

# Recent Literature on Summarization and Factuality

2023/04/28

Li Peng-Hsuan 李朋軒

# Items

- Summarization background
- Factuality
  - Generating factual summaries
  - Post-editing for factuality
  - Factuality prediction (scoring, generation re-ranking, dataset cleaning)
  - Ask questions for fact-checking
- Beyond factual errors

# Items

- Summarization background
- Factuality
  - Generating factual summaries
  - Post-editing for factuality
  - Factuality prediction (scoring, generation re-ranking, dataset cleaning)
  - Ask questions for fact-checking
- Beyond factual errors

# Summarization

- Source text/document/context → summary

---

**DOCUMENT:** Authorities said the incident took place on Sao Joao beach in Caparica, south-west of Lisbon.

The National Maritime Authority said a middle-aged man and a young girl died after they were unable to avoid the plane.

*[6 sentences with 139 words are abbreviated from here.]*

Other reports said the victims had been sunbathing when the plane made its emergency landing.

*[Another 4 sentences with 67 words are abbreviated from here.]*

Video footage from the scene carried by local broadcasters showed a small recreational plane parked on the sand, apparently intact and surrounded by beachgoers and emergency workers.

*[Last 2 sentences with 19 words are abbreviated.]*

---

---

**SUMMARY:** *A man and a child have been killed after a light aircraft made an emergency landing on a beach in Portugal.*

---

# Summarization

- Extractive summarization vs. abstractive summarization

Question: ?  
Which places seem interesting to you for visiting in this state?

Source Document: New York ?

Use the search box to look for keywords  
park 1 / 29

the Erie Canal, gave it incomparable advantages over other regions of the east coast and built its political and cultural ascendancy.

Many landmarks in New York are well known, including four of the world's ten most-visited tourist attractions in 2013: Times Square, Central Park, Niagara Falls, and Grand Central Terminal. New York is home to the Statue of Liberty. In the 21st century, New York has emerged as a global node of creativity and entrepreneurship, social tolerance, and environmental sustainability. New York has approximately 200 colleges and universities, including the State University of New York. Several have been ranked among the top 100 in the nation and world.

### History

#### Native American history

The tribes in what is now New York were predominantly Haudenosaunee and Algonquian. Long Island was divided roughly in half between the Wampanoag and Lenape. The Lenape also controlled most of the region surrounding New York Harbor. North of the Lenape was a third Algonquian nation, the Mohicans. Starting north of them, from east to west, were three Iroquoian nations: the Mohawk, the original Iroquois and the Petun. South of them, divided roughly along Appalachia, were the Susquehannock and the Erie.

Many of the Wampanoag and Mohican peoples were caught up in King Philip's War, a joint effort of many New England tribes to push Europeans off their land. After the death of their leader, Chief Philip Metacomet, most of those peoples fled inland, splitting into the Abenaki and the Schaghticoke. Many of the Mohicans remained in the region until the 1800s, however, a small group known as the Ouabano migrated southwest into West Virginia at an earlier time. They may have merged with the Shawnee.

The Mohawk and Susquehannock were the most militaristic. Trying to corner trade with the Europeans, they targeted other tribes. The Mohawk were also known for refusing white settlement on their land and banishing any of their people who converted to Christianity. They posed a major threat to the Abenaki and Mohicans, while the Susquehannock briefly conquered the Lenape in the 1600s. The most devastating event of the century, however, was the Beaver Wars. From approximately 1640–1680, Iroquoian peoples waged campaigns which extended from modern-day Michigan to Virginia against Algonquian and Siouan tribes, as well as each other. The aim was to control more land for animal trapping, a career most natives had turned to in hopes of trading with whites first. This completely changed the ethnography of the region, and most large game was hunted out before whites ever fully explored the land. Still, afterward, the Iroquois Confederacy offered shelter to refugees of the Mascouten, Erie, Choronnton, Tutelo, Saoponi, and Tuscarora nations. In the 1700s, they would also merge with the Mohawk during

Summary: New York ?

Many landmarks in New York are well known, including four of the world's ten most-visited tourist attractions in 2013: Times Square, Central Park, Niagara Falls, and Grand Central Terminal.

In contrast with New York City's urban landscape, the vast majority of the state's geographic area is dominated by meadows, forests, rivers, farms, mountains, and lakes.

Most of the southern part of the state rests on the Allegheny Plateau, which extends from the southeastern United States to the Catskill Mountains; the section in New York State is known as the Southern Tier.

Niagara Falls State Park, established in 1885, is the oldest state park in the United States and the first to be created via eminent domain.

Several U.S. national sports halls of fame are or have been situated in New York.

---

**SUMMARY:** *A man and a child have been killed after a light aircraft made an emergency landing on a beach in Portugal.*

---

**DOCUMENT:** Authorities said the incident took place on Sao Joao beach in Caparica, south-west of Lisbon.

The National Maritime Authority said a middle-aged man and a young girl died after they were unable to avoid the plane.

[6 sentences with 139 words are abbreviated from here.]

Other reports said the victims had been sunbathing when the plane made its emergency landing.

[Another 4 sentences with 67 words are abbreviated from here.]

Video footage from the scene carried by local broadcasters showed a small recreational plane parked on the sand, apparently intact and surrounded by beachgoers and emergency workers.

[Last 2 sentences with 19 words are abbreviated.]

---

# Summarization

- Standard abstractive summarization datasets
  - CNN/DailyMail
  - XSum (BBC)

# Summarization

- Factuality
  - A factual/faithful summary: a summary evidenced by its source
  - XSum: 70% gold summaries contain hallucinations

# Items

- Summarization background
- **Factuality**
  - **Generating factual summaries**
  - Post-editing for factuality
  - Factuality prediction (scoring, generation re-ranking, dataset cleaning)
  - Ask questions for fact-checking
- Beyond factual errors



# Factuality – Generating Factual Summaries

**Source:**

(...) Chris Cox, the university's director of development, said the research centre would translate discoveries made in the laboratory into new treatments. He said it would house 150 additional researchers who will be developing new ideas and treatments. Research will focus on radiation therapy, lung cancer, women's cancers, melanoma and haematological oncology. The centre is the result of a partnership between The University of Manchester, The Christie NHS Foundation Trust and Cancer Research UK. The Christie's chief executive Caroline Shaw said the funding would "help facilitate groundbreaking research right here in Manchester". (...)

**Reference:**

Plans for a new cancer research centre in Manchester have received a £12.8m funding boost from the government.

**BART (beam size=6):**

The University of Manchester has been awarded £3.5m to help fund a new cancer research centre.  
The University of Manchester has been awarded £1.3m to help fund a new cancer research centre.  
The University of Manchester has been awarded £1.2m to help fund a new cancer research centre.  
The University of Manchester has been awarded £1.5m to help fund a new cancer research centre.  
The University of Manchester has been awarded £5m to help fund a new cancer research centre.  
The University of Manchester has been awarded £4.5m to help fund a new cancer research centre.

**BART trained with rejection (greedy decoding):**

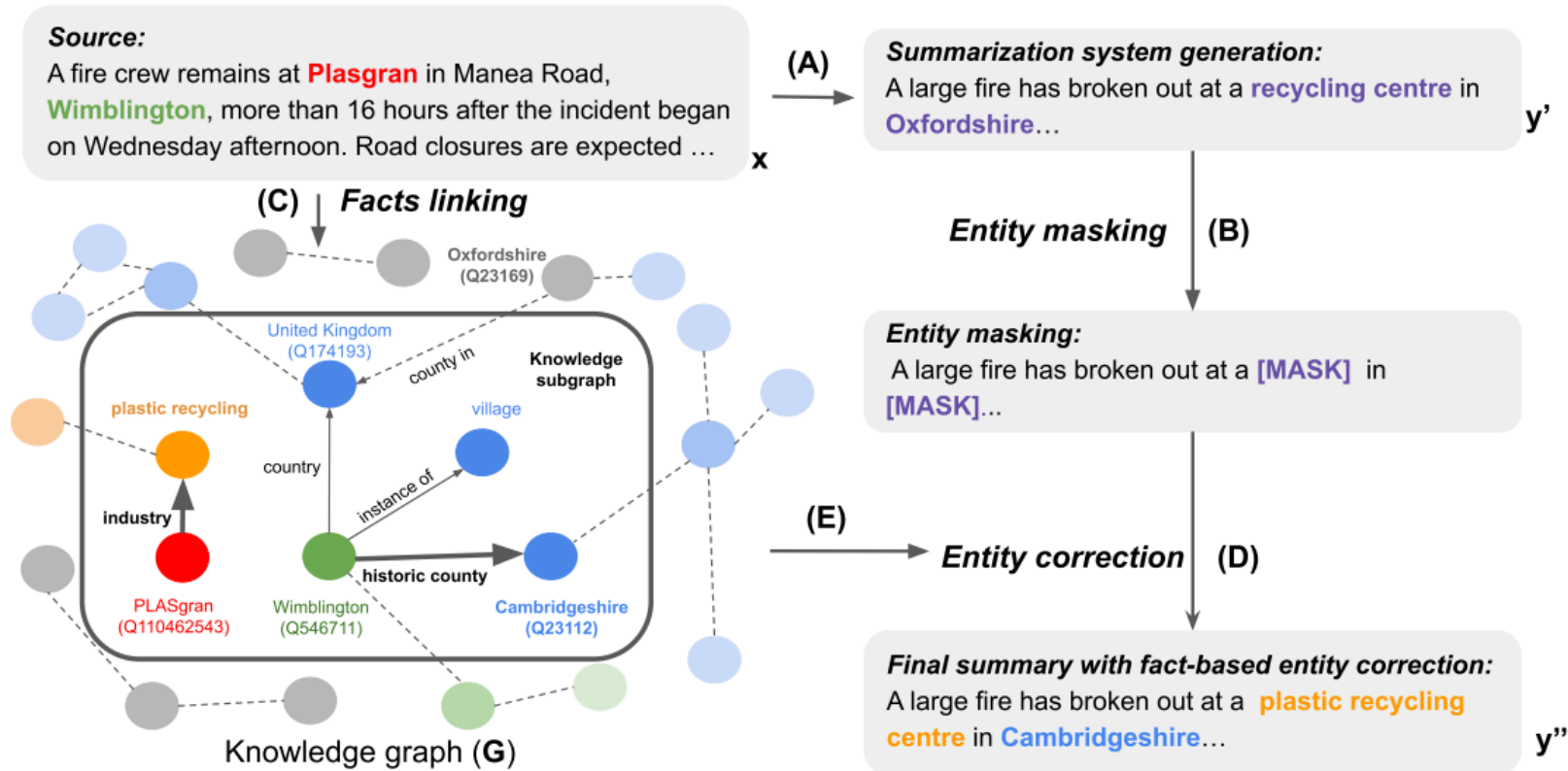
The University of Manchester has been awarded £<REJ>m to help fund its new cancer research centre.

**BART trained with rejection (regularized decoding):**

Details of a new cancer research centre at the University of Manchester have been revealed.

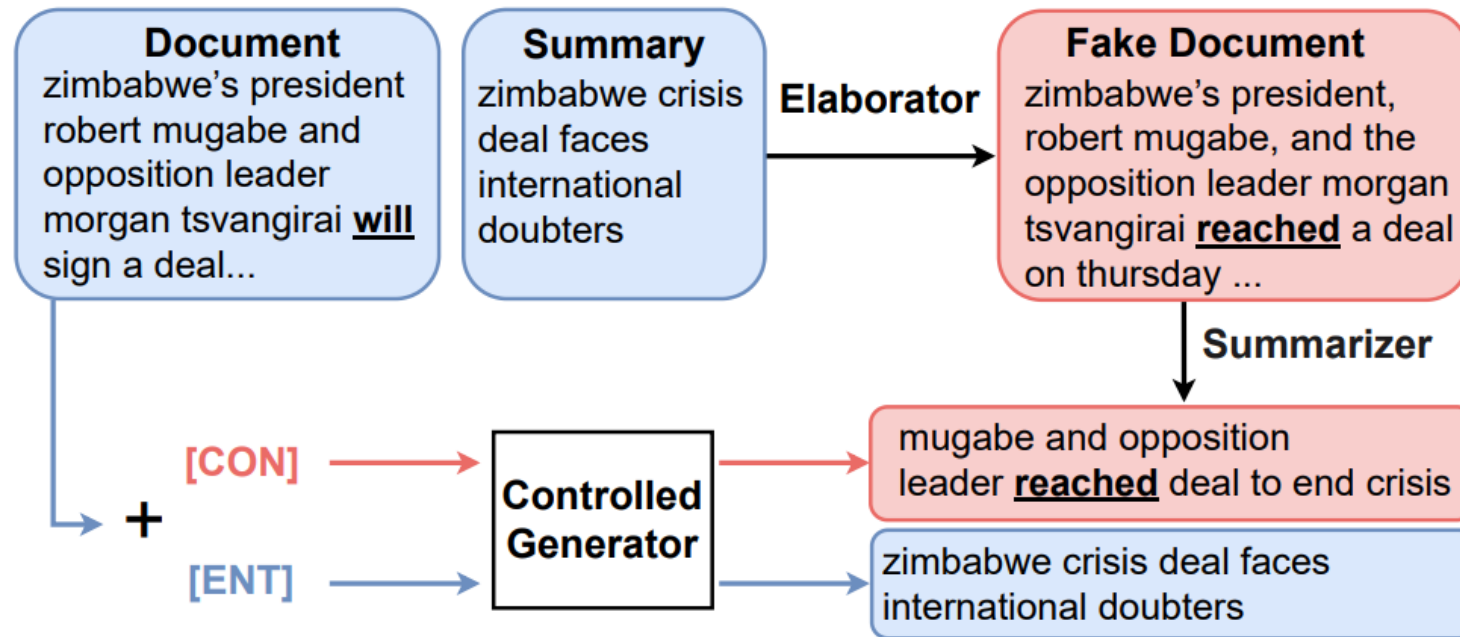
*Learning to reject noisy "gold" tokens in training set summaries*

# Factuality – Generating Factual Summaries



Two-stage summary: 1st stage for fluency; 2nd stage for filling correct entities

# Factuality – Generating Factual Summaries

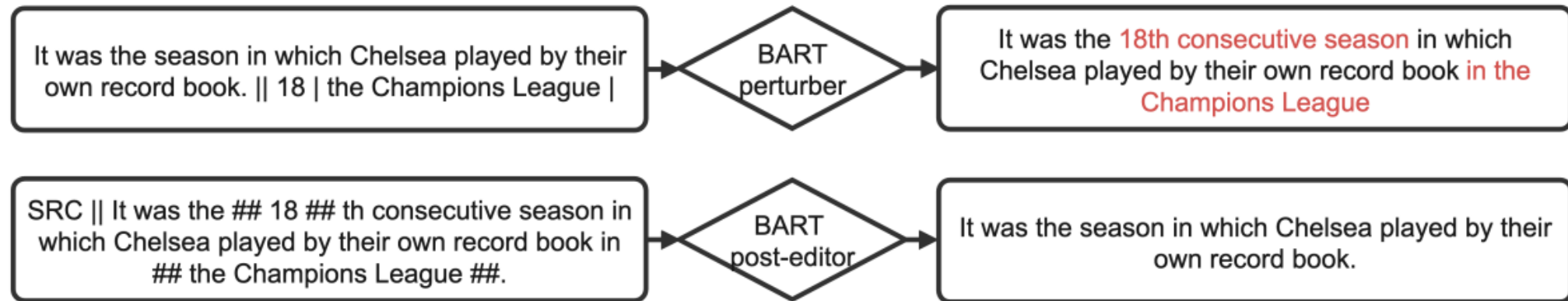


*Learning to generate both factual and hallucinated back-summarization*

# Items

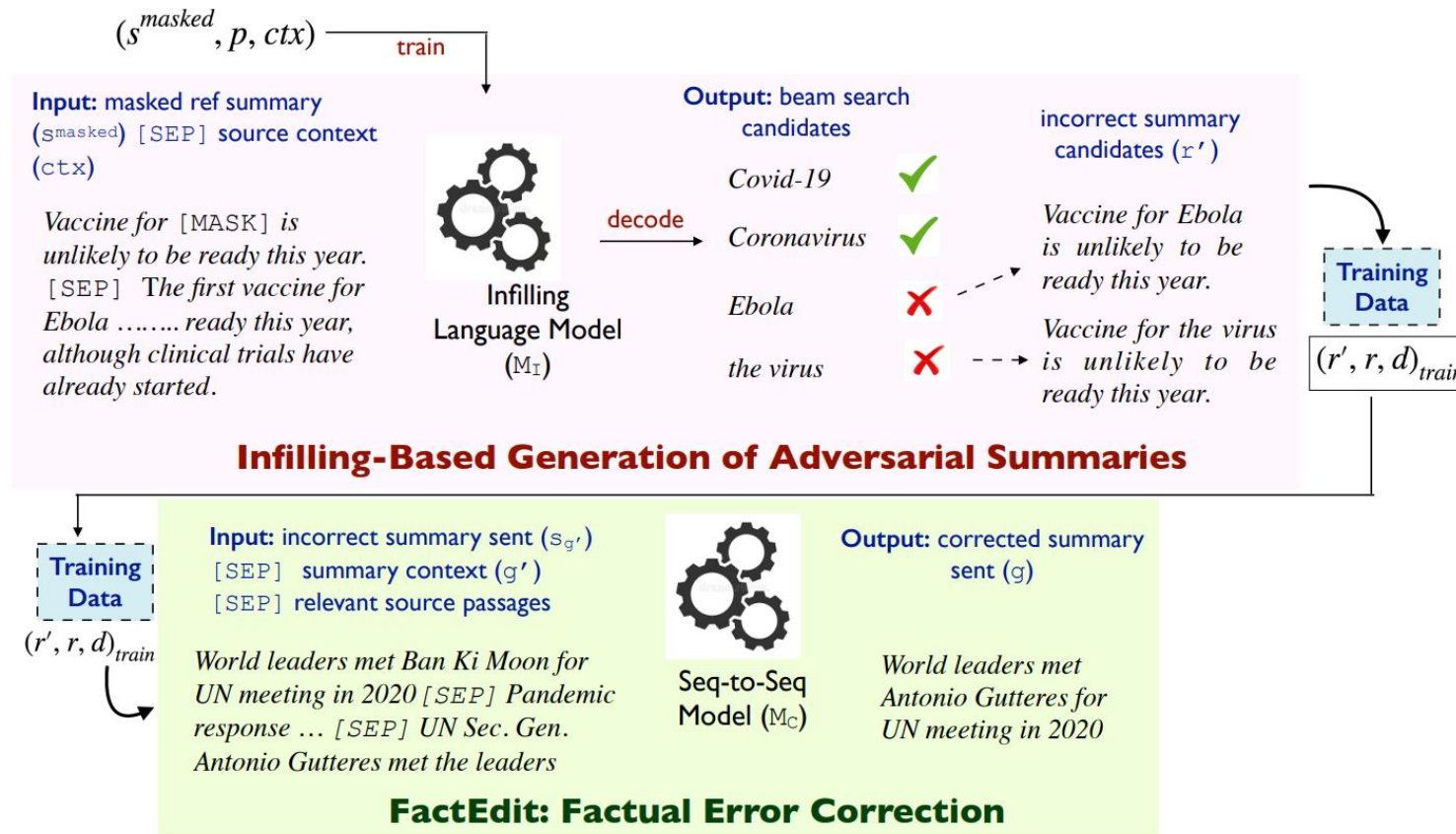
- Summarization background
- **Factuality**
  - Generating factual summaries
  - **Post-editing for factuality**
  - Factuality prediction (scoring, generation re-ranking, dataset cleaning)
  - Ask questions for fact-checking
- Beyond factual errors

# Factuality – Post-editing for Factuality



*Training a post-editor on summaries with decompression-based perturbation*

# Factuality – Post-editing for Factuality



*Training a post-editor on summaries with infilling-based perturbation*

# Items

- Summarization background
- **Factuality**
  - Generating factual summaries
  - Post-editing for factuality
  - **Factuality prediction (scoring, generation re-ranking, dataset cleaning)**
  - Ask questions for fact-checking
- Beyond factual errors

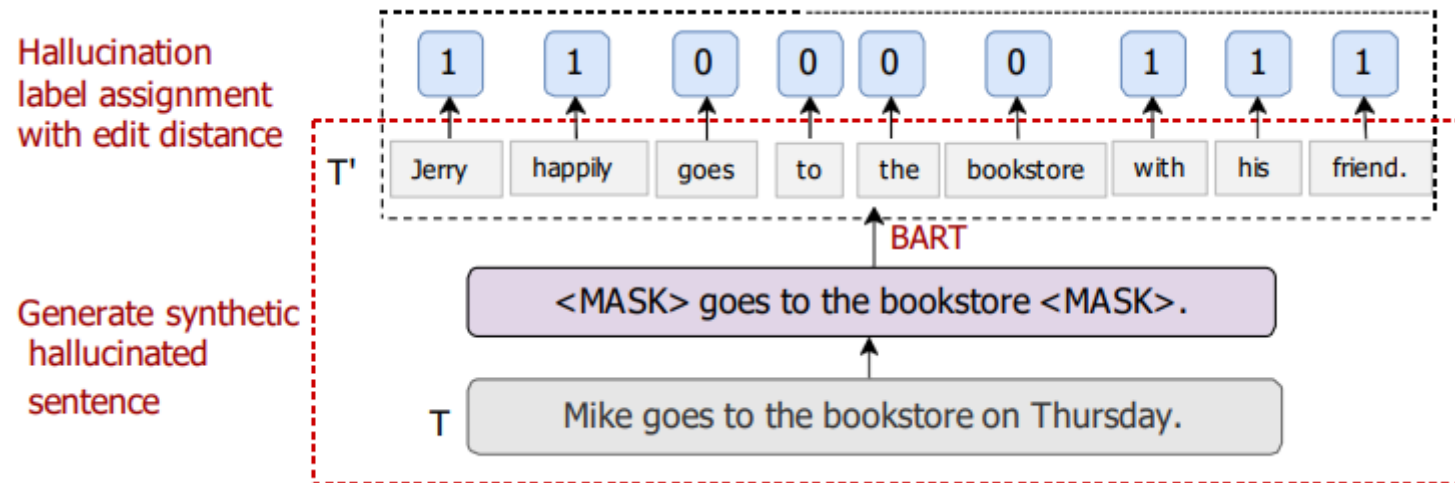
# Factuality – Factuality Prediction

Transformation	Original sentence	Transformed sentence
Paraphrasing	Sheriff Lee Baca has now decided to recall some 200 badges his department has handed out to local politicians just two weeks after the picture was released by the U.S. attorney's office in support of bribery charges against three city officials.	Two weeks after the US Attorney's Office issued photos to support bribery allegations against three municipal officials, Lee Baca has now decided to recall about 200 badges issued by his department to local politicians.
Sentence negation	Snow <b>was</b> predicted later in the weekend for Atlanta and areas even further south.	Snow <b>wasn't</b> predicted later in the weekend for Atlanta and areas even further south.
Pronoun swap	It comes after <b>his</b> estranged wife Mona Dotcom filed a \$20 million legal claim for cash and assets.	It comes after <b>your</b> estranged wife Mona Dotcom filed a \$20 million legal claim for cash and assets.
Entity swap	Charlton coach <b>Guy Luzon</b> had said on Monday: 'Alou Diarra is training with us.'	Charlton coach <b>Bordeaux</b> had said on Monday: 'Alou Diarra is training with us.'
Number swap	He says he wants to pay off the <b>\$12.6million</b> lien so he can sell the house and be done with it.	He says he wants to pay off the <b>\$3.45million</b> lien so he can sell the house and be done done with it.
Noise injection	Snow <b>was</b> predicted later in the weekend for Atlanta and areas even further south.	Snow <b>was was</b> predicted later in the weekend for Atlanta and areas <b>even</b> further south.

*Train a detector on source sentences with rule-based perturbation*

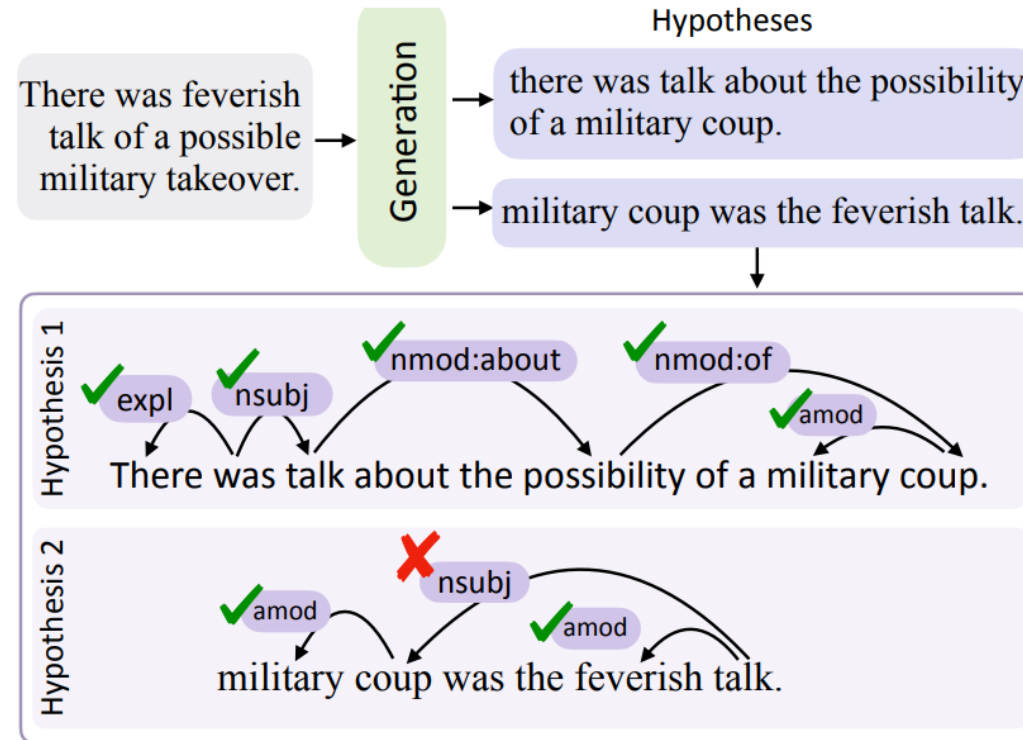


# Factuality – Factuality Prediction



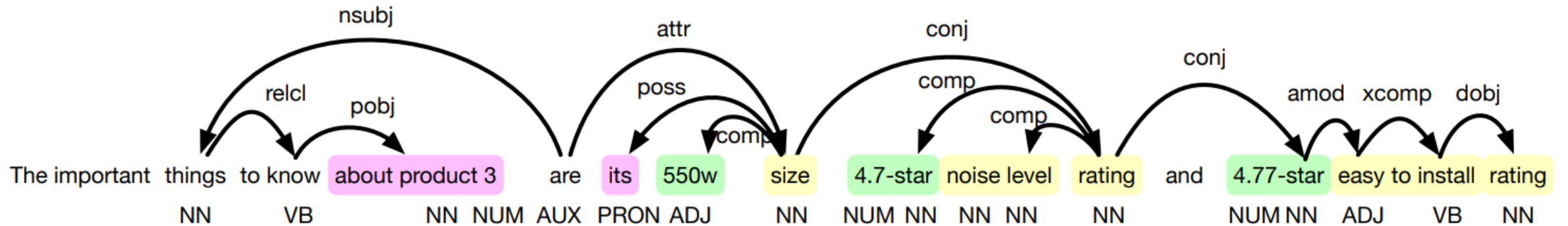
*Train a detector on summaries with generation-based perturbation*

# Factuality – Factuality Prediction



*Train a detector for paraphrase-based bad dependency arcs*

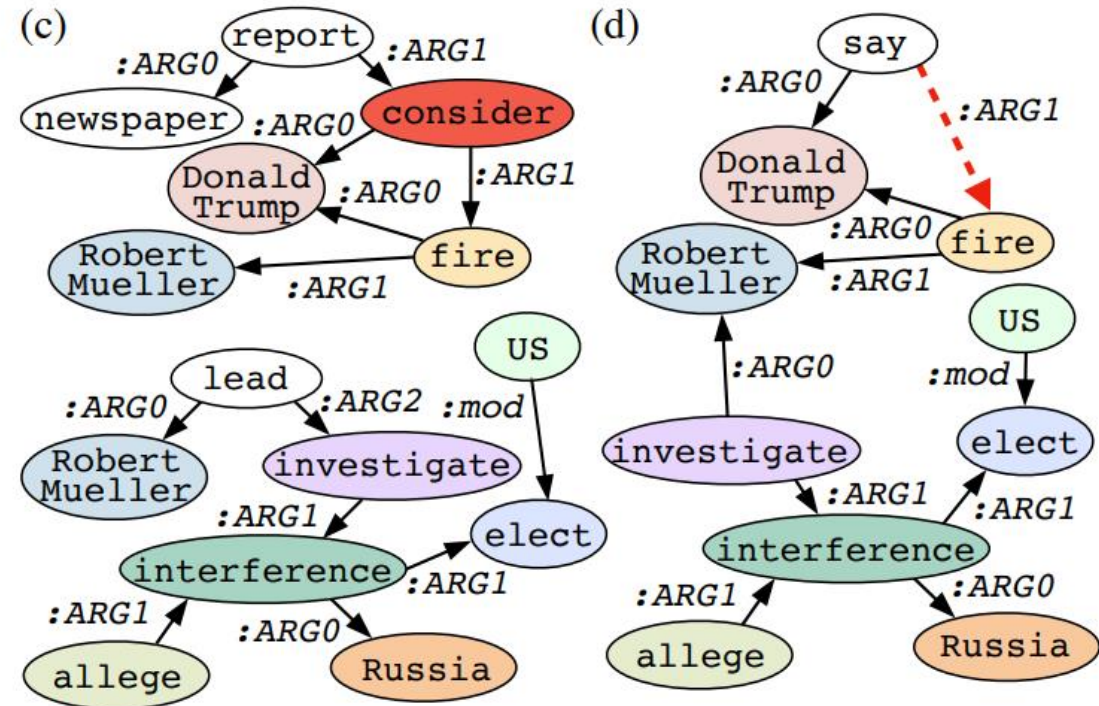
# Factuality – Factuality Prediction



*Fact-check entity-attribute pairs extracted from dependency trees*

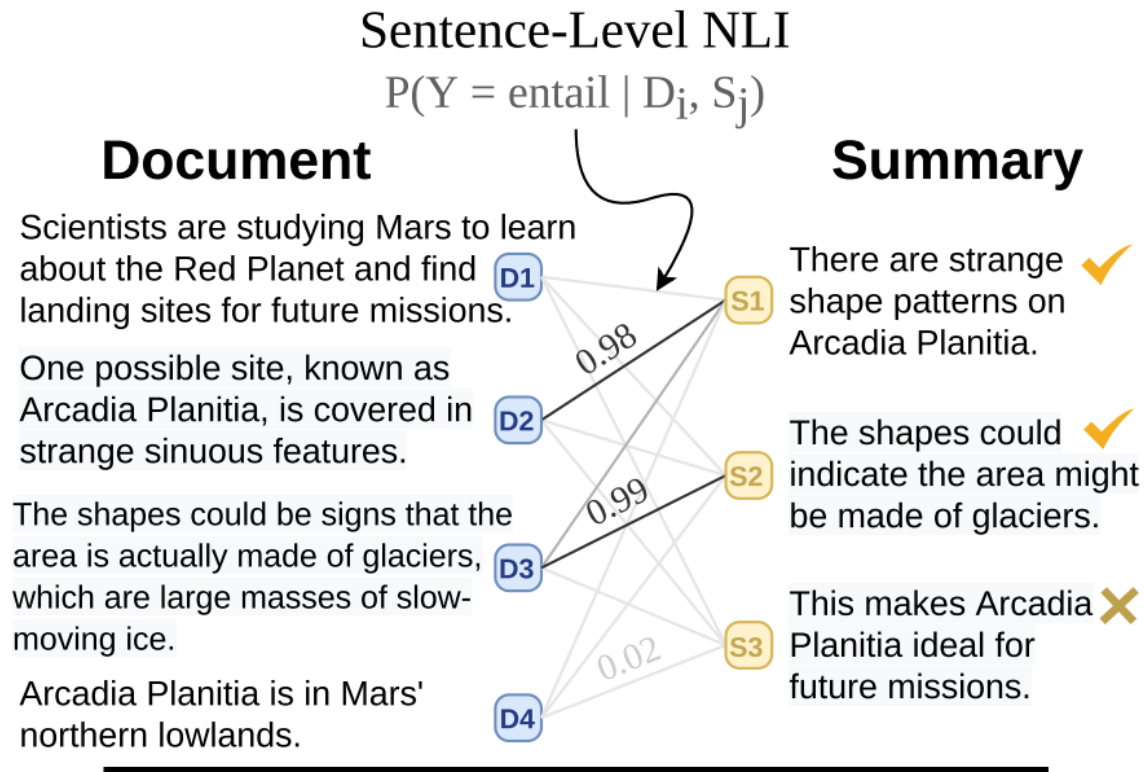
# Factuality – Factuality Prediction

- (a) [...] Mr Mueller was given the role of special counsel by the justice department to lead its investigation into alleged Russian interference in last year's US election [...] The NYT has reported that Mr Trump has considered firing Mr Mueller [...].
- (b) US President Donald Trump has said he will fire special counsel Robert Mueller, who is investigating alleged Russian interference in the US election.



*Train a detector on documents and summaries and their meaning representations*

# Factuality – Factuality Prediction

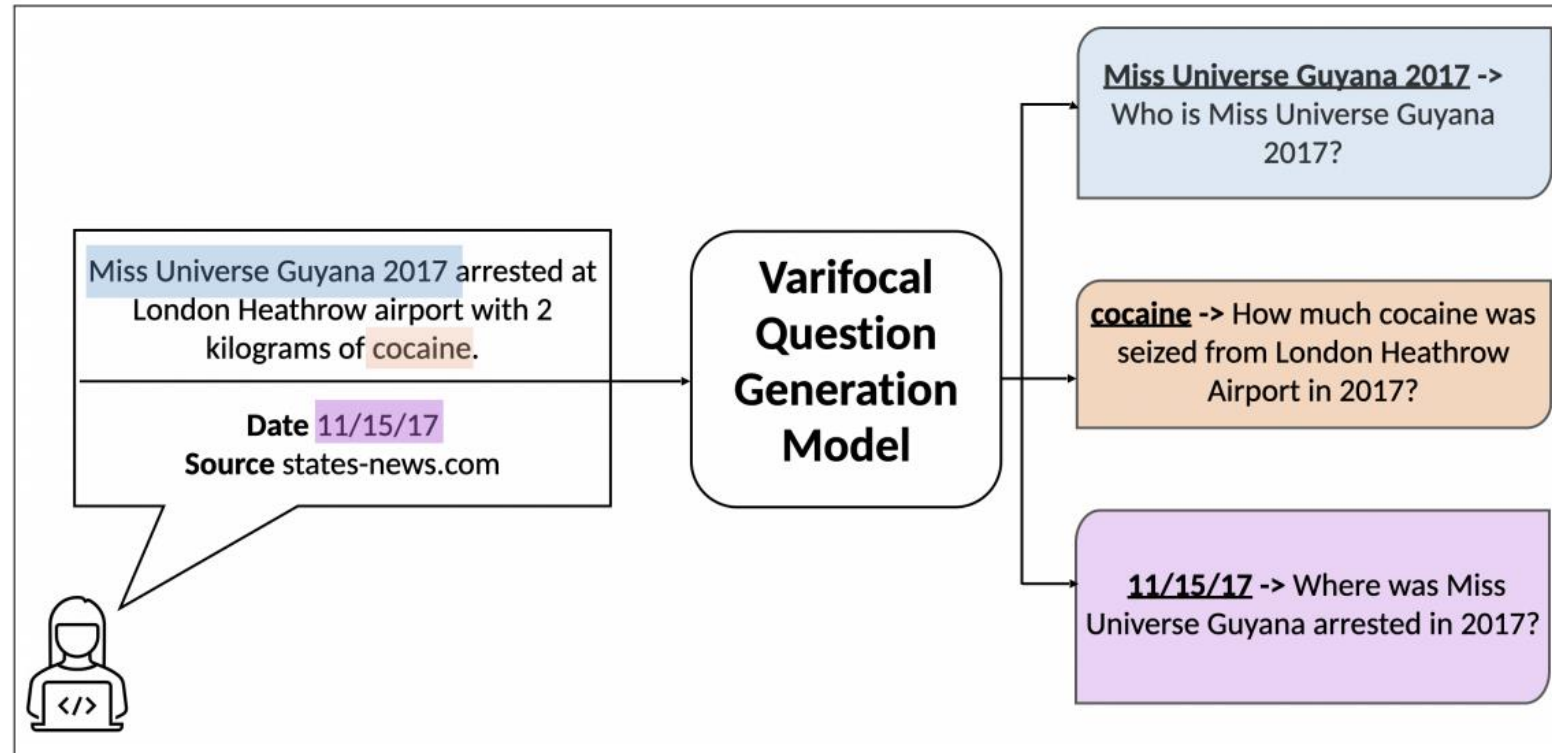


*Identifying hallucination with sentence-level document-summary entailment*

# Items

- Summarization background
- **Factuality**
  - Generating factual summaries
  - Post-editing for factuality
  - Factuality prediction (scoring, generation re-ranking, dataset cleaning)
  - Ask questions for fact-checking
- Beyond factual errors

# Factuality – Ask Questions for Fact-checking



*Training a question generator on target entities and phrases and gold answer-matched questions*

# Factuality – Ask Questions for Fact-checking

Claim	
<i>Joe Biden stated on August 31, 2020 in a speech: "When I was vice president, violent crime fell 15% in this country. ... The murder rate now is up 26% across the nation this year under Donald Trump."</i>	
Claim Decomposition: focus of this work	
Literal	<i>Q1: Did the crime rate fall by 15% during Joe Biden's presidency?</i> ✓
	<i>Q2: Did the murder rate in 2020 increase by 26% from 2019?</i> ✓
Implied	<i>Q3: Is Biden comparing crime rates from the same time interval in his statement?</i> ✗
	<i>Q4: Is violent crime rate and murder rate directly comparable?</i> ✗

Aggregation } Judgment: **half-true**

*Created a dataset of literal and implied questions for claims*



# Items

- Summarization background
- Factuality
  - Generating factual summaries
  - Post-editing for factuality
  - Factuality prediction (scoring, generation re-ranking, dataset cleaning)
  - Ask questions for fact-checking
- **Beyond factual errors**

# Beyond Factual Errors

## Excerpt from Generated Summary

*Miss Manette receives a letter from the bank informing her that information about her father's small property has been discovered. She wants to travel to France to identify him and restore him to life. Mr. Lorry explains that her father has been found under another name and is being held in a house in Paris.*

*In court, Mr. Darnay is accused of treason. However, he is acquitted after his patriot friend, Roger Cly, testifies against him.*

Corresponding excerpt from the **human-written summary** properly contextualizes the new character.

*Five years later, in 1780, a young Frenchman, named Charles Darnay, is accused of being a traitor and a spy. Lucie [Mannette] and her father are reluctant witnesses for the prosecution [...]*

EVENT UNCLEAR  
Why restore to life?  
Is her father sick?

CHARACTER UNCLEAR  
Who is Mr. Lorry?

SCENE TRANSITION  
Sudden shift in scene  
(what happened to  
Miss Mannette?)

CHARACTER UNCLEAR  
Who is Mr. Darnay?

*GPT-3 fined-tuned for summarization exhibits coherence errors*

# Beyond Factual Errors

## Context:

John Fenwick, an aspiring artist, accepts a loan from Mr. Morrison to move to London to pursue his art career.

In London, he becomes infatuated with Madame de Pastourelles, a beautiful and intelligent artist.

## Current Summary Segment

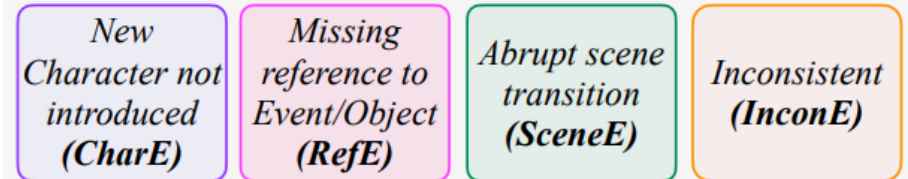
In Paris, he impresses Lord Findon with his work.

One afternoon, he writes a letter to Mrs. Morrison expressing sympathy for her husband's suicide.

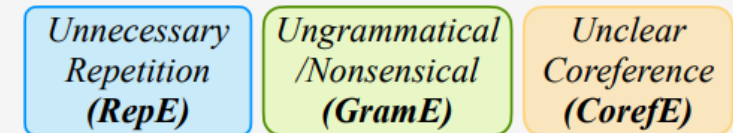
Fenwick's wife becomes frightened when a tramp threatens to kill her and her child.

One afternoon, Fenwick goes to lunch with Madam de Pastourelles, a beautiful artist, and Eugenie, a wealthy benefactor. She promises to help him help customers.

## Coherence Errors



## Language & Fluency Errors



Annotated coherence and language errors

# Backup