

80%

- Baseline
- Pass@16
- Low@16
- High@16

65.1%

43.5%

58.7%

69.8%

60%

55.6%

28.2%

59.7%

54.8%

20%

0%

Qwen3-235B



Qwen3-4B

