

---

## Rebuttal Response

---

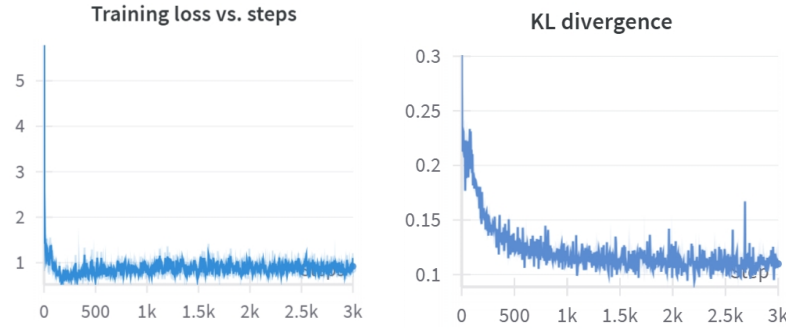


Figure 1. The evolution of the lower-level training loss and KL divergence loss in training of 10B tokens.