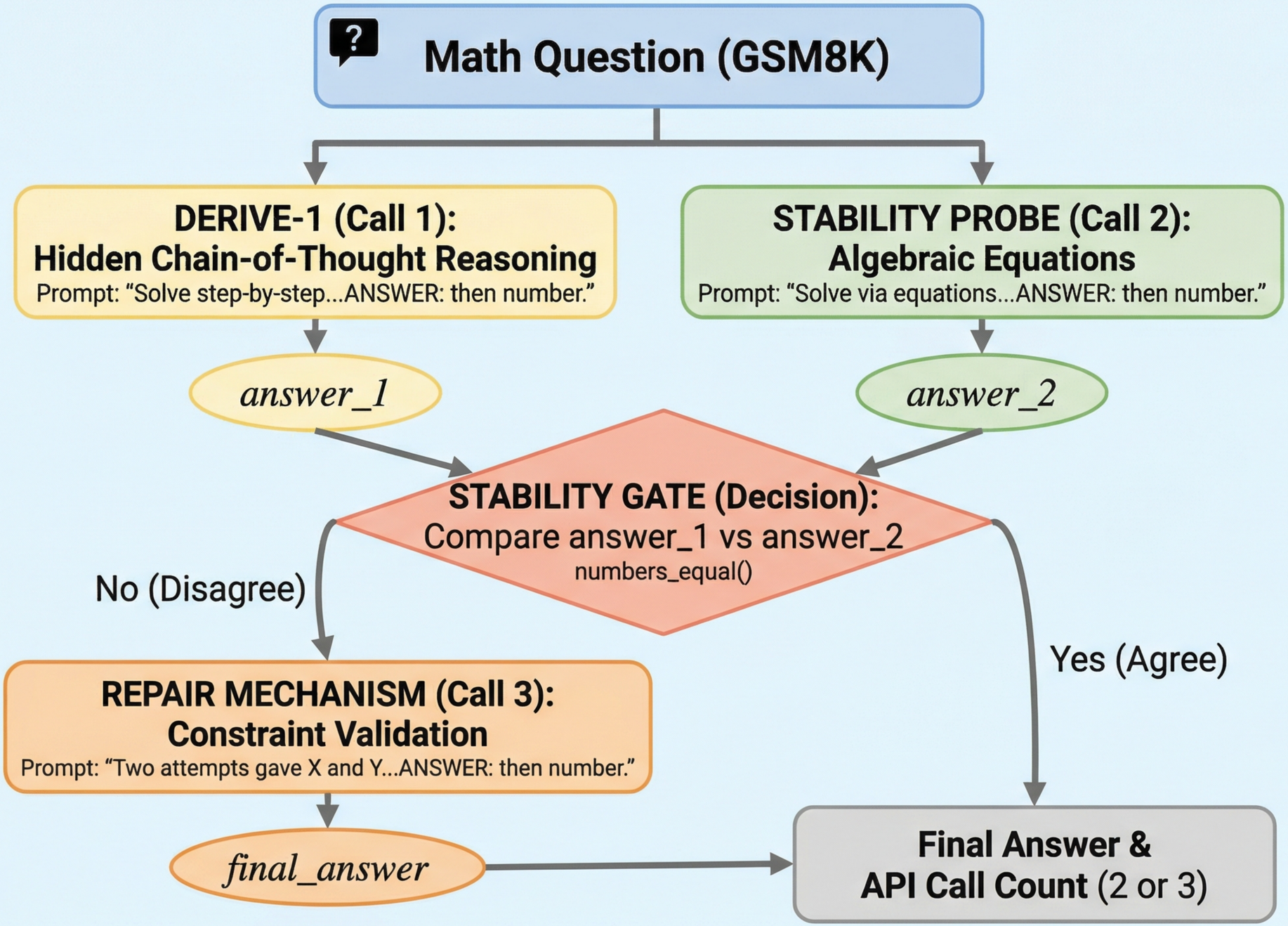

Baseline Comparison Section

ZeroShotDirect:
1 call, no reasoning

HiddenCoT:
1 call, internal CoT

SelfConsistency:
3 calls, temperature=0.7, majority voting