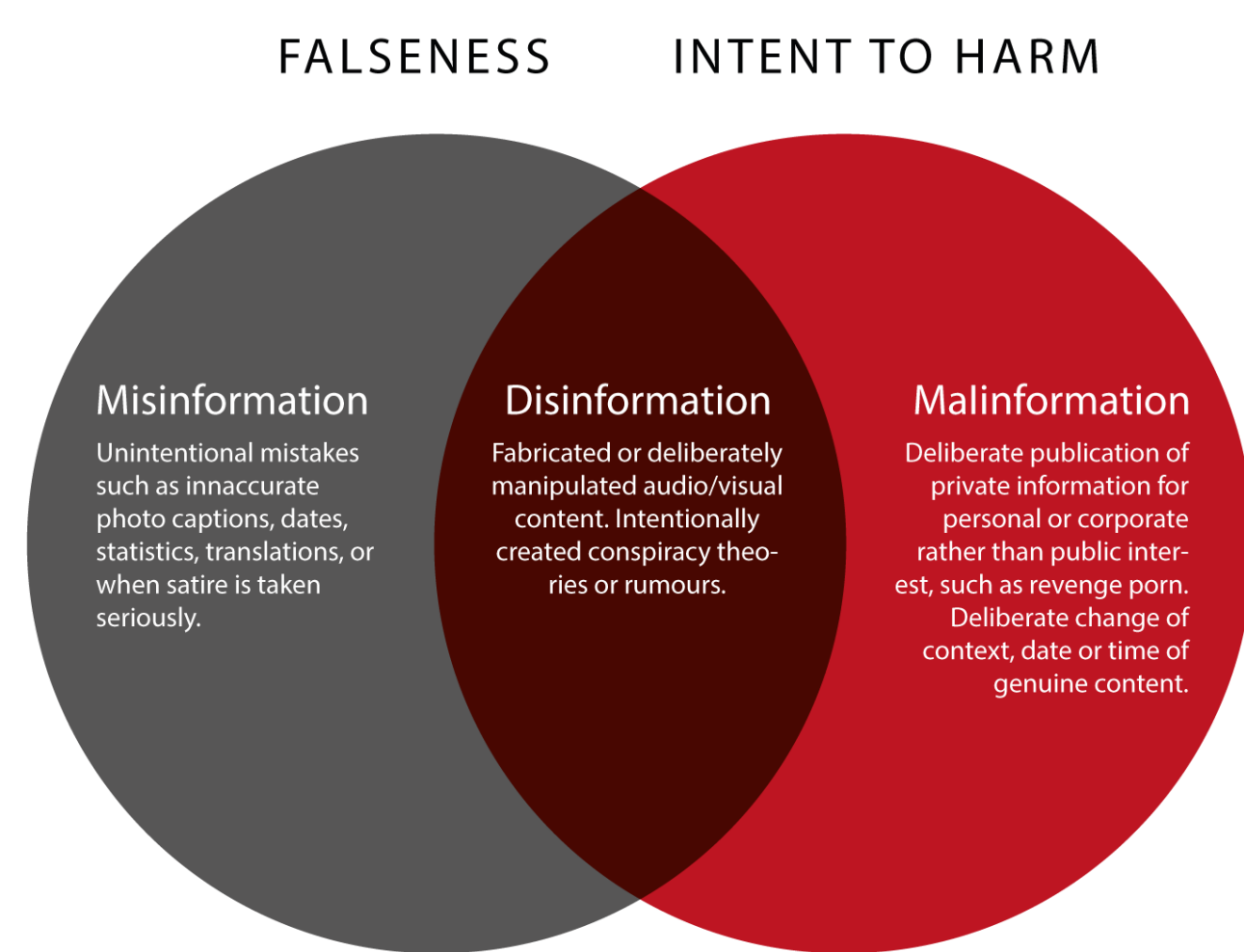




## Stance Detection and Mis/Disinformation

### TYPES OF INFORMATION DISORDER

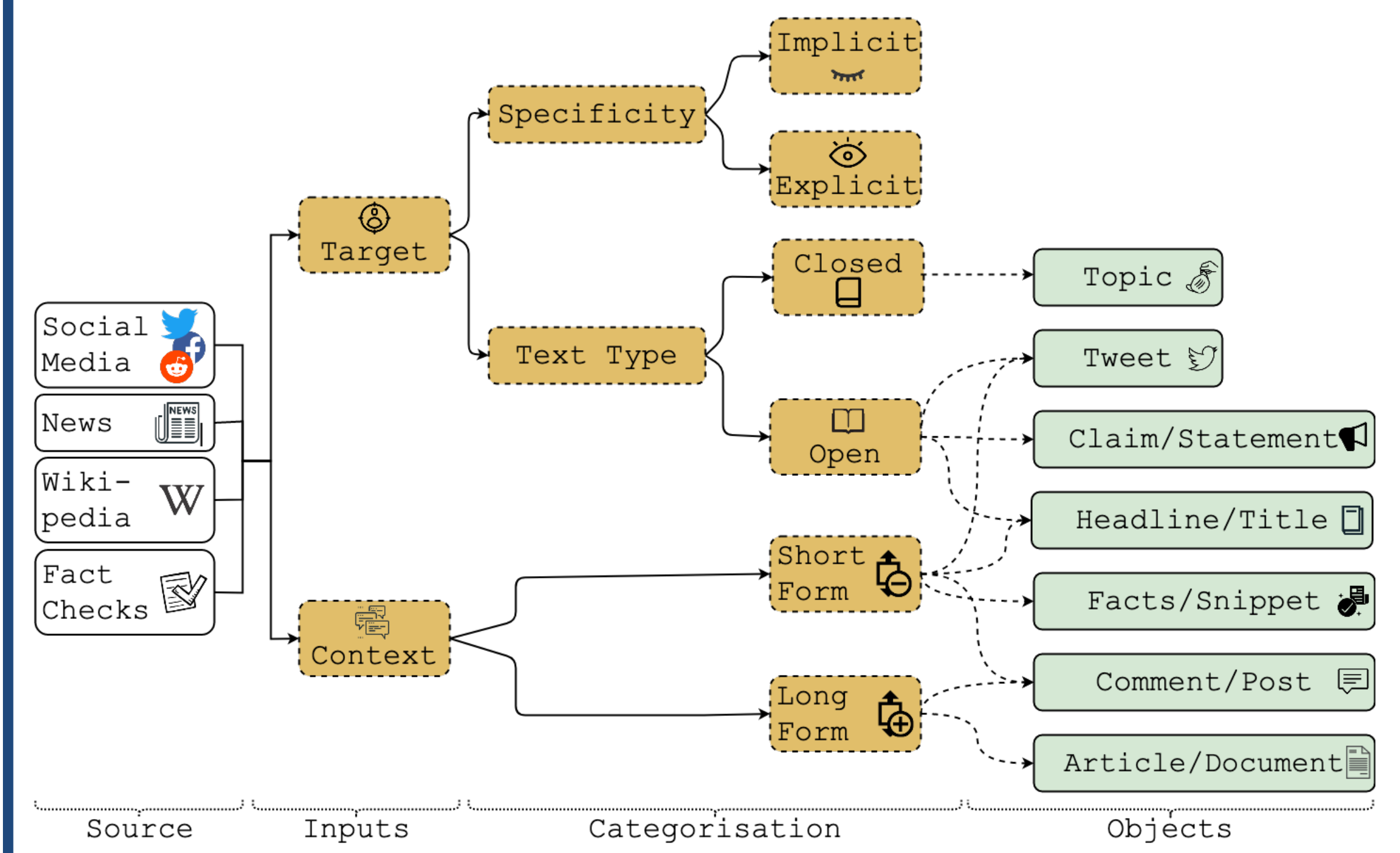


Understanding information disorder  
Wardle (2020)

**Def. (Stance):** "A public act by a social actor, achieved dialogically through overt communicative means, of simultaneously evaluating objects, positioning subjects (self and others), and aligning with other subjects, with respect to any salient dimension of the sociocultural field" Du Bois (2007)

**Def. (Stance Detection):** "For an input in the form of a piece of text and a target pair, stance detection is a classification problem where the stance of the author of the text is sought in the form of a category label from this set: Favor, Against, Neither." Küçük and Can (2020)

## Types of Stance Settings



## Summary

### What is included:

- Holistic overview** of the role that different formulations of **stance detection** play in the **detection of false content**
  - Settings** for **mis- and disinformation identification** to which **stance detection** has **successfully been applied** (datasets and task formulations)
  - Current state-of-the-art **approaches, systems, and applications**
  - Lessons learned** and important **future trends**
- Out of scope:**
- Surveying stance detection holistically, without a specific focus on veracity
  - Other closely related NLP tasks, e.g., sentiment analysis

## Stance Detection Datasets for Mis- and Disinformation Detection

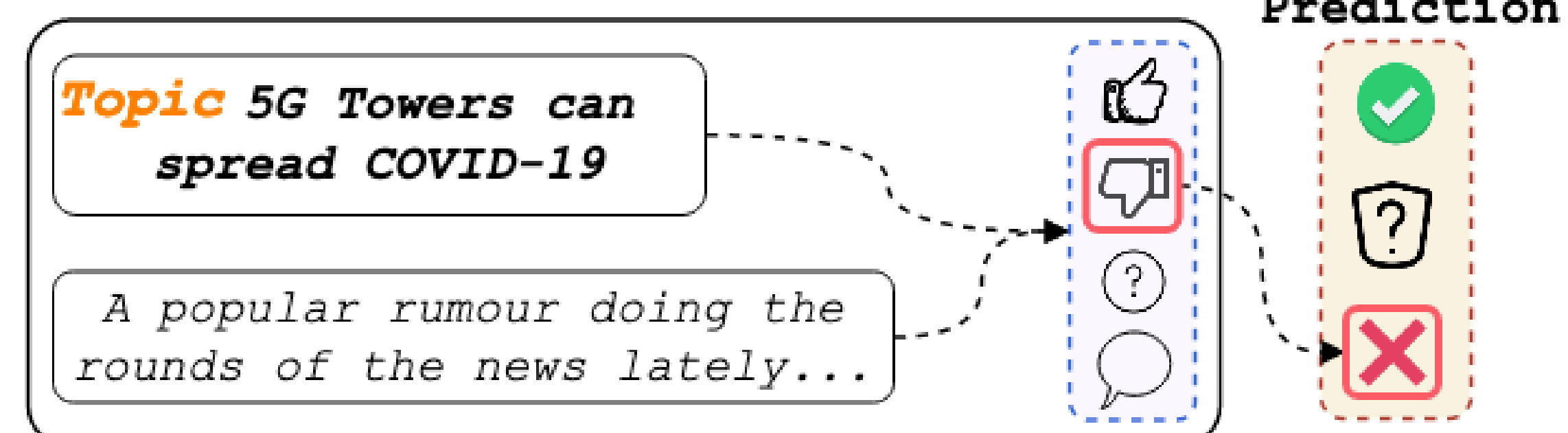
Sources: Twitter, News, Wikipedia, Reddit.

Evidence: Single, Multiple, Thread.

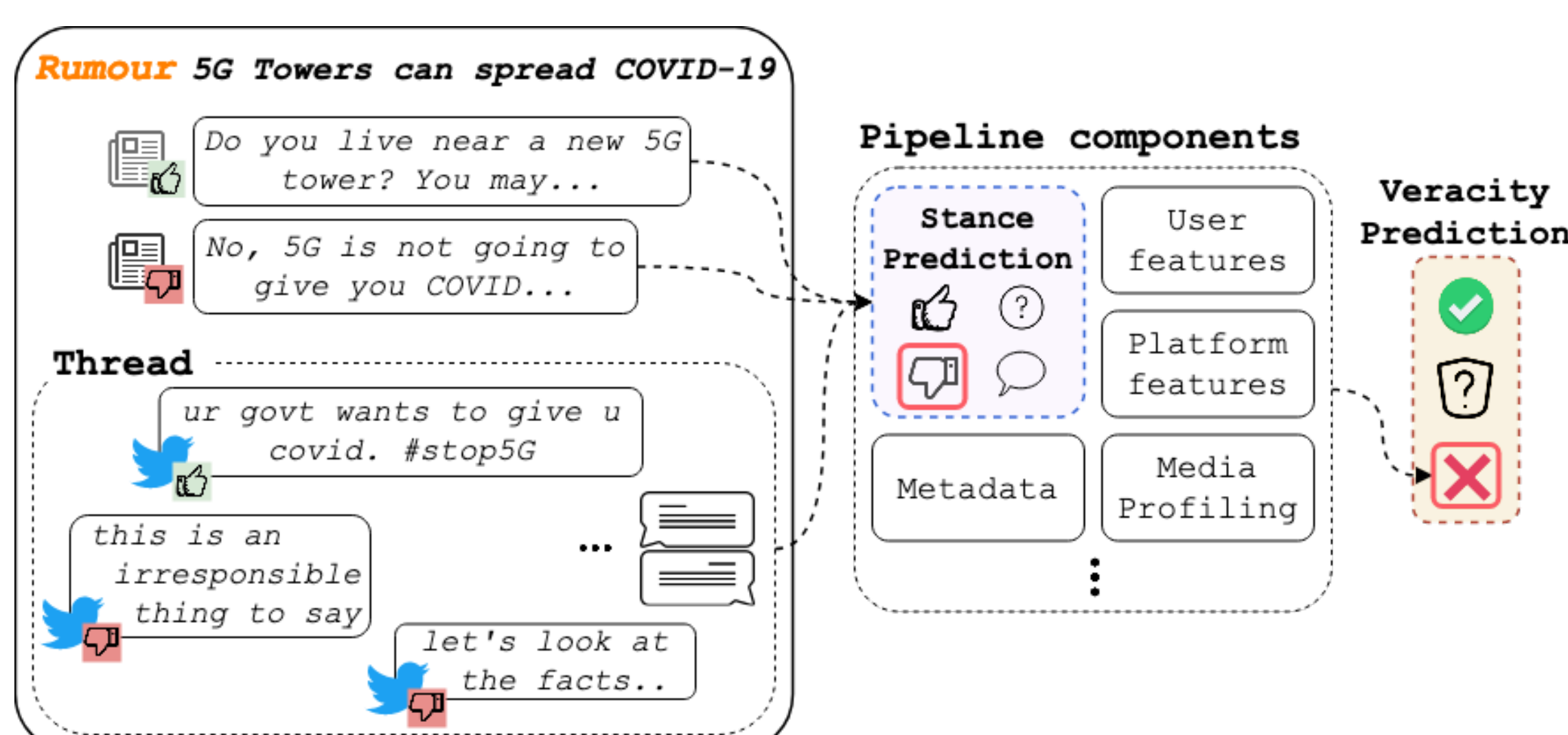
Dataset	Source(s)	Target	Context	Evidence	#Instances	Task
<b>English Datasets</b>						
<i>Rumour Has It Qazvinian</i> (2011)		Topic	Tweet		10K	Rumours
<i>PHEME Zubiaga</i> (2016)		Claim	Tweet		4.5K	Rumours
<i>Emergent Ferreira and Vlachos</i> (2016)		Headline	Article		2.6K	Rumours
<i>FNC-1 Pomerleau and Rao</i> (2017)		Headline	Article		75K	Fake news
<i>RumourEval '17 Derczynski</i> (2017)		Implicit	Tweet		7.1K	Rumours
<i>FEVER Thorne</i> (2018)		Claim	Facts		185K	Fact-checking
<i>Snopes Hanselowski</i> (2019)	Snopes	Claim	Snippets		19.5K	Fact-checking
<i>RumourEval '19 Gorrell</i> (2019)		Implicit	Post		8.5K	Rumours
<i>COVIDLies Hossain</i> (2020)		Claim	Tweet		6.8K	Misconceptions
<i>TabFact Wenhui et al.</i> (2020)		Statement	WikiTable		118K	Fact-checking
<b>Non-English Datasets</b>						
<i>Arabic FC Baly et al.</i> (2018)		Claim	Document		3K	Fact-checking
<i>DAST (Danish) Lillie</i> (2019)		Submission	Comment		3K	Rumour
<i>Croatian Bošnjak and Karan</i> (2019)		Title	Comment		0.9K	Claim verifiability
<i>ANS (Arabic) Khouja</i> (2020)		Claim	Title		3.8K	Claim verification
<i>Ara(bic)Stance Alhindi</i> (2021)		Claim	Title		4K	Claim verification

## Stance Detection Formulations

### Stance Prediction



(a) as Fact-Checking



(b) as a Component of a Fact-Checking Pipeline

## Lessons Learned and Future Trends

### Integration:

We argue for a tighter integration between stance and fact-checking.

### Dataset Size:

- A major limitation when training models
- The vast majority of datasets contain only a few thousand examples

### Multilinguality:

- Only a handful of multilingual datasets
- Small in size and do not offer cross-lingual settings (exception: Vamvas and Sennrich (2020))
- Cultural norms play a crucial role

### Shades of Truth:

- Missing notion of stance detection, but fact-checking goes beyond true/false
- Fine-grained labels are common for the related task of Sentiment Analysis

### Explainability:

- Crucial step towards adopting fully automated fact-checking
- There is a need for holistic and realistic explanations of how a fact-checking model arrived at its prediction

### Data Mixing:

- Can compensate for small dataset sizes
- Not trivial, task definitions and label inventories vary (Schiller, 2021; Hardalov, 2021)

### Multimodal Content:

- Spreading mis- and disinformation through multiple modalities is becoming increasingly popular (e.g., deepfakes, memetic warfare)
- The wisdom of the crowd in social media can be an additional information source

### Modelling the Context:

- The background of the stance-taker, e.g., previous activity, network, interests
- Characteristics of the target, e.g., funding, previously known biases, credibility

### Label Semantics:

- With time, the definition of stance and the label inventories have evolved
- The labels can share semantic similarities, but there can be mismatches in the label definitions (Hardalov, 2021; Momchil et al., 2022)