Table 35 Checklist-Based Quality by Model, Sampling Strategy, and Task Category.

(Based on $n = 10$ generated responses. Using Only GPT-4o as the Checklist-Based Quality Judge.)

Model	Sampling Strategy	A	B	C	D	E	F	G	H
gpt-4o	In-Context Regeneration (General)	3.21 (0.14)	4.63 (0.19)	4.26 (0.29)	3.89 (0.15)	4.61 (0.07)	4.40 (0.19)	4.92 (0.02)	4.33 (0.10)
gpt-4o	In-Context Regeneration (Task-Anchored)	3.47 (0.16)	4.44 (0.25)	4.23 (0.28)	4.27 (0.12)	4.55 (0.09)	4.22 (0.19)	4.78 (0.06)	4.31 (0.09)
gpt-4o	System Prompt (General)	3.62 (0.16)	4.59 (0.18)	4.22 (0.29)	3.60 (0.19)	3.82 (0.10)	3.77 (0.20)	4.64 (0.09)	4.35 (0.08)
gpt-4o	System Prompt (Task-Anchored)	3.45 (0.16)	4.53 (0.21)	4.19 (0.29)	3.33 (0.16)	3.71 (0.11)	3.23 (0.18)	4.56 (0.11)	4.06 (0.10)
claude-4-sonnet	In-Context Regeneration (General)	3.32 (0.14)	4.54 (0.19)	4.17 (0.33)	4.24 (0.13)	4.46 (0.09)	4.43 (0.18)	4.78 (0.10)	4.75 (0.04)
claude-4-sonnet	In-Context Regeneration (Task-Anchored)	3.30 (0.14)	4.48 (0.19)	4.35 (0.24)	3.71 (0.13)	4.12 (0.09)	4.20 (0.18)	4.39 (0.12)	4.46 (0.07)
claude-4-sonnet	System Prompt (General)	3.32 (0.17)	4.64 (0.12)	4.34 (0.24)	4.09 (0.17)	4.27 (0.09)	4.16 (0.18)	4.62 (0.13)	4.56 (0.07)
claude-4-sonnet	System Prompt (Task-Anchored)	3.30 (0.18)	4.58 (0.10)	4.38 (0.26)	3.77 (0.16)	3.93 (0.09)	3.83 (0.19)	4.30 (0.19)	4.33 (0.08)
gemini-2.5-flash	In-Context Regeneration (General)	2.88 (0.14)	4.57 (0.18)	4.31 (0.26)	4.13 (0.14)	4.84 (0.06)	4.50 (0.19)	4.74 (0.10)	4.38 (0.09)
gemini-2.5-flash	In-Context Regeneration (Task-Anchored)	2.92 (0.15)	4.54 (0.19)	4.19 (0.24)	3.83 (0.14)	4.76 (0.06)	4.30 (0.22)	4.56 (0.10)	4.41 (0.08)
gemini-2.5-flash	System Prompt (General)	3.16 (0.16)	4.67 (0.12)	4.33 (0.23)	4.35 (0.12)	4.49 (0.08)	3.94 (0.22)	4.78 (0.07)	4.40 (0.07)
gemini-2.5-flash	System Prompt (Task-Anchored)	2.91 (0.15)	4.67 (0.19)	4.13 (0.24)	4.27 (0.13)	4.53 (0.08)	3.34 (0.20)	4.55 (0.09)	4.20 (0.09)

Table 36 # of Functionally Diverse Responses by Model, Sampling Strategy, and Task Category.

(Using Only GPT-4o as the Functional Diversity Judge)

Model	Sampling Strategy	A	B	C	D	E	F	G	H
Llama-3.1-8B-Instruct	Temperature (t=0.1)	1.08 (0.05)	1.25 (0.14)	1.71 (0.27)	1.93 (0.18)	1.56 (0.14)	1.52 (0.23)	1.47 (0.15)	1.09 (0.04)
Llama-3.1-8B-Instruct	Temperature (t=0.5)	1.19 (0.09)	1.75 (0.25)	2.79 (0.47)	2.22 (0.18)	2.02 (0.17)	1.48 (0.24)	1.91 (0.21)	1.47 (0.10)
Llama-3.1-8B-Instruct	Temperature (t=1.0)	1.45 (0.13)	2.31 (0.30)	3.71 (0.37)	2.42 (0.20)	2.68 (0.21)	2.00 (0.29)	2.51 (0.24)	1.78 (0.13)
Online DPO (Wildchat, $\beta = 0.01$, Step 1000)	Temperature (t=0.1)	1.72 (0.16)	1.38 (0.20)	2.00 (0.33)	1.42 (0.11)	1.66 (0.12)	1.35 (0.18)	1.47 (0.11)	1.49 (0.09)
Online DPO (Wildchat, $\beta = 0.01$, Step 1000)	Temperature (t=0.5)	3.09 (0.21)	2.31 (0.30)	4.00 (0.42)	2.11 (0.18)	3.26 (0.18)	2.22 (0.31)	2.53 (0.20)	1.93 (0.14)
Online DPO (Wildchat, $\beta = 0.01$, Step 1000)	Temperature (t=1.0)	3.43 (0.21)	2.38 (0.30)	4.71 (0.16)	2.40 (0.19)	3.60 (0.20)	2.70 (0.34)	3.31 (0.22)	2.22 (0.16)
Online DPO (Wildchat, $\beta = 0.1$, Step 1000)	Temperature (t=0.1)	1.19 (0.08)	1.12 (0.09)	1.64 (0.25)	2.13 (0.16)	1.46 (0.09)	1.43 (0.23)	2.11 (0.22)	1.31 (0.07)
Online DPO (Wildchat, $\beta = 0.1$, Step 1000)	Temperature (t=0.5)	1.43 (0.12)	1.69 (0.27)	2.43 (0.42)	2.27 (0.18)	2.38 (0.15)	1.87 (0.26)	2.71 (0.23)	1.43 (0.08)
Online DPO (Wildchat, $\beta = 0.1$, Step 1000)	Temperature (t=1.0)	1.64 (0.15)	2.12 (0.26)	2.86 (0.38)	2.29 (0.18)	2.52 (0.17)	2.13 (0.31)	2.89 (0.23)	1.58 (0.10)
Online DPO (Ultrafeedback, $\beta = 0.01$, Step 1000)	Temperature (t=0.1)	2.43 (0.23)	1.31 (0.20)	1.86 (0.38)	1.05 (0.03)	1.04 (0.03)	1.48 (0.24)	1.73 (0.19)	1.25 (0.09)
Online DPO (Ultrafeedback, $\beta = 0.01$, Step 1000)	Temperature (t=0.5)	2.13 (0.20)	1.38 (0.22)	2.14 (0.43)	1.07 (0.06)	1.00 (0.00)	1.43 (0.23)	1.58 (0.18)	1.24 (0.09)
Online DPO (Ultrafeedback, $\beta = 0.01$, Step 1000)	Temperature (t=1.0)	2.17 (0.20)	1.62 (0.31)	1.86 (0.39)	1.07 (0.06)	1.02 (0.02)	1.52 (0.23)	1.69 (0.20)	1.23 (0.08)
GRPO (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=0.1)	2.79 (0.21)	1.62 (0.24)	2.71 (0.40)	2.84 (0.19)	1.12 (0.05)	1.09 (0.06)	1.22 (0.08)	1.12 (0.04)
GRPO (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=0.5)	3.55 (0.19)	2.31 (0.28)	3.29 (0.34)	3.13 (0.20)	1.18 (0.07)	1.17 (0.08)	1.47 (0.15)	1.27 (0.07)
GRPO (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=1.0)	4.00 (0.16)	2.69 (0.35)	3.93 (0.27)	3.09 (0.20)	1.16 (0.06)	1.04 (0.04)	1.67 (0.17)	1.36 (0.10)
GRPO (Wildchat, $\beta = 0.01$, Step 1000)	Temperature (t=0.1)	1.96 (0.15)	1.88 (0.24)	3.71 (0.29)	2.25 (0.16)	1.26 (0.08)	1.09 (0.06)	1.51 (0.15)	1.26 (0.09)
GRPO (Wildchat, $\beta = 0.01$, Step 1000)	Temperature (t=0.5)	2.19 (0.19)	2.06 (0.30)	3.86 (0.40)	2.29 (0.18)	1.62 (0.12)	1.17 (0.10)	1.71 (0.18)	1.41 (0.10)
GRPO (Wildchat, $\beta = 0.01$, Step 1000)	Temperature (t=1.0)	2.32 (0.19)	2.56 (0.34)	4.29 (0.24)	2.38 (0.18)	1.74 (0.15)	1.35 (0.16)	2.00 (0.20)	1.42 (0.10)
GRPO (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=0.1)	3.15 (0.19)	1.69 (0.24)	3.14 (0.42)	2.25 (0.17)	1.20 (0.06)	1.52 (0.21)	1.96 (0.20)	1.24 (0.06)
GRPO (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=0.5)	3.66 (0.20)	2.56 (0.30)	3.64 (0.32)	2.73 (0.21)	1.40 (0.11)	1.87 (0.30)	2.36 (0.22)	1.45 (0.09)
GRPO (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=1.0)	3.94 (0.18)	2.75 (0.30)	4.29 (0.27)	2.67 (0.20)	1.50 (0.13)	1.83 (0.29)	2.58 (0.23)	1.51 (0.10)

Table 37 # of Functionally Diverse Responses by Model, Sampling Strategy, and Task Category.

(Using Only GPT-4o as the Functional Diversity Judge)

Model	Sampling Strategy	A	B	C	D	E	F	G	H
Llama-3.1-8B-Instruct	Temperature (t=0.1)	1.08 (0.05)	1.25 (0.14)	1.71 (0.27)	1.93 (0.18)	1.56 (0.14)	1.52 (0.23)	1.47 (0.15)	1.09 (0.04)
Llama-3.1-8B-Instruct	Temperature (t=0.5)	1.19 (0.09)	1.75 (0.25)	2.79 (0.47)	2.22 (0.18)	2.02 (0.17)	1.48 (0.24)	1.91 (0.21)	1.47 (0.10)
Llama-3.1-8B-Instruct	Temperature (t=1.0)	1.45 (0.13)	2.31 (0.30)	3.71 (0.37)	2.42 (0.20)	2.68 (0.21)	2.00 (0.29)	2.51 (0.24)	1.78 (0.13)
GRPO (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=0.1)	2.79 (0.21)	1.62 (0.24)	2.71 (0.40)	2.84 (0.19)	1.12 (0.05)	1.09 (0.06)	1.22 (0.08)	1.12 (0.04)
GRPO (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=0.5)	3.55 (0.19)	2.31 (0.28)	3.29 (0.34)	3.13 (0.20)	1.18 (0.07)	1.17 (0.08)	1.47 (0.15)	1.27 (0.07)
GRPO (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=1.0)	4.00 (0.16)	2.69 (0.35)	3.93 (0.27)	3.09 (0.20)	1.16 (0.06)	1.04 (0.04)	1.67 (0.17)	1.36 (0.10)
GRPO (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=0.1)	3.15 (0.19)	1.69 (0.24)	3.14 (0.42)	2.25 (0.17)	1.20 (0.06)	1.52 (0.21)	1.96 (0.20)	1.24 (0.06)
GRPO (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=0.5)	3.66 (0.20)	2.56 (0.30)	3.64 (0.32)	2.73 (0.21)	1.40 (0.11)	1.87 (0.30)	2.36 (0.22)	1.45 (0.09)
GRPO (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=1.0)	3.94 (0.18)	2.75 (0.30)	4.29 (0.27)	2.67 (0.20)	1.50 (0.13)	1.83 (0.29)	2.58 (0.23)	1.51 (0.10)
GRPO w/DARLING (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=0.1)	3.47 (0.19)	1.69 (0.20)	3.14 (0.44)	2.60 (0.20)	1.70 (0.15)	1.83 (0.28)	2.11 (0.23)	1.35 (0.08)
GRPO w/DARLING (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=0.5)	3.94 (0.16)	2.25 (0.25)	3.29 (0.40)	2.75 (0.19)	2.06 (0.14)	2.22 (0.32)	2.82 (0.23)	1.64 (0.12)
GRPO w/DARLING (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=1.0)	4.26 (0.14)	3.19 (0.29)	3.64 (0.41)	2.65 (0.19)	2.52 (0.19)	2.13 (0.32)	3.53 (0.22)	1.89 (0.14)
GRPO w/DARLING (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=0.1)	2.43 (0.18)	1.19 (0.14)	2.29 (0.42)	2.29 (0.18)	1.50 (0.12)	1.26 (0.13)	1.62 (0.16)	1.10 (0.04)
GRPO w/DARLING (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=0.5)	3.42 (0.20)	1.56 (0.27)	2.86 (0.47)	2.38 (0.18)	1.58 (0.13)	1.83 (0.31)	2.13 (0.23)	1.28 (0.07)
GRPO w/DARLING (Ultrafeedback, $\beta = 0.001$, Step 1000)	Temperature (t=1.0)	3.53 (0.20)	1.75 (0.27)	3.14 (0.40)	2.25 (0.19)	1.52 (0.12)	1.83 (0.29)	2.16 (0.22)	1.33 (0.08)
GRPO w/DARLING & Task Diversity Judges (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=0.1)	2.58 (0.19)	1.19 (0.10)	2.21 (0.45)	3.07 (0.19)	3.10 (0.19)	2.70 (0.33)	3.44 (0.23)	2.11 (0.14)
GRPO w/DARLING & Task Diversity Judges (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=0.5)	3.42 (0.19)	1.88 (0.26)	2.43 (0.45)	3.29 (0.20)	3.98 (0.17)	3.43 (0.35)	3.93 (0.22)	2.92 (0.16)
GRPO w/DARLING & Task Diversity Judges (Wildchat, $\beta = 0.001$, Step 1000)	Temperature (t=1.0)	3.89 (0.18)	2.25 (0.32)	3.21 (0.38)	3.47 (0.17)	4.46 (0.13)	4.00 (0.29)	4.36 (0.19)	3.32 (0.16)
GRPO w/DARLING & Task Diversity Judges (Ultrachat, $\beta = 0.001$, Step 1000)	Temperature (t=0.1)	1.87 (0.16)	1.25 (0.11)	1.21 (0.15)	1.95 (0.17)	1.78 (0.17)	2.00 (0.31)	2.22 (0.22)	1.33 (0.08)
GRPO w/DARLING & Task Diversity Judges (Ultrachat, $\beta = 0.001$, Step 1000)	Temperature (t=0.5)	3.23 (0.21)	1.62 (0.27)	2.07 (0.37)	2.44 (0.18)	2.66 (0.18)	2.87 (0.36)	2.89 (0.24)	1.81 (0.12)
GRPO w/DARLING & Task Diversity Judges (Ultrachat, $\beta = 0.001$, Step 1000)	Temperature (t=1.0)	3.68 (0.21)	2.31 (0.28)	3.14 (0.40)	2.65 (0.19)	3.14 (0.21)	3.17 (0.35)	3.84 (0.24)	2.43 (0.16)