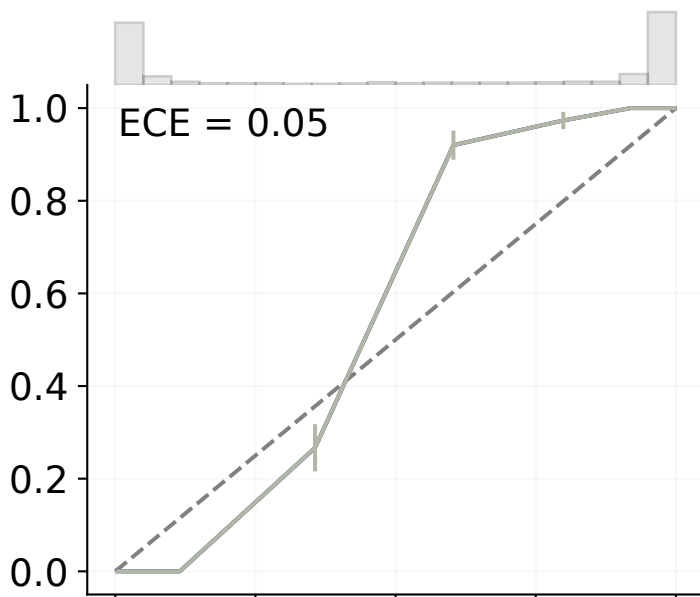
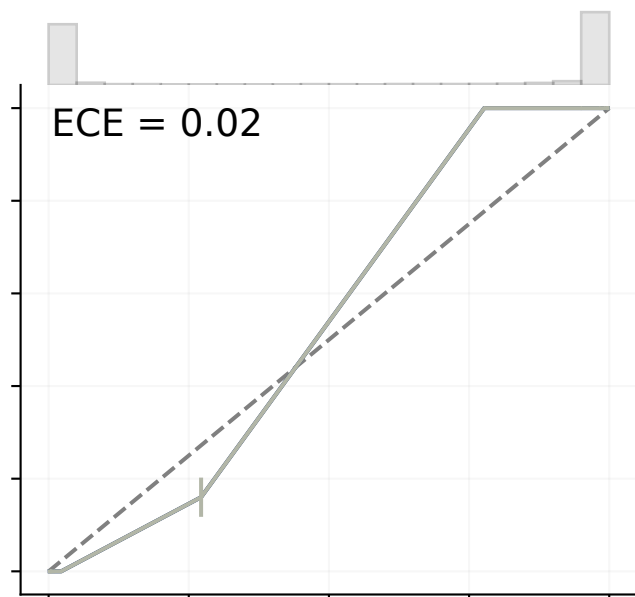


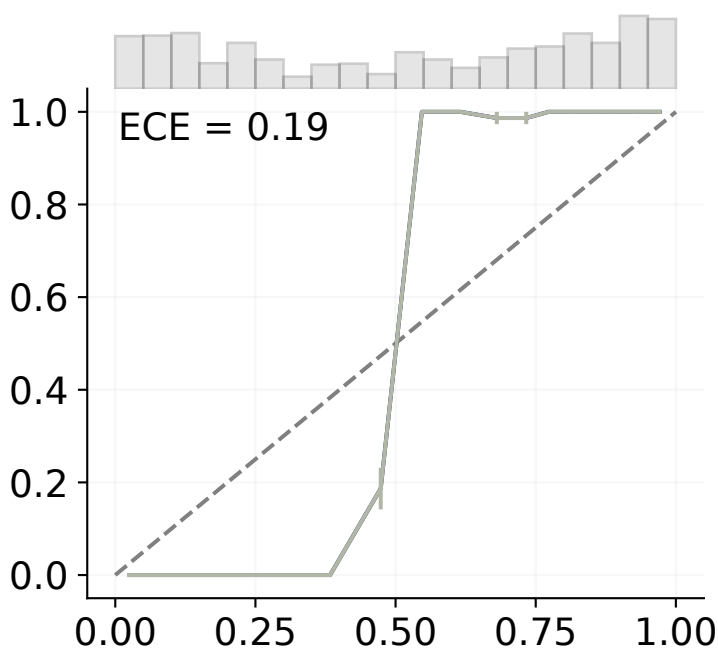
google/gemma-2-2b-it



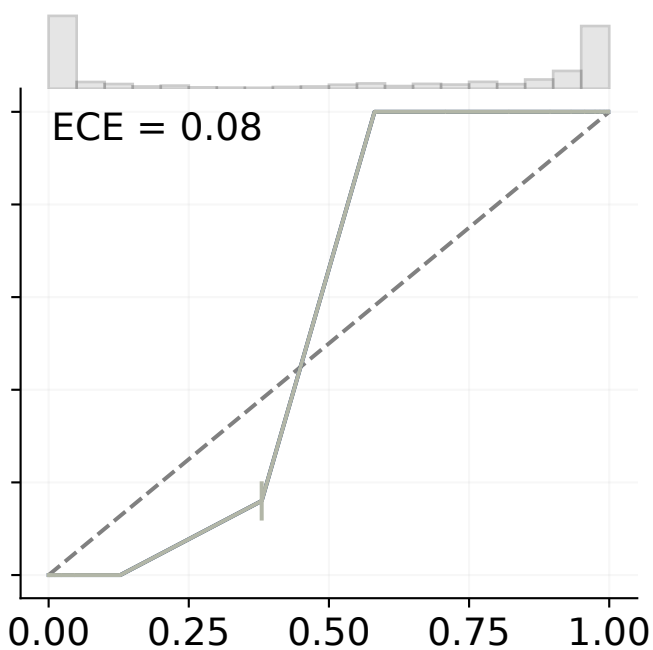
google/gemma-2-9b-it



meta-llama/Llama-3.2-1B-Instruct



meta-llama/Llama-3.2-3B-Instruct



Mean predicted probability

Proportion of positive samples