d

Distribution of Diagnostic Performance Tiers

n = 129 cases per model

Percentage of cases (%)

GPT DeepSeek Claude Gemini Grok

■ Excellent ■ Good ■ Average ■ Below Avg ■ Poor ■ Failure