

Table 14 # of Functionally Diverse Responses by Model, Sampling Strategy, and Task Category.

Model	Sampling Strategy	A	B	C	D	E	F	G	H
gpt-4o	Temperature (t=0.0)	1.57 (0.11)	1.19 (0.10)	1.86 (0.31)	1.33 (0.09)	1.22 (0.07)	1.13 (0.10)	1.18 (0.07)	1.08 (0.03)
gpt-4o	Temperature (t=1.0)	2.11 (0.19)	1.50 (0.18)	3.14 (0.46)	1.27 (0.10)	1.30 (0.10)	1.13 (0.13)	1.40 (0.13)	1.19 (0.06)
gpt-4o	Temperature (t=2.0)	2.38 (0.21)	1.69 (0.25)	3.14 (0.42)	1.31 (0.10)	1.32 (0.10)	1.09 (0.09)	1.89 (0.19)	1.24 (0.06)
gpt-4o	In-Context Regeneration (General)	2.19 (0.22)	4.94 (0.06)	5.00 (0.00)	1.24 (0.10)	1.84 (0.16)	1.35 (0.24)	2.44 (0.26)	2.84 (0.20)
gpt-4o	In-Context Regeneration (Task-Anchored)	1.06 (0.03)	5.00 (0.00)	5.00 (0.00)	1.31 (0.07)	2.82 (0.22)	1.74 (0.28)	3.31 (0.25)	3.20 (0.20)
gpt-4o	System Prompt (General)	1.94 (0.22)	5.00 (0.00)	5.00 (0.00)	1.10 (0.06)	3.80 (0.19)	2.30 (0.39)	2.96 (0.27)	3.65 (0.19)
gpt-4o	System Prompt (Task-Anchored)	1.00 (0.00)	5.00 (0.00)	5.00 (0.00)	1.69 (0.13)	4.60 (0.11)	2.95 (0.37)	4.13 (0.20)	4.20 (0.15)
claude-4-sonnet	Temperature (t=0.0)	1.06 (0.03)	1.00 (0.00)	1.29 (0.22)	1.02 (0.02)	1.40 (0.11)	1.04 (0.04)	1.13 (0.06)	1.10 (0.04)
claude-4-sonnet	Temperature (t=0.5)	1.21 (0.08)	1.06 (0.06)	1.64 (0.32)	1.00 (0.00)	1.36 (0.10)	1.17 (0.10)	1.24 (0.11)	1.14 (0.05)
claude-4-sonnet	Temperature (t=1.0)	1.28 (0.08)	1.19 (0.19)	2.14 (0.35)	1.04 (0.03)	1.40 (0.10)	1.13 (0.07)	1.47 (0.12)	1.25 (0.07)
claude-4-sonnet	In-Context Regeneration (General)	1.60 (0.13)	5.00 (0.00)	4.93 (0.07)	1.38 (0.11)	4.02 (0.19)	1.57 (0.29)	2.98 (0.27)	2.98 (0.20)
claude-4-sonnet	In-Context Regeneration (Task-Anchored)	1.08 (0.08)	5.00 (0.00)	5.00 (0.00)	2.49 (0.18)	4.94 (0.03)	3.13 (0.30)	4.62 (0.15)	4.11 (0.14)
claude-4-sonnet	System Prompt (General)	1.17 (0.10)	5.00 (0.00)	5.00 (0.00)	1.31 (0.11)	4.40 (0.14)	2.00 (0.32)	3.05 (0.27)	3.50 (0.19)
claude-4-sonnet	System Prompt (Task-Anchored)	1.08 (0.08)	5.00 (0.00)	5.00 (0.00)	1.85 (0.16)	4.86 (0.06)	2.61 (0.33)	4.20 (0.19)	4.32 (0.13)
gemini-2.5-flash	Temperature (t=0.0)	1.25 (0.07)	1.12 (0.09)	1.57 (0.17)	1.02 (0.02)	1.26 (0.06)	1.09 (0.06)	1.13 (0.05)	1.07 (0.03)
gemini-2.5-flash	Temperature (t=1.0)	2.55 (0.21)	1.81 (0.26)	3.21 (0.41)	1.13 (0.06)	2.04 (0.17)	1.22 (0.11)	1.87 (0.21)	1.40 (0.10)
gemini-2.5-flash	Temperature (t=2.0)	2.68 (0.20)	1.88 (0.20)	3.29 (0.41)	1.05 (0.04)	1.84 (0.14)	1.39 (0.14)	2.22 (0.23)	1.61 (0.12)
gemini-2.5-flash	In-Context Regeneration (General)	1.30 (0.12)	4.94 (0.06)	5.00 (0.00)	1.35 (0.11)	1.54 (0.13)	1.22 (0.18)	2.31 (0.25)	2.85 (0.20)
gemini-2.5-flash	In-Context Regeneration (Task-Anchored)	1.06 (0.06)	5.00 (0.00)	5.00 (0.00)	1.49 (0.10)	3.04 (0.19)	1.61 (0.25)	3.64 (0.23)	3.25 (0.19)
gemini-2.5-flash	System Prompt (General)	1.15 (0.11)	4.94 (0.06)	5.00 (0.00)	1.20 (0.08)	2.68 (0.21)	1.91 (0.32)	2.67 (0.27)	3.50 (0.19)
gemini-2.5-flash	System Prompt (Task-Anchored)	1.06 (0.06)	4.94 (0.06)	5.00 (0.00)	1.62 (0.15)	4.08 (0.17)	2.96 (0.39)	4.22 (0.17)	4.00 (0.16)

Table 15 # of Functionally Diverse Responses by Model, Sampling Strategy, and Task Category.

(Continued from Table 14)

Model	Sampling Strategy	A	B	C	D	E	F	G	H
Llama-3.1-8B-Instruct	Temperature (t=0.0)	1.09 (0.05)	1.25 (0.14)	1.71 (0.27)	2.40 (0.21)	2.58 (0.20)	1.48 (0.22)	1.47 (0.15)	1.11 (0.04)
Llama-3.1-8B-Instruct	Temperature (t=0.5)	1.19 (0.09)	1.81 (0.26)	2.79 (0.47)	2.75 (0.20)	3.46 (0.20)	1.70 (0.29)	1.89 (0.21)	1.49 (0.10)
Llama-3.1-8B-Instruct	Temperature (t=1.0)	1.45 (0.13)	2.38 (0.30)	3.71 (0.37)	2.89 (0.21)	4.20 (0.17)	2.17 (0.31)	2.22 (0.22)	1.75 (0.12)
Llama-3.1-8B-Instruct	In-Context Regeneration (General)	2.02 (0.18)	4.94 (0.06)	5.00 (0.00)	2.22 (0.19)	3.00 (0.21)	2.22 (0.35)	2.80 (0.25)	2.91 (0.20)
Llama-3.1-8B-Instruct	In-Context Regeneration (Task-Anchored)	1.55 (0.16)	3.94 (0.44)	4.71 (0.29)	3.11 (0.21)	4.56 (0.11)	3.43 (0.36)	4.16 (0.21)	3.72 (0.17)
Llama-3.1-8B-Instruct	System Prompt (General)	1.59 (0.17)	4.94 (0.06)	5.00 (0.00)	1.92 (0.16)	4.56 (0.12)	2.91 (0.43)	3.11 (0.25)	3.98 (0.15)
Llama-3.1-8B-Instruct	System Prompt (Task-Anchored)	1.20 (0.10)	4.25 (0.40)	5.00 (0.00)	2.75 (0.19)	4.70 (0.10)	4.13 (0.30)	3.79 (0.21)	4.35 (0.12)
Mistral-7B-Instruct-v0.3	Temperature (t=0.0)	1.81 (0.16)	1.31 (0.20)	1.71 (0.22)	3.15 (0.21)	2.80 (0.20)	1.57 (0.27)	1.53 (0.14)	1.20 (0.06)
Mistral-7B-Instruct-v0.3	Temperature (t=0.5)	2.83 (0.20)	2.00 (0.27)	3.00 (0.39)	4.05 (0.18)	3.42 (0.21)	2.04 (0.32)	2.00 (0.22)	1.28 (0.07)
Mistral-7B-Instruct-v0.3	Temperature (t=1.0)	3.68 (0.20)	2.69 (0.31)	3.43 (0.31)	4.44 (0.15)	3.98 (0.17)	2.13 (0.33)	2.51 (0.23)	1.52 (0.11)
Mistral-7B-Instruct-v0.3	In-Context Regeneration (General)	1.15 (0.08)	4.88 (0.12)	4.79 (0.21)	1.49 (0.10)	1.30 (0.09)	1.48 (0.25)	1.71 (0.19)	2.62 (0.20)
Mistral-7B-Instruct-v0.3	In-Context Regeneration (Task-Anchored)	1.57 (0.17)	4.31 (0.37)	4.64 (0.23)	1.76 (0.14)	3.00 (0.20)	2.30 (0.30)	2.04 (0.19)	2.68 (0.19)
Mistral-7B-Instruct-v0.3	System Prompt (General)	3.83 (0.24)	4.88 (0.12)	4.64 (0.29)	2.82 (0.22)	4.52 (0.12)	3.09 (0.40)	2.96 (0.28)	3.92 (0.18)
Mistral-7B-Instruct-v0.3	System Prompt (Task-Anchored)	3.08 (0.26)	4.94 (0.06)	5.00 (0.00)	3.00 (0.22)	4.88 (0.07)	4.00 (0.35)	3.69 (0.23)	4.42 (0.12)

Table 16 Vocabulary Diversity by Model, Sampling Strategy, and Task Category.

Model	Sampling Strategy	A	B	C	D	E	F	G	H
gpt-4o	Temperature (t=0.0)	0.29 (0.02)	0.31 (0.05)	0.26 (0.07)	0.41 (0.02)	0.53 (0.01)	0.53 (0.03)	0.51 (0.04)	0.48 (0.02)
gpt-4o	Temperature (t=1.0)	0.49 (0.03)	0.52 (0.06)	0.68 (0.07)	0.63 (0.01)	0.72 (0.00)	0.72 (0.01)	0.75 (0.02)	0.71 (0.01)
gpt-4o	Temperature (t=2.0)	0.54 (0.02)	0.58 (0.04)	0.71 (0.05)	0.66 (0.01)	0.75 (0.00)	0.75 (0.01)	0.79 (0.02)	0.74 (0.01)
gpt-4o	In-Context Regeneration (General)	0.61 (0.03)	0.96 (0.03)	0.95 (0.03)	0.52 (0.02)	0.72 (0.01)	0.75 (0.01)	0.84 (0.02)	0.83 (0.01)
gpt-4o	In-Context Regeneration (Task-Anchored)	0.53 (0.02)	0.96 (0.03)	0.96 (0.03)	0.53 (0.01)	0.72 (0.01)	0.76 (0.02)	0.88 (0.01)	0.86 (0.01)
gpt-4o	System Prompt (General)	0.68 (0.02)	0.78 (0.04)	0.77 (0.08)	0.55 (0.01)	0.78 (0.01)	0.82 (0.01)	0.86 (0.01)	0.85 (0.00)
gpt-4o	System Prompt (Task-Anchored)	0.51 (0.02)	0.81 (0.05)	0.75 (0.08)	0.65 (0.02)	0.79 (0.01)	0.80 (0.03)	0.87 (0.01)	0.86 (0.01)
claude-4-sonnet	Temperature (t=0.0)	0.20 (0.03)	0.07 (0.04)	0.16 (0.06)	0.28 (0.02)	0.53 (0.02)	0.43 (0.04)	0.48 (0.04)	0.42 (0.02)
claude-4-sonnet	Temperature (t=0.5)	0.35 (0.02)	0.17 (0.05)	0.29 (0.07)	0.35 (0.03)	0.63 (0.01)	0.57 (0.04)	0.62 (0.03)	0.56 (0.02)
claude-4-sonnet	Temperature (t=1.0)	0.45 (0.02)	0.23 (0.05)	0.46 (0.07)	0.42 (0.02)	0.68 (0.01)	0.63 (0.04)	0.68 (0.03)	0.63 (0.02)
claude-4-sonnet	In-Context Regeneration (General)	0.74 (0.01)	0.78 (0.05)	0.92 (0.02)	0.73 (0.01)	0.78 (0.00)	0.83 (0.01)	0.87 (0.01)	0.86 (0.01)
claude-4-sonnet	In-Context Regeneration (Task-Anchored)	0.62 (0.01)	0.75 (0.06)	0.87 (0.06)	0.76 (0.01)	0.79 (0.01)	0.82 (0.01)	0.92 (0.01)	0.87 (0.01)
claude-4-sonnet	System Prompt (General)	0.71 (0.01)	0.81 (0.01)	0.81 (0.04)	0.70 (0.01)	0.78 (0.01)	0.80 (0.01)	0.86 (0.01)	0.85 (0.00)
claude-4-sonnet	System Prompt (Task-Anchored)	0.68 (0.01)	0.81 (0.01)	0.81 (0.06)	0.75 (0.01)	0.81 (0.00)	0.82 (0.01)	0.90 (0.01)	0.86 (0.00)
gemini-2.5-flash	Temperature (t=0.0)	0.07 (0.02)	0.12 (0.05)	0.25 (0.07)	0.20 (0.02)	0.35 (0.02)	0.34 (0.03)	0.31 (0.03)	0.29 (0.02)
gemini-2.5-flash	Temperature (t=1.0)	0.38 (0.04)	0.44 (0.08)	0.70 (0.07)	0.52 (0.02)	0.76 (0.00)	0.74 (0.02)	0.76 (0.03)	0.71 (0.02)
gemini-2.5-flash	Temperature (t=2.0)	0.44 (0.04)	0.44 (0.07)	0.77 (0.04)	0.56 (0.02)	0.77 (0.00)	0.78 (0.01)	0.79 (0.03)	0.77 (0.01)
gemini-2.5-flash	In-Context Regeneration (General)	0.76 (0.02)	0.98 (0.02)	0.89 (0.04)	0.63 (0.01)	0.69 (0.01)	0.74 (0.02)	0.80 (0.03)	0.83 (0.01)
gemini-2.5-flash	In-Context Regeneration (Task-Anchored)	0.60 (0.02)	0.95 (0.03)	0.91 (0.04)	0.69 (0.01)	0.71 (0.01)	0.77 (0.02)	0.89 (0.01)	0.87 (0.01)
gemini-2.5-flash	System Prompt (General)	0.67 (0.01)	0.85 (0.05)	0.86 (0.03)	0.66 (0.01)	0.74 (0.01)	0.81 (0.01)	0.85 (0.01)	0.85 (0.00)
gemini-2.5-flash	System Prompt (Task-Anchored)	0.57 (0.02)	0.92 (0.05)	0.73 (0.07)	0.71 (0.01)	0.77 (0.01)	0.83 (0.01)	0.89 (0.01)	0.87 (0.00)