

Dataset	Property	Model	Cohen's κ
cola	grammaticality	Meta-Llama-3.1-8B-Instruct	0.420000
cola	grammaticality	Meta-Llama-3.1-70B-Instruct	0.464100
cola	grammaticality	Mistral-7B-Instruct-v0.3	0.428200
cola	grammaticality	OLMo-7B-0724-Instruct-hf	0.423400
cola	grammaticality	Starling-LM-7B-alpha	0.454900
cola	grammaticality	c4ai-command-r-v01	0.006800
cola	grammaticality	gpt-4o	0.340500
cola	grammaticality	gemini-1.5-flash-latest	0.446200
cola	grammaticality	Mixtral-8x7B-Instruct-v0.1	0.551400
cola	grammaticality	Mixtral-8x22B-Instruct-v0.1	0.542600
cola	grammaticality	c4ai-command-r-plus	0.122600
inferential-strategies	Sound Reasoning	Meta-Llama-3.1-8B-Instruct	0.134500
inferential-strategies	Sound Reasoning	Meta-Llama-3.1-70B-Instruct	0.395300
inferential-strategies	Sound Reasoning	Mistral-7B-Instruct-v0.3	0.014500
inferential-strategies	Sound Reasoning	OLMo-7B-0724-Instruct-hf	0.035100
inferential-strategies	Sound Reasoning	Starling-LM-7B-alpha	0.013000
inferential-strategies	Sound Reasoning	c4ai-command-r-v01	-0.118600
inferential-strategies	Sound Reasoning	gemini-1.5-flash-latest	0.224300
inferential-strategies	Sound Reasoning	gpt-4o	0.419000
inferential-strategies	Sound Reasoning	Mixtral-8x7B-Instruct-v0.1	0.059000
inferential-strategies	Sound Reasoning	Mixtral-8x22B-Instruct-v0.1	0.017200
inferential-strategies	Sound Reasoning	c4ai-command-r-plus	-0.016000
qags	Factual Consistency	Meta-Llama-3.1-8B-Instruct	0.581100
qags	Factual Consistency	Meta-Llama-3.1-70B-Instruct	0.696800
qags	Factual Consistency	Mistral-7B-Instruct-v0.3	0.433800
qags	Factual Consistency	OLMo-7B-0724-Instruct-hf	0.112600
qags	Factual Consistency	Starling-LM-7B-alpha	0.015100
qags	Factual Consistency	c4ai-command-r-v01	0.326600
qags	Factual Consistency	gemini-1.5-flash-latest	0.646600
qags	Factual Consistency	gpt-4o	0.722300
qags	Factual Consistency	Mixtral-8x7B-Instruct-v0.1	0.681500
qags	Factual Consistency	Mixtral-8x22B-Instruct-v0.1	0.662600
qags	Factual Consistency	c4ai-command-r-plus	0.134000
persona-chat	understandable	Meta-Llama-3.1-8B-Instruct	1.000000
persona-chat	uses knowledge	Meta-Llama-3.1-8B-Instruct	0.007400
persona-chat	understandable	Meta-Llama-3.1-70B-Instruct	0.000000
persona-chat	uses knowledge	Meta-Llama-3.1-70B-Instruct	0.473700
persona-chat	understandable	Mistral-7B-Instruct-v0.3	1.000000
persona-chat	uses knowledge	Mistral-7B-Instruct-v0.3	-0.066200
persona-chat	understandable	OLMo-7B-0724-Instruct-hf	0.000000
persona-chat	uses knowledge	OLMo-7B-0724-Instruct-hf	0.048500
persona-chat	understandable	Starling-LM-7B-alpha	0.000000

persona-chat	uses knowledge	Starling-LM-7B-alpha	-0.052600
persona-chat	understandable	c4ai-command-r-v01	0.000000
persona-chat	uses knowledge	c4ai-command-r-v01	0.013700
persona-chat	understandable	gpt-4o	0.000000
persona-chat	uses knowledge	gpt-4o	0.482800
persona-chat	understandable	gemini-1.5-flash-latest	0.000000
persona-chat	uses knowledge	gemini-1.5-flash-latest	-0.061300
persona-chat	understandable	Mixtral-8x7B-Instruct-v0.1	1.000000
persona-chat	uses knowledge	Mixtral-8x7B-Instruct-v0.1	0.080200
persona-chat	understandable	Mixtral-8x22B-Instruct-v0.1	1.000000
persona-chat	uses knowledge	Mixtral-8x22B-Instruct-v0.1	0.160400
persona-chat	understandable	c4ai-command-r-plus	1.000000
persona-chat	uses knowledge	c4ai-command-r-plus	-0.041700
<hr/>			
roscoe-cosmos	Missing Steps	Meta-Llama-3.1-8B-Instruct	0.000000
roscoe-cosmos	Contradiction	Meta-Llama-3.1-8B-Instruct	0.156900
roscoe-cosmos	Missing Steps	Meta-Llama-3.1-70B-Instruct	0.259200
roscoe-cosmos	Contradiction	Meta-Llama-3.1-70B-Instruct	0.232300
roscoe-cosmos	Missing Steps	Mistral-7B-Instruct-v0.3	0.269900
roscoe-cosmos	Contradiction	Mistral-7B-Instruct-v0.3	0.319200
roscoe-cosmos	Missing Steps	OLMo-7B-0724-Instruct-hf	-0.176800
roscoe-cosmos	Contradiction	OLMo-7B-0724-Instruct-hf	-0.183800
roscoe-cosmos	Missing Steps	Starling-LM-7B-alpha	0.028400
roscoe-cosmos	Contradiction	Starling-LM-7B-alpha	0.027400
roscoe-cosmos	Missing Steps	c4ai-command-r-v01	0.002300
roscoe-cosmos	Contradiction	c4ai-command-r-v01	-0.020000
roscoe-cosmos	Missing Steps	gemini-1.5-flash-latest	0.012700
roscoe-cosmos	Contradiction	gemini-1.5-flash-latest	0.258700
roscoe-cosmos	Missing Steps	gpt-4o	0.107400
roscoe-cosmos	Contradiction	gpt-4o	0.210400
roscoe-cosmos	Missing Steps	Mixtral-8x7B-Instruct-v0.1	0.220600
roscoe-cosmos	Contradiction	Mixtral-8x7B-Instruct-v0.1	0.153800
roscoe-cosmos	Missing Steps	Mixtral-8x22B-Instruct-v0.1	-0.025600
roscoe-cosmos	Contradiction	Mixtral-8x22B-Instruct-v0.1	0.209000
roscoe-cosmos	Missing Steps	c4ai-command-r-plus	-0.030800
roscoe-cosmos	Contradiction	c4ai-command-r-plus	-0.020000
<hr/>			
roscoe-drop	Missing Steps	Meta-Llama-3.1-8B-Instruct	0.000000
roscoe-drop	Contradiction	Meta-Llama-3.1-8B-Instruct	0.033500
roscoe-drop	Missing Steps	Meta-Llama-3.1-70B-Instruct	0.319700
roscoe-drop	Contradiction	Meta-Llama-3.1-70B-Instruct	0.218300
roscoe-drop	Missing Steps	Mistral-7B-Instruct-v0.3	0.147300
roscoe-drop	Contradiction	Mistral-7B-Instruct-v0.3	0.041100
roscoe-drop	Missing Steps	OLMo-7B-0724-Instruct-hf	0.007500
roscoe-drop	Contradiction	OLMo-7B-0724-Instruct-hf	-0.004200
roscoe-drop	Missing Steps	Starling-LM-7B-alpha	0.031300
roscoe-drop	Contradiction	Starling-LM-7B-alpha	-0.012100

roscoe-drop	Missing Steps	c4ai-command-r-v01	0.065300
roscoe-drop	Contradiction	c4ai-command-r-v01	-0.034800
roscoe-drop	Missing Steps	gpt-4o	0.349100
roscoe-drop	Contradiction	gpt-4o	0.231100
roscoe-drop	Missing Steps	gemini-1.5-flash-latest	0.052000
roscoe-drop	Contradiction	gemini-1.5-flash-latest	0.117600
roscoe-drop	Missing Steps	Mixtral-8x7B-Instruct-v0.1	0.278800
roscoe-drop	Contradiction	Mixtral-8x7B-Instruct-v0.1	-0.016300
roscoe-drop	Missing Steps	Mixtral-8x22B-Instruct-v0.1	0.280200
roscoe-drop	Contradiction	Mixtral-8x22B-Instruct-v0.1	0.110000
roscoe-drop	Missing Steps	c4ai-command-r-plus	0.058900
roscoe-drop	Contradiction	c4ai-command-r-plus	0.000000
roscoe-esnli	Missing Steps	Meta-Llama-3.1-8B-Instruct	0.000000
roscoe-esnli	Contradiction	Meta-Llama-3.1-8B-Instruct	0.280700
roscoe-esnli	Missing Steps	Meta-Llama-3.1-70B-Instruct	0.439000
roscoe-esnli	Contradiction	Meta-Llama-3.1-70B-Instruct	0.329300
roscoe-esnli	Missing Steps	Mistral-7B-Instruct-v0.3	-0.044100
roscoe-esnli	Contradiction	Mistral-7B-Instruct-v0.3	0.076200
roscoe-esnli	Missing Steps	OLMo-7B-0724-Instruct-hf	-0.106900
roscoe-esnli	Contradiction	OLMo-7B-0724-Instruct-hf	0.023100
roscoe-esnli	Missing Steps	Starling-LM-7B-alpha	0.052500
roscoe-esnli	Contradiction	Starling-LM-7B-alpha	-0.042100
roscoe-esnli	Missing Steps	c4ai-command-r-v01	-0.013100
roscoe-esnli	Contradiction	c4ai-command-r-v01	0.000000
roscoe-esnli	Missing Steps	gpt-4o	0.329300
roscoe-esnli	Contradiction	gpt-4o	0.244600
roscoe-esnli	Missing Steps	Mixtral-8x7B-Instruct-v0.1	0.018700
roscoe-esnli	Contradiction	Mixtral-8x7B-Instruct-v0.1	0.172600
roscoe-esnli	Missing Steps	gemini-1.5-flash-latest	-0.021800
roscoe-esnli	Contradiction	gemini-1.5-flash-latest	0.233500
roscoe-esnli	Missing Steps	Mixtral-8x22B-Instruct-v0.1	0.044200
roscoe-esnli	Contradiction	Mixtral-8x22B-Instruct-v0.1	0.223800
roscoe-esnli	Missing Steps	c4ai-command-r-plus	0.069500
roscoe-esnli	Contradiction	c4ai-command-r-plus	0.000000
roscoe-gsm8k	Missing Steps	Meta-Llama-3.1-8B-Instruct	0.580000
roscoe-gsm8k	Contradiction	Meta-Llama-3.1-8B-Instruct	0.145500
roscoe-gsm8k	Missing Steps	Meta-Llama-3.1-70B-Instruct	0.834000
roscoe-gsm8k	Contradiction	Meta-Llama-3.1-70B-Instruct	0.453200
roscoe-gsm8k	Missing Steps	Mistral-7B-Instruct-v0.3	0.704900
roscoe-gsm8k	Contradiction	Mistral-7B-Instruct-v0.3	0.226100
roscoe-gsm8k	Missing Steps	OLMo-7B-0724-Instruct-hf	-0.026600
roscoe-gsm8k	Contradiction	OLMo-7B-0724-Instruct-hf	0.007400
roscoe-gsm8k	Missing Steps	Starling-LM-7B-alpha	-0.025600
roscoe-gsm8k	Contradiction	Starling-LM-7B-alpha	-0.036700
roscoe-gsm8k	Missing Steps	c4ai-command-r-v01	0.233700

roscoe-gsm8k	Contradiction	c4ai-command-r-v01	0.194200
roscoe-gsm8k	Missing Steps	gpt-4o	0.839100
roscoe-gsm8k	Contradiction	gpt-4o	0.342800
roscoe-gsm8k	Missing Steps	Mixtral-8x7B-Instruct-v0.1	0.835900
roscoe-gsm8k	Contradiction	Mixtral-8x7B-Instruct-v0.1	0.329100
roscoe-gsm8k	Missing Steps	gemini-1.5-flash-latest	0.771700
roscoe-gsm8k	Contradiction	gemini-1.5-flash-latest	0.433700
roscoe-gsm8k	Missing Steps	Mixtral-8x22B-Instruct-v0.1	0.888300
roscoe-gsm8k	Contradiction	Mixtral-8x22B-Instruct-v0.1	0.344800
roscoe-gsm8k	Missing Steps	c4ai-command-r-plus	0.000000
roscoe-gsm8k	Contradiction	c4ai-command-r-plus	0.000000
<hr/>			
topical-chat	understandable	Meta-Llama-3.1-8B-Instruct	1.000000
topical-chat	uses knowledge	Meta-Llama-3.1-8B-Instruct	0.134600
topical-chat	understandable	Meta-Llama-3.1-70B-Instruct	0.000000
topical-chat	uses knowledge	Meta-Llama-3.1-70B-Instruct	-0.034500
topical-chat	understandable	Mistral-7B-Instruct-v0.3	0.000000
topical-chat	uses knowledge	Mistral-7B-Instruct-v0.3	-0.066900
topical-chat	understandable	OLMo-7B-0724-Instruct-hf	0.000000
topical-chat	uses knowledge	OLMo-7B-0724-Instruct-hf	0.055100
topical-chat	understandable	Starling-LM-7B-alpha	0.000000
topical-chat	uses knowledge	Starling-LM-7B-alpha	0.080200
topical-chat	understandable	c4ai-command-r-v01	0.000000
topical-chat	uses knowledge	c4ai-command-r-v01	0.012100
topical-chat	understandable	gemini-1.5-flash-latest	0.000000
topical-chat	uses knowledge	gemini-1.5-flash-latest	-0.061900
topical-chat	understandable	gpt-4o	0.000000
topical-chat	uses knowledge	gpt-4o	0.103300
topical-chat	understandable	Mixtral-8x7B-Instruct-v0.1	0.000000
topical-chat	uses knowledge	Mixtral-8x7B-Instruct-v0.1	0.037200
topical-chat	understandable	Mixtral-8x22B-Instruct-v0.1	0.000000
topical-chat	uses knowledge	Mixtral-8x22B-Instruct-v0.1	-0.059800
topical-chat	understandable	c4ai-command-r-plus	0.000000
topical-chat	uses knowledge	c4ai-command-r-plus	0.027800
<hr/>			
medical-safety	query risk level	Meta-Llama-3.1-8B-Instruct	0.006000
medical-safety	response type	Meta-Llama-3.1-8B-Instruct	0.009700
medical-safety	query risk level	Meta-Llama-3.1-70B-Instruct	0.020000
medical-safety	response type	Meta-Llama-3.1-70B-Instruct	-0.070700
medical-safety	query risk level	Mistral-7B-Instruct-v0.3	0.054600
medical-safety	response type	Mistral-7B-Instruct-v0.3	-0.120200
medical-safety	query risk level	OLMo-7B-0724-Instruct-hf	0.028700
medical-safety	response type	OLMo-7B-0724-Instruct-hf	-0.073700
medical-safety	query risk level	Starling-LM-7B-alpha	0.017500
medical-safety	response type	Starling-LM-7B-alpha	-0.008800
medical-safety	query risk level	c4ai-command-r-v01	0.020400
medical-safety	response type	c4ai-command-r-v01	-0.000100

medical-safety	query risk level	Mixtral-8x7B-Instruct-v0.1	0.045200
medical-safety	response type	Mixtral-8x7B-Instruct-v0.1	-0.039400
medical-safety	query risk level	gemini-1.5-flash-latest	0.024900
medical-safety	response type	gemini-1.5-flash-latest	-0.090900
medical-safety	query risk level	gpt-4o	0.034900
medical-safety	response type	gpt-4o	-0.010400
medical-safety	query risk level	Mixtral-8x22B-Instruct-v0.1	0.048900
medical-safety	response type	Mixtral-8x22B-Instruct-v0.1	-0.082900
medical-safety	query risk level	c4ai-command-r-plus	0.016800
medical-safety	response type	c4ai-command-r-plus	-0.004700
cola-grammar	Simple	Meta-Llama-3.1-8B-Instruct	0.064400
cola-grammar	Copula	Meta-Llama-3.1-8B-Instruct	0.156600
cola-grammar	Pred/SC	Meta-Llama-3.1-8B-Instruct	0.000900
cola-grammar	Result/Depictive	Meta-Llama-3.1-8B-Instruct	0.029700
cola-grammar	Particle	Meta-Llama-3.1-8B-Instruct	0.012900
cola-grammar	VP Adjunct	Meta-Llama-3.1-8B-Instruct	0.189400
cola-grammar	NP Adjunct	Meta-Llama-3.1-8B-Instruct	0.025200
cola-grammar	Temporal	Meta-Llama-3.1-8B-Instruct	0.202000
cola-grammar	Locative	Meta-Llama-3.1-8B-Instruct	0.238000
cola-grammar	Misc	Meta-Llama-3.1-8B-Instruct	0.022500
cola-grammar	Oblique	Meta-Llama-3.1-8B-Instruct	0.041700
cola-grammar	PP Arg-VP	Meta-Llama-3.1-8B-Instruct	0.134800
cola-grammar	PPArg-NP/AP	Meta-Llama-3.1-8B-Instruct	0.036800
cola-grammar	by-Phrase	Meta-Llama-3.1-8B-Instruct	0.081500
cola-grammar	Expletive	Meta-Llama-3.1-8B-Instruct	0.003200
cola-grammar	High Arity	Meta-Llama-3.1-8B-Instruct	0.008600
cola-grammar	Drop Arg	Meta-Llama-3.1-8B-Instruct	0.000900
cola-grammar	Add Arg	Meta-Llama-3.1-8B-Instruct	0.020400
cola-grammar	Passive	Meta-Llama-3.1-8B-Instruct	0.094000
cola-grammar	Imperative	Meta-Llama-3.1-8B-Instruct	0.260500
cola-grammar	Binding:Refl	Meta-Llama-3.1-8B-Instruct	0.119000
cola-grammar	Binding:Other	Meta-Llama-3.1-8B-Instruct	0.071600
cola-grammar	Matrix Q	Meta-Llama-3.1-8B-Instruct	0.166200
cola-grammar	Emb Q	Meta-Llama-3.1-8B-Instruct	0.130700
cola-grammar	Complex QP	Meta-Llama-3.1-8B-Instruct	0.218800
cola-grammar	RC	Meta-Llama-3.1-8B-Instruct	0.192200
cola-grammar	Island	Meta-Llama-3.1-8B-Instruct	0.040700
cola-grammar	CP Subj	Meta-Llama-3.1-8B-Instruct	0.008800
cola-grammar	CP Arg VP	Meta-Llama-3.1-8B-Instruct	0.073900
cola-grammar	CP Arg NP/AP	Meta-Llama-3.1-8B-Instruct	0.009800
cola-grammar	Non-finite CP	Meta-Llama-3.1-8B-Instruct	0.013800
cola-grammar	No C-izer	Meta-Llama-3.1-8B-Instruct	0.003400
cola-grammar	Deep Embed	Meta-Llama-3.1-8B-Instruct	0.039200
cola-grammar	Neg	Meta-Llama-3.1-8B-Instruct	0.309600
cola-grammar	Modal	Meta-Llama-3.1-8B-Instruct	0.853300

cola-grammar	Aux	Meta-Llama-3.1-8B-Instruct	0.350400
cola-grammar	Psuedo-Aux	Meta-Llama-3.1-8B-Instruct	0.017800
cola-grammar	Control	Meta-Llama-3.1-8B-Instruct	0.030300
cola-grammar	Raising	Meta-Llama-3.1-8B-Instruct	0.008300
cola-grammar	VP+Extract	Meta-Llama-3.1-8B-Instruct	0.018600
cola-grammar	VP arg-NP/AP	Meta-Llama-3.1-8B-Instruct	0.000800
cola-grammar	Non-finite VP Misc	Meta-Llama-3.1-8B-Instruct	0.007800
cola-grammar	Deverbal	Meta-Llama-3.1-8B-Instruct	0.000700
cola-grammar	Rel NP	Meta-Llama-3.1-8B-Instruct	0.002200
cola-grammar	Trans NP	Meta-Llama-3.1-8B-Instruct	0.000300
cola-grammar	CompX NP	Meta-Llama-3.1-8B-Instruct	0.134600
cola-grammar	NNCompd	Meta-Llama-3.1-8B-Instruct	0.025000
cola-grammar	Rel Adj	Meta-Llama-3.1-8B-Instruct	0.006800
cola-grammar	Trans Adj	Meta-Llama-3.1-8B-Instruct	0.000800
cola-grammar	Dislocation	Meta-Llama-3.1-8B-Instruct	0.010000
cola-grammar	Info Struc	Meta-Llama-3.1-8B-Instruct	0.011900
cola-grammar	Frag/Paren	Meta-Llama-3.1-8B-Instruct	0.002700
cola-grammar	Coord	Meta-Llama-3.1-8B-Instruct	0.348500
cola-grammar	Subordinate/Cond	Meta-Llama-3.1-8B-Instruct	0.081400
cola-grammar	Ellipsis/Anaphor	Meta-Llama-3.1-8B-Instruct	0.154700
cola-grammar	S-Adjunct	Meta-Llama-3.1-8B-Instruct	0.039700
cola-grammar	Quantifier	Meta-Llama-3.1-8B-Instruct	0.348600
cola-grammar	Partitive	Meta-Llama-3.1-8B-Instruct	0.185600
cola-grammar	NPI/FCI	Meta-Llama-3.1-8B-Instruct	0.123800
cola-grammar	Comparative	Meta-Llama-3.1-8B-Instruct	0.412400
cola-grammar	Sem Violation	Meta-Llama-3.1-8B-Instruct	0.009200
cola-grammar	Infl/Agr Violation	Meta-Llama-3.1-8B-Instruct	0.048400
cola-grammar	Extra/Mising Expr	Meta-Llama-3.1-8B-Instruct	0.021900
cola-grammar	Simple	Meta-Llama-3.1-70B-Instruct	0.280100
cola-grammar	Copula	Meta-Llama-3.1-70B-Instruct	0.552800
cola-grammar	Pred/SC	Meta-Llama-3.1-70B-Instruct	0.045600
cola-grammar	Result/Depictive	Meta-Llama-3.1-70B-Instruct	0.086000
cola-grammar	Particle	Meta-Llama-3.1-70B-Instruct	0.028900
cola-grammar	VP Adjunct	Meta-Llama-3.1-70B-Instruct	0.291300
cola-grammar	NP Adjunct	Meta-Llama-3.1-70B-Instruct	0.057900
cola-grammar	Temporal	Meta-Llama-3.1-70B-Instruct	0.304500
cola-grammar	Locative	Meta-Llama-3.1-70B-Instruct	0.366100
cola-grammar	Misc	Meta-Llama-3.1-70B-Instruct	0.120900
cola-grammar	Oblique	Meta-Llama-3.1-70B-Instruct	0.376900
cola-grammar	PP Arg-VP	Meta-Llama-3.1-70B-Instruct	0.394200
cola-grammar	PPArg-NP/AP	Meta-Llama-3.1-70B-Instruct	0.146400
cola-grammar	by-Phrase	Meta-Llama-3.1-70B-Instruct	0.724300
cola-grammar	Expletive	Meta-Llama-3.1-70B-Instruct	0.621000
cola-grammar	High Arity	Meta-Llama-3.1-70B-Instruct	0.262400
cola-grammar	Drop Arg	Meta-Llama-3.1-70B-Instruct	0.036600
cola-grammar	Add Arg	Meta-Llama-3.1-70B-Instruct	0.099400

cola-grammar	Passive	Meta-Llama-3.1-70B-Instruct	0.681200
cola-grammar	Imperative	Meta-Llama-3.1-70B-Instruct	0.638100
cola-grammar	Binding:Refl	Meta-Llama-3.1-70B-Instruct	0.761700
cola-grammar	Binding:Other	Meta-Llama-3.1-70B-Instruct	0.286100
cola-grammar	Matrix Q	Meta-Llama-3.1-70B-Instruct	0.543100
cola-grammar	Emb Q	Meta-Llama-3.1-70B-Instruct	0.724900
cola-grammar	Complex QP	Meta-Llama-3.1-70B-Instruct	0.572600
cola-grammar	RC	Meta-Llama-3.1-70B-Instruct	0.474800
cola-grammar	Island	Meta-Llama-3.1-70B-Instruct	0.270000
cola-grammar	CP Subj	Meta-Llama-3.1-70B-Instruct	0.144400
cola-grammar	CP Arg VP	Meta-Llama-3.1-70B-Instruct	0.339600
cola-grammar	CP Arg NP/AP	Meta-Llama-3.1-70B-Instruct	0.089500
cola-grammar	Non-finite CP	Meta-Llama-3.1-70B-Instruct	0.114300
cola-grammar	No C-izer	Meta-Llama-3.1-70B-Instruct	0.080500
cola-grammar	Deep Embed	Meta-Llama-3.1-70B-Instruct	0.183200
cola-grammar	Neg	Meta-Llama-3.1-70B-Instruct	0.606700
cola-grammar	Modal	Meta-Llama-3.1-70B-Instruct	0.933600
cola-grammar	Aux	Meta-Llama-3.1-70B-Instruct	0.359700
cola-grammar	Psuedo-Aux	Meta-Llama-3.1-70B-Instruct	0.086400
cola-grammar	Control	Meta-Llama-3.1-70B-Instruct	0.230800
cola-grammar	Raising	Meta-Llama-3.1-70B-Instruct	0.064100
cola-grammar	VP+Extract	Meta-Llama-3.1-70B-Instruct	0.155000
cola-grammar	VP arg-NP/AP	Meta-Llama-3.1-70B-Instruct	0.065000
cola-grammar	Non-finite VP Misc	Meta-Llama-3.1-70B-Instruct	0.036500
cola-grammar	Deverbal	Meta-Llama-3.1-70B-Instruct	0.009500
cola-grammar	Rel NP	Meta-Llama-3.1-70B-Instruct	0.067800
cola-grammar	Trans NP	Meta-Llama-3.1-70B-Instruct	0.022800
cola-grammar	Comp NP	Meta-Llama-3.1-70B-Instruct	0.241800
cola-grammar	NNCompd	Meta-Llama-3.1-70B-Instruct	0.124600
cola-grammar	Rel Adj	Meta-Llama-3.1-70B-Instruct	0.030500
cola-grammar	Trans Adj	Meta-Llama-3.1-70B-Instruct	0.050400
cola-grammar	Dislocation	Meta-Llama-3.1-70B-Instruct	0.052500
cola-grammar	Info Struc	Meta-Llama-3.1-70B-Instruct	0.073100
cola-grammar	Frag/Paren	Meta-Llama-3.1-70B-Instruct	0.041400
cola-grammar	Coord	Meta-Llama-3.1-70B-Instruct	0.682900
cola-grammar	Subordinate/Cond	Meta-Llama-3.1-70B-Instruct	0.137500
cola-grammar	Ellipsis/Anaphor	Meta-Llama-3.1-70B-Instruct	0.535600
cola-grammar	S-Adjunct	Meta-Llama-3.1-70B-Instruct	0.106000
cola-grammar	Quantifier	Meta-Llama-3.1-70B-Instruct	0.543500
cola-grammar	Partitive	Meta-Llama-3.1-70B-Instruct	0.550100
cola-grammar	NPI/FCI	Meta-Llama-3.1-70B-Instruct	0.304400
cola-grammar	Comparative	Meta-Llama-3.1-70B-Instruct	0.719900
cola-grammar	Sem Violation	Meta-Llama-3.1-70B-Instruct	0.077200
cola-grammar	Infl/Agr Violation	Meta-Llama-3.1-70B-Instruct	0.166800
cola-grammar	Extra/Mising Expr	Meta-Llama-3.1-70B-Instruct	0.149100
cola-grammar	Simple	OLMo-7B-0724-Instruct-hf	0.006700

cola-grammar	Copula	OLMo-7B-0724-Instruct-hf	0.036300
cola-grammar	Pred/SC	OLMo-7B-0724-Instruct-hf	0.003000
cola-grammar	Result/Depictive	OLMo-7B-0724-Instruct-hf	0.002400
cola-grammar	Particle	OLMo-7B-0724-Instruct-hf	-0.026900
cola-grammar	VP Adjunct	OLMo-7B-0724-Instruct-hf	0.068800
cola-grammar	NP Adjunct	OLMo-7B-0724-Instruct-hf	0.013000
cola-grammar	Temporal	OLMo-7B-0724-Instruct-hf	0.031200
cola-grammar	Locative	OLMo-7B-0724-Instruct-hf	0.033700
cola-grammar	Misc	OLMo-7B-0724-Instruct-hf	0.040100
cola-grammar	Oblique	OLMo-7B-0724-Instruct-hf	0.061000
cola-grammar	PP Arg-VP	OLMo-7B-0724-Instruct-hf	0.038000
cola-grammar	PPArg-NP/AP	OLMo-7B-0724-Instruct-hf	0.002600
cola-grammar	by-Phrase	OLMo-7B-0724-Instruct-hf	0.112500
cola-grammar	Expletive	OLMo-7B-0724-Instruct-hf	0.011000
cola-grammar	High Arity	OLMo-7B-0724-Instruct-hf	0.068000
cola-grammar	Drop Arg	OLMo-7B-0724-Instruct-hf	-0.029500
cola-grammar	Add Arg	OLMo-7B-0724-Instruct-hf	-0.013700
cola-grammar	Passive	OLMo-7B-0724-Instruct-hf	0.035200
cola-grammar	Imperative	OLMo-7B-0724-Instruct-hf	0.079300
cola-grammar	Binding:Refl	OLMo-7B-0724-Instruct-hf	0.020600
cola-grammar	Binding:Other	OLMo-7B-0724-Instruct-hf	0.005300
cola-grammar	Matrix Q	OLMo-7B-0724-Instruct-hf	0.076800
cola-grammar	Emb Q	OLMo-7B-0724-Instruct-hf	0.043200
cola-grammar	Complex QP	OLMo-7B-0724-Instruct-hf	0.054400
cola-grammar	RC	OLMo-7B-0724-Instruct-hf	0.065300
cola-grammar	Island	OLMo-7B-0724-Instruct-hf	0.004800
cola-grammar	CP Subj	OLMo-7B-0724-Instruct-hf	0.003200
cola-grammar	CP Arg VP	OLMo-7B-0724-Instruct-hf	0.034000
cola-grammar	CP Arg NP/AP	OLMo-7B-0724-Instruct-hf	0.003200
cola-grammar	Non-finite CP	OLMo-7B-0724-Instruct-hf	-0.003800
cola-grammar	No C-izer	OLMo-7B-0724-Instruct-hf	0.008300
cola-grammar	Deep Embed	OLMo-7B-0724-Instruct-hf	0.076000
cola-grammar	Neg	OLMo-7B-0724-Instruct-hf	0.095500
cola-grammar	Modal	OLMo-7B-0724-Instruct-hf	0.352100
cola-grammar	Aux	OLMo-7B-0724-Instruct-hf	0.017300
cola-grammar	Psuedo-Aux	OLMo-7B-0724-Instruct-hf	0.027800
cola-grammar	Control	OLMo-7B-0724-Instruct-hf	-0.000800
cola-grammar	Raising	OLMo-7B-0724-Instruct-hf	-0.000800
cola-grammar	VP+Extract	OLMo-7B-0724-Instruct-hf	0.020900
cola-grammar	VP arg-NP/AP	OLMo-7B-0724-Instruct-hf	-0.003200
cola-grammar	Non-finite VP Misc	OLMo-7B-0724-Instruct-hf	0.004000
cola-grammar	Deverbal	OLMo-7B-0724-Instruct-hf	0.000800
cola-grammar	Rel NP	OLMo-7B-0724-Instruct-hf	0.006900
cola-grammar	Trans NP	OLMo-7B-0724-Instruct-hf	0.001800
cola-grammar	Compx NP	OLMo-7B-0724-Instruct-hf	0.133000
cola-grammar	NNCompd	OLMo-7B-0724-Instruct-hf	0.016900

cola-grammar	Rel Adj	OLMo-7B-0724-Instruct-hf	0.001200
cola-grammar	Trans Adj	OLMo-7B-0724-Instruct-hf	-0.003300
cola-grammar	Dislocation	OLMo-7B-0724-Instruct-hf	0.021300
cola-grammar	Info Struc	OLMo-7B-0724-Instruct-hf	0.007400
cola-grammar	Frag/Paren	OLMo-7B-0724-Instruct-hf	-0.003800
cola-grammar	Coord	OLMo-7B-0724-Instruct-hf	0.304900
cola-grammar	Subordinate/Cond	OLMo-7B-0724-Instruct-hf	0.034300
cola-grammar	Ellipsis/Anaphor	OLMo-7B-0724-Instruct-hf	0.112800
cola-grammar	S-Adjunct	OLMo-7B-0724-Instruct-hf	0.024000
cola-grammar	Quantifier	OLMo-7B-0724-Instruct-hf	0.079800
cola-grammar	Partitive	OLMo-7B-0724-Instruct-hf	0.039700
cola-grammar	NPI/FCI	OLMo-7B-0724-Instruct-hf	0.056200
cola-grammar	Comparative	OLMo-7B-0724-Instruct-hf	0.070000
cola-grammar	Sem Violation	OLMo-7B-0724-Instruct-hf	-0.007000
cola-grammar	Infl/Agr Violation	OLMo-7B-0724-Instruct-hf	0.011900
cola-grammar	Extra/Mising Expr	OLMo-7B-0724-Instruct-hf	0.038600
cola-grammar	Simple	Mistral-7B-Instruct-v0.3	0.082200
cola-grammar	Copula	Mistral-7B-Instruct-v0.3	0.157700
cola-grammar	Pred/SC	Mistral-7B-Instruct-v0.3	0.017200
cola-grammar	Result/Depictive	Mistral-7B-Instruct-v0.3	0.014500
cola-grammar	Particle	Mistral-7B-Instruct-v0.3	0.002500
cola-grammar	VP Adjunct	Mistral-7B-Instruct-v0.3	0.038800
cola-grammar	NP Adjunct	Mistral-7B-Instruct-v0.3	0.013900
cola-grammar	Temporal	Mistral-7B-Instruct-v0.3	0.038600
cola-grammar	Locative	Mistral-7B-Instruct-v0.3	0.167100
cola-grammar	Misc	Mistral-7B-Instruct-v0.3	0.007600
cola-grammar	Oblique	Mistral-7B-Instruct-v0.3	0.033500
cola-grammar	PP Arg-VP	Mistral-7B-Instruct-v0.3	0.083000
cola-grammar	PPArg-NP/AP	Mistral-7B-Instruct-v0.3	0.010900
cola-grammar	by-Phrase	Mistral-7B-Instruct-v0.3	0.273700
cola-grammar	Expletive	Mistral-7B-Instruct-v0.3	0.117600
cola-grammar	High Arity	Mistral-7B-Instruct-v0.3	0.092100
cola-grammar	Drop Arg	Mistral-7B-Instruct-v0.3	0.005200
cola-grammar	Add Arg	Mistral-7B-Instruct-v0.3	0.029700
cola-grammar	Passive	Mistral-7B-Instruct-v0.3	0.136600
cola-grammar	Imperative	Mistral-7B-Instruct-v0.3	0.131400
cola-grammar	Binding:Refl	Mistral-7B-Instruct-v0.3	0.425100
cola-grammar	Binding:Other	Mistral-7B-Instruct-v0.3	0.082900
cola-grammar	Matrix Q	Mistral-7B-Instruct-v0.3	0.445800
cola-grammar	Emb Q	Mistral-7B-Instruct-v0.3	0.232300
cola-grammar	Complex QP	Mistral-7B-Instruct-v0.3	0.205500
cola-grammar	RC	Mistral-7B-Instruct-v0.3	0.059800
cola-grammar	Island	Mistral-7B-Instruct-v0.3	0.004200
cola-grammar	CP Subj	Mistral-7B-Instruct-v0.3	-0.005000
cola-grammar	CP Arg VP	Mistral-7B-Instruct-v0.3	0.014500
cola-grammar	CP Arg NP/AP	Mistral-7B-Instruct-v0.3	0.003100

cola-grammar	Non-finite CP	Mistral-7B-Instruct-v0.3	0.013000
cola-grammar	No C-izer	Mistral-7B-Instruct-v0.3	0.014200
cola-grammar	Deep Embed	Mistral-7B-Instruct-v0.3	0.018700
cola-grammar	Neg	Mistral-7B-Instruct-v0.3	0.310400
cola-grammar	Modal	Mistral-7B-Instruct-v0.3	0.648000
cola-grammar	Aux	Mistral-7B-Instruct-v0.3	0.184600
cola-grammar	Psuedo-Aux	Mistral-7B-Instruct-v0.3	0.052200
cola-grammar	Control	Mistral-7B-Instruct-v0.3	0.030600
cola-grammar	Raising	Mistral-7B-Instruct-v0.3	0.042900
cola-grammar	VP+Extract	Mistral-7B-Instruct-v0.3	0.004900
cola-grammar	VP arg-NP/AP	Mistral-7B-Instruct-v0.3	0.007400
cola-grammar	Non-finite VP Misc	Mistral-7B-Instruct-v0.3	0.030400
cola-grammar	Deverbal	Mistral-7B-Instruct-v0.3	0.004300
cola-grammar	Rel NP	Mistral-7B-Instruct-v0.3	0.015600
cola-grammar	Trans NP	Mistral-7B-Instruct-v0.3	0.009900
cola-grammar	Compx NP	Mistral-7B-Instruct-v0.3	0.029400
cola-grammar	NNCompd	Mistral-7B-Instruct-v0.3	0.042500
cola-grammar	Rel Adj	Mistral-7B-Instruct-v0.3	0.007900
cola-grammar	Trans Adj	Mistral-7B-Instruct-v0.3	-0.005000
cola-grammar	Dislocation	Mistral-7B-Instruct-v0.3	0.012300
cola-grammar	Info Struc	Mistral-7B-Instruct-v0.3	0.039900
cola-grammar	Frag/Paren	Mistral-7B-Instruct-v0.3	0.018200
cola-grammar	Coord	Mistral-7B-Instruct-v0.3	0.034700
cola-grammar	Subordinate/Cond	Mistral-7B-Instruct-v0.3	0.021300
cola-grammar	Ellipsis/Anaphor	Mistral-7B-Instruct-v0.3	0.186500
cola-grammar	S-Adjunct	Mistral-7B-Instruct-v0.3	0.005200
cola-grammar	Quantifier	Mistral-7B-Instruct-v0.3	0.280200
cola-grammar	Partitive	Mistral-7B-Instruct-v0.3	0.151800
cola-grammar	NPI/FCI	Mistral-7B-Instruct-v0.3	0.164900
cola-grammar	Comparative	Mistral-7B-Instruct-v0.3	0.234000
cola-grammar	Sem Violation	Mistral-7B-Instruct-v0.3	0.015900
cola-grammar	Infl/Agr Violation	Mistral-7B-Instruct-v0.3	0.027800
cola-grammar	Extra/Mising Expr	Mistral-7B-Instruct-v0.3	0.077700
cola-grammar	Simple	Starling-LM-7B-alpha	-0.013000
cola-grammar	Copula	Starling-LM-7B-alpha	0.057400
cola-grammar	Pred/SC	Starling-LM-7B-alpha	0.012400
cola-grammar	Result/Depictive	Starling-LM-7B-alpha	-0.010700
cola-grammar	Particle	Starling-LM-7B-alpha	0.022000
cola-grammar	VP Adjunct	Starling-LM-7B-alpha	0.134300
cola-grammar	NP Adjunct	Starling-LM-7B-alpha	0.048700
cola-grammar	Temporal	Starling-LM-7B-alpha	0.090900
cola-grammar	Locative	Starling-LM-7B-alpha	0.193100
cola-grammar	Misc	Starling-LM-7B-alpha	0.079500
cola-grammar	Oblique	Starling-LM-7B-alpha	-0.013100
cola-grammar	PP Arg-VP	Starling-LM-7B-alpha	0.142900
cola-grammar	PPArg-NP/AP	Starling-LM-7B-alpha	0.054300

cola-grammar	by-Phrase	Starling-LM-7B-alpha	0.271500
cola-grammar	Expletive	Starling-LM-7B-alpha	0.051300
cola-grammar	High Arity	Starling-LM-7B-alpha	0.094600
cola-grammar	Drop Arg	Starling-LM-7B-alpha	-0.025100
cola-grammar	Add Arg	Starling-LM-7B-alpha	0.017000
cola-grammar	Passive	Starling-LM-7B-alpha	0.240500
cola-grammar	Imperative	Starling-LM-7B-alpha	0.021000
cola-grammar	Binding:Refl	Starling-LM-7B-alpha	0.081700
cola-grammar	Binding:Other	Starling-LM-7B-alpha	0.013900
cola-grammar	Matrix Q	Starling-LM-7B-alpha	0.024000
cola-grammar	Emb Q	Starling-LM-7B-alpha	0.188200
cola-grammar	Complex QP	Starling-LM-7B-alpha	0.030700
cola-grammar	RC	Starling-LM-7B-alpha	0.157200
cola-grammar	Island	Starling-LM-7B-alpha	0.010000
cola-grammar	CP Subj	Starling-LM-7B-alpha	0.018200
cola-grammar	CP Arg VP	Starling-LM-7B-alpha	0.074200
cola-grammar	CP Arg NP/AP	Starling-LM-7B-alpha	0.028700
cola-grammar	Non-finite CP	Starling-LM-7B-alpha	0.018200
cola-grammar	No C-izer	Starling-LM-7B-alpha	0.034100
cola-grammar	Deep Embed	Starling-LM-7B-alpha	0.034700
cola-grammar	Neg	Starling-LM-7B-alpha	0.151000
cola-grammar	Modal	Starling-LM-7B-alpha	0.476700
cola-grammar	Aux	Starling-LM-7B-alpha	0.083200
cola-grammar	Psuedo-Aux	Starling-LM-7B-alpha	-0.014200
cola-grammar	Control	Starling-LM-7B-alpha	0.007700
cola-grammar	Raising	Starling-LM-7B-alpha	-0.000900
cola-grammar	VP+Extract	Starling-LM-7B-alpha	0.009100
cola-grammar	VP arg-NP/AP	Starling-LM-7B-alpha	0.030500
cola-grammar	Non-finite VP Misc	Starling-LM-7B-alpha	-0.006400
cola-grammar	Deverbal	Starling-LM-7B-alpha	0.028200
cola-grammar	Rel NP	Starling-LM-7B-alpha	0.018800
cola-grammar	Trans NP	Starling-LM-7B-alpha	0.006400
cola-grammar	Comp NP	Starling-LM-7B-alpha	0.074400
cola-grammar	NNCompd	Starling-LM-7B-alpha	0.008900
cola-grammar	Rel Adj	Starling-LM-7B-alpha	0.019900
cola-grammar	Trans Adj	Starling-LM-7B-alpha	0.016600
cola-grammar	Dislocation	Starling-LM-7B-alpha	0.044300
cola-grammar	Info Struc	Starling-LM-7B-alpha	0.042600
cola-grammar	Frag/Paren	Starling-LM-7B-alpha	-0.001500
cola-grammar	Coord	Starling-LM-7B-alpha	0.164200
cola-grammar	Subordinate/Cond	Starling-LM-7B-alpha	0.095800
cola-grammar	Ellipsis/Anaphor	Starling-LM-7B-alpha	0.119800
cola-grammar	S-Adjunct	Starling-LM-7B-alpha	0.034400
cola-grammar	Quantifier	Starling-LM-7B-alpha	0.197300
cola-grammar	Partitive	Starling-LM-7B-alpha	0.070600
cola-grammar	NPI/FCI	Starling-LM-7B-alpha	0.048700

cola-grammar	Comparative	Starling-LM-7B-alpha	0.188200
cola-grammar	Sem Violation	Starling-LM-7B-alpha	0.027500
cola-grammar	Infl/Agr Violation	Starling-LM-7B-alpha	0.063800
cola-grammar	Extra/Mising Expr	Starling-LM-7B-alpha	0.067000
cola-grammar	Simple	c4ai-command-r-v01	0.106900
cola-grammar	Copula	c4ai-command-r-v01	0.216700
cola-grammar	Pred/SC	c4ai-command-r-v01	-0.008100
cola-grammar	Result/Depictive	c4ai-command-r-v01	-0.005200
cola-grammar	Particle	c4ai-command-r-v01	-0.014600
cola-grammar	VP Adjunct	c4ai-command-r-v01	0.040800
cola-grammar	NP Adjunct	c4ai-command-r-v01	0.006100
cola-grammar	Temporal	c4ai-command-r-v01	0.251800
cola-grammar	Locative	c4ai-command-r-v01	0.129400
cola-grammar	Misc	c4ai-command-r-v01	-0.017200
cola-grammar	Oblique	c4ai-command-r-v01	0.009300
cola-grammar	PP Arg-VP	c4ai-command-r-v01	0.039600
cola-grammar	PPArg-NP/AP	c4ai-command-r-v01	-0.003200
cola-grammar	by-Phrase	c4ai-command-r-v01	0.248400
cola-grammar	Expletive	c4ai-command-r-v01	0.188500
cola-grammar	High Arity	c4ai-command-r-v01	0.039900
cola-grammar	Drop Arg	c4ai-command-r-v01	0.044800
cola-grammar	Add Arg	c4ai-command-r-v01	0.059700
cola-grammar	Passive	c4ai-command-r-v01	0.182200
cola-grammar	Imperative	c4ai-command-r-v01	0.028400
cola-grammar	Binding:Refl	c4ai-command-r-v01	0.340600
cola-grammar	Binding:Other	c4ai-command-r-v01	0.090200
cola-grammar	Matrix Q	c4ai-command-r-v01	0.184400
cola-grammar	Emb Q	c4ai-command-r-v01	0.086800
cola-grammar	Complex QP	c4ai-command-r-v01	0.134500
cola-grammar	RC	c4ai-command-r-v01	0.214600
cola-grammar	Island	c4ai-command-r-v01	0.023400
cola-grammar	CP Subj	c4ai-command-r-v01	0.038100
cola-grammar	CP Arg VP	c4ai-command-r-v01	0.121000
cola-grammar	CP Arg NP/AP	c4ai-command-r-v01	-0.018600
cola-grammar	Non-finite CP	c4ai-command-r-v01	0.043600
cola-grammar	No C-izer	c4ai-command-r-v01	0.041600
cola-grammar	Deep Embed	c4ai-command-r-v01	0.014200
cola-grammar	Neg	c4ai-command-r-v01	0.122500
cola-grammar	Modal	c4ai-command-r-v01	0.363100
cola-grammar	Aux	c4ai-command-r-v01	0.126200
cola-grammar	Psuedo-Aux	c4ai-command-r-v01	-0.016000
cola-grammar	Control	c4ai-command-r-v01	-0.014700
cola-grammar	Raising	c4ai-command-r-v01	0.025800
cola-grammar	VP+Extract	c4ai-command-r-v01	0.066700
cola-grammar	VP arg-NP/AP	c4ai-command-r-v01	0.004900
cola-grammar	Non-finite VP Misc	c4ai-command-r-v01	-0.004500

cola-grammar	Deverbal	c4ai-command-r-v01	0.015600
cola-grammar	Rel NP	c4ai-command-r-v01	-0.011500
cola-grammar	Trans NP	c4ai-command-r-v01	-0.035200
cola-grammar	Compx NP	c4ai-command-r-v01	0.213900
cola-grammar	NNCompd	c4ai-command-r-v01	0.009500
cola-grammar	Rel Adj	c4ai-command-r-v01	-0.024000
cola-grammar	Trans Adj	c4ai-command-r-v01	-0.034900
cola-grammar	Dislocation	c4ai-command-r-v01	0.030600
cola-grammar	Info Struc	c4ai-command-r-v01	-0.043700
cola-grammar	Frag/Paren	c4ai-command-r-v01	0.010100
cola-grammar	Coord	c4ai-command-r-v01	0.300800
cola-grammar	Subordinate/Cond	c4ai-command-r-v01	0.023500
cola-grammar	Ellipsis/Anaphor	c4ai-command-r-v01	-0.026100
cola-grammar	S-Adjunct	c4ai-command-r-v01	0.011300
cola-grammar	Quantifier	c4ai-command-r-v01	0.176600
cola-grammar	Partitive	c4ai-command-r-v01	0.102500
cola-grammar	NPI/FCI	c4ai-command-r-v01	0.105600
cola-grammar	Comparative	c4ai-command-r-v01	0.180100
cola-grammar	Sem Violation	c4ai-command-r-v01	0.075600
cola-grammar	Infl/Agr Violation	c4ai-command-r-v01	0.157400
cola-grammar	Extra/Mising Expr	c4ai-command-r-v01	0.030900
cola-grammar	Simple	Mixtral-8x7B-Instruct-v0.1	0.016200
cola-grammar	Copula	Mixtral-8x7B-Instruct-v0.1	0.399600
cola-grammar	Pred/SC	Mixtral-8x7B-Instruct-v0.1	0.041000
cola-grammar	Result/Depictive	Mixtral-8x7B-Instruct-v0.1	0.169300
cola-grammar	Particle	Mixtral-8x7B-Instruct-v0.1	0.005700
cola-grammar	VP Adjunct	Mixtral-8x7B-Instruct-v0.1	0.252500
cola-grammar	NP Adjunct	Mixtral-8x7B-Instruct-v0.1	0.102400
cola-grammar	Temporal	Mixtral-8x7B-Instruct-v0.1	0.367900
cola-grammar	Locative	Mixtral-8x7B-Instruct-v0.1	0.399500
cola-grammar	Misc	Mixtral-8x7B-Instruct-v0.1	0.148700
cola-grammar	Oblique	Mixtral-8x7B-Instruct-v0.1	0.283500
cola-grammar	PP Arg-VP	Mixtral-8x7B-Instruct-v0.1	0.394200
cola-grammar	PPArg-NP/AP	Mixtral-8x7B-Instruct-v0.1	0.089800
cola-grammar	by-Phrase	Mixtral-8x7B-Instruct-v0.1	0.648100
cola-grammar	Expletive	Mixtral-8x7B-Instruct-v0.1	0.282300
cola-grammar	High Arity	Mixtral-8x7B-Instruct-v0.1	0.338000
cola-grammar	Drop Arg	Mixtral-8x7B-Instruct-v0.1	0.041000
cola-grammar	Add Arg	Mixtral-8x7B-Instruct-v0.1	0.040400
cola-grammar	Passive	Mixtral-8x7B-Instruct-v0.1	0.542400
cola-grammar	Imperative	Mixtral-8x7B-Instruct-v0.1	0.330600
cola-grammar	Binding:Refl	Mixtral-8x7B-Instruct-v0.1	0.616800
cola-grammar	Binding:Other	Mixtral-8x7B-Instruct-v0.1	0.247900
cola-grammar	Matrix Q	Mixtral-8x7B-Instruct-v0.1	0.201300
cola-grammar	Emb Q	Mixtral-8x7B-Instruct-v0.1	0.360700
cola-grammar	Complex QP	Mixtral-8x7B-Instruct-v0.1	0.293500

cola-grammar	RC	Mixtral-8x7B-Instruct-v0.1	0.409900
cola-grammar	Island	Mixtral-8x7B-Instruct-v0.1	0.086400
cola-grammar	CP Subj	Mixtral-8x7B-Instruct-v0.1	0.059800
cola-grammar	CP Arg VP	Mixtral-8x7B-Instruct-v0.1	0.198000
cola-grammar	CP Arg NP/AP	Mixtral-8x7B-Instruct-v0.1	0.052500
cola-grammar	Non-finite CP	Mixtral-8x7B-Instruct-v0.1	0.127000
cola-grammar	No C-izer	Mixtral-8x7B-Instruct-v0.1	0.027000
cola-grammar	Deep Embed	Mixtral-8x7B-Instruct-v0.1	0.125500
cola-grammar	Neg	Mixtral-8x7B-Instruct-v0.1	0.324100
cola-grammar	Modal	Mixtral-8x7B-Instruct-v0.1	0.824300
cola-grammar	Aux	Mixtral-8x7B-Instruct-v0.1	0.362900
cola-grammar	Psuedo-Aux	Mixtral-8x7B-Instruct-v0.1	0.148400
cola-grammar	Control	Mixtral-8x7B-Instruct-v0.1	0.240800
cola-grammar	Raising	Mixtral-8x7B-Instruct-v0.1	0.058000
cola-grammar	VP+Extract	Mixtral-8x7B-Instruct-v0.1	0.122000
cola-grammar	VP arg-NP/AP	Mixtral-8x7B-Instruct-v0.1	0.033500
cola-grammar	Non-finite VP Misc	Mixtral-8x7B-Instruct-v0.1	0.030600
cola-grammar	Deverbal	Mixtral-8x7B-Instruct-v0.1	0.004200
cola-grammar	Rel NP	Mixtral-8x7B-Instruct-v0.1	0.025500
cola-grammar	Trans NP	Mixtral-8x7B-Instruct-v0.1	0.009300
cola-grammar	Compx NP	Mixtral-8x7B-Instruct-v0.1	0.110700
cola-grammar	NNCompd	Mixtral-8x7B-Instruct-v0.1	0.022100
cola-grammar	Rel Adj	Mixtral-8x7B-Instruct-v0.1	0.058800
cola-grammar	Trans Adj	Mixtral-8x7B-Instruct-v0.1	0.064300
cola-grammar	Dislocation	Mixtral-8x7B-Instruct-v0.1	0.133600
cola-grammar	Info Struc	Mixtral-8x7B-Instruct-v0.1	0.237300
cola-grammar	Frag/Paren	Mixtral-8x7B-Instruct-v0.1	0.009300
cola-grammar	Coord	Mixtral-8x7B-Instruct-v0.1	0.540500
cola-grammar	Subordinate/Cond	Mixtral-8x7B-Instruct-v0.1	0.180900
cola-grammar	Ellipsis/Anaphor	Mixtral-8x7B-Instruct-v0.1	0.312600
cola-grammar	S-Adjunct	Mixtral-8x7B-Instruct-v0.1	0.070900
cola-grammar	Quantifier	Mixtral-8x7B-Instruct-v0.1	0.252500
cola-grammar	Partitive	Mixtral-8x7B-Instruct-v0.1	0.410300
cola-grammar	NPI/FCI	Mixtral-8x7B-Instruct-v0.1	0.242100
cola-grammar	Comparative	Mixtral-8x7B-Instruct-v0.1	0.476800
cola-grammar	Sem Violation	Mixtral-8x7B-Instruct-v0.1	0.060100
cola-grammar	Infl/Agr Violation	Mixtral-8x7B-Instruct-v0.1	0.229800
cola-grammar	Extra/Mising Expr	Mixtral-8x7B-Instruct-v0.1	0.160300
cola-grammar	Simple	gpt-4o	0.404400
cola-grammar	Copula	gpt-4o	0.291600
cola-grammar	Pred/SC	gpt-4o	0.410400
cola-grammar	Result/Depictive	gpt-4o	0.477600
cola-grammar	Particle	gpt-4o	0.268500
cola-grammar	VP Adjunct	gpt-4o	0.441300
cola-grammar	NP Adjunct	gpt-4o	0.407200
cola-grammar	Temporal	gpt-4o	0.648200

cola-grammar	Locative	gpt-4o	0.489900
cola-grammar	Misc	gpt-4o	0.299000
cola-grammar	Oblique	gpt-4o	0.572000
cola-grammar	PP Arg-VP	gpt-4o	0.616900
cola-grammar	PPArg-NP/AP	gpt-4o	0.697900
cola-grammar	by-Phrase	gpt-4o	0.838200
cola-grammar	Expletive	gpt-4o	0.896700
cola-grammar	High Arity	gpt-4o	0.522500
cola-grammar	Drop Arg	gpt-4o	0.204100
cola-grammar	Add Arg	gpt-4o	0.217500
cola-grammar	Passive	gpt-4o	0.790300
cola-grammar	Imperative	gpt-4o	0.522300
cola-grammar	Binding:Refl	gpt-4o	0.743700
cola-grammar	Binding:Other	gpt-4o	0.347500
cola-grammar	Matrix Q	gpt-4o	0.801300
cola-grammar	Emb Q	gpt-4o	0.716700
cola-grammar	Complex QP	gpt-4o	0.687700
cola-grammar	RC	gpt-4o	0.779100
cola-grammar	Island	gpt-4o	0.214900
cola-grammar	CP Subj	gpt-4o	0.489600
cola-grammar	CP Arg VP	gpt-4o	0.571600
cola-grammar	CP Arg NP/AP	gpt-4o	0.219000
cola-grammar	Non-finite CP	gpt-4o	0.326500
cola-grammar	No C-izer	gpt-4o	0.378900
cola-grammar	Deep Embed	gpt-4o	0.340500
cola-grammar	Neg	gpt-4o	0.772200
cola-grammar	Modal	gpt-4o	0.914900
cola-grammar	Aux	gpt-4o	0.441700
cola-grammar	Psuedo-Aux	gpt-4o	0.417500
cola-grammar	Control	gpt-4o	0.527500
cola-grammar	Raising	gpt-4o	0.318200
cola-grammar	VP+Extract	gpt-4o	0.401400
cola-grammar	VP arg-NP/AP	gpt-4o	0.337800
cola-grammar	Non-finite VP Misc	gpt-4o	0.154700
cola-grammar	Deverbal	gpt-4o	0.285900
cola-grammar	Rel NP	gpt-4o	0.395700
cola-grammar	Trans NP	gpt-4o	0.261600
cola-grammar	Compx NP	gpt-4o	0.515700
cola-grammar	NNCompd	gpt-4o	0.441700
cola-grammar	Rel Adj	gpt-4o	0.084000
cola-grammar	Trans Adj	gpt-4o	0.317200
cola-grammar	Dislocation	gpt-4o	0.169700
cola-grammar	Info Struc	gpt-4o	0.552700
cola-grammar	Frag/Paren	gpt-4o	0.093700
cola-grammar	Coord	gpt-4o	0.754500
cola-grammar	Subordinate/Cond	gpt-4o	0.233100

cola-grammar	Ellipsis/Anaphor	gpt-4o	0.752600
cola-grammar	S-Adjunct	gpt-4o	0.372900
cola-grammar	Quantifier	gpt-4o	0.574700
cola-grammar	Partitive	gpt-4o	0.784500
cola-grammar	NPI/FCI	gpt-4o	0.580400
cola-grammar	Comparative	gpt-4o	0.786400
cola-grammar	Sem Violation	gpt-4o	0.109100
cola-grammar	Infl/Agr Violation	gpt-4o	0.322800
cola-grammar	Extra/Mising Expr	gpt-4o	0.218400
cola-grammar	Simple	Mixtral-8x22B-Instruct-v0.1	0.138800
cola-grammar	Copula	Mixtral-8x22B-Instruct-v0.1	0.319500
cola-grammar	Pred/SC	Mixtral-8x22B-Instruct-v0.1	0.168400
cola-grammar	Result/Depictive	Mixtral-8x22B-Instruct-v0.1	0.157300
cola-grammar	Particle	Mixtral-8x22B-Instruct-v0.1	0.010300
cola-grammar	VP Adjunct	Mixtral-8x22B-Instruct-v0.1	0.416400
cola-grammar	NP Adjunct	Mixtral-8x22B-Instruct-v0.1	0.148500
cola-grammar	Temporal	Mixtral-8x22B-Instruct-v0.1	0.501000
cola-grammar	Locative	Mixtral-8x22B-Instruct-v0.1	0.476500
cola-grammar	Misc	Mixtral-8x22B-Instruct-v0.1	0.212000
cola-grammar	Oblique	Mixtral-8x22B-Instruct-v0.1	0.304500
cola-grammar	PP Arg-VP	Mixtral-8x22B-Instruct-v0.1	0.336800
cola-grammar	PPArg-NP/AP	Mixtral-8x22B-Instruct-v0.1	0.186100
cola-grammar	by-Phrase	Mixtral-8x22B-Instruct-v0.1	0.812800
cola-grammar	Expletive	Mixtral-8x22B-Instruct-v0.1	0.487100
cola-grammar	High Arity	Mixtral-8x22B-Instruct-v0.1	0.401300
cola-grammar	Drop Arg	Mixtral-8x22B-Instruct-v0.1	0.043000
cola-grammar	Add Arg	Mixtral-8x22B-Instruct-v0.1	0.067700
cola-grammar	Passive	Mixtral-8x22B-Instruct-v0.1	0.756400
cola-grammar	Imperative	Mixtral-8x22B-Instruct-v0.1	0.378300
cola-grammar	Binding:Refl	Mixtral-8x22B-Instruct-v0.1	0.749600
cola-grammar	Binding:Other	Mixtral-8x22B-Instruct-v0.1	0.144300
cola-grammar	Matrix Q	Mixtral-8x22B-Instruct-v0.1	0.658700
cola-grammar	Emb Q	Mixtral-8x22B-Instruct-v0.1	0.405600
cola-grammar	Complex QP	Mixtral-8x22B-Instruct-v0.1	0.532300
cola-grammar	RC	Mixtral-8x22B-Instruct-v0.1	0.619200
cola-grammar	Island	Mixtral-8x22B-Instruct-v0.1	0.101300
cola-grammar	CP Subj	Mixtral-8x22B-Instruct-v0.1	0.089700
cola-grammar	CP Arg VP	Mixtral-8x22B-Instruct-v0.1	0.231100
cola-grammar	CP Arg NP/AP	Mixtral-8x22B-Instruct-v0.1	0.075000
cola-grammar	Non-finite CP	Mixtral-8x22B-Instruct-v0.1	0.132000
cola-grammar	No C-izer	Mixtral-8x22B-Instruct-v0.1	0.054500
cola-grammar	Deep Embed	Mixtral-8x22B-Instruct-v0.1	0.118700
cola-grammar	Neg	Mixtral-8x22B-Instruct-v0.1	0.612400
cola-grammar	Modal	Mixtral-8x22B-Instruct-v0.1	0.853600
cola-grammar	Aux	Mixtral-8x22B-Instruct-v0.1	0.365200
cola-grammar	Psuedo-Aux	Mixtral-8x22B-Instruct-v0.1	0.078500

cola-grammar	Control	Mixtral-8x22B-Instruct-v0.1	0.290900
cola-grammar	Raising	Mixtral-8x22B-Instruct-v0.1	0.120200
cola-grammar	VP+Extract	Mixtral-8x22B-Instruct-v0.1	0.123600
cola-grammar	VP arg-NP/AP	Mixtral-8x22B-Instruct-v0.1	0.183300
cola-grammar	Non-finite VP Misc	Mixtral-8x22B-Instruct-v0.1	0.075300
cola-grammar	Deverbal	Mixtral-8x22B-Instruct-v0.1	0.027300
cola-grammar	Rel NP	Mixtral-8x22B-Instruct-v0.1	0.112500
cola-grammar	Trans NP	Mixtral-8x22B-Instruct-v0.1	0.034300
cola-grammar	Compx NP	Mixtral-8x22B-Instruct-v0.1	0.353400
cola-grammar	NNCompd	Mixtral-8x22B-Instruct-v0.1	0.109700
cola-grammar	Rel Adj	Mixtral-8x22B-Instruct-v0.1	0.055400
cola-grammar	Trans Adj	Mixtral-8x22B-Instruct-v0.1	0.103700
cola-grammar	Dislocation	Mixtral-8x22B-Instruct-v0.1	0.061500
cola-grammar	Info Struc	Mixtral-8x22B-Instruct-v0.1	0.183000
cola-grammar	Frag/Paren	Mixtral-8x22B-Instruct-v0.1	0.016700
cola-grammar	Coord	Mixtral-8x22B-Instruct-v0.1	0.761300
cola-grammar	Subordinate/Cond	Mixtral-8x22B-Instruct-v0.1	0.128100
cola-grammar	Ellipsis/Anaphor	Mixtral-8x22B-Instruct-v0.1	0.160900
cola-grammar	S-Adjunct	Mixtral-8x22B-Instruct-v0.1	0.070100
cola-grammar	Quantifier	Mixtral-8x22B-Instruct-v0.1	0.595600
cola-grammar	Partitive	Mixtral-8x22B-Instruct-v0.1	0.341900
cola-grammar	NPI/FCI	Mixtral-8x22B-Instruct-v0.1	0.416200
cola-grammar	Comparative	Mixtral-8x22B-Instruct-v0.1	0.693500
cola-grammar	Sem Violation	Mixtral-8x22B-Instruct-v0.1	0.074400
cola-grammar	Infl/Agr Violation	Mixtral-8x22B-Instruct-v0.1	0.226000
cola-grammar	Extra/Mising Expr	Mixtral-8x22B-Instruct-v0.1	0.143600
cola-grammar	Simple	gemini-1.5-flash-latest	0.335900
cola-grammar	Copula	gemini-1.5-flash-latest	0.404900
cola-grammar	Pred/SC	gemini-1.5-flash-latest	0.041000
cola-grammar	Result/Depictive	gemini-1.5-flash-latest	0.158500
cola-grammar	Particle	gemini-1.5-flash-latest	0.044200
cola-grammar	VP Adjunct	gemini-1.5-flash-latest	0.211300
cola-grammar	NP Adjunct	gemini-1.5-flash-latest	0.054500
cola-grammar	Temporal	gemini-1.5-flash-latest	0.501600
cola-grammar	Locative	gemini-1.5-flash-latest	0.398200
cola-grammar	Misc	gemini-1.5-flash-latest	0.071300
cola-grammar	Oblique	gemini-1.5-flash-latest	0.193800
cola-grammar	PP Arg-VP	gemini-1.5-flash-latest	0.287200
cola-grammar	PPArg-NP/AP	gemini-1.5-flash-latest	0.075900
cola-grammar	by-Phrase	gemini-1.5-flash-latest	0.816600
cola-grammar	Expletive	gemini-1.5-flash-latest	0.604400
cola-grammar	High Arity	gemini-1.5-flash-latest	0.192400
cola-grammar	Drop Arg	gemini-1.5-flash-latest	0.028100
cola-grammar	Add Arg	gemini-1.5-flash-latest	0.042800
cola-grammar	Passive	gemini-1.5-flash-latest	0.738900
cola-grammar	Imperative	gemini-1.5-flash-latest	0.635600

cola-grammar	Binding:Refl	gemini-1.5-flash-latest	0.632500
cola-grammar	Binding:Other	gemini-1.5-flash-latest	0.203500
cola-grammar	Matrix Q	gemini-1.5-flash-latest	0.498500
cola-grammar	Emb Q	gemini-1.5-flash-latest	0.538800
cola-grammar	Complex QP	gemini-1.5-flash-latest	0.524100
cola-grammar	RC	gemini-1.5-flash-latest	0.510000
cola-grammar	Island	gemini-1.5-flash-latest	0.104500
cola-grammar	CP Subj	gemini-1.5-flash-latest	0.051400
cola-grammar	CP Arg VP	gemini-1.5-flash-latest	0.243100
cola-grammar	CP Arg NP/AP	gemini-1.5-flash-latest	0.069000
cola-grammar	Non-finite CP	gemini-1.5-flash-latest	0.120500
cola-grammar	No C-izer	gemini-1.5-flash-latest	0.100800
cola-grammar	Deep Embed	gemini-1.5-flash-latest	0.090500
cola-grammar	Neg	gemini-1.5-flash-latest	0.442600
cola-grammar	Modal	gemini-1.5-flash-latest	0.895900
cola-grammar	Aux	gemini-1.5-flash-latest	0.322700
cola-grammar	Psuedo-Aux	gemini-1.5-flash-latest	0.042800
cola-grammar	Control	gemini-1.5-flash-latest	0.338500
cola-grammar	Raising	gemini-1.5-flash-latest	0.146200
cola-grammar	VP+Extract	gemini-1.5-flash-latest	0.100600
cola-grammar	VP arg-NP/AP	gemini-1.5-flash-latest	0.080200
cola-grammar	Non-finite VP Misc	gemini-1.5-flash-latest	0.037000
cola-grammar	Deverbal	gemini-1.5-flash-latest	0.007300
cola-grammar	Rel NP	gemini-1.5-flash-latest	0.066900
cola-grammar	Trans NP	gemini-1.5-flash-latest	0.017200
cola-grammar	Comp NP	gemini-1.5-flash-latest	0.199000
cola-grammar	NNCompd	gemini-1.5-flash-latest	0.086800
cola-grammar	Rel Adj	gemini-1.5-flash-latest	0.042000
cola-grammar	Trans Adj	gemini-1.5-flash-latest	0.057400
cola-grammar	Dislocation	gemini-1.5-flash-latest	0.073600
cola-grammar	Info Struc	gemini-1.5-flash-latest	0.161300
cola-grammar	Frag/Paren	gemini-1.5-flash-latest	0.027700
cola-grammar	Coord	gemini-1.5-flash-latest	0.651200
cola-grammar	Subordinate/Cond	gemini-1.5-flash-latest	0.152200
cola-grammar	Ellipsis/Anaphor	gemini-1.5-flash-latest	0.382800
cola-grammar	S-Adjunct	gemini-1.5-flash-latest	0.069900
cola-grammar	Quantifier	gemini-1.5-flash-latest	0.617200
cola-grammar	Partitive	gemini-1.5-flash-latest	0.722900
cola-grammar	NPI/FCI	gemini-1.5-flash-latest	0.188800
cola-grammar	Comparative	gemini-1.5-flash-latest	0.770100
cola-grammar	Sem Violation	gemini-1.5-flash-latest	0.049400
cola-grammar	Infl/Agr Violation	gemini-1.5-flash-latest	0.156100
cola-grammar	Extra/Mising Expr	gemini-1.5-flash-latest	0.105900
cola-grammar	Simple	c4ai-command-r-plus	-0.060800
cola-grammar	Copula	c4ai-command-r-plus	0.062900
cola-grammar	Pred/SC	c4ai-command-r-plus	-0.033800

cola-grammar	Result/Depictive	c4ai-command-r-plus	0.044000
cola-grammar	Particle	c4ai-command-r-plus	0.030900
cola-grammar	VP Adjunct	c4ai-command-r-plus	0.172200
cola-grammar	NP Adjunct	c4ai-command-r-plus	0.076500
cola-grammar	Temporal	c4ai-command-r-plus	0.137700
cola-grammar	Locative	c4ai-command-r-plus	0.075200
cola-grammar	Misc	c4ai-command-r-plus	0.130700
cola-grammar	Oblique	c4ai-command-r-plus	-0.011500
cola-grammar	PP Arg-VP	c4ai-command-r-plus	0.290100
cola-grammar	PPArg-NP/AP	c4ai-command-r-plus	0.109400
cola-grammar	by-Phrase	c4ai-command-r-plus	0.438100
cola-grammar	Expletive	c4ai-command-r-plus	0.347100
cola-grammar	High Arity	c4ai-command-r-plus	0.008800
cola-grammar	Drop Arg	c4ai-command-r-plus	0.004200
cola-grammar	Add Arg	c4ai-command-r-plus	0.017000
cola-grammar	Passive	c4ai-command-r-plus	0.103600
cola-grammar	Imperative	c4ai-command-r-plus	0.245800
cola-grammar	Binding:Refl	c4ai-command-r-plus	0.396500
cola-grammar	Binding:Other	c4ai-command-r-plus	0.018400
cola-grammar	Matrix Q	c4ai-command-r-plus	0.106300
cola-grammar	Emb Q	c4ai-command-r-plus	0.359100
cola-grammar	Complex QP	c4ai-command-r-plus	0.212500
cola-grammar	RC	c4ai-command-r-plus	0.389200
cola-grammar	Island	c4ai-command-r-plus	0.000000
cola-grammar	CP Subj	c4ai-command-r-plus	0.041000
cola-grammar	CP Arg VP	c4ai-command-r-plus	0.184900
cola-grammar	CP Arg NP/AP	c4ai-command-r-plus	0.076600
cola-grammar	Non-finite CP	c4ai-command-r-plus	0.031000
cola-grammar	No C-izer	c4ai-command-r-plus	0.023300
cola-grammar	Deep Embed	c4ai-command-r-plus	0.151700
cola-grammar	Neg	c4ai-command-r-plus	0.576600
cola-grammar	Modal	c4ai-command-r-plus	0.489500
cola-grammar	Aux	c4ai-command-r-plus	0.227700
cola-grammar	Psuedo-Aux	c4ai-command-r-plus	0.226200
cola-grammar	Control	c4ai-command-r-plus	0.121800
cola-grammar	Raising	c4ai-command-r-plus	0.024400
cola-grammar	VP+Extract	c4ai-command-r-plus	0.047700
cola-grammar	VP arg-NP/AP	c4ai-command-r-plus	0.076400
cola-grammar	Non-finite VP Misc	c4ai-command-r-plus	0.085500
cola-grammar	Deverbal	c4ai-command-r-plus	0.033000
cola-grammar	Rel NP	c4ai-command-r-plus	0.050700
cola-grammar	Trans NP	c4ai-command-r-plus	0.044800
cola-grammar	Compx NP	c4ai-command-r-plus	0.040000
cola-grammar	NNCompd	c4ai-command-r-plus	0.070300
cola-grammar	Rel Adj	c4ai-command-r-plus	0.022500
cola-grammar	Trans Adj	c4ai-command-r-plus	0.182700

cola-grammar	Dislocation	c4ai-command-r-plus	0.057000
cola-grammar	Info Struc	c4ai-command-r-plus	-0.012300
cola-grammar	Frag/Paren	c4ai-command-r-plus	0.076100
cola-grammar	Coord	c4ai-command-r-plus	0.389500
cola-grammar	Subordinate/Cond	c4ai-command-r-plus	0.172500
cola-grammar	Ellipsis/Anaphor	c4ai-command-r-plus	0.195400
cola-grammar	S-Adjunct	c4ai-command-r-plus	0.100200
cola-grammar	Quantifier	c4ai-command-r-plus	0.161000
cola-grammar	Partitive	c4ai-command-r-plus	-0.024000
cola-grammar	NPI/FCI	c4ai-command-r-plus	0.159300
cola-grammar	Comparative	c4ai-command-r-plus	0.336700
cola-grammar	Sem Violation	c4ai-command-r-plus	0.054400
cola-grammar	Infl/Agr Violation	c4ai-command-r-plus	0.019800
cola-grammar	Extra/Mising Expr	c4ai-command-r-plus	0.118700
<hr/>			
llmbar-natural	quality _s ingle _t urn	Meta-Llama-3.1-8B-Instruct	0.565600
llmbar-natural	quality _s ingle _t urn	Meta-Llama-3.1-70B-Instruct	0.797400
llmbar-natural	quality _s ingle _t urn	Mistral-7B-Instruct-v0.3	0.302500
llmbar-natural	quality _s ingle _t urn	OLMo-7B-0724-Instruct-hf	0.243600
llmbar-natural	quality _s ingle _t urn	Starling-LM-7B-alpha	0.284600
llmbar-natural	quality _s ingle _t urn	c4ai-command-r-v01	0.594800
llmbar-natural	quality _s ingle _t urn	Mixtral-8x7B-Instruct-v0.1	0.541500
llmbar-natural	quality _s ingle _t urn	gpt-4o	0.837900
llmbar-natural	quality _s ingle _t urn	gemini-1.5-flash-latest	0.789200
llmbar-natural	quality _s ingle _t urn	Mixtral-8x22B-Instruct-v0.1	0.718200
llmbar-natural	quality _s ingle _t urn	c4ai-command-r-plus	0.555600
<hr/>			
dices-990	safety	Meta-Llama-3.1-8B-Instruct	-0.105500
dices-990	safety	Meta-Llama-3.1-70B-Instruct	-0.171900
dices-990	safety	Mistral-7B-Instruct-v0.3	-0.121800
dices-990	safety	OLMo-7B-0724-Instruct-hf	0.001400
dices-990	safety	Starling-LM-7B-alpha	-0.048900
dices-990	safety	Mixtral-8x7B-Instruct-v0.1	-0.203200
dices-990	safety	Mixtral-8x22B-Instruct-v0.1	-0.160000
dices-990	safety	c4ai-command-r-v01	-0.018300
dices-990	safety	c4ai-command-r-plus	-0.092800
dices-990	safety	gemini-1.5-flash-latest	-0.124900
dices-990	safety	gpt-4o	-0.235100
<hr/>			
dices-350-expert	safety	Meta-Llama-3.1-8B-Instruct	0.011400
dices-350-expert	safety	Meta-Llama-3.1-70B-Instruct	-0.131400
dices-350-expert	safety	Mistral-7B-Instruct-v0.3	0.011400
dices-350-expert	safety	OLMo-7B-0724-Instruct-hf	-0.062900
dices-350-expert	safety	Starling-LM-7B-alpha	0.011400
dices-350-expert	safety	Mixtral-8x7B-Instruct-v0.1	-0.114300
dices-350-expert	safety	Mixtral-8x22B-Instruct-v0.1	-0.154300
dices-350-expert	safety	c4ai-command-r-v01	0.011400

dices-350-expert	safety	c4ai-command-r-plus	0.005700
dices-350-expert	safety	gemini-1.5-flash-latest	-0.028600
dices-350-expert	safety	gpt-4o	-0.200000
dices-350-crowdsourced	safety	Meta-Llama-3.1-8B-Instruct	-0.054900
dices-350-crowdsourced	safety	Meta-Llama-3.1-70B-Instruct	-0.178700
dices-350-crowdsourced	safety	Mistral-7B-Instruct-v0.3	-0.042600
dices-350-crowdsourced	safety	OLMo-7B-0724-Instruct-hf	-0.028000
dices-350-crowdsourced	safety	Starling-LM-7B-alpha	0.012500
dices-350-crowdsourced	safety	Mixtral-8x7B-Instruct-v0.1	-0.112000
dices-350-crowdsourced	safety	Mixtral-8x22B-Instruct-v0.1	-0.084300
dices-350-crowdsourced	safety	c4ai-command-r-v01	0.007700
dices-350-crowdsourced	safety	c4ai-command-r-plus	-0.077700
dices-350-crowdsourced	safety	gemini-1.5-flash-latest	-0.019300
dices-350-crowdsourced	safety	gpt-4o	-0.219200
llmbar-adversarial	quality _{single_tturn}	Meta-Llama-3.1-8B-Instruct	-0.177100
llmbar-adversarial	quality _{single_tturn}	Meta-Llama-3.1-70B-Instruct	0.464100
llmbar-adversarial	quality _{single_tturn}	Mistral-7B-Instruct-v0.3	-0.202600
llmbar-adversarial	quality _{single_tturn}	OLMo-7B-0724-Instruct-hf	-0.100100
llmbar-adversarial	quality _{single_tturn}	Starling-LM-7B-alpha	-0.115100
llmbar-adversarial	quality _{single_tturn}	c4ai-command-r-v01	-0.195700
llmbar-adversarial	quality _{single_tturn}	Mixtral-8x7B-Instruct-v0.1	0.060600
llmbar-adversarial	quality _{single_tturn}	gpt-4o	0.578500
llmbar-adversarial	quality _{single_tturn}	gemini-1.5-flash-latest	0.287000
llmbar-adversarial	quality _{single_tturn}	Mixtral-8x22B-Instruct-v0.1	0.203500
llmbar-adversarial	quality _{single_tturn}	c4ai-command-r-plus	0.107200
toxic-chat	toxicity	Meta-Llama-3.1-8B-Instruct	0.546700
toxic-chat	jailbreaking	Meta-Llama-3.1-8B-Instruct	0.137200
toxic-chat	toxicity	Meta-Llama-3.1-70B-Instruct	0.588900
toxic-chat	jailbreaking	Meta-Llama-3.1-70B-Instruct	0.226200
toxic-chat	toxicity	Mistral-7B-Instruct-v0.3	0.578400
toxic-chat	jailbreaking	Mistral-7B-Instruct-v0.3	0.321200
toxic-chat	toxicity	OLMo-7B-0724-Instruct-hf	0.390200
toxic-chat	jailbreaking	OLMo-7B-0724-Instruct-hf	0.213200
toxic-chat	toxicity	c4ai-command-r-v01	0.344300
toxic-chat	jailbreaking	c4ai-command-r-v01	0.051400
toxic-chat	toxicity	c4ai-command-r-plus	0.528700
toxic-chat	jailbreaking	c4ai-command-r-plus	0.038200
toxic-chat	toxicity	gemini-1.5-flash-latest	0.692000
toxic-chat	jailbreaking	gemini-1.5-flash-latest	0.201100
toxic-chat	toxicity	Mixtral-8x22B-Instruct-v0.1	0.645300
toxic-chat	jailbreaking	Mixtral-8x22B-Instruct-v0.1	0.256500
toxic-chat	toxicity	Mixtral-8x7B-Instruct-v0.1	0.448900
toxic-chat	jailbreaking	Mixtral-8x7B-Instruct-v0.1	0.275700
toxic-chat	toxicity	gpt-4o	0.738800

toxic-chat	jailbreaking	gpt-4o	0.235000
toxic-chat	toxicity	Starling-LM-7B-alpha	0.447700
toxic-chat	jailbreaking	Starling-LM-7B-alpha	0.083800

Dataset	Property	Model	Spearman's ρ
switchboard-acceptability	acceptability	Meta-Llama-3.1-8B-Instruct	0.282600
switchboard-acceptability	acceptability	Meta-Llama-3.1-70B-Instruct	0.450600
switchboard-acceptability	acceptability	Mistral-7B-Instruct-v0.3	0.524100
switchboard-acceptability	acceptability	OLMo-7B-0724-Instruct-hf	0.296800
switchboard-acceptability	acceptability	Starling-LM-7B-alpha	0.128500
switchboard-acceptability	acceptability	c4ai-command-r-v01	0.525400
switchboard-acceptability	acceptability	gemini-1.5-flash-latest	0.590100
switchboard-acceptability	acceptability	gpt-4o	0.655200
switchboard-acceptability	acceptability	Mixtral-8x7B-Instruct-v0.1	0.558000
switchboard-acceptability	acceptability	Mixtral-8x22B-Instruct-v0.1	0.628400
switchboard-acceptability	acceptability	c4ai-command-r-plus	0.364600
persona-chat	engaging	Meta-Llama-3.1-8B-Instruct	0.140600
persona-chat	maintains context	Meta-Llama-3.1-8B-Instruct	0.060600
persona-chat	natural	Meta-Llama-3.1-8B-Instruct	-0.118300
persona-chat	overall	Meta-Llama-3.1-8B-Instruct	-0.152200
persona-chat	engaging	Meta-Llama-3.1-70B-Instruct	0.227100
persona-chat	maintains context	Meta-Llama-3.1-70B-Instruct	0.060700
persona-chat	natural	Meta-Llama-3.1-70B-Instruct	-0.186000
persona-chat	overall	Meta-Llama-3.1-70B-Instruct	-0.175600
persona-chat	engaging	Mistral-7B-Instruct-v0.3	0.139600
persona-chat	maintains context	Mistral-7B-Instruct-v0.3	-0.111300
persona-chat	natural	Mistral-7B-Instruct-v0.3	-0.134800
persona-chat	overall	Mistral-7B-Instruct-v0.3	-0.267700
persona-chat	engaging	OLMo-7B-0724-Instruct-hf	0.077900
persona-chat	maintains context	OLMo-7B-0724-Instruct-hf	0.028200
persona-chat	natural	OLMo-7B-0724-Instruct-hf	-0.107500
persona-chat	overall	OLMo-7B-0724-Instruct-hf	-0.222700
persona-chat	engaging	Starling-LM-7B-alpha	0.138400
persona-chat	maintains context	Starling-LM-7B-alpha	-0.016300
persona-chat	natural	Starling-LM-7B-alpha	-0.126300
persona-chat	overall	Starling-LM-7B-alpha	0.130200
persona-chat	engaging	c4ai-command-r-v01	0.237700
persona-chat	maintains context	c4ai-command-r-v01	0.088600
persona-chat	natural	c4ai-command-r-v01	0.123600
persona-chat	overall	c4ai-command-r-v01	-0.232100
persona-chat	engaging	gpt-4o	0.300100
persona-chat	maintains context	gpt-4o	0.329600
persona-chat	natural	gpt-4o	0.134200
persona-chat	overall	gpt-4o	0.127700
persona-chat	engaging	gemini-1.5-flash-latest	0.228800
persona-chat	maintains context	gemini-1.5-flash-latest	0.105700
persona-chat	natural	gemini-1.5-flash-latest	0.014000
persona-chat	overall	gemini-1.5-flash-latest	0.058000

persona-chat	engaging	Mixtral-8x7B-Instruct-v0.1	0.237700
persona-chat	maintains context	Mixtral-8x7B-Instruct-v0.1	-0.004900
persona-chat	natural	Mixtral-8x7B-Instruct-v0.1	-0.109500
persona-chat	overall	Mixtral-8x7B-Instruct-v0.1	-0.038600
persona-chat	engaging	Mixtral-8x22B-Instruct-v0.1	0.294400
persona-chat	maintains context	Mixtral-8x22B-Instruct-v0.1	0.187300
persona-chat	natural	Mixtral-8x22B-Instruct-v0.1	0.059100
persona-chat	overall	Mixtral-8x22B-Instruct-v0.1	0.104700
persona-chat	engaging	c4ai-command-r-plus	0.140600
persona-chat	maintains context	c4ai-command-r-plus	0.018200
persona-chat	natural	c4ai-command-r-plus	-0.100500
persona-chat	overall	c4ai-command-r-plus	0.202000
newsroom	Informativeness	Meta-Llama-3.1-8B-Instruct	0.502700
newsroom	Relevance	Meta-Llama-3.1-8B-Instruct	0.439400
newsroom	Fluency	Meta-Llama-3.1-8B-Instruct	0.415100
newsroom	Coherence	Meta-Llama-3.1-8B-Instruct	0.433900
newsroom	Informativeness	Meta-Llama-3.1-70B-Instruct	0.559900
newsroom	Relevance	Meta-Llama-3.1-70B-Instruct	0.572000
newsroom	Fluency	Meta-Llama-3.1-70B-Instruct	0.629900
newsroom	Coherence	Meta-Llama-3.1-70B-Instruct	0.615400
newsroom	Informativeness	Mistral-7B-Instruct-v0.3	0.277000
newsroom	Relevance	Mistral-7B-Instruct-v0.3	0.296500
newsroom	Fluency	Mistral-7B-Instruct-v0.3	0.171600
newsroom	Coherence	Mistral-7B-Instruct-v0.3	0.278200
newsroom	Informativeness	OLMo-7B-0724-Instruct-hf	0.029200
newsroom	Relevance	OLMo-7B-0724-Instruct-hf	0.005300
newsroom	Fluency	OLMo-7B-0724-Instruct-hf	-0.058200
newsroom	Coherence	OLMo-7B-0724-Instruct-hf	-0.027600
newsroom	Informativeness	Starling-LM-7B-alpha	0.282000
newsroom	Relevance	Starling-LM-7B-alpha	0.274100
newsroom	Fluency	Starling-LM-7B-alpha	0.116000
newsroom	Coherence	Starling-LM-7B-alpha	0.165400
newsroom	Informativeness	c4ai-command-r-v01	0.096300
newsroom	Relevance	c4ai-command-r-v01	0.162000
newsroom	Fluency	c4ai-command-r-v01	0.216600
newsroom	Coherence	c4ai-command-r-v01	0.164600
newsroom	Informativeness	gemini-1.5-flash-latest	0.519000
newsroom	Relevance	gemini-1.5-flash-latest	0.565400
newsroom	Fluency	gemini-1.5-flash-latest	0.591300
newsroom	Coherence	gemini-1.5-flash-latest	0.538100
newsroom	Informativeness	gpt-4o	0.615000
newsroom	Relevance	gpt-4o	0.556100
newsroom	Fluency	gpt-4o	0.593100
newsroom	Coherence	gpt-4o	0.594100
newsroom	Informativeness	Mixtral-8x7B-Instruct-v0.1	0.592100

newsroom	Relevance	Mixtral-8x7B-Instruct-v0.1	0.461200
newsroom	Fluency	Mixtral-8x7B-Instruct-v0.1	0.444300
newsroom	Coherence	Mixtral-8x7B-Instruct-v0.1	0.515600
newsroom	Informativeness	Mixtral-8x22B-Instruct-v0.1	0.492000
newsroom	Relevance	Mixtral-8x22B-Instruct-v0.1	0.466100
newsroom	Fluency	Mixtral-8x22B-Instruct-v0.1	0.410800
newsroom	Coherence	Mixtral-8x22B-Instruct-v0.1	0.393600
newsroom	Informativeness	c4ai-command-r-plus	0.411100
newsroom	Relevance	c4ai-command-r-plus	0.402900
newsroom	Fluency	c4ai-command-r-plus	0.298100
newsroom	Coherence	c4ai-command-r-plus	0.327700
<hr/>			
roscoe-cosmos	Overall Quality	Meta-Llama-3.1-8B-Instruct	0.624300
roscoe-cosmos	Coherency	Meta-Llama-3.1-8B-Instruct	0.259400
roscoe-cosmos	Overall Quality	Meta-Llama-3.1-70B-Instruct	0.677900
roscoe-cosmos	Coherency	Meta-Llama-3.1-70B-Instruct	0.424200
roscoe-cosmos	Overall Quality	Mistral-7B-Instruct-v0.3	0.714400
roscoe-cosmos	Coherency	Mistral-7B-Instruct-v0.3	0.426800
roscoe-cosmos	Overall Quality	OLMo-7B-0724-Instruct-hf	0.654300
roscoe-cosmos	Coherency	OLMo-7B-0724-Instruct-hf	0.316300
roscoe-cosmos	Overall Quality	Starling-LM-7B-alpha	0.102300
roscoe-cosmos	Coherency	Starling-LM-7B-alpha	0.154000
roscoe-cosmos	Overall Quality	c4ai-command-r-v01	0.602300
roscoe-cosmos	Coherency	c4ai-command-r-v01	0.366300
roscoe-cosmos	Overall Quality	gemini-1.5-flash-latest	0.686500
roscoe-cosmos	Coherency	gemini-1.5-flash-latest	0.446200
roscoe-cosmos	Overall Quality	gpt-4o	0.694100
roscoe-cosmos	Coherency	gpt-4o	0.441400
roscoe-cosmos	Overall Quality	Mixtral-8x7B-Instruct-v0.1	0.680300
roscoe-cosmos	Coherency	Mixtral-8x7B-Instruct-v0.1	0.386200
roscoe-cosmos	Overall Quality	Mixtral-8x22B-Instruct-v0.1	0.626900
roscoe-cosmos	Coherency	Mixtral-8x22B-Instruct-v0.1	0.398200
roscoe-cosmos	Overall Quality	c4ai-command-r-plus	0.503800
roscoe-cosmos	Coherency	c4ai-command-r-plus	0.147400
<hr/>			
roscoe-drop	Overall Quality	Meta-Llama-3.1-8B-Instruct	0.297400
roscoe-drop	Coherency	Meta-Llama-3.1-8B-Instruct	0.158100
roscoe-drop	Overall Quality	Meta-Llama-3.1-70B-Instruct	0.696000
roscoe-drop	Coherency	Meta-Llama-3.1-70B-Instruct	0.476500
roscoe-drop	Overall Quality	Mistral-7B-Instruct-v0.3	0.371200
roscoe-drop	Coherency	Mistral-7B-Instruct-v0.3	0.064000
roscoe-drop	Overall Quality	OLMo-7B-0724-Instruct-hf	0.301300
roscoe-drop	Coherency	OLMo-7B-0724-Instruct-hf	0.008000
roscoe-drop	Overall Quality	Starling-LM-7B-alpha	0.283100
roscoe-drop	Coherency	Starling-LM-7B-alpha	0.043300
roscoe-drop	Overall Quality	c4ai-command-r-v01	0.496500
roscoe-drop	Coherency	c4ai-command-r-v01	0.241500

roscoe-drop	Overall Quality	gpt-4o	0.729600
roscoe-drop	Coherency	gpt-4o	0.419600
roscoe-drop	Overall Quality	gemini-1.5-flash-latest	0.530400
roscoe-drop	Coherency	gemini-1.5-flash-latest	0.346100
roscoe-drop	Overall Quality	Mixtral-8x7B-Instruct-v0.1	0.405100
roscoe-drop	Coherency	Mixtral-8x7B-Instruct-v0.1	0.238900
roscoe-drop	Overall Quality	Mixtral-8x22B-Instruct-v0.1	0.544200
roscoe-drop	Coherency	Mixtral-8x22B-Instruct-v0.1	0.333100
roscoe-drop	Overall Quality	c4ai-command-r-plus	0.370100
roscoe-drop	Coherency	c4ai-command-r-plus	0.056900
dailydialog-acceptability	acceptability	Meta-Llama-3.1-8B-Instruct	0.612700
dailydialog-acceptability	acceptability	Meta-Llama-3.1-70B-Instruct	0.604500
dailydialog-acceptability	acceptability	Mistral-7B-Instruct-v0.3	0.476600
dailydialog-acceptability	acceptability	OLMo-7B-0724-Instruct-hf	0.065600
dailydialog-acceptability	acceptability	Starling-LM-7B-alpha	0.086700
dailydialog-acceptability	acceptability	c4ai-command-r-v01	0.227500
dailydialog-acceptability	acceptability	gpt-4o	0.686400
dailydialog-acceptability	acceptability	gemini-1.5-flash-latest	0.633900
dailydialog-acceptability	acceptability	Mixtral-8x7B-Instruct-v0.1	0.625700
dailydialog-acceptability	acceptability	Mixtral-8x22B-Instruct-v0.1	0.551200
dailydialog-acceptability	acceptability	c4ai-command-r-plus	0.524600
roscoe-esnli	Overall Quality	Meta-Llama-3.1-8B-Instruct	0.297400
roscoe-esnli	Coherency	Meta-Llama-3.1-8B-Instruct	0.073900
roscoe-esnli	Overall Quality	Meta-Llama-3.1-70B-Instruct	0.515600
roscoe-esnli	Coherency	Meta-Llama-3.1-70B-Instruct	0.294000
roscoe-esnli	Overall Quality	Mistral-7B-Instruct-v0.3	0.405000
roscoe-esnli	Coherency	Mistral-7B-Instruct-v0.3	0.233100
roscoe-esnli	Overall Quality	OLMo-7B-0724-Instruct-hf	0.226200
roscoe-esnli	Coherency	OLMo-7B-0724-Instruct-hf	-0.008000
roscoe-esnli	Overall Quality	Starling-LM-7B-alpha	0.156600
roscoe-esnli	Coherency	Starling-LM-7B-alpha	0.065100
roscoe-esnli	Overall Quality	c4ai-command-r-v01	0.427000
roscoe-esnli	Coherency	c4ai-command-r-v01	0.127600
roscoe-esnli	Overall Quality	gpt-4o	0.657200
roscoe-esnli	Coherency	gpt-4o	0.315200
roscoe-esnli	Overall Quality	Mixtral-8x7B-Instruct-v0.1	0.401600
roscoe-esnli	Coherency	Mixtral-8x7B-Instruct-v0.1	0.228400
roscoe-esnli	Overall Quality	gemini-1.5-flash-latest	0.495000
roscoe-esnli	Coherency	gemini-1.5-flash-latest	0.203200
roscoe-esnli	Overall Quality	Mixtral-8x22B-Instruct-v0.1	0.507700
roscoe-esnli	Coherency	Mixtral-8x22B-Instruct-v0.1	0.260500
roscoe-esnli	Overall Quality	c4ai-command-r-plus	0.145300
roscoe-esnli	Coherency	c4ai-command-r-plus	0.025600
roscoe-gsm8k	Overall Quality	Meta-Llama-3.1-8B-Instruct	0.860400

roscoe-gsm8k	Coherency	Meta-Llama-3.1-8B-Instruct	0.652100
roscoe-gsm8k	Overall Quality	Meta-Llama-3.1-70B-Instruct	0.906300
roscoe-gsm8k	Coherency	Meta-Llama-3.1-70B-Instruct	0.748500
roscoe-gsm8k	Overall Quality	Mistral-7B-Instruct-v0.3	0.759800
roscoe-gsm8k	Coherency	Mistral-7B-Instruct-v0.3	0.500800
roscoe-gsm8k	Overall Quality	OLMo-7B-0724-Instruct-hf	0.151700
roscoe-gsm8k	Coherency	OLMo-7B-0724-Instruct-hf	0.056100
roscoe-gsm8k	Overall Quality	Starling-LM-7B-alpha	0.551200
roscoe-gsm8k	Coherency	Starling-LM-7B-alpha	0.369300
roscoe-gsm8k	Overall Quality	c4ai-command-r-v01	0.755600
roscoe-gsm8k	Coherency	c4ai-command-r-v01	0.645700
roscoe-gsm8k	Overall Quality	gpt-4o	0.907800
roscoe-gsm8k	Coherency	gpt-4o	0.733100
roscoe-gsm8k	Overall Quality	Mixtral-8x7B-Instruct-v0.1	0.882700
roscoe-gsm8k	Coherency	Mixtral-8x7B-Instruct-v0.1	0.702900
roscoe-gsm8k	Overall Quality	gemini-1.5-flash-latest	0.898400
roscoe-gsm8k	Coherency	gemini-1.5-flash-latest	0.725300
roscoe-gsm8k	Overall Quality	Mixtral-8x22B-Instruct-v0.1	0.910900
roscoe-gsm8k	Coherency	Mixtral-8x22B-Instruct-v0.1	0.706100
roscoe-gsm8k	Overall Quality	c4ai-command-r-plus	0.823000
roscoe-gsm8k	Coherency	c4ai-command-r-plus	0.539100
<hr/>			
topical-chat	engaging	Meta-Llama-3.1-8B-Instruct	0.213500
topical-chat	maintains context	Meta-Llama-3.1-8B-Instruct	0.230900
topical-chat	natural	Meta-Llama-3.1-8B-Instruct	0.195300
topical-chat	overall	Meta-Llama-3.1-8B-Instruct	-0.047800
topical-chat	engaging	Meta-Llama-3.1-70B-Instruct	0.397900
topical-chat	maintains context	Meta-Llama-3.1-70B-Instruct	0.199000
topical-chat	natural	Meta-Llama-3.1-70B-Instruct	0.331000
topical-chat	overall	Meta-Llama-3.1-70B-Instruct	0.207000
topical-chat	engaging	Mistral-7B-Instruct-v0.3	0.298700
topical-chat	maintains context	Mistral-7B-Instruct-v0.3	0.273800
topical-chat	natural	Mistral-7B-Instruct-v0.3	0.160800
topical-chat	overall	Mistral-7B-Instruct-v0.3	0.433700
topical-chat	engaging	OLMo-7B-0724-Instruct-hf	0.242200
topical-chat	maintains context	OLMo-7B-0724-Instruct-hf	-0.126000
topical-chat	natural	OLMo-7B-0724-Instruct-hf	-0.074800
topical-chat	overall	OLMo-7B-0724-Instruct-hf	0.267100
topical-chat	engaging	Starling-LM-7B-alpha	0.136300
topical-chat	maintains context	Starling-LM-7B-alpha	0.105700
topical-chat	natural	Starling-LM-7B-alpha	-0.018000
topical-chat	overall	Starling-LM-7B-alpha	0.355100
topical-chat	engaging	c4ai-command-r-v01	0.076900
topical-chat	maintains context	c4ai-command-r-v01	0.158300
topical-chat	natural	c4ai-command-r-v01	0.050700
topical-chat	overall	c4ai-command-r-v01	-0.018700

topical-chat	engaging	gemini-1.5-flash-latest	0.029200
topical-chat	maintains context	gemini-1.5-flash-latest	0.314400
topical-chat	natural	gemini-1.5-flash-latest	0.182000
topical-chat	overall	gemini-1.5-flash-latest	0.162600
topical-chat	engaging	gpt-4o	0.245900
topical-chat	maintains context	gpt-4o	0.278400
topical-chat	natural	gpt-4o	0.279900
topical-chat	overall	gpt-4o	0.217600
topical-chat	engaging	Mixtral-8x7B-Instruct-v0.1	0.083800
topical-chat	maintains context	Mixtral-8x7B-Instruct-v0.1	0.472100
topical-chat	natural	Mixtral-8x7B-Instruct-v0.1	0.105200
topical-chat	overall	Mixtral-8x7B-Instruct-v0.1	0.165100
topical-chat	engaging	Mixtral-8x22B-Instruct-v0.1	0.099400
topical-chat	maintains context	Mixtral-8x22B-Instruct-v0.1	0.189300
topical-chat	natural	Mixtral-8x22B-Instruct-v0.1	0.101900
topical-chat	overall	Mixtral-8x22B-Instruct-v0.1	0.136900
topical-chat	engaging	c4ai-command-r-plus	0.074700
topical-chat	maintains context	c4ai-command-r-plus	0.192800
topical-chat	natural	c4ai-command-r-plus	0.166600
topical-chat	overall	c4ai-command-r-plus	0.116400
<hr/>			
summeval	coherence	Meta-Llama-3.1-8B-Instruct	0.263800
summeval	consistency	Meta-Llama-3.1-8B-Instruct	0.330200
summeval	fluency	Meta-Llama-3.1-8B-Instruct	0.183400
summeval	relevance	Meta-Llama-3.1-8B-Instruct	0.401800
summeval	coherence	Meta-Llama-3.1-70B-Instruct	0.552900
summeval	consistency	Meta-Llama-3.1-70B-Instruct	0.471500
summeval	fluency	Meta-Llama-3.1-70B-Instruct	0.227500
summeval	relevance	Meta-Llama-3.1-70B-Instruct	0.499600
summeval	coherence	Mistral-7B-Instruct-v0.3	0.478500
summeval	consistency	Mistral-7B-Instruct-v0.3	0.458700
summeval	fluency	Mistral-7B-Instruct-v0.3	0.229400
summeval	relevance	Mistral-7B-Instruct-v0.3	0.451000
summeval	coherence	OLMo-7B-0724-Instruct-hf	0.040000
summeval	consistency	OLMo-7B-0724-Instruct-hf	0.094300
summeval	fluency	OLMo-7B-0724-Instruct-hf	0.048600
summeval	relevance	OLMo-7B-0724-Instruct-hf	0.063900
summeval	coherence	Starling-LM-7B-alpha	0.202300
summeval	consistency	Starling-LM-7B-alpha	0.116200
summeval	fluency	Starling-LM-7B-alpha	0.103400
summeval	relevance	Starling-LM-7B-alpha	0.175100
summeval	coherence	c4ai-command-r-v01	0.211000
summeval	consistency	c4ai-command-r-v01	0.060500
summeval	fluency	c4ai-command-r-v01	0.109000
summeval	relevance	c4ai-command-r-v01	0.155500
summeval	coherence	gpt-4o	0.384500

summeval	consistency	gpt-4o	0.350800
summeval	fluency	gpt-4o	0.267200
summeval	relevance	gpt-4o	0.406500
summeval	coherence	Mixtral-8x7B-Instruct-v0.1	0.464100
summeval	consistency	Mixtral-8x7B-Instruct-v0.1	0.503700
summeval	fluency	Mixtral-8x7B-Instruct-v0.1	0.453200
summeval	relevance	Mixtral-8x7B-Instruct-v0.1	0.479400
summeval	coherence	gemini-1.5-flash-latest	0.409800
summeval	consistency	gemini-1.5-flash-latest	0.367000
summeval	fluency	gemini-1.5-flash-latest	0.357900
summeval	relevance	gemini-1.5-flash-latest	0.394600
summeval	coherence	Mixtral-8x22B-Instruct-v0.1	0.581400
summeval	consistency	Mixtral-8x22B-Instruct-v0.1	0.616400
summeval	fluency	Mixtral-8x22B-Instruct-v0.1	0.445000
summeval	relevance	Mixtral-8x22B-Instruct-v0.1	0.521900
summeval	coherence	c4ai-command-r-plus	0.168100
summeval	consistency	c4ai-command-r-plus	0.204900
summeval	fluency	c4ai-command-r-plus	0.266300
summeval	relevance	c4ai-command-r-plus	0.136600
<hr/>			
wmt-human-en-de	quality	Meta-Llama-3.1-8B-Instruct	0.114000
wmt-human-en-de	quality	Meta-Llama-3.1-70B-Instruct	0.365700
wmt-human-en-de	quality	Mistral-7B-Instruct-v0.3	0.358100
wmt-human-en-de	quality	OLMo-7B-0724-Instruct-hf	-0.027600
wmt-human-en-de	quality	Starling-LM-7B-alpha	0.151500
wmt-human-en-de	quality	c4ai-command-r-v01	0.150200
wmt-human-en-de	quality	Mixtral-8x7B-Instruct-v0.1	0.198500
wmt-human-en-de	quality	gpt-4o	0.633600
wmt-human-en-de	quality	gemini-1.5-flash-latest	0.459800
wmt-human-en-de	quality	Mixtral-8x22B-Instruct-v0.1	0.505800
wmt-human-en-de	quality	c4ai-command-r-plus	0.424500
<hr/>			
wmt-human-zh-en	quality	Meta-Llama-3.1-8B-Instruct	0.144000
wmt-human-zh-en	quality	Meta-Llama-3.1-70B-Instruct	0.394000
wmt-human-zh-en	quality	Mistral-7B-Instruct-v0.3	0.394800
wmt-human-zh-en	quality	OLMo-7B-0724-Instruct-hf	0.013400
wmt-human-zh-en	quality	Starling-LM-7B-alpha	0.146300
wmt-human-zh-en	quality	c4ai-command-r-v01	0.149300
wmt-human-zh-en	quality	Mixtral-8x7B-Instruct-v0.1	0.251100
wmt-human-zh-en	quality	gpt-4o	0.543900
wmt-human-zh-en	quality	gemini-1.5-flash-latest	0.407900
wmt-human-zh-en	quality	Mixtral-8x22B-Instruct-v0.1	0.476000
wmt-human-zh-en	quality	c4ai-command-r-plus	0.421000
<hr/>			
wmt-23-en-de	quality	Meta-Llama-3.1-70B-Instruct	0.136600
wmt-23-en-de	quality	Mistral-7B-Instruct-v0.3	0.183500
wmt-23-en-de	quality	OLMo-7B-0724-Instruct-hf	-0.052500

wmt-23-en-de	quality	Meta-Llama-3.1-8B-Instruct	0.075700
wmt-23-en-de	quality	Starling-LM-7B-alpha	-0.091300
wmt-23-en-de	quality	Mixtral-8x7B-Instruct-v0.1	0.172200
wmt-23-en-de	quality	c4ai-command-r-v01	0.194500
wmt-23-en-de	quality	Mixtral-8x22B-Instruct-v0.1	0.231800
wmt-23-en-de	quality	c4ai-command-r-plus	0.220100
wmt-23-en-de	quality	gpt-4o	0.216800
wmt-23-en-de	quality	gemini-1.5-flash-latest	0.161100
wmt-23-zh-en	quality	Meta-Llama-3.1-70B-Instruct	0.144600
wmt-23-zh-en	quality	Mistral-7B-Instruct-v0.3	0.148700
wmt-23-zh-en	quality	OLMo-7B-0724-Instruct-hf	0.009100
wmt-23-zh-en	quality	Meta-Llama-3.1-8B-Instruct	0.024600
wmt-23-zh-en	quality	Starling-LM-7B-alpha	0.014200
wmt-23-zh-en	quality	Mixtral-8x7B-Instruct-v0.1	0.150100
wmt-23-zh-en	quality	c4ai-command-r-v01	0.135100
wmt-23-zh-en	quality	Mixtral-8x22B-Instruct-v0.1	0.186300
wmt-23-zh-en	quality	c4ai-command-r-plus	0.151600
wmt-23-zh-en	quality	gpt-4o	0.169000
wmt-23-zh-en	quality	gemini-1.5-flash-latest	0.140300
recipe-crowd-sourcing-data	grammar	Meta-Llama-3.1-8B-Instruct	0.453600
recipe-crowd-sourcing-data	fluency	Meta-Llama-3.1-8B-Instruct	0.194200
recipe-crowd-sourcing-data	verbosity	Meta-Llama-3.1-8B-Instruct	0.347500
recipe-crowd-sourcing-data	structure	Meta-Llama-3.1-8B-Instruct	0.301300
recipe-crowd-sourcing-data	success	Meta-Llama-3.1-8B-Instruct	0.325200
recipe-crowd-sourcing-data	overall	Meta-Llama-3.1-8B-Instruct	0.415800
recipe-crowd-sourcing-data	grammar	Meta-Llama-3.1-70B-Instruct	0.734400
recipe-crowd-sourcing-data	fluency	Meta-Llama-3.1-70B-Instruct	0.696600
recipe-crowd-sourcing-data	verbosity	Meta-Llama-3.1-70B-Instruct	0.719500
recipe-crowd-sourcing-data	structure	Meta-Llama-3.1-70B-Instruct	0.555300
recipe-crowd-sourcing-data	success	Meta-Llama-3.1-70B-Instruct	0.605900
recipe-crowd-sourcing-data	overall	Meta-Llama-3.1-70B-Instruct	0.667100
recipe-crowd-sourcing-data	grammar	Mistral-7B-Instruct-v0.3	0.301500
recipe-crowd-sourcing-data	fluency	Mistral-7B-Instruct-v0.3	NaN
recipe-crowd-sourcing-data	verbosity	Mistral-7B-Instruct-v0.3	0.316900
recipe-crowd-sourcing-data	structure	Mistral-7B-Instruct-v0.3	0.334500
recipe-crowd-sourcing-data	success	Mistral-7B-Instruct-v0.3	0.282500
recipe-crowd-sourcing-data	overall	Mistral-7B-Instruct-v0.3	0.140000
recipe-crowd-sourcing-data	grammar	OLMo-7B-0724-Instruct-hf	0.001100
recipe-crowd-sourcing-data	fluency	OLMo-7B-0724-Instruct-hf	0.012300
recipe-crowd-sourcing-data	verbosity	OLMo-7B-0724-Instruct-hf	0.170800
recipe-crowd-sourcing-data	structure	OLMo-7B-0724-Instruct-hf	0.080900
recipe-crowd-sourcing-data	success	OLMo-7B-0724-Instruct-hf	0.131200
recipe-crowd-sourcing-data	overall	OLMo-7B-0724-Instruct-hf	0.187300
recipe-crowd-sourcing-data	grammar	Starling-LM-7B-alpha	0.287000
recipe-crowd-sourcing-data	fluency	Starling-LM-7B-alpha	0.046900

recipe-crowd-sourcing-data	verbosity	Starling-LM-7B-alpha	-0.088400
recipe-crowd-sourcing-data	structure	Starling-LM-7B-alpha	-0.182700
recipe-crowd-sourcing-data	success	Starling-LM-7B-alpha	-0.017600
recipe-crowd-sourcing-data	overall	Starling-LM-7B-alpha	0.189500
recipe-crowd-sourcing-data	grammar	Mixtral-8x7B-Instruct-v0.1	0.693600
recipe-crowd-sourcing-data	fluency	Mixtral-8x7B-Instruct-v0.1	0.694300
recipe-crowd-sourcing-data	verbosity	Mixtral-8x7B-Instruct-v0.1	0.773200
recipe-crowd-sourcing-data	structure	Mixtral-8x7B-Instruct-v0.1	0.116600
recipe-crowd-sourcing-data	success	Mixtral-8x7B-Instruct-v0.1	0.598400
recipe-crowd-sourcing-data	overall	Mixtral-8x7B-Instruct-v0.1	0.550000
recipe-crowd-sourcing-data	grammar	c4ai-command-r-v01	-0.053700
recipe-crowd-sourcing-data	fluency	c4ai-command-r-v01	-0.088800
recipe-crowd-sourcing-data	verbosity	c4ai-command-r-v01	0.286600
recipe-crowd-sourcing-data	structure	c4ai-command-r-v01	0.042900
recipe-crowd-sourcing-data	success	c4ai-command-r-v01	0.435100
recipe-crowd-sourcing-data	overall	c4ai-command-r-v01	-0.284500
recipe-crowd-sourcing-data	grammar	Mixtral-8x22B-Instruct-v0.1	0.774400
recipe-crowd-sourcing-data	fluency	Mixtral-8x22B-Instruct-v0.1	0.608400
recipe-crowd-sourcing-data	verbosity	Mixtral-8x22B-Instruct-v0.1	0.571000
recipe-crowd-sourcing-data	structure	Mixtral-8x22B-Instruct-v0.1	0.625500
recipe-crowd-sourcing-data	success	Mixtral-8x22B-Instruct-v0.1	0.327400
recipe-crowd-sourcing-data	overall	Mixtral-8x22B-Instruct-v0.1	0.680000
recipe-crowd-sourcing-data	grammar	c4ai-command-r-plus	0.315800
recipe-crowd-sourcing-data	fluency	c4ai-command-r-plus	0.546000
recipe-crowd-sourcing-data	verbosity	c4ai-command-r-plus	0.442300
recipe-crowd-sourcing-data	structure	c4ai-command-r-plus	0.286300
recipe-crowd-sourcing-data	success	c4ai-command-r-plus	-0.207100
recipe-crowd-sourcing-data	overall	c4ai-command-r-plus	0.530900
recipe-crowd-sourcing-data	grammar	gemini-1.5-flash-latest	0.693700
recipe-crowd-sourcing-data	fluency	gemini-1.5-flash-latest	0.644000
recipe-crowd-sourcing-data	verbosity	gemini-1.5-flash-latest	0.836400
recipe-crowd-sourcing-data	structure	gemini-1.5-flash-latest	0.655100
recipe-crowd-sourcing-data	success	gemini-1.5-flash-latest	0.610800
recipe-crowd-sourcing-data	overall	gemini-1.5-flash-latest	0.590800
recipe-crowd-sourcing-data	grammar	gpt-4o	0.848400
recipe-crowd-sourcing-data	fluency	gpt-4o	0.718600
recipe-crowd-sourcing-data	verbosity	gpt-4o	0.786600
recipe-crowd-sourcing-data	structure	gpt-4o	0.813500
recipe-crowd-sourcing-data	success	gpt-4o	0.763200
recipe-crowd-sourcing-data	overall	gpt-4o	0.760500
