NEURAL INFORMATION PROCESSING SYSTEMS

UNIVERSITY OF CAMBRIDGE

Queen Mary University of London

AVerImaTeC: A Dataset for Automatic Verification of Image-Text Claims with Evidence from the Web

Rui Cao♡, Zifeng Ding♡, Zhijiang Guo♡, Michael Schlichtkrull♢, Andreas Vlachos♡
University of Cambridge♡, Queen Mary University of London♢

Fact-checking

- Definition: assessing and arguing for the factuality of claims
  - Curb the spread of misinformation: great efforts from journalisms

Fact-checking

- Definition: assessing and arguing for the factuality of claims
  - Curb the spread of misinformation
- Vast volume of online information: automated fact-checking (AFC)

AFC Datasets

- Most datasets focus exclusively on textual claims
  - Approximately **80%** of online claims are multimodal involving both text and media

Dufour, Nicholas, et al. "Ammeba: A large-scale survey and dataset of media-based misinformation in-the-wild." *arXiv preprint arXiv:2405.11697* 1.8 (2024).

AFC Datasets

- Most datasets focus exclusively on textual claims
  - Approximately **80%** of online claims are multimodal involving both text and media
  - Images are the most prevalent media type

Dufour, Nicholas, et al. "Ammeba: A large-scale survey and dataset of media-based misinformation in-the-wild." *arXiv preprint arXiv:2405.11697* 1.8 (2024).

AFC Datasets

- Most datasets focus exclusively on textual claims
- Existing image-text AFC datasets
  - Many are synthetic: discrepancies between synthetic data and real-world data

Zeng, Fengzhu, et al. "Multimodal misinformation detection by learning from synthetic data with multimodal LLMs." *arXiv preprint arXiv:2409.19656* (2024).

AFC Datasets

- Most datasets focus exclusively on textual claims
- Existing image-text AFC datasets
  - Many are synthetic
  - Real image-text claims: context-dependent

AFC Datasets

- Most datasets focus exclusively on textual claims
- Existing image-text AFC datasets
  - Many are synthetic
  - Real image-text claims: context-dependence
  - Lack annotated evidence: impossible to evaluation models' reasoning process

AVerImaTeC: Automated VERification of IMAge-TExt Claim

- 1,297 real-world image-text claims annotated with evidence from the web

AVerImaTeC: Automated VERification of IMAge-TExt Claim

- 1,297 real-world image-text claims annotated with evidence from the web
- We employ crowdworkers to decompose the verification of image-text claims into a sequence of question-answering with evidence from the web

AVerImaTeC: Automated VERification of IMAge-TExt Claim

- 1,297 real-world image-text claims annotated with evidence from the web
- We employ crowdworkers to decompose the verification of image-text claims into a sequence of question-answering with evidence from the web
- We assess the consistency of the annotation in AVerImaTeC via inter-annotator studies, achieving a $\kappa = 0.742$ on verdicts and 74.7% consistency on QA pairs

AVerImaTeC

- The rationale for verification in the fact-checking article decomposed into a sequence of QA pairs
  - Potentially multimodal, evidence from the web

AVerImaTeC

- Abundant metadata

Claim Text: *Kamala Harris with her parents and she is not a black American.*

Claim Image: 

Claim Date: *2020.08.12*
Claim Source: *Facebook*
Refuting Reasons: *Misuse of images; Textually refuted*
Image Misuse: *Out-of-Context*
Claim Type: *...*

Q1: *What is the date of the claim image being published?*
Related Image to Q1: 

A1: *The image was taken in 2016.*

Q2: *Were Kamala Harris parents alive in 2016?*
A2: *Her mother died in 2009.*

Q3: *Who are the people shown in the Image of the claim?*
Related Image to Q3: 

A3: *Harris (center) was with her supporters Suneil Parulekar (left) and Rohini Parulekar (right) at the 2016 Pratham gala.*

Q4: *What is Kamala Harris's ethnic background?*
A4: *She has Jamaican and Indian parents.*

Q5: *Can a person be seen as a black American if they have either Indian or Jamaican parents?*
A5: *Yes. Not all black people are African American.*

Verdict: *Refuted*

Justification: *The image was interpreted out-of-context. The image was taken in 2016, while Harris' mother died in 2009. It is impossible that she was with her parents in the image. Besides, another evidence proves the image shows Harris was with her supporters at the 2016 Pratham gala, rather than with her parents. Harris has Jamaican and Indian heritage, proving she is a black American. Therefore, the textual part of the claim, "she is not a black American" is refuted.*

AVerImaTeC

- A veracity label based on retrieved evidence

Claim Text: *Kamala Harris with her parents and she is not a black American.*

Claim Image: 

Claim Date: *2020.08.12*

Claim Source: *Facebook*

Refuting Reasons: *Misuse of images; Textually refuted*

Image Misuse: *Out-of-Context*

Claim Type: *...*

Q1: *What is the date of the claim image being published?*

Related Image to Q1: 

A1: *The image was taken in 2016.*

Q2: *Were Kamala Harris parents alive in 2016?*

A2: *Her mother died in 2009.*

Q3: *Who are the people shown in the Image of the claim?*

Related Image to Q3: 

A3: *Harris (center) was with her supporters Suneil Parulekar (left) and Rohini Parulekar (right) at the 2016 Pratham gala.*

Q4: *What is Kamala Harris's ethnic background?*

A4: *She has Jamaican and Indian parents.*

Q5: *Can a person be seen as a black American if they have either Indian or Jamaican parents?*

A5: *Yes. Not all black people are African American.*

Verdict: *Refuted*

Justification: *The image was interpreted out-of-context. The image was taken in 2016, while Harris' mother died in 2009. It is impossible that she was with her parents in the image. Besides, another evidence proves the image shows Harris was with her supporters at the 2016 Pratham gala, rather than with her parents. Harris has Jamaican and Indian heritage, proving she is a black American. Therefore, the textual part of the claim, "she is not a black American" is refuted.*

AVerImaTeC

- A textual justification explaining how the verdict is reached

Evaluation

- Retrieved evidence: a separate reference-based evaluation for textual and visual evidence

Evaluation

- Retrieved evidence: a separate reference-based evaluation for textual and visual evidence
- Veracity: conditional verdict accuracy which measures the correctness of predicted verdicts only when the associated evidence score exceeds a predefined threshold

Experiment Findings

- Models generally generate essential questions for claim verification, but struggle with evidence retrieval, especially image-related evidence.

Baselines consist of an LLM and an MLLM, performing different tasks

LLM	MLLM	Q-Eval	Evid-Eval	Veracity (.2/.3/.4)			Justifications (.2/.3/.4)		
Paralleled Question Generation									
Gemini	Gemini	0.42	0.15	0.15	0.13	0.08	0.15	0.11	0.07
Qwen	Qwen-VL	0.43	0.18	0.09	0.08	0.05	0.13	0.11	0.07
Gemma	Gemma	0.39	0.21	0.14	0.12	0.09	0.17	0.14	0.10
Qwen	LLaVA	0.37	0.16	0.09	0.08	0.05	0.12	0.10	0.06
Dynamic Question Generation									
Gemini	Gemini	0.33	0.22	0.17	0.16	0.12	0.17	0.16	0.11
Qwen	Qwen-VL	0.27	0.12	0.10	0.09	0.05	0.09	0.08	0.05
Gemma	Gemma	0.27	0.19	0.15	0.13	0.10	0.15	0.13	0.09
Qwen	LLaVA	0.32	0.16	0.13	0.11	0.08	0.11	0.10	0.08
Hybrid Question Generation									
Gemini	Gemini	0.36	0.19	0.18	0.17	0.10	0.17	0.15	0.09
Qwen	Qwen-VL	0.37	0.16	0.11	0.09	0.06	0.12	0.10	0.06
Gemma	Gemma	0.26	0.25	0.16	0.15	0.11	0.19	0.17	0.12
Qwen	LLaVA	0.30	0.17	0.09	0.09	0.07	0.12	0.11	0.07

Experiment Findings

- Ablation studies proved models with ground-truth evidence perform well in verdict prediction

Evid. Source	LLM	MLLM	Refuted	Supported	NEE	Conflict.	Overall	Justi.
Ground-Truth	Gemini	Gemini	0.84	0.92	0.52	0.00	0.82	0.50
	Qwen	Qwen-VL	0.87	0.47	0.62	0.00	0.78	0.44
	Gemma	Gemma	0.63	0.84	0.62	0.00	0.64	0.49
	Qwen	LLaVA	0.55	0.47	0.90	0.00	0.55	0.43
No Search	Qwen	Qwen-VL	0.01	0.02	0.14	0.00	0.02	0.04

Experiment Findings

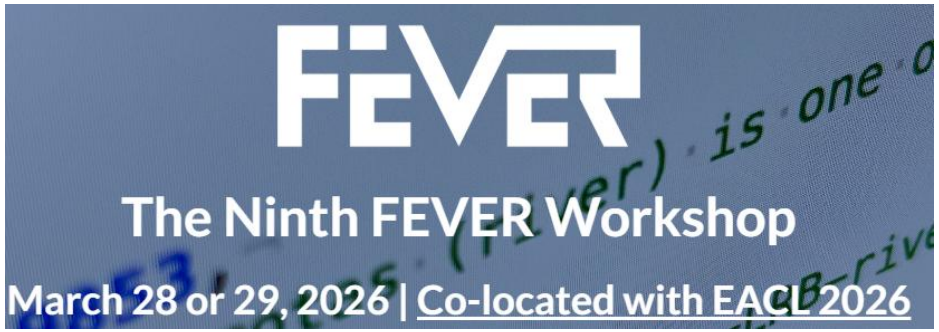
- Web-based information retrieval is important in real-world image-text claim verification

Evid. Source	LLM	MLLM	Refuted	Supported	NEE	Conflict.	Overall	Justi.
Ground-Truth	Gemini	Gemini	0.84	0.92	0.52	0.00	0.82	0.50
	Qwen	Qwen-VL	0.87	0.47	0.62	0.00	0.78	0.44
	Gemma	Gemma	0.63	0.84	0.62	0.00	0.64	0.49
	Qwen	LLaVA	0.55	0.47	0.90	0.00	0.55	0.43
No Search	Qwen	Qwen-VL	0.01	0.02	0.14	0.00	0.02	0.04

Conclusion

Dataset

Code



Conclusion

Dataset

Code