Accuracy by Exam and Agent for GPT-4