EMONA: Event-level Moral Opinions in News Articles

Yuanyuan Lei, Md Messal Monem Miah, Ayesha Qamar, Sai Ramana Reddy, Jonathan Tong, Haotian Xu, Ruihong Huang

Computer Science and Engineering, Texas A&M University, College Station, TX

Introduction

- Most previous research on moral frames focused on social media short texts and annotated for an entire social media posts
- The journalists express their stances through moral judgment towards events
- propose a new task to understand **event-level** moral opinions in news articles

Moral Foundation Theory

- The social moral principles are categorized into five dimensions, each with a positive and a negative judgment: Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion, Purity/Degradation

Motivations

- **RQ1: can event-level moral opinions reflect article-level ideology?**

Liberal: Texas Gov. Rick Perry is **broadcasting** [non-moral] a new campaign ad that **contrast** [cheating] Christianity with homosexuality.

Conservative: Texas Gov. **claimed** [non-moral] in an ad that President Obama is waging a **war** [degradation] on religion, and he can **defend** [purity] faith.

- **RQ2: can event-level moral opinions explain sentence-level media bias?**

- **RQ3: can event-level moral opinions uncover implicit opinions?**

Example: The new president Trump **told** [non-moral] American: "The time for empty **talk** [cheating] is over. Now arrives the hour of **action** [authority]."

Data Sources

We intentionally select three different components to form our dataset. Overall, the dataset comprises 400 articles, 10k sentences, 283k words.

- **AllSides:** provides article-level ideology label (left, center, right)

12 topics x 5 articles x 3 stances = 180 articles

- **BASIL:** provides sentence-level media bias label

5 triplets x 3 media x 10 years = 150 articles

- **MPQA 3.0:** annotates entity/event-level general opinions, retain 70 articles

Data Annotation

	Pairwise			Majority			llama-2	gpt-3.5	gpt-4
	Min	Max	Mean	Min	Max	Mean	Mean	Mean	Mean
Event Extraction	0.7670	0.8182	**0.7938**	0.8656	0.8877	0.8772	0.4464	0.5870	0.7113
Moral Identification	0.4797	0.5846	**0.5276**	0.6483	0.7814	0.7017	0.3374	0.4064	0.4707
Ten Moral Classification	0.6682	0.7876	**0.7230**	0.8279	0.9245	0.8679	0.3685	0.4068	0.5546

Table 1: Pairwise Cohen's kappa inter-agreement scores among five annotators, and their agreement with majority voting label. The right three columns show the agreement between large language models and human annotators

Dataset Analysis

	# Doc	# Sent	# Event	# (%) Moral
ALL	400	10912	45199	9613 (21.27)
AllSides	180	5448	22527	5628 (24.98)
BASIL	150	3811	16200	2586 (15.96)
MPQA 3.0	70	1653	6472	1399 (21.62)

Table 2: EMONA dataset contains 45199 events, 9613 (21.27%) moral events
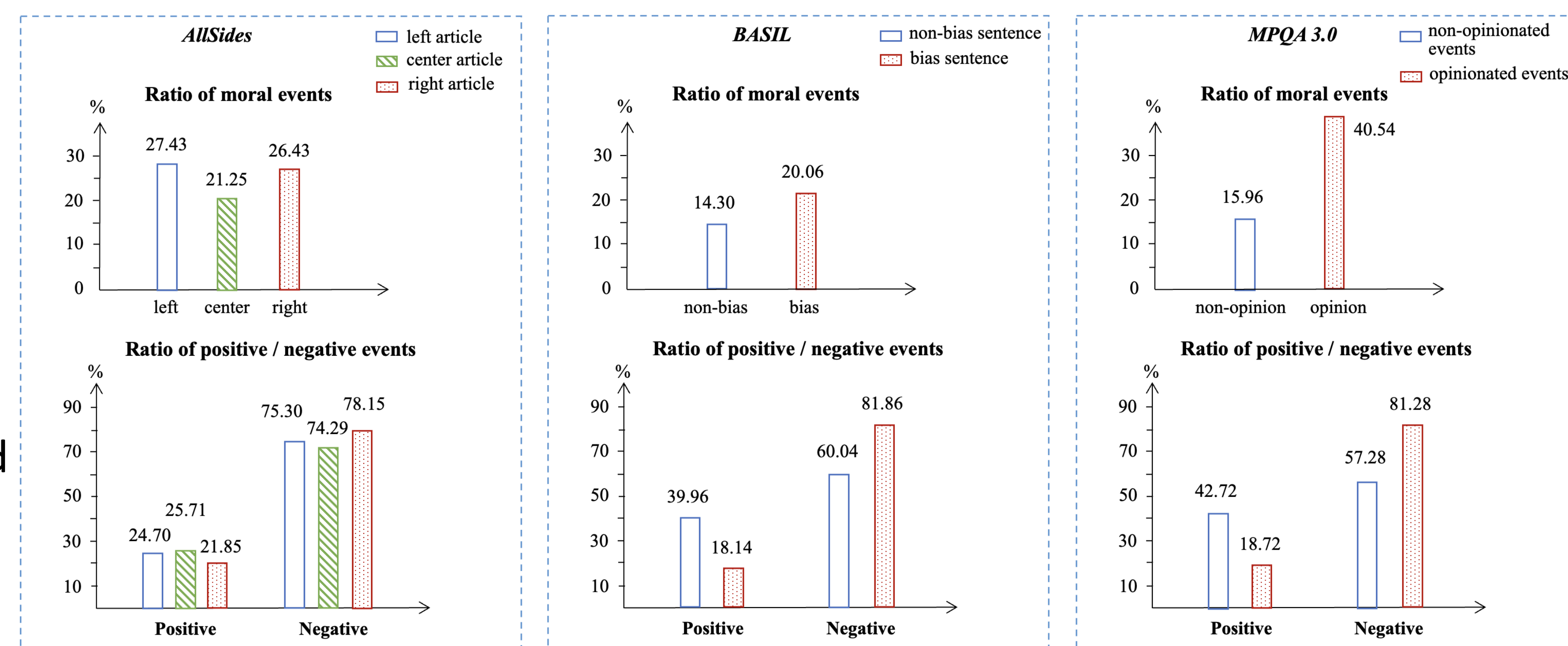

Figure 1: Analysis of event-level moral opinions with article-level ideology, sentence-level media bias, and event-level general opinions

Intrinsic Evaluation

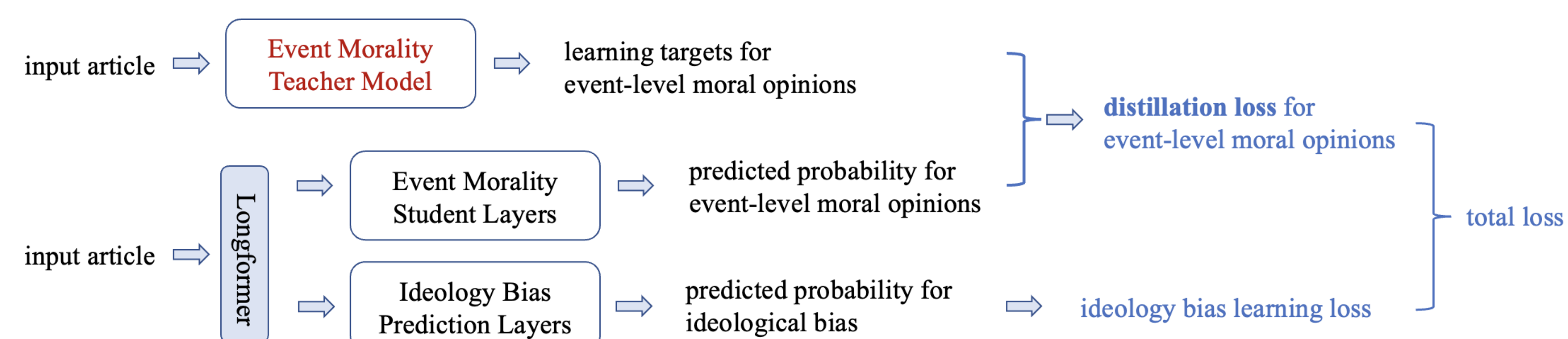We build baseline models for the following tasks:

- Event moral identification: identify whether an event bears moral judgments
- Event moral classification: classify 10 moral labels + non-moral for events
- End-to-end system: predict label for every word in a plain article

	event moral identification			event moral classification			end-to-end system		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
lexicon	36.49	27.80	31.56	18.21	19.39	16.75	-	-	-
gpt-3.5-turbo	41.45	58.90	48.66	24.20	27.46	22.08	22.98	23.14	20.61
gpt-4	59.25	60.91	60.06	35.08	32.68	30.83	30.64	32.73	30.06
longformer	61.81	65.48	63.59	46.32	36.56	39.50	44.30	35.47	38.42

Table 3: Performance of intrinsic evaluations on the EMONA dataset

Extrinsic Evaluation

We propose to incorporate event-level moral opinions to three downstream tasks: (1) Article-level ideology classification (2) Sentence-level media bias identification (3) Event-level general opinion identification

	article-level ideology			sentence-level media bias			event-level opinions		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
longformer baseline	83.69	84.65	84.11	46.81	45.65	46.22	56.73	57.71	57.21
+ event-level moral opinions	87.15	88.03	87.46	50.79	49.60	50.19	60.93	62.94	61.92

Table 4: Performance of extrinsic evaluations on AllSides, BASIL, MPQA 3.0