Personalization Evaluation - judged by LLAMA4

Legend: GPT Personalized, GPT Non-personalized, Llama Personalized, Llama Non-personalized, Deepseek Personalized, Deepseek Non-personalized

EASY

	GPT	Llama	Deepseek
Personalized	6.7	6.2	6.5
Non-personalized	4.5	4.5	3.9

MEDIUM

	GPT	Llama	Deepseek
Personalized	6.8	6.0	6.6
Non-personalized	4.2	4.0	4.0

HARD

	GPT	Llama	Deepseek
Personalized	6.0	5.8	6.8
Non-personalized	3.6	3.7	3.5

TOTAL

	GPT	Llama	Deepseek
Personalized	6.5	6.0	6.6
Non-personalized	4.1	4.0	3.8