# Open-weight models on FormationEval v0.1 (32 models)



| Model | Accuracy (%) |
|---|---|
| glm-4.7 | 98.6% |
| deepseek-r1 | 96.2% |
| deepseek-v3.2 | 94.9% |
| llama-4-scout | 93.1% |
| mistral-medium-3.1 | 93.1% |
| qwen3-235b-a22b-2507 | 93.1% |
| qwen3-30b-a3b-thinking-2507 | 93.1% |
| qwen3-14b | 92.9% |
| qwen3-32b | 92.1% |
| gpt-oss-120b | 90.7% |
| qwen3-vl-8b-thinking | 90.3% |
| mistral-small-3.2-24b-instruct | 89.3% |
| gpt-oss-20b | 89.3% |
| mistral-small-24b-instruct-2501 | 88.7% |
| qwen3-8b | 88.7% |
| phi-4-reasoning-plus | 87.7% |
| ministral-14b-2512 | 87.7% |
| qwen3-vl-8b-instruct | 87.5% |
| glm-4-32b | 87.3% |
| ministral-8b-2512 | 86.9% |
| gemma-3-27b-it | 85.3% |
| deepseek-r1-0528-qwen3-8b | 85.1% |
| gemma-3-12b-it | 82.2% |
| nemotron-nano-9b-v2 | 79.6% |
| ministral-3b-2512 | 79.2% |
| mistral-nemo | 78.8% |
| nemotron-3-nano-30b-a3b | 77.4% |
| nemotron-nano-12b-v2-vl | 77.4% |
| gemma-3n-e4b-it | 75.2% |
| llama-3.1-8b-instruct | 72.5% |
| gemma-3-4b-it | 71.3% |
| llama-3.2-3b-instruct | 57.6% |