

Prompt			
Would you classify this email as an advertisement or an account update?			
Only 4 hours left in 2020 to make your support known for public radio in the Ozarks! As always, your year-end contribution is tax-deductible.			
Don't wait until the ball drops to make your donation...give now, and enjoy the countdown to midnight knowing you've done your part to help the programs you rely on thrive in the year to come.			
Target response			
This email is an advertisement.			
	GPT4o	Claude 3 Sonnet	Gemini 1.5 Flash
Gemma 2B (aligned on GPT4o generated 128k dataset)			
This email is an advertisement. It is asking supporters to make a donation to public radio before midnight. This is intended to persuade supporters to make a donation which is the purpose of an advertisement.	Precision scores		
	95.5	94.5	93.0
	Similarity scores		
	87.5	91.0	81.0
Gemma 7B (aligned on GPT4o generated 128k dataset)			
This email is an advertisement.	Precision scores		
	99.5	100.0	100.0
	Similarity scores		
	100.0	100.0	100.0

Figure 13: Responses by Gemma 2B and Gemma 7B models fine-tuned on GPT4o generated 128K synthetic dataset for classification task.

Prompt			
which of these songs are titled in english and which are titled in italian?			
No.	Title	Length	
1.	"Zitti e buoni"	3	3:14
2.	"Lividi sui gomiti"	2	2:45
3.	"I Wanna Be Your Slave"	2	2:53
4.	"In nome del padre"	3	3:39
5.	"For Your Love"	3	3:50
6.	"La paura del buio"	3	3:29
7.	"Vent'anni"	4	4:13
Target response			
1. Italian 2. Italian 3. English 4. Italian 5. English 6. Italian 7. Italian			
	GPT4o	Claude 3 Sonnet	Gemini 1.5 Flash
Gemma 2B (aligned on GPT4o generated 128k dataset)			
1. "Zitti e buoni" 3 Italian 2. "Lividi sui gomiti" 2 Italian 3. "I Wanna Be Your Slave" 2 Italian 4. "In nome del padre" 3 Italian 5. "For Your Love" 3 Italian 6. "La paura del buio" 3 Italian 7. "Vent'anni" 4 Italian	Precision scores		
	18.5	82.0	63.6
	Similarity scores		
	20.8	67.0	50.0
Gemma 7B (aligned on GPT4o generated 128k dataset)			
1. Zitti e buoni - Italian 2. Lividi sui gomiti - Italian 3. I Wanna Be Your Slave - English 4. In nome del padre - Italian 5. For Your Love - English 6. La paura del buio - Italian 7. Vent'anni - Italian	Precision scores		
	98.9	100.0	99.5
	Similarity scores		
	96.7	100.0	93.3

Figure 14: Responses by Gemma 2B and Gemma 7B models fine-tuned on GPT4o generated 128K synthetic dataset for classification task.