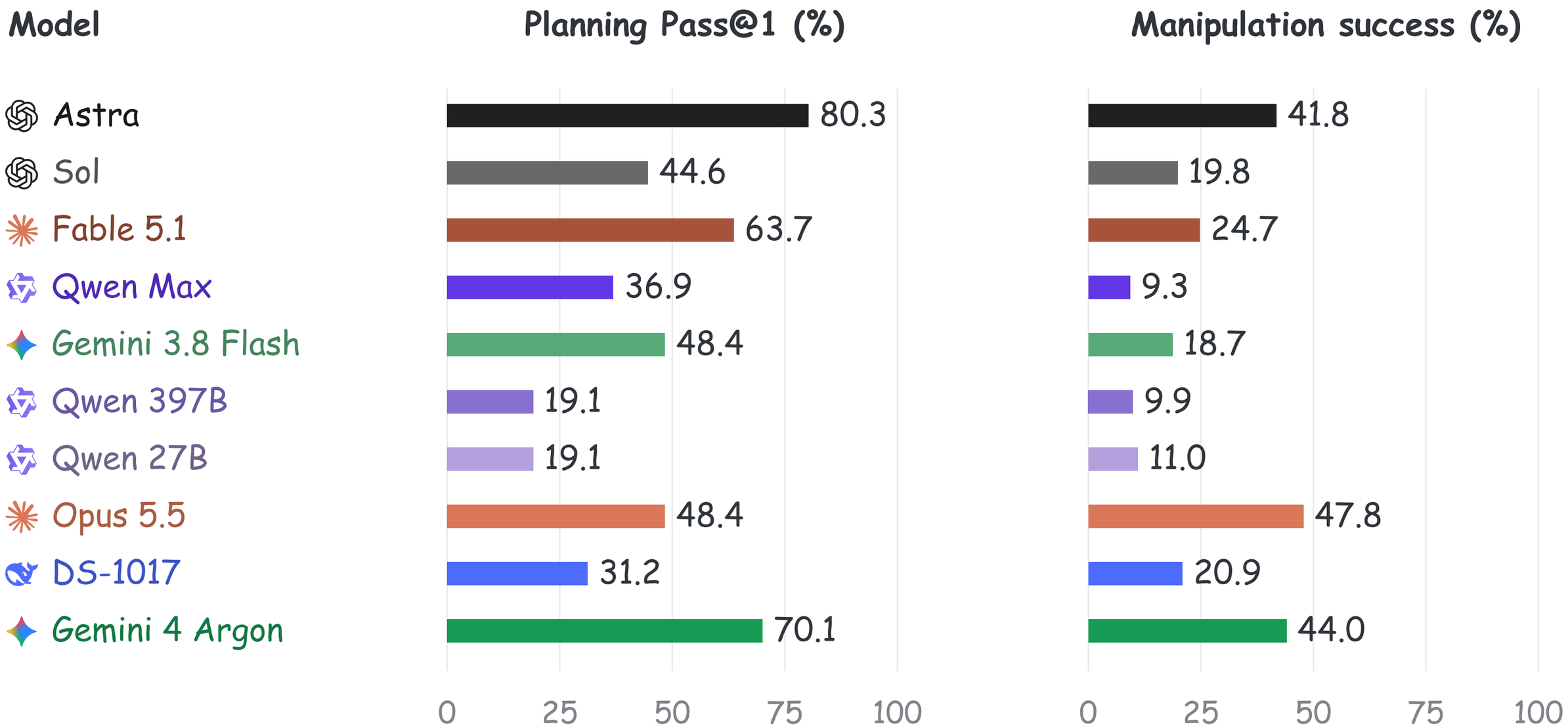


Table 2 · one run per task · source-specific evaluation budgets