

Histogram of Zephyr Perplexity Ratios on Harmful and Harmless Subsets of FairPrism (32 epochs)

