**Link to published version**: https://www.aclweb.org/anthology/D19-1500/

**Abstract**:

Privacy policies are long and complex documents that are difficult for users to read and understand. Yet, they have legal effects on how user data can be collected, managed

and used. Ideally, we would like to empower users to inform themselves about the issues that matter to them, and enable them to selectively explore these issues. We present PrivacyQA, a corpus consisting of 1750 questions about the privacy policies of mobile applications, and over 3500 expert annotations of relevant answers. We observe that a strong neural baseline underperforms human performance by almost 0.3 F1 on PrivacyQA, suggesting considerable room for improvement for future systems. Further, we use this dataset to categorically identify challenges to question answerability, with domain-general implications for any question answering system. The PrivacyQA corpus offers a challenging corpus for question answering, with genuine real world utility.