

Supplementary Table 3. Predicted probabilities of the GPT-4o model's theme detection in each of the 102 videos

Video	True Label	Nicotine Vaping	Cannabis Vaping	Fashion	Entertainment	Healthy Lifestyle
1	fashion	0.05	0.05	0.8	0.1	0.1
2	fashion	0	0	0.8	0.2	0.1
3	fashion	0	0	0.8	0.1	0.1
4	fashion	0	0	0.8	0.1	0.1
5	fashion	0	0	0.8	0.1	0.1
6	fashion	0	0	0.9	0.1	0.1
7	fashion	0.1	0.1	0.8	0.2	0.2
8	fashion	0	0	0.8	0.1	0.1
9	fashion	0	0	0.8	0.1	0.1
10	fashion	0	0	0.9	0.1	0.1
11	fashion & healthy lifestyle	0	0	0.8	0.1	0.7
12	fashion	0	0	0.8	0.1	0.1
13	fashion	0.05	0.05	0.9	0.1	0.1
14	fashion	0	0	0.9	0.1	0.2
15	fashion	0	0	0.8	0.1	0.1
16	fashion	0	0	0.8	0.1	0.1
17	fashion	0.05	0.05	0.8	0.1	0.1
18	fashion	0	0	0.9	0.1	0.1
19	nicotine vaping & fashion	0.1	0.1	0.6	0.2	0.2
20	fashion	0	0	0.8	0.1	0.1
21	healthy lifestyle	0	0	0	0	1
22	healthy lifestyle	0	0	0	0	1
23	nicotine vaping & healthy lifestyle	0.9	0.1	0.5	0.2	0.3
24	healthy lifestyle	0	0	0	0	1
25	nicotine vaping & healthy lifestyle	0.9	0.1	0	0	1
26	healthy lifestyle	0	0	0	0	1
27	nicotine vaping & healthy lifestyle	0.9	0.1	0	0.1	0.8
28	healthy lifestyle	0	0	0	0	1
29	healthy lifestyle	0	0	0	0	1
30	nicotine vaping & healthy lifestyle	0.9	0.1	0.1	0.2	1
31	nicotine vaping & healthy lifestyle	0	0	0	0.1	1
32	nicotine vaping & healthy lifestyle	0.9	0.1	0.2	0.1	0.8
33	nicotine vaping & healthy lifestyle	0.9	0.5	0	0.1	0.8

34	healthy lifestyle	0	0	0	0	1
35	nicotine vaping & healthy lifestyle	1	0.1	0.1	0.2	0.5
36	healthy lifestyle	0	0	0.8	0.1	0.7
37	healthy lifestyle	0	0	0	0	1
38	healthy lifestyle	0	0	0	0	1
39	nicotine vaping & healthy lifestyle	1	0	0	0	1
40	healthy lifestyle	0	0	0.1	0	1
41	nicotine vaping	0.9	0.3	0	0.1	0
42	nicotine vaping	0.9	0.1	0	0	0
43	nicotine vaping	1	0	0	0	0
44	nicotine vaping	1	0.1	0	0.2	0
45	nicotine vaping	0	0	0	0	0
46	nicotine vaping	1	0	0	0	0
47	nicotine vaping	1	0.1	0	0.2	0
48	nicotine vaping	0.9	0.1	0	0.2	0
49	nicotine vaping	0.9	0.1	0	0.8	0
50	nicotine vaping	0.9	0.1	0.2	0.8	0.1
51	nicotine vaping	1	0.1	0.2	0.1	0.1
52	nicotine vaping	0.9	0.1	0	0	0
53	nicotine vaping	0.8	0.1	0.3	0.2	0.1
54	nicotine vaping	1	0	0	0	0
55	nicotine vaping	0.8	0.7	0.2	0.3	0.1
56	nicotine vaping	0	0	0.1	0.1	0.1
57	nicotine vaping	1	0.1	0	0.2	0
58	nicotine vaping	1	0	0.1	0.1	0
59	nicotine vaping	0.9	0.1	0	0.5	0
60	nicotine vaping	1	0.1	0	0.1	0
61	nicotine vaping	1	0.1	0	0	0
62	nicotine vaping	0.9	0.1	0	0.2	0
63	entertainment	0	0	0	1	0
64	entertainment	0	0	0	1	0
65	entertainment	0	0	0	1	0
66	entertainment	0	0	0	1	0
67	entertainment	0	0	0	1	0
68	entertainment	0	0	0	1	0
69	entertainment	0.1	0.1	0.1	1	0.1
70	entertainment	0	0	0	1	0
71	entertainment	0	0	0	1	0

72	entertainment	0	0	0	1	0
73	entertainment	0	0	0	1	0
74	entertainment	0	0	0	1	0
75	entertainment	0.9	0.1	0	1	0
76	entertainment	0	0	0	1	0
77	entertainment	0	0	0	1	0
78	entertainment	0	0	0.2	1	0
79	entertainment	0	0	0	1	0
80	entertainment	0.05	0.05	0.05	1	0.05
81	entertainment	0	0	0.1	1	0
82	entertainment	0	0	0	1	0
83	cannabis vaping	0.8	0.2	0.1	0.3	0.1
84	cannabis vaping	0	1	0	0	0
85	cannabis vaping	0.9	0.5	0	0.7	0
86	cannabis vaping & healthy lifestyle	0.8	0.7	0	0.2	0.9
87	cannabis vaping	0.1	0.9	0	0.2	0
88	cannabis vaping	0.9	0.1	0	0.2	0
89	cannabis vaping	0	1	0	0	0
90	cannabis vaping	0.1	1	0	0	0
91	cannabis vaping	0	1	0	0	0
92	cannabis vaping	0.1	0.9	0	0	0
93	cannabis vaping	0	1	0	0	0
94	cannabis vaping	0	1	0	0.1	0
95	cannabis vaping	0	1	0	0	0
96	cannabis vaping	0.1	1	0	0	0
97	cannabis vaping	0.1	1	0	0	0
98	cannabis vaping	0.9	0.8	0	0.2	0
99	cannabis vaping	0.8	0.8	0.1	0.6	0.1
100	cannabis vaping	0.8	0.7	0.1	0.2	0.1
101	cannabis vaping	0	1	0	0	0
102	cannabis vaping	0.1	0.9	0	0	0

Video IDs are represented by the numeric values 1 through 102. The presence of the themes is evident from the label in the *True Label* column (e.g., *nicotine vaping & healthy lifestyle* signifies the presence of the nicotine vaping and healthy lifestyle themes in a video).

The GPT-4o accuracy metrics reported in the manuscript were calculated as the proportion of the correct predictions out of the total number of predictions (102) in the 102 videos for each theme (nicotine vaping, cannabis vaping, fashion, entertainment, healthy lifestyle). A predicted probability threshold of above 50% (0.5) was used to evaluate the model's accuracy. Predicted probability was considered correct if a truly present theme in each video was identified with a predicted probability of above 50%, or if a truly absent theme was identified with a predicted probability below or equal to 50%.