

Interrater-Reliabilität unter Variation der Anfangssortierung Llama 3.3 70B Instruct vs. DeepSeek-R1

