

NewsMet: A 'Do It All' dataset of contemporary Metaphors in News headlines

Rohan Joseph¹, Timothy Liu², Aik Beng Ng², Simon See^{1,2}, Sunny Rai^{1,3}

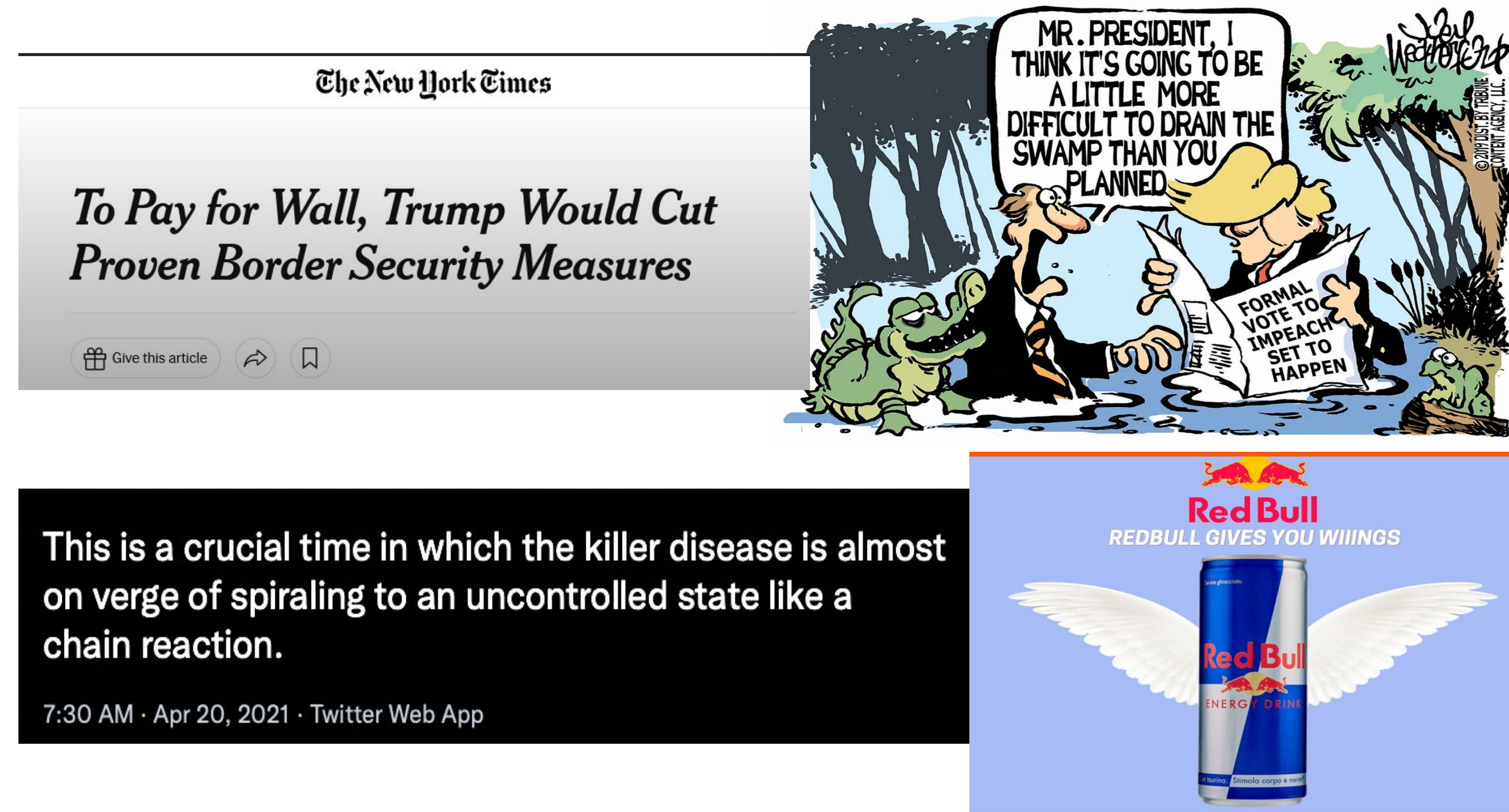
rohan18545@mechyd.ac.in

What is NewsMet?

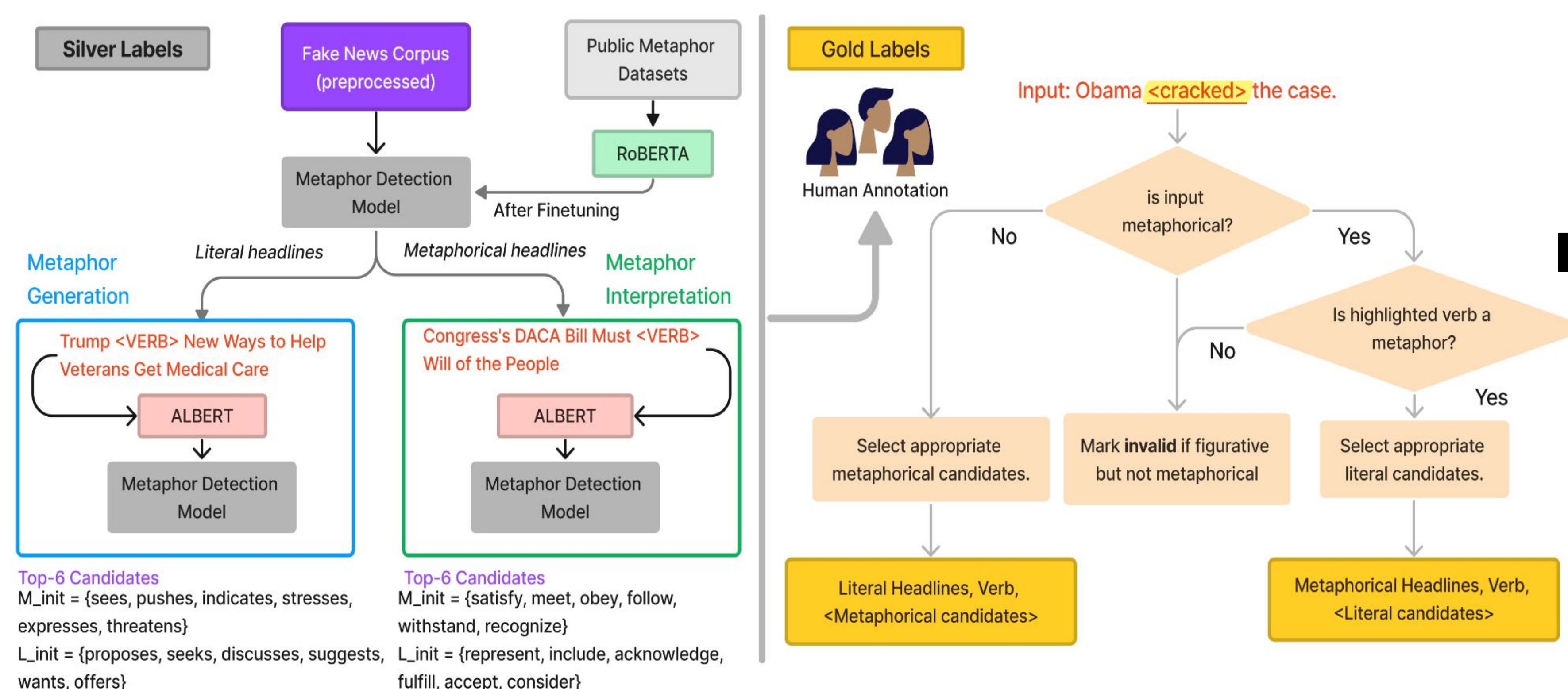
- High quality dataset of 2669 headlines to **detect**, **interpret** and **generate** metaphors.
- Headlines from sources such as NYT, Rawstory and Breitbart.
- Contemporary headlines published between 2017 to 2018.
- Labelled with SVO triplet + Metaphoricity of Verb

Why is Metaphor Processing important?

- Metaphors are everywhere!
- Highly creative constructs.
- Continuously evolving usage.
 - Love **heals** soul → Amtrak Dining Car **heals** nation
- Culture-dependent. Want to test Multilingual LLMs? Use Metaphors!



Proposed Approach



Group	D _{gold}	D _{gold+}	
Invalid	314	-	-
H _m & V _m	389	1009	
H _m & V _l	205	-	
H _m & (V _m V _l)	594	1009	=1603
H _l & V _l	611	455	=1066
	=1205	=1464	= 2669

Examples

Ex 1: Handling <subject> of Police Killing **Spurs** <metaphor> Grand Jury Inquiry <object> Into Prosecutor

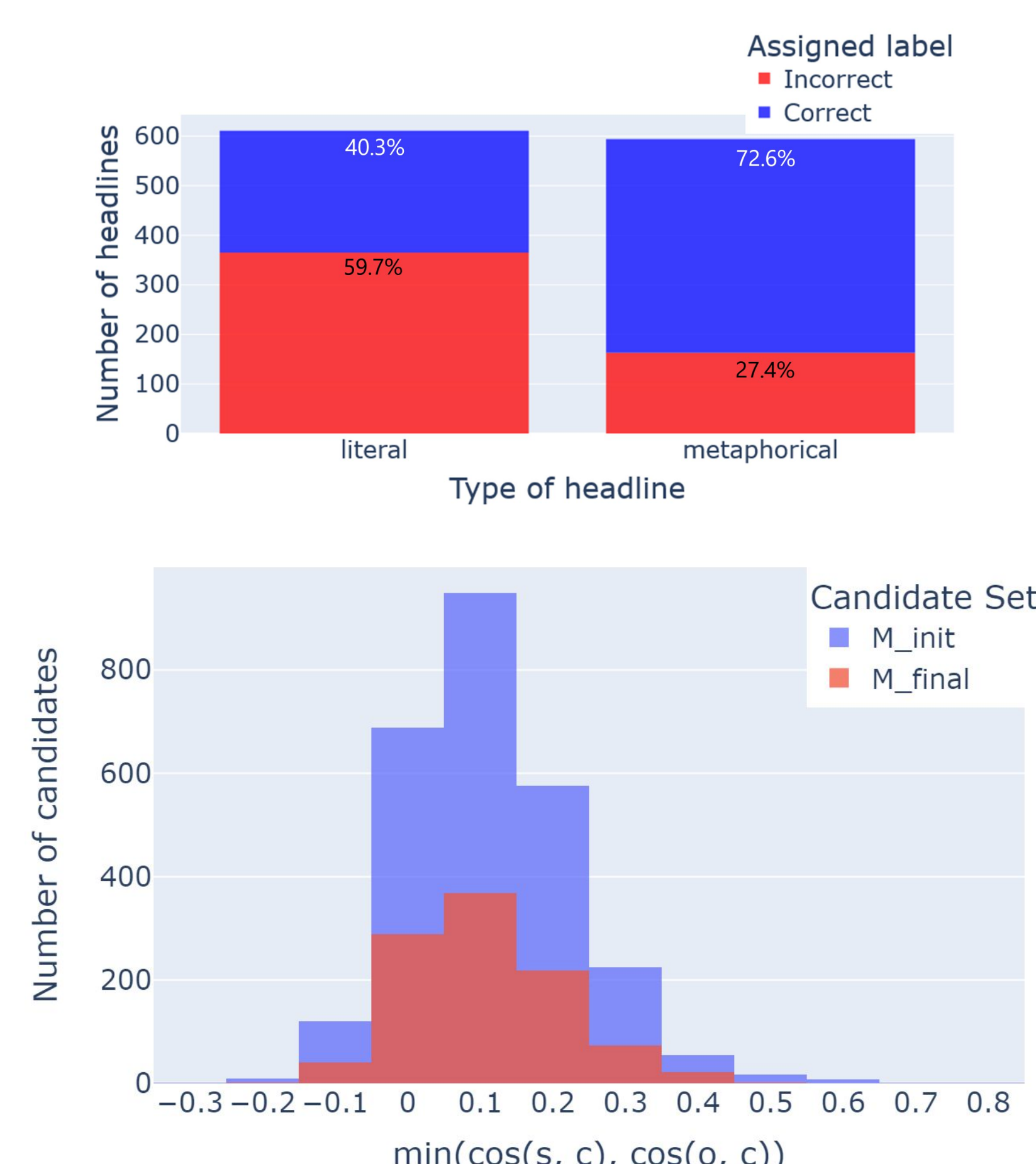
Literal Candidates: [prompting, causing]

Ex 2: The Pneumonic Plague <subject> **Killed** <literal verb> 50 Million People <object> in the Middle Ages and Now It's Back (VIDEO)

Metaphorical Candidates: [strangled, murdered, wiped]

Findings and Results

- Finetuned LLM model has a high (>50%) false positive rate when labelling metaphors!
- LLMs offer candidates with low metaphoricity however, semantic appropriateness is a challenge!
- LLMs are good at generating candidates from diverse SOURCE DOMAINS. 84% of metaphorical headlines with diversity score >=4.



An example for calculating *diversity*

Headline (focus verb)	Clusters _i	diversity
Far Cry 5': Coop Mission has a massive problem with mission progress	{ confronts, faces } ₁ , { creates, causes } ₂ , { finds } ₃ , { tackles } ₄	4

- Even a small Gold Labelled Dataset from new corpora improved quality of predictions!

Model	A	P	R	F1
M _{met_det}	0.54	0.70	0.35	0.46
M _{D_{imbal}+D_{gold+}}	0.61	0.68	0.57	0.62

Takeaways

- Use caution when deploying finetuned LLMs for labelling new corpora.
- Proposed metrics such as diversity and metaphoricity serve as heuristics to measure quality of generated metaphors.
- Use our Annotation pipeline to semi-automate annotating new corpora.
- When to use NewsMet? - Evaluate Metaphor processing systems/ LLMs, Hyprepartisan News, Figurative text translation (Read Paper for more details)

Dataset available at

https://github.com/AxleBlaze3/NewsMet_Metaphor_Dataset

