

End-to-end neural language understanding

- Instead of decomposing the problem into its parts (coreference, semantic role labeling), can we teach a neural network to understand texts using a large collection of inputs and output?
- Natural language inference
- Question answering (“machine reading comprehension”)

Natural Language Inference

Premise

A soccer game with multiple males playing.

Some men are playing a sport.

Entailment

Hypothesis

A man inspects the uniform of a figure.

The man is sleeping.

Contradiction

How can we generate a large number of examples?

(Dagan et al., 2006; Bowman et al., 2015)

Crowdsourcing: generation

This task will involve reading a line from a non-fiction article and writing three sentences that relate to it. The line will describe a situation or event. Using only this description and what you know about the world:

- Write one sentence that is definitely correct about the situation or event in the line.
- Write one sentence that might be correct about the situation or event in the line.
- Write one sentence that is definitely incorrect about the situation or event in the line.

(Williams et al., 2018)

Crowdsourcing: validation

Met my first girlfriend that way.	FACE-TO-FACE contradiction C C N C	I didn't meet my first girlfriend until later.
8 million in relief in the form of emergency housing.	GOVERNMENT neutral N N N N	The 8 million dollars for emergency housing was still not enough to solve the problem.
Now, as children tend their gardens, they have a new appreciation of their relationship to the land, their cultural heritage, and their community.	LETTERS neutral N N N N	All of the children love working in their gardens.
At 8:34, the Boston Center controller received a third transmission from American 11	9/11 entailment E E E E	The Boston Center controller got a third transmission from American 11.
I am a lacto-vegetarian.	SLATE neutral N N E N	I enjoy eating cheese too much to abstain from dairy.
someone else noticed it and i said well i guess that's true and it was somewhat melodious in other words it wasn't just you know it was really funny	TELEPHONE contradiction C C C C	No one noticed and it wasn't funny at all.

(Williams et al., 2018)

MultiNLI corpus stats

Genre	#Examples		
	Train	Dev.	Test
<i>SNLI</i>	<i>550,152</i>	<i>10,000</i>	<i>10,000</i>
FICTION	77,348	2,000	2,000
GOVERNMENT	77,350	2,000	2,000
SLATE	77,306	2,000	2,000
TELEPHONE	83,348	2,000	2,000
TRAVEL	77,350	2,000	2,000
9/11	0	2,000	2,000
FACE-TO-FACE	0	2,000	2,000
LETTERS	0	2,000	2,000
OUP	0	2,000	2,000
VERBATIM	0	2,000	2,000
MultiNLI Overall	392,702	20,000	20,000

(Williams et al., 2018)

The Stanford Question Answering Dataset (SQuAD): extractive question answering

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

What causes precipitation to fall?

gravity

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

graupel

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

within a cloud

(23215 paragraphs
from top-ranked
Wikipedia pages)

(Rajpurkar et al., 2016)

The Stanford Question Answering Dataset (SQuAD): extractive question answering

Paragraph 1 of 43

Spend around 4 minutes on the following paragraph to ask 5 questions! If you can't ask 5 questions, ask 4 or 3 (worse), but do your best to ask 5. Select the answer from the paragraph by clicking on 'Select Answer', and then highlight the smallest segment of the paragraph that answers the question.

Oxygen is a chemical element with symbol O and atomic number 8. It is a member of the chalcogen group on the periodic table and is a highly reactive nonmetal and oxidizing agent that readily forms compounds (notably oxides) with most elements. By mass, oxygen is the third-most abundant element in the universe, after hydrogen and helium. At standard temperature and pressure, two atoms of the element bind to form dioxygen, a colorless and odorless diatomic gas with the formula O₂.

2. Diatomic oxygen gas constitutes 20.8% of the Earth's atmosphere. However, monitoring of atmospheric oxygen levels show a global downward trend, because of fossil-fuel burning. Oxygen is the most abundant element by mass in the Earth's crust as part of oxide compounds such as silicon dioxide, making up almost half of the crust's mass.

When asking questions, **avoid using** the same words/phrases as in the paragraph. Also, you are encouraged to pose **hard questions**.

Ask a question here. Try using your own words

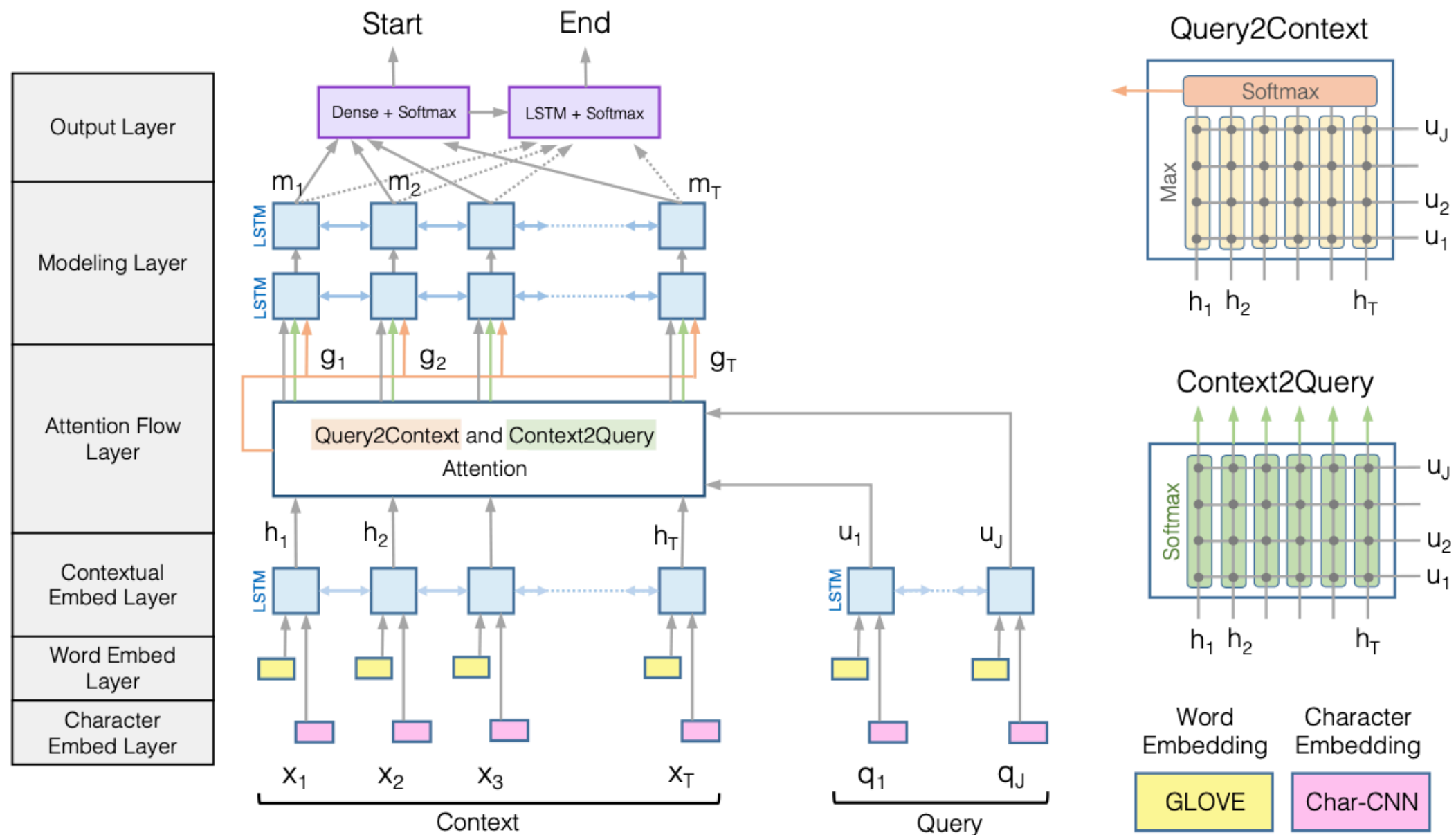
Select Answer

Ask a question here. Try using your own words

Select Answer

(Rajpurkar et al., 2016)

“Bidirectional Attention Flow” (BiDAF) for extractive question answering



(Seo et al., 2017)

What do end-to-end systems really learn? Adversarial examples

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

What do end-to-end systems really learn? Adversarial examples

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**

(Jia et al., 2017)

What do end-to-end systems really learn? Adversarial examples

	Match Single	Match Ens.	BiDAF Single	BiDAF Ens.
Original	71.4	75.4	75.5	80.0
ADDSSENT	27.3	29.4	34.3	34.2
ADDONESSENT	39.0	41.8	45.7	46.9
ADDANY	7.6	11.7	4.8	2.7
ADDCOMMON	38.9	51.0	41.7	52.6

SQuAD 2.0

Unanswerable questions with plausible (but incorrect) answers:

Article: Endangered Species Act

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

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

Plausible Answer: *later laws*

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

Plausible Answer: *Bald Eagle Protection Act*

SQuAD 2.0

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance <i>Stanford University</i> (Rajpurkar & Jia et al. '18)	86.831	89.452
1 <div>Apr 06, 2020</div>	SA-Net on Albert (ensemble) <i>QIANXIN</i>	90.724	93.011
2 <div>May 05, 2020</div>	SA-Net-V2 (ensemble) <i>QIANXIN</i>	90.679	92.948
2 <div>Apr 05, 2020</div>	Retro-Reader (ensemble) <i>Shanghai Jiao Tong University</i> http://arxiv.org/abs/2001.09694	90.578	92.978
3 <div>Jul 31, 2020</div>	ATRLP+PV (ensemble) <i>Hithink RoyalFlush</i>	90.442	92.877

“Multi-hop” reasoning

Paragraph A, Return to Olympus:

[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

Paragraph B, Mother Love Bone:

[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood’s personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band’s debut album, “Apple”, thus ending the group’s hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of “Apple”?

A: Malfunkshun

Supporting facts: 1, 2, 4, 6, 7

(Yang et al 2018)

Natural Questions

- Questions typed into the Google search engine + a Wikipedia page that Google returned as a result (Kwiatkowski et al. 2019)
- Annotators identify a paragraph in the Wikipedia page that contains the answer (or say the answer isn't in the page), and identify the answer
- The question is natural in that it wasn't generated based on a passage
- Questions may be repetitive and memorizable (Lewis et al 2020)

Natural questions are often underdetermined (> 50% of the time!)



AMBIGQA Input

When did harry potter and the sorcerer's stone movie come out?

Harry Potter and the Philosopher's Stone (film)

From Wikipedia, the free encyclopedia

The film had its world premiere at the Odeon Leicester Square in London on 4 November 2001, with the cinema arranged to resemble Hogwarts School. (...) The film was released to cinemas in the United Kingdom and United States on 16 November 2001.

AMBIGQA Output

Q: When did harry potter and the sorcerer's stone movie come out at the Odeon Leicester Square?

A: 4 November 2001

Q: When did harry potter and the sorcerer's stone movie come out in cinemas?

A: 16 November 2001

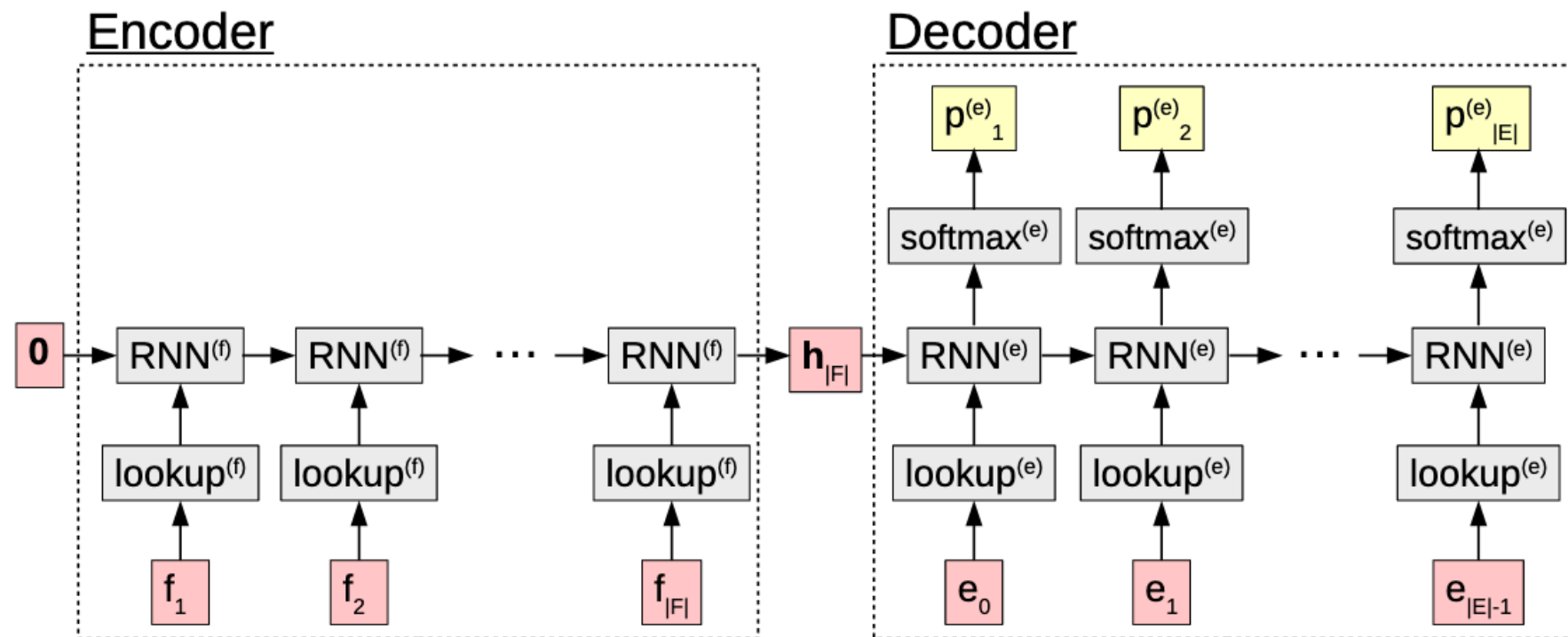


(Min et al 2020)

Challenges in end-to-end question answering

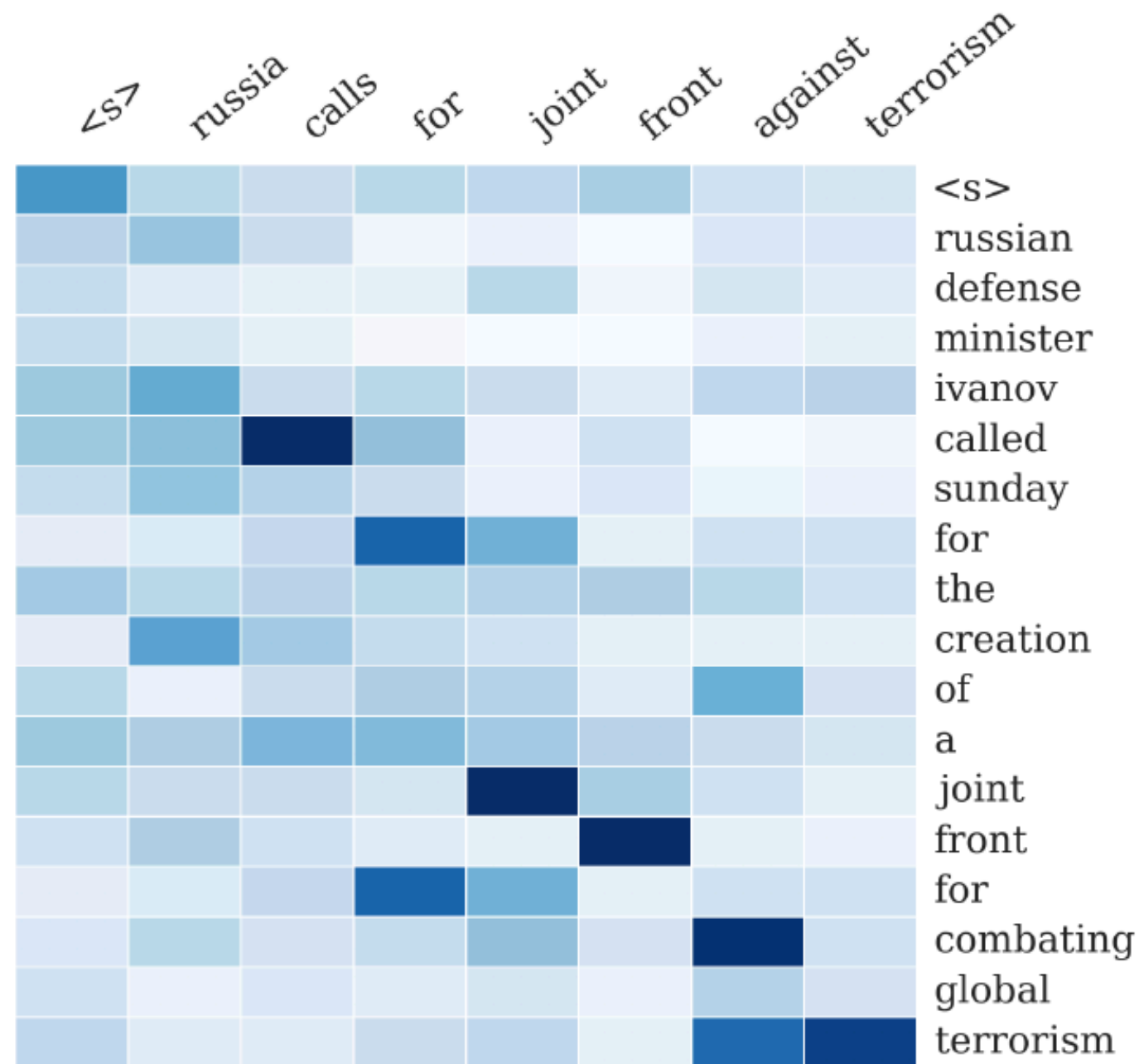
- Do the datasets really measure general question answering abilities?
- Generalization across data-sets (Talmor et al 2019)
- Simple baselines often perform too well (Chen and Durrett 2019)
- Extending to open-domain question answering system that need to also retrieve the document

End-to-end neural language generation



- Summarization: document \rightarrow summary
- Dialog: history \rightarrow response
- Captioning: image \rightarrow caption...

Neural summarization



(Rush et al 2015)

Making neural summarization more extractive

Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, *muhammadu buhari* told cnn's christiane amannpour that **he plans to aggressively fight corruption that has long plagued nigeria** and go after the root of the nation's unrest. *buhari* said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, **he said his administration is confident it will be able to thwart criminals** and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. *buhari* defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. **the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.**

Baseline Seq2Seq + Attention: UNK UNK says his administration is confident it will be able to **destabilize nigeria's economy**. UNK says his administration is confident it will be able to thwart criminals and other **nigerians**. **he says the country has long nigeria and nigeria's economy.**

Pointer-Gen: *muhammadu buhari* says he plans to aggressively fight corruption **in the northeast part of nigeria**. he says he'll "rapidly give attention" to curbing violence **in the northeast part of nigeria**. he says his administration is confident it will be able to thwart criminals.

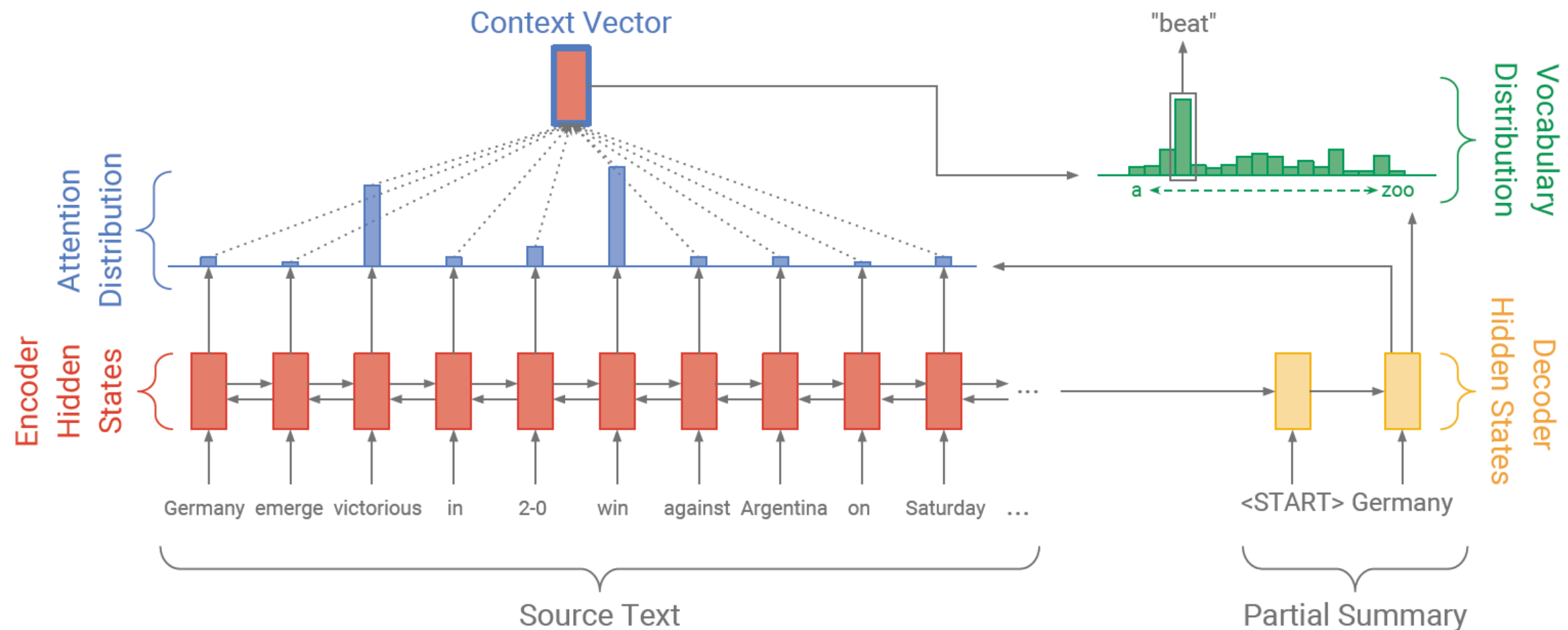
Pointer-Gen + Coverage: *muhammadu buhari* says he plans to aggressively fight corruption that has long plagued nigeria. he says his administration is confident it will be able to thwart criminals. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.

Factual
errors

Repetitions

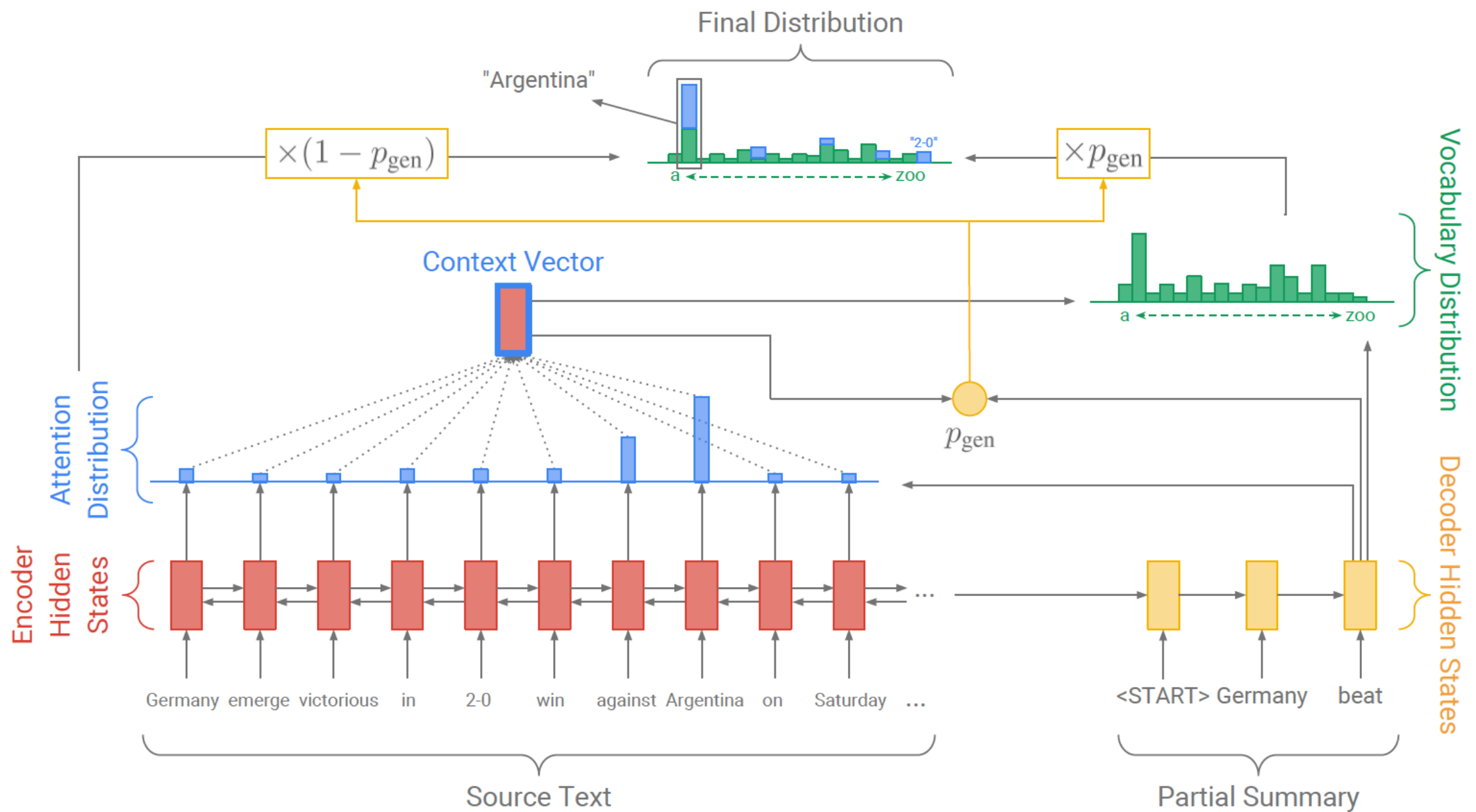
(See et al
2017)

Copying for summarization



(See et al 2017)

Copying for summarization



(Gu et al 2016; See et al 2017)

Coverage

Explicitly remind the model of all its previous attention vectors:

$$c_t = \sum_{t'=0}^{t-1} a^{t'}$$

$$e_i^t = v^T \tanh(W_h h_i + W_s s_t + w_c c_t^t + b)$$

Enforce divergence between the current attention vector and the sum of the previous ones:

$$\text{covloss} = \sum_i \min(a_i^t, c_i^t)$$

(Gu et al 2016; See et al 2017)

Copying for summarization

Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, *muhammadu buhari* told cnn's christiane amannpour that **he plans to aggressively fight corruption that has long plagued nigeria** and go after the root of the nation's unrest. *buhari* said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, **he said his administration is confident it will be able to thwart criminals** and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. *buhari* defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. **the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.**

Baseline Seq2Seq + Attention: UNK UNK says his administration is confident it will be able to **destabilize nigeria's economy**. UNK says his administration is confident it will be able to thwart criminals and other **nigerians**. **he says the country has long nigeria and nigeria's economy.**

Pointer-Gen: *muhammadu buhari* says he plans to aggressively fight corruption **in the northeast part of nigeria**. he says he'll "rapidly give attention" to curbing violence **in the northeast part of nigeria**. he says his administration is confident it will be able to thwart criminals.

Pointer-Gen + Coverage: *muhammadu buhari* says he plans to aggressively fight corruption that has long plagued nigeria. he says his administration is confident it will be able to thwart criminals. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.

(See et al
2017)

“Hallucination” in state-of-the-art summarization systems

GOLD	Zac Goldsmith will contest the 2016 London mayoral election for the Conservatives, it has been announced.	
DOCUMENT:	The Richmond Park and North Kingston MP said he was “honoured” after winning 70% of the 9,227 votes cast using an online primary system. He beat London Assembly Member Andrew Boff, MEP Syed Kamall and London’s deputy mayor for crime and policing Stephen Greenhalgh. Mr Goldsmith’s main rival is likely to be Labour’s Sadiq Khan. (2 sentences with 59 words are abbreviated here.) Mr Goldsmith, who was the favourite for the Tory nomination, balloted his constituents earlier this year to seek permission to stand. At the very point of his entry into the race for London mayor, Zac Goldsmith’s decision revealed two big characteristics. (5 sentences with 108 words are abbreviated here.) Mr Goldsmith - who first entered Parliament in 2010 - told the BBC’s Daily Politics that he hoped his environmental record would appeal to Green and Lib Dem voters and he also hoped to “reach out” to UKIP supporters frustrated with politics as usual and the UK’s relationship with the EU. Zac Goldsmith Born in 1975, educated at Eton and the Cambridge Centre for Sixth-form Studies (5 sentences with 76 words are abbreviated here.) Mr Goldsmith, who has confirmed he would stand down from Parliament if he became mayor, triggering a by-election, said he wanted to build on current mayor Boris Johnson’s achievements. (3 sentences with 117 words are abbreviated here.) Both Mr Khan and Mr Goldsmith oppose a new runway at Heathrow airport, a fact described by the British Chambers of Commerce as “depressing”. (1 sentences with 31 words is abbreviated here.) Current mayor Boris Johnson will step down next year after two terms in office. He is also currently the MP for Uxbridge and South Ruislip, having been returned to Parliament in May. Some Conservatives have called for an inquiry into the mayoral election process after only 9,227 people voted - compared with a 87,884 turnout for the Labour contest. (4 sentences with 121 words are abbreviated here.)	
PTGEN	UKIP leader Nigel Goldsmith has been elected as the new mayor of London to elect a new Conservative MP.	[45.7, 6.1, 28.6]
TCONVS2S	Former London mayoral candidate Zac Goldsmith has been chosen to stand in the London mayoral election.	[50.0, 26.7, 37.5]
TRANS2S	Former London mayor Sadiq Khan has been chosen as the candidate to be the next mayor of London.	[35.3, 12.5, 23.5]
GPT-TUNED	Conservative MP Zac Goldwin’s bid to become Labour’s candidate in the 2016 London mayoral election.	[42.4, 25.8, 36.4]
BERTS2S	Zac Goldsmith has been chosen to contest the London mayoral election.	[66.7, 40.0, 51.9]

(Maynez et al 2020)

Hallucination in state-of-the-art summarization systems

Models	Hallucinated		
	I	E	$I \cup E$
PTGEN	19.9	63.3	75.3
TCONVS2S	17.7	71.5	78.5
TRANS2S	19.1	68.1	79.3
BERTS2S	16.9	64.1	73.1
GOLD	7.4	73.1	76.9

- “Intrinsic hallucinations” combine concepts from the document in a way that is unsupported by the document
- Gold “summaries” from standard datasets also contain facts that are not in the document!

(Maynez et al 2020)