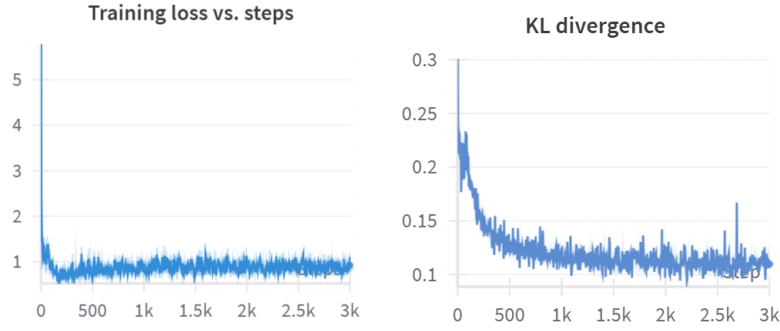# Rebuttal Response

Figure 1. The evolution of the lower-level training loss and KL divergence loss in training of 10B tokens. Proxy model size: 31m, target LLM size: 410M.
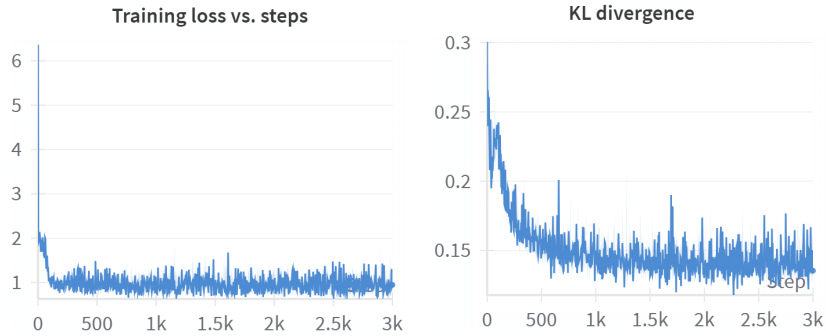

Figure 2. The evolution of the lower-level training loss and KL divergence loss in training of 10B tokens. Proxy model size: 160m, target LLM size: 410M.