

Distribution of Safety Stability Index  
(Llama 3.1 8B Instruct, 14.3% below threshold)

