Deidentification of PII in knowledge graphs using Direct Preference Optimization

Divyaksh Shukla
Department of Computer Science and Engineering
IIT Kanpur
Kanpur, India
Email: divyaksh@cse.iitk.ac.in

***Abstract*—Large language models (LLMs) have learned to capture language effectively. However, their responses to certain queries or instructions are not preferrable to humans. Reinforcement learning from human feedback provided a solution to generate preferred responses but was complex to train and shown to drift away from the base LLM to maximize rewards. We propose to parameterize the reward function of the RLHF loop, leading to the elimination of the reward, thus reducing the problem to maximum likelihood estimation of the aligned model with respect to human preferences. We test our strategy on generating graphs (via node-relationship triplets) on legal documents and aligning the responses to redact personally identifiable information, thus, generating de-identified document graphs. Such graphs can be used by lawyers for prior case retrieval.**

REFERENCES