

Gemma-7B-it on harmful_test dataset [68, 78) (117/313 samples, 37.4%)

