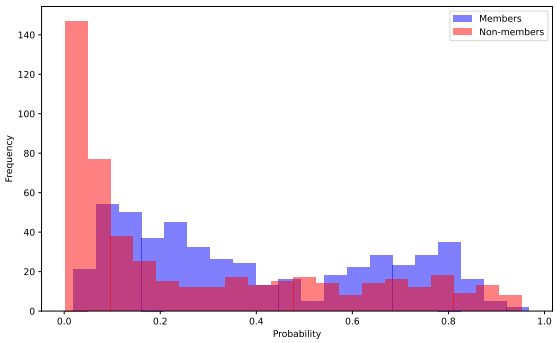
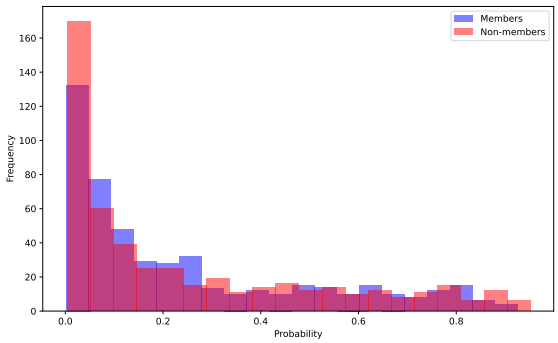



(a) $\varepsilon = \infty$




(b) $\varepsilon < 1$

**Fig. 1**: **Membership Inference on Vicuna 7B with trec.** We follow the membership inference attacks proposed in [1] (e.g., Figure 2). We use Vicuna-7B prompted with 500 different one-shot examples from trec. We present the prediction probabilities of the ground truth label for members and non-members (randomly sampled points from the validation set). (a) The output probabilities for members are significantly higher than for non-members when the prompts are selected without any privacy protection ($\varepsilon = \infty$). However, (b) the difference is substantially reduced, with similar probability values for members and non-members, when we select the prompts with privacy using PromptPATE [1] (in this case with $\varepsilon < 1$).

## 1. REFERENCES

[1] *Flocks of Stochastic Parrots: Differentially Private Prompt Learning for Large Language Models.* Haonan Duan, Adam Dziedzic, Nicolas Papernot, Franziska Boenisch. NeurIPS 2023.