Paper Title –Center, Bold, Arial Narrow, Size 16

Rovena Llapushi1\*, Hersi Kopani2

1 Author Affiliation (10pt regular, center align), mail, ORCID number

2 [hersi.kopani@gmail.com](mailto:hersi.kopani@gmail.com) , 0009-0007-8018-1452

**Abstract:** Hate speech on social media appears to be inevitable in our era of ever-more connectivity. This presents a distinct challenge because it is hard to design automated techniques to identify such speech, especially for low-resource languages.

We present PAR database, derived from YouTube comments associated with politics, actuality, and reality show content. Our study compares the performance of two models: Google's Flan-T5 and GPT 3.5. We evaluate their ability to detect hate speech across three topics and analyze their performance on each individual topic. We also evaluate the effect of translating a comment from Albanian Jargon to standard Albanian on the model’s performance. This analysis is important because both models utilize the translated English version as the target language for distinguishing such comments.

Lastly, we debate whether large language models like GPT 3.5 are required for this kind of task. Is using a model this scale for transfer learning really that beneficial, even if the model has no knowledge of the Albanian language? Or should we focus more of our attention on engineering and text annotation?