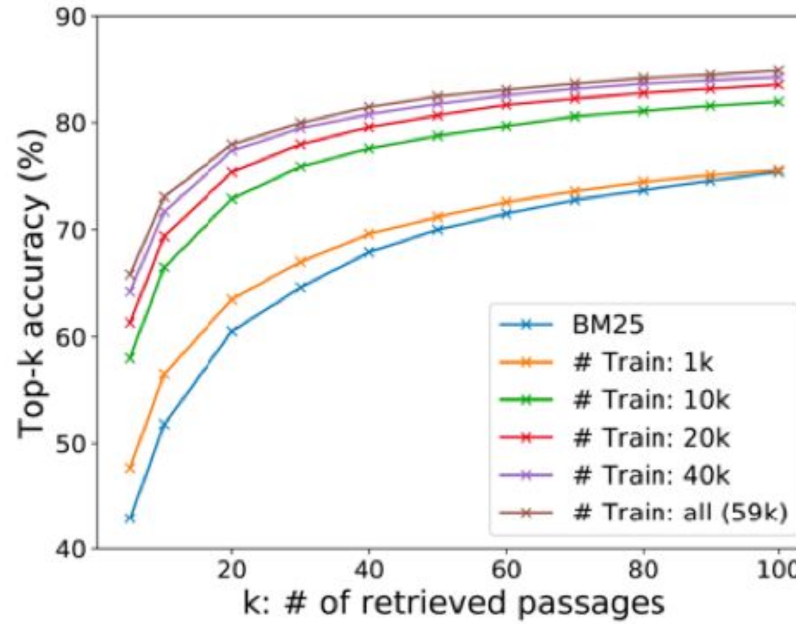
Dense Passage Retrieval



Karpukhin et al, 2020.

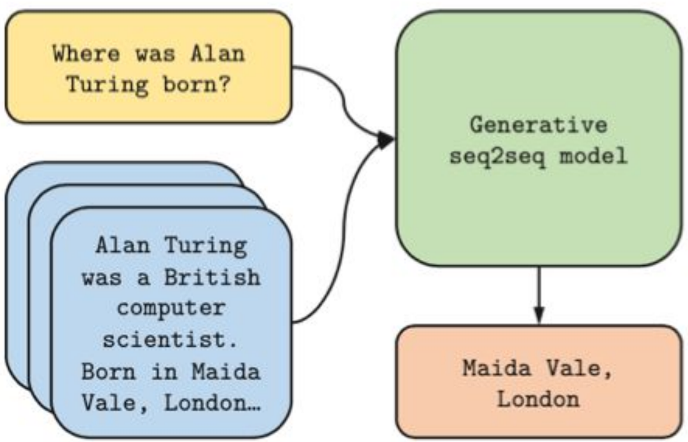
Dense Passage Retrieval





Karpukhin et al, 2020.

Dense Retrieval + Generative Models

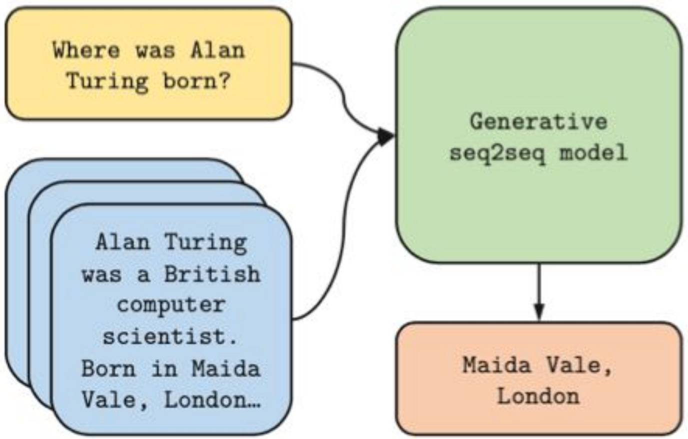


Fusion-in-decoder (FID)

DPR + T5

Izacard and Grave, 2020.

Dense Retrieval + Generative Models



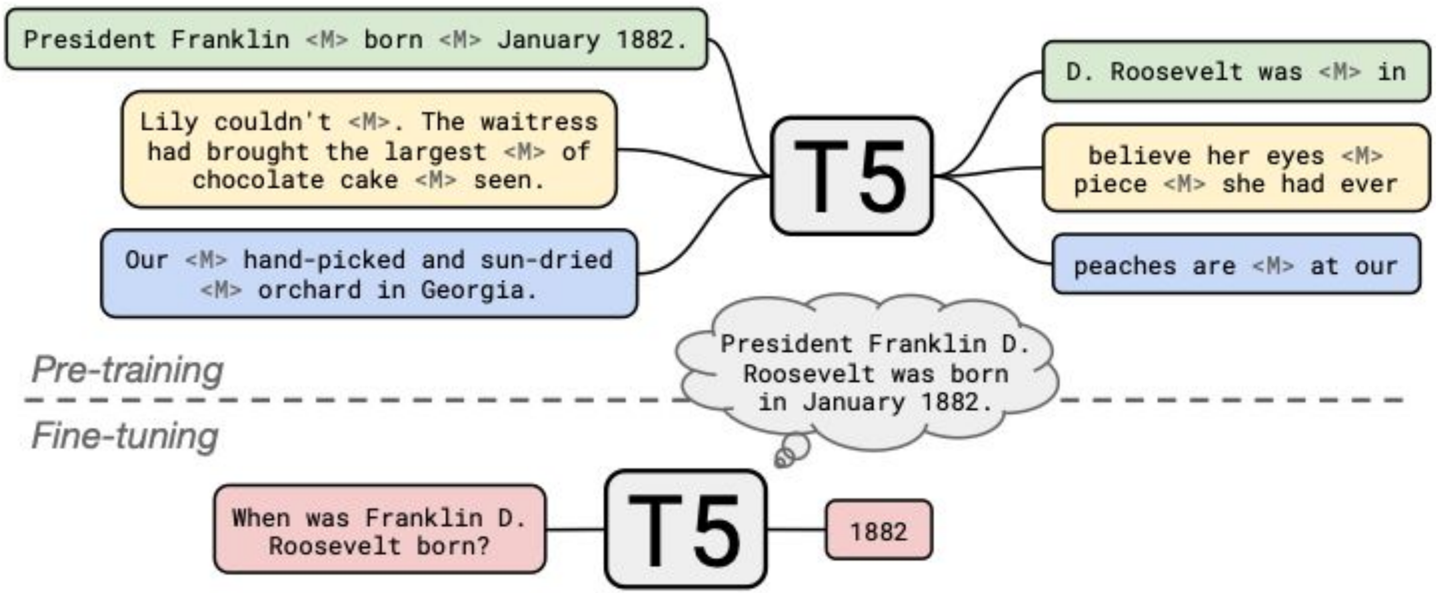
Model	NaturalQuestions	TriviaQA	
ORQA (Lee et al., 2019)	31.3	45.1	-
REALM (Guu et al., 2020)	38.2	-	-
DPR (Karpukhin et al., 2020)	41.5	57.9	-
SpanSeqGen (Min et al., 2020)	42.5	-	-
RAG (Lewis et al., 2020)	44.5	56.1	68.0
T5 (Roberts et al., 2020)	36.6	-	60.5
GPT-3 few shot (Brown et al., 2020)	29.9	-	71.2
Fusion-in-Decoder (base)	48.2	65.0	77.1
Fusion-in-Decoder (large)	**51.4**	**67.6**	**80.1**

Fusion-in-decoder (FID)

DPR + T5

Izacard and Grave, 2020.

Generative Models for Open-Domain QA

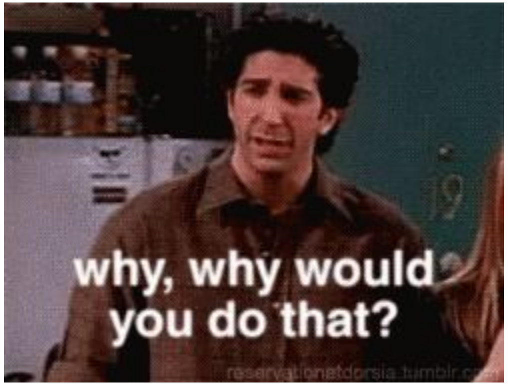


Roberts et al, 2020.

Answering Why-Questions

Matt and Sarah were pregnant.
They wanted to announce it in a fun way.
They wrote it on a cake.
They invited their friends over.
When their friends saw the cake, they were excited.

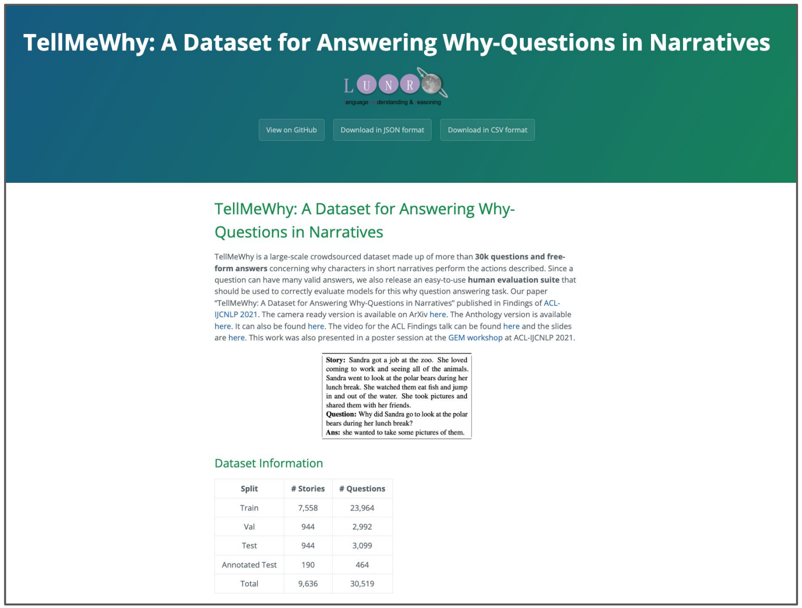
Why were Matt and Sarah pregnant?



Knowing why is important for reasoning about events

Lal et al, 2021.

Answering Why-Questions



TellMeWhy: A Dataset for Answering Why-Questions in Narratives

View on GitHub | Download in JSON format | Download in CSV format

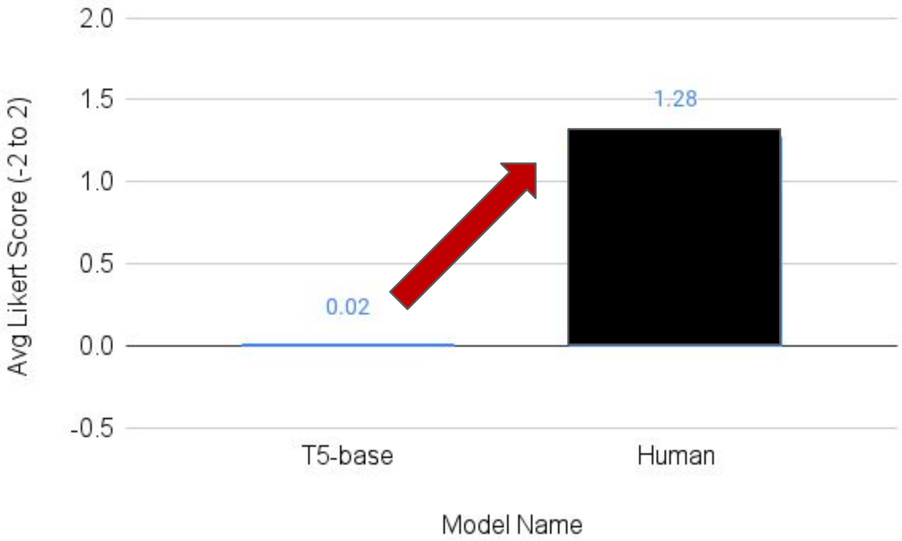
TellMeWhy: A Dataset for Answering Why-Questions in Narratives

TellMeWhy is a large-scale crowdsourced dataset made up of more than **30k questions and free-form answers** concerning why characters in short narratives perform the actions described. Since a question can have many valid answers, we also release an easy-to-use **human evaluation suite** that should be used to correctly evaluate models for this why question answering task. Our paper "TellMeWhy: A Dataset for Answering Why-Questions in Narratives" published in Findings of ACL-IJCNLP 2021. The camera ready version is available on ArXiv here. The Anthology version is available here. It can also be found here. The video for the ACL Findings talk can be found here and the slides are here. This work was also presented in a poster session at the GEM workshop at ACL-IJCNLP 2021.

Story: Sandra got a job at the zoo. She loved coming to work and seeing all of the animals. Sandra went to look at the polar bears during her lunch break. She watched them eat fish and jump in and out of the water. She took pictures and shared them with her friends.
Question: Why did Sandra go to look at the polar bears during her lunch break?
Ans: she wanted to take some pictures of them.

Dataset Information

Split	# Stories	# Questions
Train	7,558	23,964
Val	944	2,992
Test	944	3,099
Annotated Test	190	464
Total	9,636	30,519



Lal et al, 2021.

Using Commonsense to Answer Why-Questions

Matt and Sarah were pregnant.
They wanted to announce it in a fun way.
They wrote it on a cake.
They invited their friends over.
When their friends saw the cake, they were excited.

Q: Why were Matt and Sarah pregnant?

Commonsense Knowledge:
- ❑ become pregnant to have babies
- ❑ can become pregnant from sexual intercourse

T5

They wanted to have a baby

Lal et al, 2022.

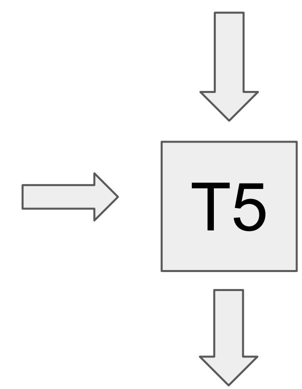
Using Commonsense to Answer Why-Questions

Matt and Sarah were pregnant.
They wanted to announce it in a fun way.
They wrote it on a cake.
They invited their friends over.
When their friends saw the cake, they were excited.

Q: Why were Matt and Sarah pregnant?

- Larger Model
- External Commonsense Resource

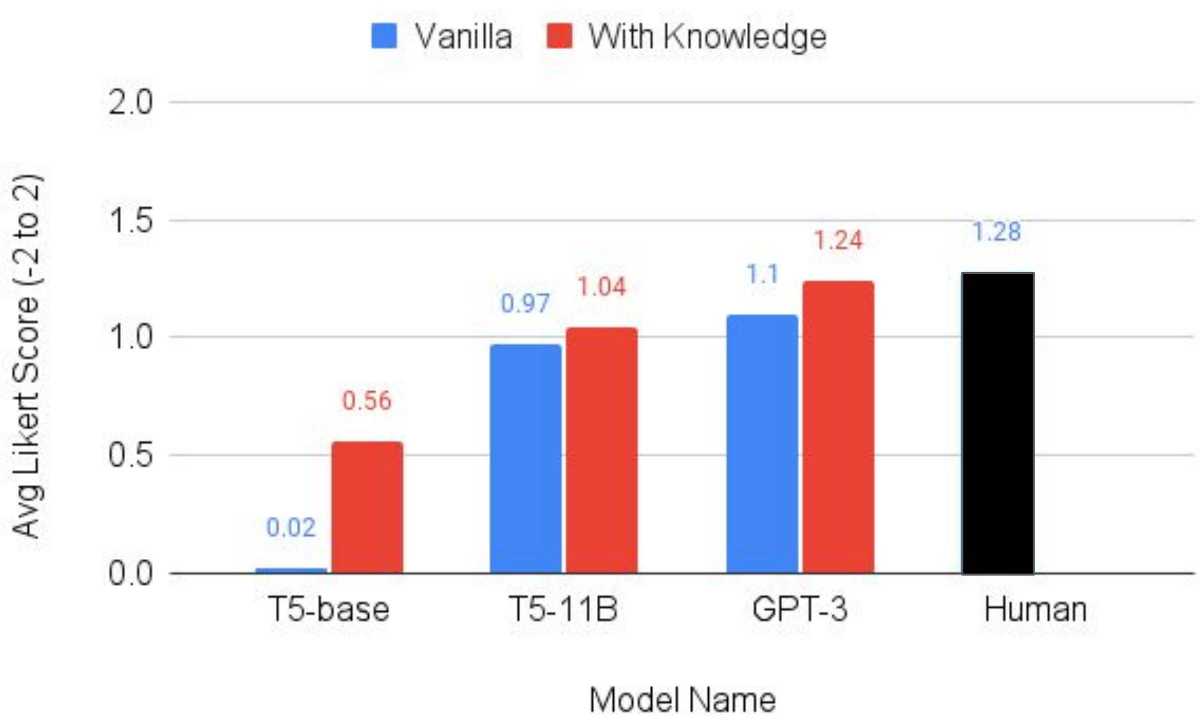
Commonsense Knowledge:
- ❑ become pregnant to have babies
- ❑ can become pregnant from sexual intercourse



They wanted to have a baby

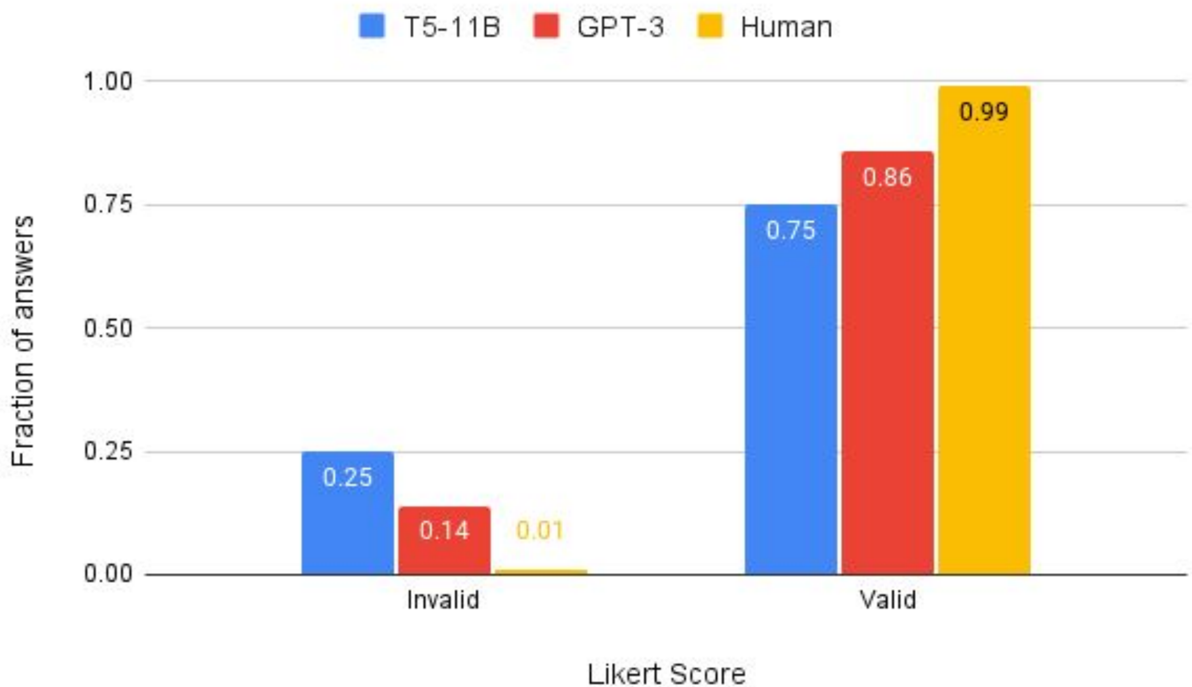
Lal et al, 2022.

How Good are Models?



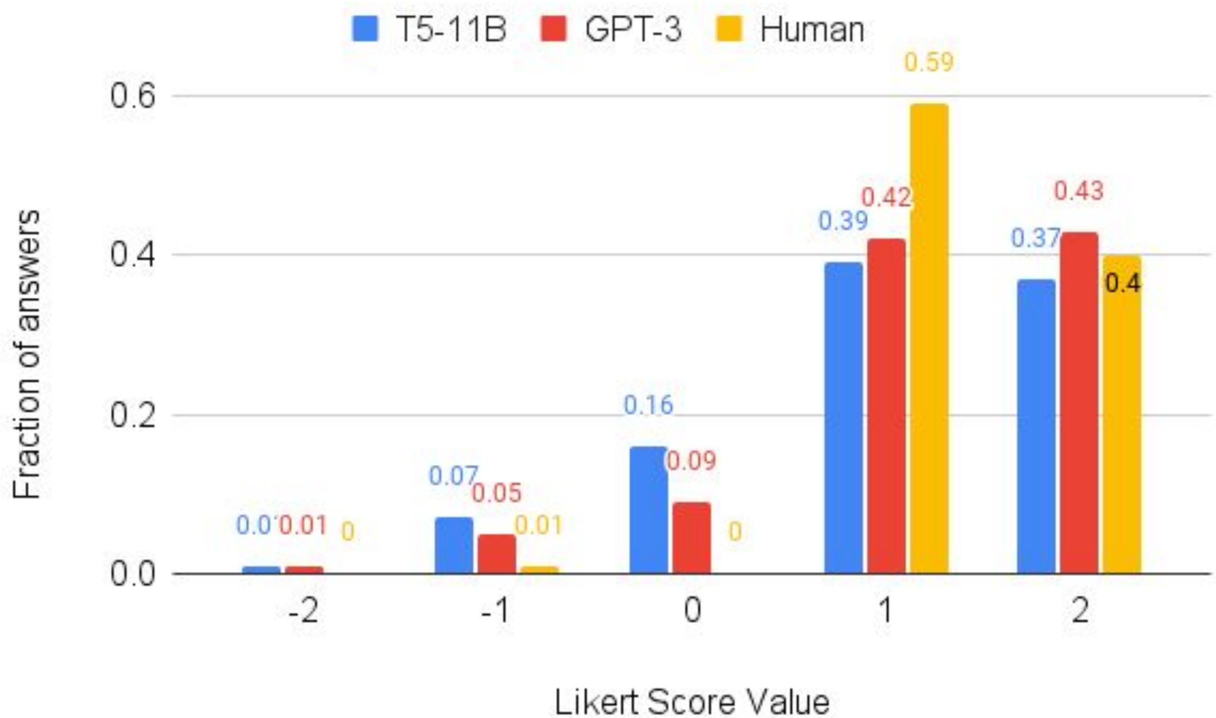
Lal et al, 2022.

Looking Closer at Likert Values



Lal et al, 2022.

Looking Closer at Likert Values



Lal et al, 2022.

Multi-Step Questions Answering

Questions to answer which we need multiple steps of reasoning.

What are Multi-Step Questions

Questions to answer which we need multiple steps of reasoning.

Where did OpenAI's CEO go for undergrad?

- Who is OpenAI's CEO ⇒ Sam Altman
- Where did Sam Altman go for undergrad? ⇒ Stanford University

Stanford University

Why should we care about Multi-Step Questions?

Why should we care about Multi-Step Questions?

Many of our day-to-day information needs require multi-step reasoning.

Why should we care about Multi-Step Questions?

Many of our day-to-day information needs require multi-step reasoning.

Which vegetarian restaurants near me are open if I’ve a peanut allergy?

- Find a list of open restaurants near me.
- Select the ones which have vegetarian options in the menu.
- Select the ones which have peanut-free options in the menu.

Why should we care about Multi-Step Questions?

Many of our day-to-day information needs require multi-step reasoning.

Can I finish GOT Season 7 if I’ve 10 hours this weekend?

- Get a list of episodes and duration of GOT from season 7.
- Sum the time duration of GOT for all the episodes.
- Check if the total duration is less than 10 hours.

To satisfy such information needs, we need models that perform multi-step reasoning.

What are the Challenges of Multi-Step QA?

Challenges of Multi-Step QA

Challenges of Multi-Step QA

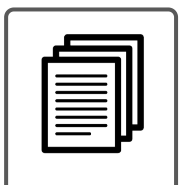
Reading Comprehension QA

Where did OpenAI's CEO go for undergrad?

Challenges of Multi-Step QA

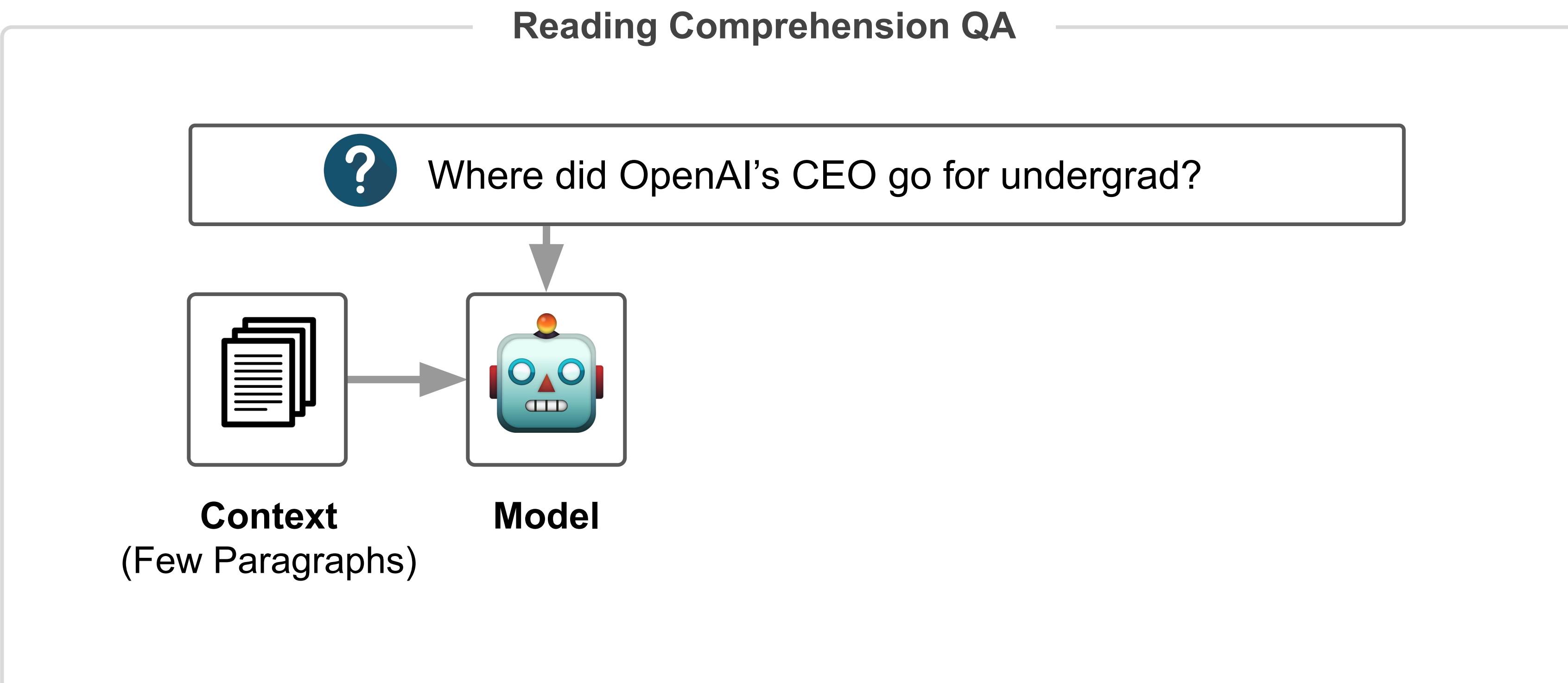
Reading Comprehension QA

Where did OpenAI's CEO go for undergrad?

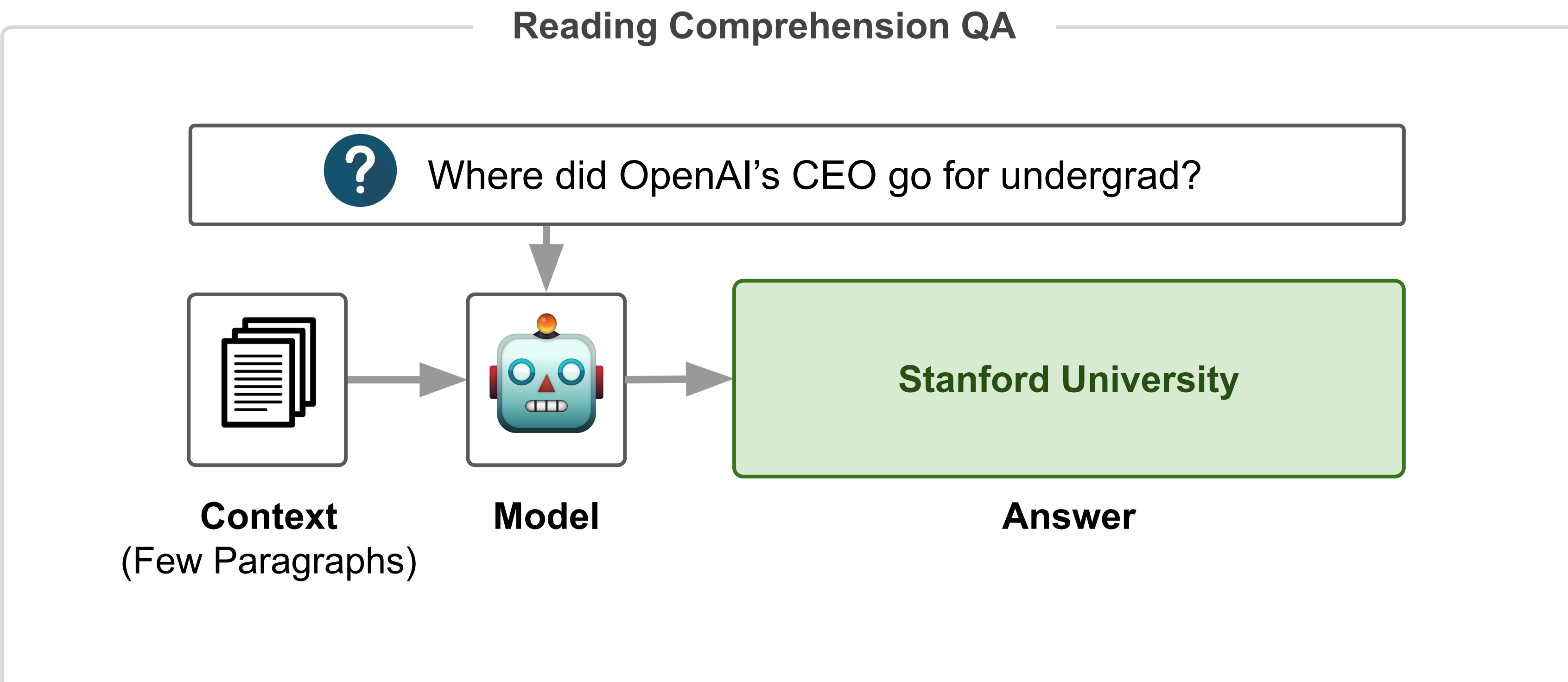


Context
(Few Paragraphs)

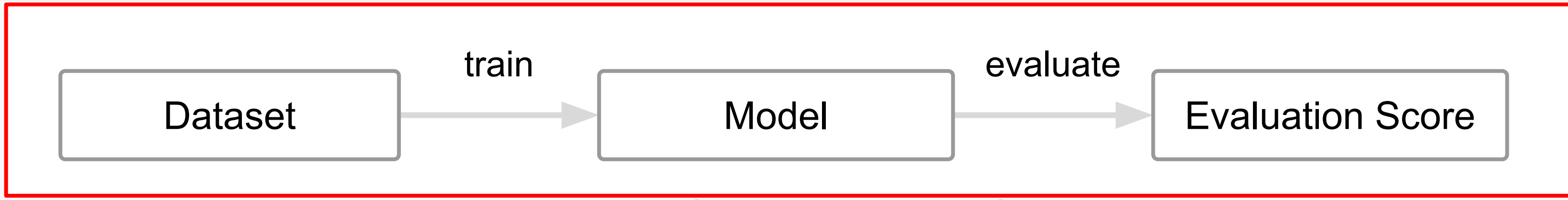
Challenges of Multi-Step QA

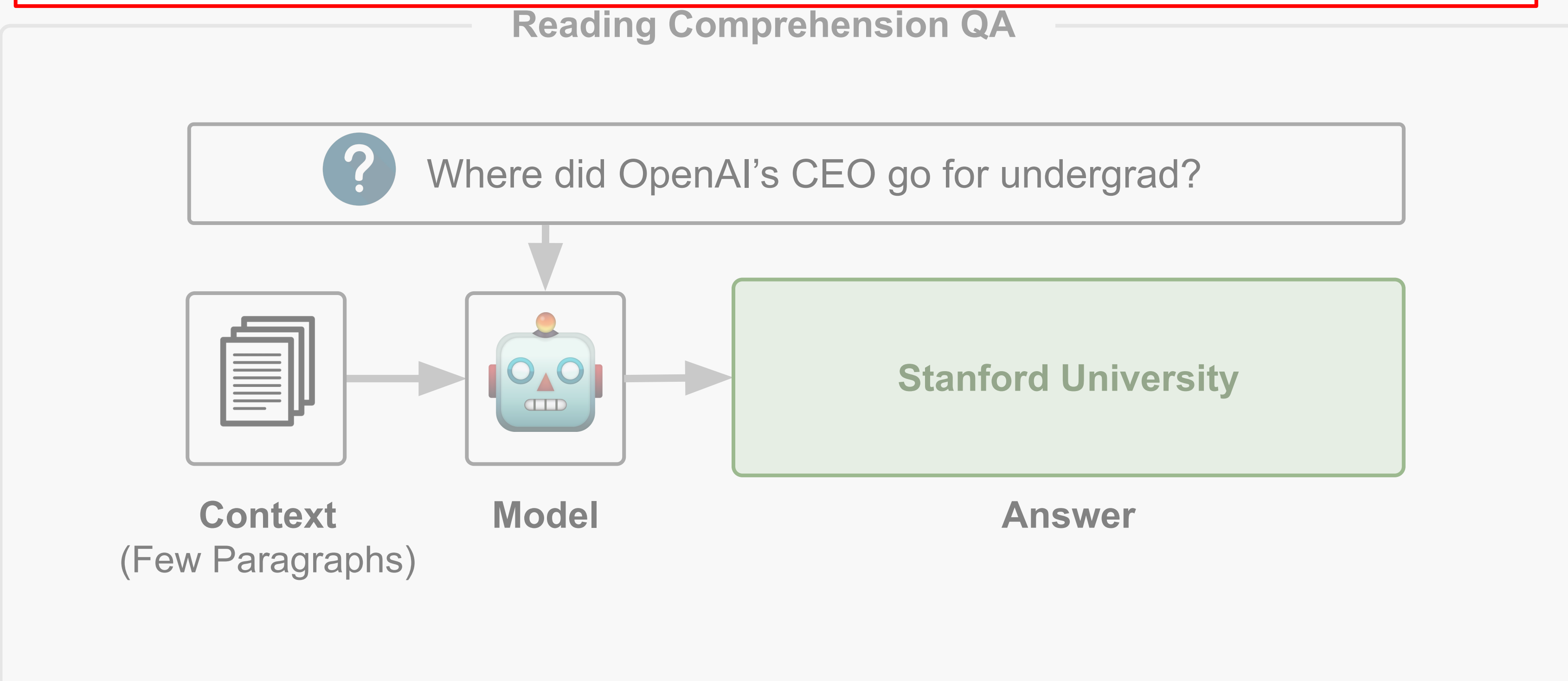


Challenges of Multi-Step QA



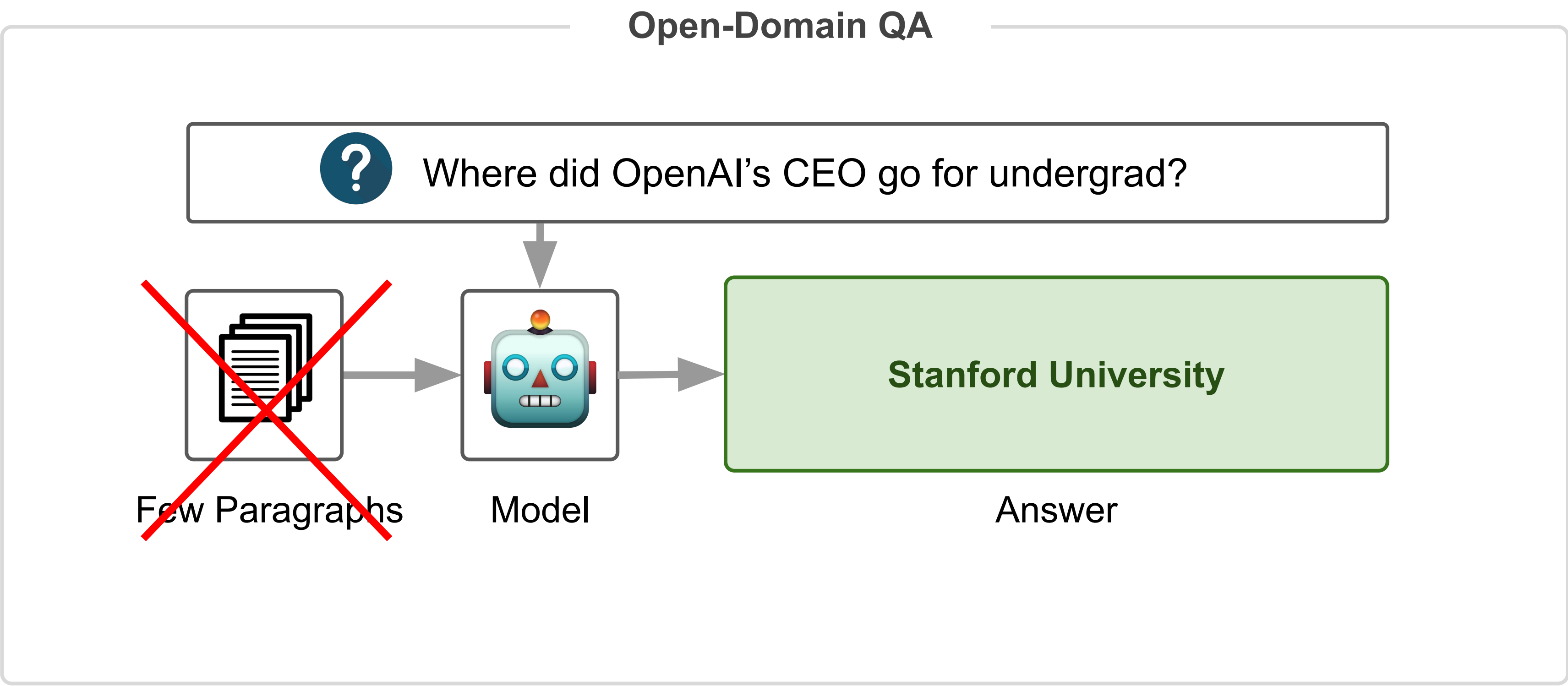
Challenges of Multi-Step QA





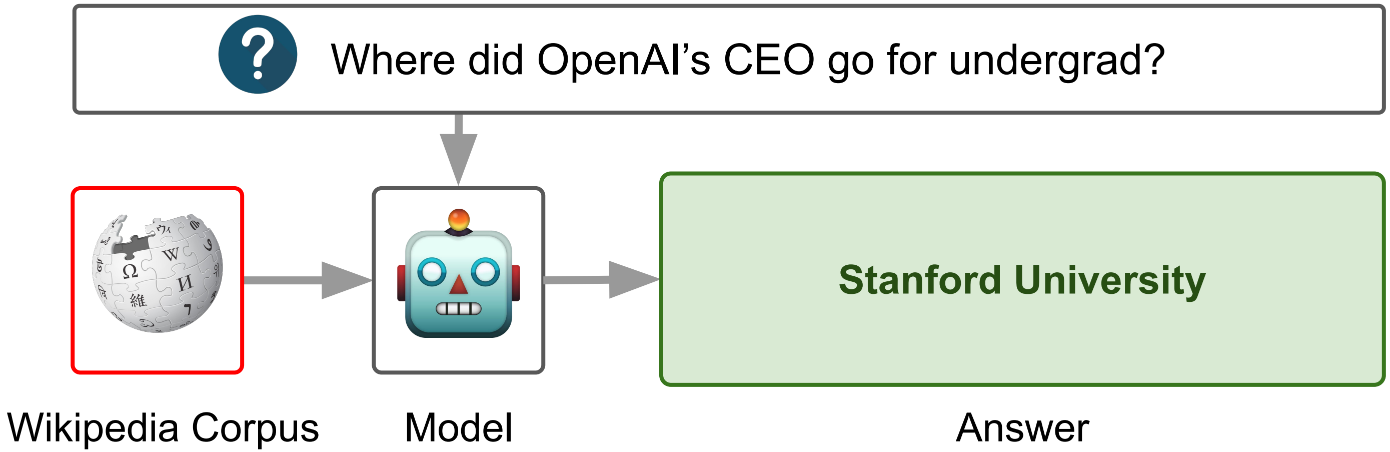
Challenges of Multi-Step QA

Challenges of Multi-Step QA

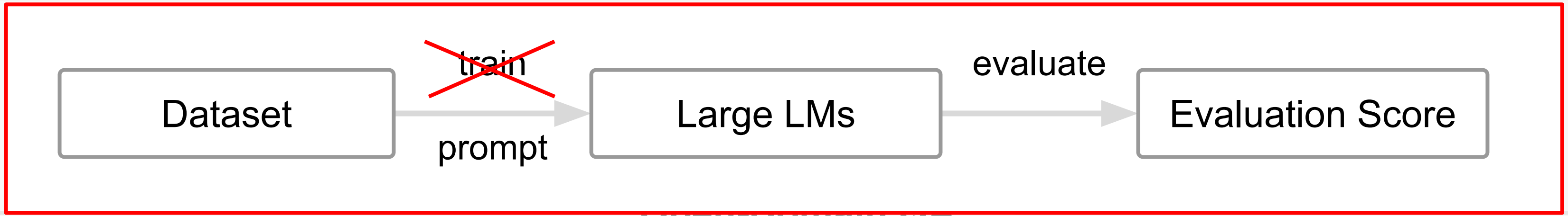


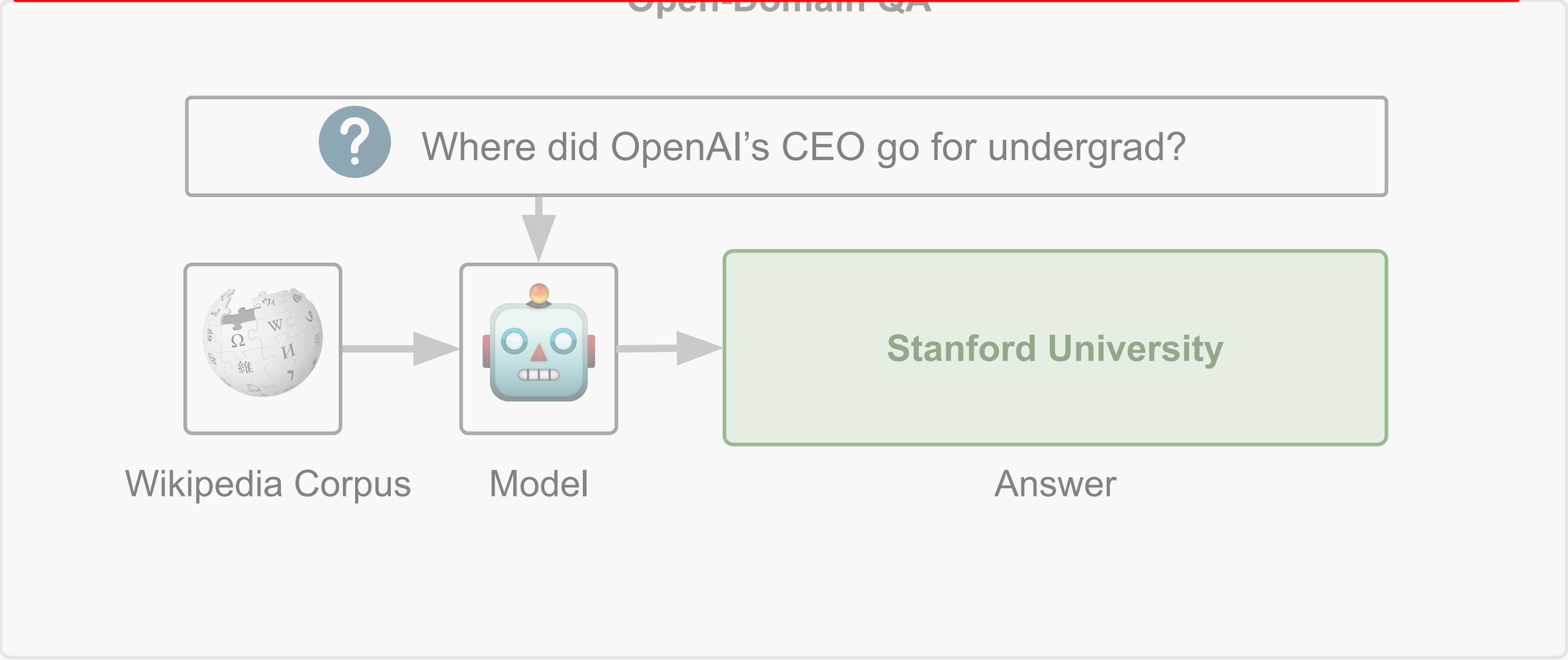
Challenges of Multi-Step QA

Open-Domain QA

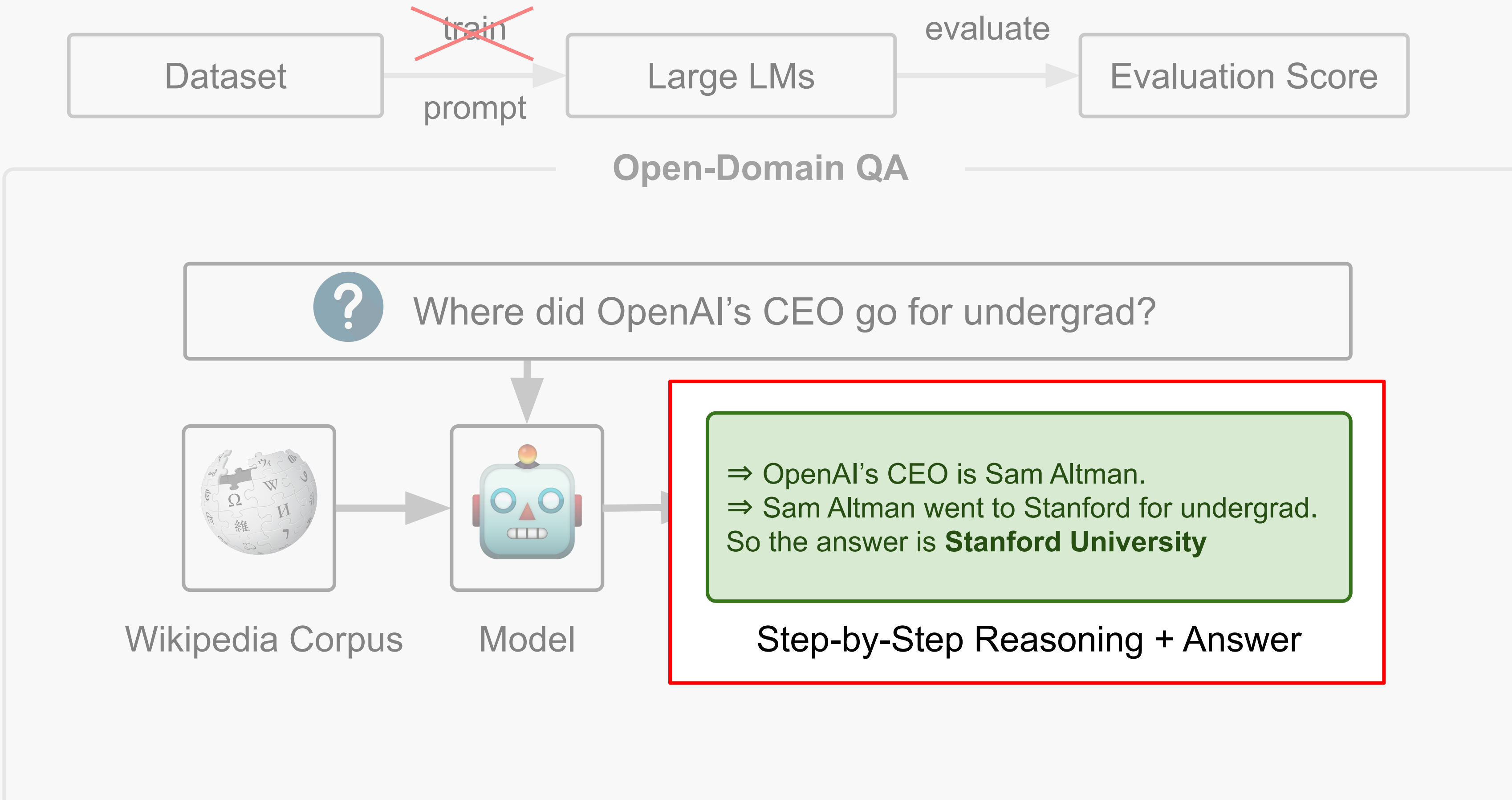


Challenges of Multi-Step QA





Challenges of Multi-Step QA



State of Few-Shot Multi-Step Open-Domain QA

Model	HpQA^{Br}	HpQA	2WikiMQA	MQ^{2H}
InterAug	– | –	30.3 | –	– | –	– | –
ReAct	– | –	35.1 | –	– | –	– | –
SelfAsk	– | –	– | –	40.1 | –	15.2 | –
DecomP	– | 50.0	– | –	– | 59.3	– | –
IRCoT QA	**45.8 | 58.5**	**49.3 | 60.7**	**57.7 | 68.0**	**34.2 | 43.8**

EM | F1

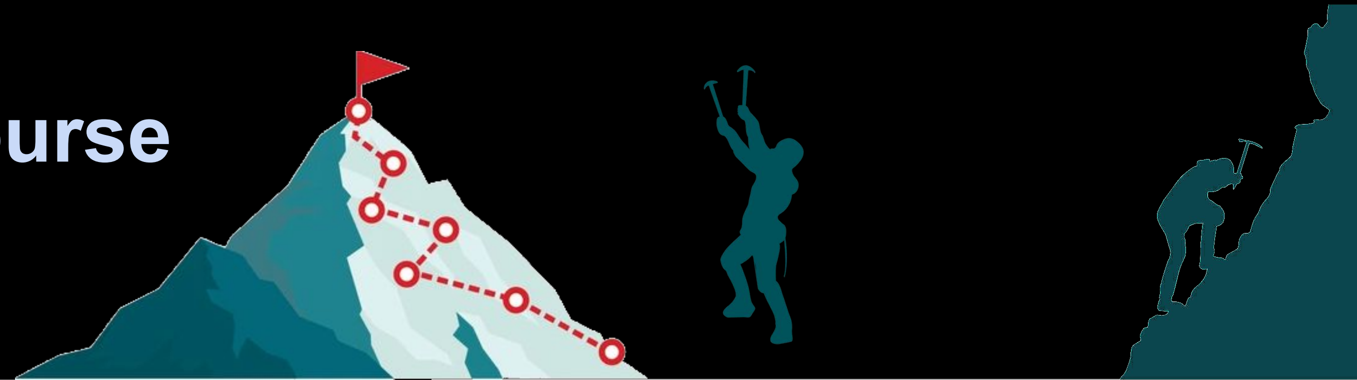
⇒ InterAug: (Internet-augmented LMs through few-shot prompting for ODQA) Lazaridou et. al.
⇒ ReAct: (ReAct: Synergizing Reasoning and Acting in Language Models) Yao et. al
⇒ SelfAsk: (Measuring and Narrowing the Compositionality Gap in Language Models) Press et. al.
⇒ DecomP: (Decomposed Prompting: A Modular Approach for Solving Complex Tasks) Khot et. al.

Machine Translation

CSE538 - Spring 2024

Guest Lecturer: Yash Kumar Lal

NLP, The Course



Overall NLP Concept

I. Syntax

Introduction to NLP; Tokenization; Words Corpora

One-hot, and Multi-hot encoding.
Parts-of-Speech; Named Entities;

Parsing; Verbal Predicates;Dependency Parsing

II. Semantics

Dependency Parsing; Word Sense Disambiguation

Vector Semantics (Embeddings), Word2vec

Probabilistic Language Models
Ngram Classifier, Topic Modeling

Overall NLP Concept

III. Language Modeling

Ethical Considerations

Masked Language Modeling (autoencoding)

Generative Language Modeling (autoregressive)

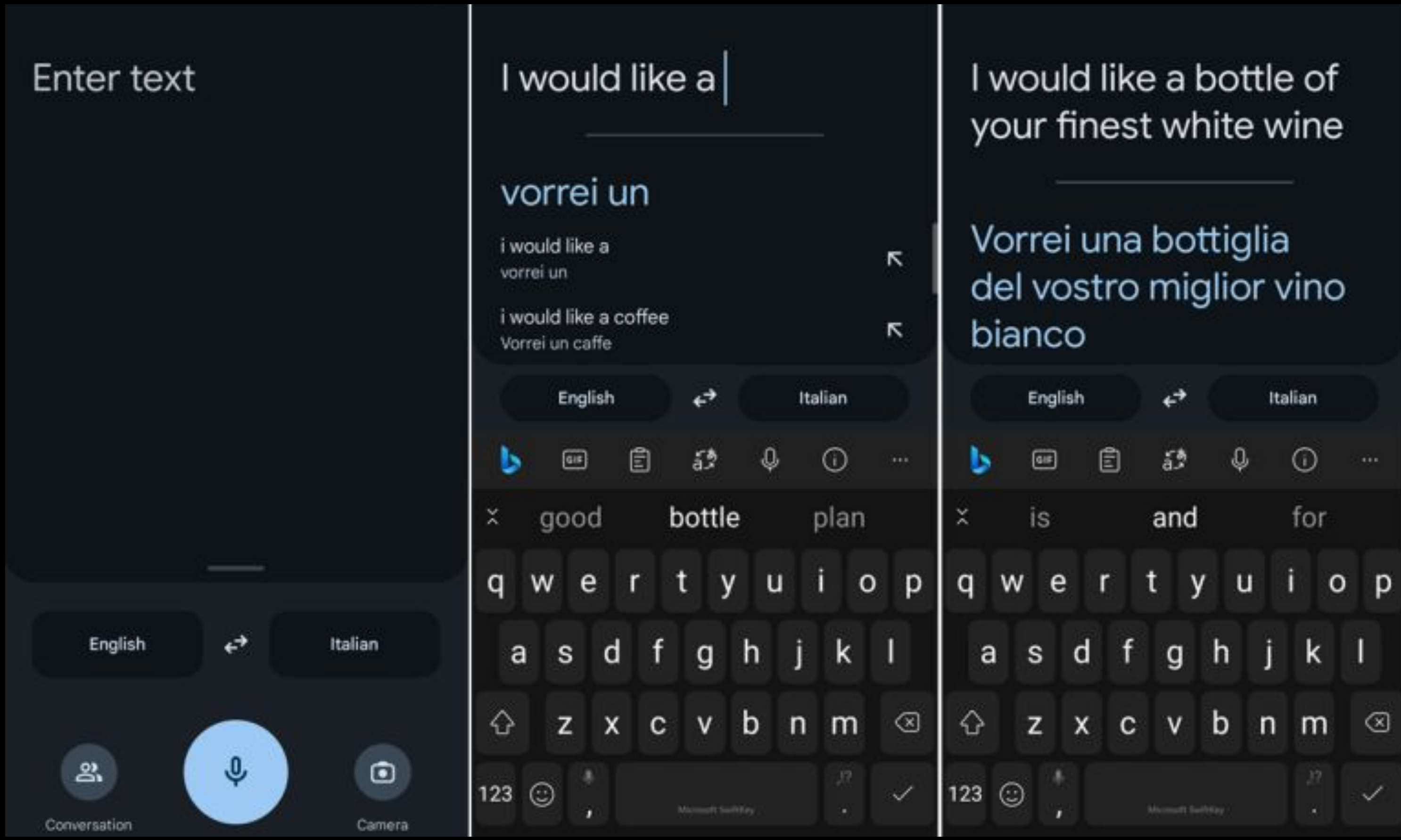
Applying LMs

IV. Applications

Language and Psychology
(advanced sentiment)

Speech and Audio Processing, Dialog (chatbots)

Question Answering, Translation


Enter text
English
Italian
Conversation
Camera
I would like a
vorrei un
i would like a
vorrei un
i would like a coffee
Vorrei un caffe
English
Italian
good
bottle
plan
I would like a bottle of your finest white wine
Vorrei una bottiglia del vostro miglior vino bianco
English
Italian
is
and
for

Translate this sentence

Aquellos vestidos caros son italianos.

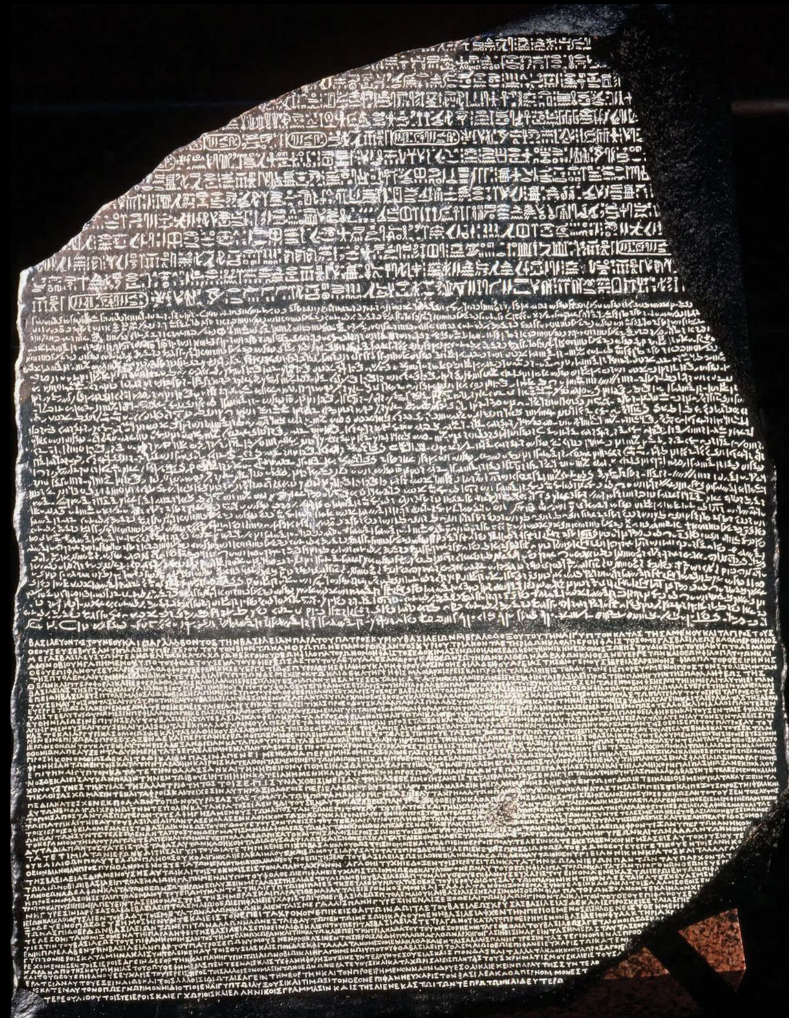
your | music | wines | are

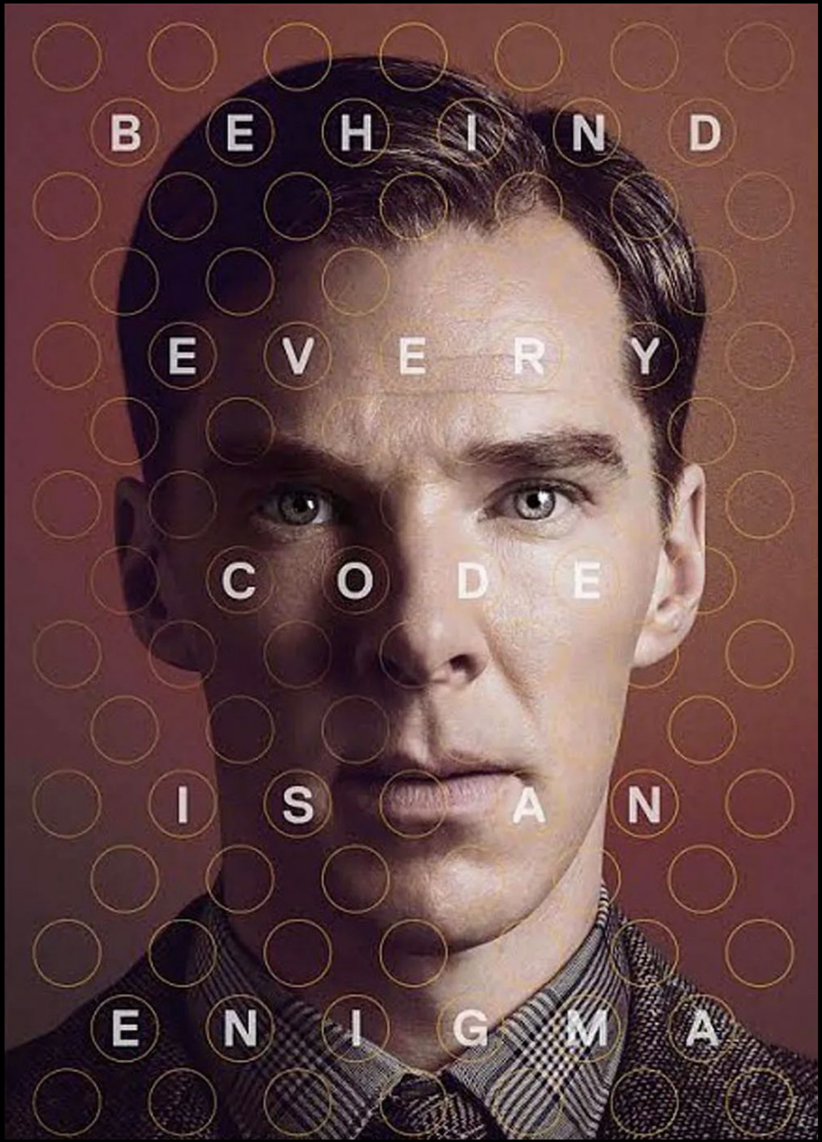
these | dresses | Those

expensive | Italian

CHECK

duolingo




BEHIND
EVERY
CODE
IS AN
ENIGMA

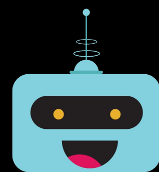
Machine Translation

X: Ces robes chères sont italiennes

French

Machine Translation

X: Ces robes chères sont italiennes



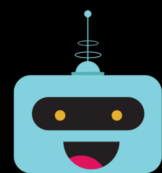
French

English

Machine Translation

X: Ces robes chères sont italiennes

French



Y: Those expensive dresses are Italian

English

History of Machine Translation

- Rosetta Stone to understand Egyptian hieroglyphics
- World War 2 code-breaking efforts (Alan Turing)
- 1940s-mid 1960s: Rule-based systems (RBMT)
- 1990s-2013: Statistical systems (SMT)
- 2014+: Neural systems (NMT)

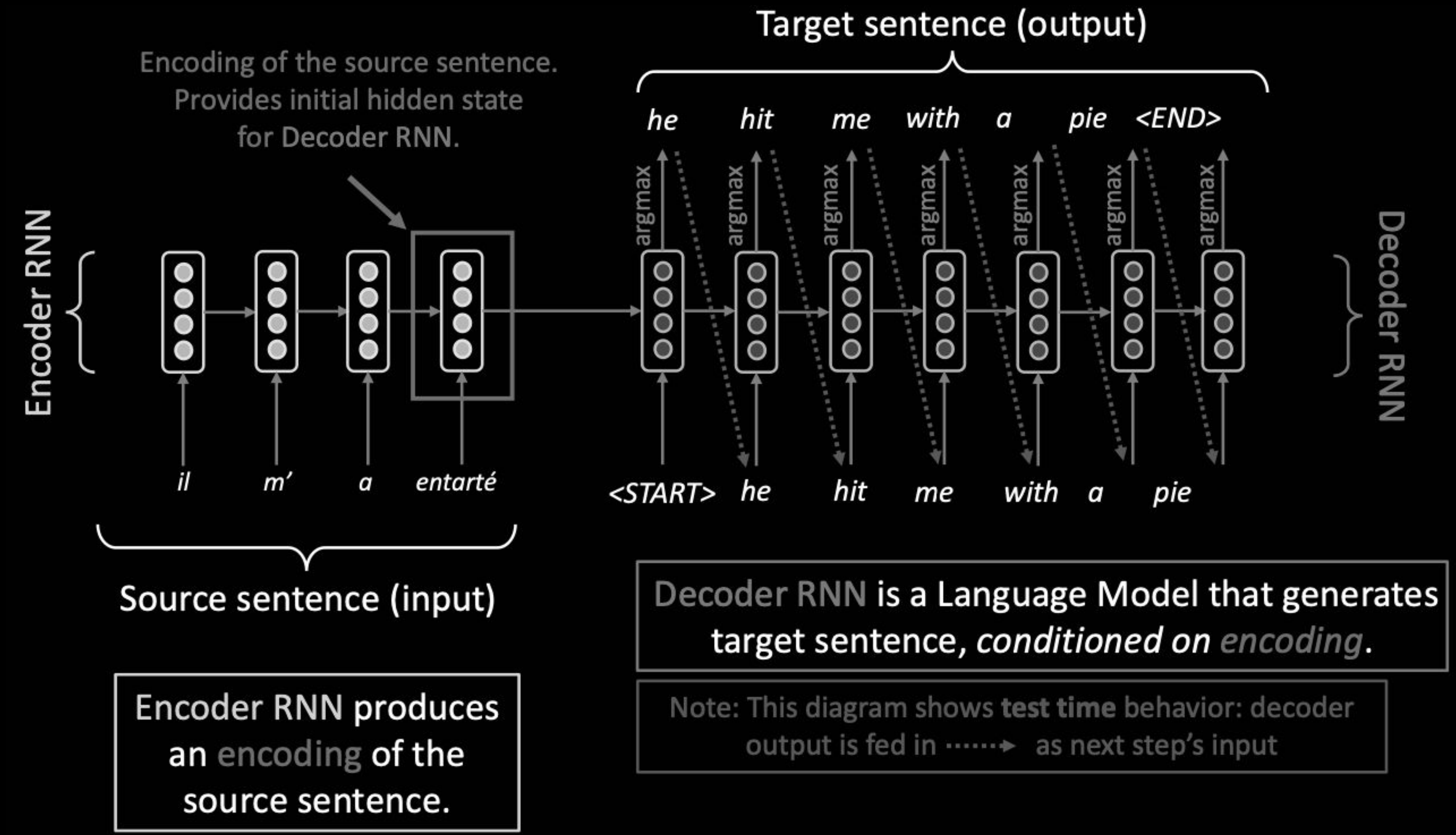
History of Machine Translation

- Rosetta Stone to understand Egyptian hieroglyphics
- World War 2 code-breaking efforts (Alan Turing)
- 1940s-mid 1960s: Rule-based systems (RBMT)
- 1990s-2013: Statistical systems (SMT)
- 2014+: Neural systems (NMT)

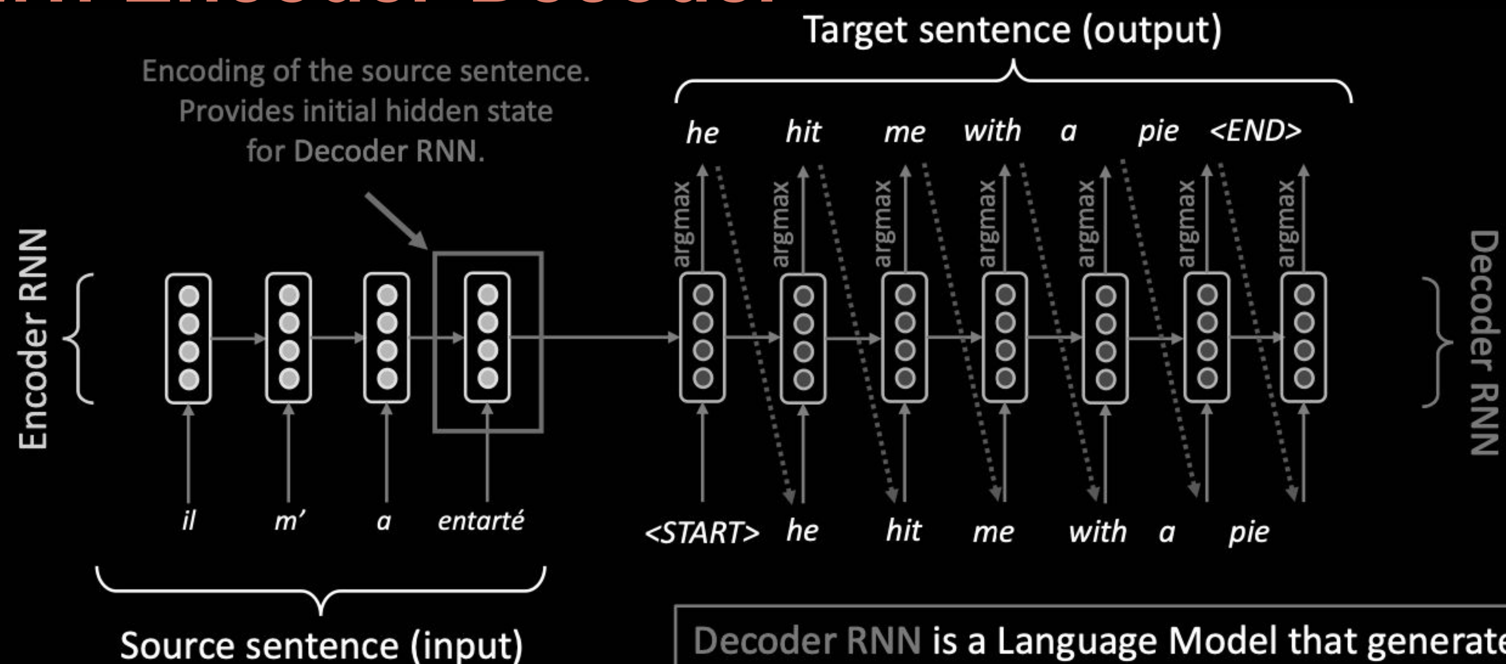
Neural Machine Translation

- Encoder-Decoder Models
- Attention Mechanism
- Transformers

NMT: Encoder-Decoder



NMT: Encoder-Decoder



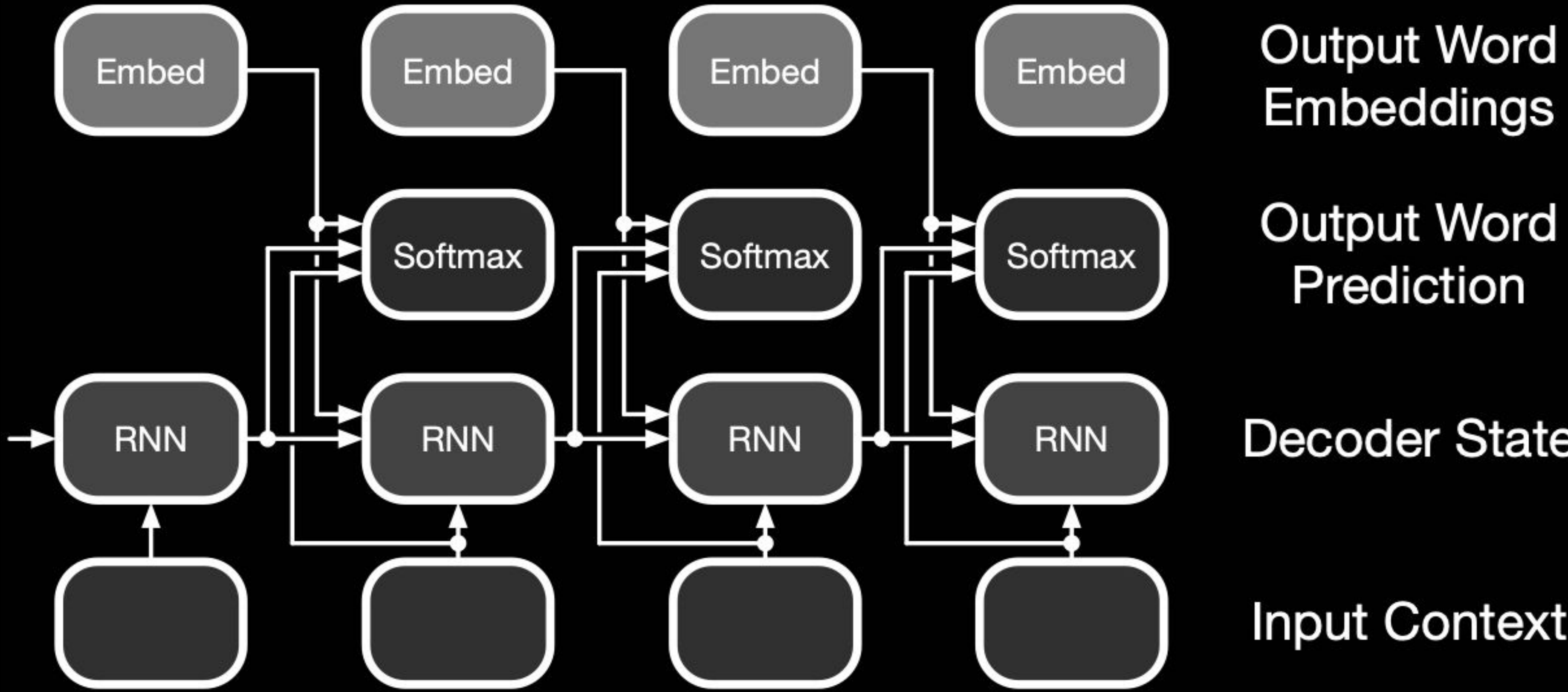
Encoder RNN produces an encoding of the source sentence.

Decoder RNN is a Language Model that generates target sentence, *conditioned on encoding*.

Note: This diagram shows **test time** behavior: decoder output is fed in as next step’s input

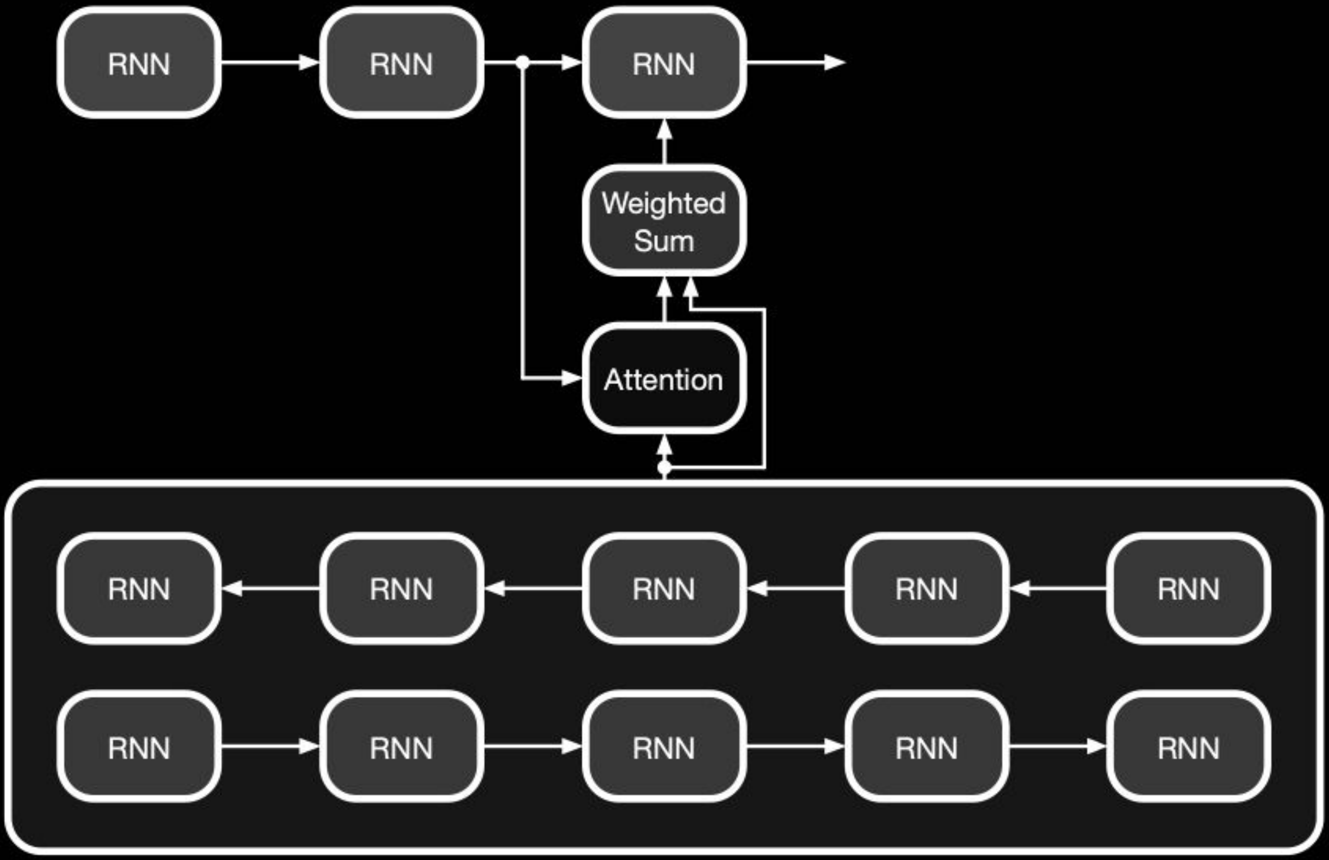
$$P(y|x) = P(y_1|x)\,P(y_2|y_1,x)\,P(y_3|y_1,y_2,x)\dots\underbrace{P(y_T|y_1,\dots,y_{T-1},x)}_{\text{Probability of next target word, given target words so far and source sentence } x}$$

NMT: Encoder-Decoder



Sutskever et al., 2014

NMT: Attention



Decoder State

Input Context

Attention

Right-to-Left Encoder

Left-to-Right Encoder

Bahdanau et al., 2014

NMT: Attention

Model	All	No UNK°
RNNencdec-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNencdec-50	17.82	26.71
RNNsearch-50	26.75	34.16
RNNsearch-50⋆	28.45	36.15
Moses	33.30	35.63

NMT is close to or outperforms SMT!

Bahdanau et al., 2014

NMT: Attention

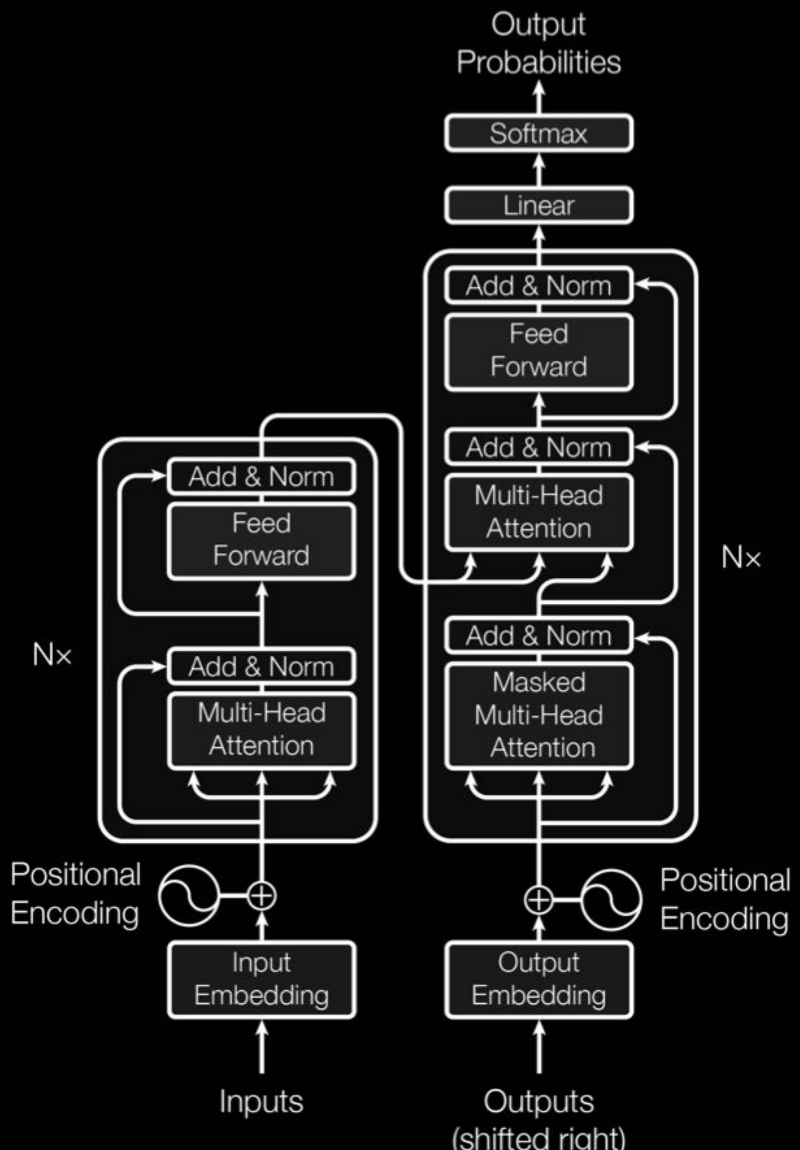
Model	All	No UNK°
RNNencdec-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNencdec-50	17.82	26.71
RNNsearch-50	26.75	34.16
RNNsearch-50*	28.45	36.15
Moses	33.30	35.63

NMT is close to or outperforms SMT!

Sequential computation = extremely time-taking
Cannot handle very long sequence lengths

Bahdanau et al., 2014

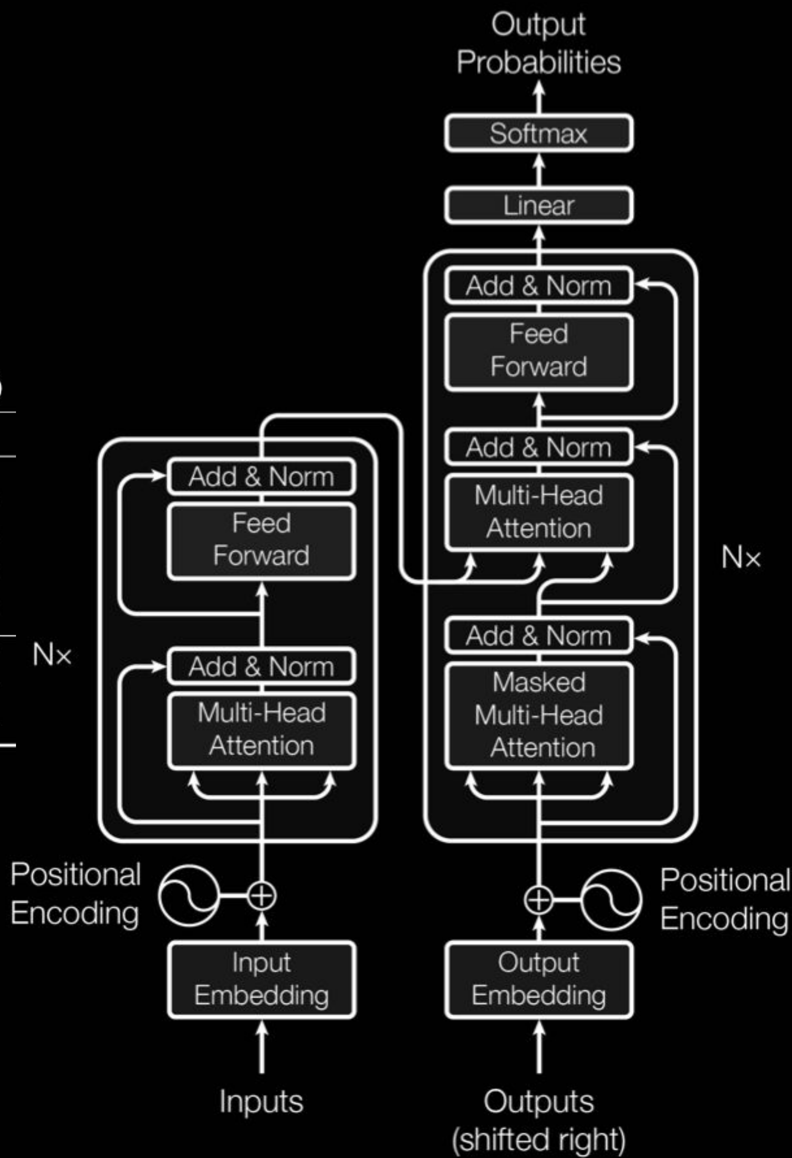
NMT: Transformer



Vaswani et al., 2017

NMT: Transformer

Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet [18]	23.75			
Deep-Att + PosUnk [39]		39.2		$1.0 \cdot 10^{20}$
GNMT + RL [38]	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [32]	26.03	40.56	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [39]		40.4		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [38]	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	26.36	**41.29**	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	$\mathbf{3.3 \cdot 10^{18}}$	
Transformer (big)	**28.4**	**41.8**	$2.3 \cdot 10^{19}$	



Vaswani et al., 2017

Training NMT Systems

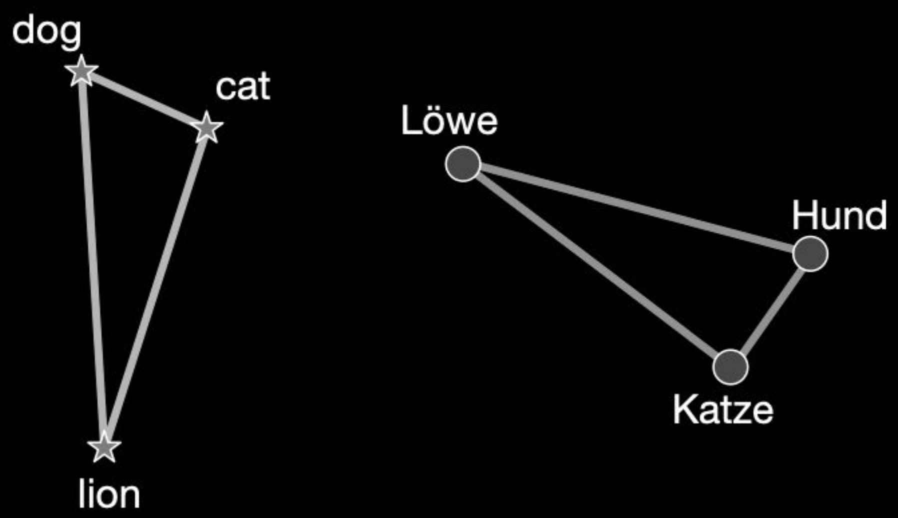
- Parallel Data
- Transfer Learning
- Data Augmentation using Backtranslation
- Unsupervised Methods

Backtranslation

- Train a system in the reverse language translation
- Use this to translate target side monolingual data (synthetic data)
- Combine generated synthetic parallel data with real parallel data to build the final system
- Can be done iteratively: Iterative Backtranslation

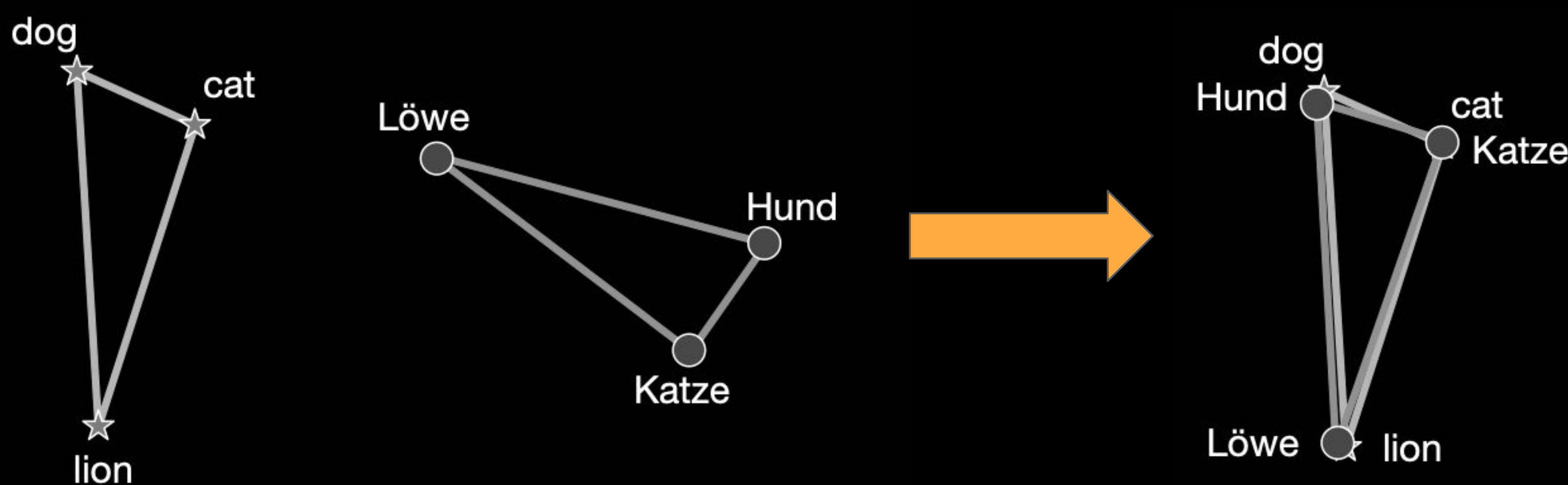
Sennrich et al., 2016
Hoang et al., 2018

Unsupervised Machine Translation



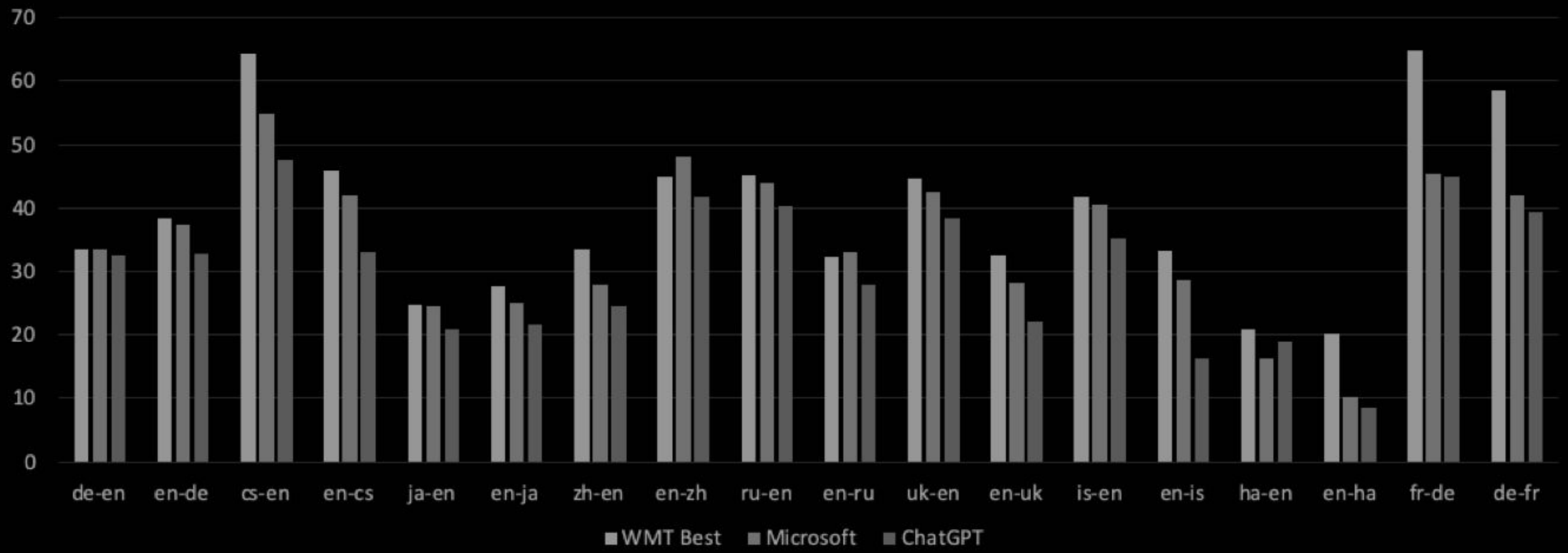
Embedding spaces for different language have similar shape

Unsupervised Machine Translation



How can we rotate the triangles to match up?

MT and LLMs



Despite VERY large-scale pretraining data, LLMs are not great

Challenges

- Reliable Evaluation
- Low-resource MT
- Multilinguality
- Document+ level translation

Human Evaluation

Given: MT output, reference translation

Task: Assess quality of MT output

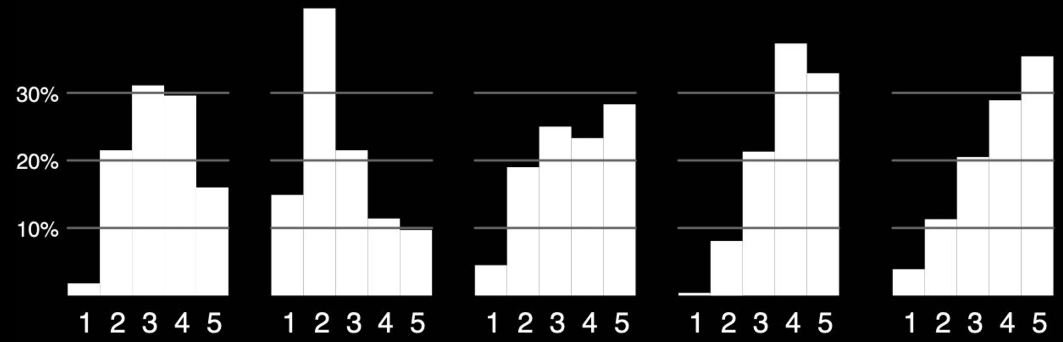
- Adequacy: Does output convey same meaning as input?
- Fluency: Is the output fluent in the target language?

Human Evaluation

Given: MT output, reference translation

Task: Assess quality of MT output

- Adequacy: Does output convey same meaning as input?
- Fluency: Is the output fluent in the target language?



Annotators disagree too; AND they use Google Translate themselves!!!

Automatic Evaluation

Given: MT output, reference translation

Task: Compute similarity between them

Automatic Evaluation

- Lexical metrics
- Trained metrics

Automatic Evaluation

- Lexical metrics
  - Word Error Rate (WER)

$$\text{WER} = \frac{\textit{substitutions} + \textit{insertions} + \textit{deletions}}{\textit{reference-length}}$$

Think Levenshtein Distance!

Automatic Evaluation

- Lexical metrics
  - BLEU

$$\text{BLEU} = \min\left(1, \frac{\textit{output-length}}{\textit{reference-length}}\right) \left(\prod_{i=1}^{4} \textit{precision}_i\right)^{\frac{1}{4}}$$

Papineni et al., 2002

Automatic Evaluation

- Lexical metrics
  - BLEU

$$\frac{correct}{output\text{-}length}$$

$$\text{BLEU} = \min\left(1, \frac{output\text{-}length}{reference\text{-}length}\right) \left(\prod_{i=1}^{4} precision_i\right)^{\frac{1}{4}}$$

Sophisticated 4-gram Precision

Papineni et al., 2002

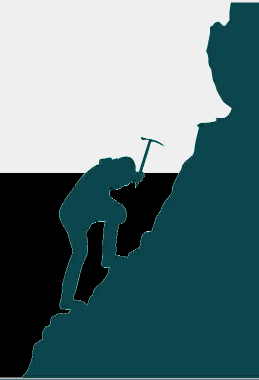
Automatic Evaluation

- Lexical metrics
  - Ignore relevance of words
  - Operate on local level
  - Human translators often have poor BLEU scores!

Low Resource Machine Translation

- 7000+ languages in the world
- <100,000 parallel sentences for most language pairs
- Massively multilingual MT - 1600 languages
- Datasets to advance this: FLORES, NLLB

NLP, The Course



Overall NLP Concept

I. Syntax

Introduction to NLP; Tokenization; Words Corpora

One-hot, and Multi-hot encoding.
Parts-of-Speech; Named Entities;

Parsing; Verbal Predicates;Dependency Parsing

II. Semantics

Dependency Parsing; Word Sense Disambiguation

Vector Semantics (Embeddings), Word2vec

Probabilistic Language Models
Ngram Classifier, Topic Modeling

Overall NLP Concept

III. Language Modeling

Ethical Considerations

Masked Language Modeling (autoencoding)

Generative Language Modeling (autoregressive)

Applying LMs

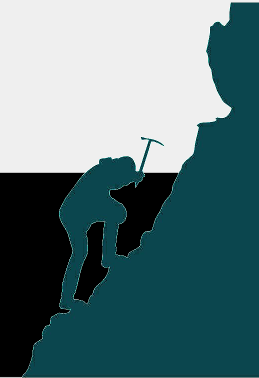
IV. Applications

Language and Psychology
(advanced sentiment)

Speech and Audio Processing

Question Answering, Translation

NLP, The Course



Overall NLP Concept

I. Syntax

Introduction to NLP; Tokenization; Words Corpora

One-hot, and Multi-hot encoding.
Parts-of-Speech; Named Entities;

Parsing; Verbal Predicates;Dependency Parsing

II. Semantics

Dependency Parsing; Word Sense Disambiguation

Vector Semantics (Embeddings), Word2vec

Probabilistic Language Models
Ngram Classifier, Topic Modeling

Overall NLP Concept

III. Language Modeling

Ethical Considerations

Masked Language Modeling (autoencoding)

Generative Language Modeling (autoregressive)

Applying LMs

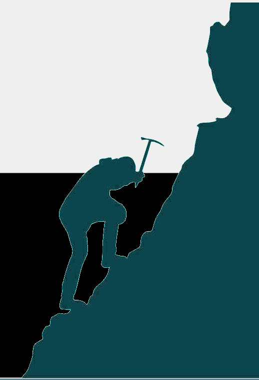
IV. Applications

Language and Psychology
(advanced sentiment)

Speech and Audio Processing

Question Answering, Translation

NLP, The Course



Overall NLP Concept

I. Syntax

Introduction to NLP; Tokenization; Words Corpora

One-hot, and Multi-hot encoding.
Parts-of-Speech; Named Entities;

Parsing; Verbal Predicates;Dependency Parsing

II. Semantics

Dependency Parsing; Word Sense Disambiguation

Vector Semantics (Embeddings), Word2vec

Probabilistic Language Models
Ngram Classifier, Topic Modeling

Overall NLP Concept

III. Language Modeling

Ethical Considerations

Masked Language Modeling (autoencoding)

Generative Language Modeling (autoregressive)

Applying LMs

IV. Applications

Language and Psychology
(advanced sentiment)

Speech and Audio Processing

Question Answering, Translation

NLP, The Course

Overall NLP Concept	Computation or ML
I. Syntax | Classification	
Introduction to NLP; Tokenization; Words Corpora	Regular Expressions; Edit Distance
One-hot, and Multi-hot encoding. Parts-of-Speech; Named Entities;	Maximum Entropy Classifier (LogReg), Gradient Descent,
Parsing; Verbal Predicates;Dependency Parsing	Cross Validation; Regularization Accuracy Metrics; Shift Reduce
II. Semantics | Probabilistic Models	
Dependency Parsing; Word Sense Disambiguation	Term Probabilities; N-d Vectors
Vector Semantics (Embeddings), Word2vec	LDA, Skipgram Model
Probabilistic Language Models Ngram Classifier, Topic Modeling	markov assumption, chain rule, smoothing

Overall NLP Concept	Computation or ML
III. Language Modeling | Transformers	
Ethical Considerations	Model cards, Pred Bias Frmwrk
Masked Language Modeling (autoencoding)	Neural Networks; Backprop Cross-Entropy Loss Self-Attention,
Generative Language Modeling (autoregressive)	Positional encodings The Transformer: Beam Search
Applying LMs	Fine-Tuning, zero-/few-shot, Instruction tuning
IV. Applications | Custom Statistical or Symbolic	
Language and Psychology (advanced sentiment)	Differential Language Analysis; Adaptive Modeling; Human LMing
Speech and Audio Processing,	Wave Transforms; RNNs
Question Answering, Translation	Multihop Reasoning