

Can Humans Identify Domains?

Maria Barrett · Max Müller-Eberstein · Elisa Bassignana
Amalie Brogaard Pauli · Mike Zhang · Rob van der Goot



TGeGUM

- 9.1k sentences from GUM (Zeldes, 2017);
- Single sentence/prose level annotations (x3);
- 11 genre classes (source type);
- 10/100 topic classes as per Dewey (1979).

Introduction & Motivation

- **Textual domain** is a crucial property within the Natural Language Processing (NLP) community due to its effects on downstream model performance.
- The concept itself is, in practice, referring to any non-typological property, such as genre, topic, medium or style of a document.
- We investigate the core notion of domains via human proficiency in identifying related intrinsic textual properties, specifically the concepts of:
 - **Genre (communicative purpose);**
 - **Topic (subject matter).**

① "And this is what Luther writes to Erasmus."
Gold: conversation

② "Arrange rack in the middle of the oven."
Gold: wikihow

③ "What sort of a nose did Cleopatra have?"
Gold: interview

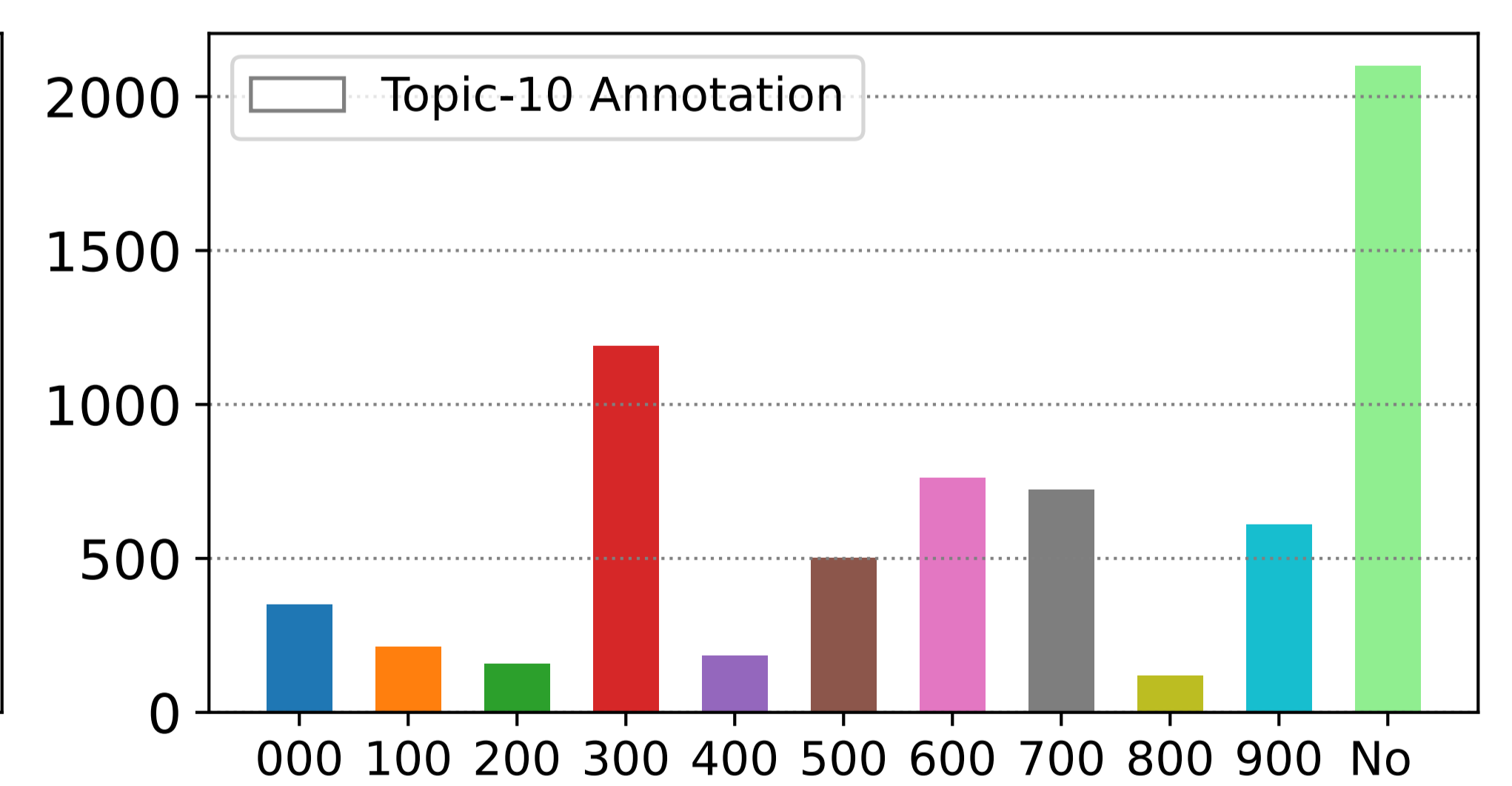
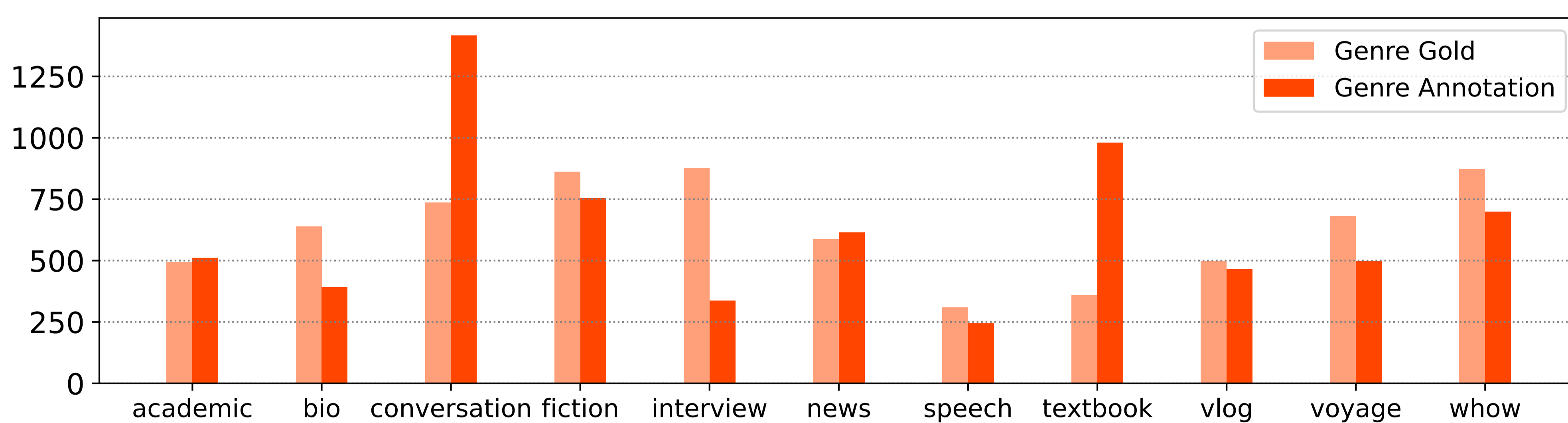
1. textbook
2. wikihow
3. textbook

1. textbook
2. wikihow
3. interview

1. vlog
2. conversation
3. conversation

Data

	Instances	
	Sentence	Prose
Train	6,911	1,358
Dev.	1,117	217
Test	1,096	221
Total	9,124	1,796



"This is... HEY!"

1. Majority: vlog (0.67)

2. PerLabel-Regr.: vlog (0.33), conversation (0.00), ... (0.67)

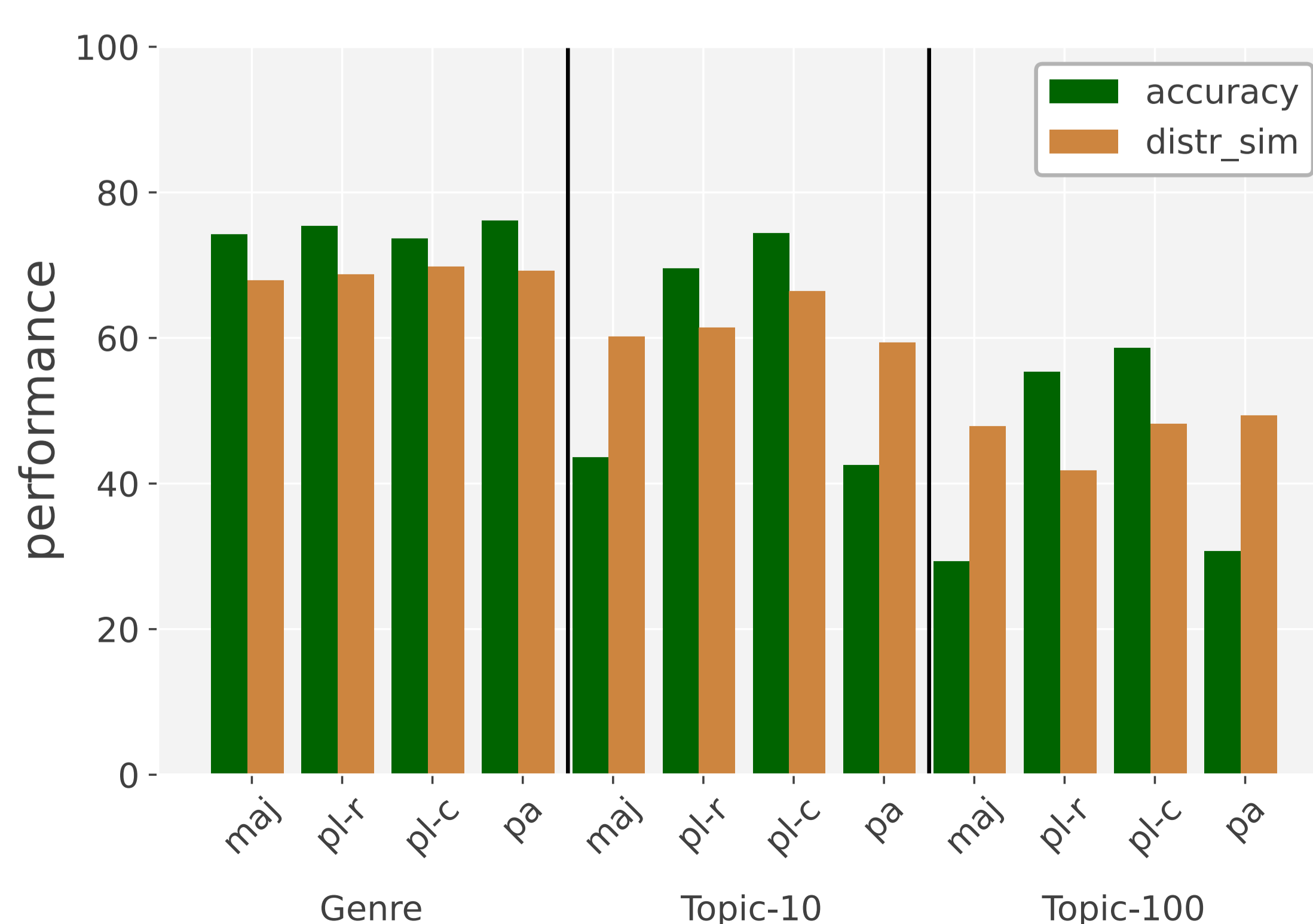
3. PerLabel-Class.: "0.33", "0.00", "0.67", ...

4. PerAnnotator: vlog, conversation

Head

Model

This is ... HEY !



Takeaways

- With a Fleiss' kappa of at most **0.53 on the sentence level** and **0.66 at the prose level**, it is evident that despite the ubiquity of domains in NLP, there is little human consensus on how to define them.
- By training classifiers to perform the same task, we find that **this uncertainty also extends to NLP models.**