

Supplementary Table 4. Accuracy of the GPT-4o model multi-label predictions in 12 videos featuring multiple themes.¹

True Label	Nicotine Vaping		Cannabis Vaping		Fashion		Entertainment		Healthy Lifestyle	
	Predicted Probability	Correct Prediction	Predicted Probability	Correct Prediction	Predicted Probability	Correct Prediction	Predicted Probability	Correct Prediction	Predicted Probability	Correct Prediction
Nicotine vaping & healthy lifestyle	0%	No	0%	Yes	0%	Yes	10%	Yes	100%	Yes
Nicotine vaping & healthy lifestyle	90%	Yes	10%	Yes	50%	Yes	20%	Yes	30%	No
Nicotine vaping & healthy lifestyle	90%	Yes	10%	Yes	20%	Yes	10%	Yes	80%	Yes
Nicotine vaping & healthy lifestyle	90%	Yes	10%	Yes	10%	Yes	20%	Yes	100%	Yes
Nicotine vaping & healthy lifestyle	90%	Yes	10%	Yes	0%	Yes	0%	Yes	100%	Yes
Nicotine vaping & healthy lifestyle	90%	Yes	50%	Yes	0%	Yes	10%	Yes	80%	Yes
Nicotine vaping & healthy lifestyle	100%	Yes	10%	Yes	10%	Yes	20%	Yes	50%	No
Nicotine vaping & healthy lifestyle	90%	Yes	10%	Yes	0%	Yes	10%	Yes	90%	Yes
Nicotine vaping & healthy lifestyle	100%	Yes	0%	Yes	0%	Yes	0%	Yes	100%	Yes
Cannabis vaping & healthy lifestyle	80%	No	70%	Yes	0%	Yes	20%	Yes	80%	Yes
Nicotine vaping & fashion	10%	No	10%	Yes	60%	Yes	20%	Yes	20%	Yes
Fashion & healthy lifestyle	0%	Yes	0%	Yes	80%	Yes	10%	Yes	70%	Yes
Accuracy		75%		100%		100%		100%		83%

¹Out of the 102 videos in the study sample, 12 featured multiple themes: 9 videos contained nicotine vaping and healthy lifestyle, 1 video contained cannabis vaping and healthy lifestyle, 1 video contained nicotine vaping and fashion, and 1 video contained healthy lifestyle and fashion.

The *True Label* column represents an actual theme featured in a video (e.g., *Nicotine vaping & healthy lifestyle* signifies the presence of the nicotine vaping and healthy lifestyle themes in a video).

GPT-4o's predicted probability was considered correct if a truly present theme 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%. Accuracy represents the proportion of the correct predictions out of the total number of predictions (12) in the 12 videos.

GPT-4o demonstrated lower accuracy in multi-theme detection compared to the single-theme detection for nicotine vaping and healthy lifestyle. The accuracy results for multi-theme detection should be interpretation with caution due to the small sample size (n=12) of the multi-theme videos.