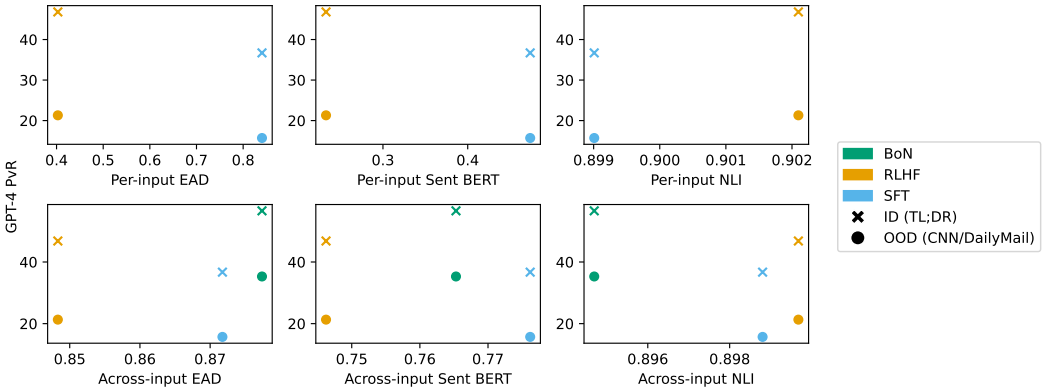

Figure 25: We plot the diversity vs gpt4 win rate trade-off for the summarisation task, across in-distribution and out-of-distribution winrates and per- and across-input diversity metrics.

L GENERALISATION VS DIVERSITY TRADE-OFF PLOTS

Fig. 25 shows the tradeoff between diversity and win rate in the summarisation task, across the three policy types we investigate. This reinforces the inherent tradeoff between generalisation and diversity present in existing language model fine-tuning techniques.