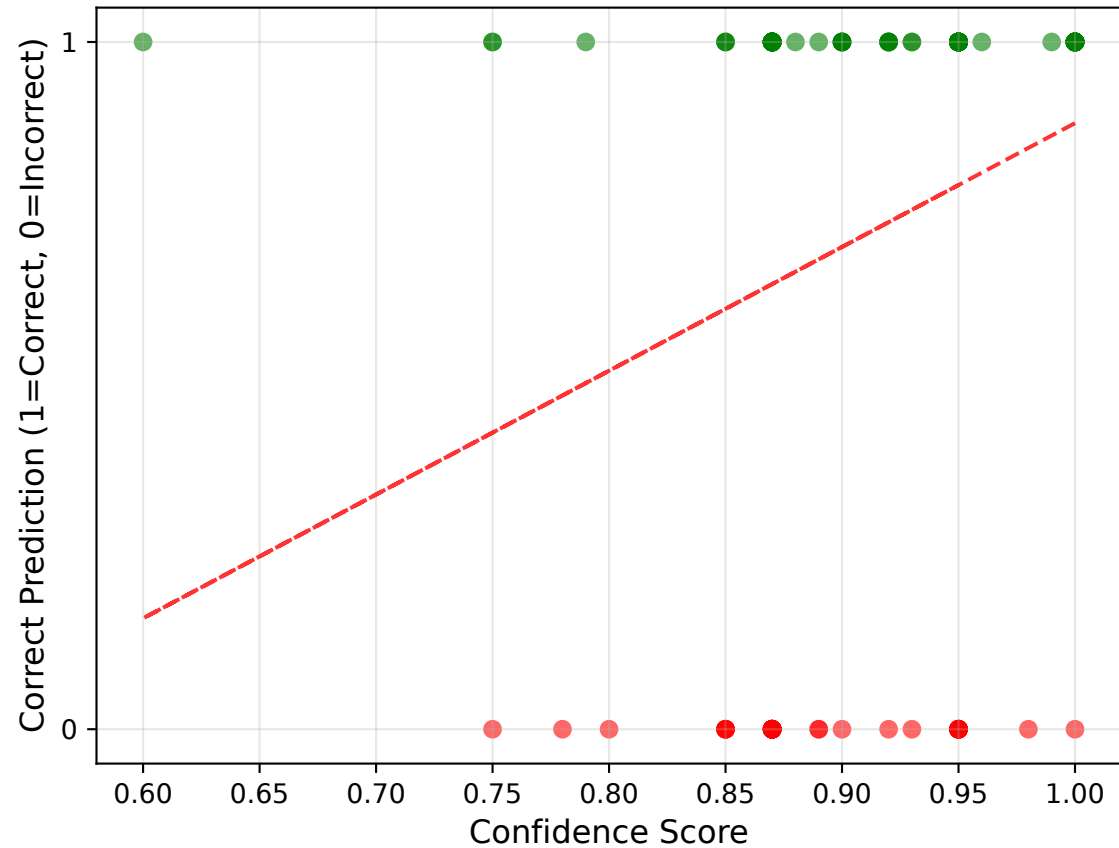
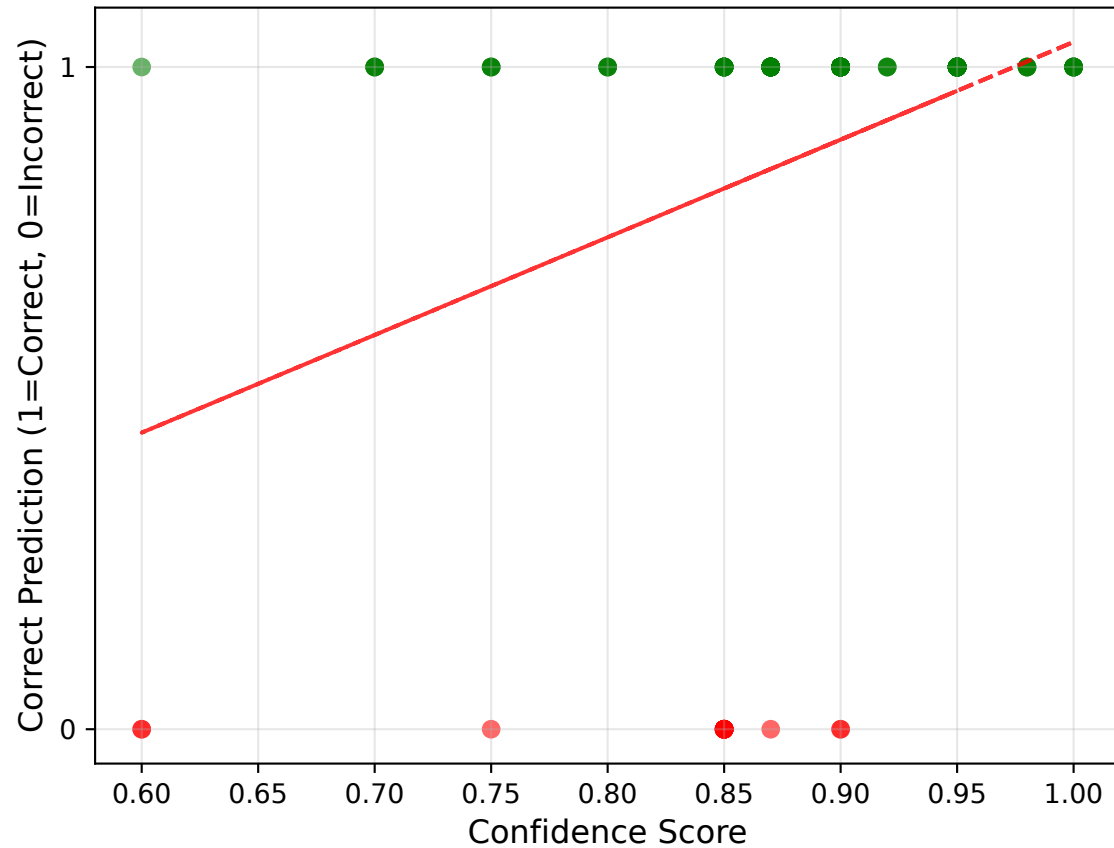


Qwen: Confidence vs Accuracy



GPT: Confidence vs Accuracy



Orchestrator: Confidence vs Accuracy

