

Interrater-Reliabilität unter Variation der Anfangssortierung Llama 3.3 vs. DeepSeek R1

