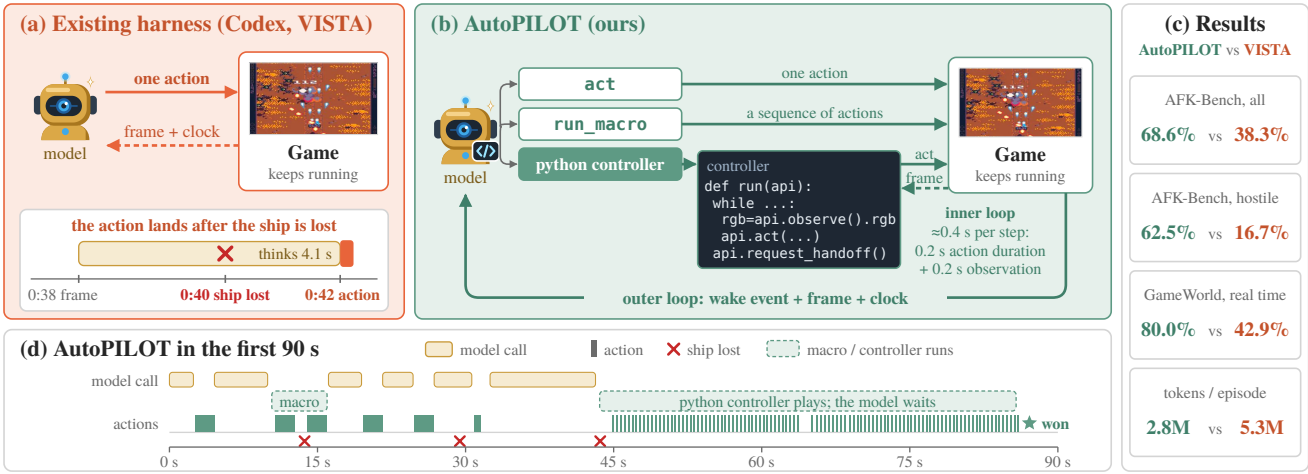
(a) Existing harness (Codex, VISTA)
one action
frame + clock
model
Game
keeps running
the action lands after the ship is lost
thinks 4.1 s
0:38 frame
0:40 ship lost
0:42 action
(b) AutoPILOT (ours)
act
run_macro
python controller
one action
a sequence of actions
model
controller
def run(api):
while ...:
rgb=api.observe().rgb
api.act(...)
api.request_handoff()
act
frame
Game
keeps running
inner loop
≈0.4 s per step:
0.2 s action duration
+ 0.2 s observation
outer loop: wake event + frame + clock
(c) Results
AutoPILOT vs VISTA
AFK-Bench, all
68.6% vs 38.3%
AFK-Bench, hostile
62.5% vs 16.7%
GameWorld, real time
80.0% vs 42.9%
tokens / episode
2.8M vs 5.3M
(d) AutoPILOT in the first 90 s
model call
action
ship lost
macro / controller runs
model call
macro
python controller plays; the model waits
actions
won
0 s
15 s
30 s
45 s
60 s
75 s
90 s