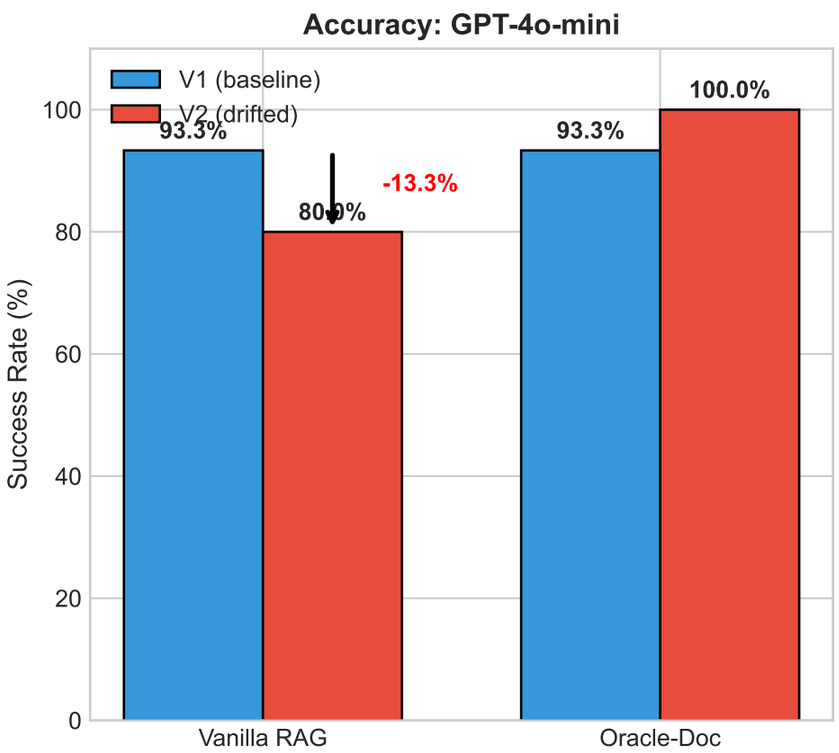
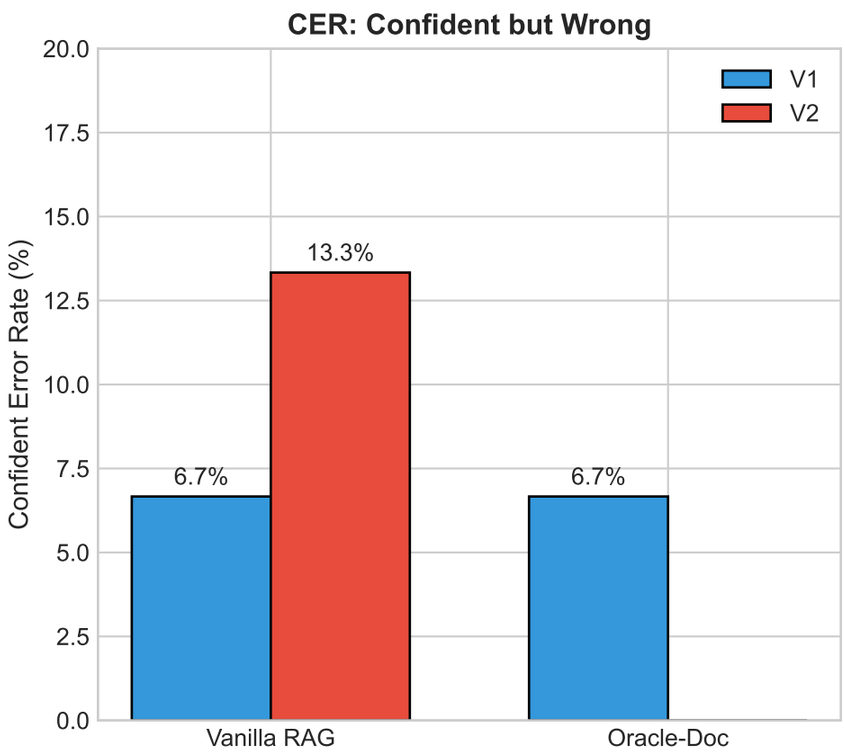
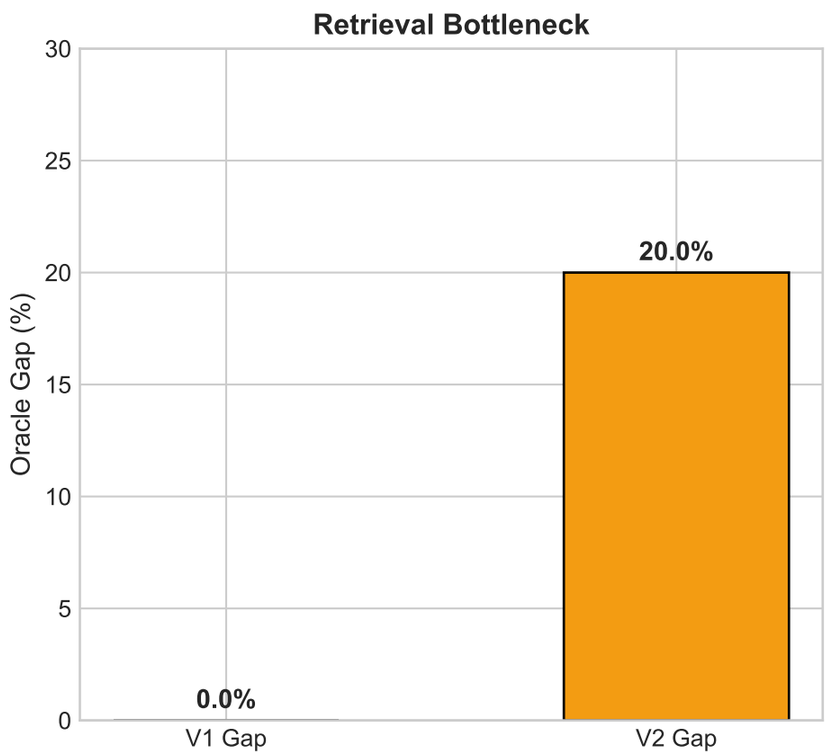

Accuracy: GPT-4o-mini
V1 (baseline)
V2 (drifted)
Success Rate (%)
93.3%
80.0%
-13.3%
93.3%
100.0%
Vanilla RAG
Oracle-Doc


CER: Confident but Wrong
V1
V2
Confident Error Rate (%)
6.7%
13.3%
6.7%
Vanilla RAG
Oracle-Doc


Retrieval Bottleneck
Oracle Gap (%)
0.0%
20.0%
V1 Gap
V2 Gap
Gap = Oracle_Acc - RAG_Acc
(Higher = more retrieval failures)