

Table 6: Additional comparison between BiSCOPE and three more baselines on both normal and paraphrased datasets with in-distribution and out-of-distribution (OOD) settings.

		Normal Dataset					Paraphrased Dataset				Normal	Paraphrased
Method		GPT-3.5 Turbo	GPT-4 Turbo	Claude-3 Sonnet	Claude-3 Opus	Gemini 1.0-pro	GPT-3.5 Turbo	GPT-4 Turbo	Claude-3 Sonnet	Claude-3 Opus	Avg.	Avg.
Arxiv	OpenAI Detector	0.7889	0.6660	0.6673	0.6673	0.6976	0.7062	0.6654	0.6673	0.6673	0.6409	0.6705
	Binoculars	0.9097	0.9135	0.9256	0.9699	0.9560	0.6617	0.6971	0.8112	0.8672	0.8681	0.5835
	GhostBuster	0.9716	0.9886	0.9815	0.9813	0.9571	0.9700	0.9943	0.9814	0.9856	0.7999	0.9672
	BiSCOPE	0.9870	0.9928	0.9796	0.9885	0.9708	0.9769	0.9800	0.9625	0.9870	0.8324	0.8534
	BiSCOPE*	0.9928	0.9943	0.9869	0.9913	0.9797	0.9870	0.9859	0.9593	0.9884	0.8771	0.8740
Yelp	OpenAI Detector	0.7286	0.6668	0.6668	0.6616	0.6798	0.7240	0.6668	0.6668	0.6668	0.6328	0.6563
	Binoculars	0.7295	0.6665	0.7583	0.8260	0.6885	0.6683	0.6655	0.6908	0.7284	0.7702	0.6681
	GhostBuster	0.8193	0.8369	0.8746	0.8644	0.8625	0.8174	0.8649	0.9271	0.9145	0.6917	0.8452
	BiSCOPE	0.9023	0.9405	0.9652	0.9532	0.9486	0.9064	0.9473	0.9814	0.9789	0.8835	0.9523
	BiSCOPE*	0.9010	0.9452	0.9658	0.9570	0.9545	0.9102	0.9530	0.9830	0.9757	0.8791	0.9505
Creative	OpenAI Detector	0.6666	0.6671	0.6669	0.6671	0.6271	0.6671	0.6671	0.6671	0.6671	0.6406	0.5475
	Binoculars	0.9945	0.9681	0.9814	0.9866	0.9880	0.9627	0.8381	0.9348	0.9540	0.8655	0.8738
	GhostBuster	0.9965	0.9821	0.9834	0.9834	0.9920	0.9861	0.9786	0.9871	0.9865	0.8854	0.9012
	BiSCOPE	0.9985	0.9950	0.9960	0.9930	0.9964	0.9955	0.9945	0.9955	0.9940	0.8913	0.9707
	BiSCOPE*	0.9975	0.9955	0.9955	0.9945	0.9970	0.9955	0.9955	0.9950	0.9945	0.8967	0.9513
Essay	OpenAI Detector	0.7069	0.6664	0.6667	0.6669	0.6426	0.6669	0.6662	0.6667	0.6664	0.6141	0.5300
	Binoculars	0.9995	0.9970	0.9945	0.9960	0.9978	0.9920	0.9607	0.9787	0.9955	0.8675	0.9429
	GhostBuster	0.9995	0.9950	0.9960	0.9965	0.9967	0.9916	0.9861	0.9880	0.9930	0.8772	0.9435
	BiSCOPE	1.0000	0.9990	0.9985	0.9970	0.9994	0.9965	0.9990	0.9990	0.9980	0.7701	0.9435
	BiSCOPE*	1.0000	0.9990	0.9985	0.9975	0.9989	0.9975	0.9990	0.9985	0.9985	0.7791	0.9292
Code	OpenAI Detector	0.7213	0.6977	0.6916	0.6542	0.6666	0.7514	0.6639	0.6639	0.6695	0.5325	0.5767
	Binoculars	0.7073	0.6512	0.6612	0.6653	0.6624	0.7101	0.6338	0.8041	0.7179	0.6727	0.6771
	GhostBuster	0.8524	0.7942	0.6556	0.6749	0.3860	0.8662	0.7729	0.7757	0.5390	0.5662	0.6790
	BiSCOPE	0.9665	0.9655	0.8528	0.6069	0.7809	0.9659	0.9464	0.9691	0.9250	0.6934	0.8999
	BiSCOPE*	0.9692	0.9586	0.8526	0.6620	0.7741	0.9597	0.9435	0.9600	0.9222	0.6877	0.9024

Table 7: Ablation results of different segmentation methods in multi-point splitting of BiSCOPE.

		Normal Dataset					Paraphrased Dataset				Normal	Paraphrased
Method		GPT-3.5 Turbo	GPT-4 Turbo	Claude-3 Sonnet	Claude-3 Opus	Gemini 1.0-pro	GPT-3.5 Turbo	GPT-4 Turbo	Claude-3 Sonnet	Claude-3 Opus	Avg.	Avg.
Arxiv	Every 50% Text	0.9754	0.9871	0.9635	0.9855	0.9637	0.9587	0.9756	0.9486	0.9767	0.9750	0.9649
	Every 25% Text	0.9813	0.9914	0.9662	0.9884	0.9690	0.9675	0.9769	0.9622	0.9797	0.9792	0.9716
	Every 10% Text (In Paper)	0.9870	0.9928	0.9796	0.9885	0.9708	0.9769	0.9800	0.9625	0.9870	0.9837	0.9766
	Every 50% Text	0.8922	0.9314	0.9584	0.9471	0.9337	0.8921	0.9359	0.9779	0.9704	0.9326	0.9441
	Every 25% Text	0.9002	0.9381	0.9651	0.9527	0.9466	0.9041	0.9452	0.9817	0.9757	0.9405	0.9517
Yelp	Every 10% Text (In Paper)	0.9023	0.9405	0.9652	0.9532	0.9486	0.9064	0.9473	0.9814	0.9789	0.9420	0.9535
	Every 50% Text	0.9985	0.9960	0.9940	0.9955	0.9958	0.9950	0.9955	0.9935	0.9930	0.9960	0.9943
	Every 25% Text	0.9980	0.9955	0.9960	0.9930	0.9970	0.9960	0.9955	0.9950	0.9935	0.9959	0.9950
	Every 10% Text (In Paper)	0.9985	0.9950	0.9960	0.9930	0.9964	0.9955	0.9945	0.9955	0.9940	0.9958	0.9949
	Every 50% Text	1.0000	0.9990	0.9965	0.9970	0.9994	0.9975	0.9995	0.9975	0.9975	0.9984	0.9980
Creative	Every 25% Text	1.0000	0.9990	0.9985	0.9980	0.9994	0.9965	0.9990	0.9985	0.9980	0.9990	0.9980
	Every 10% Text (In Paper)	1.0000	0.9990	0.9985	0.9970	0.9994	0.9965	0.9990	0.9990	0.9980	0.9988	0.9981
	Every 50% Text	0.8564	0.8790	0.7706	0.5933	0.6479	0.8798	0.8752	0.9211	0.8427	0.7495	0.8797
	Every 25% Text	0.9532	0.9333	0.8115	0.6184	0.7088	0.9363	0.9322	0.9470	0.8856	0.8050	0.9252
	Every 10% Text (In Paper)	0.9665	0.9655	0.8528	0.6069	0.7809	0.9659	0.9464	0.9691	0.9250	0.8345	0.9516

Table 8: Performance results when not using the completion prompt in BiSCOPE.

		Normal Dataset					Paraphrased Dataset				Normal	Paraphrased
Method		GPT-3.5 Turbo	GPT-4 Turbo	Claude-3 Sonnet	Claude-3 Opus	Gemini 1.0-pro	GPT-3.5 Turbo	GPT-4 Turbo	Claude-3 Sonnet	Claude-3 Opus	Avg.	Avg.
Arxiv	w/o Completion Prompt	0.9813	0.9914	0.9767	0.9900	0.9692	0.9714	0.9855	0.9621	0.9766	0.9817	0.9739
	w/ Completion Prompt	0.9870	0.9928	0.9796	0.9885	0.9708	0.9769	0.9800	0.9625	0.9870	0.9837	0.9766
Yelp	w/o Completion Prompt	0.9008	0.9471	0.9686	0.9577	0.9523	0.9058	0.9576	0.9872	0.9817	0.9453	0.9581
	w/ Completion Prompt	0.9023	0.9405	0.9652	0.9532	0.9486	0.9064	0.9473	0.9814	0.9789	0.9420	0.9535
Creative	w/o Completion Prompt	0.9980	0.9950	0.9950	0.9925	0.9952	0.9945	0.9955	0.9950	0.9940	0.9952	0.9948
	w/ Completion Prompt	0.9985	0.9950	0.9960	0.9930	0.9964	0.9955	0.9945	0.9955	0.9940	0.9958	0.9949
Essay	w/o Completion Prompt	1.0000	0.9995	0.9990	0.9975	0.9994	0.9975	0.9990	0.9990	0.9975	0.9991	0.9982
	w/ Completion Prompt	1.0000	0.9990	0.9985	0.9970	0.9994	0.9965	0.9990	0.9990	0.9980	0.9988	0.9981
Code	w/o Completion Prompt	0.9786	0.9752	0.8327	0.6703	0.7653	0.9720	0.9532	0.9717	0.9210	0.8444	0.9545
	w/ Completion Prompt	0.9665	0.9655	0.8528	0.6069	0.7809	0.9659	0.9464	0.9691	0.9250	0.8345	0.9516

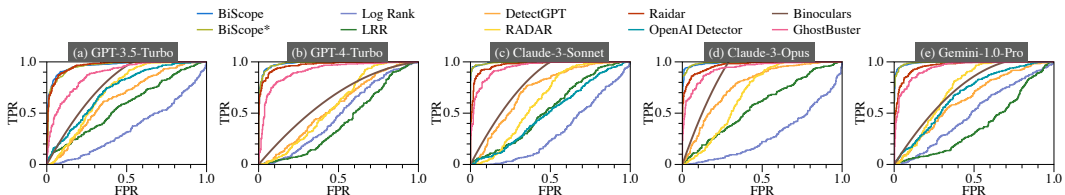


Figure 9: ROC curves of BiSCOPE and all the baselines.