

Table 1: TSU of TriviaQA Dataset

TriviaQA	Two-Parameter				Three-Parameter			Multi-Parameter		
	TSU (0.3, 0.7)	TSU (0.7, 1)	TSU (1, 1.4)	TSU (1.4, 2)	TSU (0.3, 0.7, 1)	TSU (0.7, 1, 1.4)	TSU (1, 1.4, 2)	TSU (0.3-1.4)	TSU (0.7-2)	TSU (0.3-2)
Inv-Entropy	86.7%	90.5%	93.5%	95.0%	85.4%	83.2%	81.9%	80.9%	78.6%	77.1%
NI Entropy	73.8%	80.1%	89.2%	90.0%	70.4%	66.2%	61.9%	54.3%	51.7%	49.2%
NR-Inv-Entropy	75.6%	78.3%	81.9%	84.0%	70.2%	68.5%	65.4%	63.2%	60.3%	58.7%
NRI-Entropy	62.8%	68.2%	75.5%	78.3%	58.1%	56.2%	55.7%	55.4%	52.4%	50.2%
L-Perplexity	39.2%	42.5%	46.7%	50.0%	30.2%	28.3%	25.4%	23.7%	21.9%	20.1%
L-Entropy	49.8%	57.3%	68.4%	72.5%	45.1%	43.0%	42.0%	40.5%	38.3%	36.2%
VU	53.4%	59.2%	63.3%	65.0%	51.1%	49.2%	48.0%	47.0%	45.2%	44.0%
Semantic Entropy	66.3%	70.2%	74.7%	78.0%	60.3%	59.1%	58.9%	58.8%	56.1%	54.9%
Lexical sim.	80.2%	86.4%	91.5%	93.0%	75.6%	74.2%	73.4%	72.8%	70.1%	68.5%
DegMat	84.1%	87.2%	89.7%	91.0%	80.3%	79.1%	78.2%	77.3%	75.0%	73.2%

Table 2: TSU of SimpleQA Dataset

SimpleQA	Two-Parameter				Three-Parameter			Multi-Parameter		
	TSU (0.3, 0.7)	TSU (0.7, 1)	TSU (1, 1.4)	TSU (1.4, 2)	TSU (0.3, 0.7, 1)	TSU (0.7, 1, 1.4)	TSU (1, 1.4, 2)	TSU (0.3-1.4)	TSU (0.7-2)	TSU (0.3-2)
Inv-Entropy	90.7%	91.2%	91.9%	92.5%	89.3%	88.0%	87.5%	85.2%	83.9%	83.3%
NI Entropy	53.3%	60.1%	72.1%	74.0%	52.0%	51.2%	50.4%	49.8%	48.2%	47.0%
NR-Inv-Entropy	60.0%	65.1%	78.3%	80.0%	59.4%	58.2%	58.0%	57.6%	56.3%	55.1%
NRI-Entropy	60.7%	64.3%	69.8%	72.0%	59.2%	58.4%	54.1%	53.2%	52.5%	50.8%
L-Perplexity	34.5%	42.0%	59.1%	62.0%	33.1%	31.5%	30.2%	28.3%	26.9%	25.4%
L-Entropy	46.2%	54.0%	66.0%	68.9%	44.0%	43.2%	42.0%	41.7%	40.3%	39.0%
VU	38.9%	37.2%	10.4%	12.0%	36.3%	34.1%	33.0%	32.1%	31.2%	29.8%
Semantic Entropy	63.7%	66.0%	69.5%	72.0%	62.2%	60.1%	59.3%	58.2%	57.1%	55.9%
Lexical sim.	87.3%	88.0%	90.4%	91.5%	86.2%	85.0%	84.3%	82.6%	81.4%	80.7%
DegMat	81.2%	82.4%	88.6%	89.5%	80.7%	79.1%	77.9%	75.4%	73.6%	72.0%

Table 3: AUROC, PRR and Brier of TriviaQA Dataset

TriviaQA	AUROC (\uparrow)		PRR (\uparrow)		Brier (\downarrow)	
	GPT-3.5	LLaMA-8B	GPT-3.5	LLaMA-8B	GPT-3.5	LLaMA-8B
Inv-Entropy	0.810 \pm 0.050	0.795 \pm 0.043	0.907 \pm 0.036	0.871 \pm 0.041	0.117 \pm 0.020	0.151 \pm 0.018
NI Entropy	0.639 \pm 0.057	0.713 \pm 0.052	0.828 \pm 0.044	0.769 \pm 0.056	0.156 \pm 0.021	0.182 \pm 0.019
NR-Inv-Entropy	0.804 \pm 0.051	0.790 \pm 0.045	0.906 \pm 0.032	0.843 \pm 0.046	0.121 \pm 0.022	0.156 \pm 0.020
NRI-Entropy	0.636 \pm 0.060	0.700 \pm 0.054	0.825 \pm 0.045	0.747 \pm 0.062	0.155 \pm 0.022	0.190 \pm 0.019
L-Perplexity	0.760 \pm 0.029	0.654 \pm 0.067	0.890 \pm 0.039	0.836 \pm 0.049	0.145 \pm 0.019	0.153 \pm 0.019
L-Entropy	0.780 \pm 0.031	0.679 \pm 0.058	0.905 \pm 0.034	0.853 \pm 0.041	0.150 \pm 0.017	0.157 \pm 0.017
VU	0.800 \pm 0.028	0.780 \pm 0.032	0.895 \pm 0.033	0.865 \pm 0.035	0.170 \pm 0.020	0.167 \pm 0.019
Semantic Entropy	0.670 \pm 0.028	0.652 \pm 0.068	0.885 \pm 0.037	0.835 \pm 0.049	0.158 \pm 0.018	0.161 \pm 0.019
SAR	–	0.668 \pm 0.056	–	0.871 \pm 0.035	–	0.160 \pm 0.018
Lexical sim.	0.774 \pm 0.043	0.632 \pm 0.060	0.812 \pm 0.050	0.825 \pm 0.047	0.165 \pm 0.020	0.146 \pm 0.023
DegMat	0.790 \pm 0.030	0.665 \pm 0.060	0.880 \pm 0.036	0.850 \pm 0.046	0.166 \pm 0.018	0.159 \pm 0.018

Table 4: AUROC, PRR and Brier of SciQ Dataset

SciQ	AUROC (\uparrow)		PRR (\uparrow)		Brier (\downarrow)	
	GPT-3.5	LLaMA-8B	GPT-3.5	LLaMA-8B	GPT-3.5	LLaMA-8B
Inv-Entropy	0.754 \pm 0.053	0.751 \pm 0.051	0.852 \pm 0.047	0.846 \pm 0.041	0.156 \pm 0.021	0.161 \pm 0.018
NI Entropy	0.593 \pm 0.061	0.602 \pm 0.067	0.755 \pm 0.054	0.705 \pm 0.064	0.197 \pm 0.017	0.206 \pm 0.017
NR-Inv-Entropy	0.712 \pm 0.056	0.642 \pm 0.059	0.816 \pm 0.052	0.751 \pm 0.055	0.174 \pm 0.021	0.198 \pm 0.016
NRI-Entropy	0.565 \pm 0.063	0.615 \pm 0.065	0.717 \pm 0.062	0.706 \pm 0.064	0.197 \pm 0.018	0.206 \pm 0.018
L-Perplexity	0.700 \pm 0.028	0.692 \pm 0.053	0.800 \pm 0.035	0.802 \pm 0.050	0.185 \pm 0.017	0.177 \pm 0.017
L-Entropy	0.730 \pm 0.029	0.752 \pm 0.045	0.845 \pm 0.036	0.842 \pm 0.044	0.188 \pm 0.016	0.158 \pm 0.016
VU	0.740 \pm 0.030	0.730 \pm 0.032	0.840 \pm 0.032	0.830 \pm 0.035	0.190 \pm 0.018	0.180 \pm 0.017
Semantic Entropy	0.710 \pm 0.031	0.699 \pm 0.049	0.830 \pm 0.034	0.808 \pm 0.049	0.185 \pm 0.017	0.179 \pm 0.017
SAR	–	0.661 \pm 0.060	–	0.755 \pm 0.055	–	0.192 \pm 0.018
Lexical sim.	0.610 \pm 0.027	0.629 \pm 0.051	0.800 \pm 0.035	0.719 \pm 0.054	0.191 \pm 0.018	0.146 \pm 0.023
DegMat	0.700 \pm 0.028	0.610 \pm 0.057	0.810 \pm 0.038	0.716 \pm 0.066	0.195 \pm 0.017	0.186 \pm 0.018