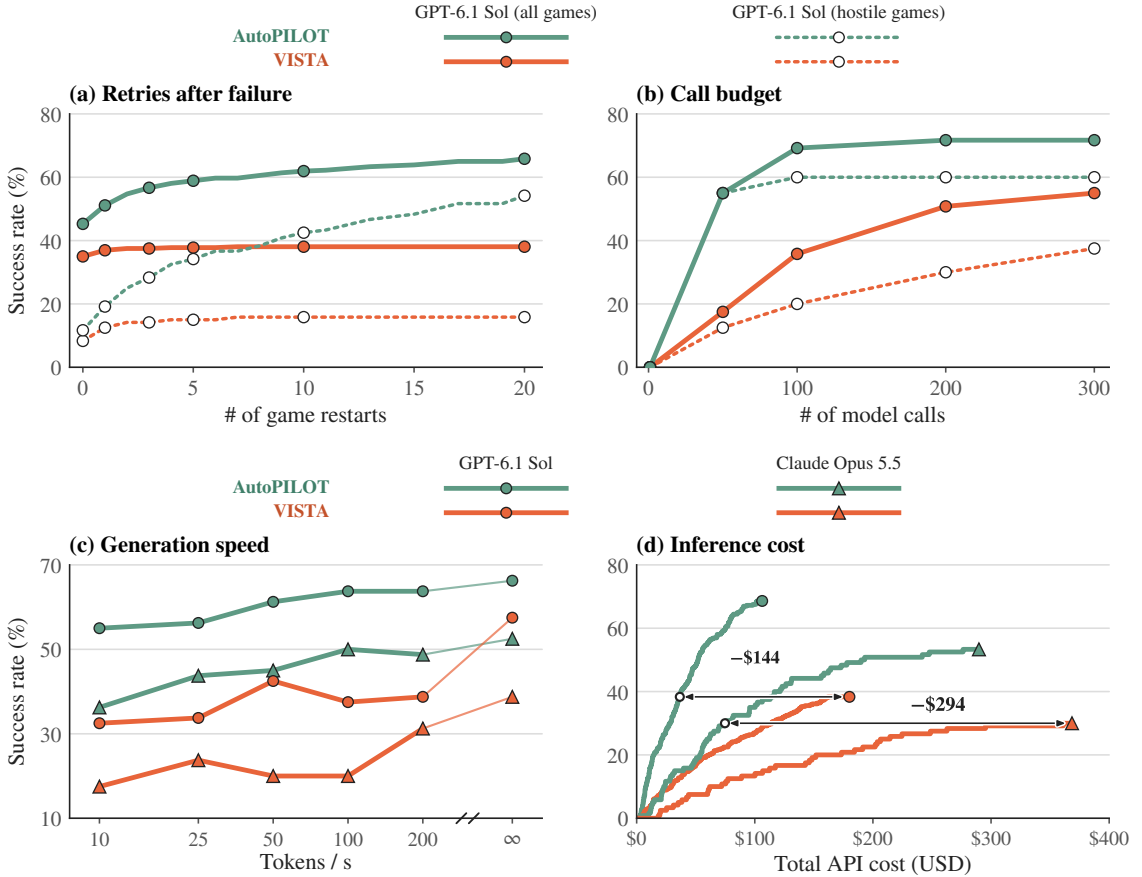
GPT-6.1 Sol (all games)
GPT-6.1 Sol (hostile games)
AutoPILOT
VISTA
(a) Retries after failure
Success rate (%)
# of game restarts
(b) Call budget
# of model calls
GPT-6.1 Sol
Claude Opus 5.5
AutoPILOT
VISTA
(c) Generation speed
Success rate (%)
Tokens / s
(d) Inference cost
−$144
−$294
Total API cost (USD)