

Question answering & reading comprehension

CMSC 723 / LING 723 / INST 725

Hal Daumé III [he/him]

3 Dec 2019

(many slides c/o Marine Carpuat or Dan Jurafsky)

Announcements, logistics

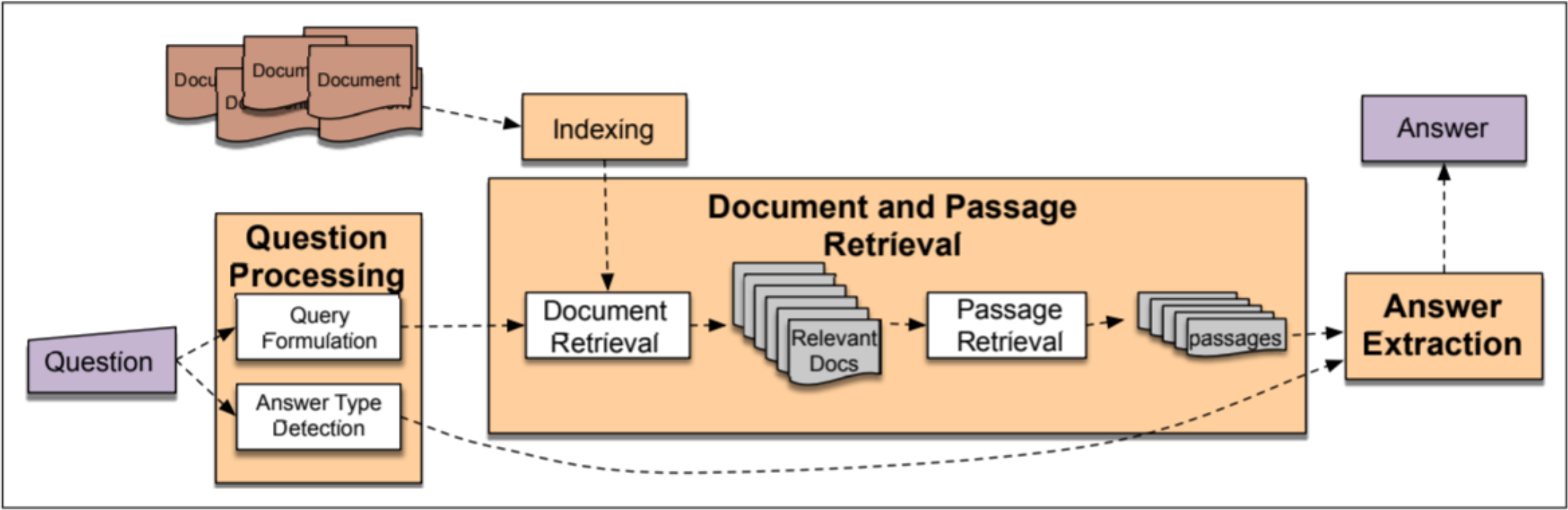
- Homework 4/5:
 - Everything out by now, deadline last day of classes (Dec 9)
 - Last part is EC
- Plan for P5 deadline/exam period
 - One poster per team
 - Students should spend ~50% of time at their own poster, ~50% of time looking at others
 - Hal/TAs will go quickly through each poster, so be sure to have a 4min spiel prepared
 - If for some reason you won't be there, please arrange with me by Dec 06

Today

- Question answering
- Reading comprehension

Factoid question answering

Question	Answer
Where is the Louvre Museum located?	in Paris, France
What's the abbreviation for limited partnership?	L.P.
What are the names of Odin's ravens?	Huginn and Muninn
What currency is used in China?	the yuan
What kind of nuts are used in marzipan?	almonds
What instrument does Max Roach play?	drums
What's the official language of Algeria?	Arabic
How many pounds are there in a stone?	14



TREC vs the web

Answer extraction

- Extract anything of the right “type”

“Who is the prime minister of India?”
Manmohan Singh, Prime Minister of India, had told left leaders that the deal would not be renegotiated.

“How tall is Mt. Everest?”
The official height of Mount Everest is 29029 feet

- Fails for many types

Tag	Example
ABBREVIATION	
abb	What’s the abbreviation for limited partnership?
exp	What does the “c” stand for in the equation E=mc2?
DESCRIPTION	
definition	What are tannins?
description	What are the words to the Canadian National anthem?
manner	How can you get rust stains out of clothing?
reason	What caused the Titanic to sink?
ENTITY	
animal	What are the names of Odin’s ravens?
body	What part of your body contains the corpus callosum?
color	What colors make up a rainbow?
creative	In what book can I find the story of Aladdin?
currency	What currency is used in China?
disease/medicine	What does Salk vaccine prevent?
event	What war involved the battle of Chapultepec?
food	What kind of nuts are used in marzipan?
instrument	What instrument does Max Roach play?
lang	What’s the official language of Algeria?
letter	What letter appears on the cold-water tap in Spain?
other	What is the name of King Arthur’s sword?
plant	What are some fragrant white climbing roses?
product	What is the fastest computer?
religion	What religion has the most members?
sport	What was the name of the ball game played by the Mayans?
substance	What fuel do airplanes use?
symbol	What is the chemical symbol for nitrogen?
technique	What is the best way to remove wallpaper?
term	How do you say “ Grandma” in Irish?
vehicle	What was the name of Captain Bligh’s ship?
word	What’s the singular of dice?
HUMAN	
description	Who was Confucius?
group	What are the major companies that are part of Dow Jones?
ind	Who was the first Russian astronaut to do a spacewalk?
title	What was Queen Victoria’s title regarding India?
LOCATION	
city	What’s the oldest capital city in the Americas?
country	What country borders the most others?
mountain	What is the highest peak in Africa?
other	What river runs through Liverpool?
state	What states do not have state income tax?
NUMERIC	
code	What is the telephone number for the University of Colorado?
count	About how many soldiers died in World War II?
date	What is the date of Boxing Day?
distance	How long was Mao’s 1930s Long March?
money	How much did a McDonald’s hamburger cost in 1963?
order	Where does Shanghai rank among world cities in population?
other	What is the population of Mexico?
period	What was the average life expectancy during the Stone Age?
percent	What fraction of a beaver’s life is spent swimming?
temp	How hot should the oven be when making Peachy Oat Muffins?
speed	How fast must a spacecraft travel to escape Earth’s gravity?
size	What is the size of Argentina?
weight	How many pounds are there in a stone?

Feature-based supervised learning

- Answer-type match
- Pattern match

Pattern	Question	Answer
<AP> such as <QP>	What is autism?	“, <u>developmental disorders</u> such as autism”
<QP>, a <AP>	What is a caldera?	“the Long Valley caldera, a <u>volcanic crater</u> 19 miles long”

- # of matched question keywords / distance
- Novelty factor
- Apposition features
- Punctuation location

From QA to Reading Comprehension

Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in **Houston, Texas**, she performed in various **singing and dancing** competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, *Dangerously in Love* (**2003**), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".

Q: "In what city and state did Beyoncé grow up?"

A: "**Houston, Texas**"

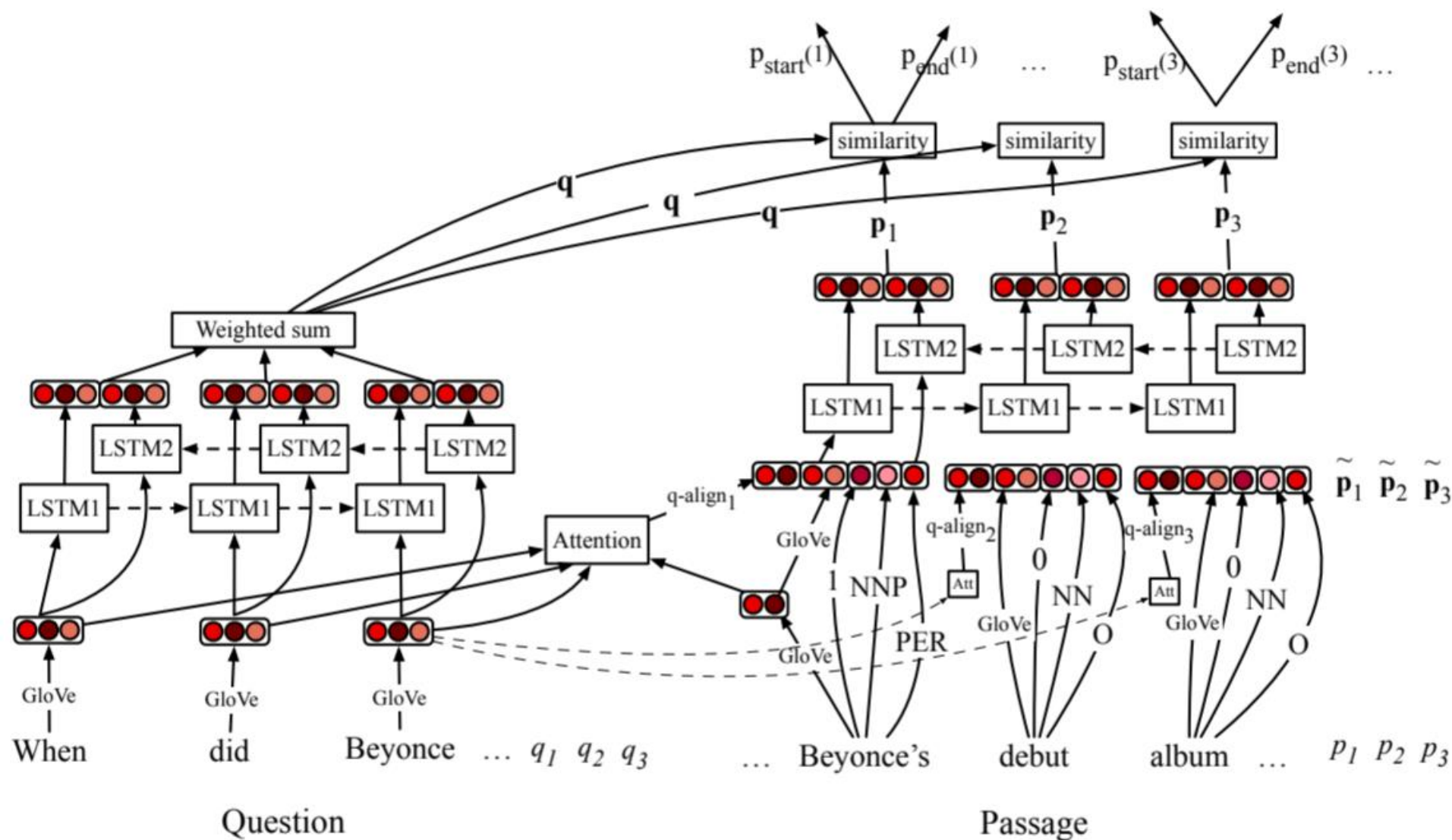
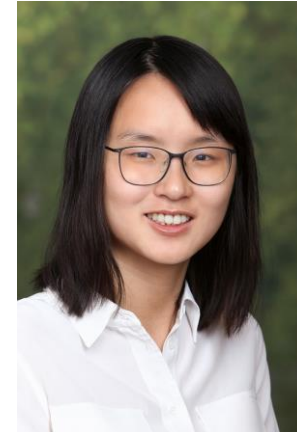
Q: "What areas did Beyoncé compete in when she was growing up?"

A: "**singing and dancing**"

Q: "When did Beyoncé release *Dangerously in Love*?"

A: "**2003**"

Neural reading comprehension



Reading comprehension with BERT...

Side-note on BERT, ELMO, etc...



“Word2vec is the sriracha
sauce of deep learning!”
– Chris Manning





Words mean different things
in different contexts



Structural learning from 2005 to
today

Large “free” text corpora provide training data

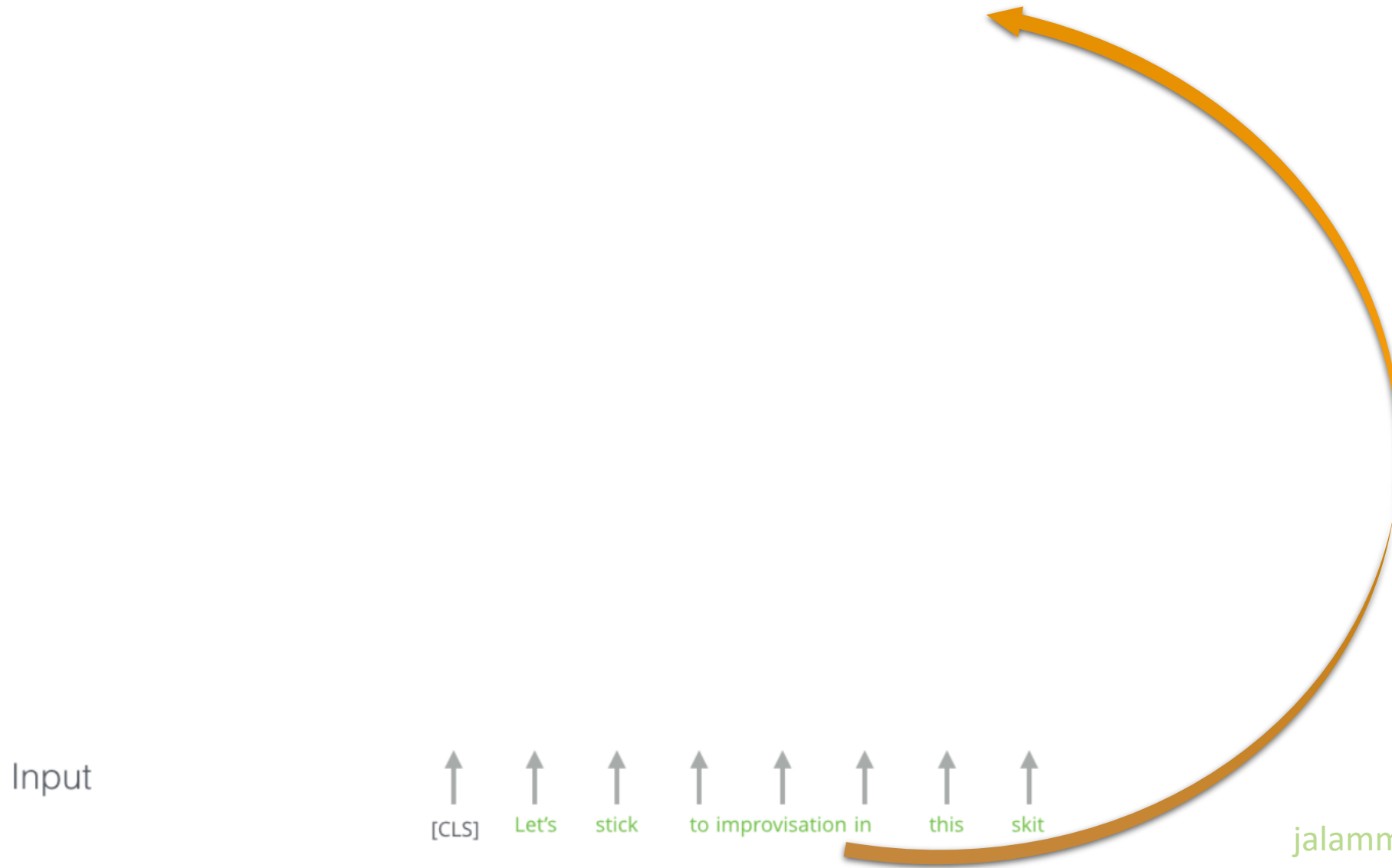
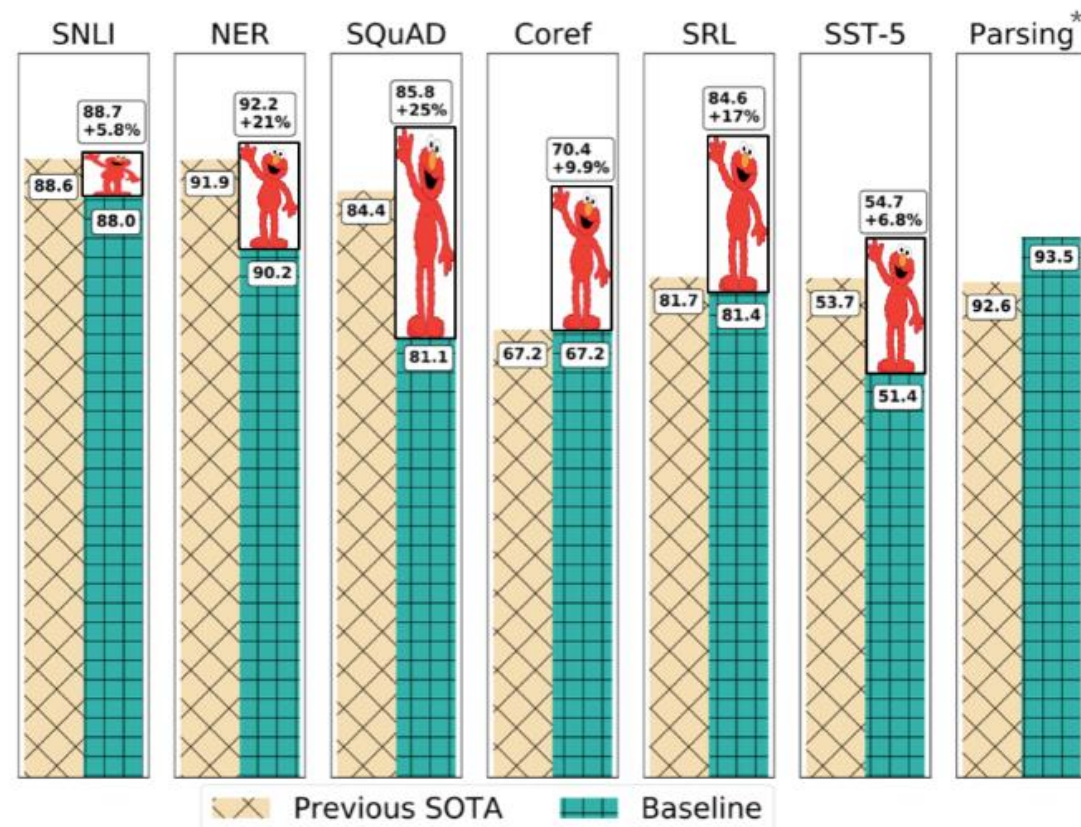
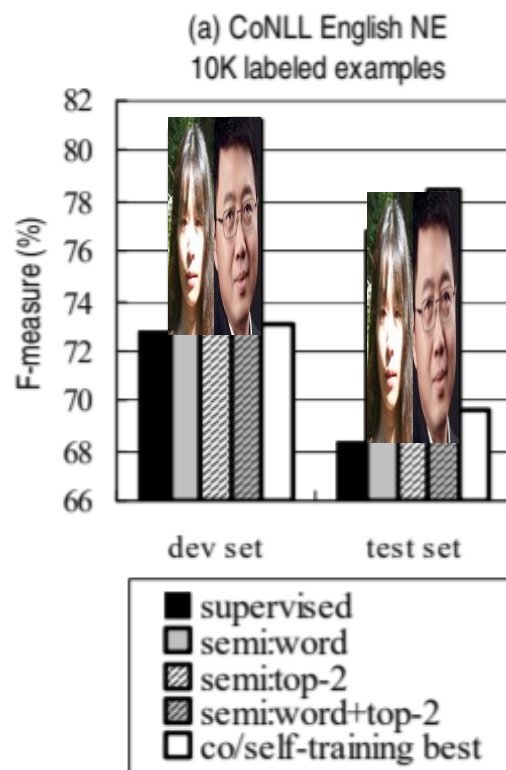
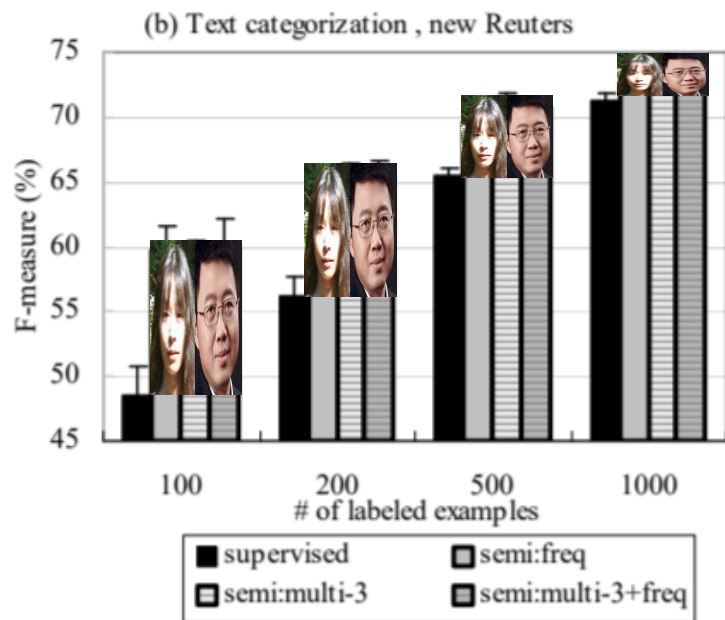


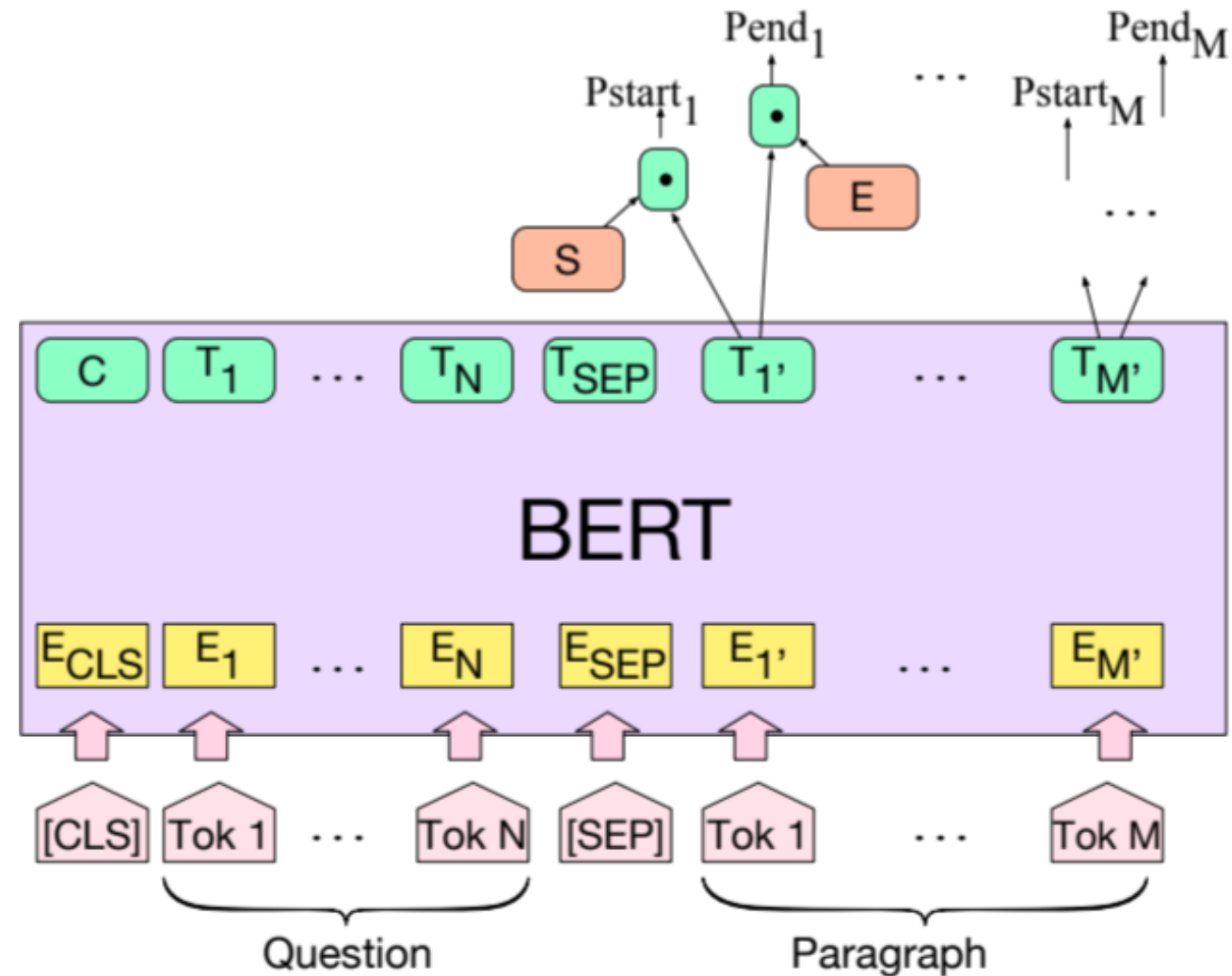
Image credit: Jay Alammam
jalammam.github.io/illustrated-bert



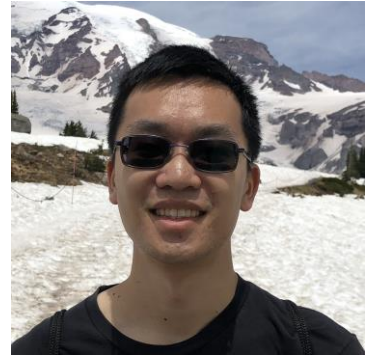
*Kitaev and Klein, ACL 2018 (see also Joshi et al., ACL 2018)

Measurable improvements across many tasks

BERT-based Reading Comprehension



Breaking Reading Comprehension



Article: Super Bowl 50

Paragraph: *“Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver’s Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.”*

Question: *“What is the name of the quarterback who was 38 in Super Bowl XXXIII?”*

Original Prediction: John Elway

Prediction under adversary: Jeff Dean

AddSent and AddAny

Article: **Nikola Tesla**

Paragraph: "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 enroll 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."

Question: "What city did Tesla move to in 1880?"
Answer: *Prague*
Model Predicts: *Prague*

AddAny

Randomly initialize d words:

spring attention income getting reached

↓ Greedily change one word

spring attention income other reached

↓ Repeat many times

Adversary Adds: **tesla move move other george**

Model Predicts: *george*

AddSent

What city did *Tesla* move to in 1880?

Prague

↓ (Step 1)
Mutate
question

↓ (Step 2)
Generate
fake answer

What city did *Tadakatsu* move to in 1881?

Chicago

↓ (Step 3)
Convert into
statement

Tadakatsu moved the city of Chicago to in 1881.

↓ (Step 4)
Fix errors with
crowdworkers,
verify resulting
sentences with
other crowdworkers

Adversary Adds: **Tadakatsu moved to the city of Chicago in 1881.**

Model Predicts: *Chicago*

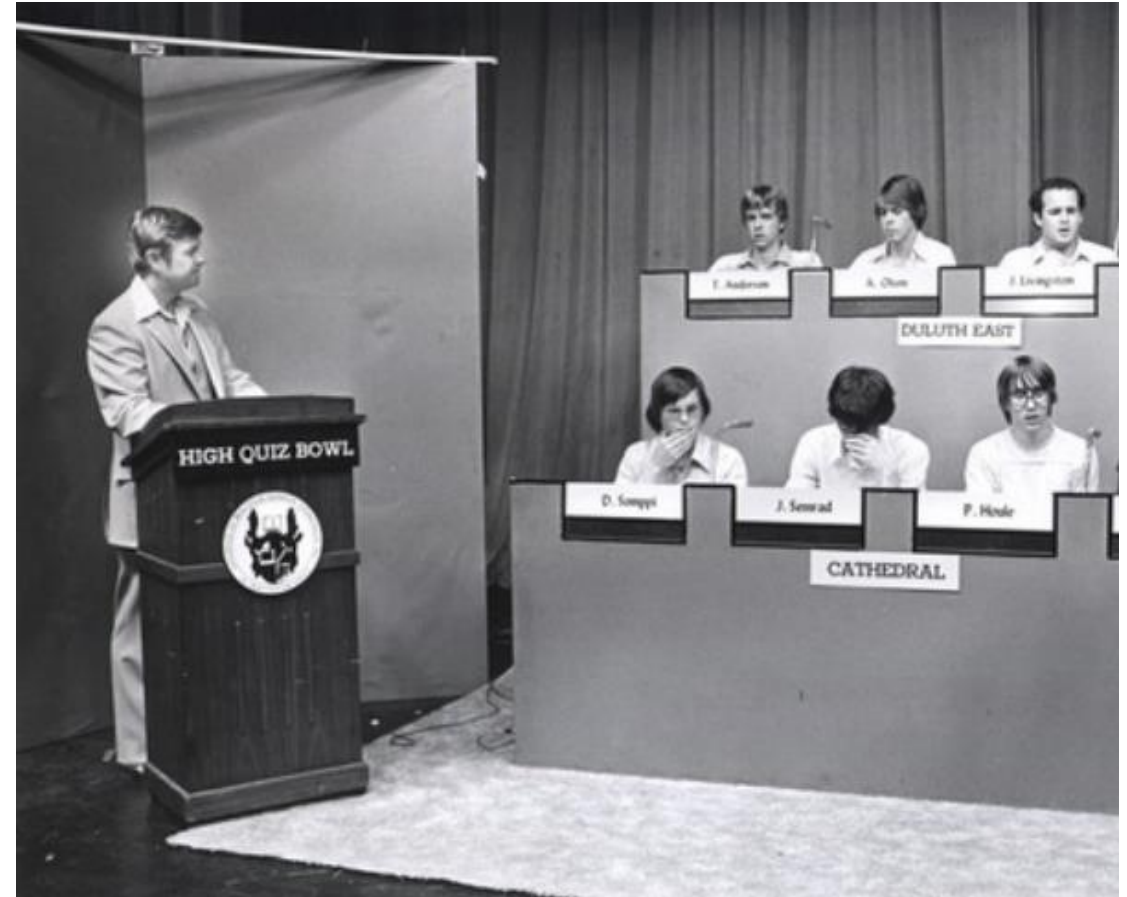
Impact on system performance

Model	Original	ADDSSENT	ADDONESSENT
ReasoNet-E	81.1	39.4	49.8
SEDT-E	80.1	35.0	46.5
BiDAF-E	80.0	34.2	46.9
Mnemonic-E	79.1	46.2	55.3
Ruminating	78.8	37.4	47.7
jNet	78.6	37.9	47.0
Mnemonic-S	78.5	46.6	56.0
ReasoNet-S	78.2	39.4	50.3
MPCM-S	77.0	40.3	50.0
SEDT-S	76.9	33.9	44.8
RaSOR	76.2	39.5	49.5
BiDAF-S	75.5	34.3	45.7
Match-E	75.4	29.4	41.8
Match-S	71.4	27.3	39.0
DCR	69.3	37.8	45.1
Logistic	50.4	23.2	30.4

Other question answering settings

Quizbowl

- Game called “quiz bowl”
- Two teams play each other
- Moderator reads a question
- When a team knows the answer, they buzz in
- **If right**, they get points; **otherwise**, rest of the question is read to the other team
- Hundreds of teams in the US alone
- Example . . .



Quizbowl example

- With Leo Szilard, he invented a doubly-eponymous

Quizbowl example

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory

Quizbowl example

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so

Quizbowl example

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients

Quizbowl example

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by

Quizbowl example

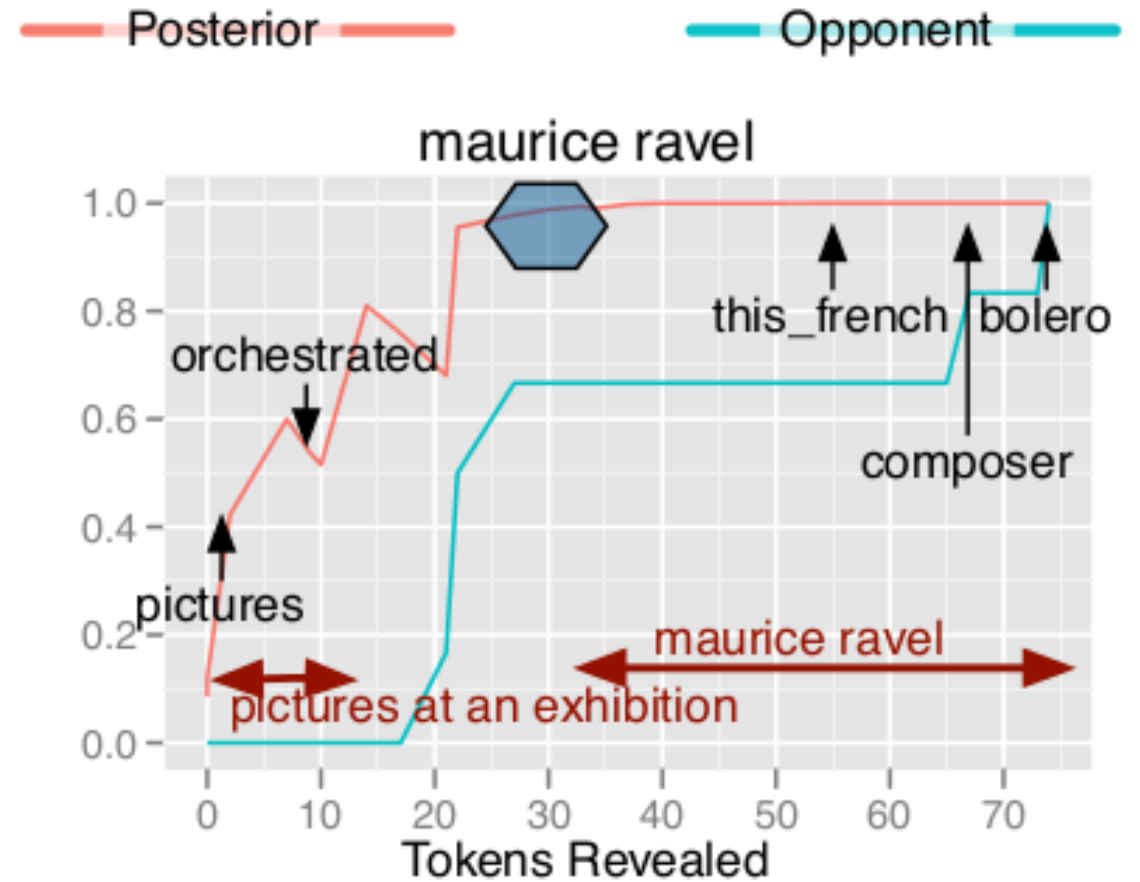
- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating

Quizbowl example

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating the special and general theories of relativity?

Evaluation methodology

- Mechanical Turk to collect human data
- 7000 questions were answered in the first day
- Over 43000 questions were answered in the space of two weeks
- Total of 461 unique users
- Leaderboard to encourage users



Prediction



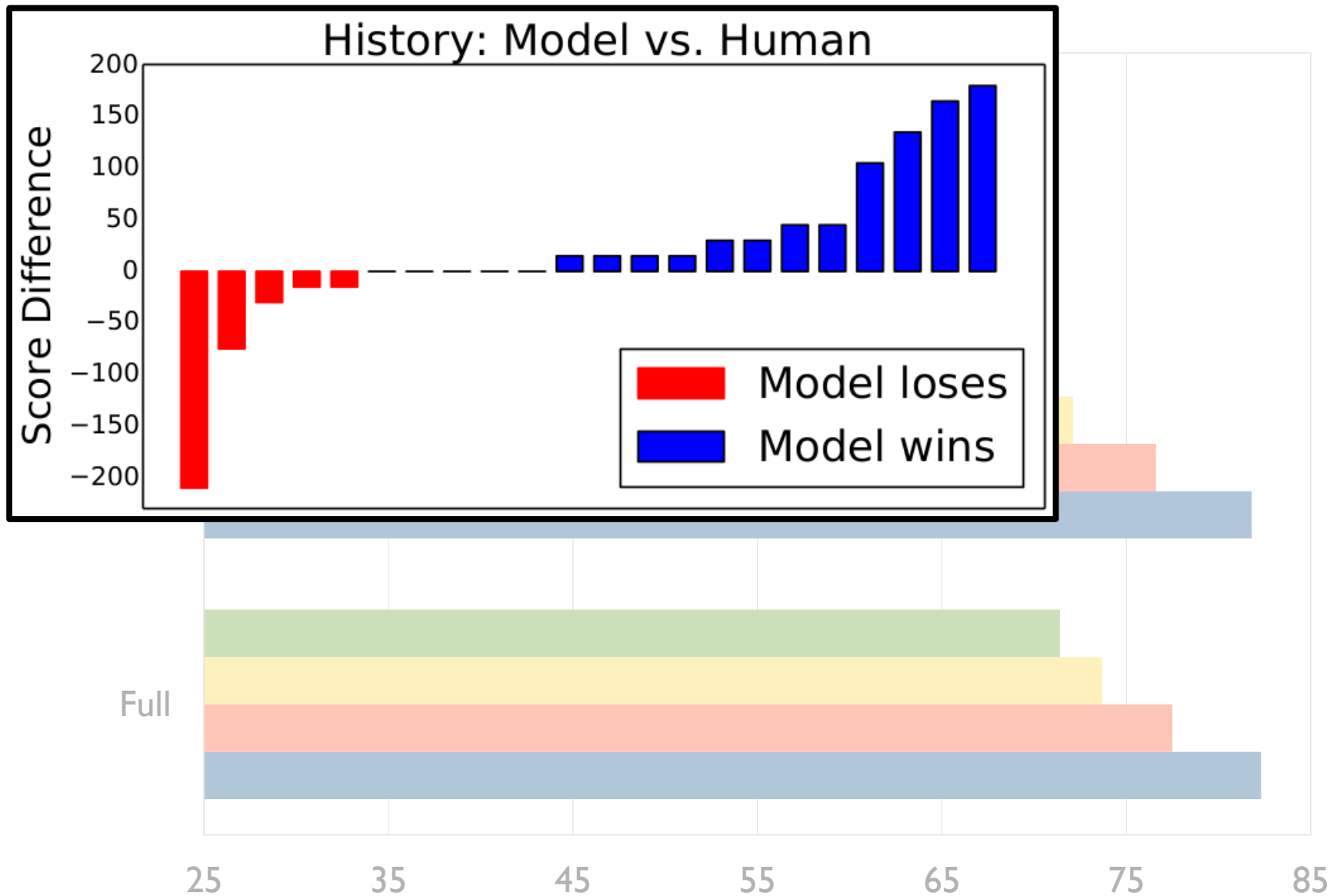
Buzz



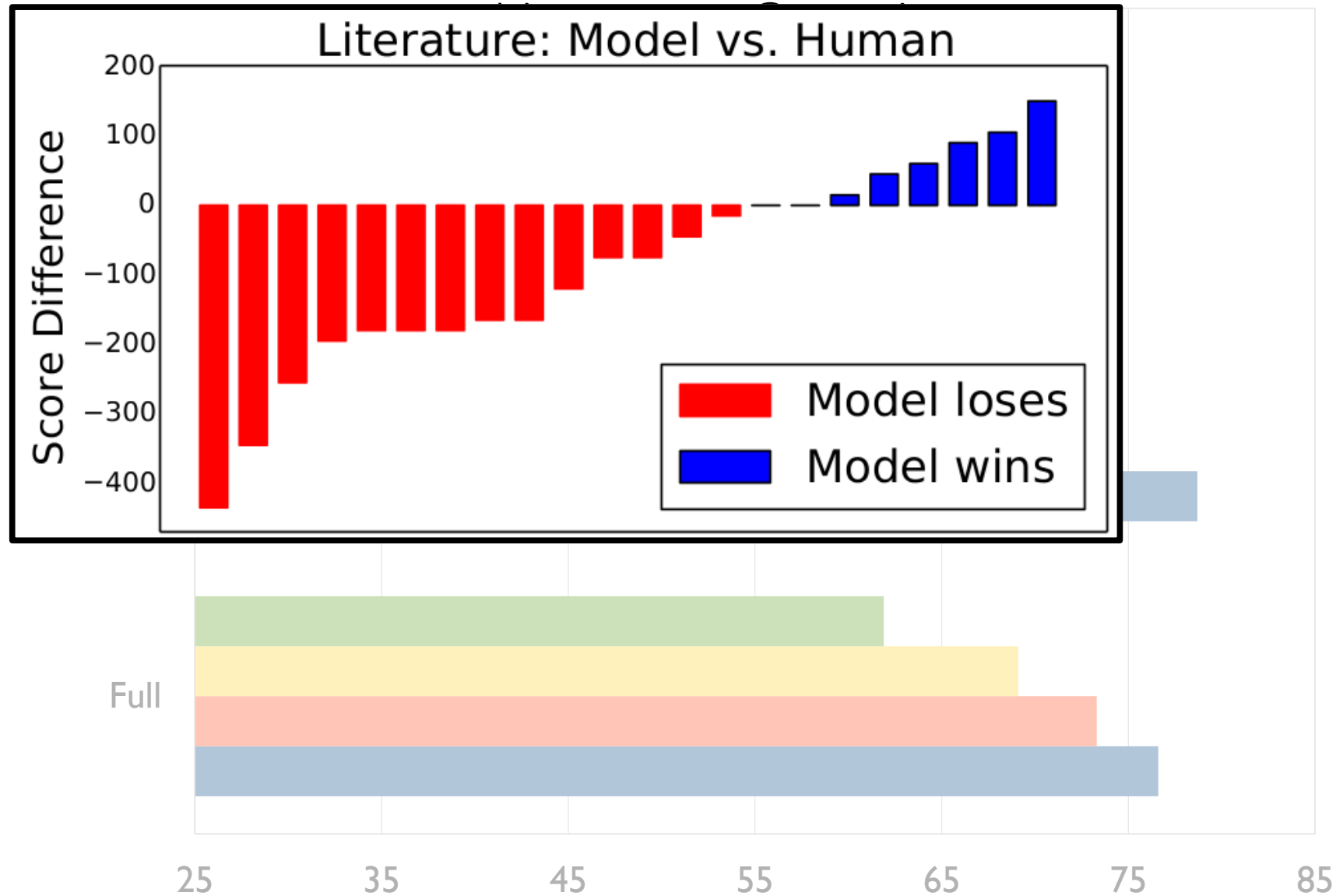
Observation



Results on question-answering task



Results on question-answering task



But the true test...



But the true test... RESULTS!



Question generation



“No computer is ever going to ask a new, reasonable question. It takes trained people to do that.”

Grace Hopper

A Computer Scientist and United States Navy
Rear Admiral



Question generation

- Reading comprehension question generation

vs

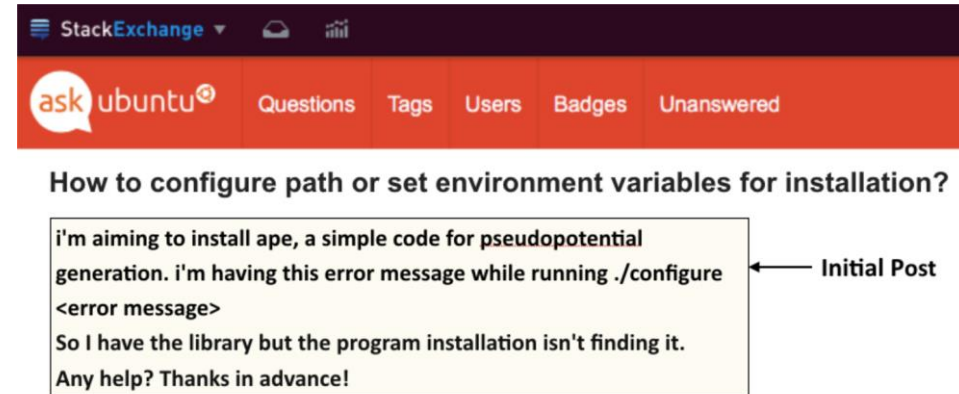
- Clarification question generation

dataset and modeling insight

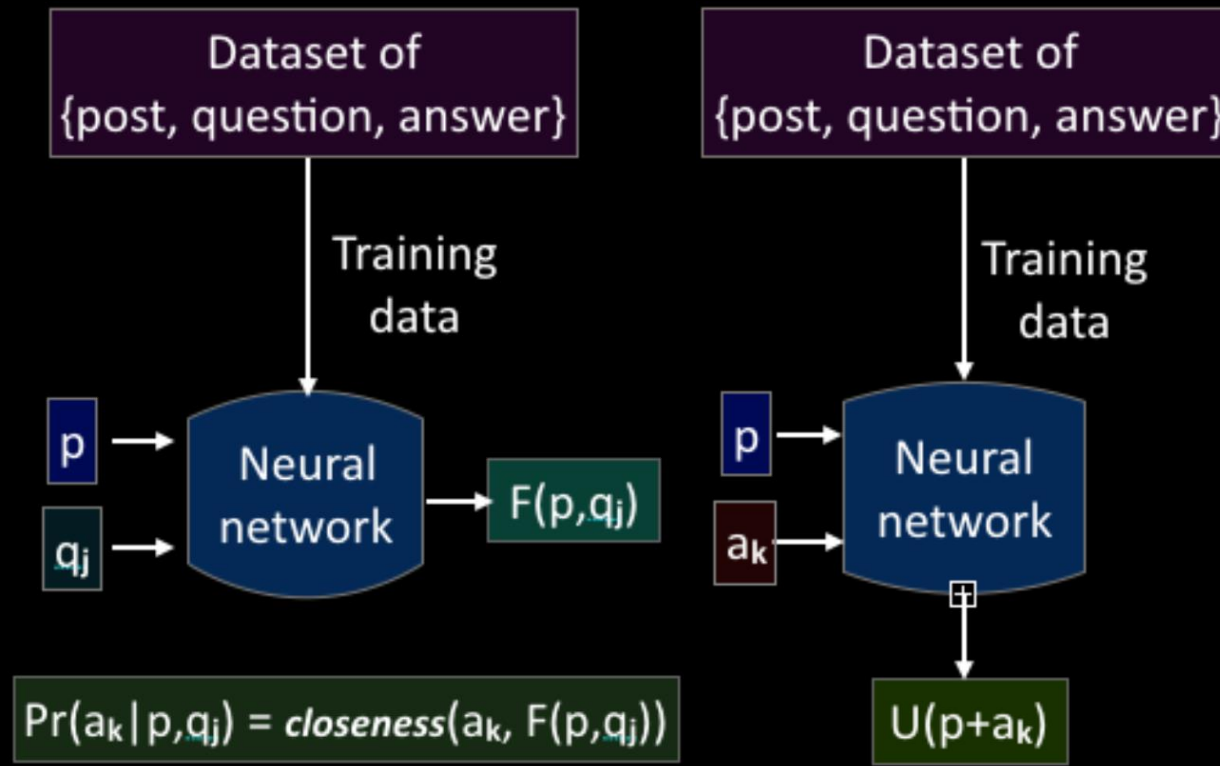
- There *is* such a thing as a bad question!

- a) What version of Ubuntu do you have?
- b) Is the moon waxing or waning?
- c) Are you running Ubuntu 14.10 kernel 4.4.0-59-generic on an x86 64 architecture?

- Post/clarification/answer triples from S.E. (via edit history)

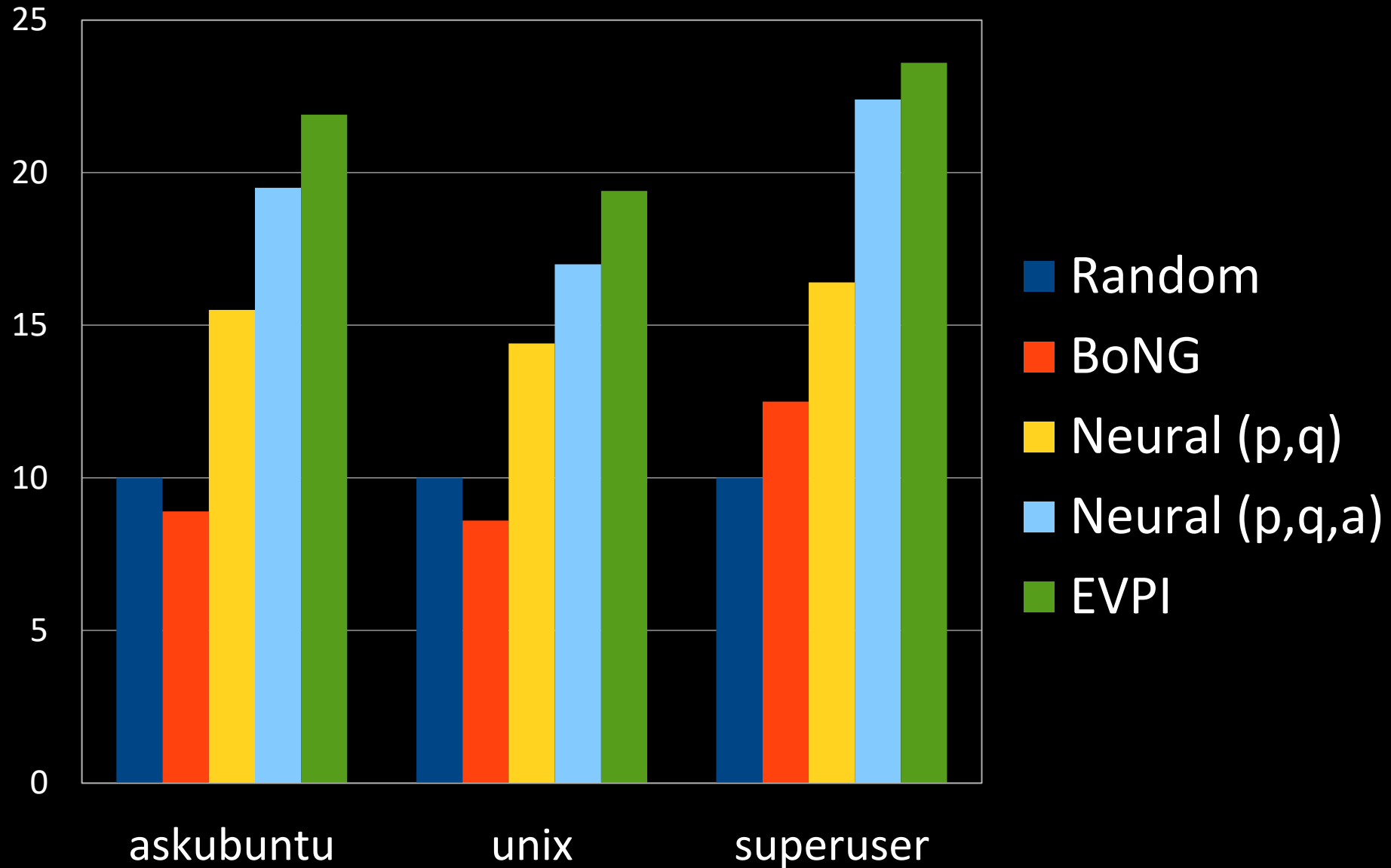


expected value of perfect information



$$\arg \max_{q \in Q} \sum_{a \in A} \mathbb{P}[a | p, q] U(p + a)$$

results: accuracy



interactive question *asking*

START

tell me more
about **BOOK**

tell me more
about **SECTION**

1st Epistle to
the Corinthians?



Question: The final section of this book includes the Aramaic phrase "anathema marana tha." Its first section quotes Isaiah 29 and says that "God chose the foolish things of the world to shame the wise. This book ends a discussion on the qualities that define love with the lines "And now these three remain: faith, hope, and love." But the greatest of these is love." For 10 points, name this Pauline epistle addressed to a Greek city.

ANSWER: First Epistle to the Corinthians [require both parts; prompt on partial answers]

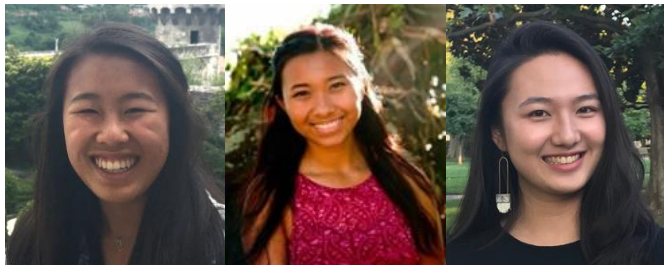
Question: The final section of this book includes the Aramaic phrase "anathema marana tha." Its first section quotes Isaiah 29 and says that "God chose the foolish things of the world to shame the wise. This **book** ends a discussion on the qualities that define love with the lines "And now these three remain: faith, hope, and love." But the greatest of these is love." For 10 points, name this Pauline epistle addressed to a Greek city.

ANSWER: First Epistle to the Corinthians [require both parts; prompt on partial answers]

Question: The final section of this book includes the Aramaic phrase "anathema marana tha." Its first **section** quotes Isaiah 29 and says that "God chose the foolish things of the world to shame the wise. This **book** ends a discussion on the qualities that define love with the lines "And now these three remain: faith, hope, and love." But the greatest of these is love." For 10 points, name this Pauline epistle addressed to a Greek city.

ANSWER: First Epistle to the Corinthians [require both parts; prompt on partial answers]

CORRECT !



Overall

- Question answering vs reading comprehension
- BERT
- Different styles of QA