Multimodal Multi-Document Evidence Summarization In Fact-Checking Task

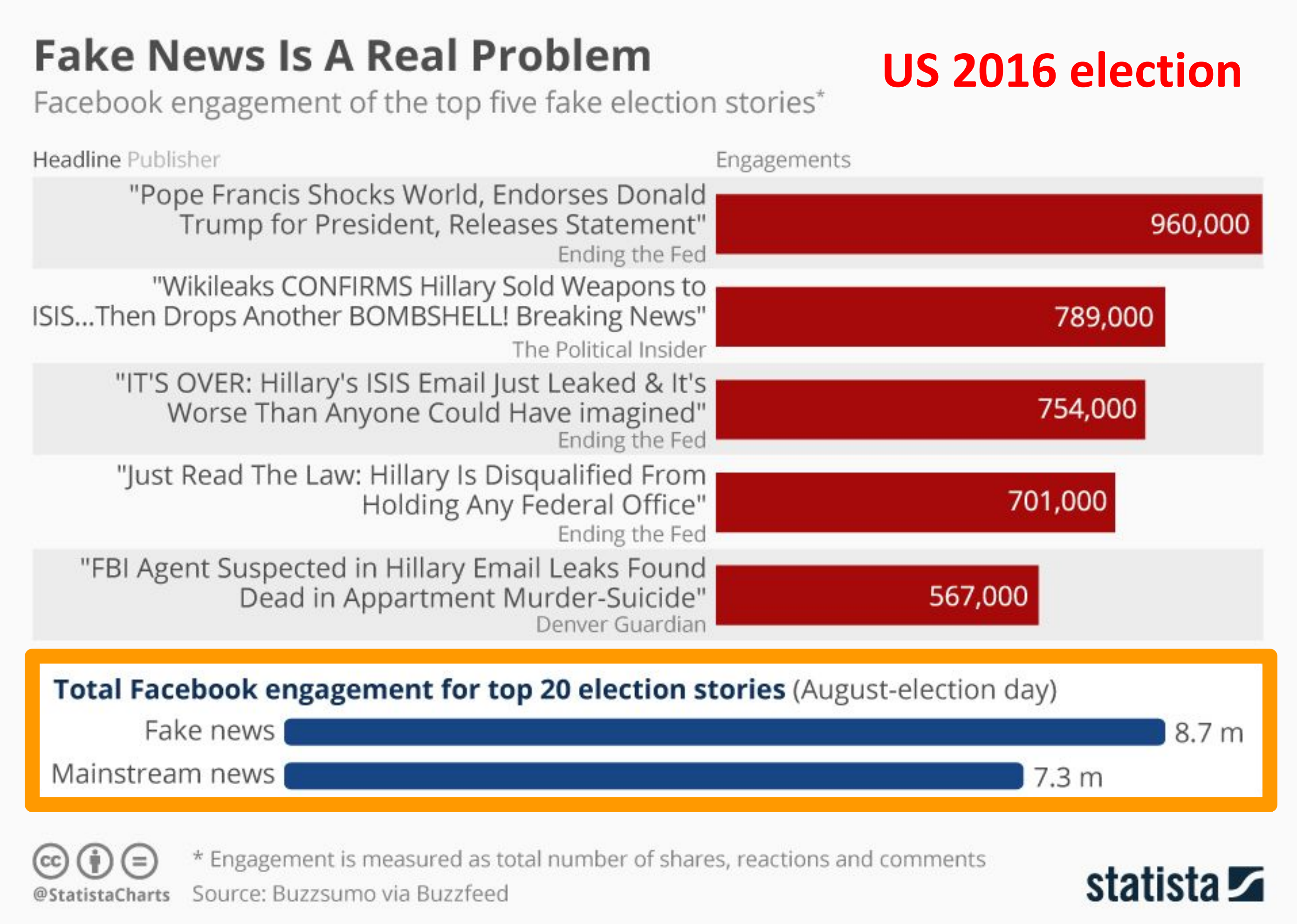
Ting-Chih Chen

12/1/2023

Agenda

- Research Background
- Limitations & Challenges
  - Multimedia Event Extraction
  - Multimodal Summarization
- Methods:
  - Evidence Summarization
  - Multimodal Multi-doc Claim Generation
- Results
- Conclusion

US 2016 election



Research Background

- Research indicates that verifying all aspects of a 200 word claim can require 4 hours of effort

Ref: The State of Fact-Checking in Science Journalism

Research Background

- **Claim: Hamas releases the hostages during truce with Israel**

Hamas says it wants to extend its four-day truce with Israel, which has entered its third day and has now seen the release of three groups of Israeli hostages from Gaza and three groups of Palestinian prisoners and detainees from Israeli jails. …

The fragile ceasefire between Israel and Hamas was back on track Sunday as the militants freed 17 more hostages, including 14 Israelis and the first American, in a third exchange under a four-day truce that the U.S. said it hoped would be extended. In turn, Israel released 39 Palestinian prisoners, all young men. …

Abigail Edan, a 4-year-old with American and Israeli citizenship who was taken hostage on Oct. 7, was back in Israel on Sunday, President Biden said in a news conference. **The Red Cross was transporting a group of Hamas's captives, the third of four groups planned to be exchanged for Palestinian prisoners in as many days, to Israeli authorities**. …

Research Background

- **Claim: Israelis rally in the playground**

Hamas says it wants to extend its four-day truce with Israel, which has entered its third day and has now seen the release of three groups of Israeli hostages from Gaza and three groups of Palestinian prisoners and detainees from Israeli jails. …

The fragile ceasefire between Israel and Hamas was back on track Sunday as the militants freed 17 more hostages, including 14 Israelis and the first American, in a third exchange under a four-day truce that the U.S. said it hoped would be extended. In turn, Israel released 39 Palestinian prisoners, all young men. …

Abigail Edan, a 4-year-old with American and Israeli citizenship who was taken hostage on Oct. 7, was back in Israel on Sunday, President Biden said in a news conference. **The Red Cross was transporting a group of Hamas's captives, the third of four groups planned to be exchanged for Palestinian prisoners in as many days, to Israeli authorities.** …

Research Background

- **Claim: Israelis rally in the playground**

We focus on generating the claim-specific summary, including the necessary information

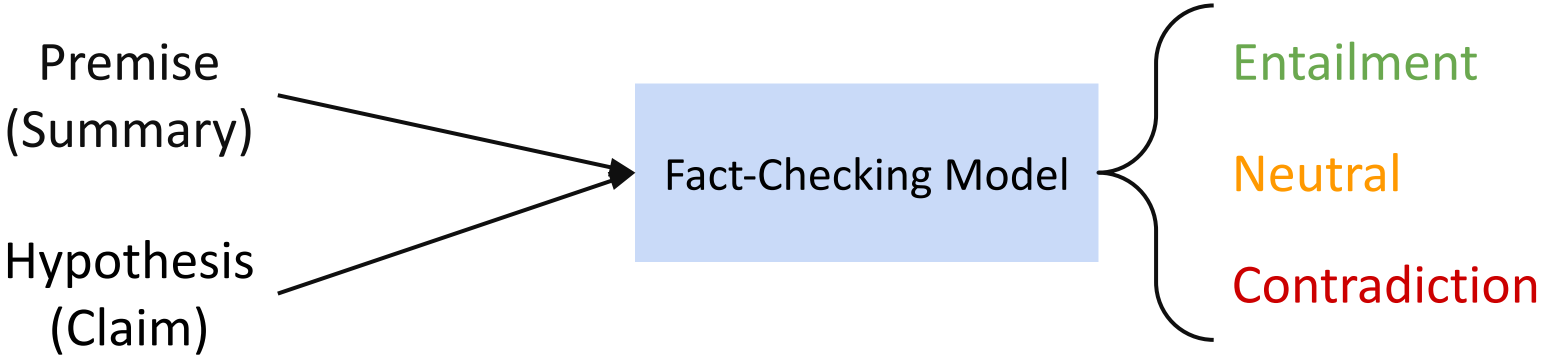
Hama
its fou
which
and h
three
from C
Palestinia
detainees from Israeli jails. …

prisoners, all young men. …

and
e on
sident
d
as's
ed to
ers in as
many days, to Israeli authorities. …

Fact-Checking Task

- Goal is to evaluate the truthfulness of statements.
- The task requires a system to decide whether a piece of text (the **hypothesis**) can be logically deduced from another piece of text accepted as true (the **premise**)



Fact-Checking Task

Entailment:

- Premise: A soccer game with multiple males playing.
- Hypothesis: Some men are playing a sport.

Neutral:

- Premise: A smiling costumed woman is holding an umbrella.
- Hypothesis: A happy woman in a fairy costume holds an umbrella.

Contradiction:

- Premise: A man inspects the uniform of a figure in some East Asian country.
- Hypothesis: The man is sleeping.

Limitations & Challenges

- Multimedia event extraction is the existing potential method to integrate multimodal information
- It can extract the events and entities from the multimedia and simply introduce what type of event is

Challenge: Although we can use this method to understand the relationship between events and entities, we cannot use this to do fact-checking task

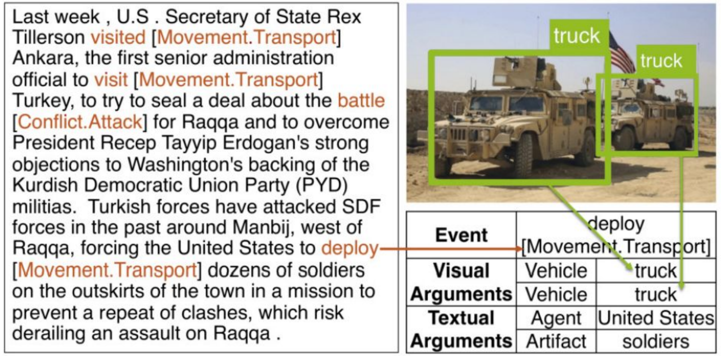


Figure 1: An example of Multimedia Event Extraction. An event mention and some event arguments (*Agent* and *Person*) are extracted from text, while the vehicle arguments can only be extracted from the image.

Ref: Cross-media Structured Common Space for Multimedia Event Extraction

Limitations & Challenges

- Multimodal summarization is a main method to summary the concept from the text, images and videos
- It can not only transfer images and videos information but also transfer the the information in the table

Challenge: Although text summarization is good for illustrating the events in the videos, we must generate the claim-specific summary for fact-checking

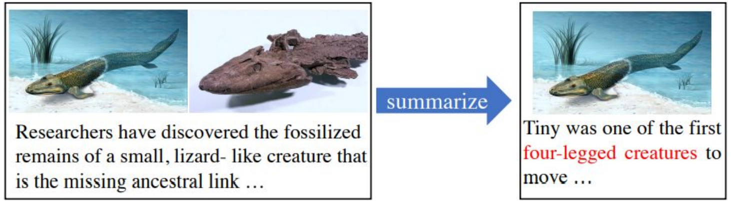


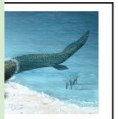
Figure 1: The illustration of our proposed task – Multimodal Summarization with Multimodal Output (MSMO). The image can help better understand the text in the red font.

Ref: MSMO: Multimodal Summarization with Multimodal Output

Limitations & Challenges

- Multimodal summarization is a main method to summary the concept f
- I
  v
  t

Core challenges: we want to think about a method that can understand the multimodal data and it can reason the relationships between the events and entities.



l task –
l Output
tand the

Ch
is g
vid
claim-specific summary for fact-checking

Agenda

- Research Background
- Limitations & Challenges
  - Multimedia Event Extraction
  - Multimodal Summarization
- Methods:
  - Evidence Summarization
  - Multimodal Multi-doc Claim Generation
- Results
- Conclusion

MetaSumPerceiver

- Goal: Generate the claim-specific summaries for fact-checking from multimodal multi-document dataset
- Input: the set of multimodal documents, images, and a claim

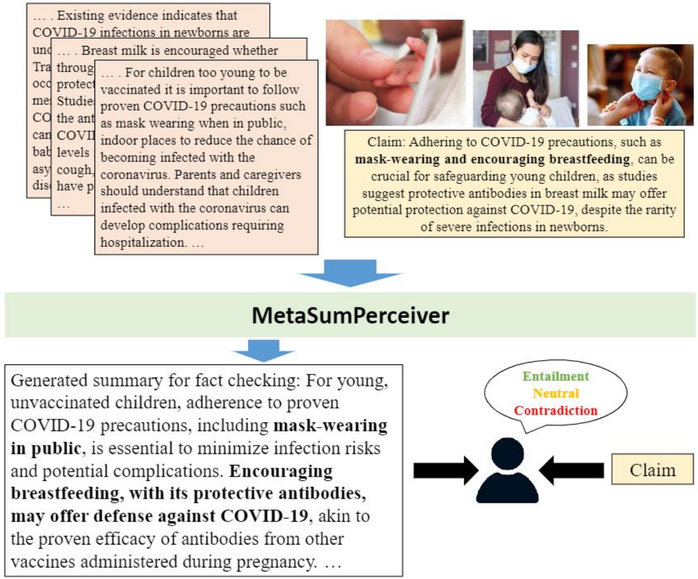
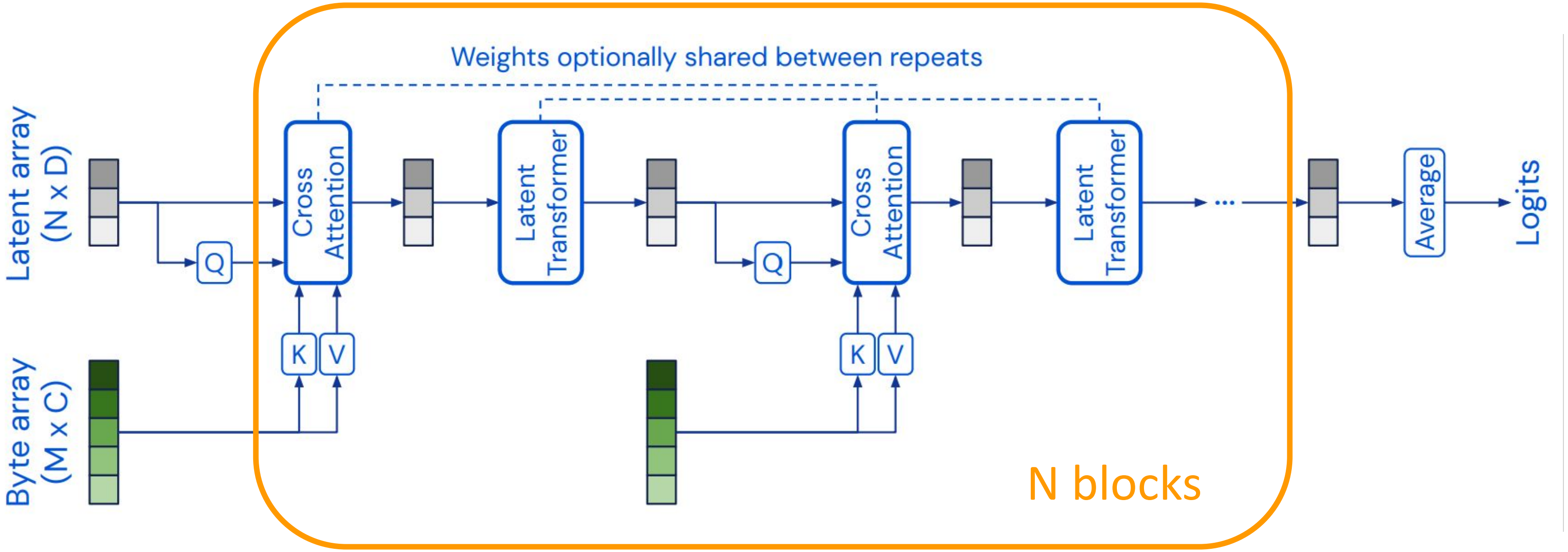


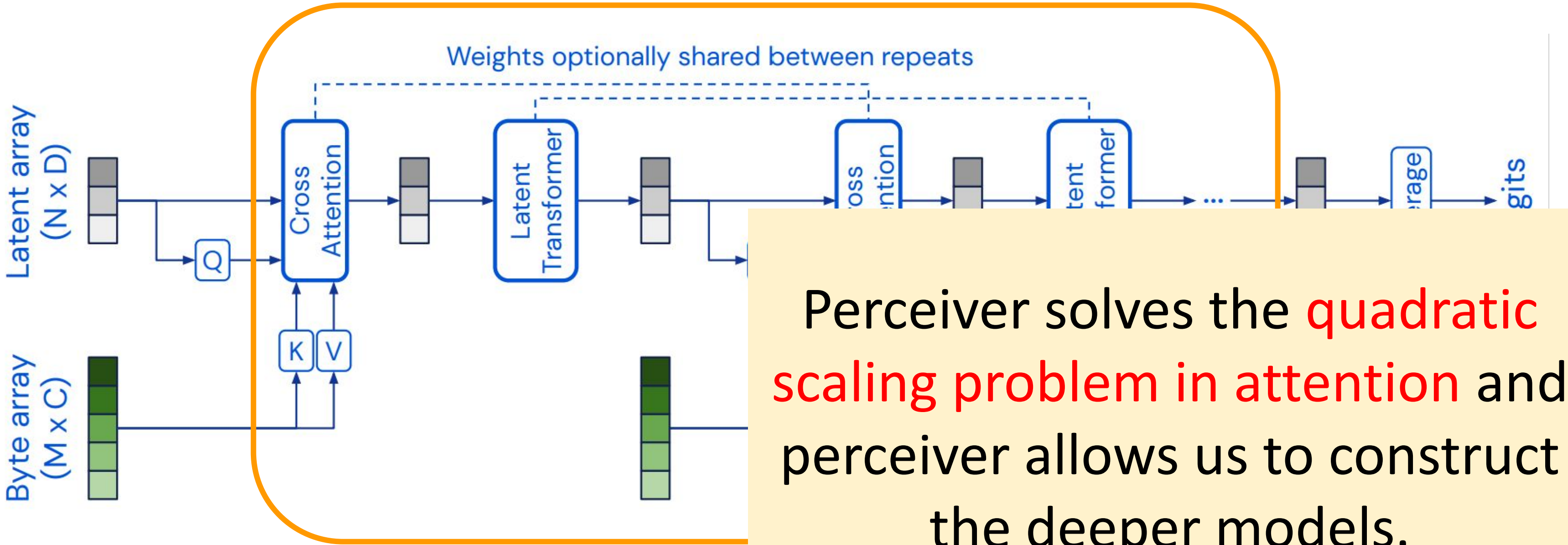
Figure 1: Overview of MetaSumPerceiver: Using inputs such as documents, images, and claims, MetaSumPerceiver generates summaries to facilitate fact-checking. In this example, the summary for fact-checking provides evidence and establishes that the claim in question is entailed by the evidence.

Perceiver



Ref: Perceiver: General Perception with Iterative Attention

Perceiver



Perceiver solves the quadratic scaling problem in attention and perceiver allows us to construct the deeper models.

Ref: Perceiver: General Perception with Iterative Attention

MetaSumPerceiver

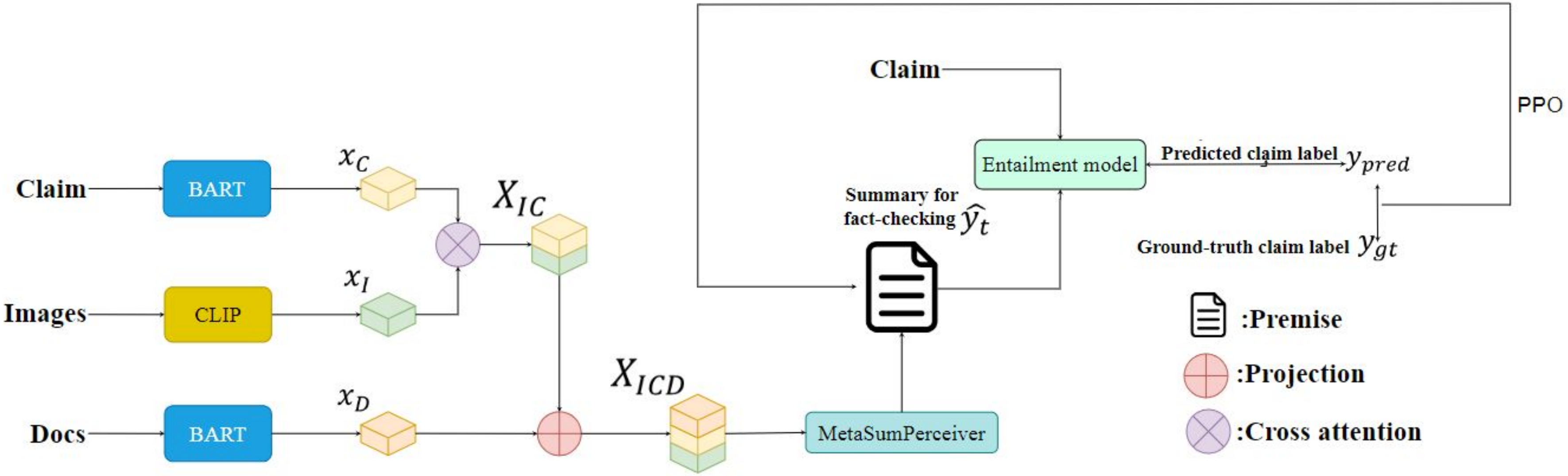


Figure 2: Overview of MetaSumPerceiver: This figure illustrates the process of generating a summary for fact-checking using MetaSumPerceiver, integrating a fixed entailment model for accurate truthfulness labeling. Furthermore, it highlights how PPO is employed to continually refine the summary during the fact-checking process.

MetaSumPerceiver

Design to generate the summary from multimodal data

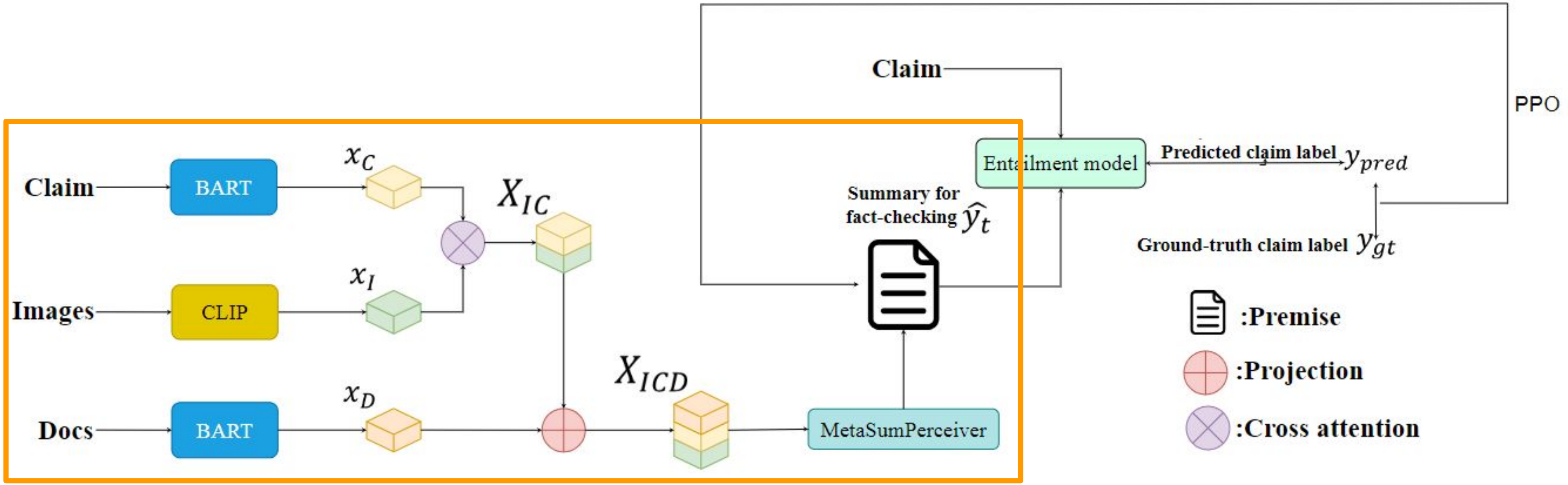


Figure 2: Overview of MetaSumPerceiver: This figure illustrates the process of generating a summary for fact-checking using MetaSumPerceiver, integrating a fixed entailment model for accurate truthfulness labeling. Furthermore, it highlights how PPO is employed to continually refine the summary during the fact-checking process.

MetaSumPerceiver

RL for generating the claim-specific summary

To create a claim-specific summary, we need an **entailment model** that mimics human judgment to assess the summary's relevance to the claim

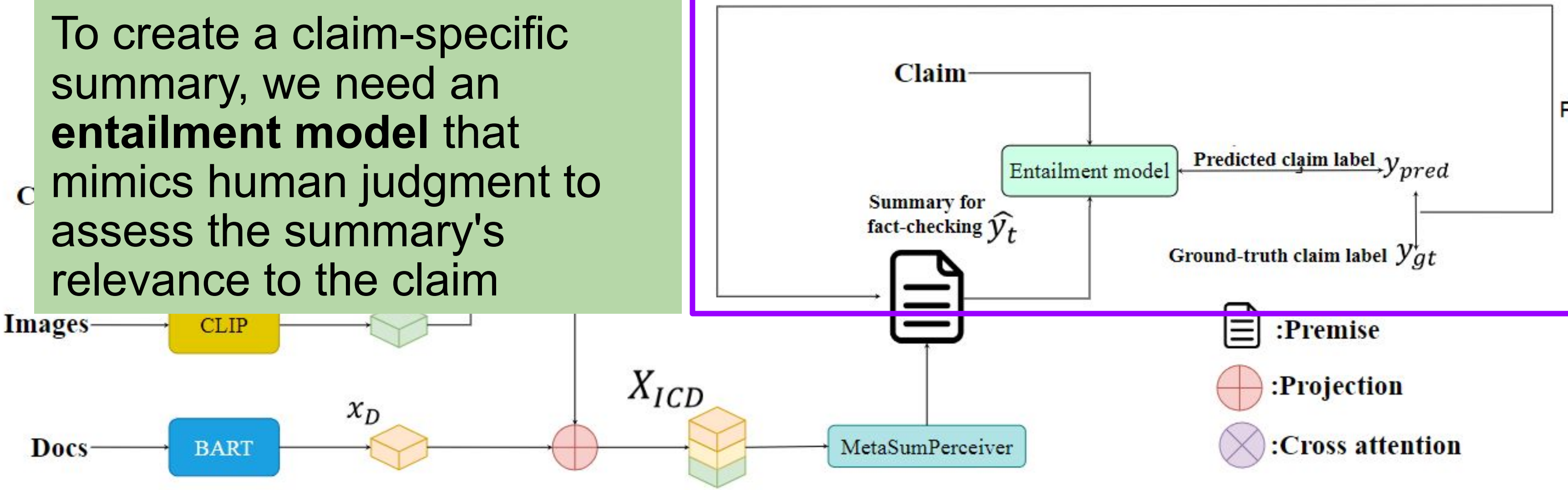


Figure 2: Overview of MetaSumPerceiver: This figure illustrates the process of generating a summary for fact-checking using MetaSumPerceiver, integrating a fixed entailment model for accurate truthfulness labeling. Furthermore, it highlights how PPO is employed to continually refine the summary during the fact-checking process.

Training Strategy

- Preprocessing
  - Text: we obtain the text embeddings from BART
  - Image: we extract the visual features from CLIP
- After preprocessing, we perform a cross-attention between image and claim
- Then, we project X_{IC} into X_D

$$X_{IC} = ATTN(Q_{x_C}, K_{x_I}, V_{x_I})$$

$$X_{ICD} = [proj(X_{IC}, \theta)^\intercal, X_D^\intercal]^\intercal$$

$$\mathcal{L}_{\text{sum}} = -\sum_{t=1}^{T}\sum_{i=1}^{N} y_{t_i}\log(\hat{y}_{t_i})$$

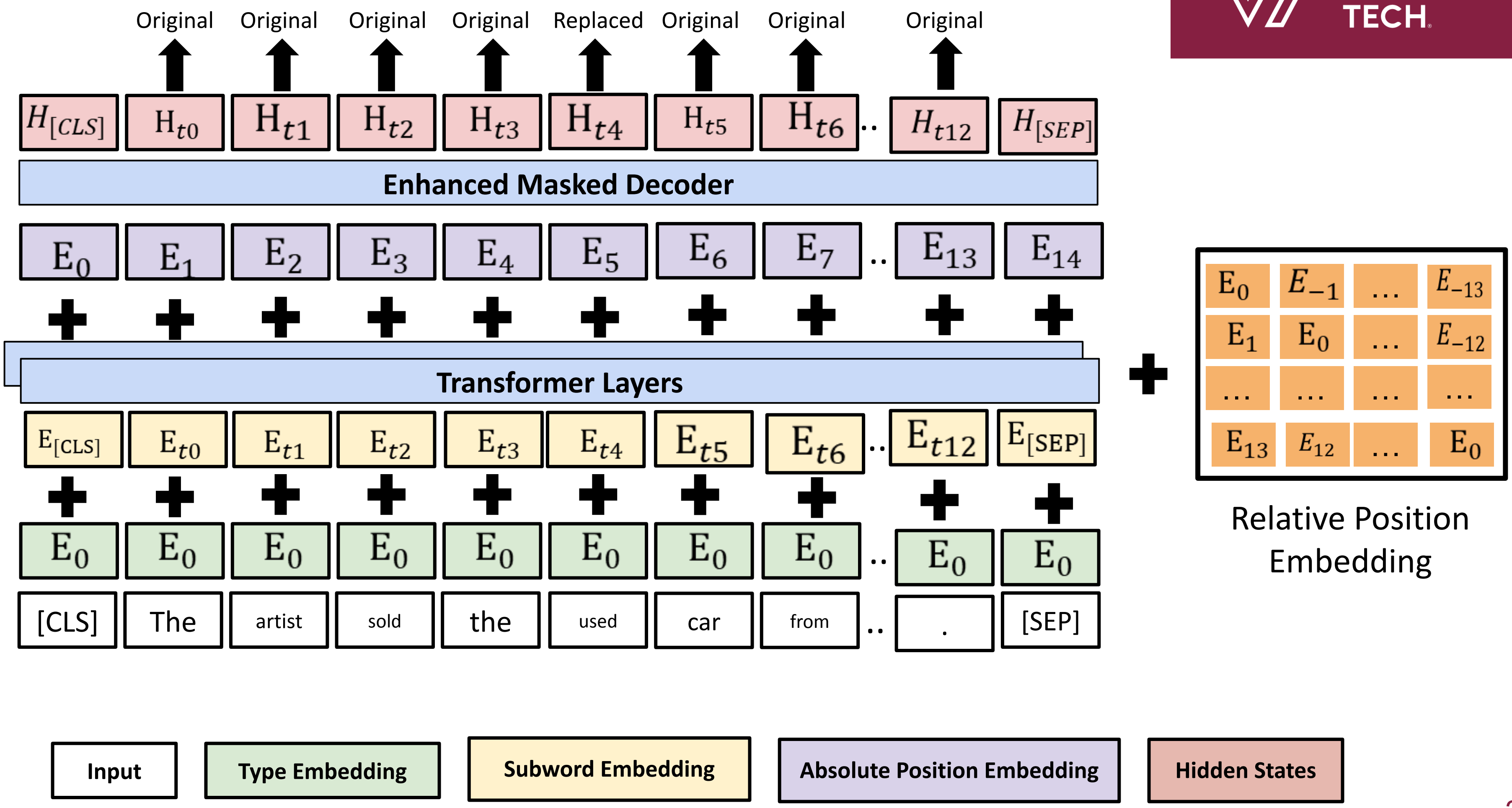
x_C : claim x_I : image X_D : doc y_{t_i}: GT summary $\hat{y}_{t_i}$: Predicted summary

Entailment model

- We use DeBERTa v3 to be our entailment model

Table 3: Comparison results on the GLUE development set.

Model #Train	CoLA Mcc 8.5k	QQP Acc 364k	MNLI-m/mm Acc 393k	SST-2 Acc 67k	STS-B Corr 7k	QNLI Acc 108k	RTE Acc 2.5k	MRPC Acc 3.7k	Avg.
$\text{BERT}_{\text{large}}$	60.6	91.3	86.6/-	93.2	90.0	92.3	70.4	88.0	84.05
$\text{RoBERTa}_{\text{large}}$	68.0	92.2	90.2/90.2	96.4	92.4	93.9	86.6	90.9	88.82
$\text{XLNet}_{\text{large}}$	69.0	92.3	90.8/90.8	**97.0**	92.5	94.9	85.9	90.8	89.15
$\text{ELECTRA}_{\text{large}}$	69.1	92.4	90.9/-	96.9	92.6	95.0	88.0	90.8	89.46
$\text{DeBERTa}_{\text{large}}$	70.5	92.3	91.1/91.1	96.8	92.8	95.3	88.3	91.9	90.00
$\text{DeBERTaV3}_{\text{large}}$	**75.3**	**93.0**	**91.8/91.9**	96.9	**93.0**	**96.0**	**92.7**	**92.2**	**91.37**



Ref: DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing

Proximal Policy Optimization

- Score from the reward model

This value is derived from the trained entailment classifier

$$r(x_C, \hat{y}_t) = P(y_{gt}|x_C, \hat{y}_t) - 0.5 * \Sigma_{y_{gt} \neq y_{pred}} \boxed{P(y_{pred}|x_C, \hat{y}_t)}, \qquad (3.5)$$

The objective is to maximize the likelihood that the generated summary for fact-checking contains the facts necessary for the model to predict the claim's ground truth label

x_C: claim $\hat{y}_t$: generated summary y_{gt}: GT label y_{pred}: Predicted label

Proximal Policy Optimization

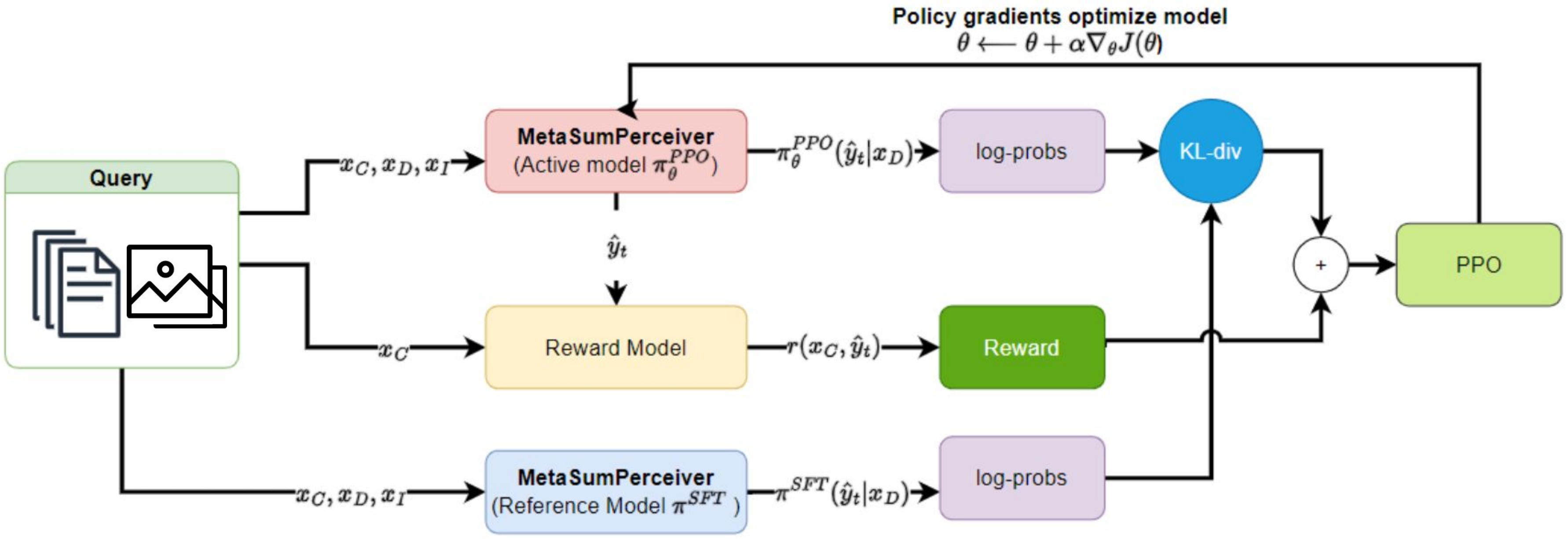
- We employ PPO as our policy gradient method for RL. PPO adds an additional term to the reward function, which imposes a penalty determined by the KL divergence between the trained RL policy summarizer and the initial supervised summarizer

$$r_{total} = r(x_C, \hat{y}_t) - \eta KL(\pi_\phi^{PPO}(\hat{y}_t|x_D), \pi^{SFT}(\hat{y}_t|x_D)), \tag{3.6}$$

This coefficient functions as an entropy boost, enhancing exploration throughout the policy domain and urging the model to engage in a diverse set of actions

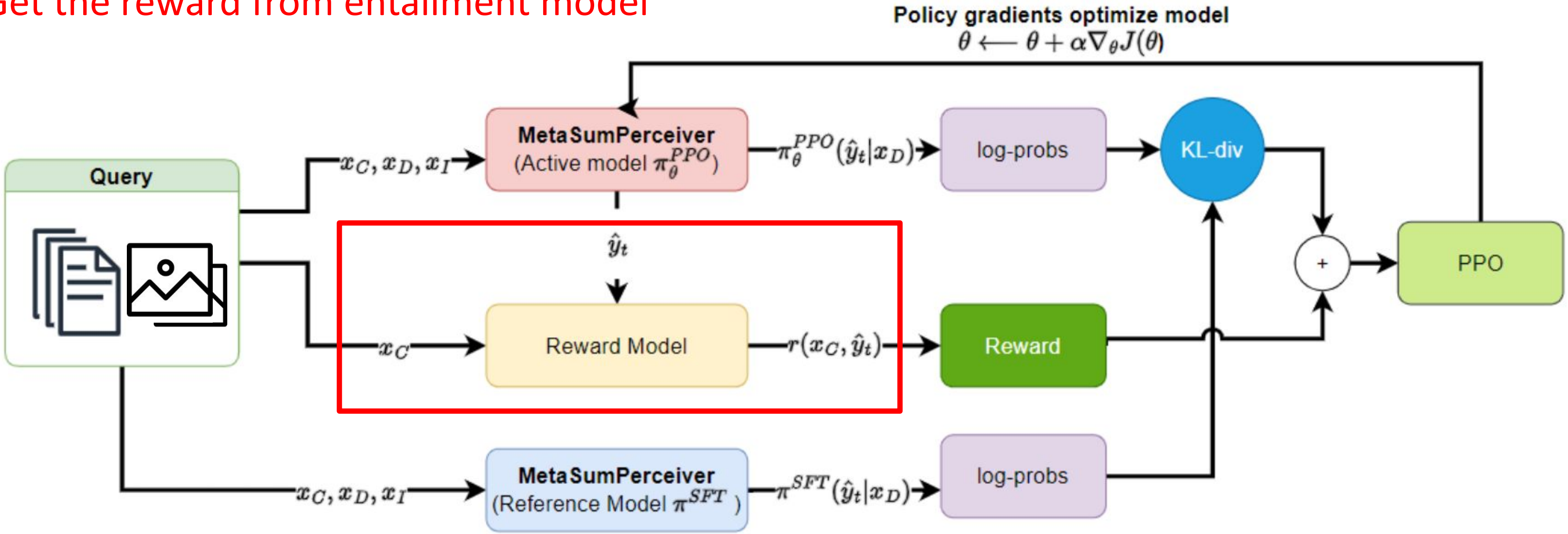
π_ϕ^{PPO}: trained RL policy summarizer π^{SFT}: initial supervised summarizer

Proximal Policy Optimization



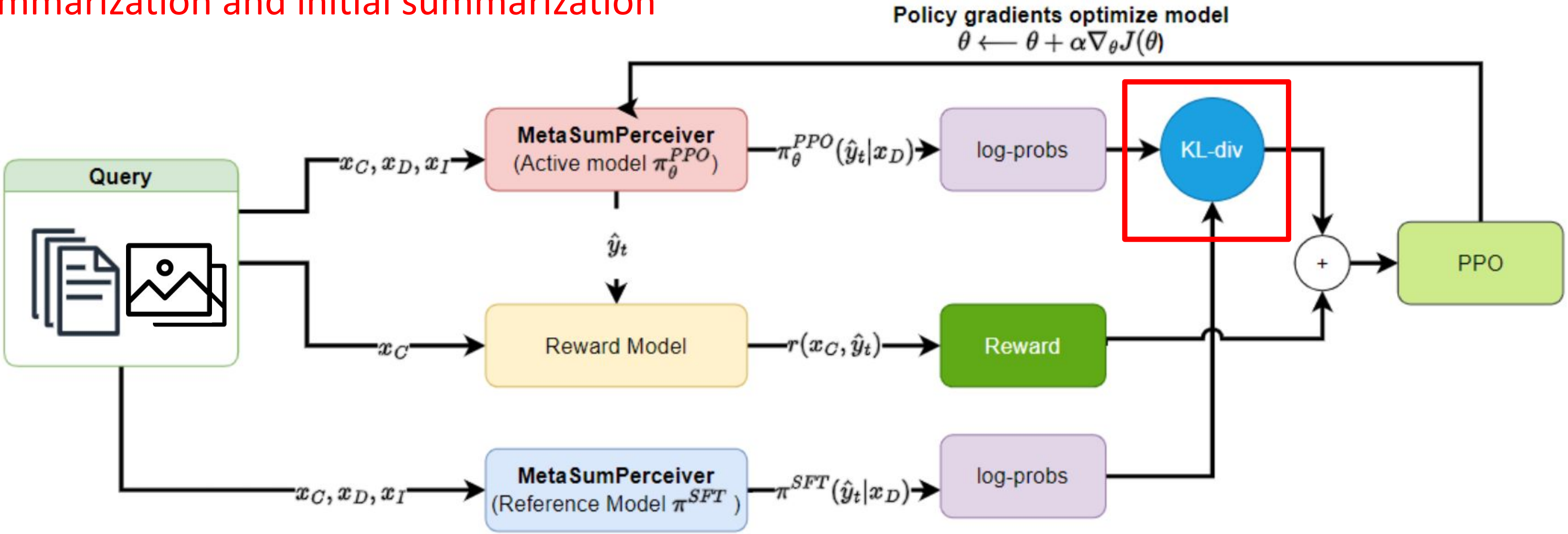
Proximal Policy Optimization

Get the reward from entailment model



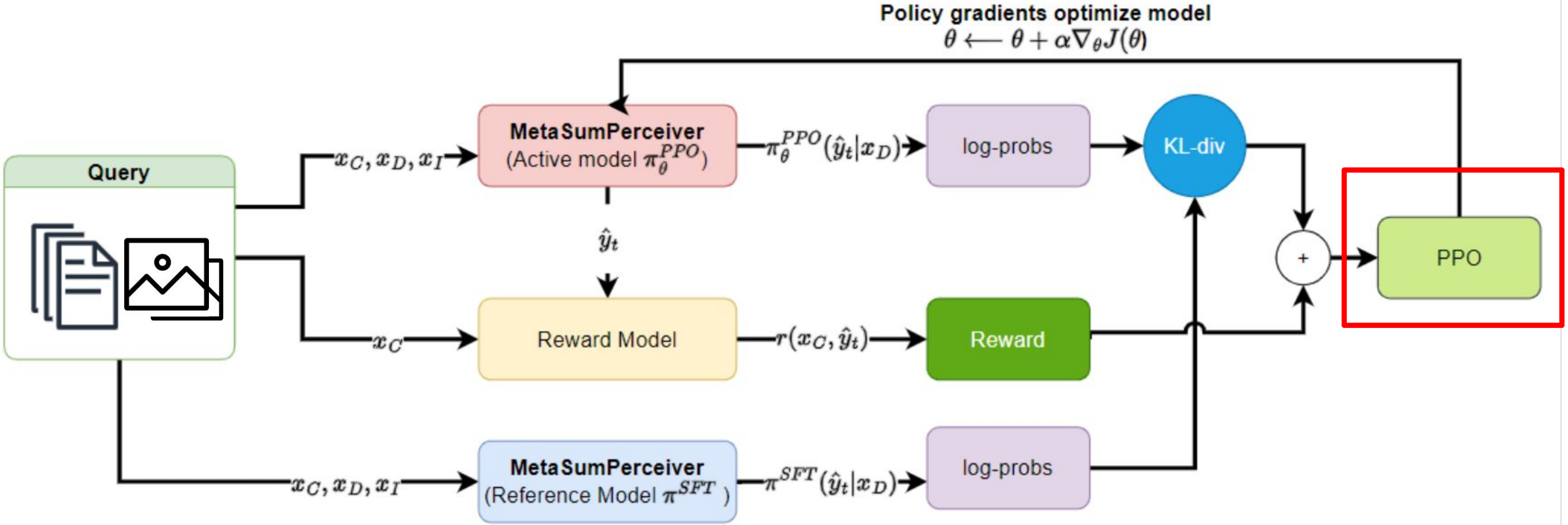
Proximal Policy Optimization

Compute the KL-div between trained summarization and initial summarization



Proximal Policy Optimization

Update summarization model



MOCHEG Benchmark

- MOCHEG is a large-scale dataset with multimodal fact checking task.
- Their claims are annotated from the humans
- Tasks:
  - Evidence retrieval
  - Claim verification
  - Explanation generation
- Classes:
  - Supported
  - NEI(Not Enough Info)
  - Refuted

Figure 1: An example of end-to-end multimodal fact-checking and explanation generation.

Ref: End-to-End Multimodal Fact-Checking and Explanation Generation: A Challenging Dataset and Models

Claim Verification

- Our method exhibits superior performance, achieving a SOTA 48.2 F-score in the MOCHEG dataset
- Compare with MOCHEG, MOCHEG's classifier relies on fixed thresholds, which may not be optimal for every situation
- Our approach involves generating summaries for fact-checking via reinforcement learning with fixed entailment models

Table 4.2: Performance of claim verification in MOCHEG. DeBERTAV3 and Llama 2 represent the fixed entailment models. Gold Evidence denotes ground truth text and image evidence while System Evidence means automatically retrieved text and image evidence.

Setting	**F-score (%)**
Our w/ Text Evidence $\rightarrow$ DeBERTAV3	42.7
Our w/ Text and Image Evidence $\rightarrow$ DeBERTAV3	45.1
Our w/ Text Evidence $\rightarrow$ Llama 2	43.9
Our w/ Text and Image Evidence $\rightarrow$ Llama 2	**48.2**
MOCHEG w/ Text Evidence	42.7
MOCHEG w/ Image Evidence	40.9
MOCHEG w/ Text and Image Evidence	**44.0**
Human w/o Evidence	20.0
Human w/ System Evidence	62.0
Human w/ Gold Evidence	**70.0**

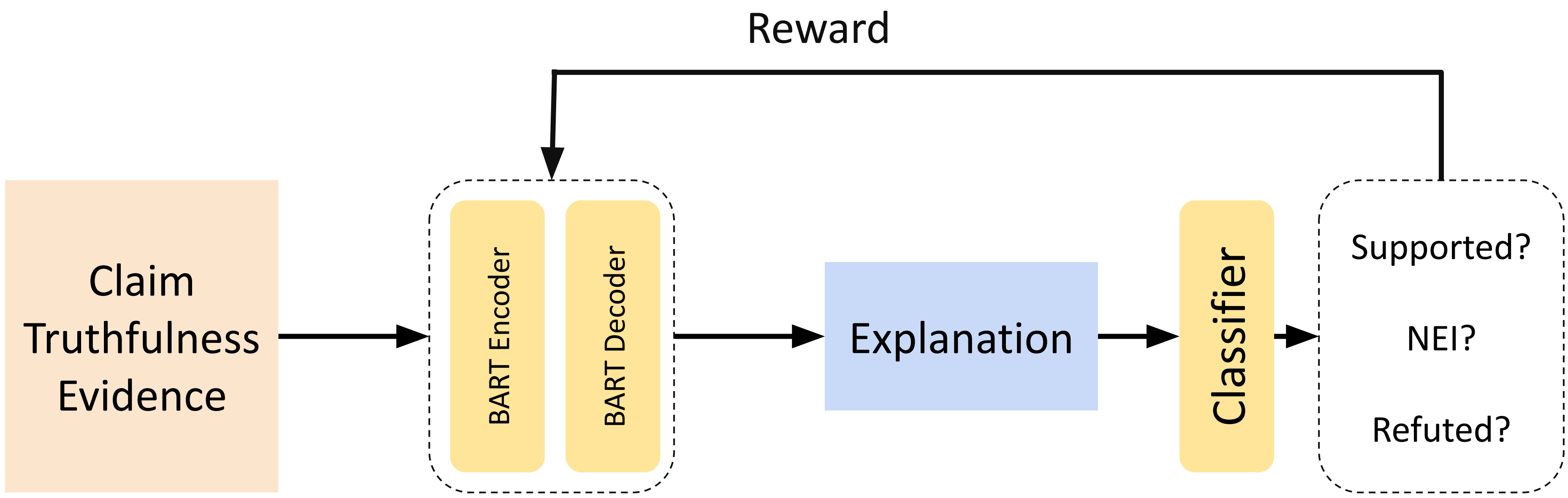
Claim Verification

- The zero-shot Llama 2 entailment surrogate model surpasses DeBERTAV3 in performance
- However, in the refuted and NEI labels, performance of Llama 2 is not better than DeBERTAV3

Table 4.1: Performance of claim verification in MOCHEG with our method. We separately calculate the precision and recall in supported, refuted, and NEI claim labels. We compare our method with published baselines in Table 4.2.

Setting	Accuracy (%)	Precision (%) Supported	Precision (%) Refuted	Precision (%) NEI	Recall (%) Supported	Recall (%) Refuted	Recall (%) NEI
Our w/ Text Evidence → DeBERTAV3	43.7	79.2	66.9	**33.9**	40.5	30.6	25.8
Our w/ Text and Image Evidence → DeBERTAV3	50.8	83.4	**69.3**	27.3	42.9	34.2	30.9
Our w/ Text Evidence → Llama 2	46.7	80.4	68.1	31.5	37.2	35.4	31.5
Our w/ Text and Image Evidence → Llama 2	**53.7**	**87.3**	60.3	32.4	**48.3**	**36.9**	**34.8**

Explanation Generation



To ensure the explanation is consistent with the truthfulness label. The method uses RL to optimize the generation model.

Explanation Generation

- Our model outperforms MOCHEG's evidence-retrieval based method on the rationale generation task
- Comparison:
  - MOCHEG relies on retrieval from a pool of multimodal documents
  - Our method based on summarization may rephrase the same evidence

Table 4.3: Performace of explanation generation. Our system outperforms MOCHEG on equivalent settings.

Setting	ROUGE 1 (%)	ROUGE 2 (%)	ROUGE L (%)	BLEU (%)	BERTScore (%)
MOCHEG w/ Gold Evidence, Gold Truthfulness	**45.5**	**27.3**	**35.4**	**21.8**	**89.0**
MOCHEG w/ Gold Evidence, System Truthfulness	43.8	26.3	34.1	20.8	88.8
MOCHEG w/ System Evidence, Gold Truthfulness	35.5	17.4	**26.0**	**10.9**	87.0
MOCHEG w/ System Evidence, System Truthfulness	33.8	16.5	24.8	10.0	86.9
Our w/ System Evidence, Gold Truthfulness	**36.7**	**17.9**	25.7	10.7	**87.3**
Our w/ System Evidence, System Truthfulness	34.3	16.8	25.4	10.4	87.1

Claim	Summary for fact-checking	Image evidence	Truthfulness
#1 On March 21, 2021, an adviser to former U.S. President Donald Trump said he was preparing to launch his own social media platform.	… Donald Trump has been banned from Twitter, Facebook, and Instagram since the Jan. 6 riot at the Capitol, but a senior advise to the former president says he plans to launch **his own social media site in the next few months** …		**Supported**
#2 During the pandemic, I have been in D.C. voting regularly.	… Rep. Mark Pocan of Wisconsin says he's been " in DC voting regularly" since the COVID-19 pandemic hit Congress in May 2020, **but his office says that's a bit of a stretch.** …		**NEI**
#3 U.S. Sen. Elizabeth Warren said or argued to the effect that 'taxpayers must fund sex reassignment surgery.	Warren, like many politicians, has supported policies aimed at promoting equality and non-discrimination for the LGBTQ+ community, **but attributing such a specific statement to her lacks verifiable sources.** …		**Rejected**
#4 By spring 2020, the sun had entered a lockdown period where its solar activity decreased to the point that famine, earthquakes, and freezing weather threatened life on Earth.	… Solar minimums are important because **they can affect how satellites orbit the Earth, as well as the intensity of cosmic rays and the brightness of sun's rays....**		**Rejected**

Figure 4.1: Examples of Multimodal Fact-Checking. The truthfulness column shows gold labels.

Ablation

- We conducted ablation experiments for claim verification on our Multi-News-Fact-Checking dataset
- We achieve balanced accuracy in both precision and recall, underscoring our method's ability to clearly differentiate between truthful and untruthful labels without bias in predictions

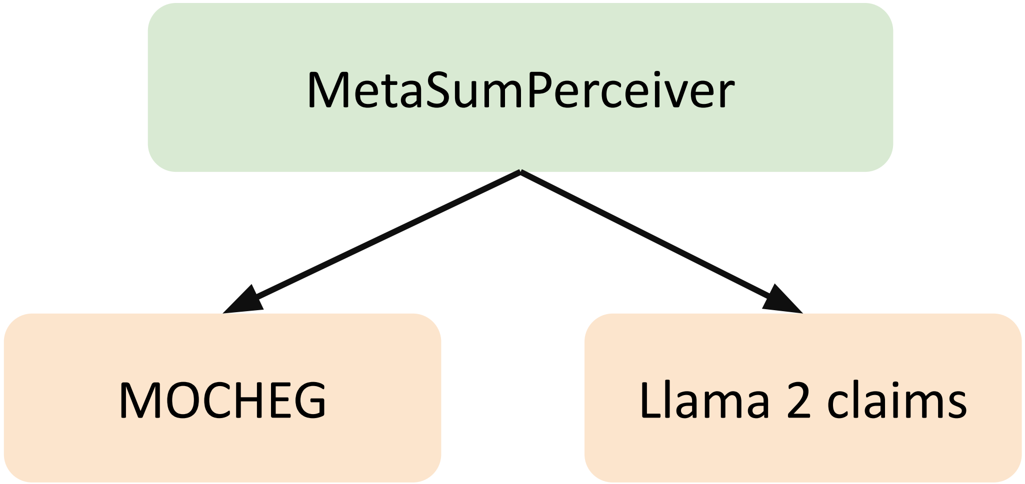
Table 4.7: Performance of claim verification in Multi-News-Fact-Checking dataset. DeBERTAV3 and Llama 2 serve as the fixed entailment models. Gold Evidence refers to claim labels based on gold standards, whereas System Evidence indicates our predicted claim labels.

Setting	**F-score (%)**
Our w/ DeBERTAV3	39.9
Our w/ Llama 2	$\boxed{\textbf{43.4}}$
Our w/ Llama 2(No RL)	41.8
PEGASUS w/ DeBERTAV3	25.4
PEGASUS w/ Llama 2	$\boxed{\textbf{30.8}}$
T5 large w/ DeBERTAV3	28.5
T5 large w/ Llama 2	$\boxed{\textbf{32.7}}$
Human w/o Evidence	23.0
Human w/ System Evidence	65.0
Human w/ Gold Evidence	**76.0**

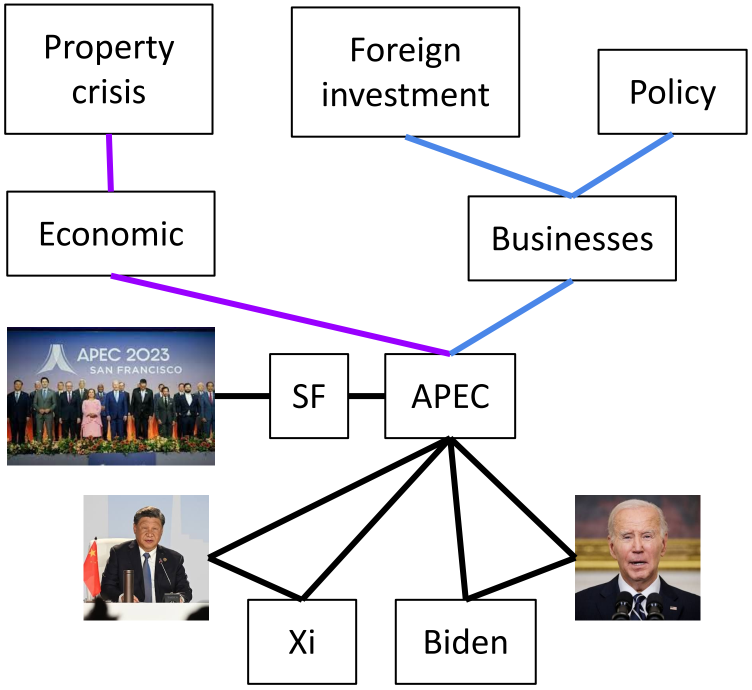
Ablation

Table 4.8: Performance of claim verification in Multi-News-Fact-Checking dataset. We compare our method with Llama 2, other offline summarization models.

Setting	Accuracy (%)	Precision (%) Entailment	Precision (%) Contradiction	Precision (%) Neutral	Recall (%) Entailment	Recall (%) Contradiction	Recall (%) Neutral
PEGASUS → DeBERTAV3	33.2	64.2	14.7	21.5	37.3	12.4	11.9
PEGASUS → Llama 2	39.5	37.4	23.1	42.8	27.6	24.3	24.0
T5 large → DeBERTAV3	34.8	62.8	17.5	26.2	33.0	18.5	18.2
T5 large → Llama 2	37.2	40.2	32.8	**48.0**	30.5	26.4	26.8
Our → DeBERTAV3	36.7	**75.5**	28.9	27.5	41.0	21.7	**47.2**
Our (No RL) → Llama 2	42.6	41.0	**53.7**	34.6	54.8	37.8	29.6
Our → Llama 2	$\boxed{\mathbf{45.6}}$	49.2	48.7	33.6	**56.9**	**44.1**	28.4



Now, we have the model and evaluate on MOCHEG and Llama 2 claims. However, we want to understand the performance of our model in multimodal multi-doc claims. We create the multimodal multi-doc claim generation.



Agenda

- Research Background
- Limitations & Challenges
  - Multimedia Event Extraction
  - Multimodal Summarization
- Methods:
  - Evidence Summarization
  - Multimodal Multi-doc Claim Generation
- Results
- Conclusion

KG2Claim - Pipeline

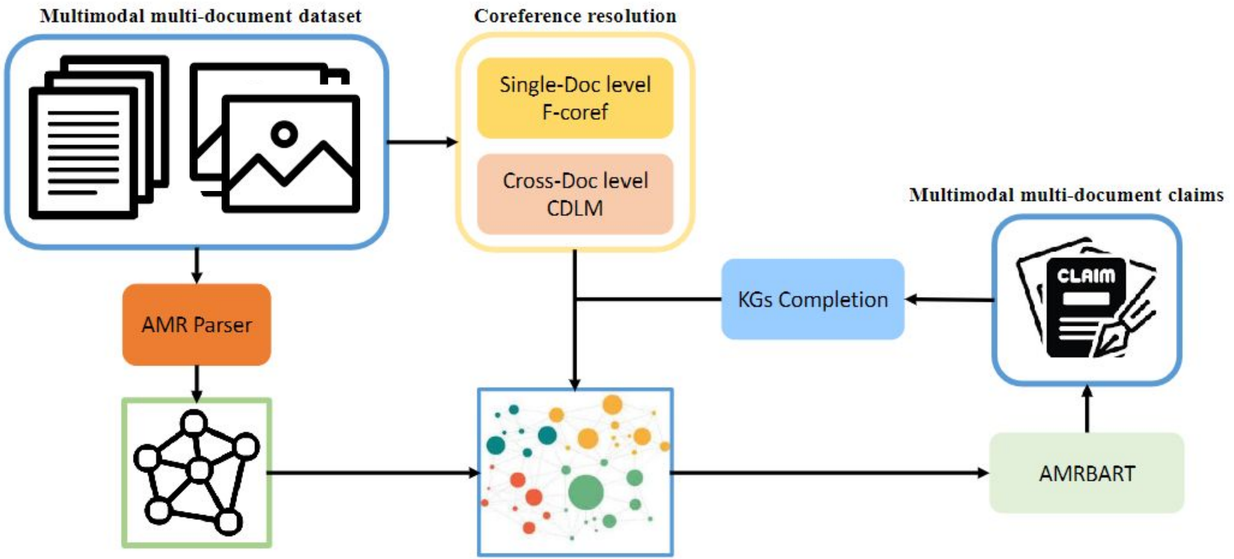


Figure 1.1: Pipeline of AMRKG2Claim: The elements encompass an AMR parser, single-document and cross-document coreference resolution, Knowledge graphs completion with LLMs, and AMRBART, generating the multimodal multi-document claims.

AMR Parsing

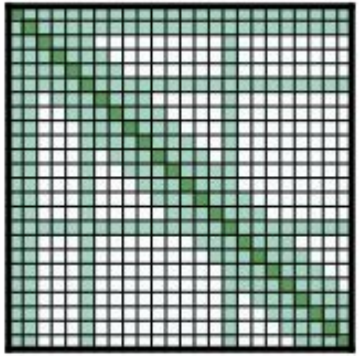
- NewsStories dataset:
    - 350,000 clusters, and each cluster has over 2 news articles saying same news
    - Each cluster has images, articles and videos

	GoodNews [4]	NYTimes 800K [36]	Visual News [24]	NewsStories Unfiltered	NewsStories Filtered
# Media channels	1	1	4	28,215	46
# Story clusters	-	-	-	-	350,000
# Articles	257,033	444,914	623,364	31,362,735	931,679
# Images	462,642	792,971	1,080,595	22,905,000	754,732
# Videos	0	0	0	1,020,363	333,357
Avg article length	451	974	773	446	584

Ref: NewsStories: Illustrating articles with visual summaries

Single-document coreference resolution

- We use F-coref to extract the coreference resolution in the single document
- F-coref:
  - Longformer encoder
  - Mention scoring function f_m
  - Pairwise antecedent scoring function f_a
- The advantage of longformer:
  - Reducing the time complexity of self-attention
  - Extending the input sequence length



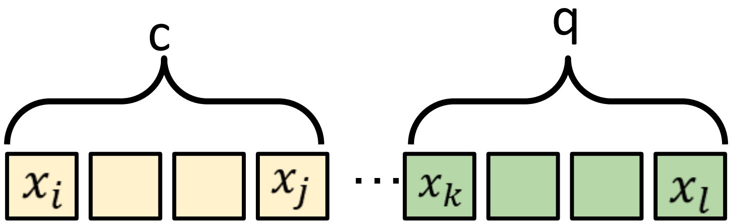
(d) Global+sliding window

Ref: F-coref: Fast, Accurate and Easy to Use Coreference Resolution and Longformer: The Long-Document Transformer

Single-document coreference resolution

- After encoding the text into vectors, F-coref will compute mention scoring function and antecedent scoring function
- Final pairwise score for a coreference link is composed by
  - The score of q being a mention, c being a mention
  - The score of how likely is c being an antecedent of q

$$c = (x_i, x_j), q = (x_k, x_l)$$



$$F(c,q) = \begin{cases} f_m(c) + f_m(q) + f_a(c,q) & c \neq \varepsilon \\ 0 & c = \varepsilon \end{cases}$$

where ε is the null antecedent.

Ref: F-coref: Fast, Accurate and Easy to Use Coreference Resolution

Cross-document coreference resolution

- Pretraining approach in CDLM:
  - Pretraining over sets of related docs that contain the overlapping information

Doc 1: "*Harry Shearer* is ***suing*** *Vivendi's* **Universal Music** for $125 million for allegedly fraudulent ..."

Doc 2:"...*Harry Shearer* ***alleges*** **parent company of Universal Music and StudioCanal** withheld millions..."

Doc 3: "*Shearer* was then ***joined in the lawsuit*** with ***StudioCanal*** and its French parent ***Vivendi*** by his co-stars"

Ref: CDLM: Cross-Document Language Modeling

CDLM

The final produced pairwise-mention representation

After finishing coreference resolution, we add **:coref** in our KGs

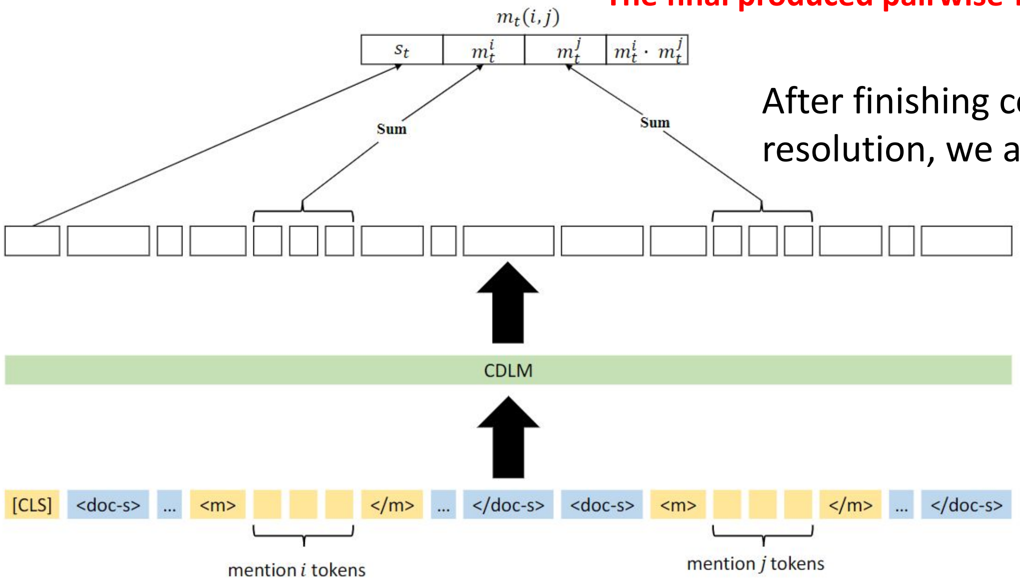
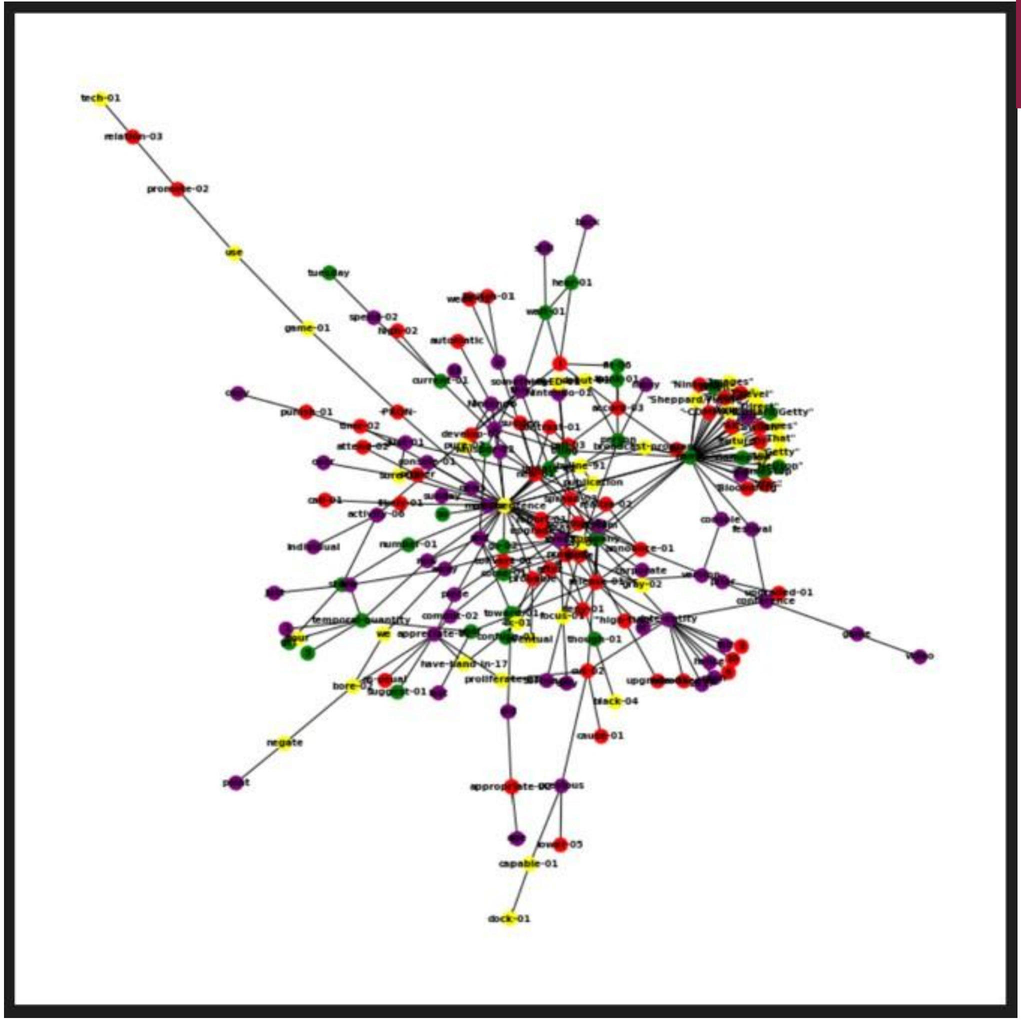
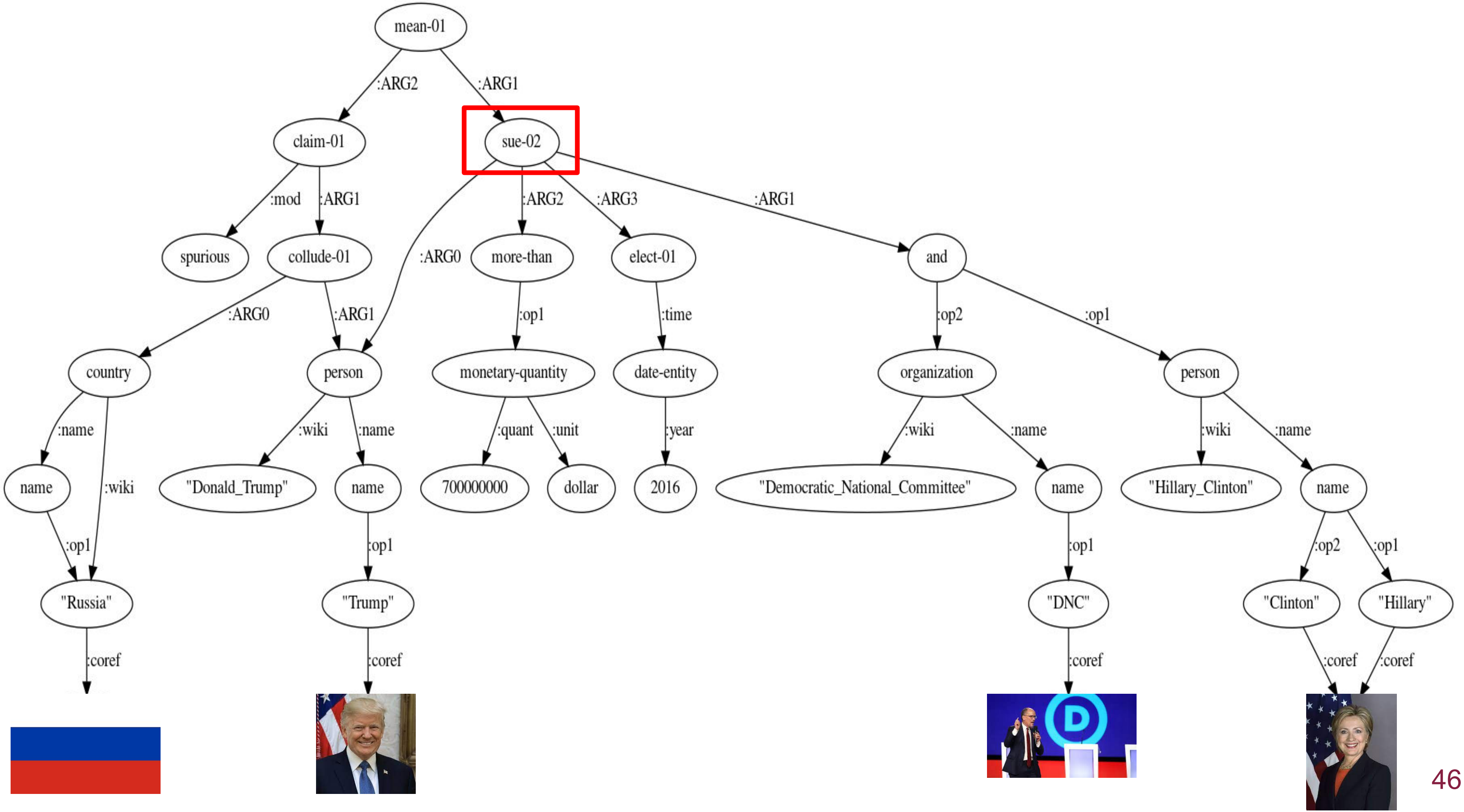


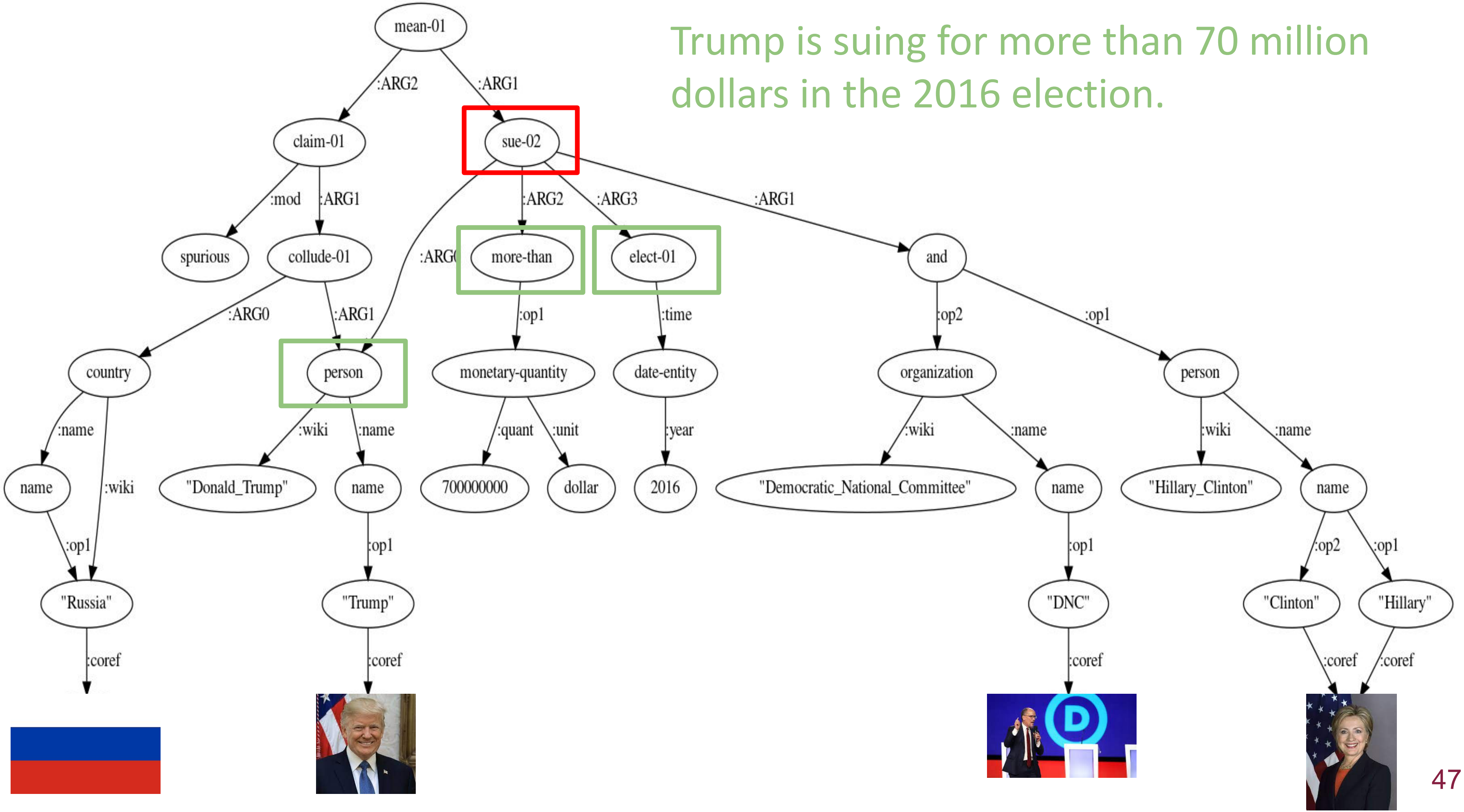
Figure 3.1: The CDLM model utilizes pairwise mention representation for coreference resolution. m_t^i, m_t^j and s_t are the cross-document contextualized representation vectors for mentions i and j, and of the [CLS] pairwise-mention representation. The tokens colored in yellow represent global attention, and tokens colored in blue represent local attention.

Ref: CDLM: Cross-Document Language Modeling

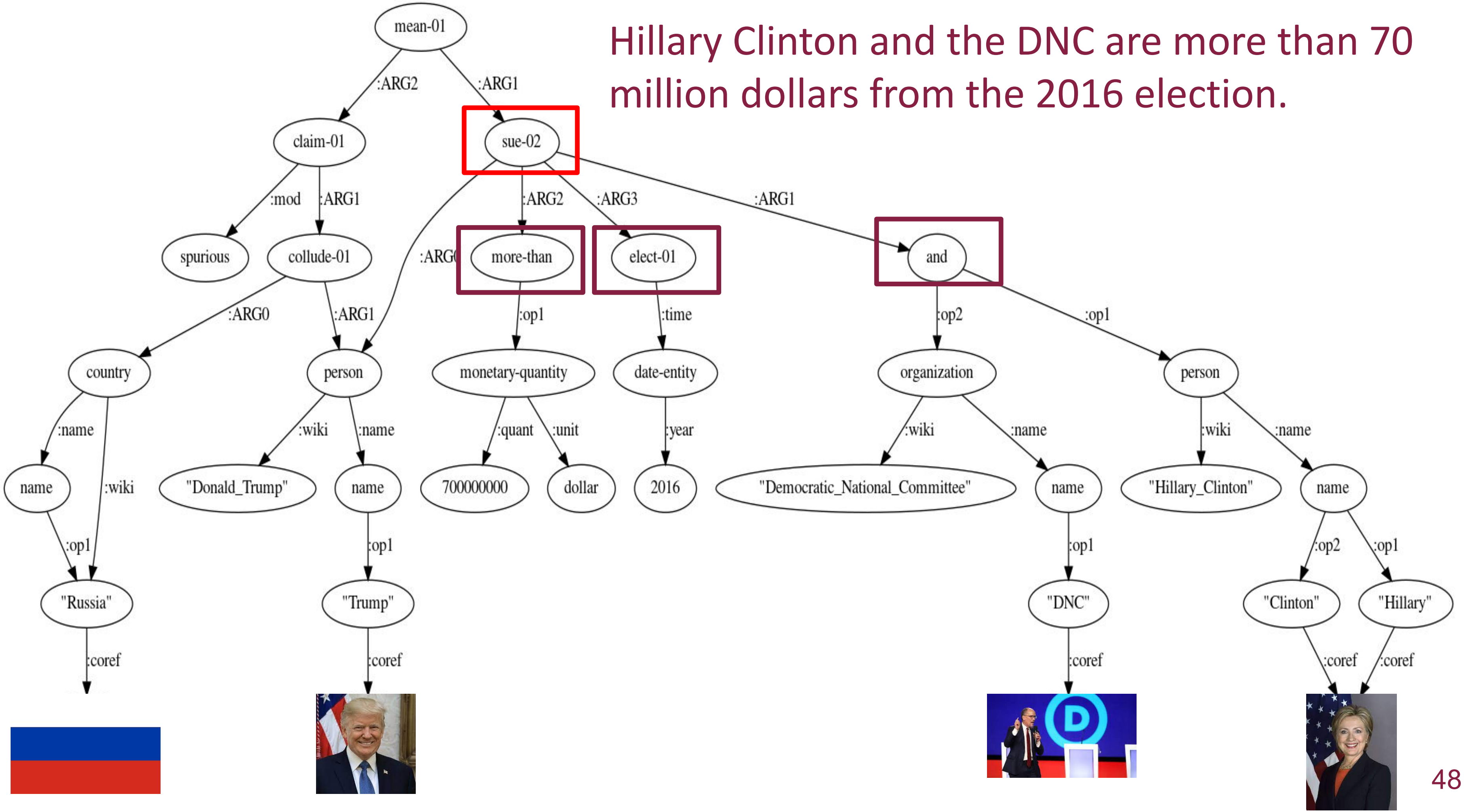

tech-01
relation-03
use
game-01
individual
negate
capable-01
dock-01
cause-01
black-04
festival
have-hand-in-17


mean-01
:ARG2
:ARG1
claim-01
sue-02
:mod
:ARG1
spurious
collude-01
:ARG0
:ARG2
:ARG3
:ARG1
more-than
elect-01
and
:ARG0
:ARG1
:op1
:time
:op2
:op1
country
person
monetary-quantity
date-entity
organization
person
:name
:wiki
:wiki
:name
:quant
:unit
:year
:wiki
:name
:wiki
:name
name
"Donald_Trump"
name
700000000
dollar
2016
"Democratic_National_Committee"
name
"Hillary_Clinton"
name
:op1
:op1
:op1
:op2
:op1
"Russia"
"Trump"
"DNC"
"Clinton"
"Hillary"
:coref
:coref
:coref
:coref
:coref

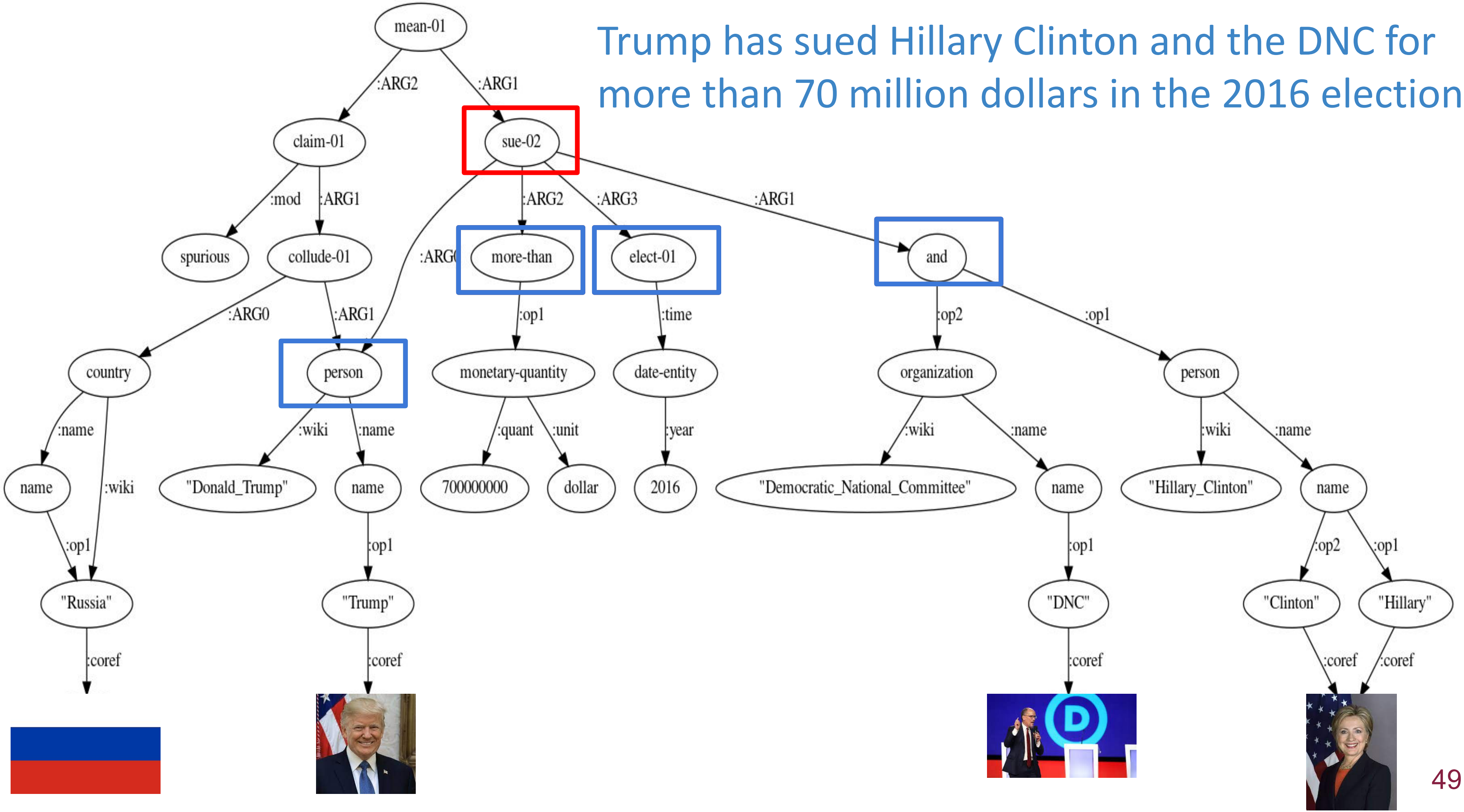
Trump is suing for more than 70 million dollars in the 2016 election.



Hillary Clinton and the DNC are more than 70 million dollars from the 2016 election.

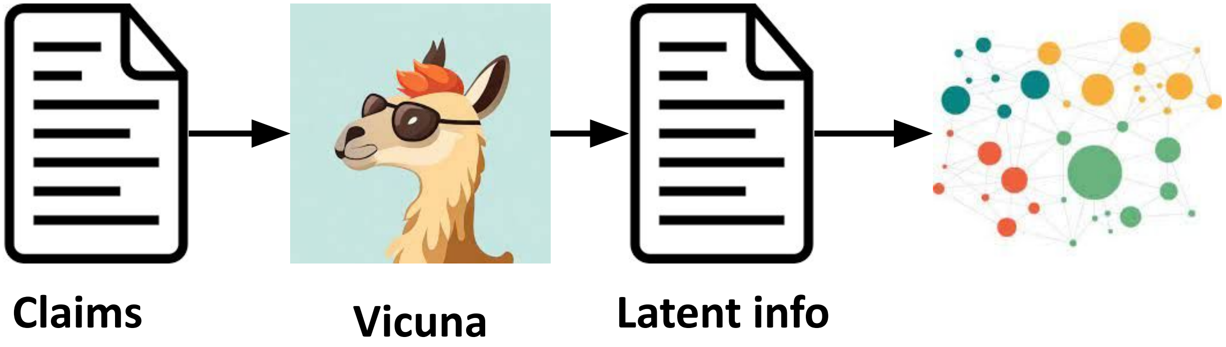


Trump has sued Hillary Clinton and the DNC for more than 70 million dollars in the 2016 election.



KG Completion

- After generating the claims from AMRBART, we feed these claims into Vicuna to induce the latent information
- To furnish the KG, we parse the induced claims and insert into the KG



Ref: Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality

KG Completion

- **Existing KG:**
  - <Ronaldo, Plays for, Manchester United>
  - <Messi, Plays for, Paris Saint-Germain>
- **Content:**
  - In a shocking turn of events, Lionel Messi has decided to part ways with Paris Saint-Germain (PSG) and join the ranks of Manchester United. The legendary footballer expressed his excitement about the new chapter in his career with the English club.
- **Updated KG:**
  - <Ronaldo, Plays for, Manchester United>
  - <Messi, Plays for, Manchester United>
  - <Messi, Formerly played for, Paris Saint-Germain>

Claims Analysis

- To make sure the claims are check-worthy, we test the claims on the claim detection task
- Classes:
  - UFS: These are factual claims but not check-worthy
  - CFS: These sentences contain factual claims that the general public will be interested in learning about their veracity
  - NFS: These sentences do not contain any factual claims

Table 4.1: Performance of claim detection.

Claim classes	(%)
Unimportant Factual Sentence (UFS)	17.67
Check-worthy Factual Sentence (CFS)	$\boxed{68.6}$
Non-factual Sentence (NFS)	13.71

Ref: A Benchmark Dataset of Check-worthy Factual Claims

Claims Truthfulness Label Test

- Our assumption is that these claims are entailed by the news
- To verify this, we test these claims with MetaSumPerceiver
- We suspect that the discrepancy may arise from the edge labels in the KGs

Truthfulness labels	(%)
Entailment label	74.3
Neutral label	8.24
Contradiction label	17.46

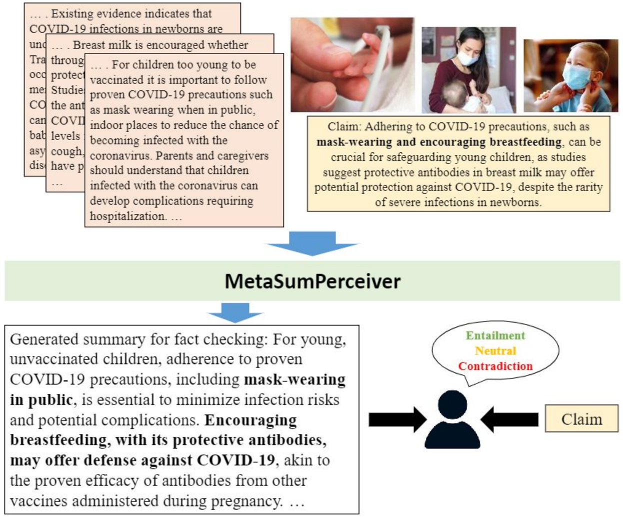
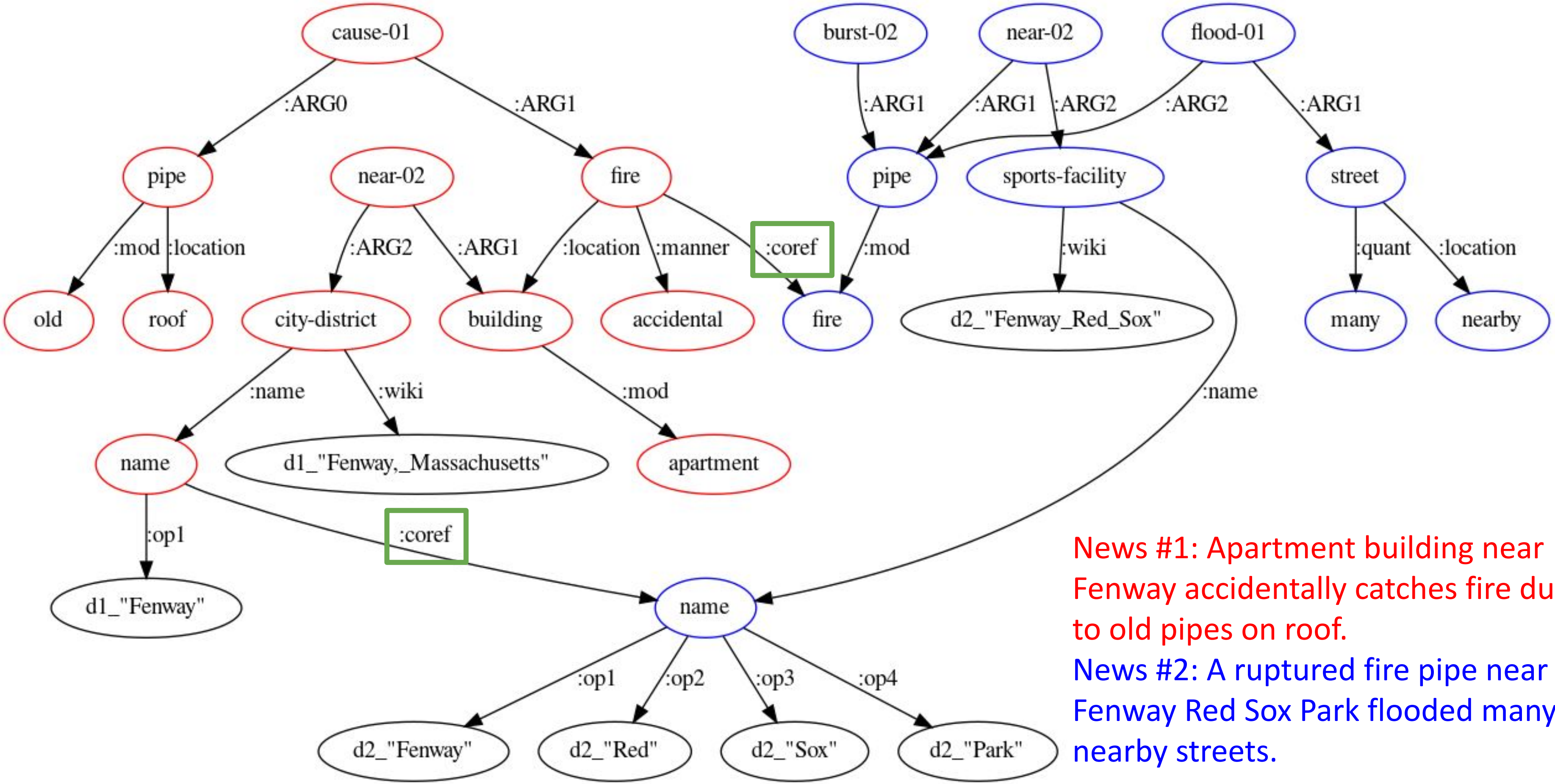


Figure 1: Overview of MetaSumPerceiver: Using inputs such as documents, images, and claims, MetaSumPerceiver generates summaries to facilitate fact-checking. In this example, the summary for fact-checking provides evidence and establishes that the claim in question is entailed by the evidence.

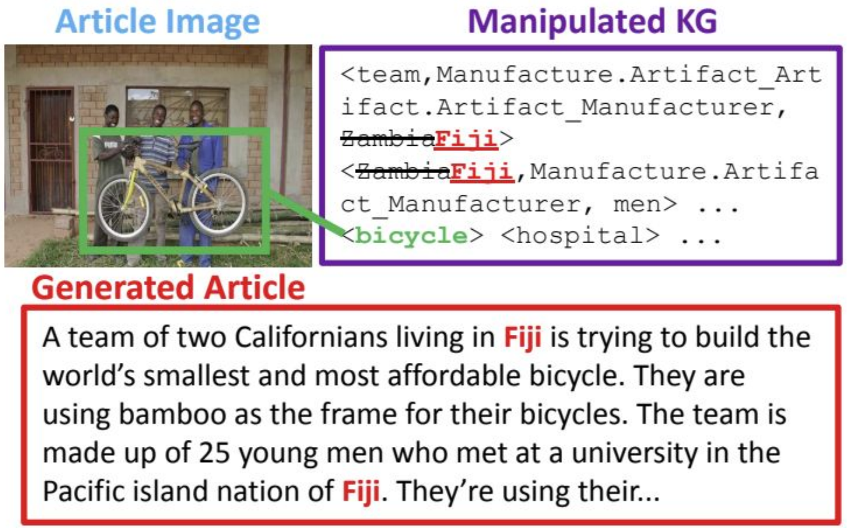


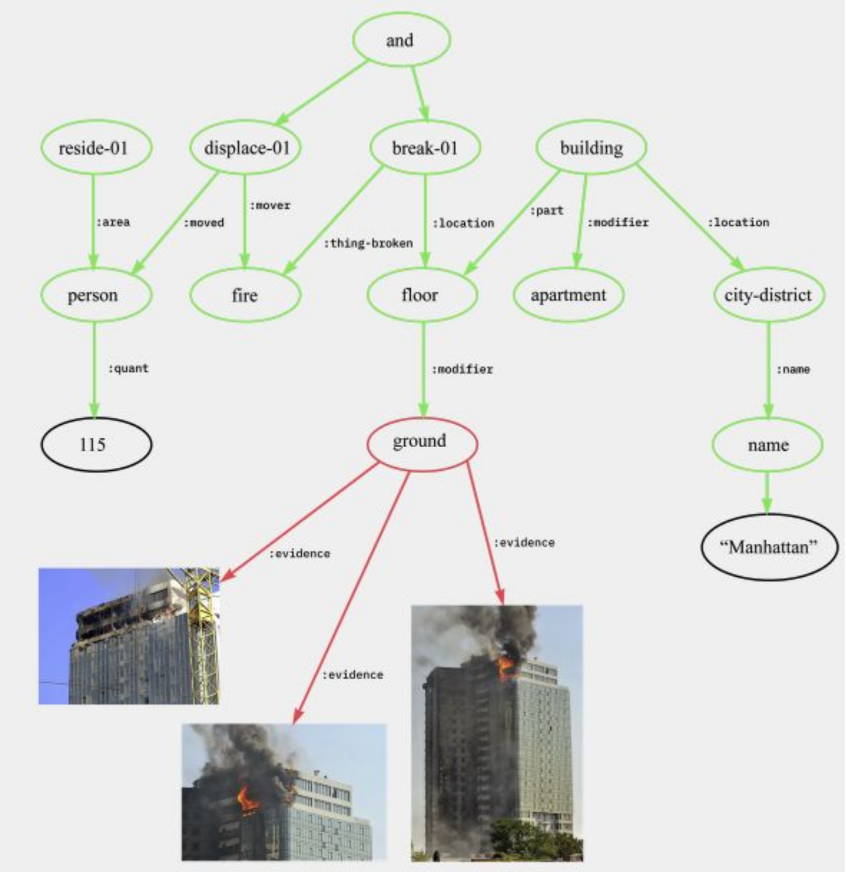
News #1: Apartment building near Fenway accidentally catches fire due to old pipes on roof.

News #2: A ruptured fire pipe near Fenway Red Sox Park flooded many nearby streets.

Future work

- Fine-grained entailment model
- Manipulated claims





Ref: InfoSurgeon: Cross-Media Fine-grained Information Consistency Checking for Fake News Detection

Conclusion

- We introduce:
  - MetaSumPerceiver, a summarization model designed to produce concise, informative summaries for claim fact-checking from complex multimodal datasets
  - KG2Claim, a text generation pipeline to produce the claims from the knowledge graphs. Our text generation approach can generate claims related to multimodal multi-document information
- Our approach surpasses the SOTA method by 4.2% in the claim verification task on the MOCHEG
- We also demonstrate the effectiveness of our generated claims for fact-checking tasks, showcasing the strong performance of our model in this regard

MetaSumPerceiver: Multimodal Multi-Document Evidence Summarizer for Fact-Checking

Anonymous EACL submission

Abstract

Fact-checking real world claims often requires reviewing multiple multimodal documents in order to assess the claim's truthfulness, a highly laborious and time consuming task. In this paper, we present a summarization model crafted to generate claim-specific summaries useful for fact-checking from multimodal multi-document datasets. The model takes inputs in the form of documents, images, and a claim, with the objective of assisting in fact-checking tasks. We introduce a dynamic perceiver-based model that is able to handle inputs from multiple modalities of arbitrary lengths. To train our model, we leverage a novel reinforcement learning-based entailment objective in order to generate summaries that provide evidence distinguishing between different truthfulness labels. To assess the efficacy of our approach, we conduct experiments on both an existing benchmark as well as a new dataset of multidocument claims which we contribute. Our approach outperforms the SOTA approach by 4.2% in the

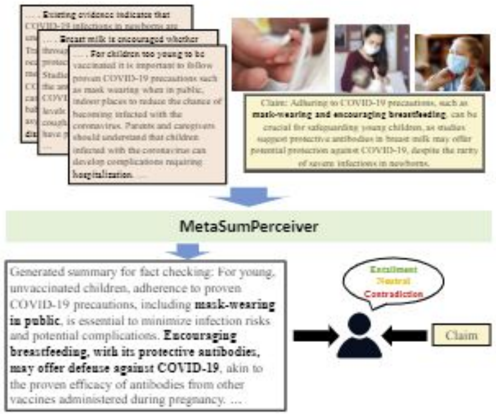


Figure 1: Overview of MetaSumPerceiver: Using inputs such as documents, images, and claims, MetaSumPerceiver generates summaries to facilitate fact-checking. In this example, the summary for fact-checking provides evidence and establishes that the claim in question is entailed by the evidence.

thank you