92

Signal-Driven Observation for Long-Horizon Web Agents

Shubham Gaur sgaur2@ucsc.edu
Ian Lane ialane@ucsc.edu
Department of Computer Science and Engineering, UC Santa Cruz

UC SANTA CRUZ BaskinEngineering
BayLearn 2026
Santa Clara University · 15 Oct 2026

Paper **Portfolio**

POSITION
Observe when the page changes, not after every action.

EVIDENCE
Many steps barely change the page, across 5 datasets.

PAYOFF
Reading only on signals: ~27% fewer planner tokens.

01 Agents re-read every page

After every action the agent ingests the full page (**20K–80K tokens**), changed or not.

~50%
WebArena success, even for frontier agents
Zhou et al. 2023

~½
of WebArena-Lite runs get stuck mid-task
Gemini-2.5-Pro · Wang et al. 2026

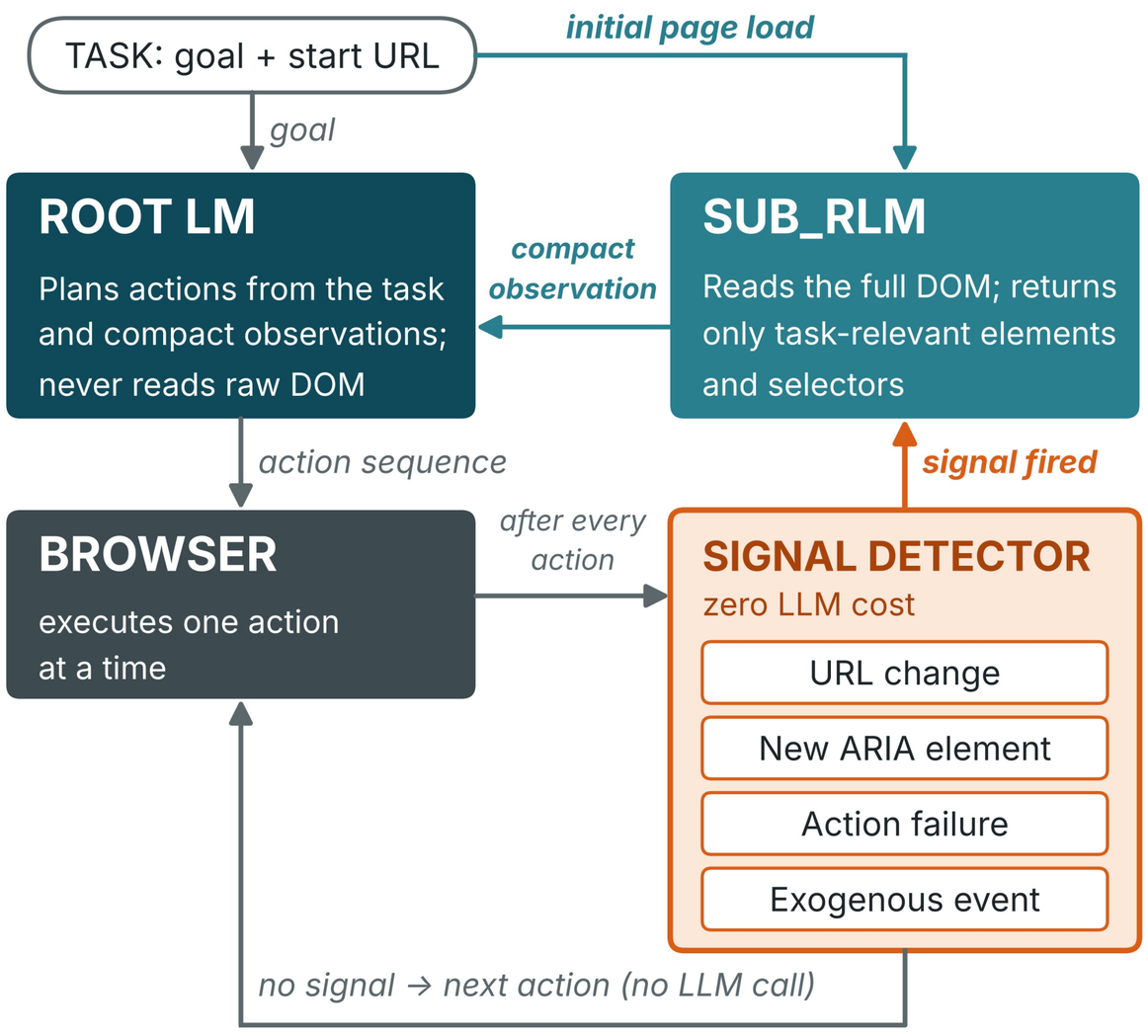
We hypothesize re-reading drives three failures:

Context rot
stale tokens crowd out reasoning

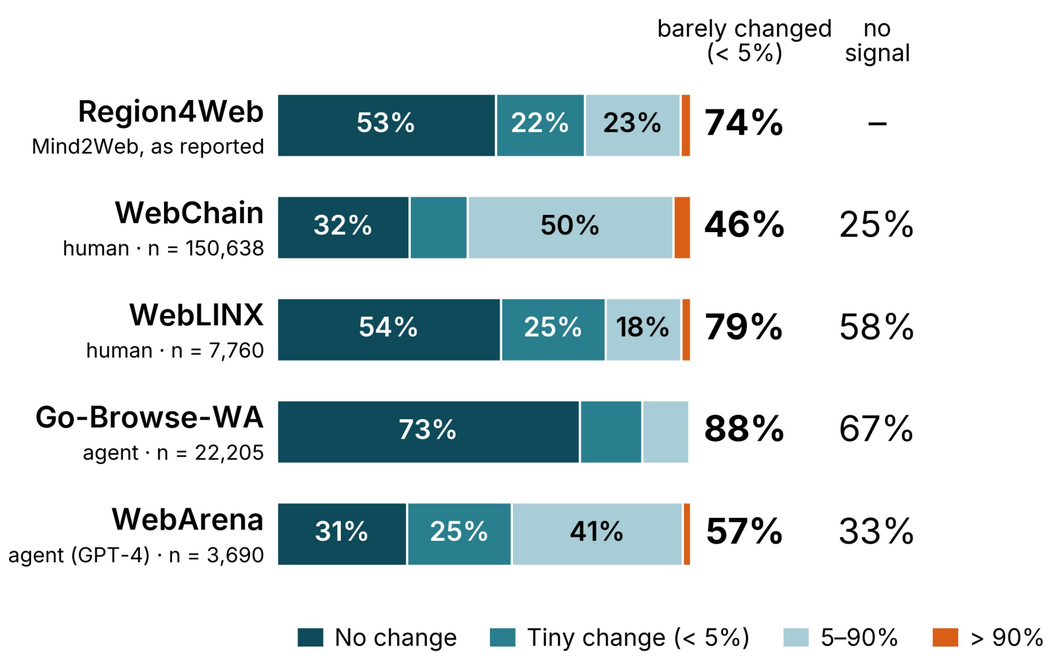
Loop-trapping
revisits a state it already saw

Goal drift
the objective gets buried

02 SDO: decide *when* to look

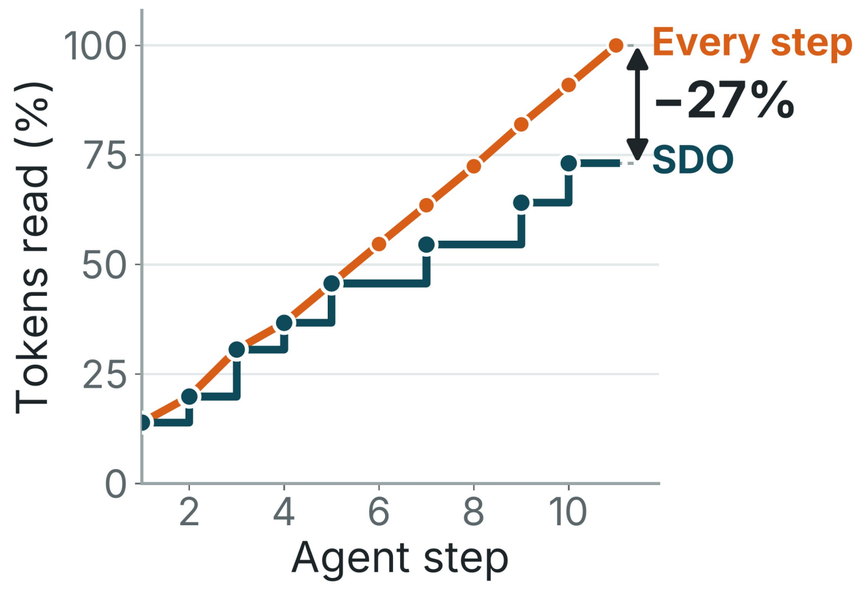


03 Many steps barely change the page



Share of same-page steps by fraction of page elements changed. *No signal* (all steps): no URL change and no new interactive element. Region4Web as reported (Kwon & Lee 2026).

04 Signal-only reading: ~¼ fewer tokens



One typical WebArena (GPT-4) task. Median saving 27% over 266 tasks (IQR 9–45%); Go-Browse agents 68%. Upper bound: only the URL and new-element signals modeled.

05 How much vs. when

System	Re-observes
FocusAgent, ACON	every step, compressed
PageDigest	on URL change
WebChallenger	on URL change or modal
SDO (ours)	**on any of 4 signals**

WHAT SDO STILL HAS TO SOLVE

Silent changes: a price updated by script fires no signal.

Whose change? the page reacting to us vs. on its own.

Errors between signals: a wrong click that "works" goes unnoticed.

Over-filtering: the sub-call may drop what the planner needs.

NEXT

Build SDO in BrowserGym → evaluate on WebArena-Lite → live sites

Zhou et al. 2023 (WebArena) · Wang et al. 2026 (subgoal-driven agents) · Kwon & Lee 2026 (Region4Web) · Zhang et al. 2025 (Recursive LMs) · Hwang et al. 2026 (WebChallenger) · Gandhi & Neubig 2025 (Go-Browse) · Data: Mind2Web, WebChain, WebLINX, WebArena