Question answering

CS685 Fall 2020

Advanced Natural Language Processing

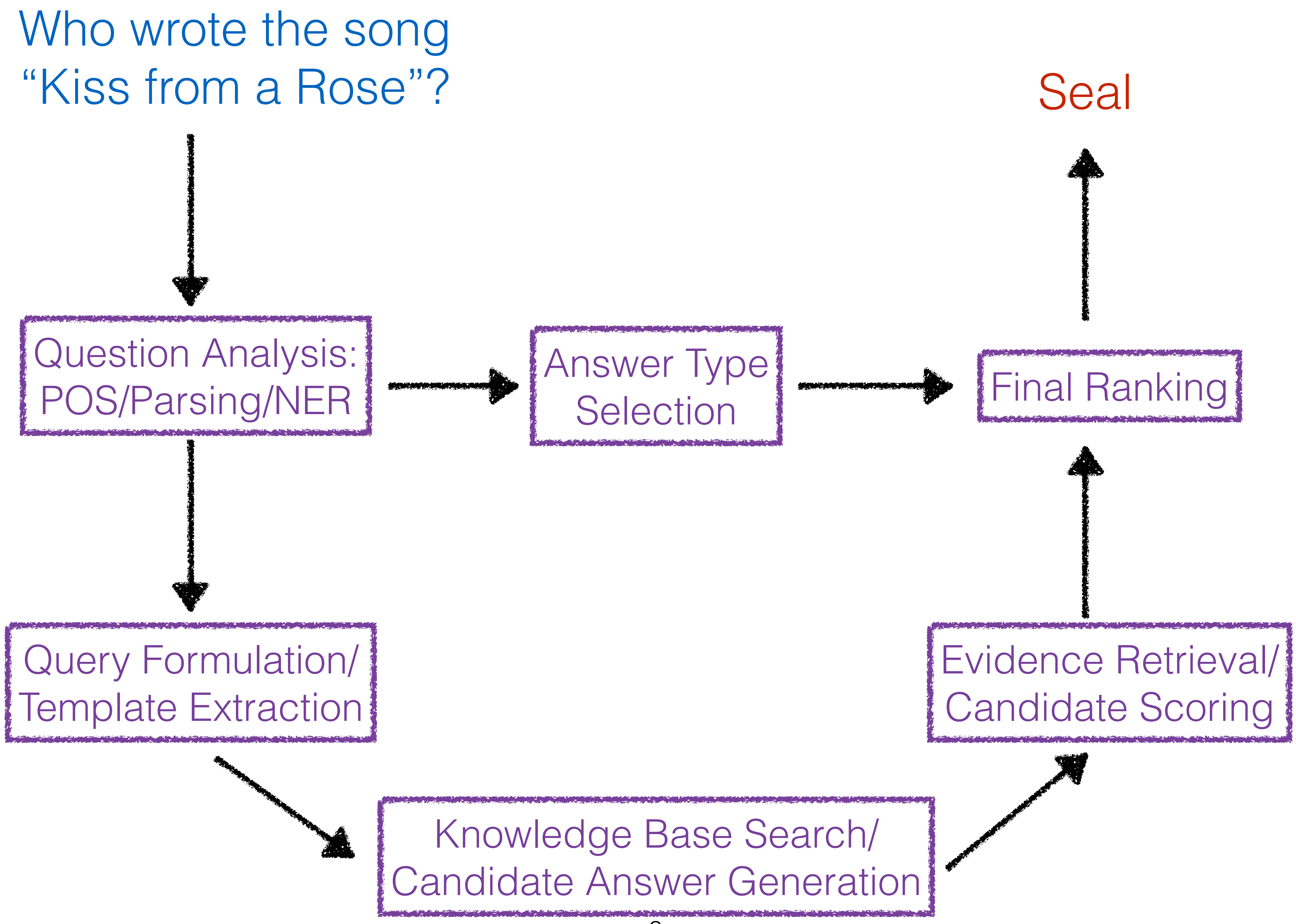
Mohit Iyyer

College of Information and Computer Sciences
University of Massachusetts Amherst

some slides from Jordan Boyd-Graber, Jacob Devlin, and Chris Manning

Stuff from last time

- HW0 grades published, good job!
- HW1 coming soon :)
- Project proposal feedback in early October
- Exam pushed back to end of October
- Thanks to whoever posted all those Notability tips in the anonymous form!



Who wrote the song “Kiss from a Rose”?

Question Analysis: POS/Parsing/NER

Answer Type Selection

Final Ranking

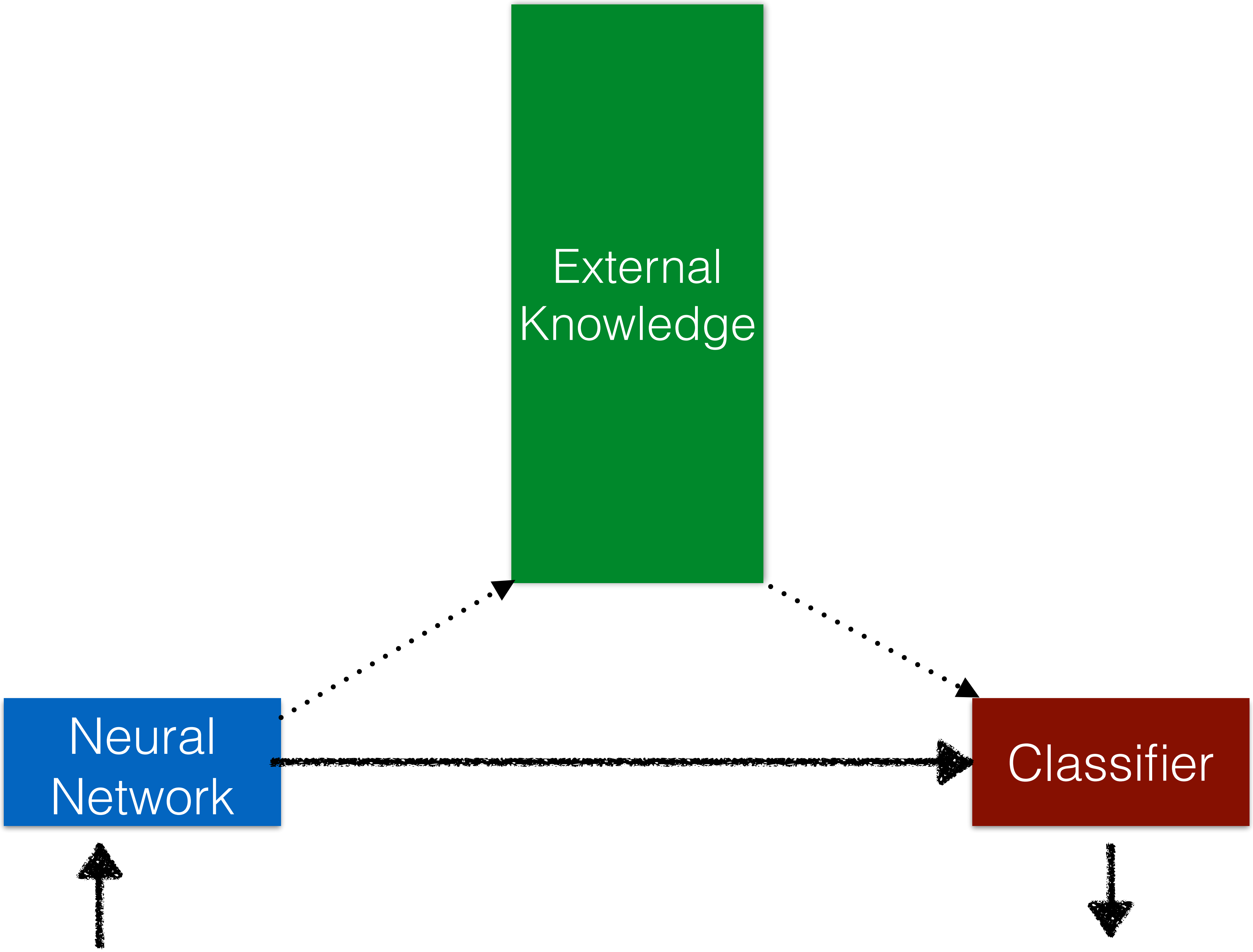
Seal

Query Formulation/ Template Extraction

Knowledge Base Search/ Candidate Answer Generation

Evidence Retrieval/ Candidate Scoring

Can we replace all of these modules with a single neural network?



External Knowledge

Neural Network

Classifier

Who wrote the song “Kiss from a Rose”?

Seal

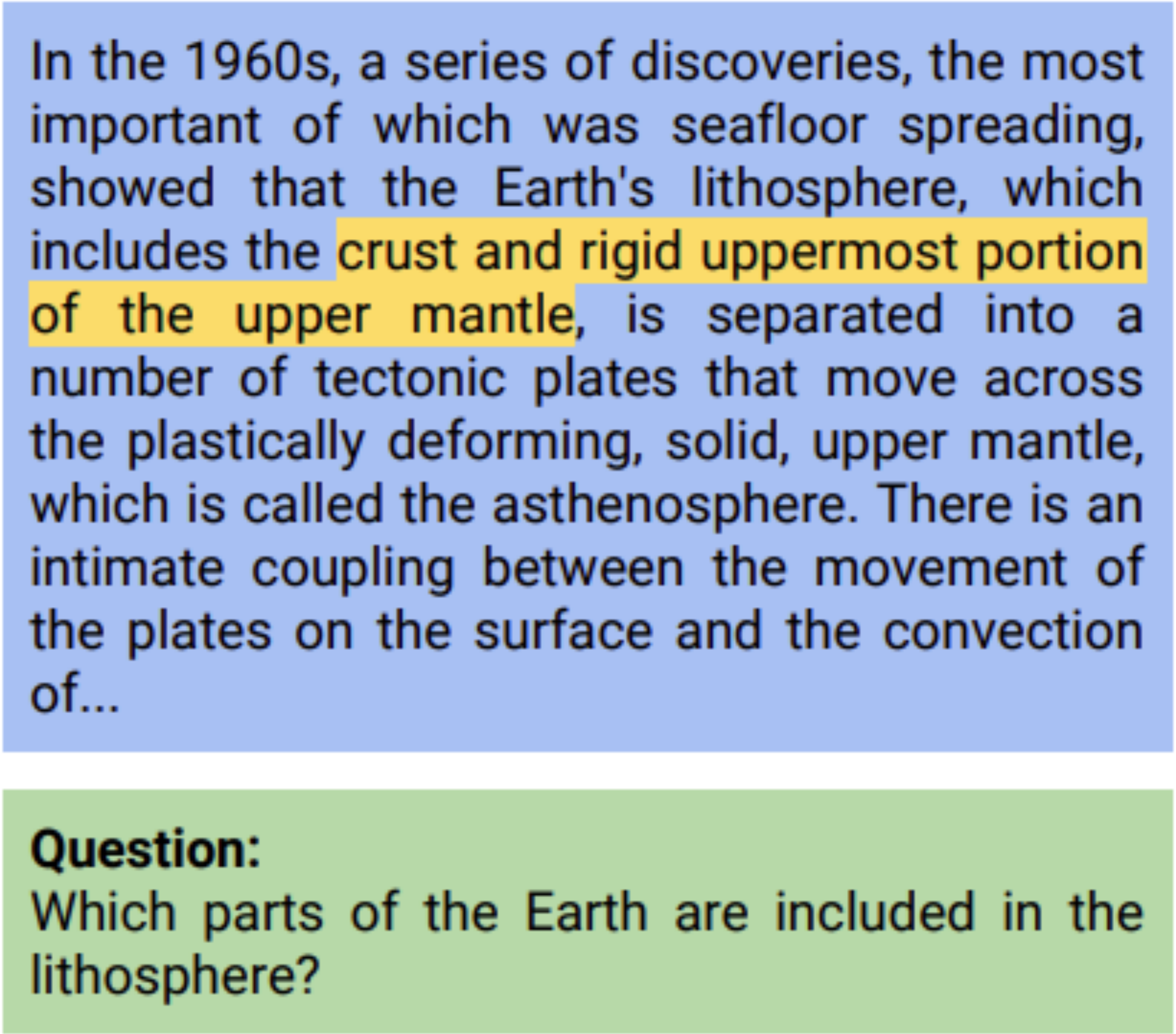
- factoid QA: the answer is a single entity / numeric
    - “who wrote the book “Dracula”?

- non-factoid QA: answer is free text
    - “why is Dracula so evil?”

- QA subtypes (could be factoid or non-factoid):
    - semantic parsing: question is mapped to a logical form which is then executed over some database
        - “how many people did Dracula bite?”
    - reading comprehension: answer is a span of text within a document (could be factoid or non-factoid)
    - community-based QA: question is answered by multiple web users (e.g., Yahoo! Answers)
    - visual QA: questions about images

Machine reading
(“reading comprehension”)

SQuAD



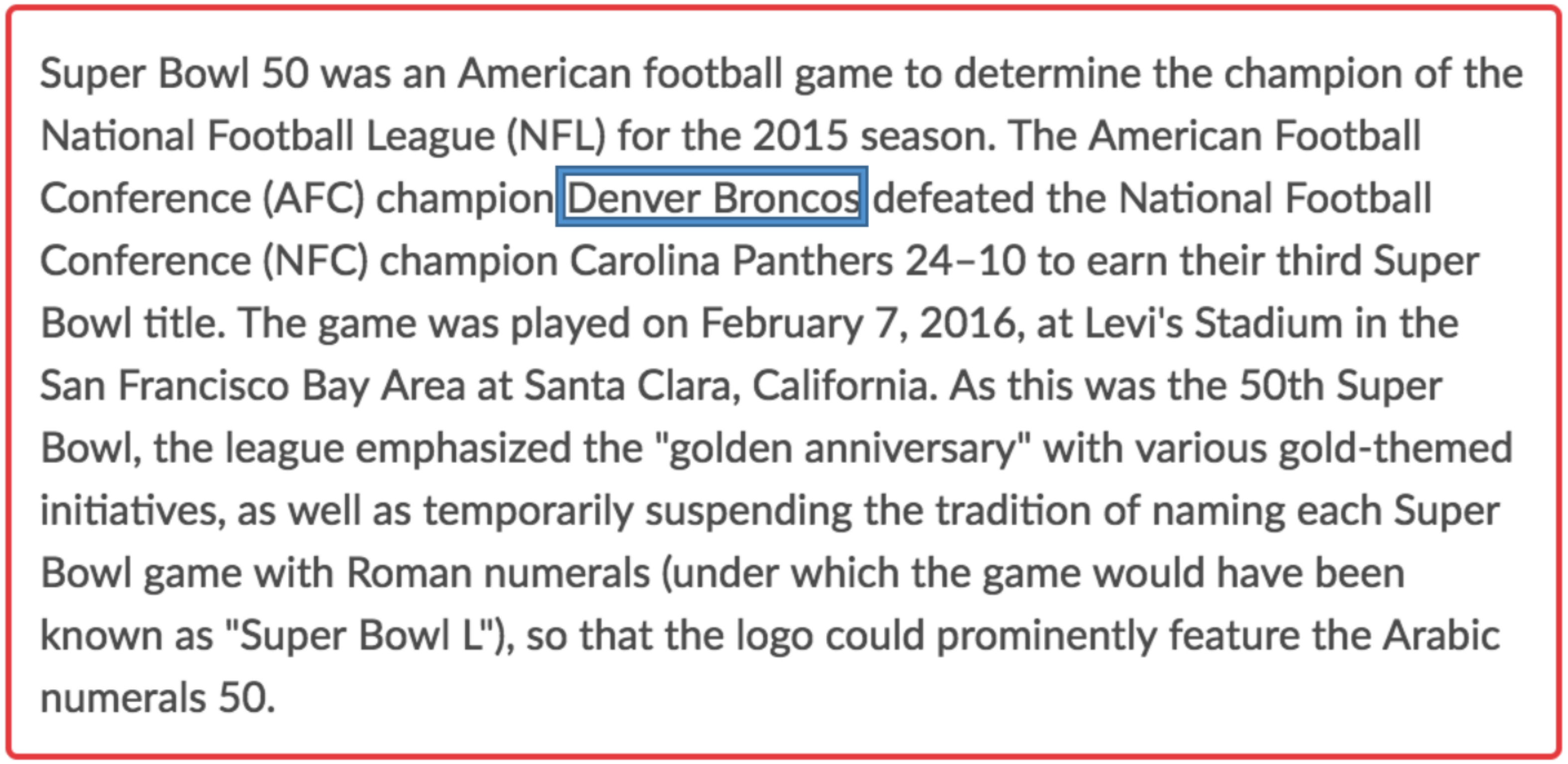
In the 1960s, a series of discoveries, the most important of which was seafloor spreading, showed that the Earth's lithosphere, which includes the crust and rigid uppermost portion of the upper mantle, is separated into a number of tectonic plates that move across the plastically deforming, solid, upper mantle, which is called the asthenosphere. There is an intimate coupling between the movement of the plates on the surface and the convection of...

Question:
Which parts of the Earth are included in the lithosphere?

Let’s look at the DRQA model (Chen et al., ACL 2017)

(pre-BERT)

Big idea



Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California. As this was the 50th Super Bowl, the league emphasized the "golden anniversary" with various gold-themed initiatives, as well as temporarily suspending the tradition of naming each Super Bowl game with Roman numerals (under which the game would have been known as "Super Bowl L"), so that the logo could prominently feature the Arabic numerals 50.

Q: Which NFL team represented the AFC at Super Bowl 50?

A: Denver Broncos

Start and End Probabilities

$$P_{\text{start}}(i) \propto \exp\{\vec{p}_i W_s \vec{q}\} \tag{1}$$

$$P_{\text{end}}(i) \propto \exp\{\vec{p}_i W_e \vec{q}\} \tag{2}$$

1. A vector representing our question
2. Vector representing each word in the query text
3. Parameter: here’s the start/end of the answer

Start and End Probabilities

$$P_{\text{start}}(i) \propto \exp\{\vec{p}_i W_s \vec{q}\} \tag{1}$$

$$P_{\text{end}}(i) \propto \exp\{\vec{p}_i W_e \vec{q}\} \tag{2}$$

1. A vector representing our question
2. Vector representing each word in the query text
3. Parameter: here's the start/end of the answer

Start and End Probabilities

$$P_{\text{start}}(i) \propto \exp\{\vec{p}_i W_s \vec{q}\} \tag{1}$$

$$P_{\text{end}}(i) \propto \exp\{\vec{p}_i W_e \vec{q}\} \tag{2}$$

1. A vector representing our question
2. Vector representing each word in the query text
3. Parameter: here’s the start/end of the answer

Start and End Probabilities

$$P_{\text{start}}(i) \propto \exp\left\{\vec{p}_i W_s \vec{q}\right\} \tag{1}$$

$$P_{\text{end}}(i) \propto \exp\left\{\vec{p}_i W_e \vec{q}\right\} \tag{2}$$

1. A vector representing our question
2. Vector representing each word in the query text
3. Parameter: here's the start/end of the answer

Start and End Probabilities

$$P_{\text{start}}(i) \propto \exp\left\{\vec{p}_i W_s \vec{q}\right\} \tag{1}$$

$$P_{\text{end}}(i) \propto \exp\left\{\vec{p}_i W_e \vec{q}\right\} \tag{2}$$

1. A vector representing our question
2. Vector representing each word in the query text
3. Parameter: here’s the start/end of the answer

Start and End Probabilities

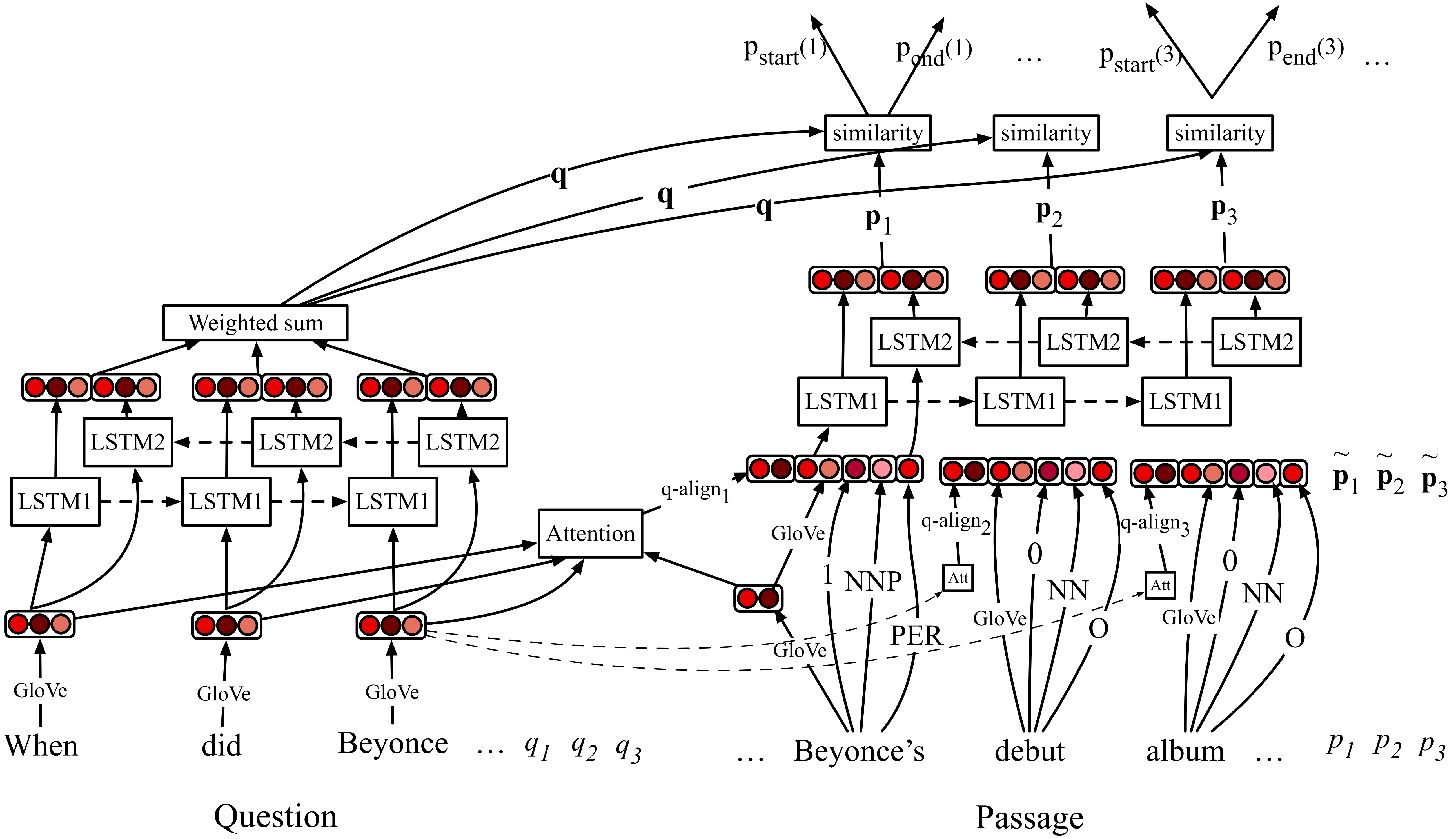
$$P_{\text{start}}(i) \propto \exp\{\vec{p}_i W_s \vec{q}\} \quad (1)$$

$$P_{\text{end}}(i) \propto \exp\{\vec{p}_i W_e \vec{q}\} \quad (2)$$

1. A vector representing our question
2. Vector representing each word in the query text
3. Parameter: here's the start/end of the answer

How does this work at test-time?

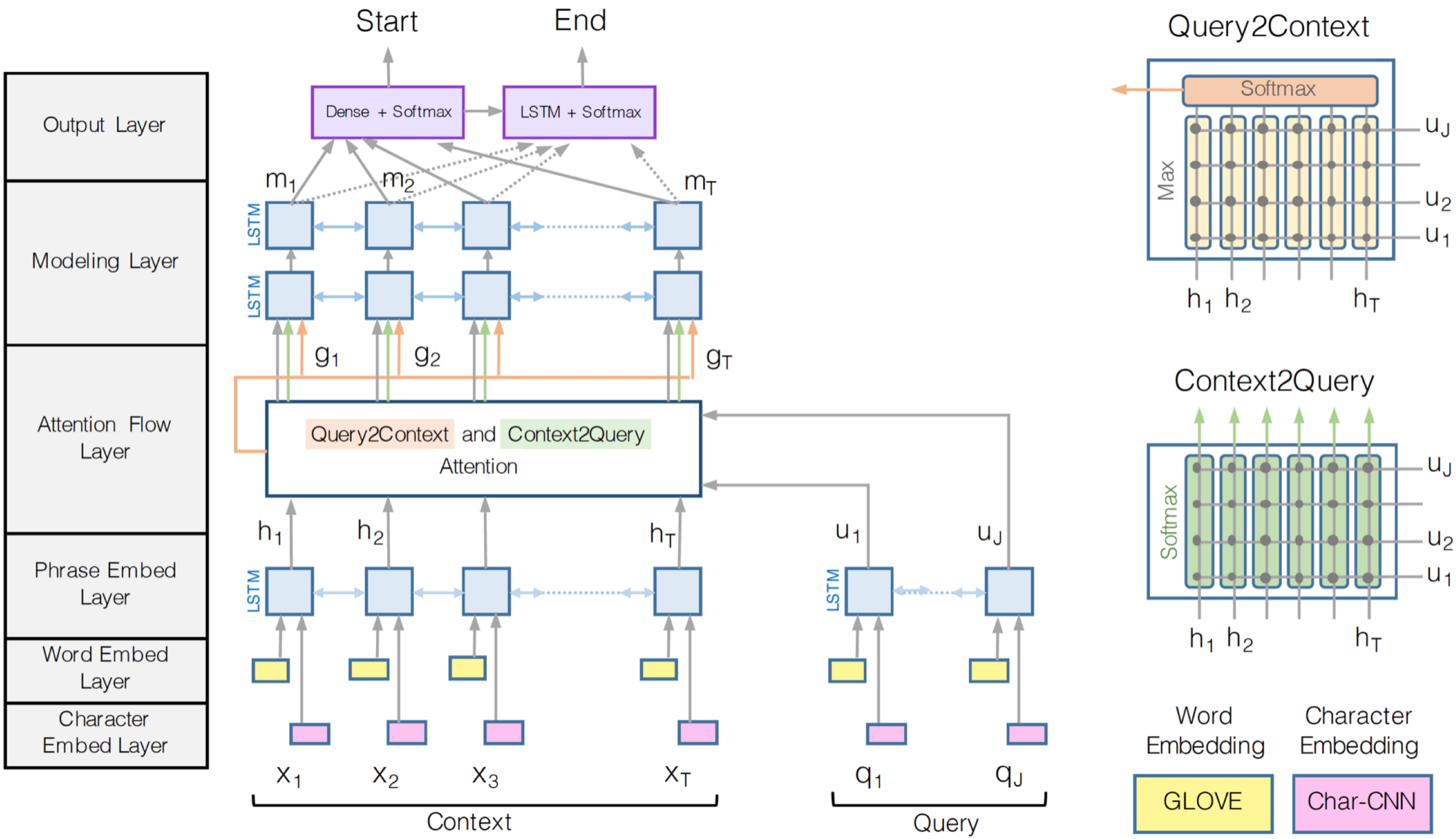
Stanford Attentive Reader++



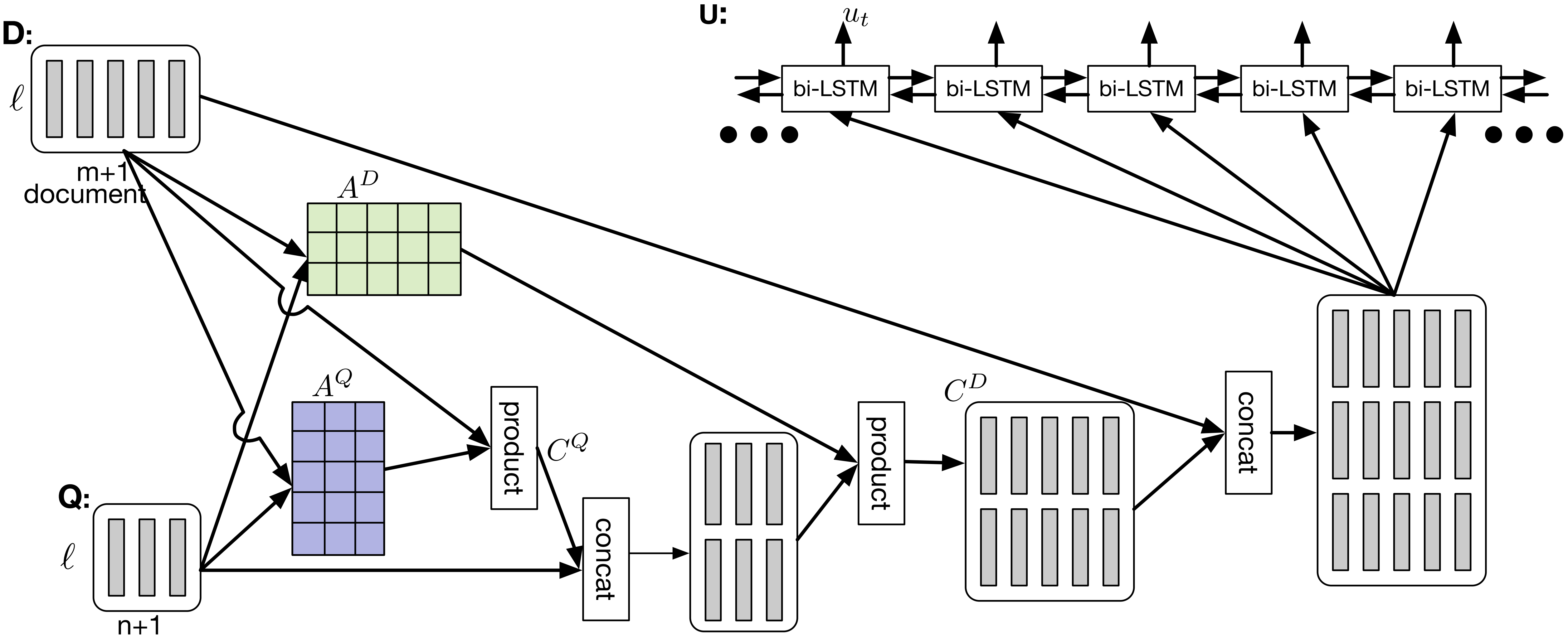
Training objective: $\mathcal{L} = -\sum \log P^{(\text{start})}(a_{\text{start}}) - \sum \log P^{(\text{end})}(a_{\text{end}})$

5. BiDAF: Bi-Directional Attention Flow for Machine Comprehension

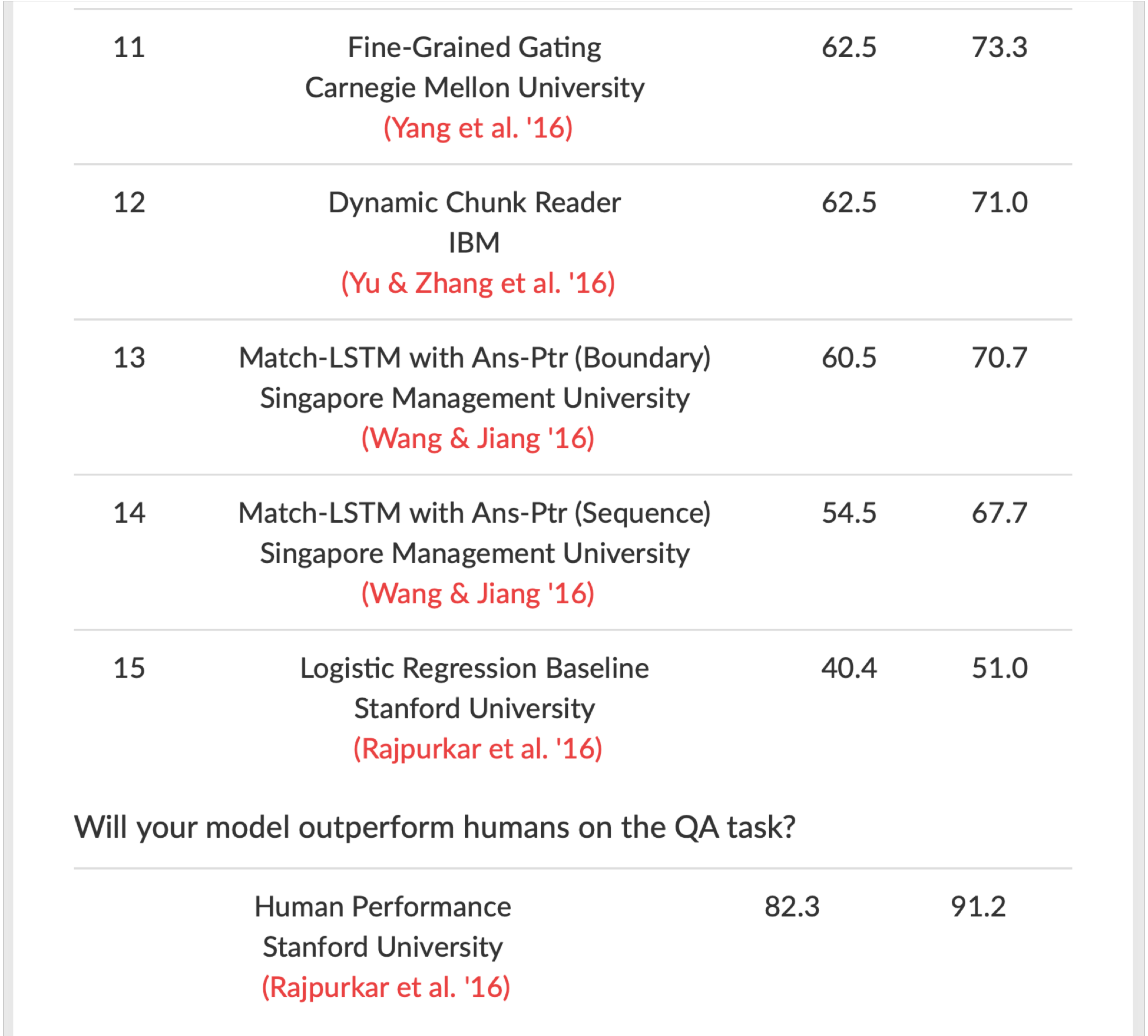
(Seo, Kembhavi, Farhadi, Hajishirzi, ICLR 2017)



Coattention Encoder



SQuAD v1.1 leaderboard, end of 2016 (Dec 6)



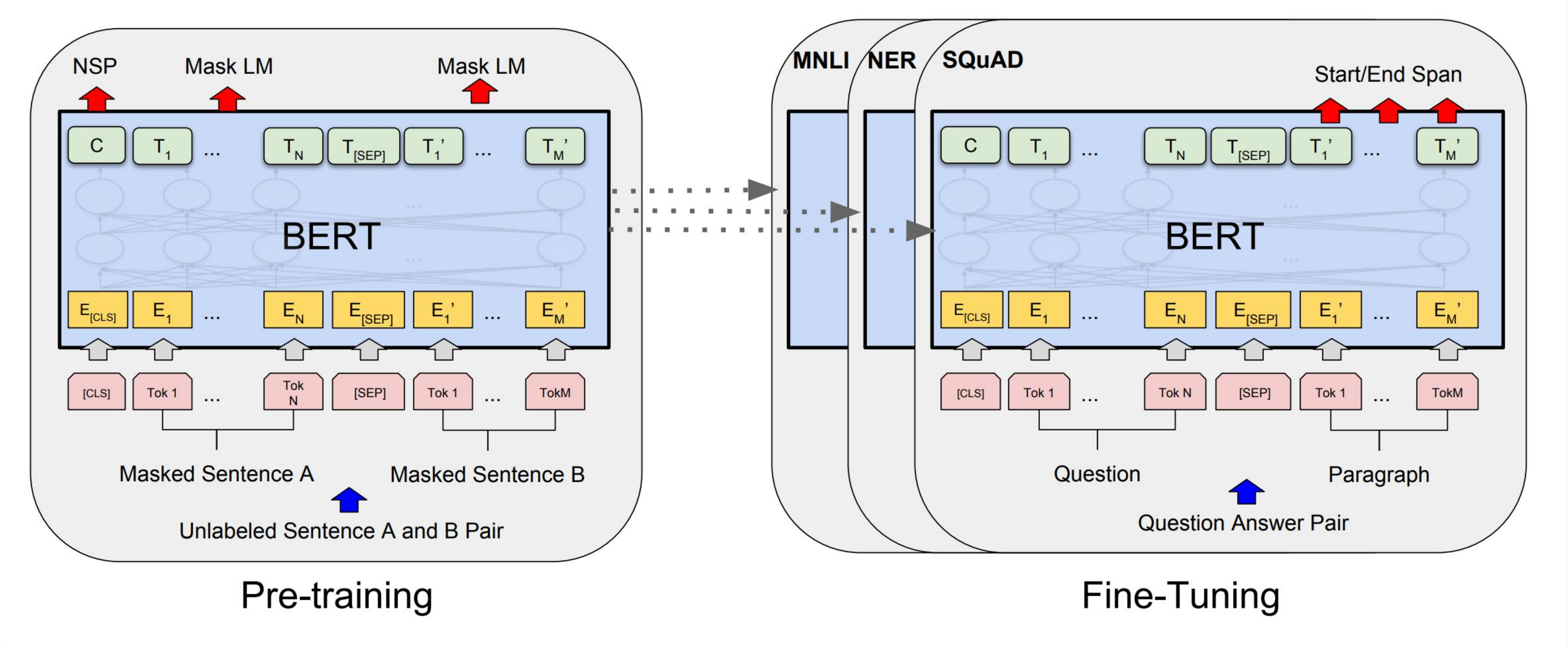
		EM	F1
11	Fine-Grained Gating Carnegie Mellon University (Yang et al. '16)	62.5	73.3
12	Dynamic Chunk Reader IBM (Yu & Zhang et al. '16)	62.5	71.0
13	Match-LSTM with Ans-Ptr (Boundary) Singapore Management University (Wang & Jiang '16)	60.5	70.7
14	Match-LSTM with Ans-Ptr (Sequence) Singapore Management University (Wang & Jiang '16)	54.5	67.7
15	Logistic Regression Baseline Stanford University (Rajpurkar et al. '16)	40.4	51.0

Will your model outperform humans on the QA task?

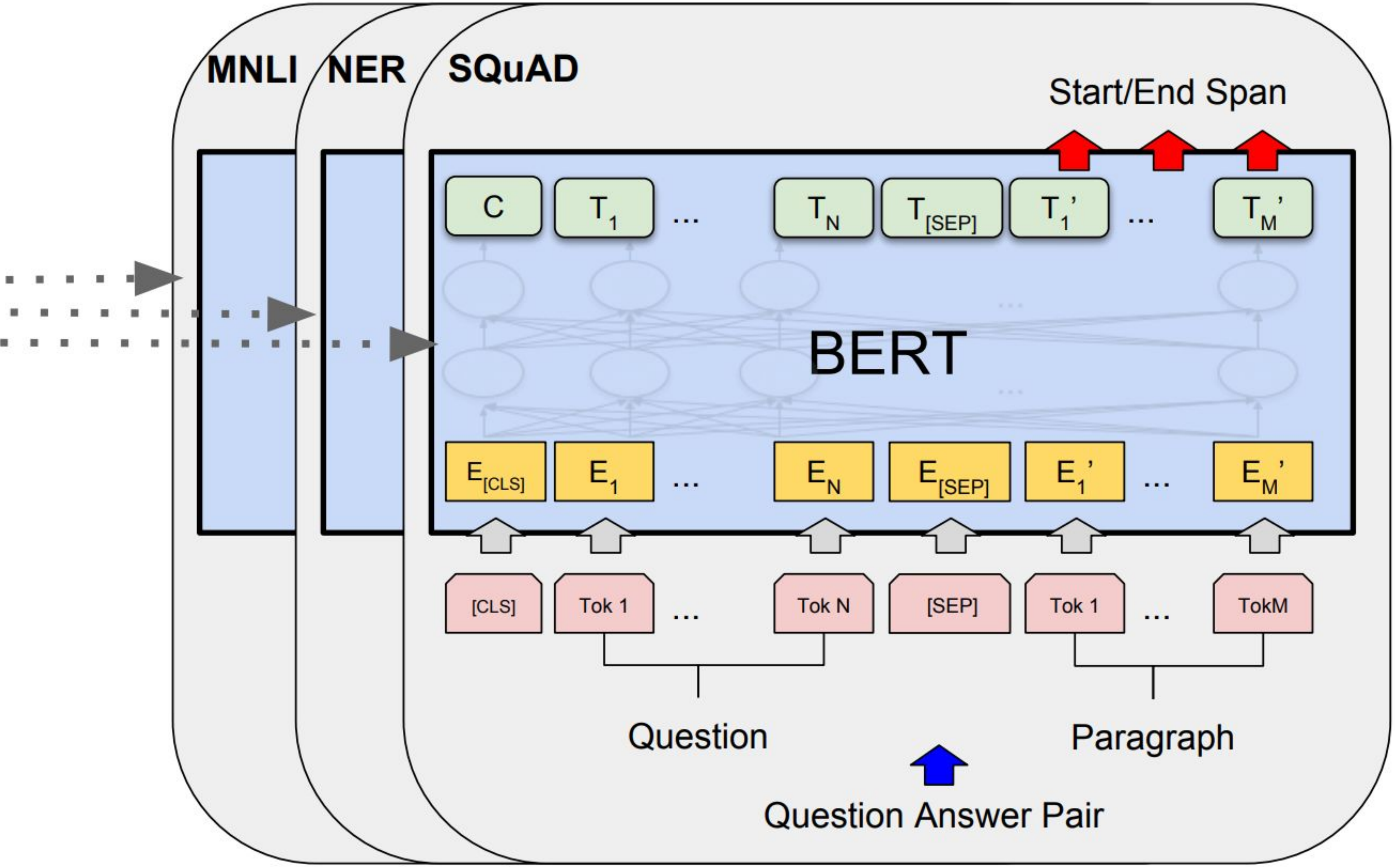
Human Performance Stanford University (Rajpurkar et al. '16)	82.3	91.2

All of these models are trained from scratch on the SQuAD training set!!!

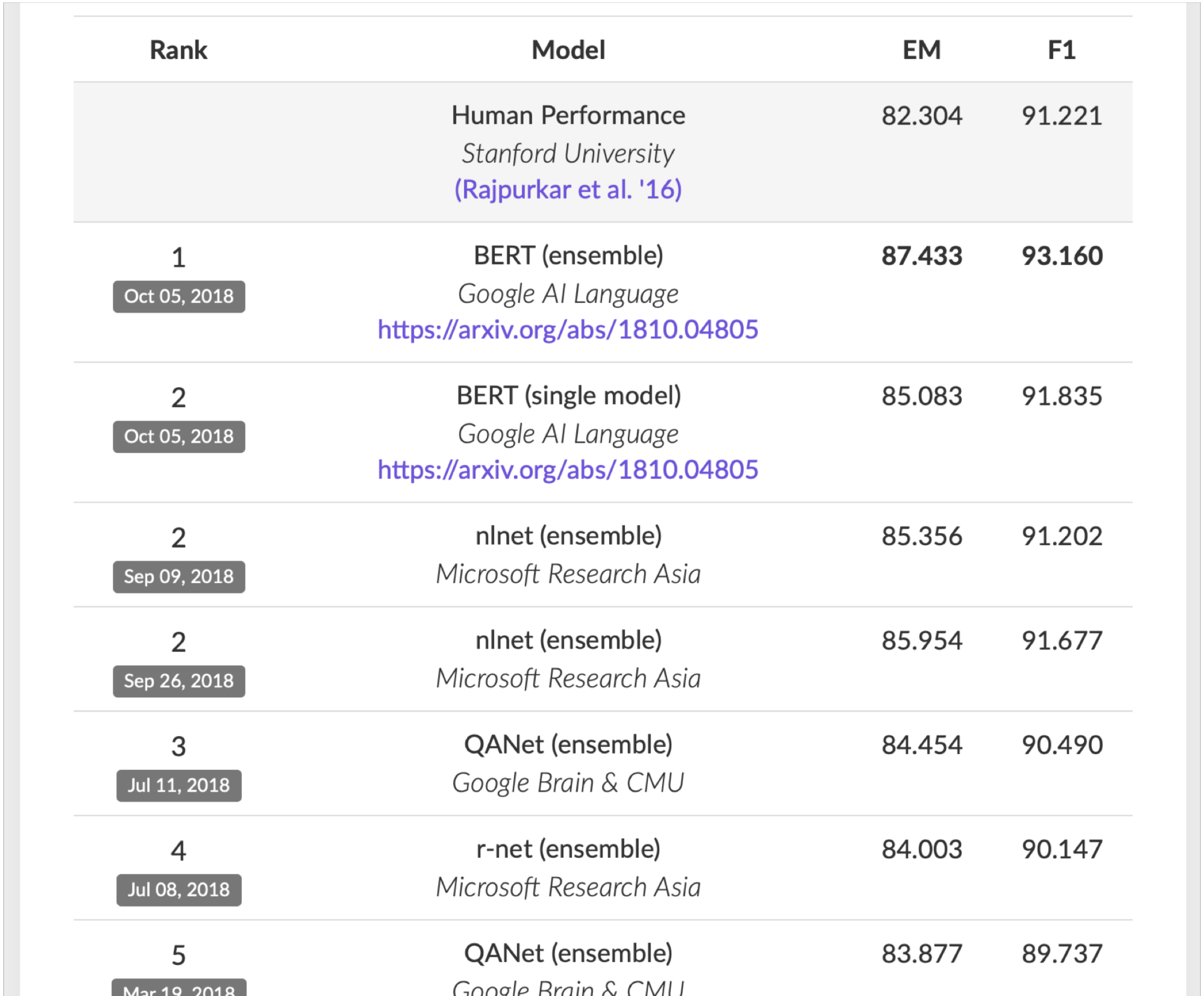
Fine-Tuning Procedure



Simply concatenate the question and paragraph into a single sequence, pass through BERT, and apply a softmax layer on the final layer token representations to predict start/end answer span boundaries



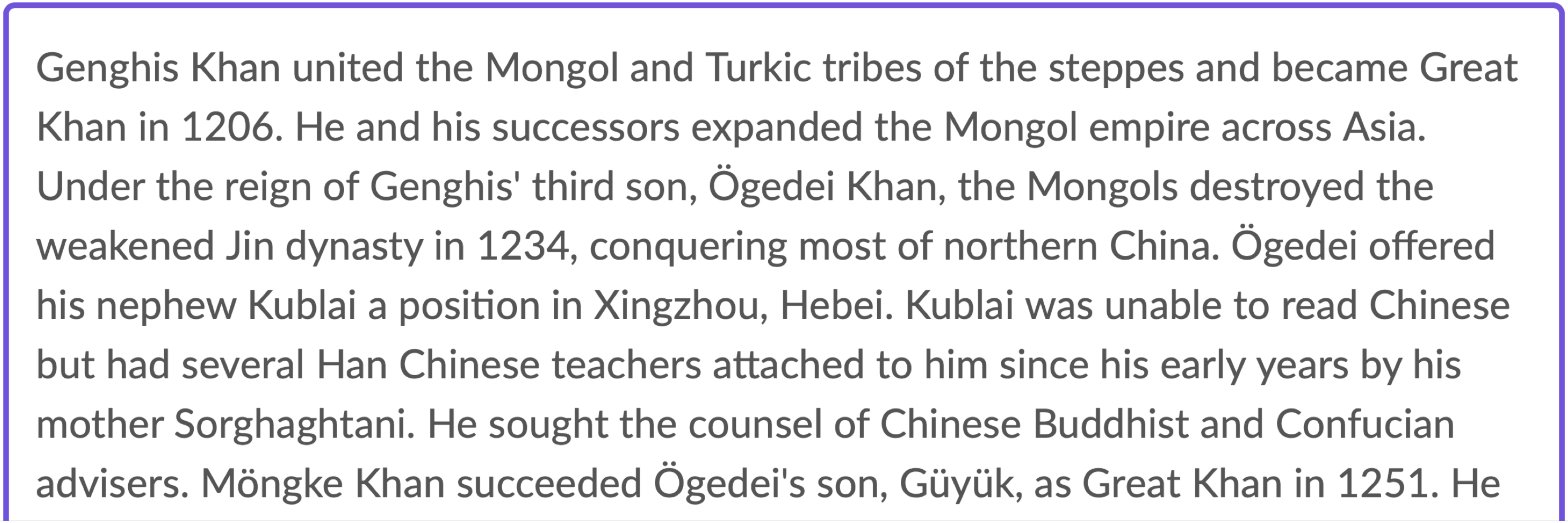
SQuAD v1.1 leaderboard, 2019-02-07 – it’s solved!



Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) *Google AI Language* https://arxiv.org/abs/1810.04805	**87.433**	**93.160**
2 Oct 05, 2018	BERT (single model) *Google AI Language* https://arxiv.org/abs/1810.04805	85.083	91.835
2 Sep 09, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.356	91.202
2 Sep 26, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.954	91.677
3 Jul 11, 2018	QANet (ensemble) *Google Brain & CMU*	84.454	90.490
4 Jul 08, 2018	r-net (ensemble) *Microsoft Research Asia*	84.003	90.147
5 Mar 19, 2018	QANet (ensemble) *Google Brain & CMU*	83.877	89.737

Transfer learning via BERT made most of the task-specific QA architectures obsolete

SQuAD 2.0 Example



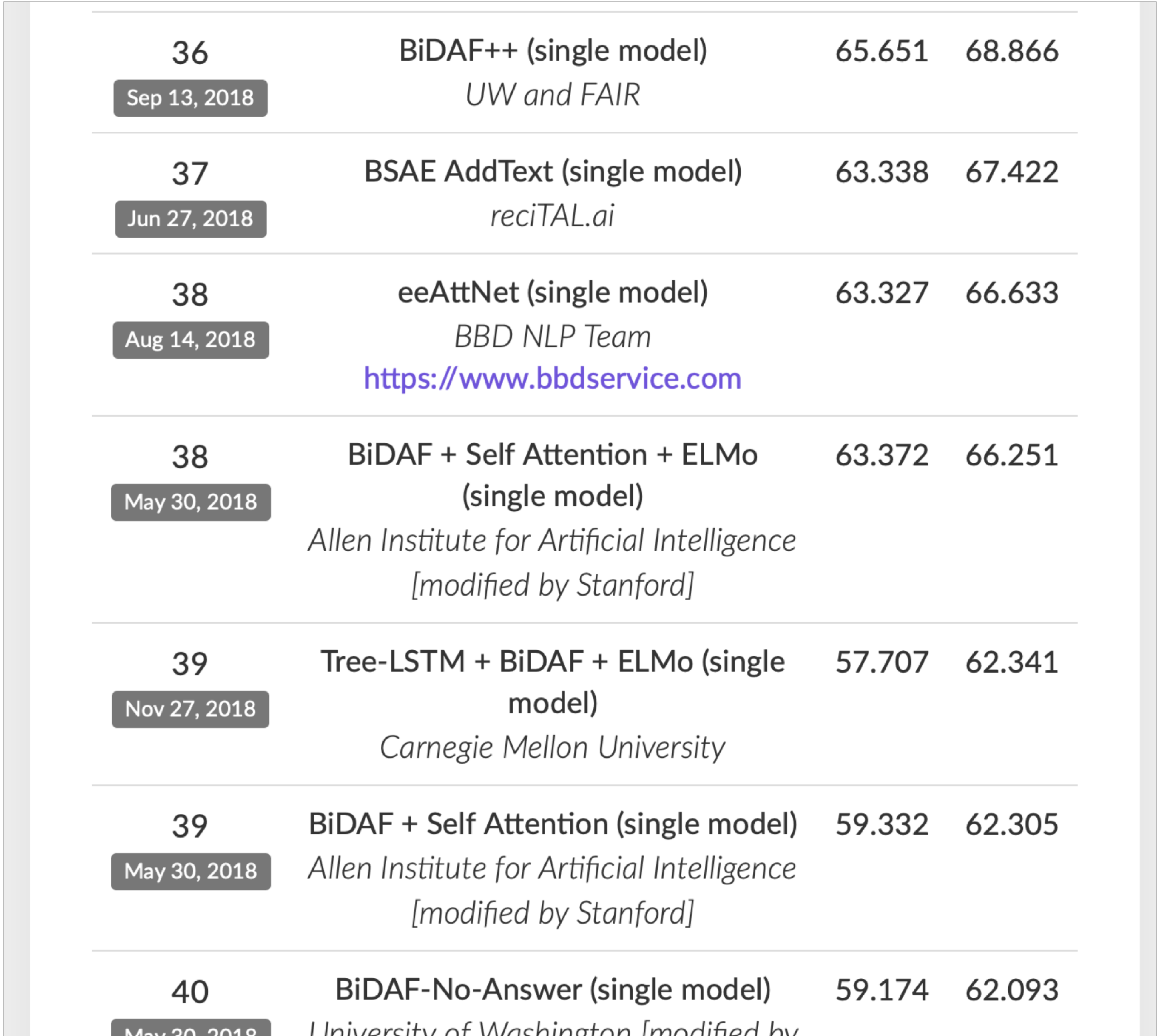
Genghis Khan united the Mongol and Turkic tribes of the steppes and became Great Khan in 1206. He and his successors expanded the Mongol empire across Asia. Under the reign of Genghis' third son, Ögedei Khan, the Mongols destroyed the weakened Jin dynasty in 1234, conquering most of northern China. Ögedei offered his nephew Kublai a position in Xingzhou, Hebei. Kublai was unable to read Chinese but had several Han Chinese teachers attached to him since his early years by his mother Sorghaghtani. He sought the counsel of Chinese Buddhist and Confucian advisers. Möngke Khan succeeded Ögedei's son, Güyük, as Great Khan in 1251. He

When did Genghis Khan kill Great Khan?

Gold Answers: <No Answer>

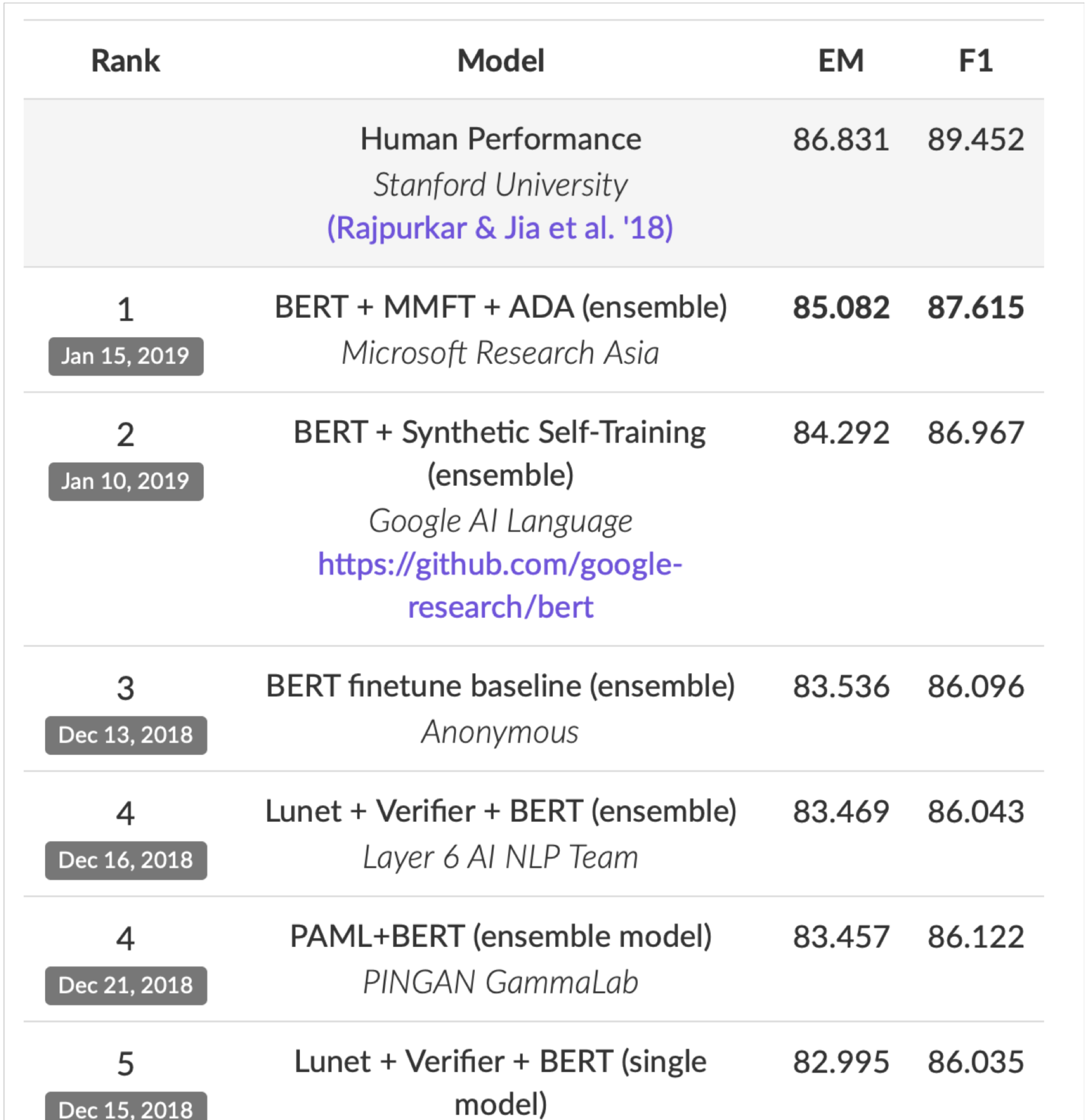
Prediction: 1234 [from Microsoft nlnet]

SQuAD 2.0 leaderboard, 2019-02-07



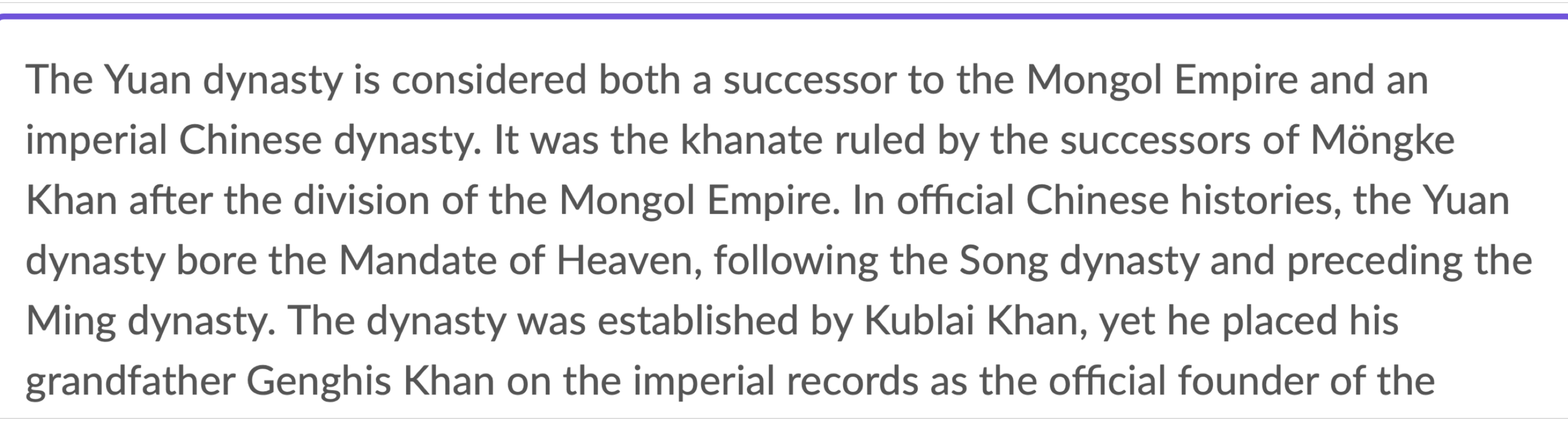
		EM	F1
36 Sep 13, 2018	BiDAF++ (single model) *UW and FAIR*	65.651	68.866
37 Jun 27, 2018	BSAE AddText (single model) *reciTAL.ai*	63.338	67.422
38 Aug 14, 2018	eeAttNet (single model) *BBD NLP Team* https://www.bbdservice.com	63.327	66.633
38 May 30, 2018	BiDAF + Self Attention + ELMo (single model) *Allen Institute for Artificial Intelligence [modified by Stanford]*	63.372	66.251
39 Nov 27, 2018	Tree-LSTM + BiDAF + ELMo (single model) *Carnegie Mellon University*	57.707	62.341
39 May 30, 2018	BiDAF + Self Attention (single model) *Allen Institute for Artificial Intelligence [modified by Stanford]*	59.332	62.305
40 May 30, 2018	BiDAF-No-Answer (single model) *University of Washington [modified by*	59.174	62.093

SQuAD 2.0 leaderboard, 2019-02-07



Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Jan 15, 2019	BERT + MMFT + ADA (ensemble) *Microsoft Research Asia*	**85.082**	**87.615**
2 Jan 10, 2019	BERT + Synthetic Self-Training (ensemble) *Google AI Language* https://github.com/google-research/bert	84.292	86.967
3 Dec 13, 2018	BERT finetune baseline (ensemble) *Anonymous*	83.536	86.096
4 Dec 16, 2018	Lunet + Verifier + BERT (ensemble) *Layer 6 AI NLP Team*	83.469	86.043
4 Dec 21, 2018	PAML+BERT (ensemble model) *PINGAN GammaLab*	83.457	86.122
5 Dec 15, 2018	Lunet + Verifier + BERT (single model)	82.995	86.035

Good systems are great, but still basic NLU errors



The Yuan dynasty is considered both a successor to the Mongol Empire and an imperial Chinese dynasty. It was the khanate ruled by the successors of Möngke Khan after the division of the Mongol Empire. In official Chinese histories, the Yuan dynasty bore the Mandate of Heaven, following the Song dynasty and preceding the Ming dynasty. The dynasty was established by Kublai Khan, yet he placed his grandfather Genghis Khan on the imperial records as the official founder of the

What dynasty came before the Yuan?

Gold Answers: ① Song dynasty ② Mongol Empire
③ the Song dynasty

Prediction: Ming dynasty [BERT (single model) (Google AI)]

SQuAD limitations

- SQuAD has a number of other key limitations too:
  - Only span-based answers (no yes/no, counting, implicit why)
  - Questions were constructed looking at the passages
    - Not genuine information needs
    - Generally greater lexical and syntactic matching between questions and answer span than you get IRL
  - Barely any multi-fact/sentence inference beyond coreference

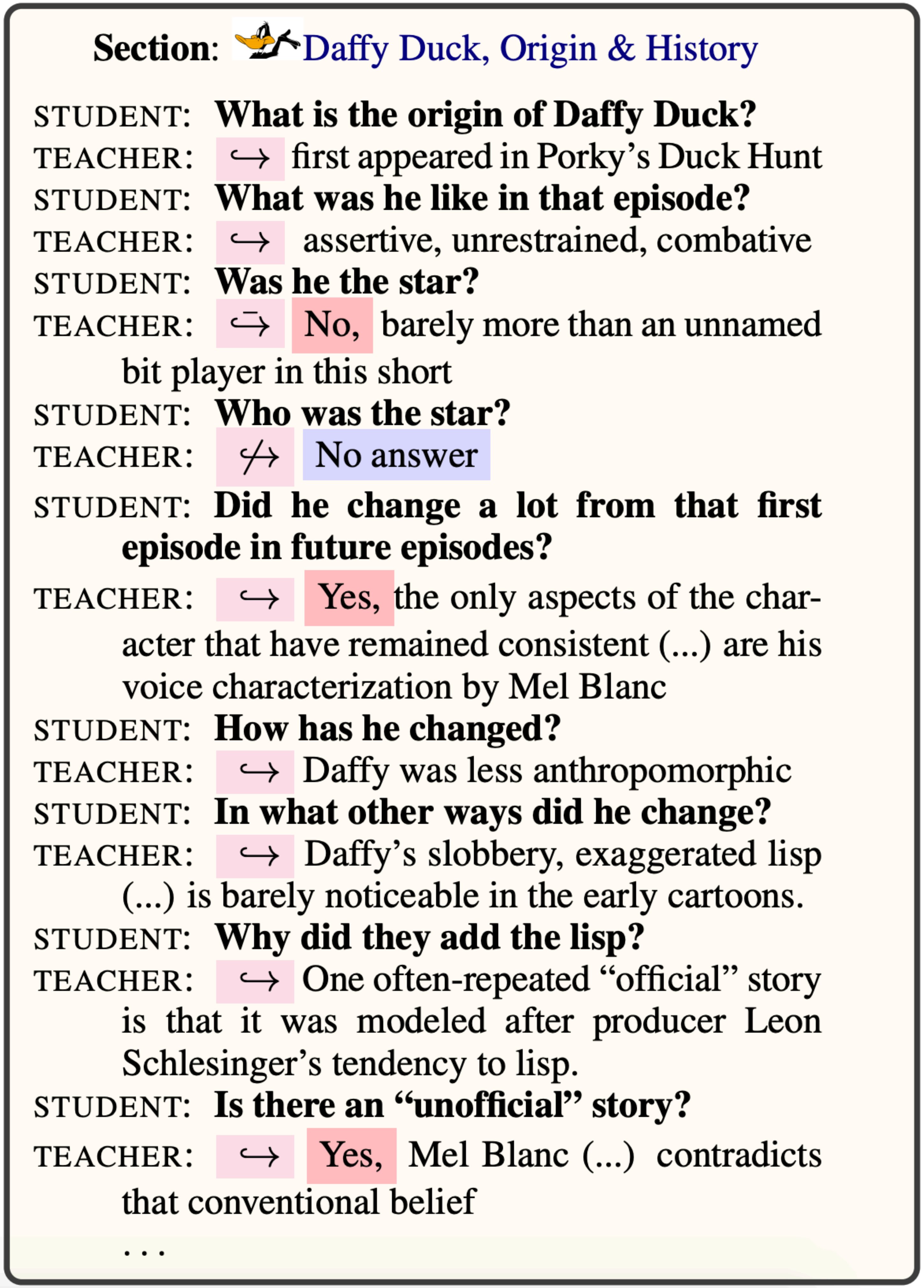
- Nevertheless, it is a well-targeted, well-structured, clean dataset
  - It has been the most used and competed on QA dataset
  - It has also been a useful starting point for building systems in industry (though in-domain data always really helps!)

Several variants of the SQuAD style setup (all easily portable to BERT :)

Conversational question answering: Multiple questions about the same document (answers still spans from the document)

datasets: QuAC, CoQA, CSQA, etc

How do we use BERT to solve this task?



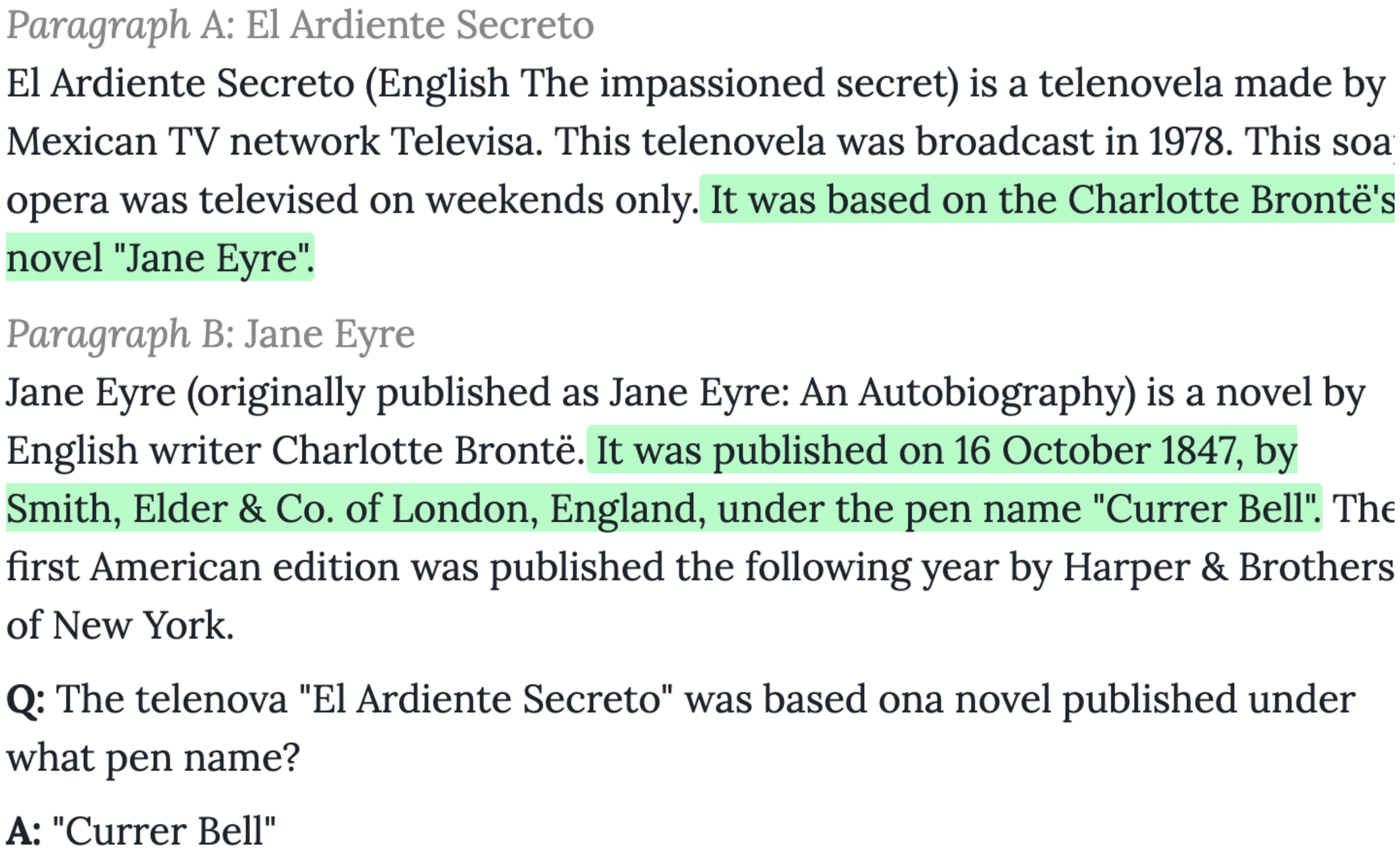
Section: Daffy Duck, Origin & History

STUDENT: **What is the origin of Daffy Duck?**
TEACHER: ↪ first appeared in Porky's Duck Hunt
STUDENT: **What was he like in that episode?**
TEACHER: ↪ assertive, unrestrained, combative
STUDENT: **Was he the star?**
TEACHER: ↪̄ No, barely more than an unnamed bit player in this short
STUDENT: **Who was the star?**
TEACHER: ↪̸ No answer
STUDENT: **Did he change a lot from that first episode in future episodes?**
TEACHER: ↪ Yes, the only aspects of the character that have remained consistent (...) are his voice characterization by Mel Blanc
STUDENT: **How has he changed?**
TEACHER: ↪ Daffy was less anthropomorphic
STUDENT: **In what other ways did he change?**
TEACHER: ↪ Daffy's slobbery, exaggerated lisp (...) is barely noticeable in the early cartoons.
STUDENT: **Why did they add the lisp?**
TEACHER: ↪ One often-repeated "official" story is that it was modeled after producer Leon Schlesinger's tendency to lisp.
STUDENT: **Is there an "unofficial" story?**
TEACHER: ↪ Yes, Mel Blanc (...) contradicts that conventional belief
. . .

Multi-hop question answering:

Requires models to perform more “reasoning” over the document

datasets: HotpotQA, QAngaroo



Paragraph A: El Ardiente Secreto

El Ardiente Secreto (English The impassioned secret) is a telenovela made by Mexican TV network Televisa. This telenovela was broadcast in 1978. This soa opera was televised on weekends only. It was based on the Charlotte Brontë's novel "Jane Eyre".

Paragraph B: Jane Eyre

Jane Eyre (originally published as Jane Eyre: An Autobiography) is a novel by English writer Charlotte Brontë. It was published on 16 October 1847, by Smith, Elder & Co. of London, England, under the pen name "Currer Bell". The first American edition was published the following year by Harper & Brothers of New York.

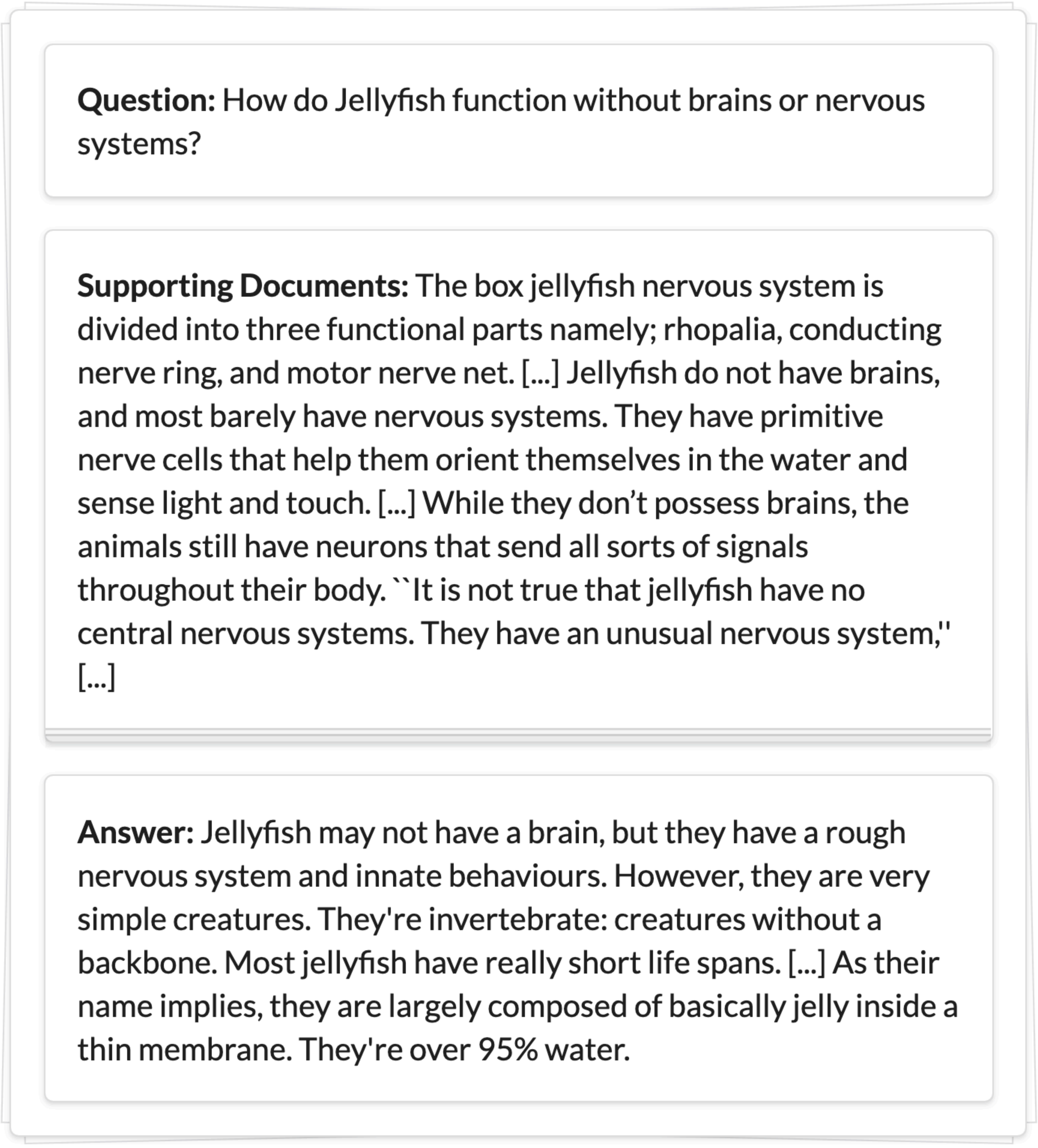
Q: The telenova "El Ardiente Secreto" was based ona novel published under what pen name?

A: "Currer Bell"

long-form question answering: Answers must be *generated*, not extracted

datasets: ELI5, NarrativeQA, etc

More on these later!



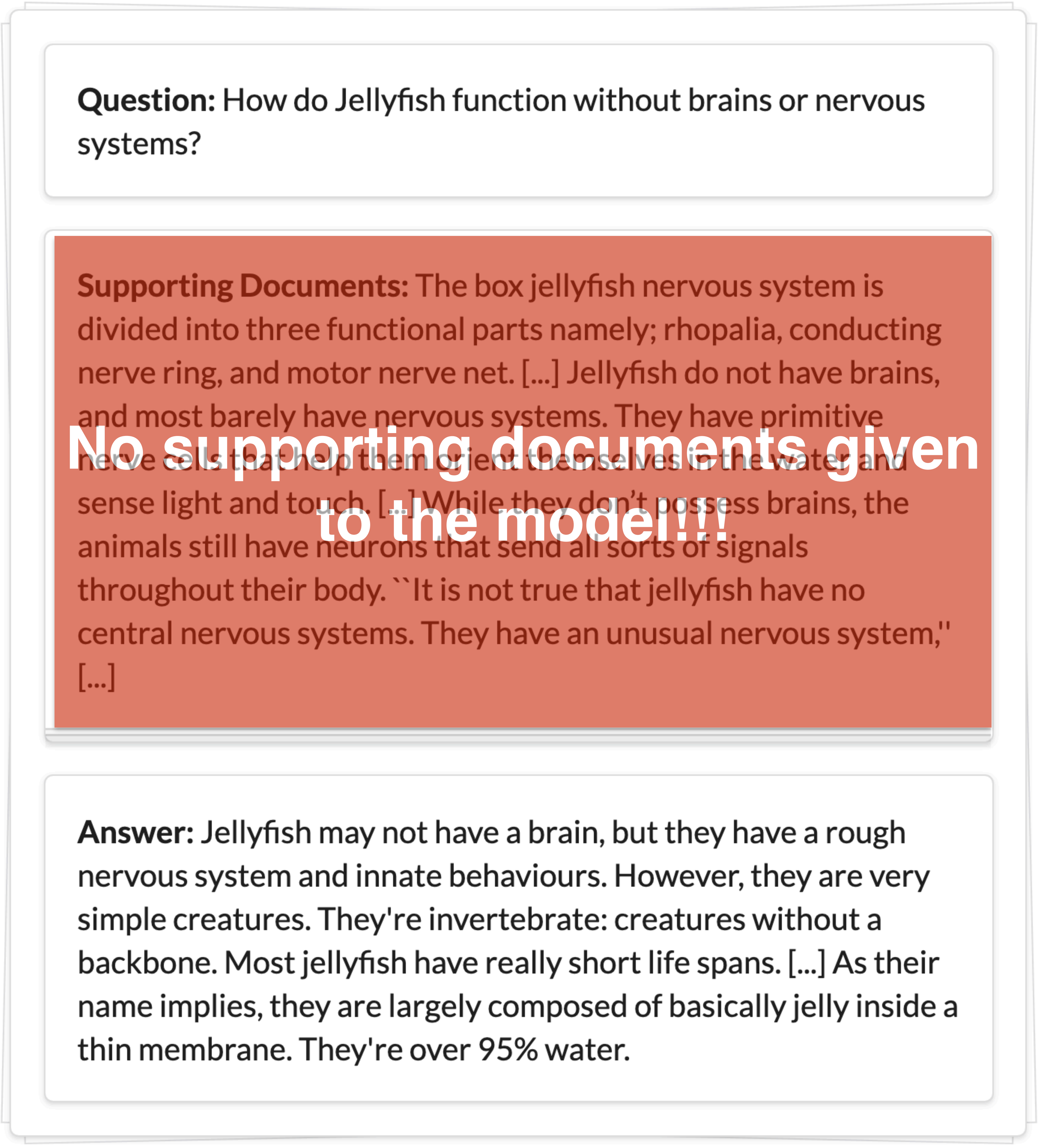
Question: How do Jellyfish function without brains or nervous systems?

Supporting Documents: The box jellyfish nervous system is divided into three functional parts namely; rhopalia, conducting nerve ring, and motor nerve net. [...] 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 don't possess brains, the animals still have neurons that send all sorts of signals throughout their body. ``It is not true that jellyfish have no central nervous systems. They have an unusual nervous system,'' [...]

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. [...] As their name implies, they are largely composed of basically jelly inside a thin membrane. They're over 95% water.

open-domain question answering: a model must retrieve relevant documents and use them to generate an answer (no evidence given!)

The future of QA?



Question: How do Jellyfish function without brains or nervous systems?

Supporting Documents: The box jellyfish nervous system is divided into three functional parts namely; rhopalia, conducting nerve ring, and motor nerve net. [...] 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 don't possess brains, the animals still have neurons that send all sorts of signals throughout their body. ``It is not true that jellyfish have no central nervous systems. They have an unusual nervous system,'' [...]

No supporting documents given to the model!!!

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. [...] As their name implies, they are largely composed of basically jelly inside a thin membrane. They're over 95% water.

All of these QA tasks are very similar… can we share information across different datasets to improve our performance across the board? (more next time!)

finally… a real-world example of deploying QA models

Quiz Bowl

what is quiz bowl?

- a trivia game that contains questions about famous entities (e.g., novels, battles, countries)

- developed a deep learning system, QANTA, to play quiz bowl

- one of the first applications of deep learning to question answering

Iyyer et al., EMNLP 2014 & ACL 2015

This author described a "plank in reason" breaking and hitting a "world at every plunge" in a poem which opens "I felt a funeral in my brain."

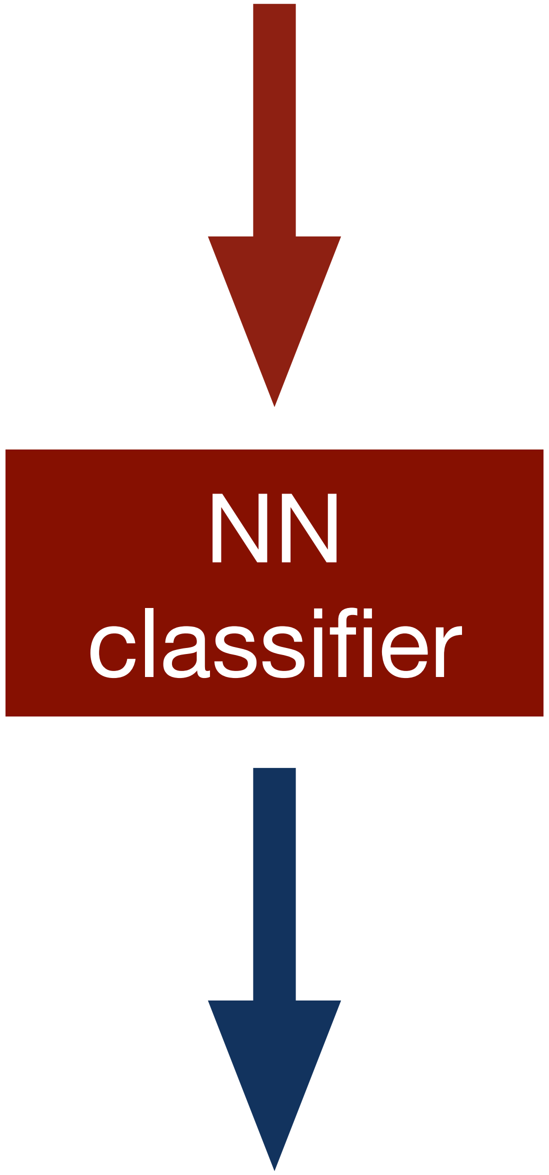
She wrote that "the stillness round my form was like the stillness in the air" in "I heard a fly buzz when I died."

She wrote about a scarcely visible roof and a cornice that was "but a mound" in a poem about a carriage ride with Immortality and Death.

For 10 points, name this reclusive "Belle of Amherst" who wrote "Because I could not stop for Death."

A: Emily Dickinson

… name this reclusive "Belle of Amherst”…

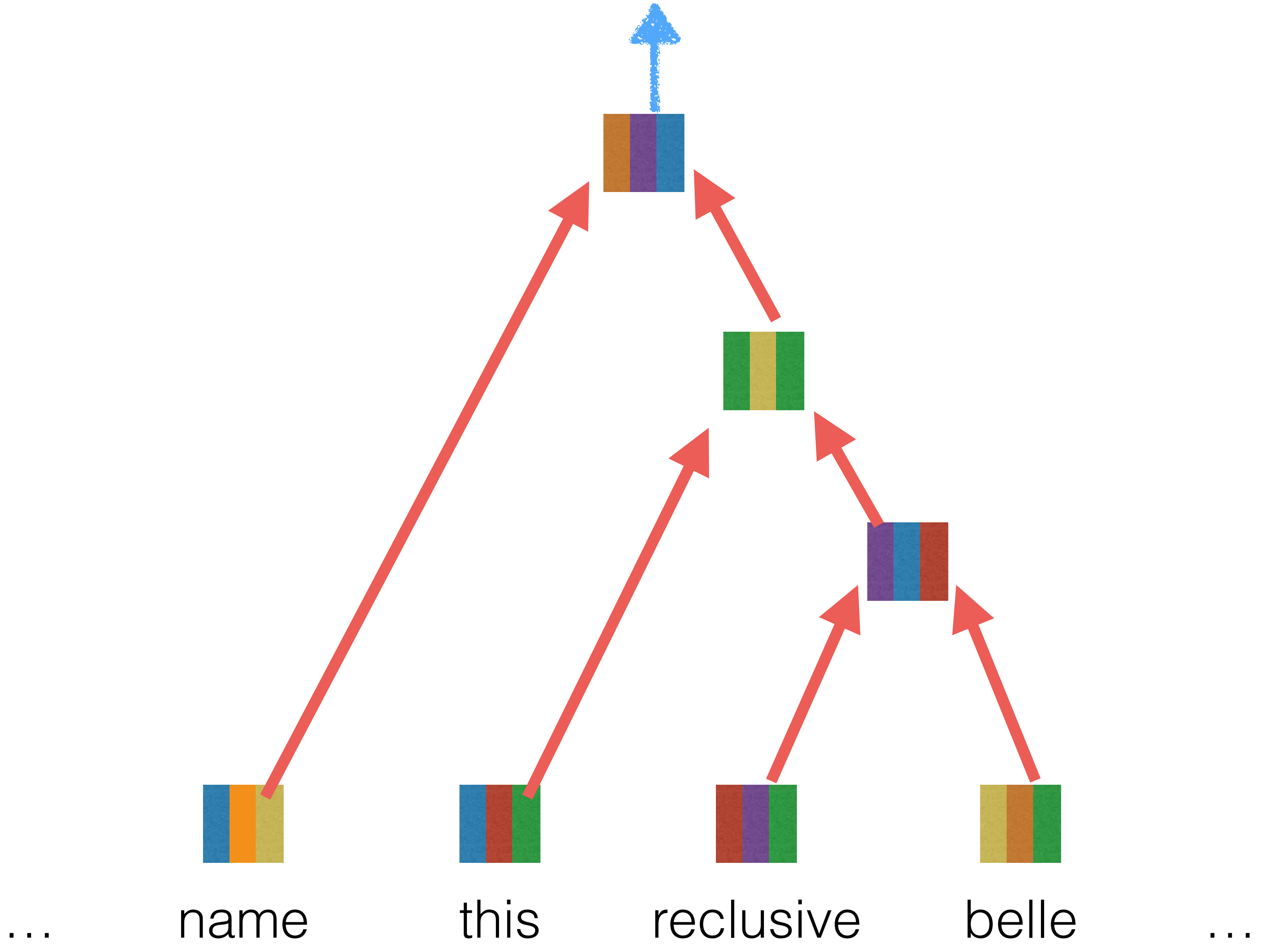


NN classifier

Emily Dickinson

dependency-tree NNs

softmax: predict **Emily Dickinson** out of a set of ~5000 answers



… name this reclusive belle …

Iyyer et al., EMNLP 2014

simple discourse-level representations by averaging

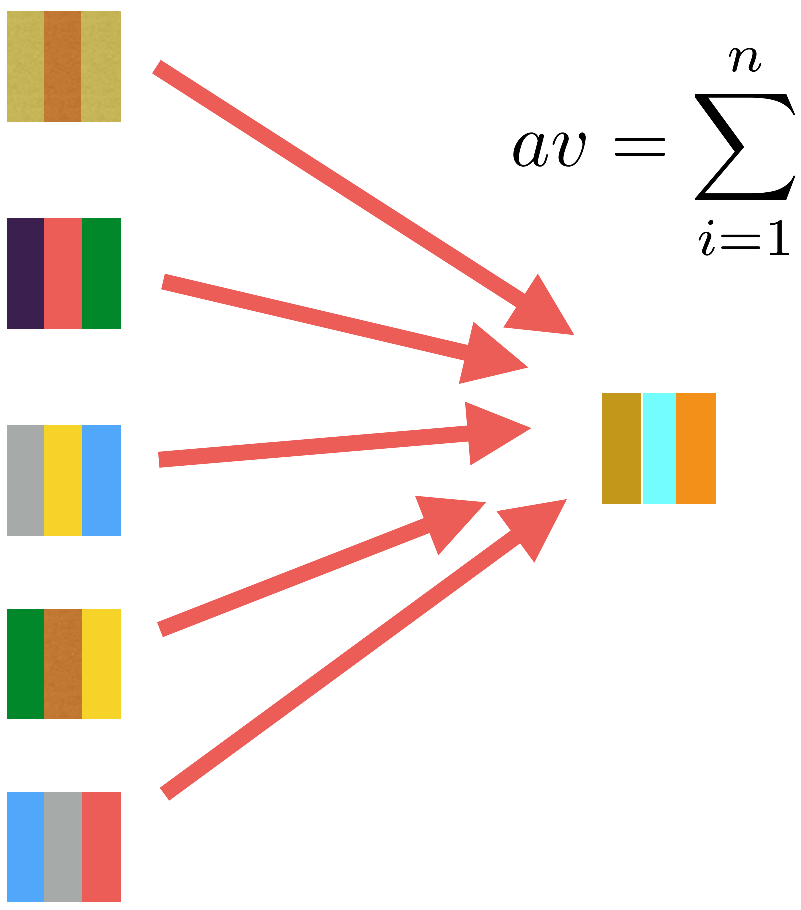
In one novel, one of these figures antagonizes an impoverished family before leaping into an active volcano.

Another of these figures titles a novella in which General Spielsdorf describes the circumstances of his niece Bertha Reinfeldt's death to the narrator, Laura.

In addition to Varney and Carmilla, another of these figures sails on the Russian ship Demeter in order to reach London.

That figure bites Lucy Westenra before being killed by a coalition including Jonathan Harker and Van Helsing.

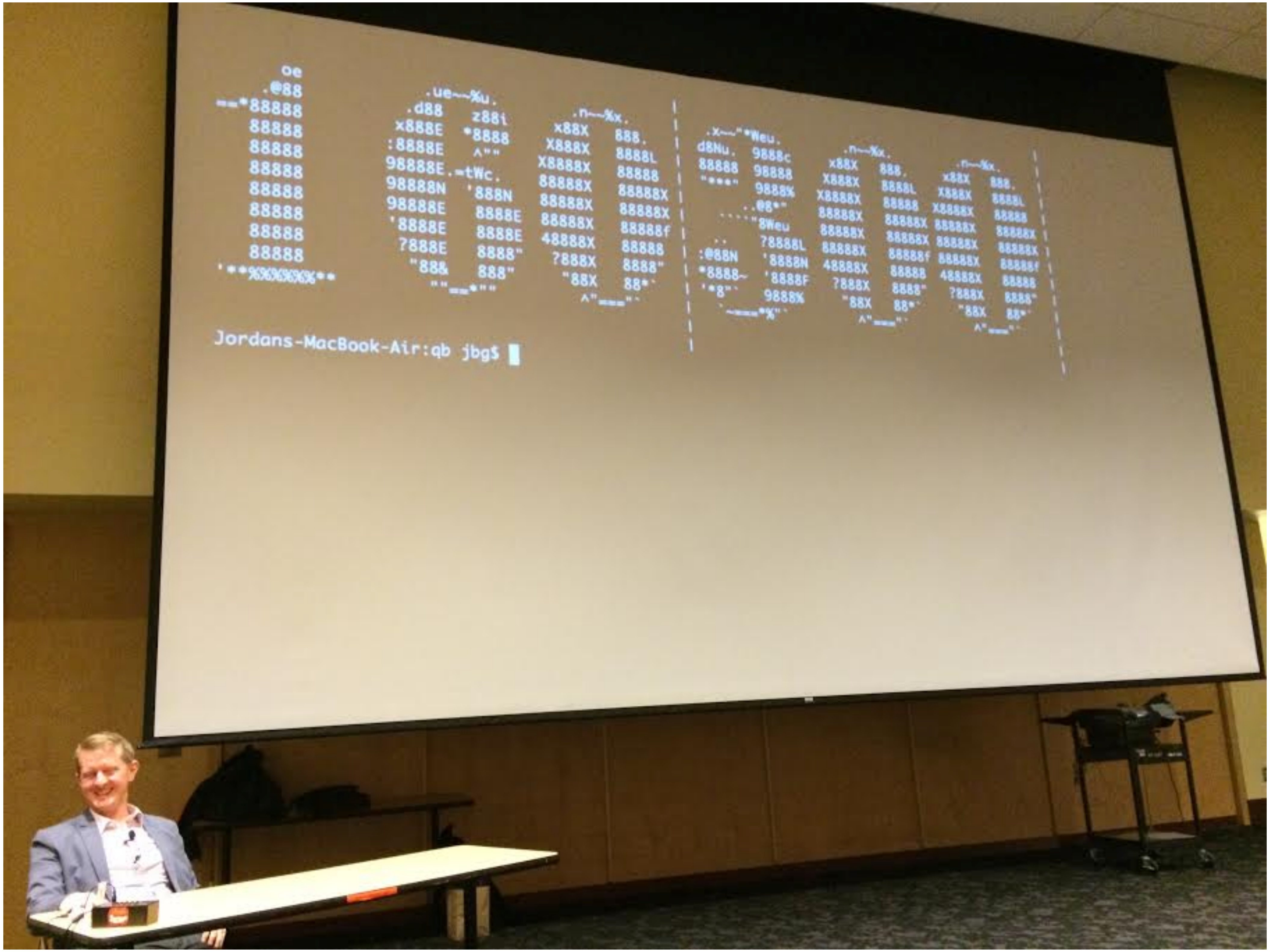
For 10 points, identify these bloodsucking beings most famously exemplified by Bram Stoker's Dracula.



$$av = \sum_{i=1}^{n} \frac{c_i}{n}$$

Of course, nowadays we would just put these questions into BERT and place a classifier over the [CLS] token to predict the answer!

2015: defeated Ken Jennings 300-160

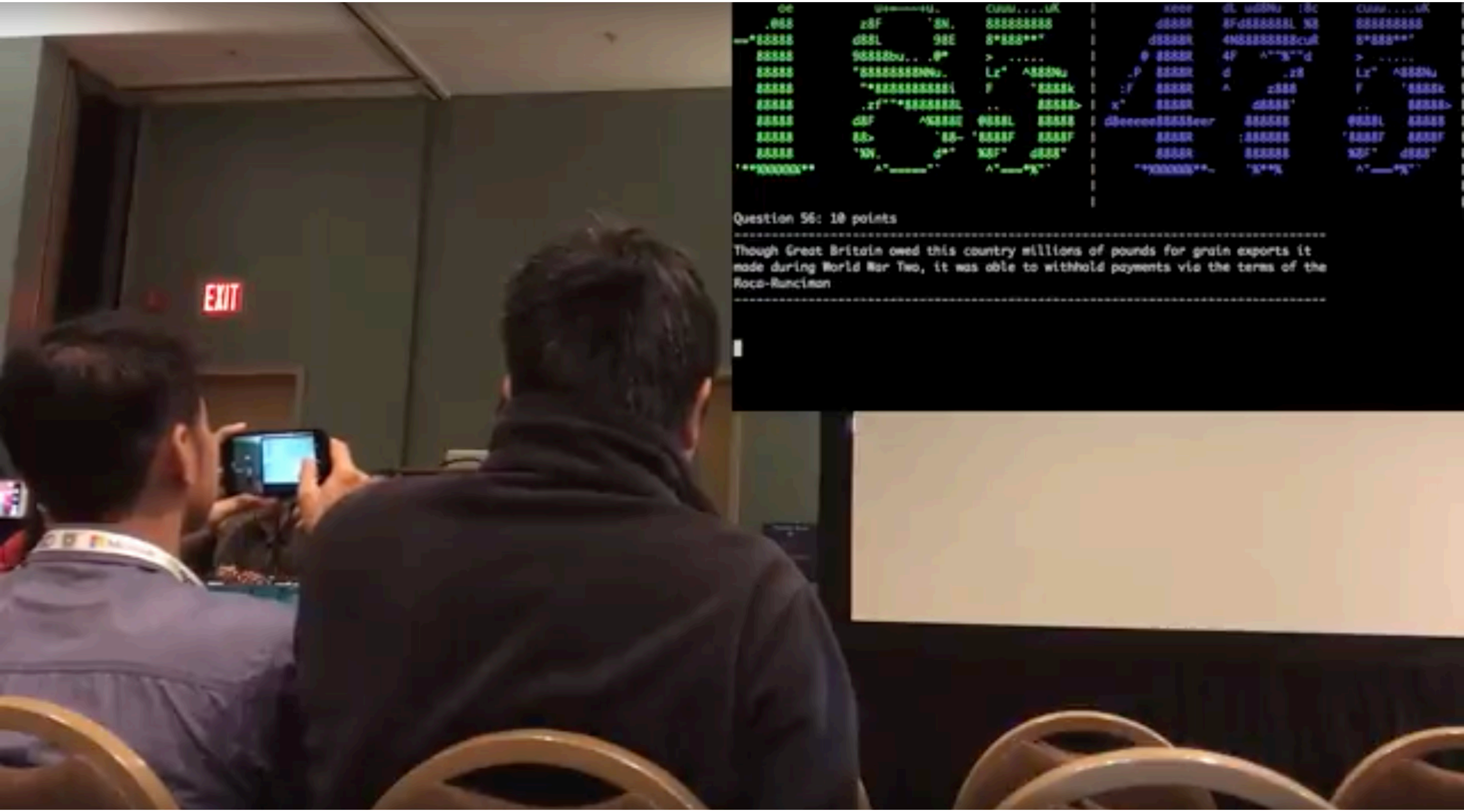


```
Jordans-MacBook-Air:qb jbg$
```

2016: lost to top quiz bowlers 345-145

2017: beat top quiz bowlers 260-215

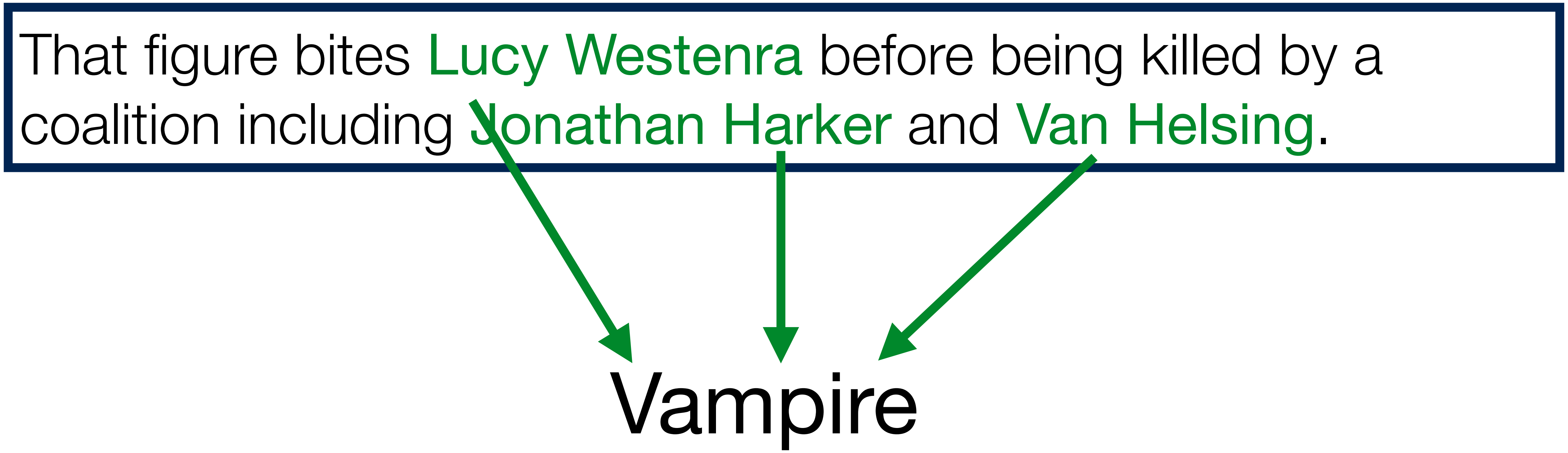
late 2017: crushed top team 475-185



```
Question 56: 10 points
--------------------------------------------------------------------------------
Though Great Britain owed this country millions of pounds for grain exports it
made during World War Two, it was able to withhold payments via the terms of the
Roca-Runciman
--------------------------------------------------------------------------------
```

deep learning ~ memorization

during training, QANTA becomes very good at associating named entities in questions with answers…



That figure bites Lucy Westenra before being killed by a coalition including Jonathan Harker and Van Helsing.

Vampire

deep learning ~ memorization

during training, QANTA becomes very good at associating named entities in questions with answers…

In one novel, one of these figures antagonizes an impoverished family before leaping into an active volcano.

???

These types of questions are still beyond the capabilities of our models