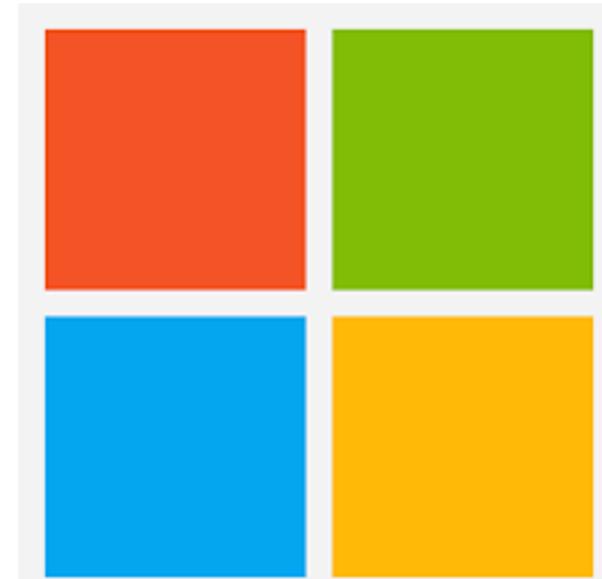




# Evaluating Inter-Bilingual Semantic Parsing for Indian Languages

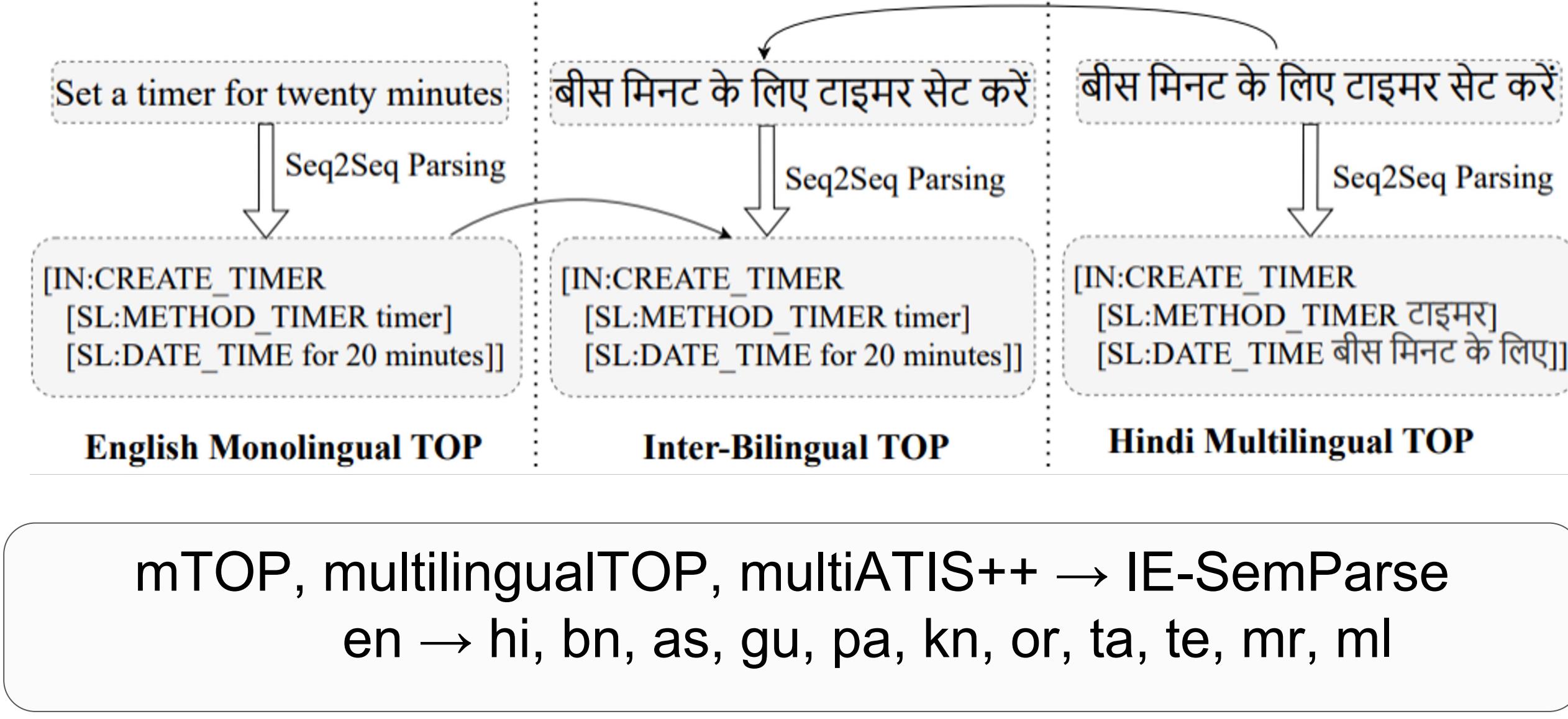
Divyanshu Aggarwal<sup>1\*</sup>, Vivek Gupta<sup>2</sup>, Anoop Kunchukuttan<sup>3,4</sup>

<sup>1</sup>American Express, AI Labs; <sup>2</sup>University of Utah; <sup>3</sup>AI4Bharat; <sup>4</sup>Microsoft India



## Introduction

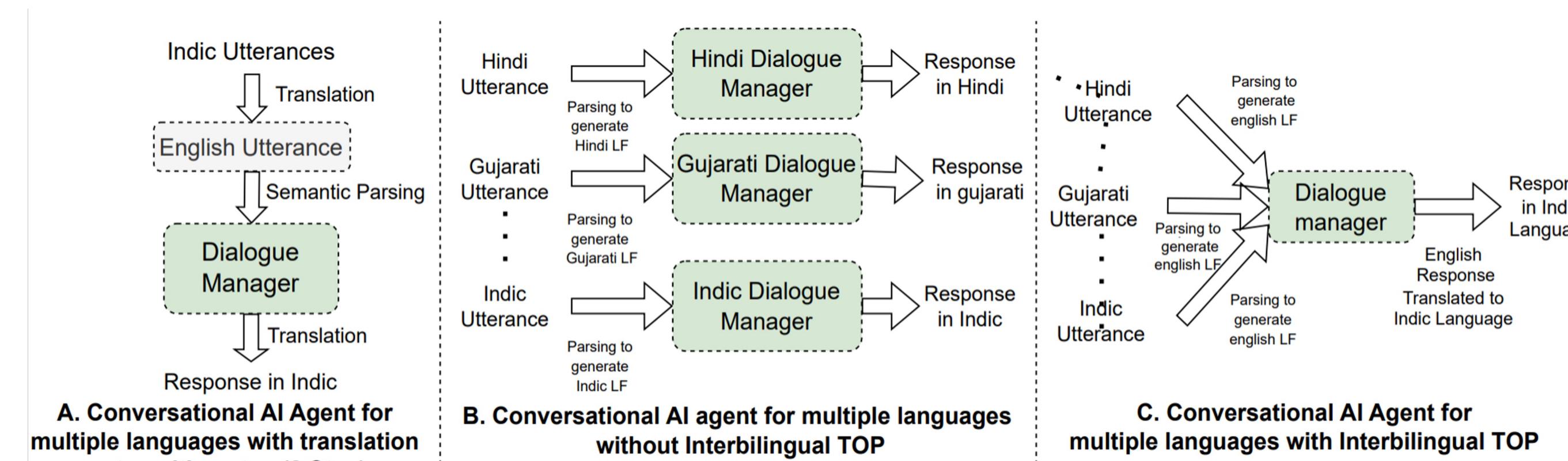
- We propose IE-SemParse a bilingual semantic parsing suite for consisting of 3 datasets.
- The dataset is created by translating utterances of mTOP, multilingualTOP and multiATIS++. We preserve the original logical forms with English slot values.



## Our Contributions

- A novel task called Inter-Bilingual TOP, involving multilingual utterances as input and English logical forms as output.
- A dataset called IE-SEMPARSE, which encompasses 11 Indo-Dravidian languages, representing approximately 22% of the world's population.
- Exploration of various seq2seq models and different train-test strategies.

## Benefits of Inter-bilingual Semantic Parsing



- Approach A: Translate to English then parse to logical form.
- Approach B: Separate parser and dialogue manager for each language
- Approach C: Inter-bilingual Semantic Parsing.

We show that Inter-bilingual Semantic Parsing is a good middle ground approach to enhance model's multilingual semantic parsing ability and reduce system latency and redundancy.

## Data Creation

- IE-mTOP:** This dataset is a translated version of the multi-domain TOP-v2 dataset. English utterances were translated to Indic languages using IndicTrans (Ramesh et al., 2021), while preserving the logical forms.
- IE-multilingualTOP:** This dataset is from the multilingual TOP dataset, where utterances were translated, and logical forms were decoupled using the pytext library.
- IE-multiATIS++:** This dataset comes from the multi-ATIS++, where utterances were translated, and the logical forms were generated from labelled dictionaries and decoupled.

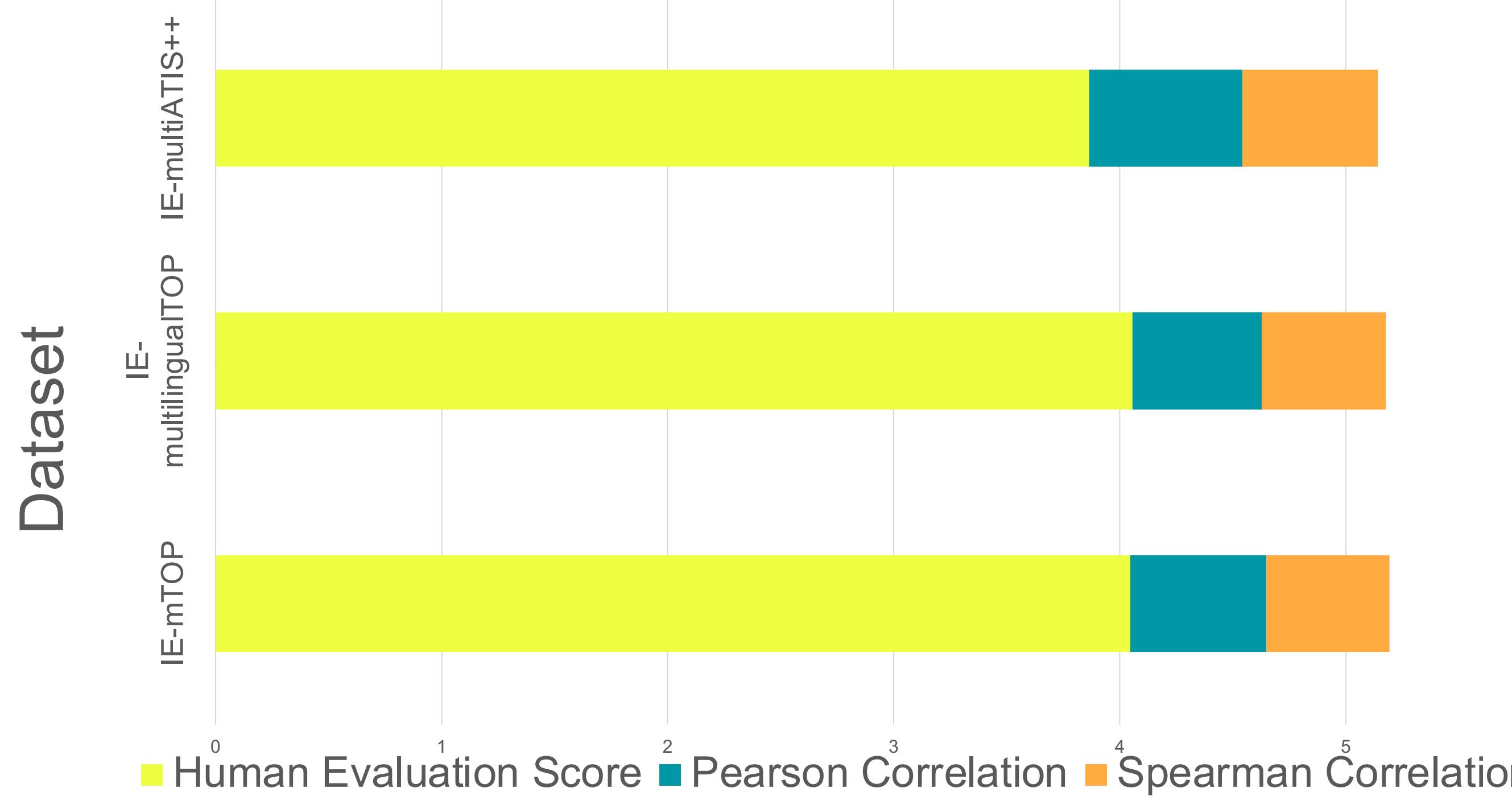
**Note:** Pre and Post-processing steps are elaborated in paper.

## Dataset Evaluation

### Automatic Evaluation Benchmarks

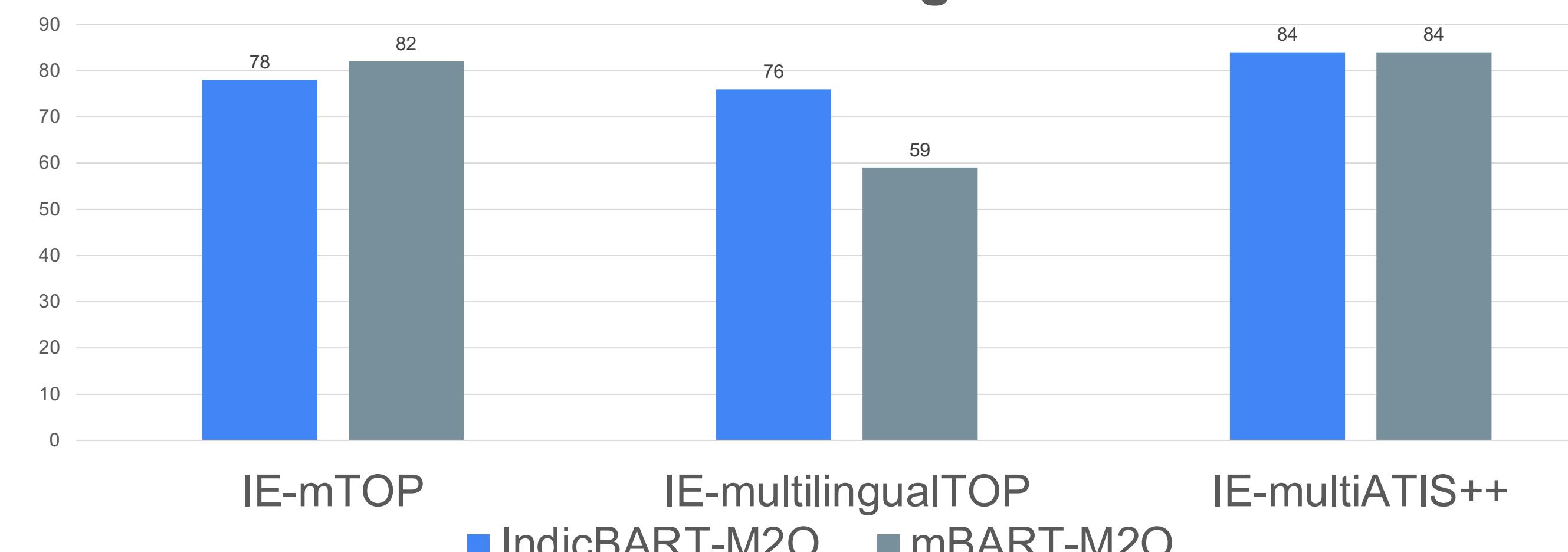
Score	Samanantar	IE-mTOP	IE-multilingualTOP	IE-multiATIS++
BertScore	0.85	0.86	0.98	0.86
CometScore	0.12	0.13	0.14	0.13
BT_BertScore	0.96	0.93	0.92	0.92

### Human Evaluation Benchmarks



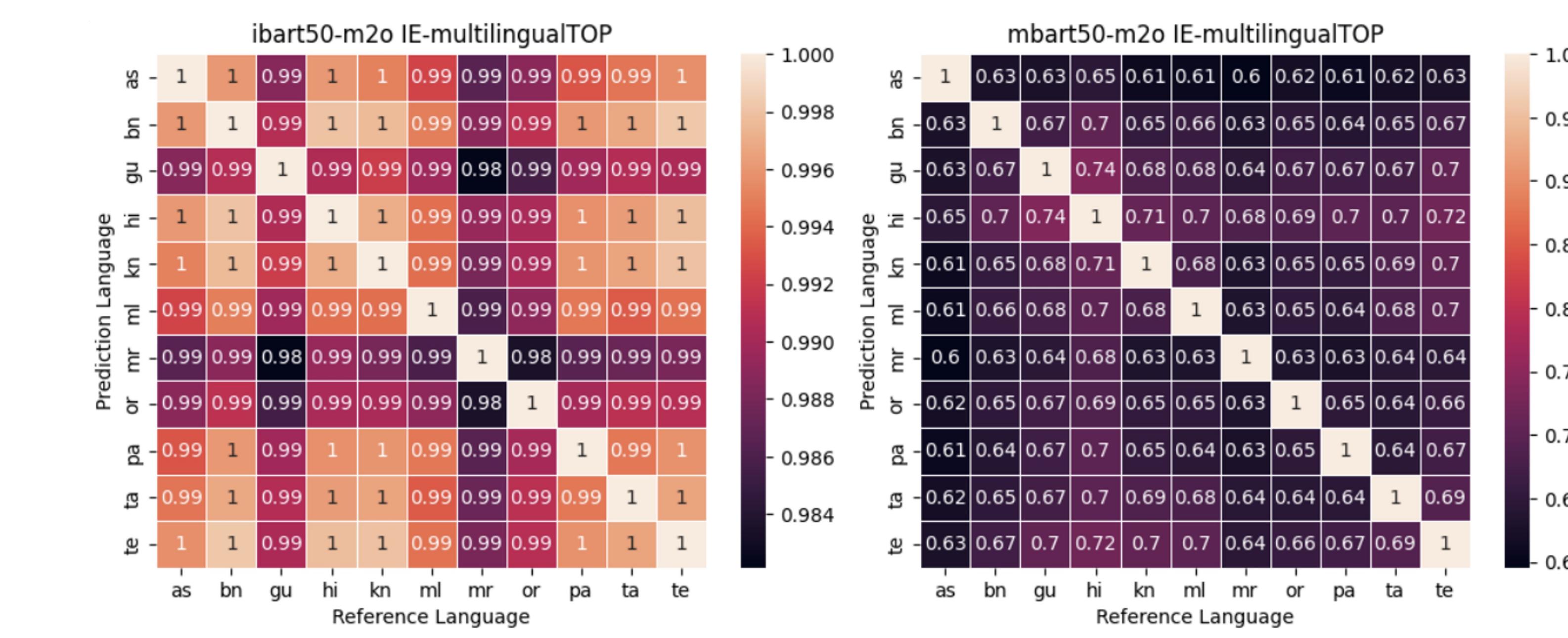
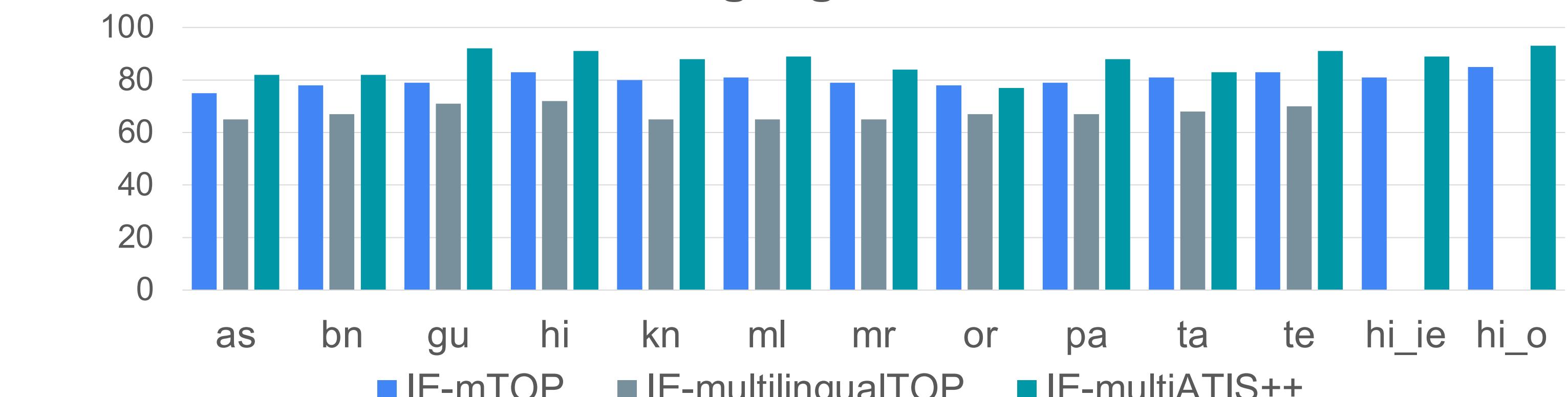
## Results and Analysis

### Model Wise Average Scores

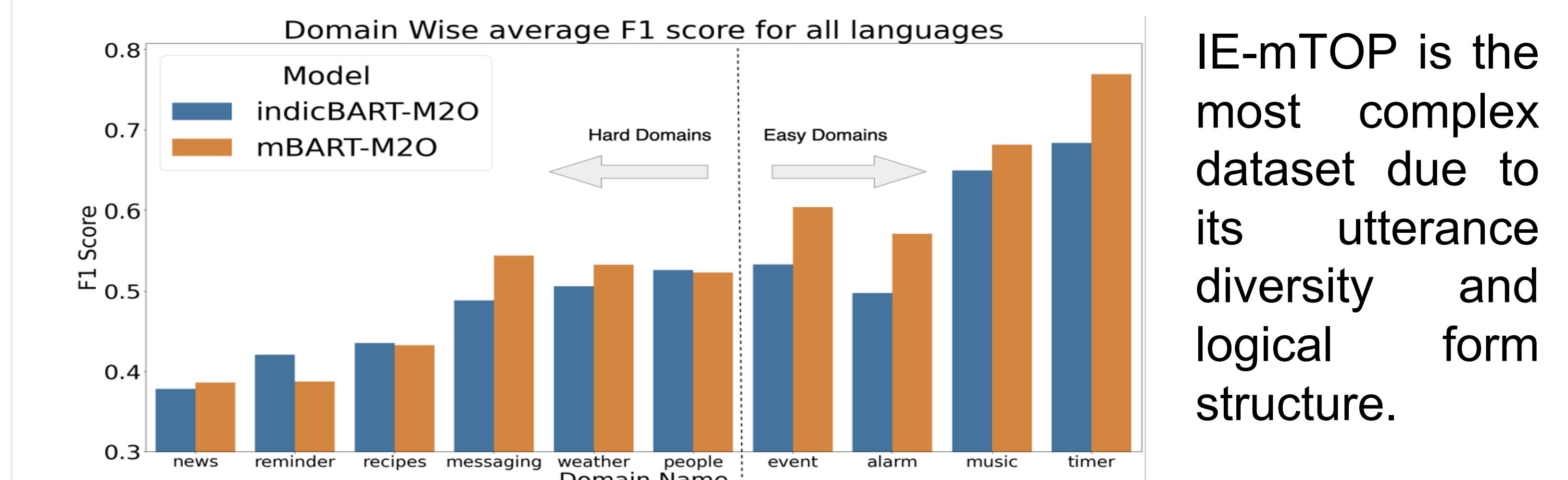


## Results and Analysis Continued

### Language Wise Scores



mBART perform very highly on high resource languages skewing the average score. IndicBART is consistent across languages.



IE-mTOP is the most complex dataset due to its utterance diversity and logical form structure.

Code: <https://iesemparse.github.io/>