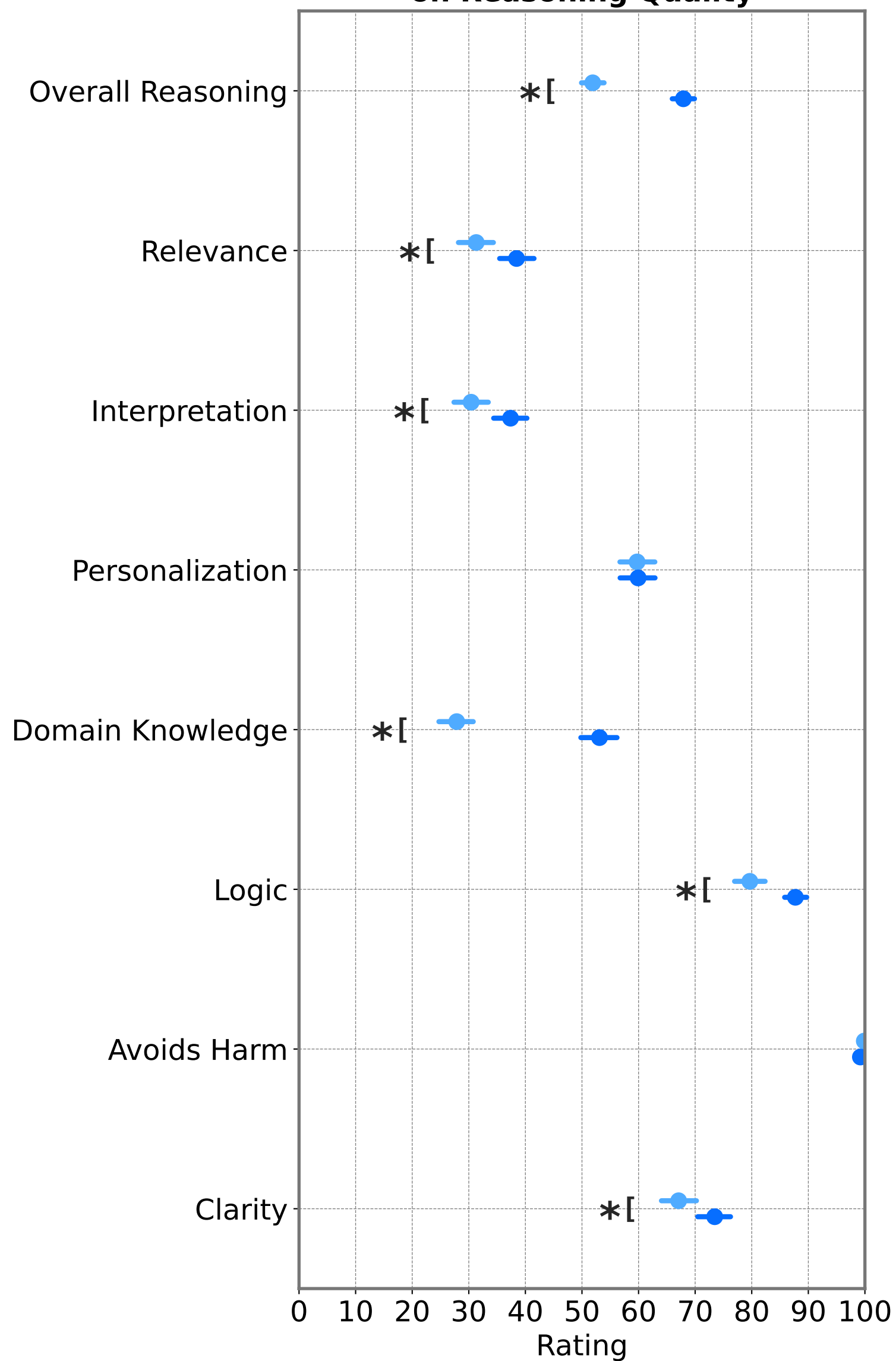
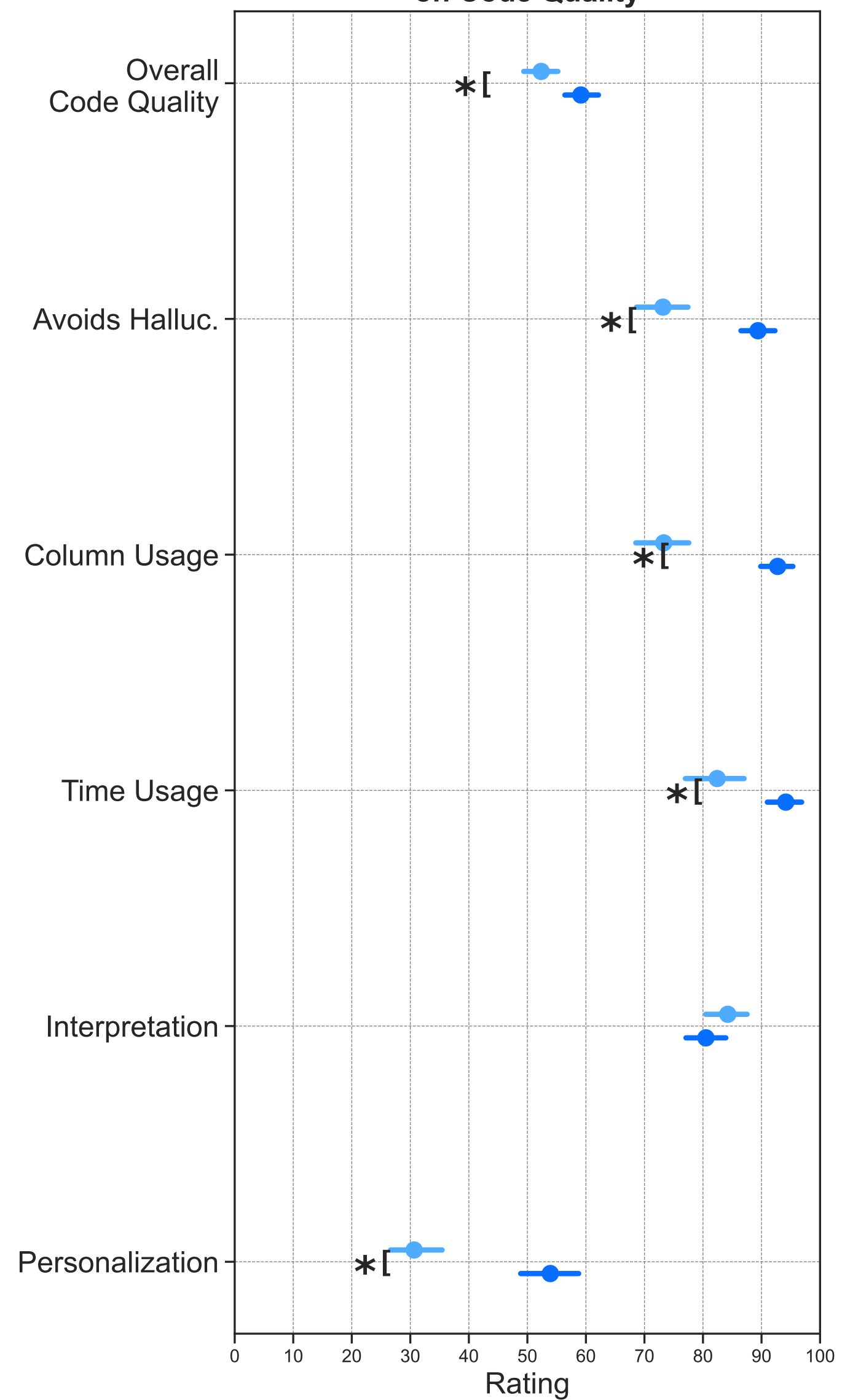


**(B) Human Eval of Open Queries
on Reasoning Quality**



**(C) Expert Eval of Open Queries
on Code Quality**



● PHIA ● Code Generation ● Numerical Reasoning