Rui Huang and Wenhao Luo
AAAI-23 Student Abstract Chairs

SUBMISSION: 4743
TITLE: Can You Answer This? - Exploring Zero-Shot QA Generalization Capabilities in Large Language Models

---------------------- REVIEW 1 --------------------
SUBMISSION: 4743
TITLE: Can You Answer This? - Exploring Zero-Shot QA Generalization Capabilities in Large Language Models
AUTHORS: Saptarshi Sengupta, Shreya Ghosh, Preslav Nakov and Prasenjit Mitra

----------- Summary -----------
SCORE: 2 (accept)
----- TEXT:
This paper explored the zero-shot capabilities of Transformer-based Language Models for question answering and the performance change towards domain drift. They focused on two aspects: length and polysemous words in different senses. They examined some popular models and raised conclusions.

Based on experimental observations, the authors proposed some possible reasons for the poor performance of TLMs on zero-shot QA generalization, which has a guiding significance for future work to design transformer-based models.

---------------------- REVIEW 2 --------------------
SUBMISSION: 4743
TITLE: Can You Answer This? - Exploring Zero-Shot QA Generalization Capabilities in Large Language Models
AUTHORS: Saptarshi Sengupta, Shreya Ghosh, Preslav Nakov and Prasenjit Mitra

----------- Summary -----------
SCORE: -1 (weak reject)
----- TEXT:
The abstract addresses the very relevant problem of how transformer based language models can be applied to domain-specific datasets. Despite their general success in general domain problems, domain-specific tasks remain difficult due to polysemous words with rare usages, detailed answer requirements (in a QA setting) and fine nuances between different requests.

While the presented abstract provides promising initial insights, further investigations of polysemous words and homographs would be greatly appreciated. Additionally, representations of not just individual words, but entire sentences might also be of interest. For example, in a biomedical setup, explanations (or questions regarding) different drug therapies where only a single word is different (e.g., the name of the drug), would need to yield significantly different embeddings to separate the topic of the requests.

Despite being valuable on its own, it would be great if the authors would not only provide an analysis, but also provide guidance. E.g., as alluded to in the abstract, recurrent or convolutional approaches might be more capable for closed-domain-datasets in which very specific knowledge is required. When should such an approach be chosen over a transformer based approach, especially in light of limited domain specific data.

The preliminary analysis is interesting, however, the presented abstract lacks deeper insights as to how such problems could be addressed.