F1 scores depend only on recall 1.0 0.81 0.76 0.64 0.47 0.5 F1 Score Precision Recall GPT-4o Claude 3.5 Sonnet Llama 3.1 Instruct GPT-40-Mini