

Open Domain Question Answering

What's the difference between Open Domain QA and Machine Reading?

Standard Machine Reading



QUESTION

What temperature is a fever?



CONTEXT

Although a fever (pyrexia) could be considered any body temperature above the normal temperature of 98.6 degrees Fahrenheit (98.6 F or 37 C), medically, a person is not considered to have a significant fever until the temperature is above 100.4 F (38.0 C).

Standard Machine Reading



QUESTION

What temperature is a fever?



CONTEXT

Although a fever (pyrexia) could be considered any body temperature above the normal temperature of 98.6 degrees Fahrenheit (98.6 F or 37 C), medically, a person is not considered to have a significant fever until the temperature is above **100.4** F (38.0 C).



ANSWER

100.4



Open Domain Question Answering



QUESTION

What temperature is a fever?



CONTEXT

Entire Wikipedia, Entire Web



ANSWER

100.4

Why do we care about Open Domain?

- In real life, we don't know where the answer is

Why do we care about Open Domain?

- In real life, we don't know where the answer is
- Information exists in many different locations on the web
 - consult multiple sources
 - contradictions? multiple references?

Why is Open Domain Difficult?

- Processing very large amounts of information

Why is Open Domain Difficult?

- Processing very large amounts of information
- Questions could be asked on any arbitrary topic

Why is Open Domain Difficult?

- Processing very large amounts of information
- Questions could be asked on any arbitrary topic

Why is Open Domain Difficult?

- Processing very large amounts of information
- Questions could be asked on any arbitrary topic
- Encompasses both retrieval and generation

Part 1: Knowledge Bases

Knowledge Bases

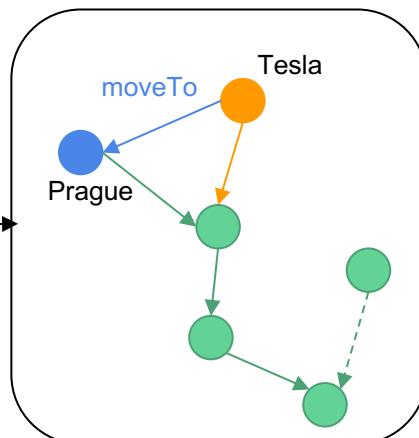
- KB: structured repository of knowledge (usually relational DB)
- Goal: encode knowledge so that it can be queried efficiently
- Scale can be huge: billions of facts, millions of entities
- KB can be generic or specific
- Examples: Cyc, WikiData, DBPedia, Google KG, GeneOntology, IMDB, etc.
- Key challenge is their construction!
 - Manually: Crowdsourcing, paid experts
 - Automatically: Information extraction or Automatic KB Construction



Automatic Knowledge Base Construction

In January 1880, two of Tesla's uncles put together enough money to help him leave Gospic for Prague where he was to study. Unfortunately, he arrived too late to enrol at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses.

[Text]



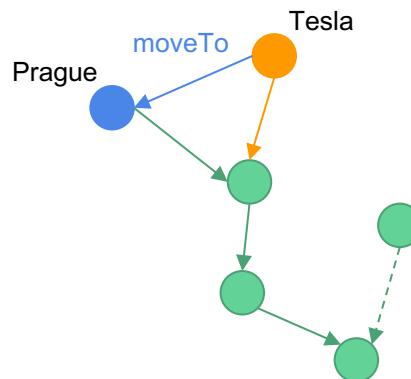
[Meaning]



[Information Need]

Knowledge Graph Construction

In January 1880, two of Tesla's uncles put together enough money to help him leave Gospic for Prague where he was to study. Unfortunately, he arrived too late to enrol at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses.



What city did Tesla move to in 1880?

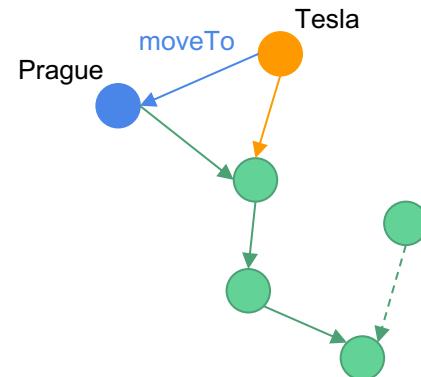
Prague

Knowledge Graph Construction

In January 1880, two of Tesla's uncles put together enough money to help him leave Gospic for Prague where he was to study. Unfortunately, he arrived too late to enrol at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses.

[Text]

X

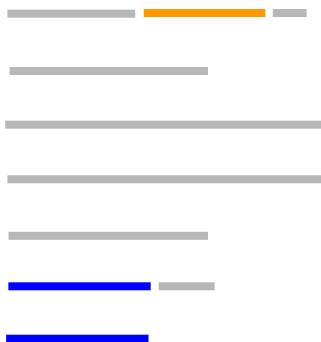


[Meaning]

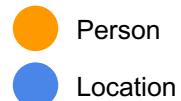
Y

Entity Extraction

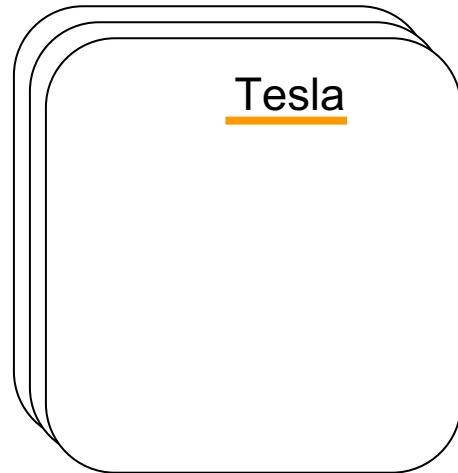
Two of Tesla's
uncles put
together enough
money to help
him leave
Gospic for
Prague



- Linear Chain CRF
- Bi-directional RNNs
- Hybrid RNN & CRFs



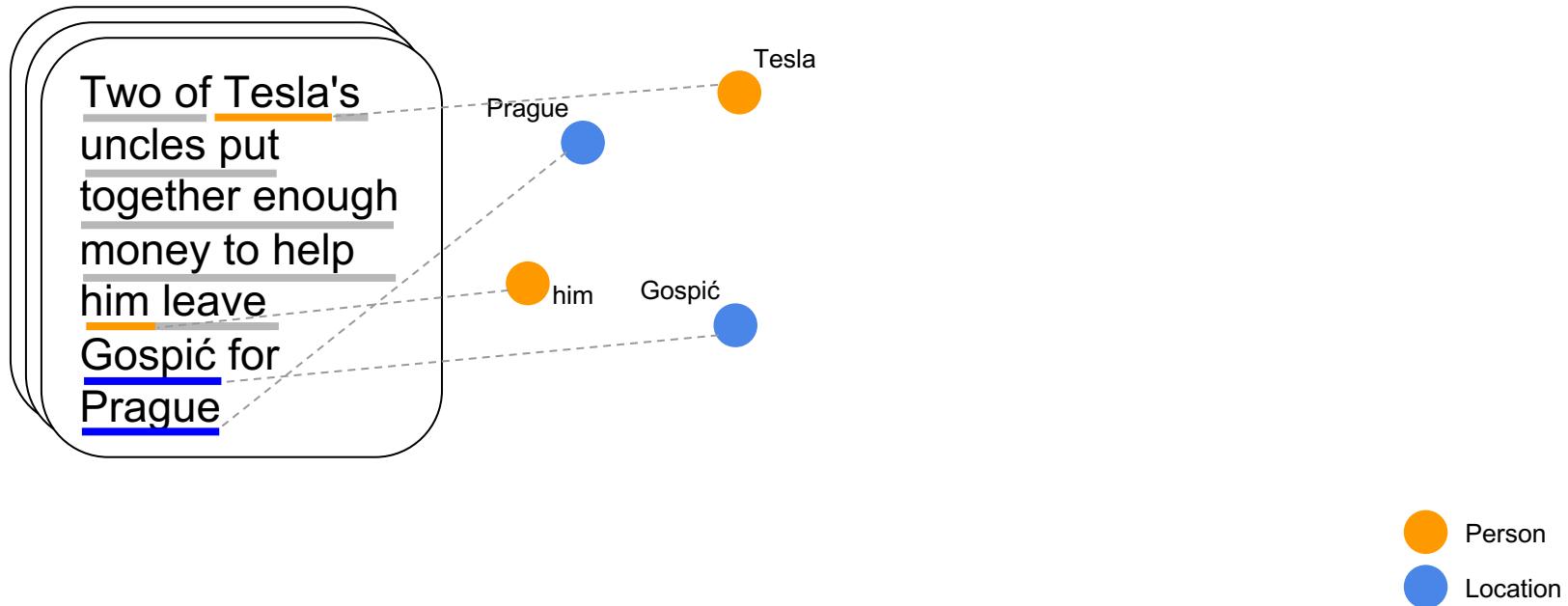
Challenge: Ambiguity



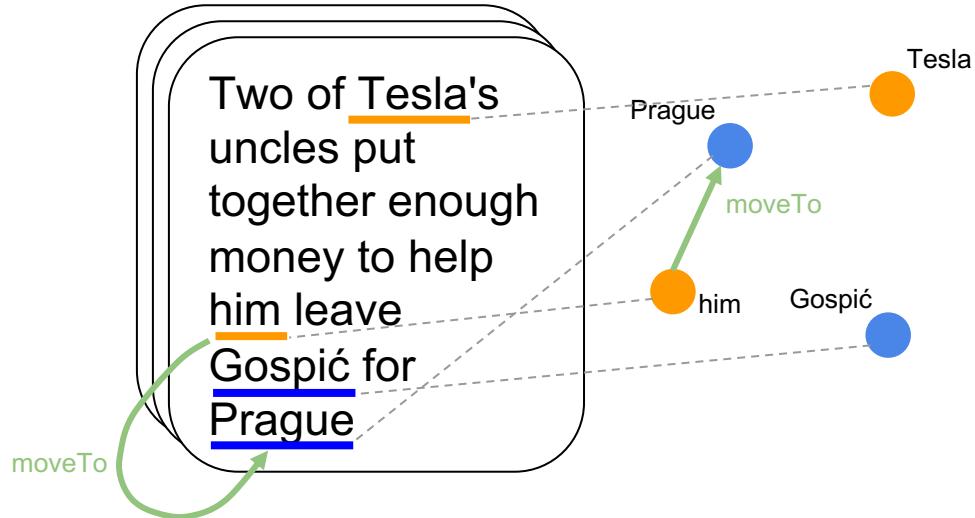
● Person?

● Brand?

Instantiate Nodes



Relation Extraction



- Neural Classification
- Distant Supervision

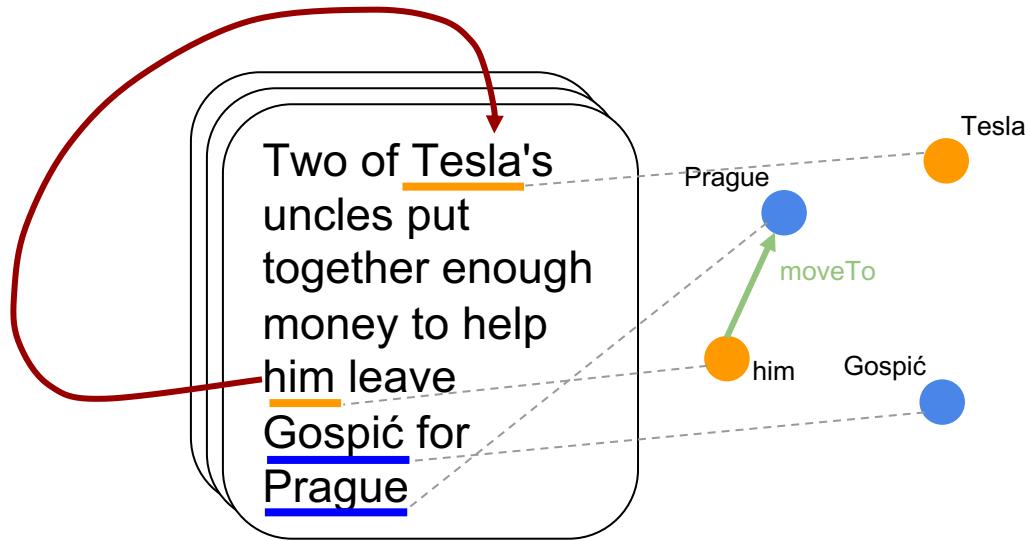
Challenge: Variation

Two of Tesla's uncles put together enough money to help **him leave Gospic for Prague**

Two of Tesla's uncles put together enough money to help **him move to Prague**

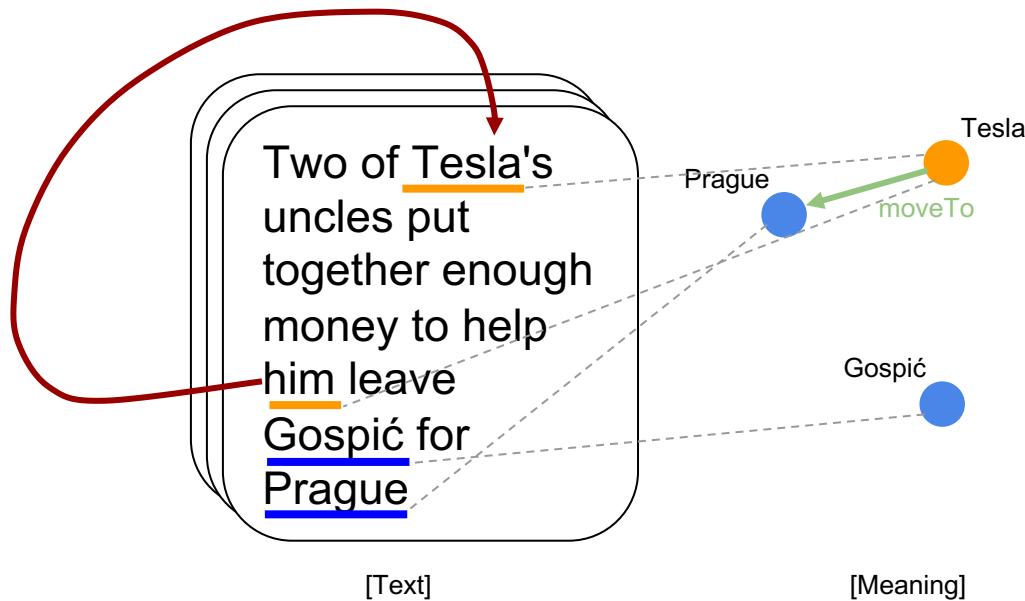
Two of Tesla's uncles put together enough money to help **him settle in Prague**

Coreference Resolution

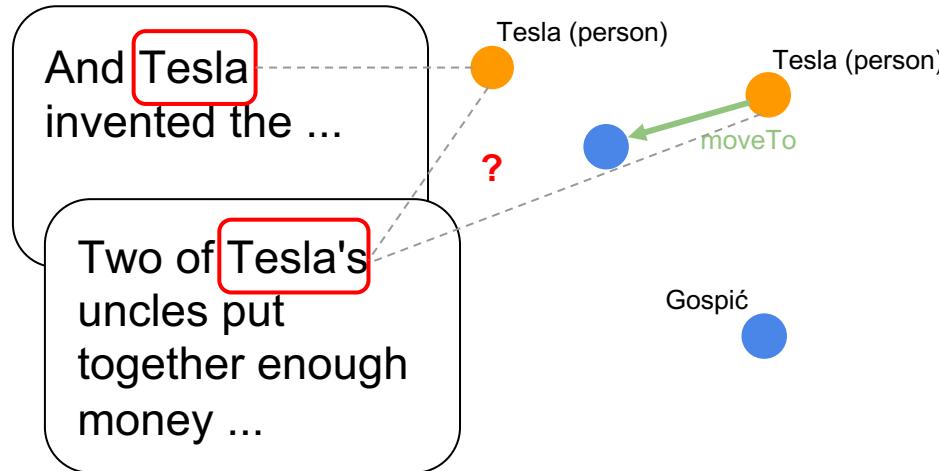


- Neural Classification
- Latent Variables

Collapsing Nodes

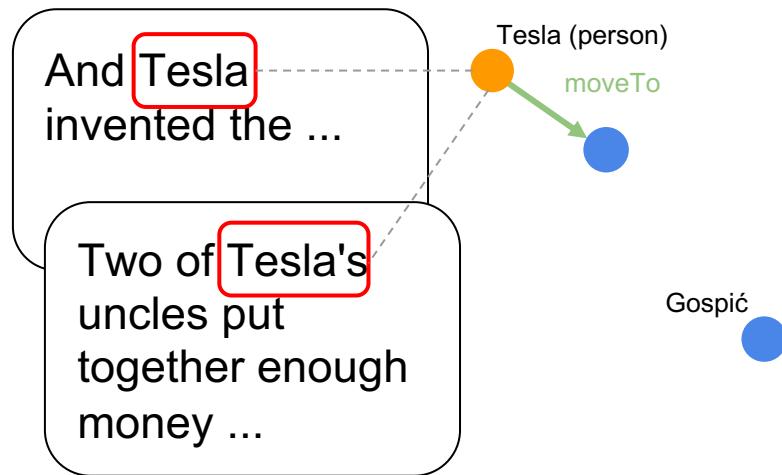


Entity Linking



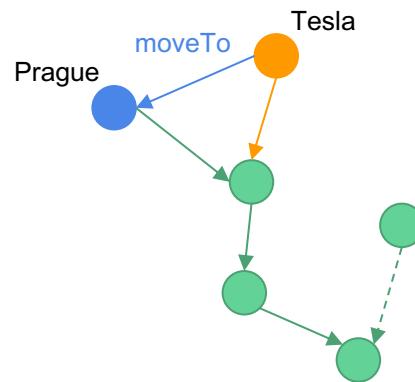
- Reranking ...
- Embeddings ...

Collapsing



Strengths

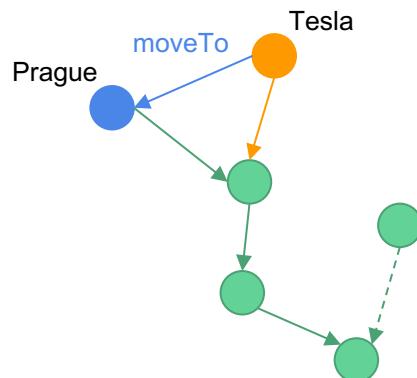
In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague where he was to study. Unfortunately, he arrived too late to enrol at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses.



- Supports Reasoning
- Fast access
- Generalisation
- Interpretable
- Existing KBs can serve as supervision signal!

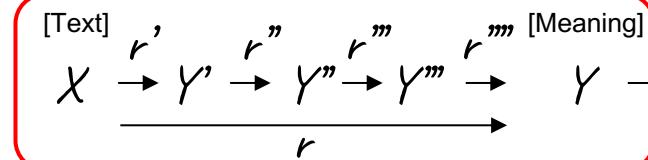
Weakness: Cascading errors

In January 1880, two of Tesla's uncles put together enough money to help him leave Gospic for Prague where he was to study. Unfortunately, he arrived too late to enrol at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses.



What city did Tesla move to in 1880?

Prague

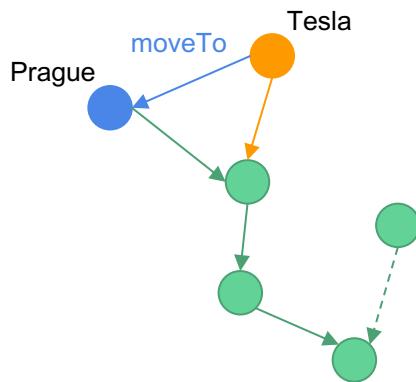


a



Weakness: Cascading errors

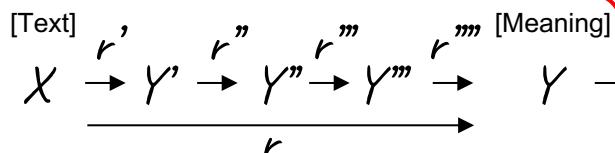
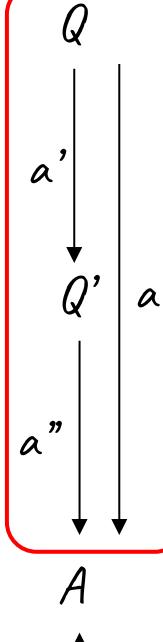
In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague where he was to study. Unfortunately, he arrived too late to enrol at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses.



What city did Tesla move to in 1880?

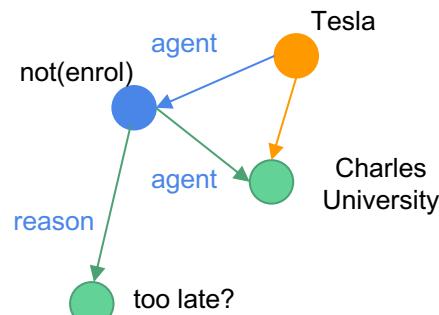
moveTo(Tesla,X)?

Prague



Weakness: Engineering Schemas and Formalisms

Unfortunately, he arrived too late to enrol at Charles University



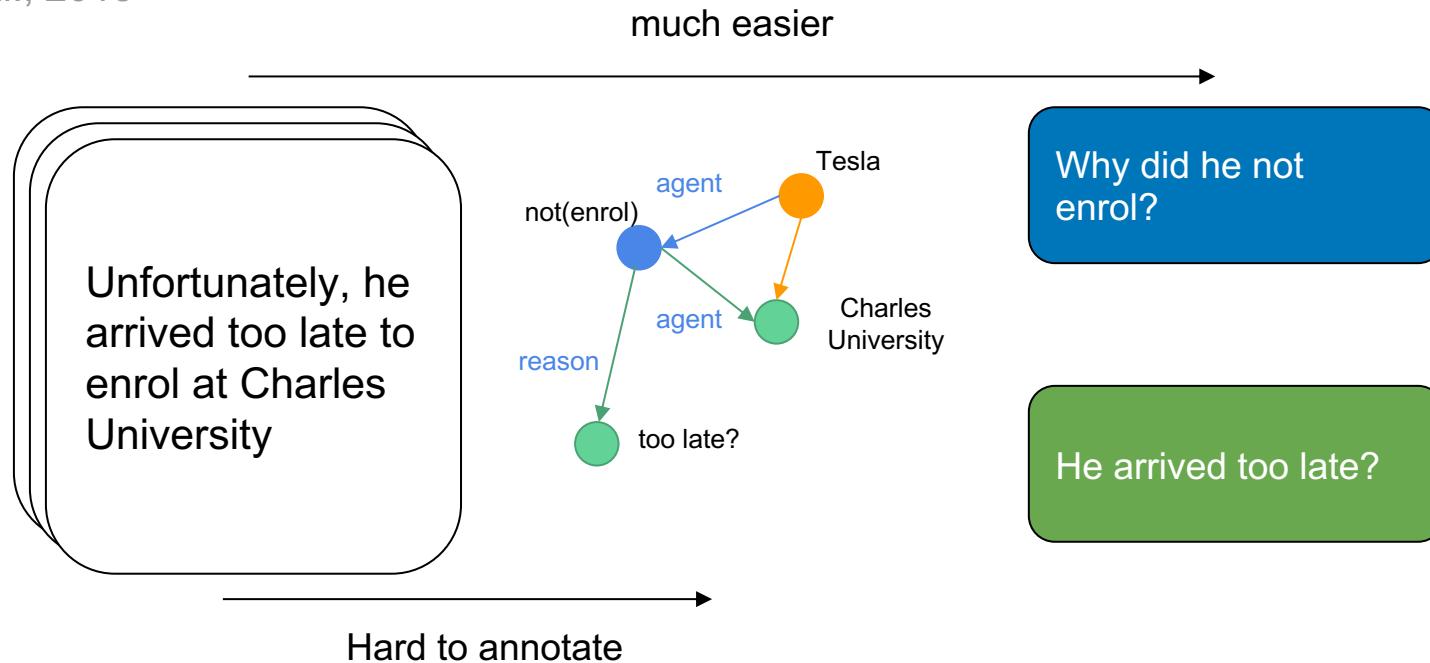
Why did he not enrol?

He arrived too late?

getting this right is hard

Weakness: Annotation

He et al., 2015



Structured Representations

- Advantages

- Fast access
- Scalable
- Interpretable
- Supports reasoning
- Universality of representations: independent of question

- Disadvantages

- Less robust to variation in language
- Cascading errors
- Schema engineering
- Annotation requires experts

Part 2: End to End

End to End Open Domain QA Model

Dr QA, Chen et al 2017

What temperature is a fever?



SELECT EVIDENCE

Fever

From Wikipedia, the free encyclopedia

This article is about the medical condition. For other uses, see [Fever \(disambiguation\)](#).

"Feverish" redirects here. For the Pure album, see [Pure \(Canadian band\) § Feverish](#).

Fever, also referred to as **pyrexia**, is defined as having a **temperature** above the **normal range** due to an increase in the body's temperature **set point**.^{[1][6][7]} There is not a single agreed-upon upper limit for normal temperature with sources using values between 37.2 and 38.3 °C (99.0 and 100.9 °F) in humans.^{[1][2][8]} The increase in set point triggers increased **muscle contractions** and causes a feeling of **cold** or **chills**.^[3] This results in greater heat production and efforts to conserve heat.^[4] When the set point temperature returns to normal, a person feels hot, becomes **flushed**, and may begin to **sweat**.^[4] Rarely a fever may trigger a **febrile seizure**, with this being more common in young children.^[5] Fevers do not typically go higher than 41 to 42 °C (105.8 to 107.6 °F).^[7]

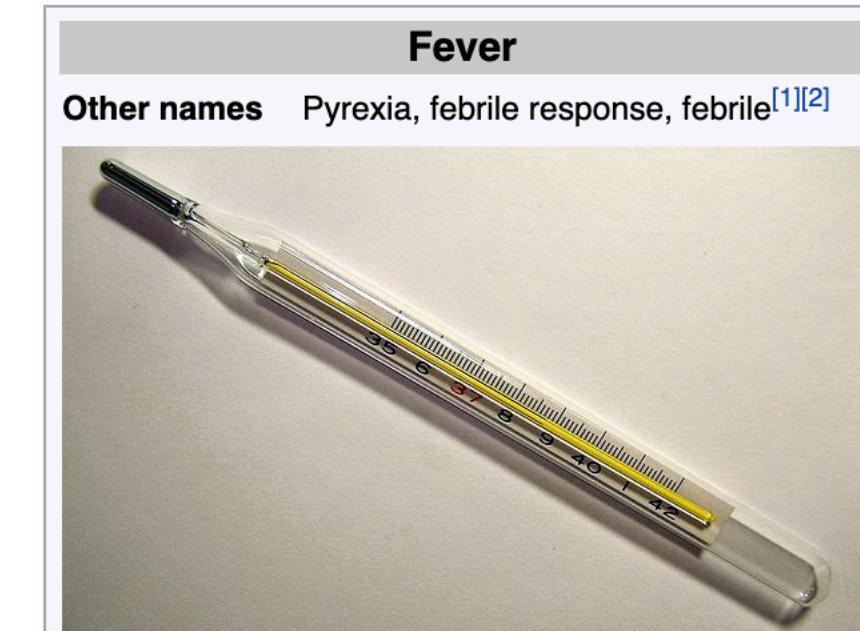
A fever can be caused by many **medical conditions** ranging from non-serious to **life-threatening**.^[12] This includes **viral**, **bacterial**, and **parasitic infections**—such as **influenza**, the **common cold**, **meningitis**, **urinary tract infections**, **appendicitis**, **COVID-19**, and **malaria**.^{[12][13]} Non-infectious causes include **vasculitis**, **deep vein thrombosis**, **connective tissue disease**, side effects of medication, and **cancer**.^{[12][14]} It differs from **hyperthermia**, in that hyperthermia is an increase in body temperature over the temperature set point, due to either too much heat production or not enough **heat loss**.^[2]

Treatment to reduce fever is generally not required.^{[3][9]} Treatment of associated pain and inflammation, however, may be useful and help a person rest.^[9] Medications such as **ibuprofen** or **paracetamol** (acetaminophen) may help with this as well as lower temperature.^{[9][10]} Measures such as putting a cool damp cloth on the forehead and having a slightly warm bath are not useful and may simply make a person more uncomfortable.^[9] Children younger than three months require medical attention, as might people with serious medical problems such as a **compromised immune system** or people with other symptoms.^[15] **Hyperthermia** does require treatment.^[3]

Fever is one of the most common **medical signs**.^[3] It is part of about 30% of healthcare visits by children^[3] and occurs in up to 75% of adults who are seriously sick.^[11] While fever evolved as a defense mechanism, treating fever does not appear to worsen outcomes.^{[16][17]} Fever is often viewed with greater concern by parents and healthcare professionals than is usually deserved, a phenomenon known as **fever phobia**.^{[3][18]}

[Contents](#) [hide]

- 1 Associated symptoms
- 2 Diagnosis

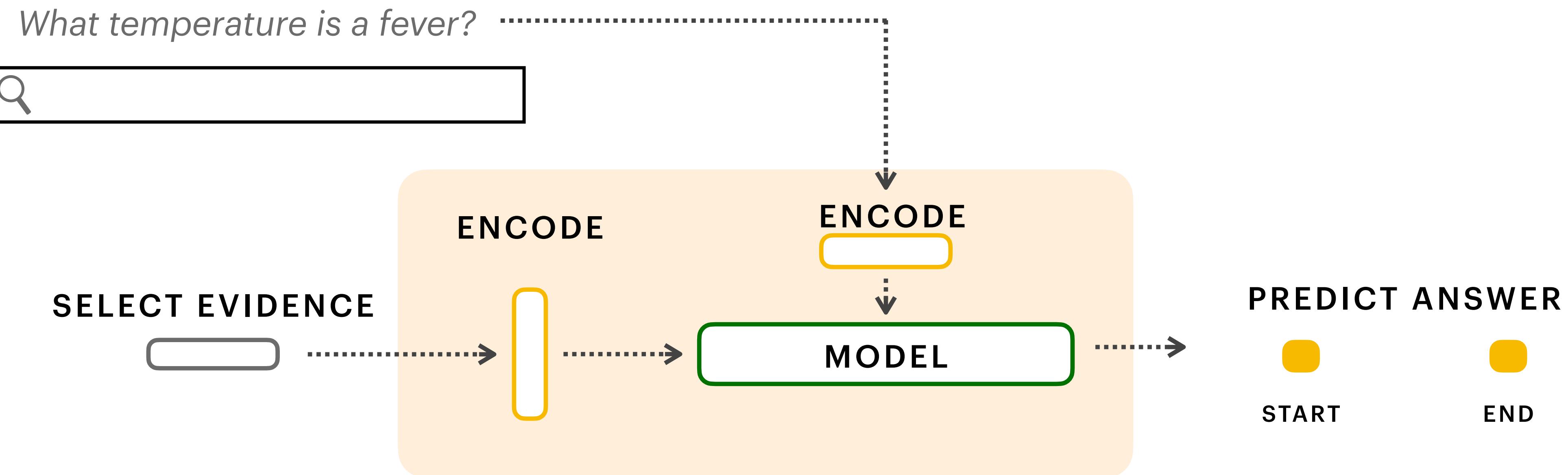


An analog **medical thermometer** showing a temperature of 38.7 °C or 101.7 °F

Specialty	Infectious disease, pediatrics
Symptoms	Initially: shivering, feeling cold, chills ^[3] Later: flushed, sweating ^[4]
Complications	Febrile seizure ^[5]
Causes	Virus, bacteria, increase in the body's temperature set point ^{[6][7]}
Diagnostic	Temperature > between 37.2 and 42 °C

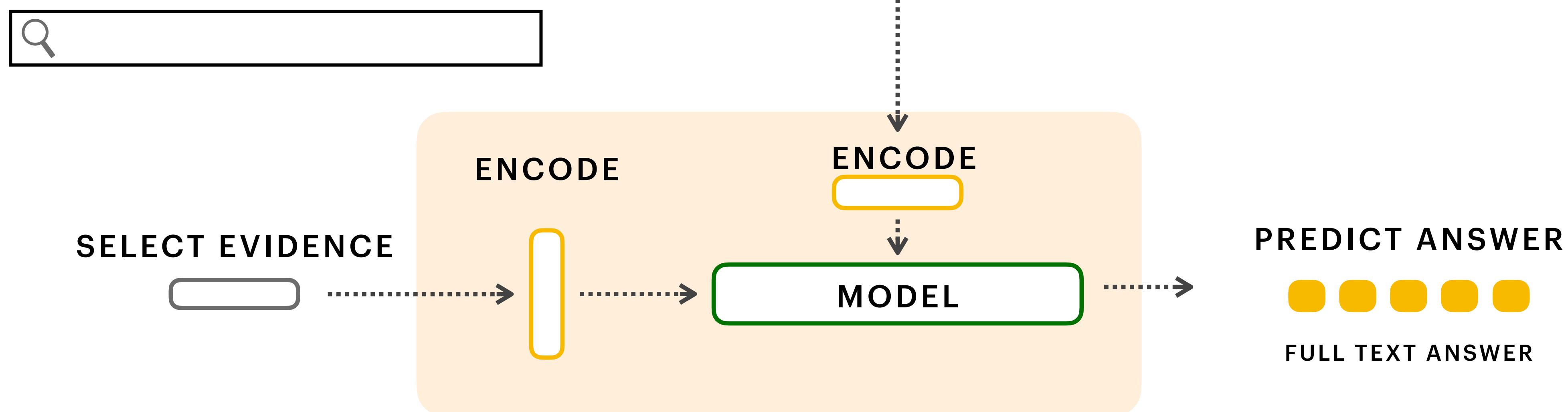
End to End Open Domain QA Model

Dr QA, Chen et al 2017

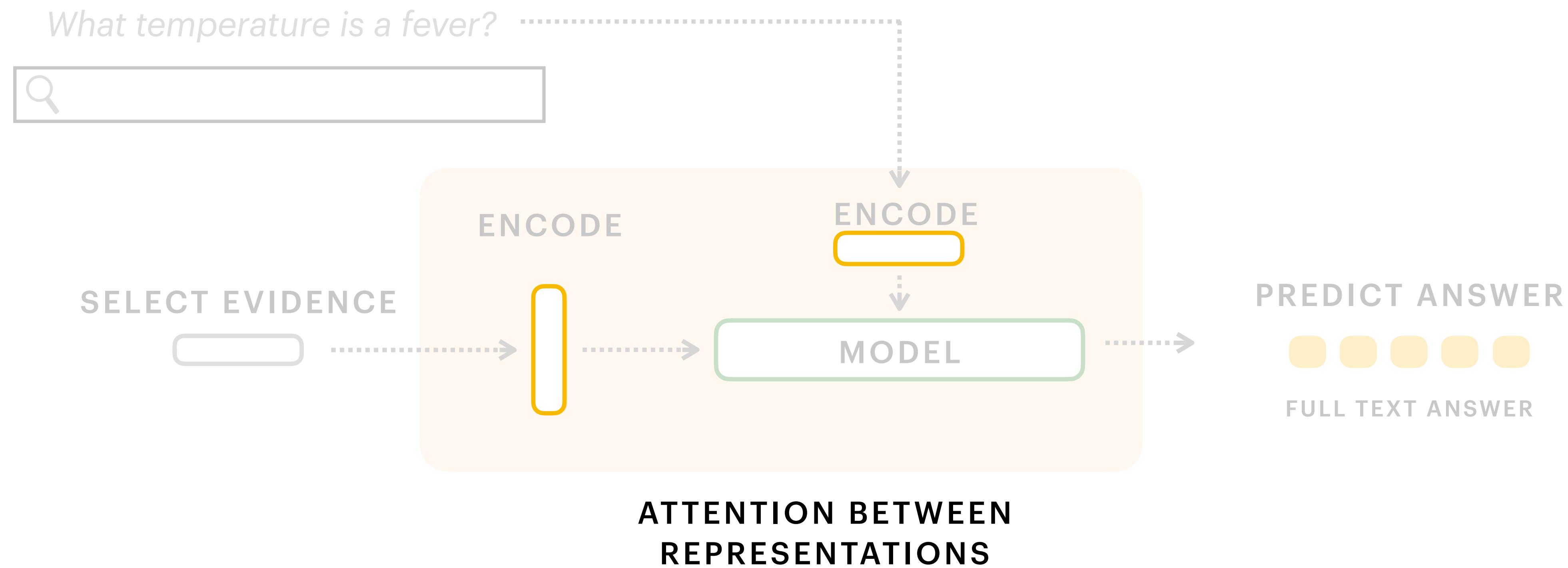


Can also be Generative

What temperature is a fever?

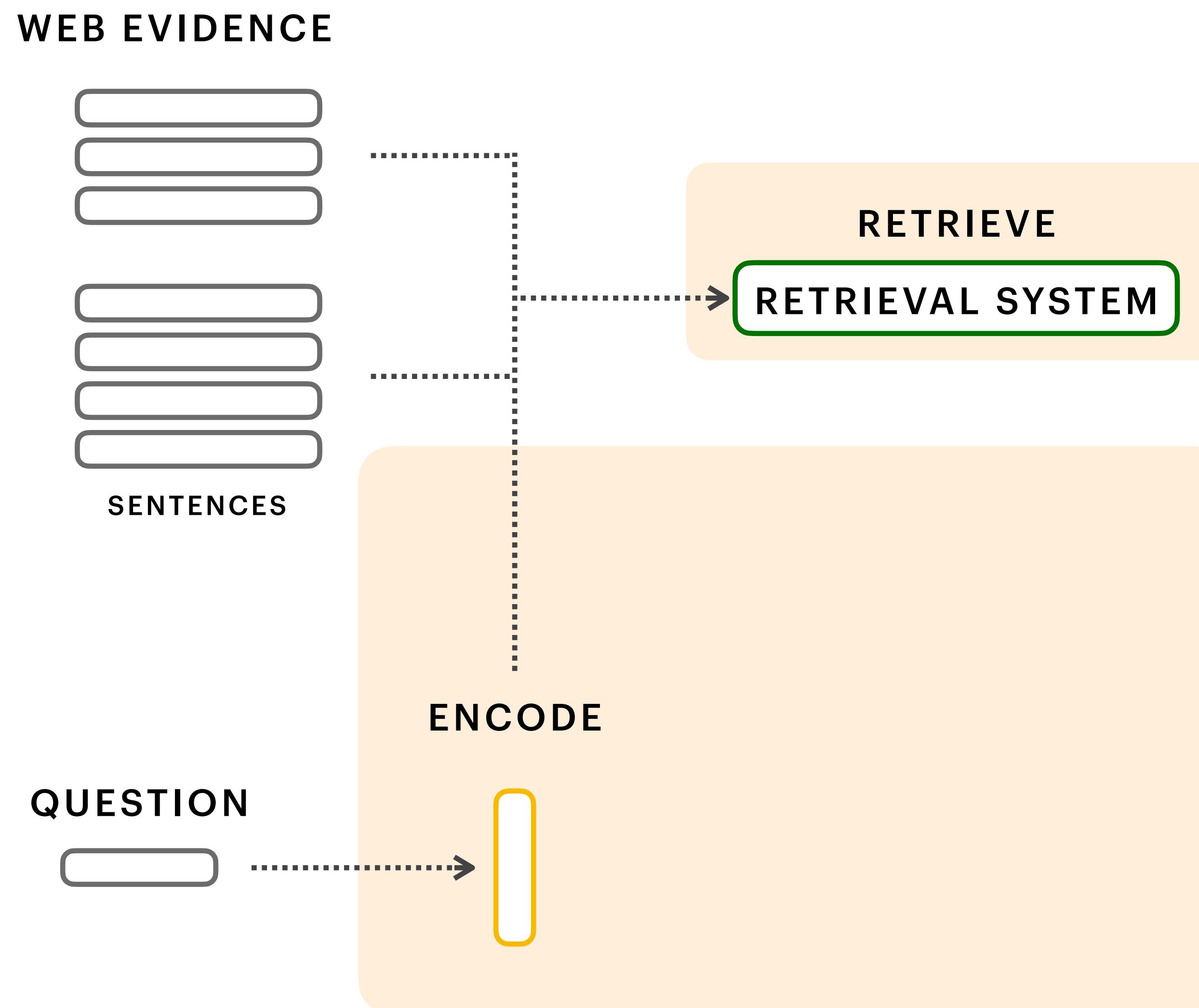


Usually model interaction between question and evidence



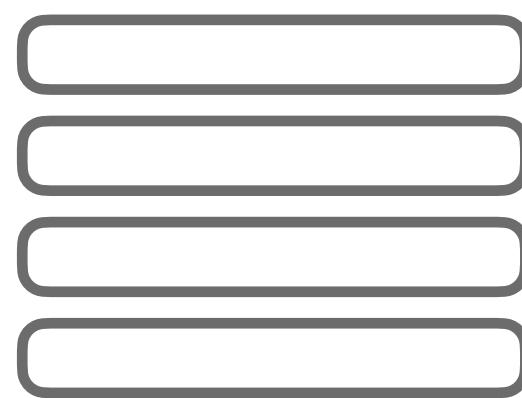
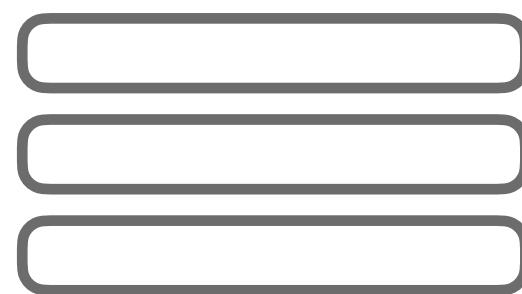
Retrieval + Generation

REALM, Guu et al
RAG, Lewis et al



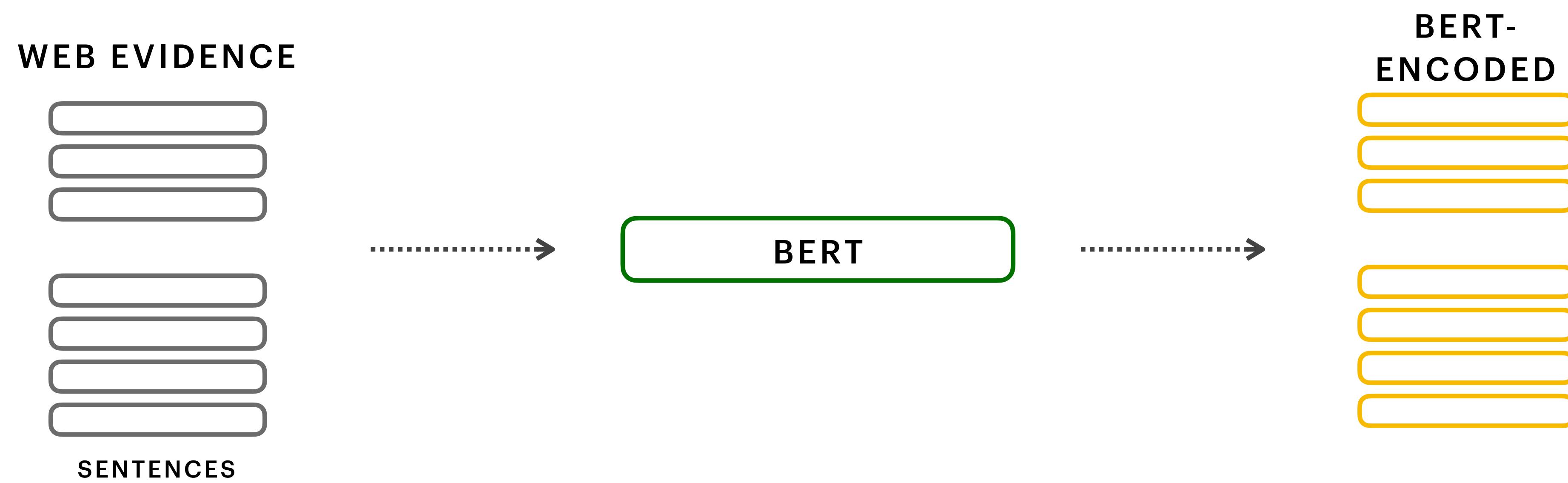
How does the Retrieval System work?

WEB EVIDENCE

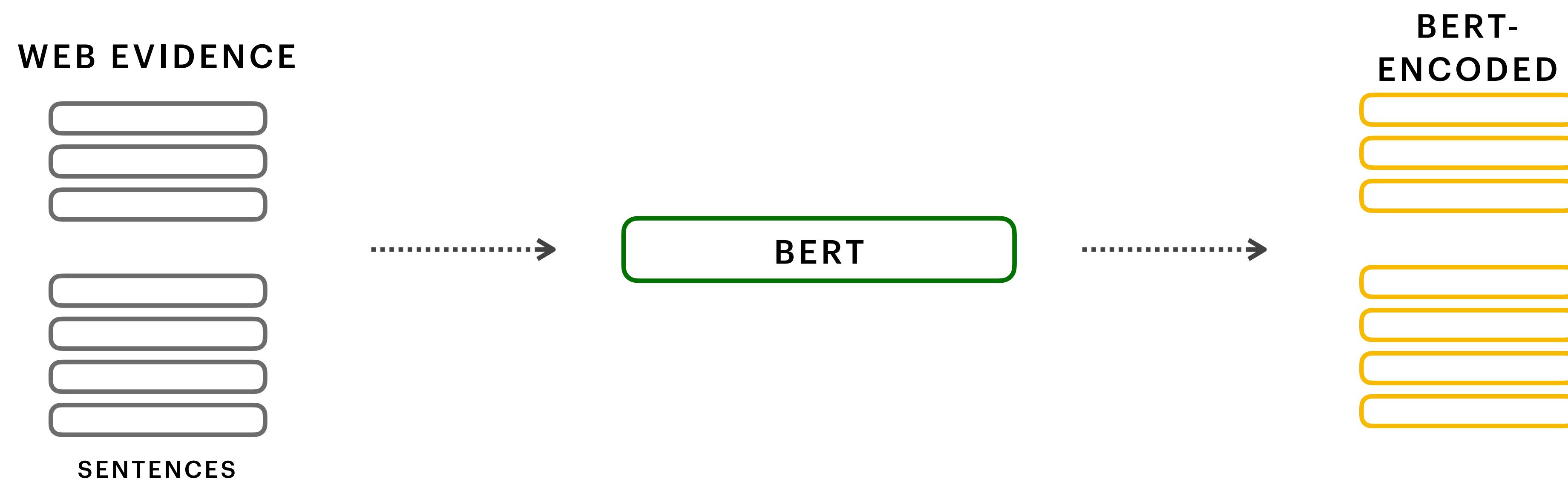


SENTENCES

How does the Retrieval System work?



Note: this can be pre-computed, if you have a fixed knowledge set



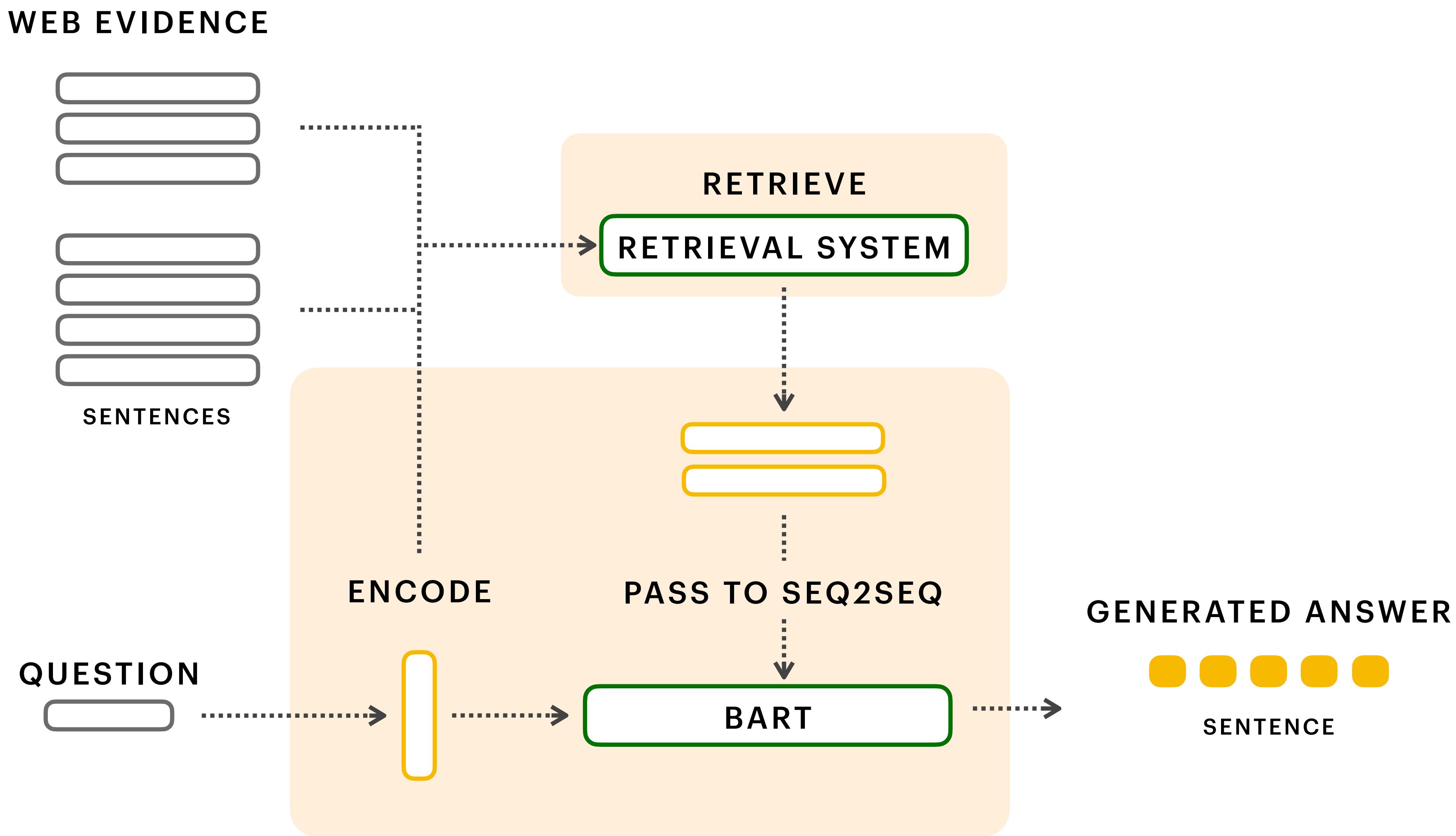
How do you actually do the retrieval?



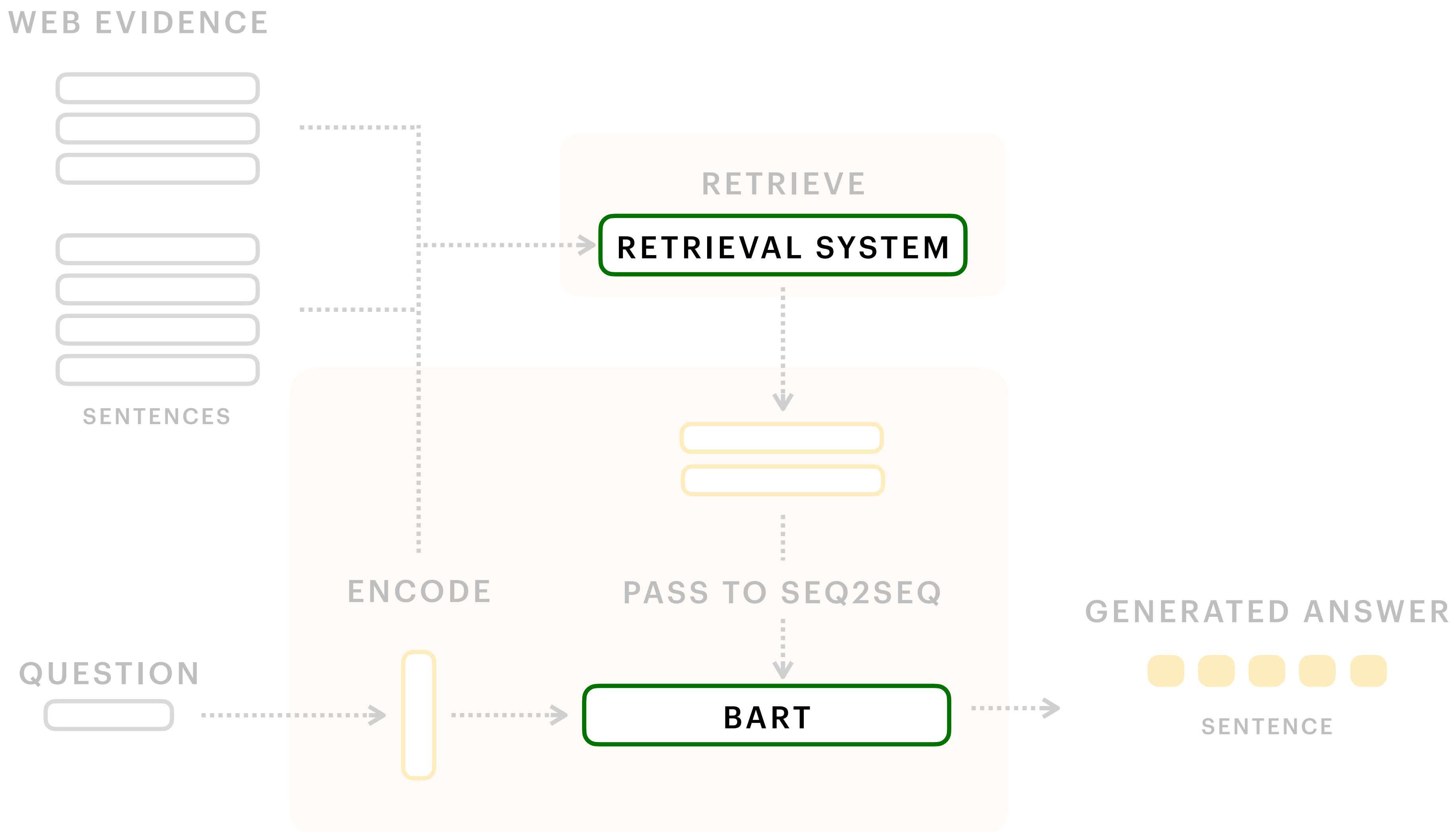
They are vectors, so you can do inner product search



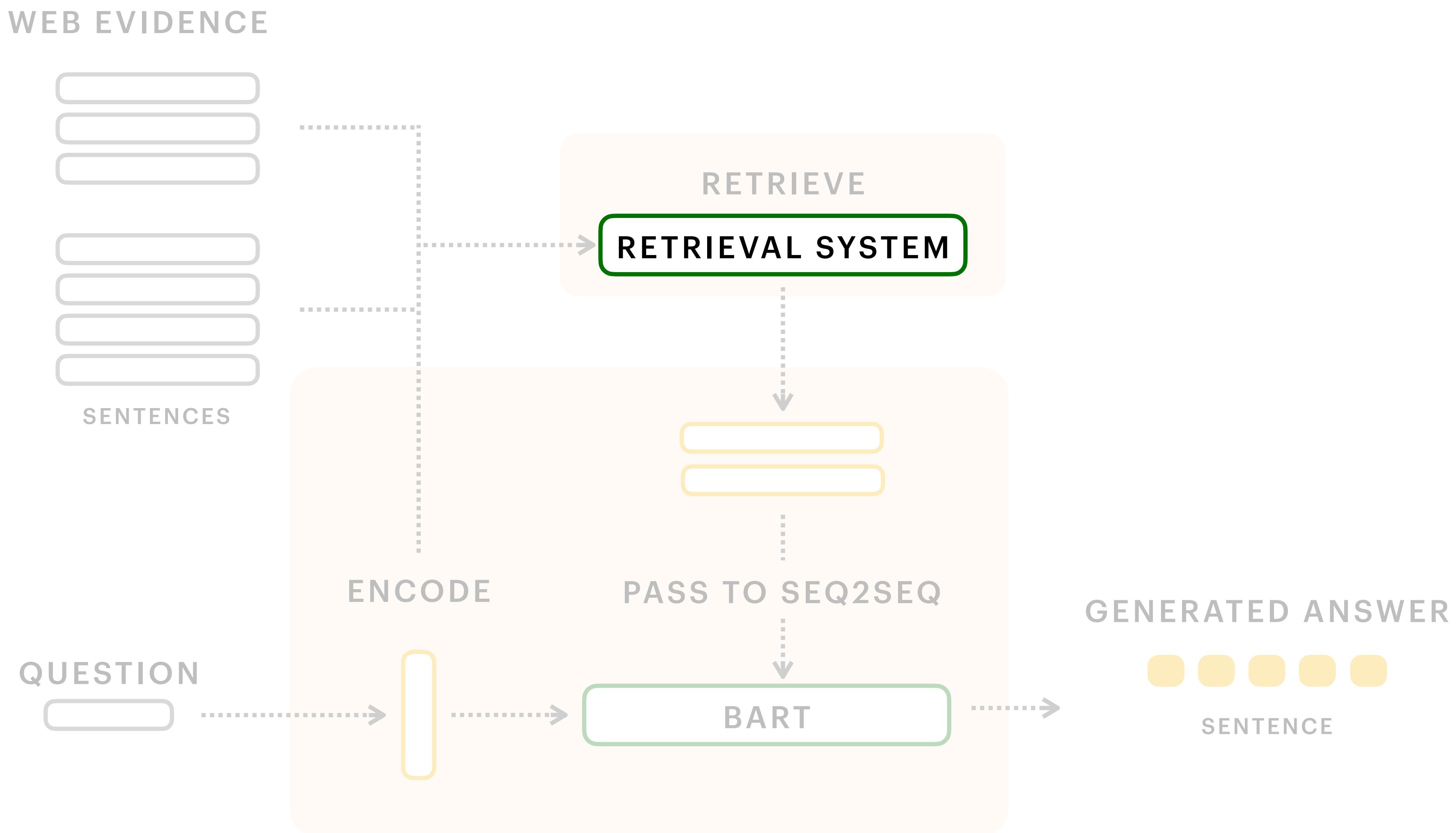
Add the Generative Part



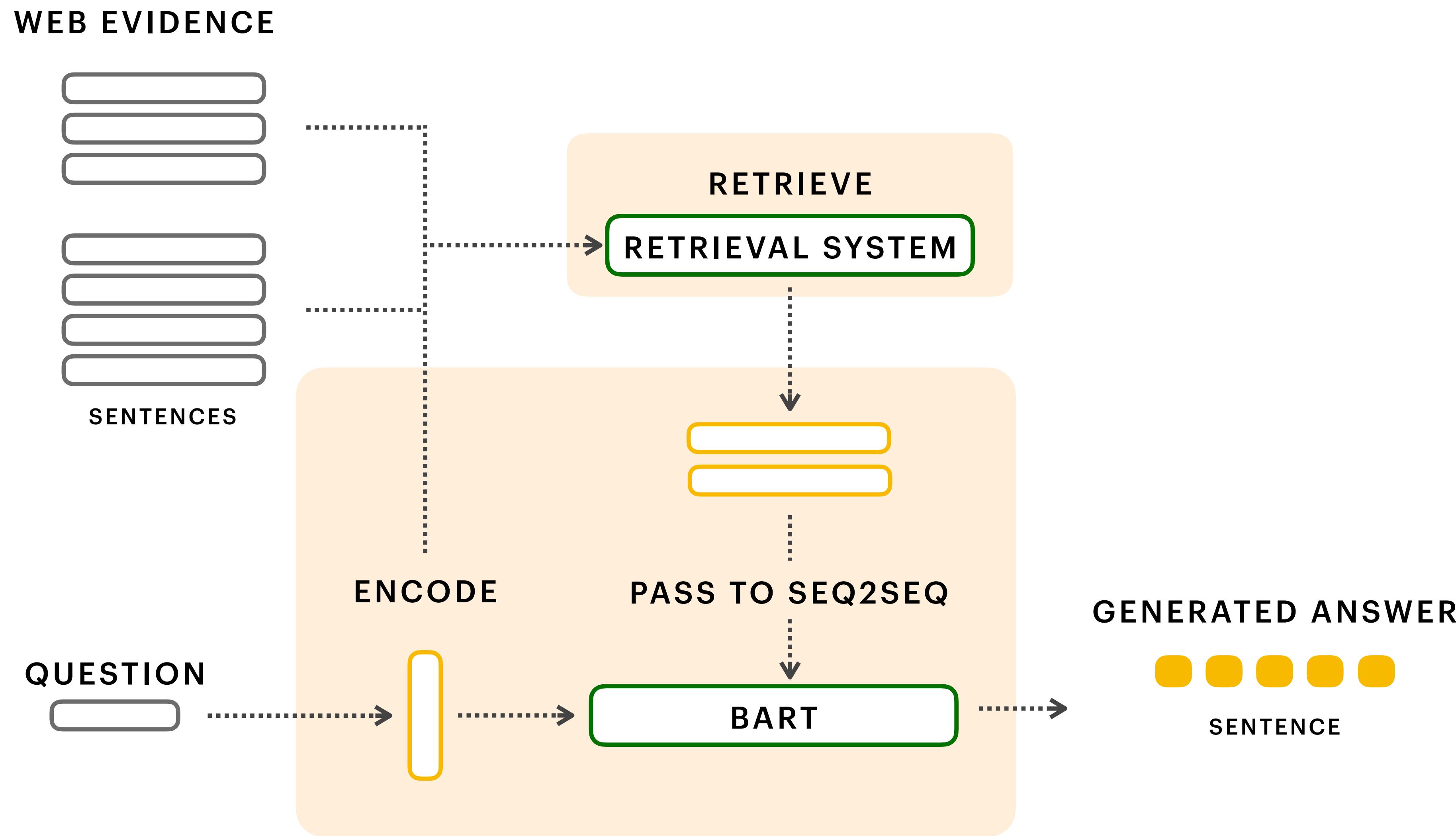
Which parts are pretrained?



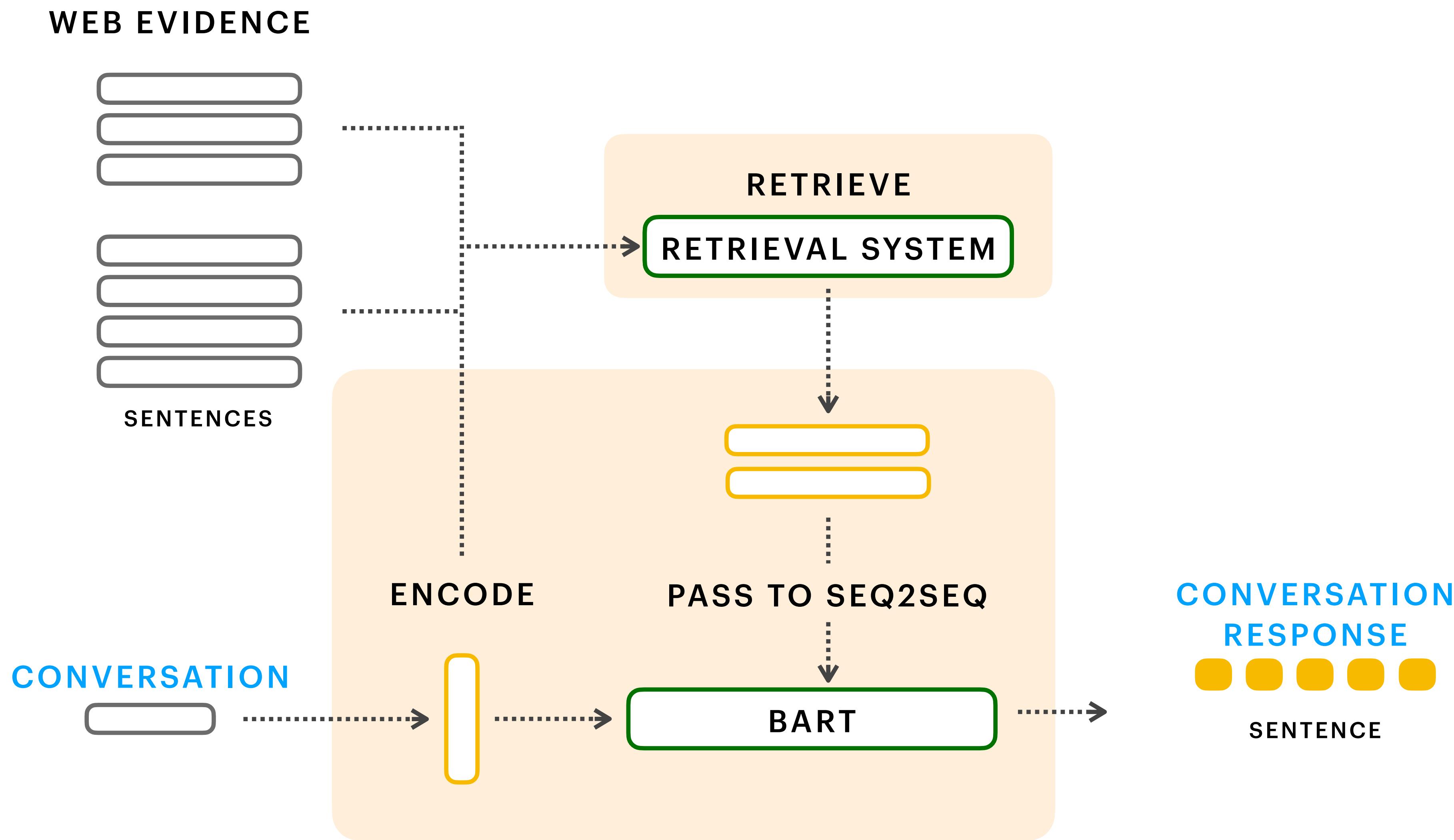
Which parts don't get backprop?

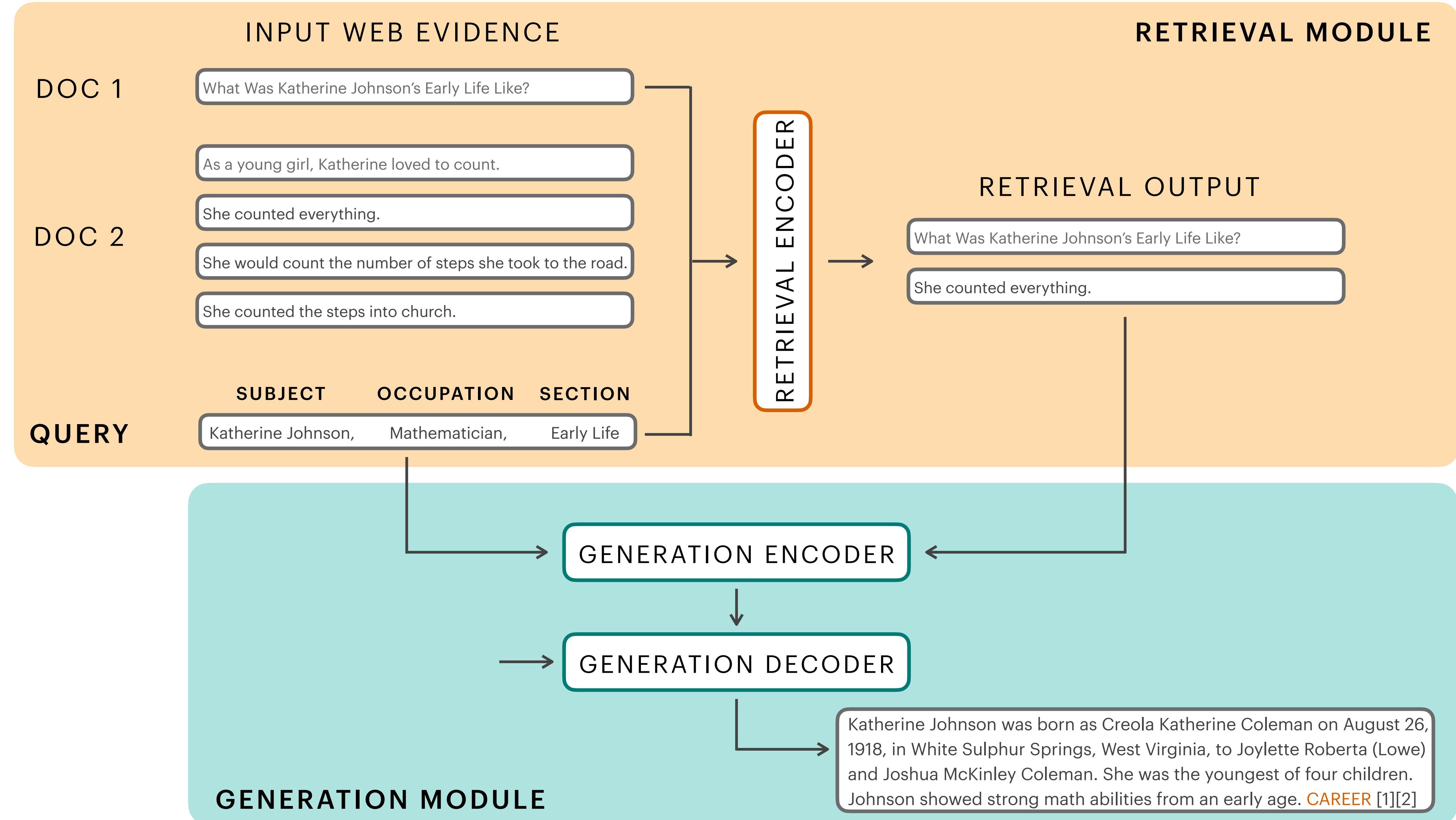


Reflection Moment: Generic Architecture

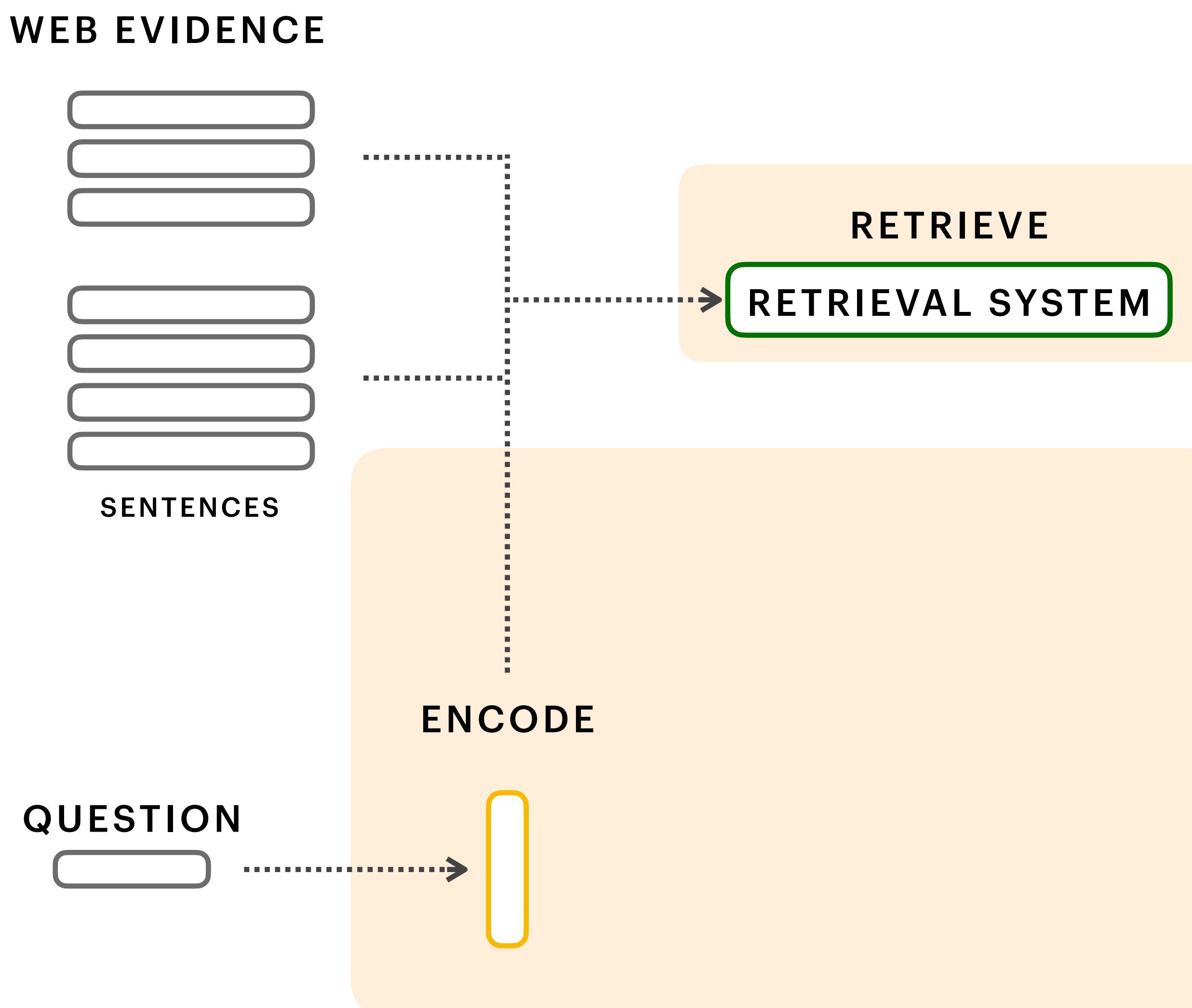


Reflection Moment: Generic Architecture





Challenges: Retrieval Part really matters

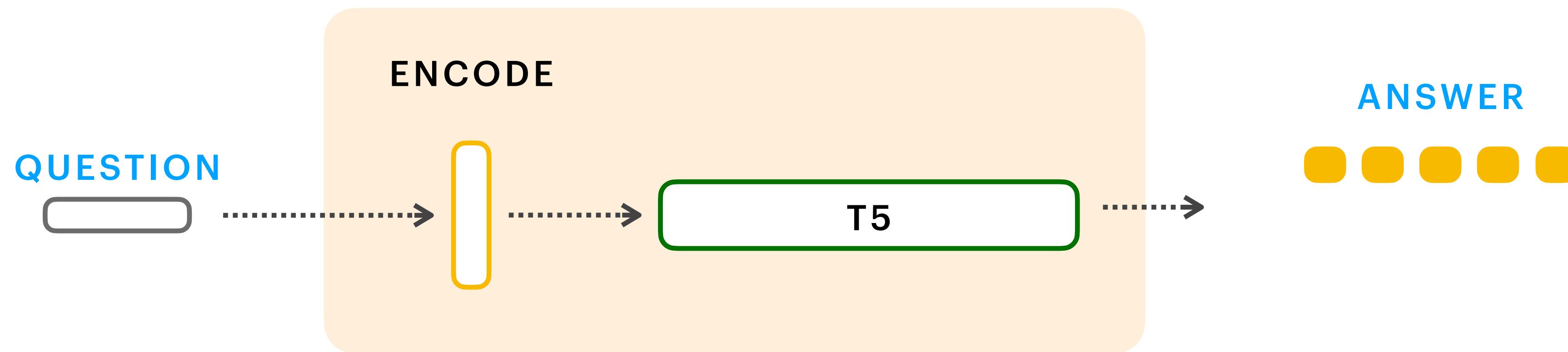


No Retrieval: Large-Scale Pretraining Only

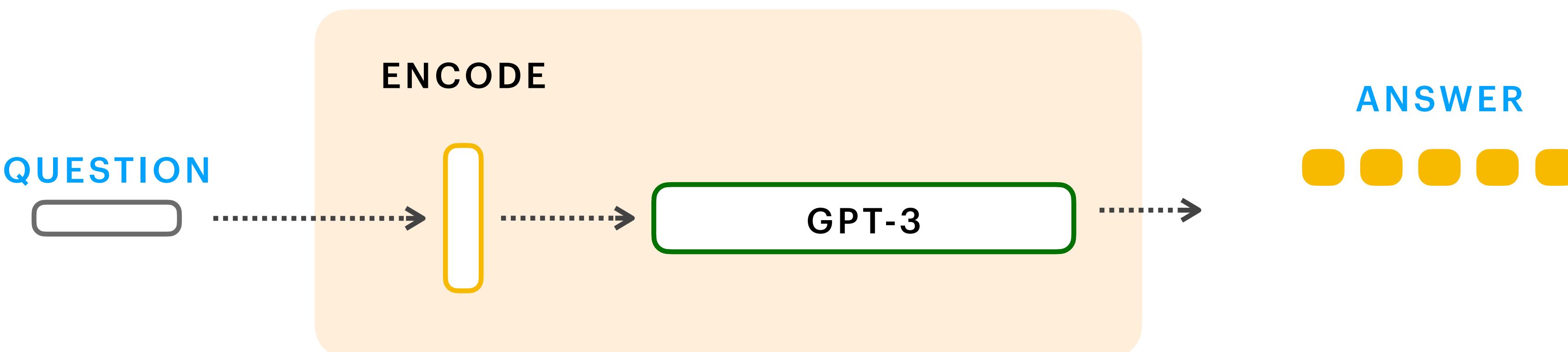
LAMA, Petroni et al 2019

Query	Answer	Generation
Francesco Bartolomeo Conti was born in ____.	Florence	Rome [-1.8], Florence [-1.8], Naples [-1.9], Milan [-2.4], Bologna [-2.5]
Adolphe Adam died in ____.	Paris	Paris [-0.5], London [-3.5], Vienna [-3.6], Berlin [-3.8], Brussels [-4.0]
English bulldog is a subclass of ____.	dog	dogs [-0.3], breeds [-2.2], dog [-2.4], cattle [-4.3], sheep [-4.5]
The official language of Mauritius is ____.	English	English [-0.6], French [-0.9], Arabic [-6.2], Tamil [-6.7], Malayalam [-7.0]
Patrick Oboya plays in ____ position.	midfielder	centre [-2.0], center [-2.2], midfielder [-2.4], forward [-2.4], midfield [-2.7]
Hamburg Airport is named after ____.	Hamburg	Hess [-7.0], Hermann [-7.1], Schmidt [-7.1], Hamburg [-7.5], Ludwig [-7.5]
The original language of Mon oncle Benjamin is ____.	French	French [-0.2], Breton [-3.3], English [-3.8], Dutch [-4.2], German [-4.9]
Dani Alves plays with ____ .	Barcelona	Santos [-2.4], Porto [-2.5], Sporting [-3.1], Brazil [-3.3], Portugal [-3.7]
Paul Toungui is a ____ by profession .	politician	lawyer [-1.1], journalist [-2.4], teacher [-2.7], doctor [-3.0], physician [-3.7]
Sodium sulfide consists of ____.	sodium	water [-1.2], sulfur [-1.7], sodium [-2.5], zinc [-2.8], salt [-2.9]
Gordon Scholes is a member of the ____ political party.	Labor	Labour [-1.3], Conservative [-1.6], Green [-2.4], Liberal [-2.9], Labor [-2.9]
Kenya maintains diplomatic relations with ____.	Uganda	India [-3.0], Uganda [-3.2], Tanzania [-3.5], China [-3.6], Pakistan [-3.6]
iPod Touch is produced by ____.	Apple	Apple [-1.6], Nokia [-1.7], Sony [-2.0], Samsung [-2.6], Intel [-3.1]
Bailey Peninsula is located in ____.	Antarctica	Antarctica [-1.4], Bermuda [-2.2], Newfoundland [-2.5], Alaska [-2.7], Canada [-3.1]
JDK is developed by ____.	Oracle	IBM [-2.0], Intel [-2.3], Microsoft [-2.5], HP [-3.4], Nokia [-3.5]
Carl III used to communicate in ____.	Swedish	German [-1.6], Latin [-1.9], French [-2.4], English [-3.0], Spanish [-3.0]
Sunshine Coast, British Columbia is located in ____.	Canada	Canada [-1.2], Alberta [-2.8], Yukon [-2.9], Labrador [-3.4], Victoria [-3.4]
Pope Clement VII has the position of ____ .	pope	cardinal [-2.4], Pope [-2.5], pope [-2.6], President [-3.1], Chancellor [-3.2]
Joe Cocker is represented by music label ____.	Capitol	EMI [-2.6], BMG [-2.6], Universal [-2.8], Capitol [-3.2], Columbia [-3.3]
London Jazz Festival is located in ____.	London	London [-0.3], Greenwich [-3.2], Chelsea [-4.0], Camden [-4.6], Stratford [-4.8]
Border TV is owned by ____.	ITV	Sky [-3.1], ITV [-3.3], Global [-3.4], Frontier [-4.1], Disney [-4.3]
The native language of Mammootty is ____.	Malayalam	Malayalam [-0.2], Tamil [-2.1], Telugu [-4.8], English [-5.2], Hindi [-5.6]
The Sharon Cuneta Show was created in ____.	Philippines	Manila [-3.2], Philippines [-3.6], February [-3.7], December [-3.8], Argentina [-4.0]

Question Answering with Large-Scale Pretraining



Question Answering with GPT-3



Context → Q: What school did burne hogarth establish?

A:

Target Completion → School of Visual Arts

Downsides of Knowledge Base type Systems

- Small scale conditions
- Requires expert knowledge
- Brittle with noisy, ambiguous data
- Not much applicative success

Benefits of End to End Systems

- Can scale to very large datasets
- Can be used and designed by non domain experts
- Robust to noise and ambiguity
- Game Changers in multiple applications

Downsides of End to End Systems

- Very data hungry
- Not interpretable
- Often large, slow...
- Can be very vulnerable to retrieval quality

Part 3: Challenges in Question Answering

Challenge 1: Complexity of Questions

Standard Question Answering Datasets



QUESTION

What temperature is a fever?



CONTEXT

Although a fever (pyrexia) could be considered any body temperature above the normal temperature of 98.6 degrees Fahrenheit (98.6 F or 37 C), medically, a person is not considered to have a significant fever until the temperature is above 100.4 F (38.0 C).

Standard Question Answering Datasets



QUESTION

What temperature is a fever?



CONTEXT

Although a fever (pyrexia) could be considered any body temperature above the normal temperature of 98.6 degrees Fahrenheit (98.6 F or 37 C), medically, a person is not considered to have a significant fever until the temperature is above **100.4** F (38.0 C).

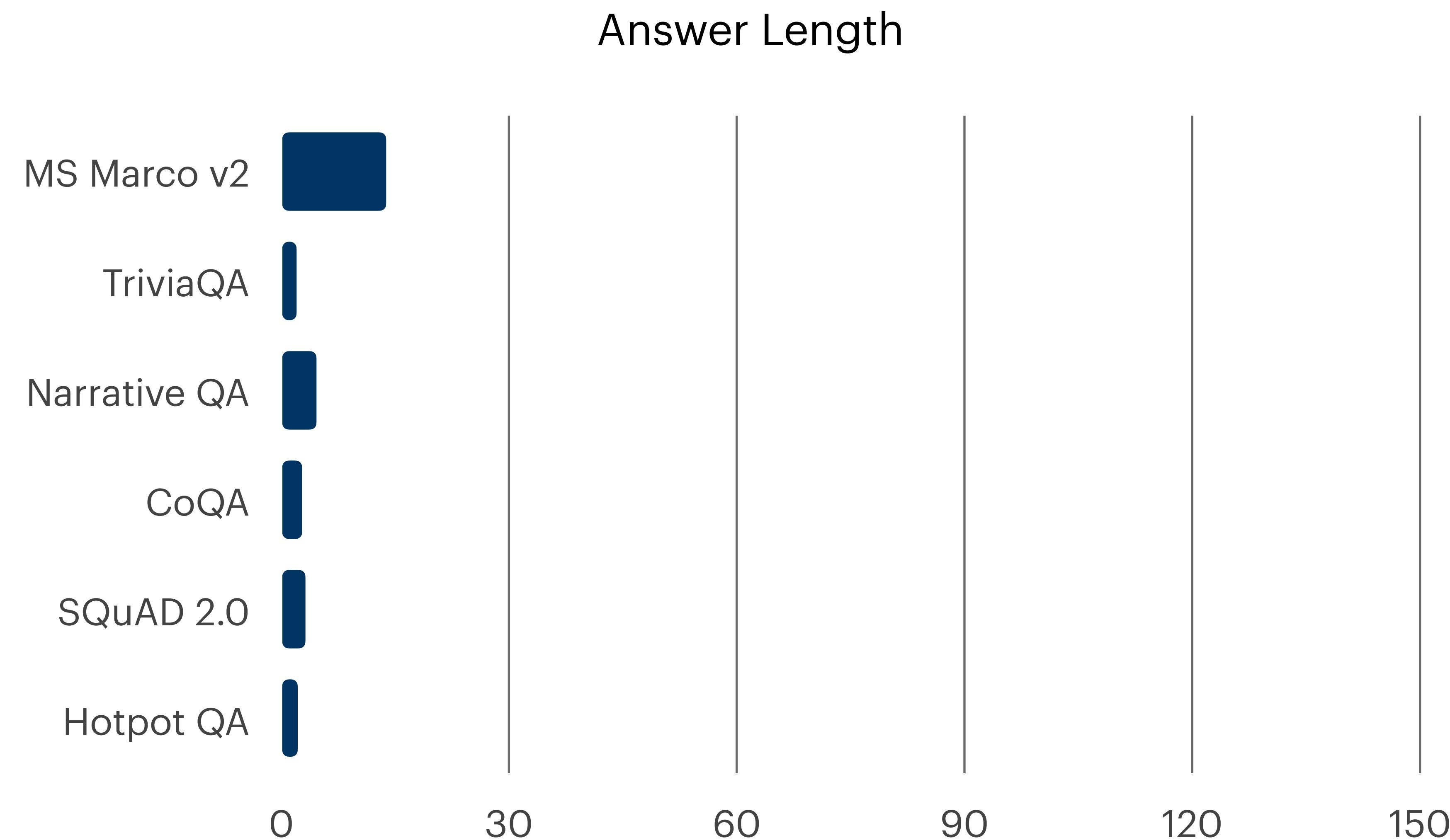


ANSWER

100.4



(some) Current QA Datasets



Why is moving beyond short answers difficult?

- Lack of Data

Why is moving beyond short answers difficult?

- Lack of Data
- How can we Generate Fluent Output

Why is moving beyond short answers difficult?

- Lack of Data
- How can we Generate Fluent Output
- Evaluation

What causes a fever?

Why are fevers dangerous?

Why do you need to bring your temperature down?

Long Form Question Answering



QUESTION

Why do you need to bring your temperature down?



ANSWER

Up to a point, having a fever is a good thing when you're fighting an infection as in the case of sepsis (infection in the blood). Many pathogens don't fare well in even a degree or two of average raised temperature, while your body is much more resilient. It's still a pretty serious condition on its own, and sepsis is frequently fatal regardless of the not only the body's attempts to fight it, but with medical intervention.

The problems in general however, start when the fever is too high, or just high for too long. Your body will release something called chaperone molecules that help your proteins fold correctly, but there will still be errors and it's more energetically expensive. This chaperone molecules also have limits, and past a certain point your body fails on a number of levels.

Explain Like I'm Five Dataset

↑ E r/explainlikeimfive

90
↓

ELI5: why do you need to bring your temperature down?

I read somewhere that the reason our temp rises when we are sick is to help out immune system. if so, why try to lower it

28 Comments

SORT BY BEST ▾

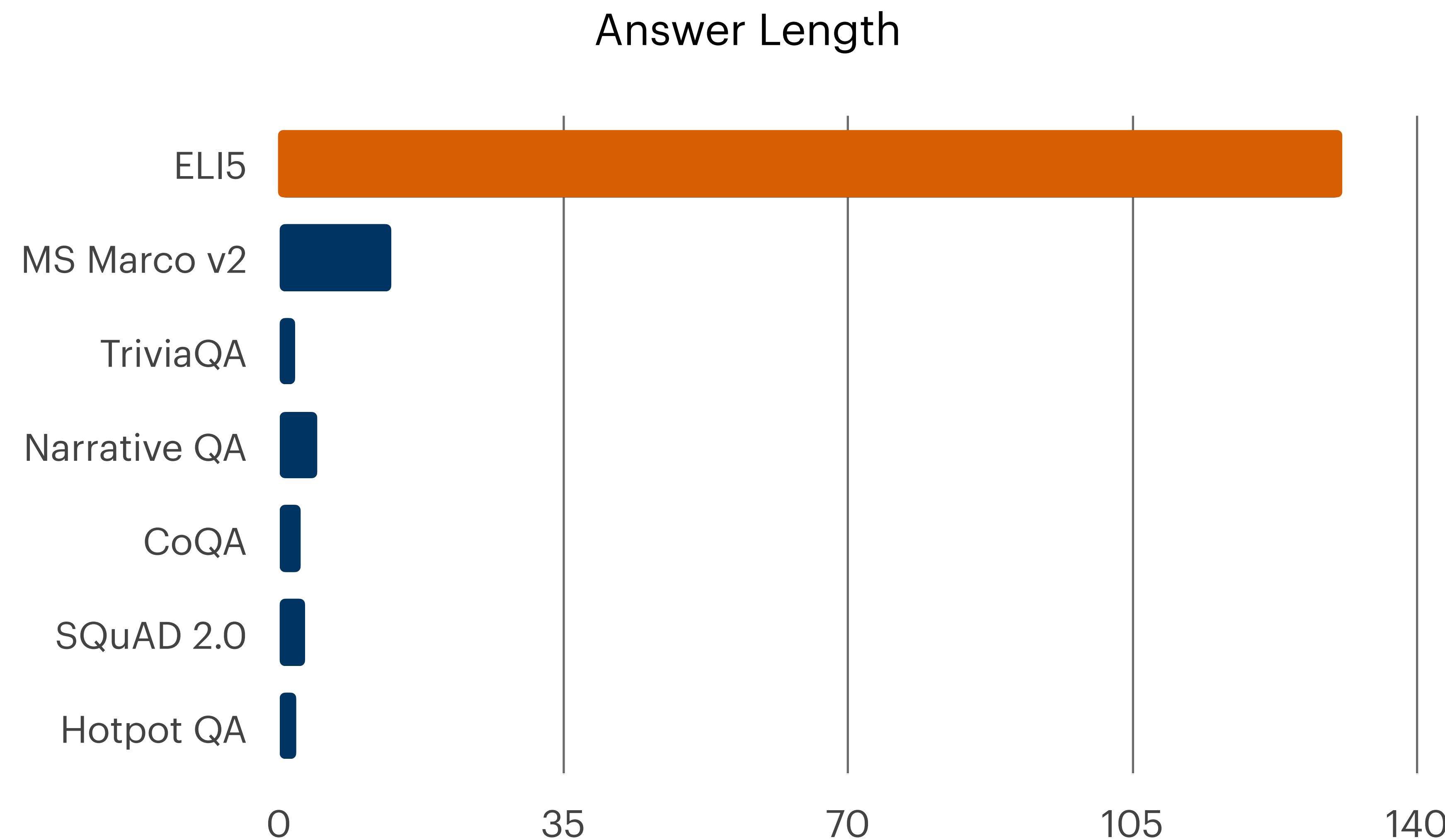
↑
↓

Up to a point, having a fever is a good thing when you're fighting an infection as in the case of sepsis (infection in the blood). Many pathogens don't fare well in even a degree or two of average raised temperature, while your body is much more resilient. It's still a pretty serious condition on its own, and sepsis is frequently fatal regardless of the not only the body's attempts to fight it, but with medical intervention.

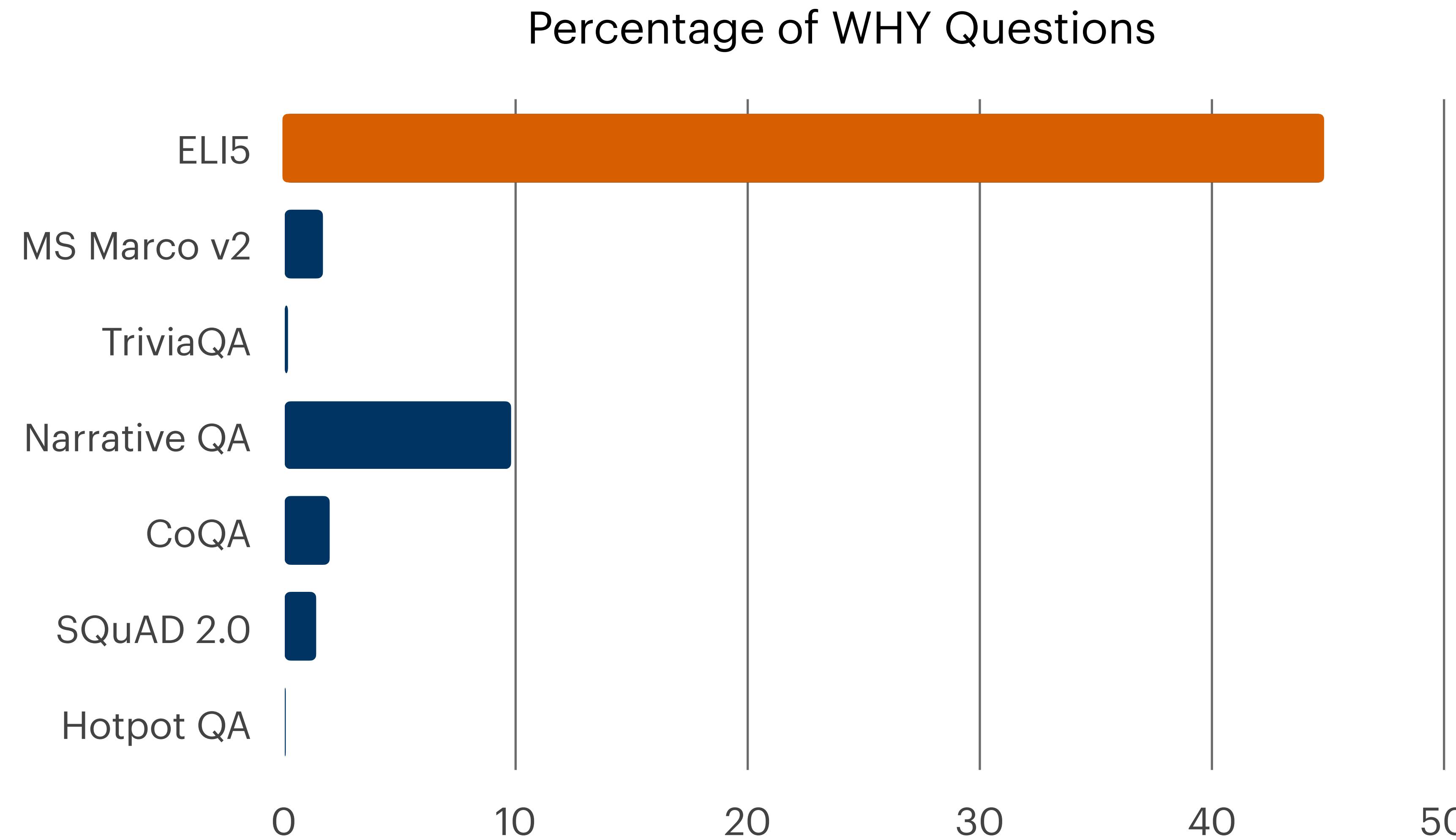
The problems in general however, start when the fever is too high, or just high for too long. Your body will release something called chaperone molecules that help your proteins fold correctly, but there will still be errors and it's more energetically expensive. These chaperone molecules also have limits, and past a certain point your body fails on a number of levels.

For one, a lot of what your cells do is interact with, transport, and produce proteins. The function of a protein is determined by its three dimensional structure, and it gets that through a process of folding. This is a process which can go wrong, and heat makes it far more likely to go wrong. Past a certain point critical proteins will start to unfold (denature) as in exposure to cooking methods. Needless to say, this does you no favors.

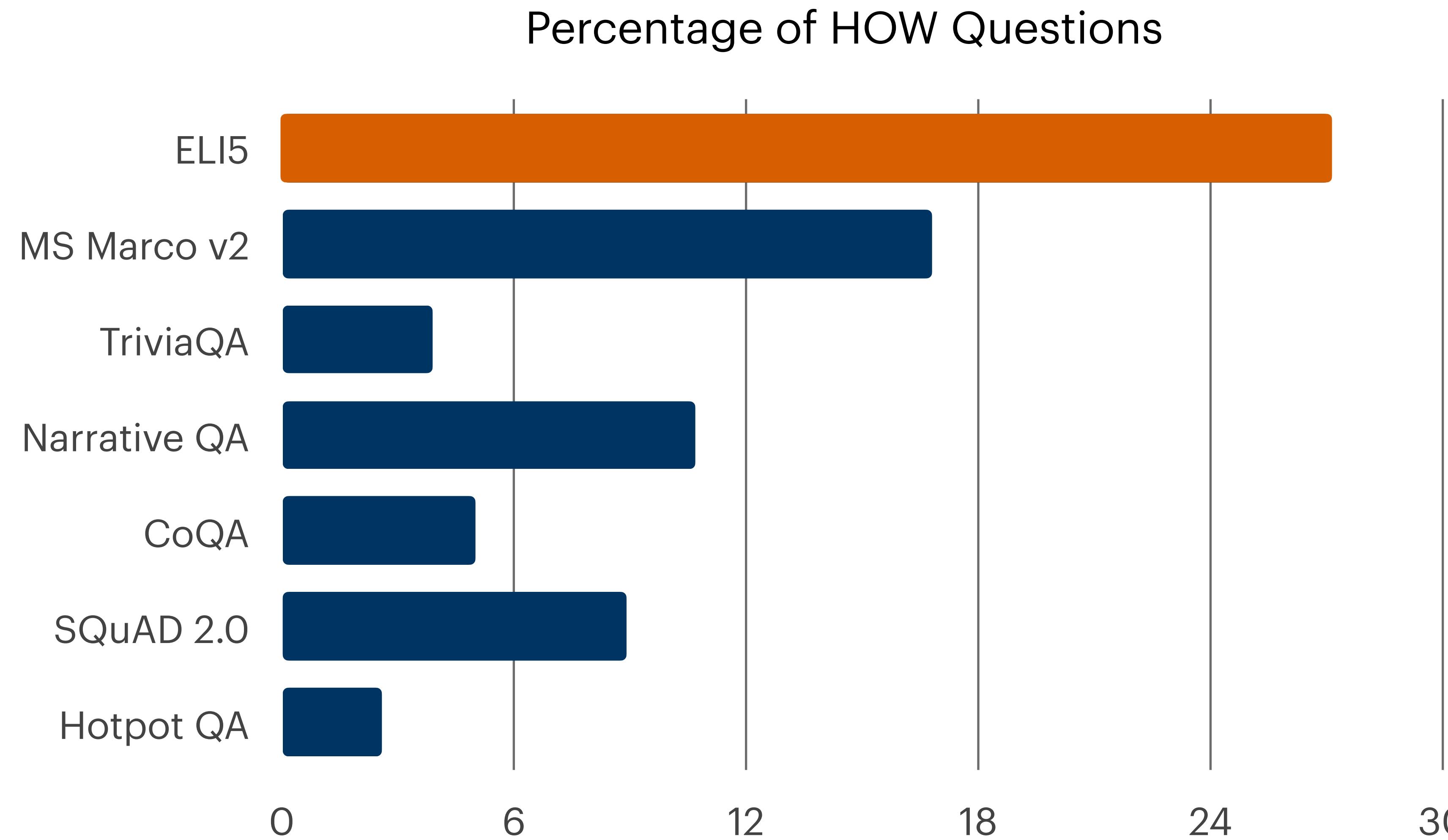
Explain Like I'm Five: Comparison



Explain Like I'm Five Dataset



Explain Like I'm Five Dataset



Why are How and Why Questions Important?

- Natural, human-written questions look different than Crowdsourcing
- Challenging reasoning steps

Challenge 2: Unanswerable Questions

Many questions are unanswerable



QUESTION

Why does my husband read so much faster than I do?

Why was there an electrical failure in my apartment building on Saturday night?

Why does Louis have such a good fashion sense?

How can models deal with this?

- “Unsure” response

How can models deal with this?

- “Unsure” response
- Ask clarifying questions

How can models deal with this?

- “Unsure” response
- Ask clarifying questions
- Use model confidence to understand if the question should be answered

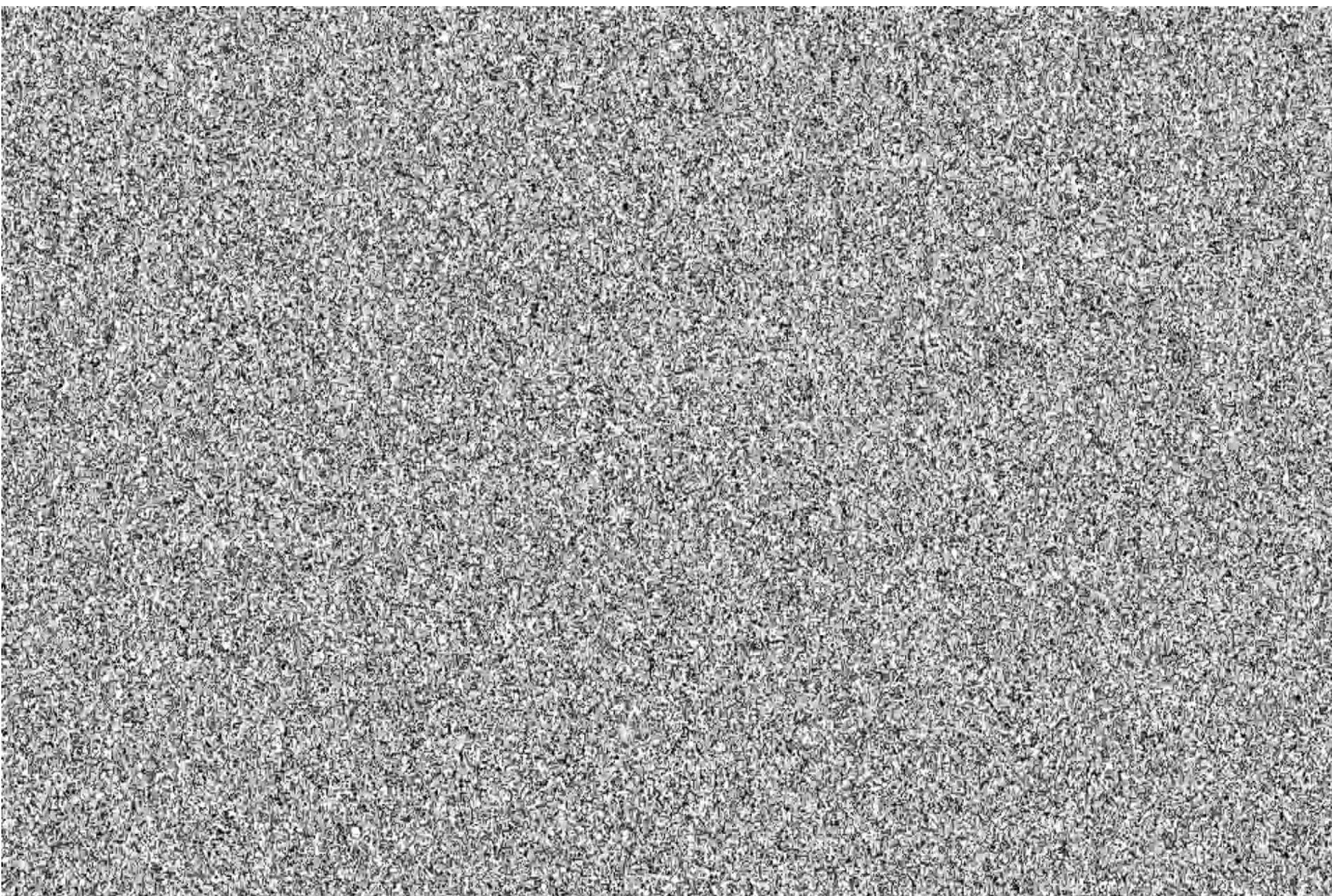
Challenge 3: Robustness

Bias in Question Answering



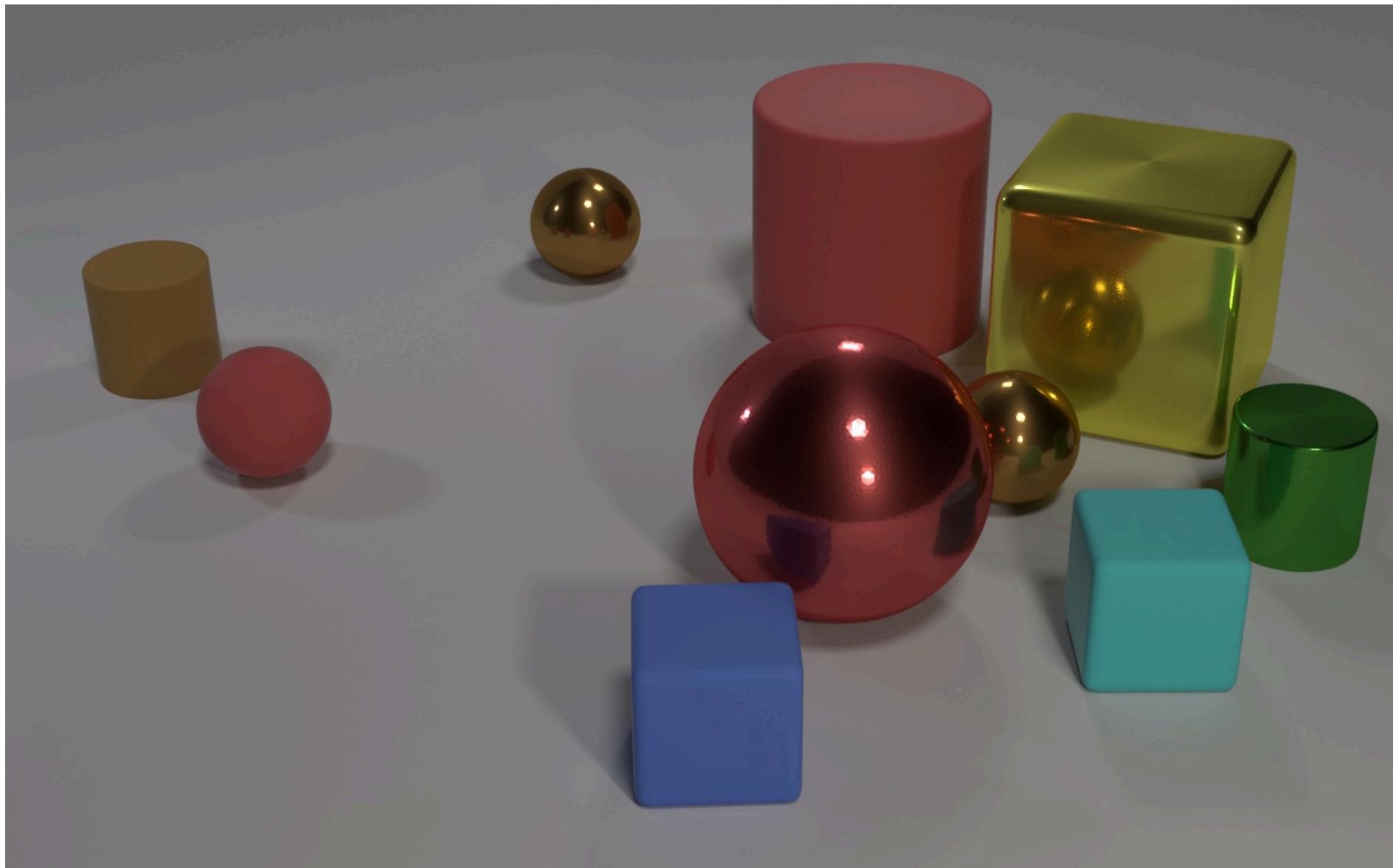
What color is the grass?

Bias in Question Answering



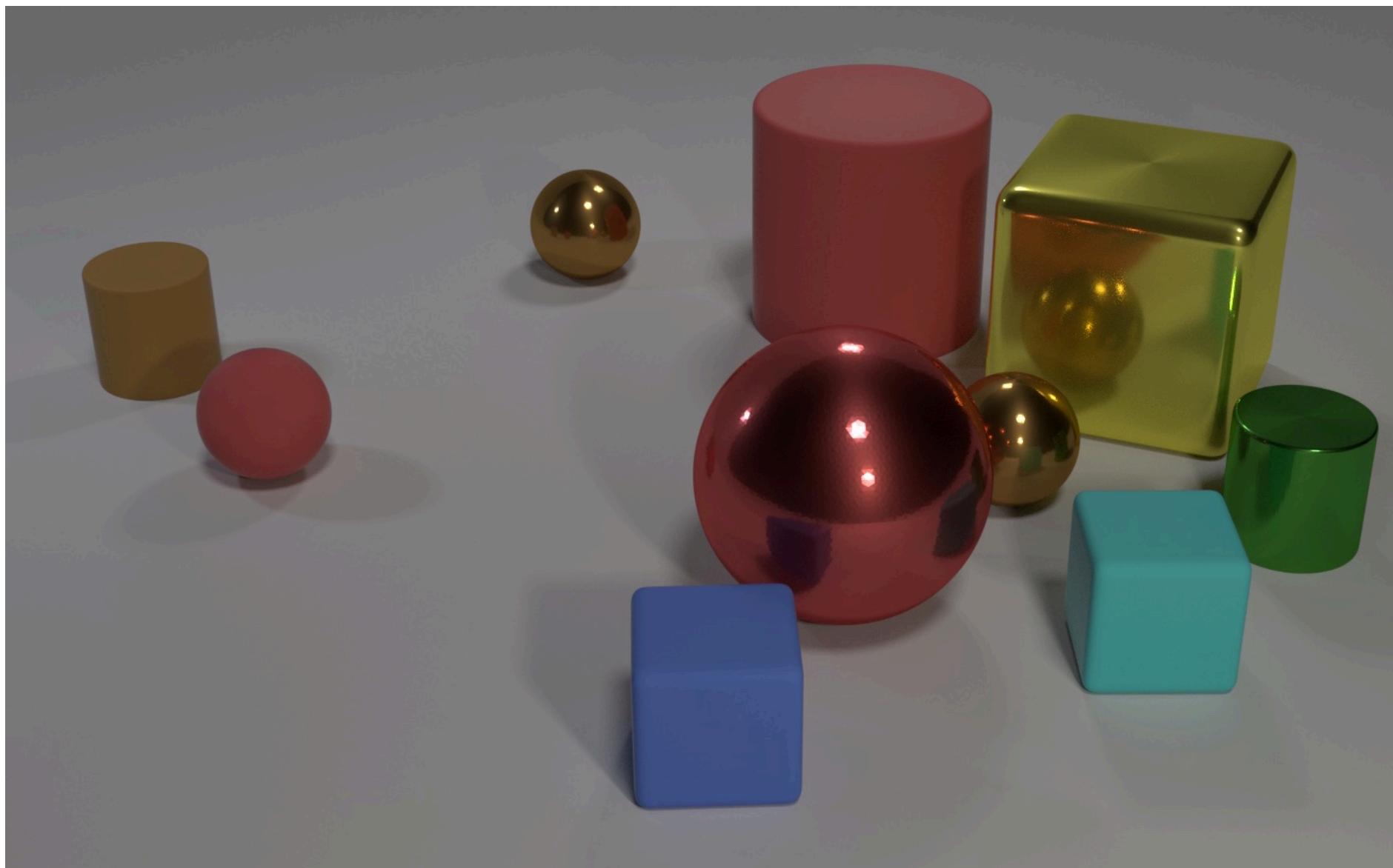
What color is the grass?

Bias in Question Answering



What color is the large sphere that is left of a blue cube that is beneath a small shiny thing that ...?

Bias in Question Answering



What color is the large sphere that is left of a blue cube that is beneath a small shiny thing that?

SQuAD Dataset



QUESTION

When did a storm hit Richard's fleet?



ANSWER

1911

SQuAD Dataset



QUESTION

When did a storm hit Richard's fleet?



CONTEXT

In April **1191** Richard the Lion-hearted left Messina with a large fleet in order to reach Acre. But a storm dispersed the fleet. After some searching, it was discovered that the boat carrying his sister and his fiancée Berengaria was anchored on the south coast of Cyprus, together with the wrecks of several other ships, including the treasure ship.



ANSWER

1191

Bias in Question Answering

In April **1191** Richard the Lion-hearted left Messina with a large fleet in order to reach Acre. But a storm dispersed the fleet. After some searching, it was discovered that the boat carrying his sister and his fiancée Berengaria was anchored on the south coast of Cyprus, together with the wrecks of several other ships, including the treasure ship.

When did a storm hit Richard's fleet?

Biased Squad: Dates

Training Time: Context with One Date

In April **1191** Richard the Lion-hearted left Messina with a large fleet in order to reach Acre. But a storm dispersed the fleet. After some searching, it was discovered that the boat carrying his sister and his fiancée Berengaria was anchored on the south coast of Cyprus, together with the wrecks of several other ships, including the treasure ship.

Inference Time: Context with Multiple Dates

In April **1191** Richard the Lion-hearted left Messina with a large fleet in order to reach Acre. But a storm dispersed the fleet. After some searching, he left in May **1192**, it was discovered that the boat carrying his sister and his fiancée Berengaria was anchored on the south coast of Cyprus, together with the wrecks of several other ships, including the treasure ship. After arriving back in Europe in **1193**, ...

Adversarial Squad

Inference Time: Adversarial Sentence Insertion

What we now call gravity was not identified as a universal force until the work of Isaac Newton. [...] Galileo was instrumental in describing the characteristics of falling objects [...] this acceleration due to gravity towards the surface of the Earth is usually designated as and has a magnitude of about 9.81 meters per second squared [...], and points toward the center of the Earth. [...] Distractor: **Object falls about 5 times faster on Mars.**

Challenge 4: Dealing with Information at Scale

Challenges when models read the web

- What if different pieces of information disagree?

Challenges when models read the web

- What if different pieces of information disagree?
- Which sources of information are reliable?

Challenges when models read the web

- What if different pieces of information disagree?
- Which sources of information are reliable?
- Are the sources biased?

Challenges when models read the web

- What if different pieces of information disagree?
- Which sources of information are reliable?
- Are the sources biased?
- Are they representative?

Challenges when models read the web

- What if different pieces of information disagree?
- Which sources of information are reliable?
- Are the sources biased?
- Are they representative?

Challenge 5: Multilinguality

Multilingual Question Answering

- Not a lot of work...

Challenge 6: Explainability

How did we get this answer???



QUESTION

What temperature is a fever?



CONTEXT

Although a fever (pyrexia) could be considered any body temperature above the normal temperature of 98.6 degrees Fahrenheit (98.6 F or 37 C), medically, a person is not considered to have a significant fever until the temperature is above **100.4** F (38.0 C).



ANSWER

100.4



Summary: Where Models Work Well

- Question is answerable
- Relevant paragraph/text is given, and not too long
- Answer is not too complex
- Same domain between training and testing

Overall Thoughts: What Happened to QA in 2018-2019?

- Is your dataset testing what you want to test? Is your dataset challenging enough?

Overall Thoughts: What Happened to QA in 2018-2019?

- Is your dataset testing what you want to test? Is your dataset challenging enough?
- What is the strongest baseline possible? How far do simple baselines go?

Possible Projects in Question Answering

- Question Answering in other languages
- Analysis of how well pretrained models are doing via prompt engineering
- Use Retrieval for your own task
- What kinds of explanations do users like?
- etc