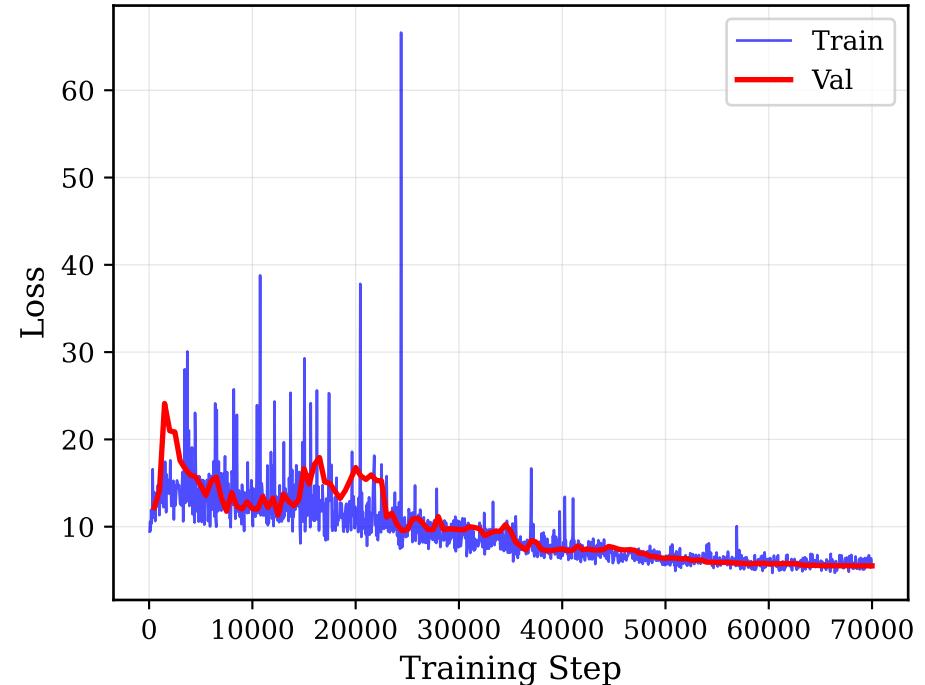
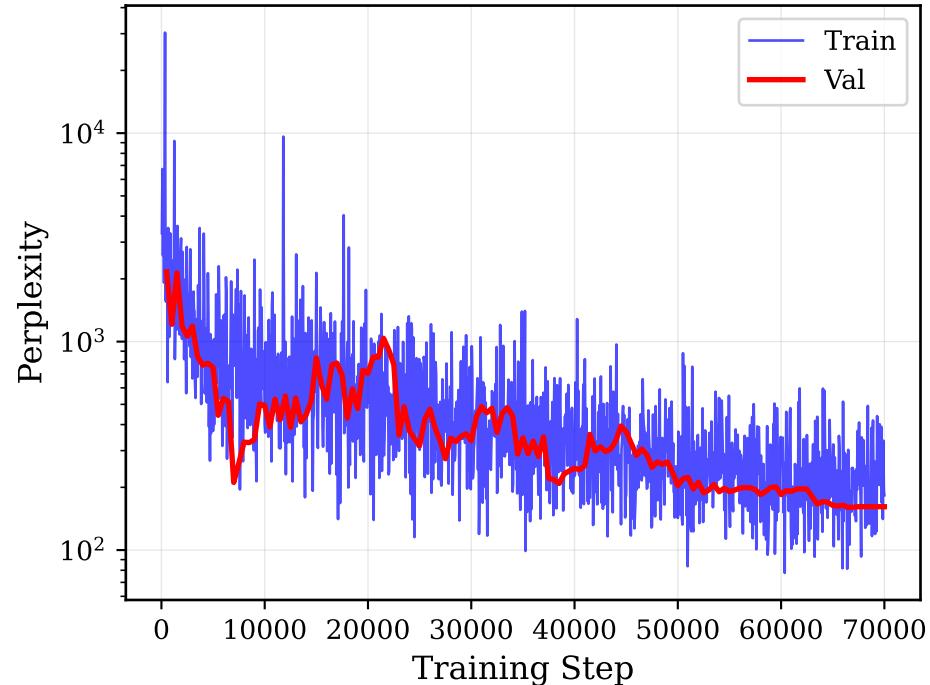


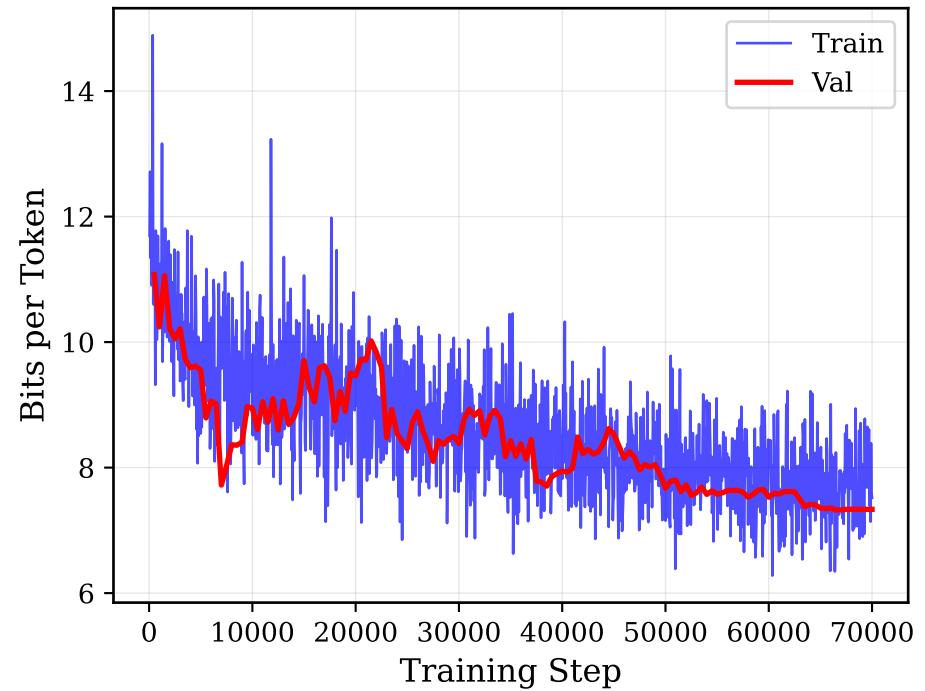
(a) Total Loss



(b) Perplexity



(c) Bits per Token



(d) Gradient Norms

