Table 6: **(Q1)**: Metrics to assess the goodness of *layer 1* and *31* to detect personas (Level 2) in *Mistral-7B-Instruct*. Metrics: Silhouette Score (SH), Calinski-Harabasz Score (CH), Euclidean Distance (ED), Davies-Bouldin Score (DB).

Model	Dimension	layer	SH (↑)	CH (↑)	DB (↓)	ED (↑)
Mistral-7B-Instruct	AGREE	1	0.370 ± 0.082	176.5 ± 110.62	0.690 ± 0.084	0.110 ± 0.006
		31	0.251 ± 0.000	164.1 ± 0.00	1.743 ± 0.0008	5.403 ± 0.003
	CONSC	1	0.242 ± 0.113	96.8 ± 17.79	1.753 ± 0.505	0.063 ± 0.02
		31	0.171 ± 0.010	114.6 ± 9.23	2.140 ± 0.086	4.249 ± 0.122
	OPEN	1	0.423 ± 0.020	176.8 ± 13.84	1.006 ± 0.596	0.107 ± 0.023
		31	0.235 ± 0.000	194.2 ± 0.01	1.693 ± 0.000	5.147 ± 0.000
	EXTRA	1	0.459 ± 0.041	172.6 ± 39.78	0.558 ± 0.096	0.122 ± 0.020
		31	0.327 ± 0.000	225.2 ± 0.000	1.102 ± 0.000	8.020 ± 0.000
	NEURO	1	0.392 ± 0.109	123.9 ± 7.74	0.877 ± 0.573	0.107 ± 0.032
		31	0.203 ± 0.001	137.4 ± 3.31	1.873 ± 0.143	5.182 ± 0.271
	VIRTUE	1	0.389 ± 0.098	110.1 ± 7.74	0.961 ± 0.666	0.104 ± 0.032
		31	0.272 ± 0.000	211.5 ± 1.18	1.448 ± 0.056	5.499 ± 0.079
	RELAT	1	0.468 ± 0.077	378.1 ± 226.82	0.794 ± 0.070	0.119 ± 0.005
		31	0.226 ± 0.013	126.5 ± 13.76	1.370 ± 0.081	6.602 ± 0.295
	DEONT	1	0.392 ± 0.085	201.0 ± 120.04	0.740 ± 0.175	0.123 ± 0.011
		31	0.242 ± 0.001	155.4 ± 0.31	1.587 ± 0.082	5.631 ± 0.118
	UTILI	1	0.445 ± 0.064	188.0 ± 62.33	0.563 ± 0.087	0.135 ± 0.012
		31	0.319 ± 0.000	289.4 ± 0.000	1.215 ± 0.000	6.743 ± 0.000
	NIHIL	1	0.529 ± 0.000	436.5 ± 0.000	0.555 ± 0.000	0.130 ± 0.000
		31	0.177 ± 0.011	132.1 ± 17.07	2.058 ± 0.139	5.067 ± 0.278
	CONS	1	0.360 ± 0.098	98.9 ± 25.45	1.253 ± 0.559	0.083 ± 0.026
		31	0.230 ± 0.004	152.9 ± 0.31	1.843 ± 0.023	4.927 ± 0.072
	LIBER	1	0.511 ± 0.026	119.4 ± 49.65	0.495 ± 0.100	0.119 ± 0.002
		31	0.189 ± 0.026	119.4 ± 5.45	1.862 ± 0.326	4.769 ± 0.798
	LGBTQ	1	0.448 ± 0.041	136.8 ± 40.46	0.585 ± 0.101	0.122 ± 0.006
		31	0.255 ± 0.000	215.0 ± 0.01	1.478 ± 0.002	6.092 ± 0.006
	IMMI	1	0.437 ± 0.034	140.0 ± 93.28	0.558 ± 0.046	0.126 ± 0.009
		31	0.200 ± 0.006	149.7 ± 11.71	1.928 ± 0.080	4.945 ± 0.155