

**A Averaged Performance of the Baseline Dynamic Prompt Model**

Precision			Recall	F <sub>1</sub> -score	Precision			Recall	F <sub>1</sub> -score
GPT 4	5-shot	16.40	50.00	24.70	Llama3	5-shot	12.56	56.32	20.55
		17.67	50.62	26.20			13.68	55.81	21.97
		17.48	53.09	26.30			12.14	59.52	20.16
		23.91	54.32	33.20			14.25	59.79	23.02
	AVG	18.865	52.0075	27.60		AVG	13.1575	57.86	21.425
	10-shot	25.41	58.75	35.47		10-shot	24.51	59.52	34.72
		21.67	54.32	30.99			19.63	62.35	29.86
		21.43	59.26	31.48			23.96	57.44	33.81
		20.47	54.32	29.73			21.38	60.43	31.59
	AVG	22.245	56.6625	31.9175		AVG	22.37	59.935	32.495
	20-shot	28.74	58.63	38.57		20-shot	27.22	57.65	36.98
		27.03	57.54	36.78			23.24	52.44	32.21
		26.52	59.79	36.74			22.12	52.08	31.06
		28.65	59.02	38.57			25.49	53.06	34.44
	AVG	27.735	58.745	37.665		AVG	24.5175	53.8075	33.6725

**Table 1.** Averaged performance of the baseline dynamic prompt model on the REDDIT-IMPACTS dataset across different shot settings.

		Precision	Recall	F <sub>1</sub> -score			Precision	Recall	F <sub>1</sub> -score
GPT 4	5-shot	69.97	91.43	79.27	Llama3	5-shot	68.09	84.35	75.35
		68.32	88.3	77.29			69.70	76.67	73.02
		64.58	90.15	75.25			71.00	79.15	74.86
		71.59	91.39	80.29			67.10	73.25	70.04
	AVG	68.615	90.3175	78.025		AVG	68.9725	78.355	73.3175
	10-shot	75.11	90.84	82.23		10-shot	72.38	79.18	75.63
		74.41	90.32	81.60			73.31	74.24	73.77
		74.13	85.71	79.50			71.15	77.05	73.98
		77.65	86.35	81.73			73.40	81.17	77.2
	AVG	75.325	88.305	81.265		AVG	72.56	77.91	75.145
	20-shot	75.36	88.33	81.33		20-shot	75.12	74.84	74.98
		73.13	91.74	81.38			72.03	71.88	71.96
		71.84	91.1	80.33			76.88	77.20	77.04
		77.94	85.54	81.57			77.64	78.40	78.02
	AVG	74.5675	89.1775	81.1525		AVG	75.4175	75.58	75.5

**Table 2.** Averaged performance of the baseline dynamic prompt model on the BC5CDR dataset across different shot settings.

		Precision	Recall	F <sub>1</sub> -score			Precision	Recall	F <sub>1</sub> -score
GPT 4	5-shot	62.38	63.78	63.07	Llama3	5-shot	57.69	68.39	62.58
		61.88	62.19	62.03			61.83	61.33	61.58
		65.26	68.14	66.67			58.24	68.06	62.77
		62.71	62.36	62.54			59.44	71.29	64.82
	AVG	63.0575	64.1175	63.5775		AVG	59.3	67.2675	62.9375
	10-shot	66.38	74.24	70.09		10-shot	59.13	71.63	64.78
		68.37	74.86	71.47			62.68	63.8	63.23
		66.72	75.33	70.77			61.71	67.58	64.51
		66.46	73.34	69.73			62.24	62.84	62.54
	AVG	66.9825	74.4425	70.515		AVG	61.44	66.4625	63.765
	20-shot	70.41	69.84	70.12		20-shot	63.89	62.04	62.95
		70.11	71.24	70.67			63.96	62.83	63.39
		71.45	73.39	72.41			59.22	61.71	60.44
		70.64	70.80	70.72			60.96	61.88	61.42
	AVG	70.6525	71.3175	70.98		AVG	62.0075	62.115	62.05

**Table 3.** Averaged performance of the baseline dynamic prompt model on the MIMIC III dataset across different shot settings.

		Precision	Recall	F <sub>1</sub> -score			Precision	Recall	F <sub>1</sub> -score
GPT 4	5-shot	46.17	50.52	48.25	Llama3	5-shot	33.88	40.67	36.97
		45.69	49.07	47.32			32.26	39.14	35.37
		40.24	43.65	41.88			36.38	28.49	31.95
		47.96	52.83	50.28			40.73	30.53	34.9
	AVG	45.015	49.0175	46.9325		AVG	35.8125	34.7075	34.7975
	10-shot	52.98	52.67	52.82		10-shot	42.79	31.08	35.66
		53.71	53.36	53.54			40.18	28.73	36.27
		52.99	50.35	51.64			35.75	32.05	33.76
		53.85	52.3	53.06			39.95	34.11	36.71
	AVG	53.3825	52.17	52.765		AVG	39.6675	31.4925	35.6
	20-shot	54.57	54.73	54.65		20-shot	40.59	42.07	41.32
		44.86	45.43	45.14			40.87	42.8	41.81
		51.6	51.75	51.68			41.48	43.02	42.24
		55.7	57.24	56.46			39.88	42.41	41.1
	AVG	51.6825	52.2875	51.9825		AVG	40.705	42.575	41.6175

**Table 4.** Averaged performance of the baseline dynamic prompt model on the NCBI dataset across different shot settings.

		Precision	Recall	F <sub>1</sub> -score			Precision	Recall	F <sub>1</sub> -score
GPT 4	5-shot	27.24	62.73	37.99	Llama3	5-shot	24.94	73.03	37.18
		26.71	58.58	36.69			25.95	61.96	36.58
		29.23	60.47	39.41			26.11	73.57	38.54
		25.86	58.41	35.85			26.55	59.62	36.74
	AVG	27.26	60.0475	37.485		AVG	25.8875	67.045	37.26
	10-shot	26.92	58.26	36.82		10-shot	25.67	70.61	37.65
		26.1	59.15	36.22			25.1	59.19	35.15
		24.41	58.56	34.46			25.76	70.06	37.67
		29.19	60.82	39.45			25.73	57.46	35.54
	AVG	26.655	59.1975	36.7375		AVG	25.565	64.33	36.5025
	20-shot	27.86	59.12	37.87		20-shot	26.18	63.59	37.09
		25.02	60.64	35.42			26.63	61.14	37.1
		29.33	61.71	39.76			25.93	63.57	36.84
		30.18	61.66	40.52			27.54	70.85	39.66
	AVG	28.0975	60.7825	38.3925		AVG	26.57	64.7875	37.6725

**Table 5.** Averaged performance of the baseline dynamic prompt model on the Med-Mentions dataset across different shot settings.