

Advancing Multimodal Fact-Checking Against Climate Misinformation: A Benchmark Dataset and Comparison of Lightweight Models

Omar El Baf, Quentin Senatore, Amira Mouakher Laure Berti-Equille

UMR ESPACE-DEV, Univ Montpellier, IRD, Univ Antilles, Univ Guyane, Univ Réunion, Univ Nouvelle-Calédonie, Univ Perpignan, Montpellier, France

Contact : laure.berthi@ird.fr, amira.mouakher@univ-perp.fr

Context

- Proliferation of misinformation and disinformation in particular related to Climate Change
- People are more likely to trust textual information when it is combined with images, so multimodal misinformation is potentially more harmful.
- AI systems struggle to detect climate misinformation due to the lack of diverse, high-quality multimodal benchmark datasets.

Limitations of Current Datasets

- ✗ *Limited reasoning depth*: No multi-hop or cross-modal inference
- ✗ *Quality issues*: Imbalanced labels and noisy or weakly verified claims
- ✗ *Lack of scale and diversity*: Most datasets remain narrow in topic or platform coverage
- ✗ *Few temporal datasets*: Limited resources for video-based or evolving misinformation

Existing Datasets for Climate Fact-Checking

Dataset	Modality	Size	Key Features / Notes
CLIMATE-FEVER (2020)	Text	1,535 claims	Extends FEVER; uses Wikipedia evidence; suffers from class imbalance and low-quality claims.
CLIMATEX (2023)	Text	8,094 statements	Extracted from IPCC reports; annotated with graded certainty; evaluates model confidence in scientific assertions.
MM-CLAIMS (2022)	Text + Image	86k tweets (3.4k labeled)	Combines textual and visual features; demonstrates improved claim detection.
Bai et al. (2024)	Text + Image	49,316 tweets	Climate tweets with paired images; explores reasoning over contradictory cross-modal signals.
MULTICLIMATE (2024)	Video + Text	4,209 (from 100 videos)	Video-based stance detection; includes transcripts and frames; introduces temporal reasoning challenges.

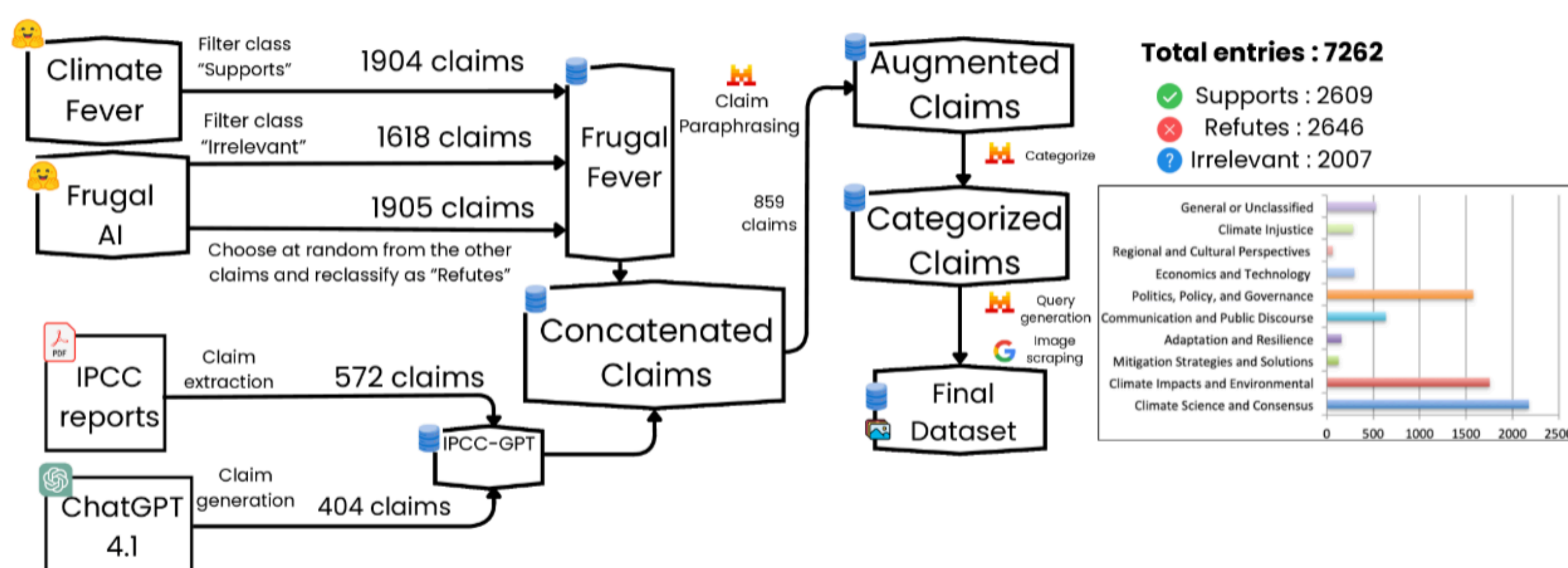


Scan me!

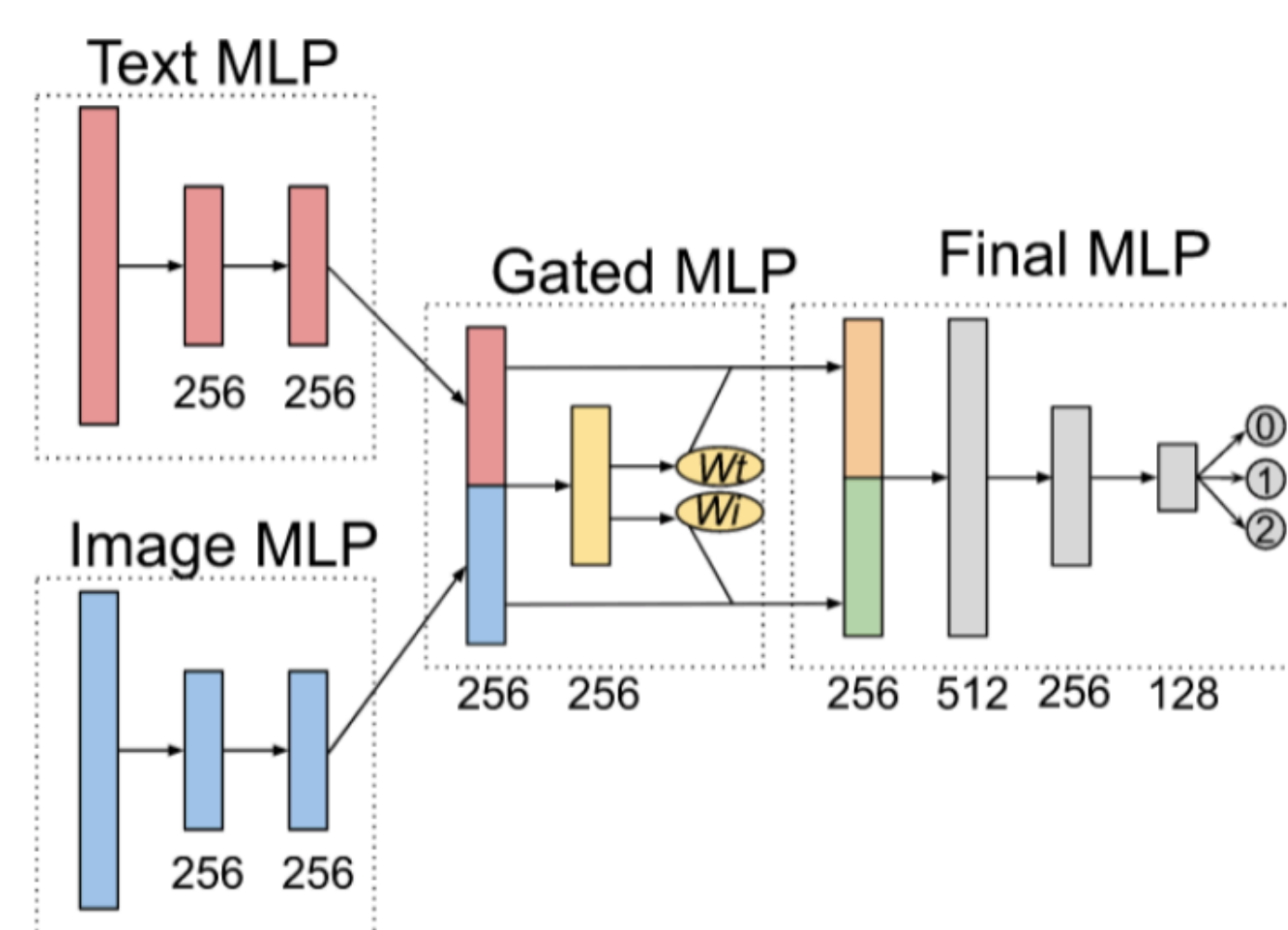
Two Contributions to Climate Fact-Checking

https://github.com/LaureBerti/TIGER_M4FC

TIGER: A balanced, extensible multimodal dataset for climate fact-checking



M4FC: Lightweight MLP-based models with efficient performance



Experimental Results

Results on CLIMATE FEVER (extended with images)

- Accuracy remains below 70%, with low F1-scores \Rightarrow dataset is challenging and noisy.
- Tree-based and MLP models train extremely fast ($< 20s$), but likely underfit
- M4FC requires longer training (≈ 1 min) but shows higher stability
- Indicates weaker signal-to-label alignment in CLIMATE FEVER

Encoding			Model	Accuracy	F1-score	Time (s)			CO ₂ (g)
Encoder	Visual	Textual				Encoding	Training	Inference	
Jina-Clip v2	EVA02-L14	Jina XLM-RoBERTa	MLP	63.5 \pm 1.68	51.08 \pm 2.16	17 160	7.42	0.03	0.51
			Random Forest	69.38 \pm 2.78	56.07 \pm 4.88	17 160	17.89	0.02	0.93
			Gradient Boosting	70.17 \pm 2.77	54.91 \pm 4.64	17 160	12.93	0.004	1.21
CLIP	ViT-B/16	CLIP Transformer	M4FC	64.87 \pm 0.68	35.92 \pm 6.93	10 080	64.39	0.18	7.09
CLIP	ResNet-50	CLIP Transformer	M4FC	65.26 \pm 0.68	41.28 \pm 7.61	9 420	39.98	0.17	19.87
CLIP	ResNet-50x4	CLIP Transformer	M4FC	65.15 \pm 0.79	42.69 \pm 6.82	19 500	63.73	0.18	13.41

Results on TIGER

- All models achieve strong, stable performance ($> 83\%$ accuracy, balanced F1-scores).
- Higher training times for CLIP encoders reflect richer cross-modal alignments.
- M4FC matches or outperforms tree-based models with smoother generalization.
- ResNet-50x4 achieves same accuracy as ResNet-50 but with lower carbon footprint.

Encoding			Model	Accuracy	F1-score	Time (s)			CO ₂ (g)
Encoder	Visual	Textual				Encoding	Training	Inference	
Jina-Clip v2	EVA02-L14	Jina XLM-RoBERTa	MLP	83.14 \pm 1.10	82.73 \pm 1.16	19 020	9.97	0.04	0.56
			Random Forest	83.33 \pm 3.10	82.67 \pm 3.27	19 020	20.97	0.004	1.03
			Gradient Boosting	83.48 \pm 0.78	82.91 \pm 0.80	19 020	38.06	0.006	1.34
CLIP	ViT-B/16	CLIP Transformer	M4FC	84.41 \pm 0.82	83.96 \pm 0.82	11 100	175.22	1.13	7.85
CLIP	ResNet-50	CLIP Transformer	M4FC	84.84 \pm 0.77	84.39 \pm 0.82	10 440	493.20	2.45	21.99
CLIP	ResNet-50x4	CLIP Transformer	M4FC	84.84 \pm 0.78	84.40 \pm 0.80	22 500	332.53	1.15	14.84