Method	Wiki					AG News					Xsum				
	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**
Loss Attack	0.614	0.577	0.593	0.605	0.597	0.591	0.529	0.554	0.580	0.564	0.628	0.564	0.577	0.594	0.591
Neighbour Attack	0.647	0.612	0.621	0.627	0.627	0.622	0.587	0.594	0.610	0.603	0.612	0.547	0.571	0.582	0.578
DetectGPT	0.623	0.587	0.603	0.619	0.608	0.611	0.579	0.582	0.603	0.594	0.603	0.541	0.563	0.577	0.571
Min-K%	0.658	0.623	0.629	0.643	0.638	0.629	0.604	0.607	0.619	0.615	0.621	0.562	0.588	0.594	0.591
Min-K%++	0.623	0.613	0.645	0.648	0.635	0.635	0.609	0.623	0.631	0.625	0.627	0.556	0.589	0.604	0.594
LiRA-Base	0.710	0.681	0.694	0.709	0.699	0.658	0.634	0.641	0.657	0.648	0.776	0.718	0.734	0.759	0.747
LiRA-Candidate	0.769	0.726	0.735	0.748	0.744	0.717	0.690	0.708	0.714	0.707	0.823	0.772	0.785	0.809	0.797
SPV-MIA	**0.975**	**0.929**	**0.932**	**0.951**	**0.938**	**0.949**	**0.885**	**0.898**	**0.903**	**0.909**	**0.944**	**0.897**	**0.918**	**0.937**	**0.924**

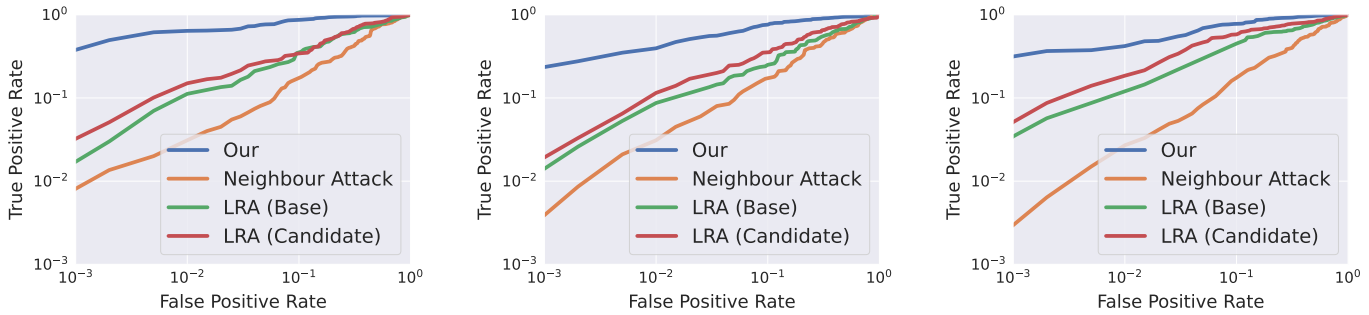
Table 1: Evaluation of all baselines and SPV-MIA using **AUC scores**.

Method	Wiki					AG News					Xsum				
	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**
Loss Attack	1.3%	1.2%	1.1%	1.4%	1.2%	1.3%	1.0%	1.3%	1.2%	1.2%	1.5%	1.0%	1.0%	1.1%	1.2%
Neighbour Attack	4.1%	3.6%	2.8%	3.4%	3.5%	3.6%	2.7%	2.8%	3.1%	3.1%	3.2%	2.4%	2.5%	2.7%	2.7%
DetectGPT	3.7%	3.1%	2.6%	3.2%	3.2%	3.3%	2.4%	2.6%	2.7%	2.8%	3.0%	2.1%	2.4%	2.6%	2.5%
Min-K%	4.4%	4.3%	3.4%	3.7%	4.0%	3.7%	3.4%	3.8%	3.6%	3.6%	3.4%	2.5%	2.7%	3.1%	2.9%
Min-K%++	3.7%	4.2%	3.8%	3.9%	3.9%	4.0%	3.3%	3.9 %	4.1%	3.8%	3.1%	2.8%	3.2%	3.4%	3.1%
LiRA-Base	12.5%	11.3%	10.7%	11.2%	11.4%	9.2%	8.0%	8.3%	8.7%	8.6%	13.5%	9.3%	10.7%	12.2%	11.4%
LiRA-Candidate	16.3%	14.3%	14.8%	15.0%	15.1%	12.2%	9.4%	10.6%	11.5%	10.9%	19.4%	10.9%	14.5%	18.5%	15.8%
SPV-MIA	**67.3%**	**55.4%**	**57.6%**	**64.2%**	**61.1%**	**42.9%**	**34.8%**	**37.6%**	**39.5%**	**38.7%**	**42.1%**	**38.6%**	**40.7%**	**42.0%**	**40.9%**

Table 2: Evaluation of all baselines and SPV-MIA using **TPR@1%FPR**.

Method	Wiki					AG News					Xsum				
	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**	GPT-2	GPT-J	Falcon	LLaMA	**Avg.**
Loss Attack	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%	0.2%	0.1%	0.1%	0.1%
Neighbour Attack	1.2%	0.6%	0.5%	0.8%	0.8%	0.8%	0.4%	0.3%	0.4%	0.5%	0.5%	0.3%	0.3%	0.3%	0.4%
DetectGPT	0.9%	0.4%	0.5%	0.6%	0.6%	0.6%	0.2%	0.3%	0.4%	0.4%	0.4%	0.2%	0.2%	0.3%	0.3%
Min-K%	1.4%	0.6%	0.7%	0.9%	0.9%	1.2%	0.4%	0.7%	0.6%	0.7%	0.5%	0.3%	0.4%	0.4%	0.4%
Min-K%++	1.2%	0.7%	0.9%	1.1%	1.0%	1.4%	0.9%	1.1%	1.0%	1.1%	0.8%	0.4%	0.7%	0.6%	0.6%
LiRA-Base	1.9%	1.4%	1.5%	1.7%	1.6%	1.8%	1.4%	1.3%	1.4%	1.5%	3.1%	2.5%	2.6%	3.5%	2.9%
LiRA-Candidate	3.7%	2.5%	2.8%	3.2%	3.1%	2.3%	1.8%	1.7%	1.9%	1.9%	4.7%	3.4%	3.8%	5.1%	4.3%
SPV-MIA	**39.1%**	**28.9%**	**32.7%**	**37.8%**	**34.6%**	**25.3%**	**17.3%**	**18.7%**	**23.5%**	**21.2%**	**34.4%**	**27.6%**	**28.9%**	**31.5%**	**30.6%**

Table 3: Evaluation of all baselines and SPV-MIA using **TPR@0.1%FPR**.



(a) Wiki (b) AG News (c) Xsum

Figure 1: **Full log-scale ROC curves** of SPV-MIA and the three representative baselines on LLaMAs fine-tuned over three datasets.

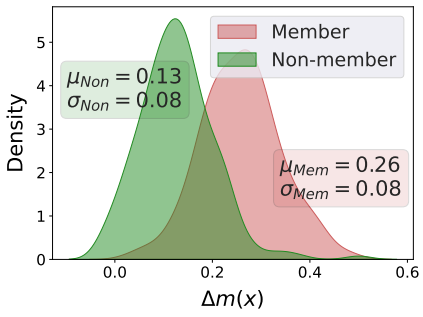


Figure 2: The distributions of member and non-member records w.r.t MIA metric score $\Delta m(x)$.