





*Research Article*

# Do AIs know what the most important issue is? Using language models to code open-text social survey responses at scale

Jonathan Mellon <sup>1</sup>, Jack Bailey <sup>2</sup>, Ralph Scott <sup>3</sup>, James Breckwoldt<sup>2</sup>, Marta Miori<sup>2</sup>, and Phillip Schmedeman <sup>1</sup>

## Abstract

Can artificial intelligence accurately label open-text survey responses? We compare the accuracy of six large language models (LLMs) using a few-shot approach, three supervised learning algorithms (SVM, DistilRoBERTa, and a neural network trained on BERT embeddings), and a second human coder on the task of categorizing “most important issue” responses from the British Election Study Internet Panel into 50 categories. For the scenario where a researcher lacks existing training data, the accuracy of the highest-performing LLM (Claude-1.3: 93.9%) neared human performance (94.7%) and exceeded the highest-performing supervised approach trained on 1000 randomly sampled cases (neural network: 93.5%). In a scenario where previous data has been labeled but a researcher wants to label novel text, the best LLM’s (Claude-1.3: 80.9%) few-shot performance is only slightly behind the human (88.6%) and exceeds the best supervised model trained on 576,000 cases (DistilRoBERTa: 77.8%). PaLM-2, Llama-2, and the SVM all performed substantially worse than the best LLMs and supervised models across all metrics and scenarios. Our results suggest that LLMs may allow for greater use of open-ended survey questions in the future.