WP29 — Visual Evidence Appendix

Simon-Pierre Boucher, UQO

October 2026

Companion to UQO Working Paper 29. All captures are authentic (native `screencapture`); modes, hashes and selection rules follow the screenshot manifest. Capture failures are listed without images.

1 Visual evidence gallery

Every image below is an unmodified or cropped capture of a real surface, recorded in the screenshot manifest (`evidence/manifests/screenshots.csv`) with its capture mode, tool, UTC time, SHA-256 of the original and of the public copy, and the run and event it documents. Capture failures are listed with their reason and have no image. Hashes attest file integrity relative to the manifest; they do not independently prove the moment of capture, which is triangulated with the session and event timestamps in the state store.

S001 (no image) **S001** — mode `viewer_postrun`; 2026-10-30T03:42:35 UTC; run RD0002; tool /usr/sbin/screencapture (macOS 27.0); original SHA-256 532a5184efbfe3de…; public SHA-256 –…; none. CAPTURE FAILURE: window-level capture of Safari produced the browser chrome without the page content (web content is rendered out of process); retained as a record of the failure, not usable as evidence. Evidence viewer after the smoke batch: four runs evaluated (RD0001-RD0004) with stored costs and verdicts.

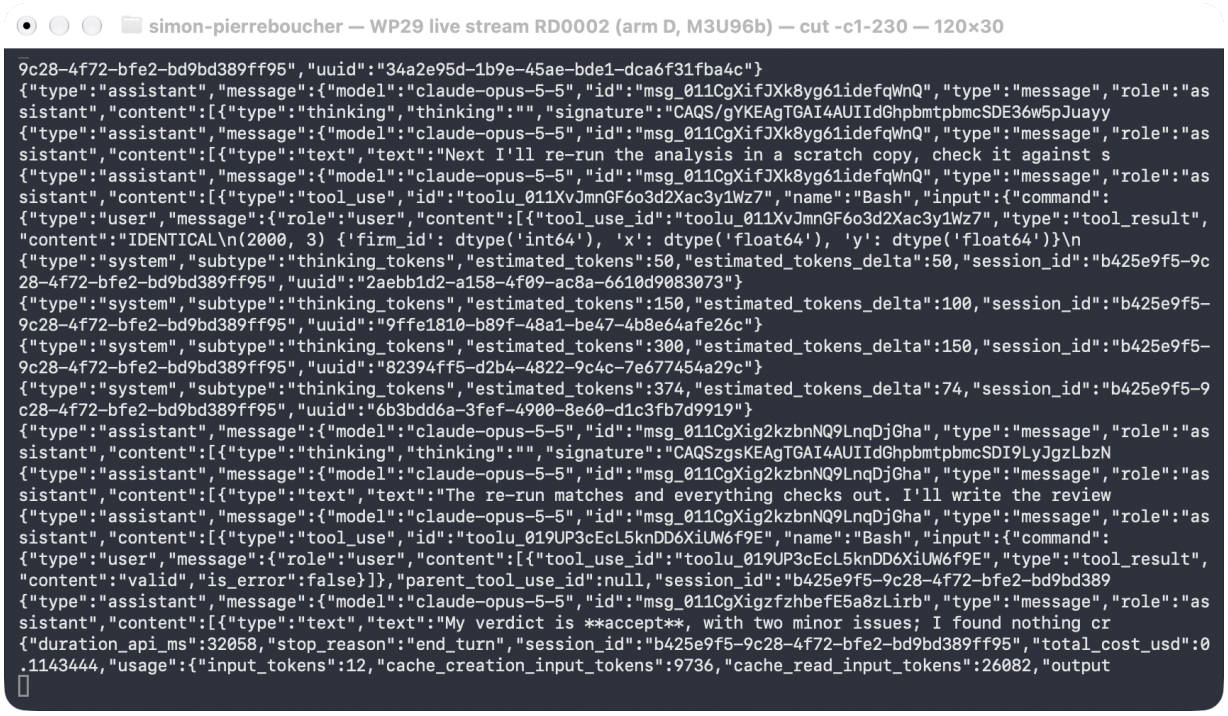


Figure 1: **S002** — mode `native_live`; 2026-10-30T03:34:33 UTC; run RD0002; tool /usr/sbin/screencapture (macOS 27.0); original SHA-256 ff42539ba7cc1a8b…; public SHA-256 ff42539ba7cc1a8b…; copied without modification. Native Terminal window tailing the stream-json output of executor session RD0002 (arm D) on M3U96b, captured immediately after the run ended (the executor session had finished 41 s earlier and the run was marked evaluated 1.4 s before the capture; the window shows the tail of the live stream file).

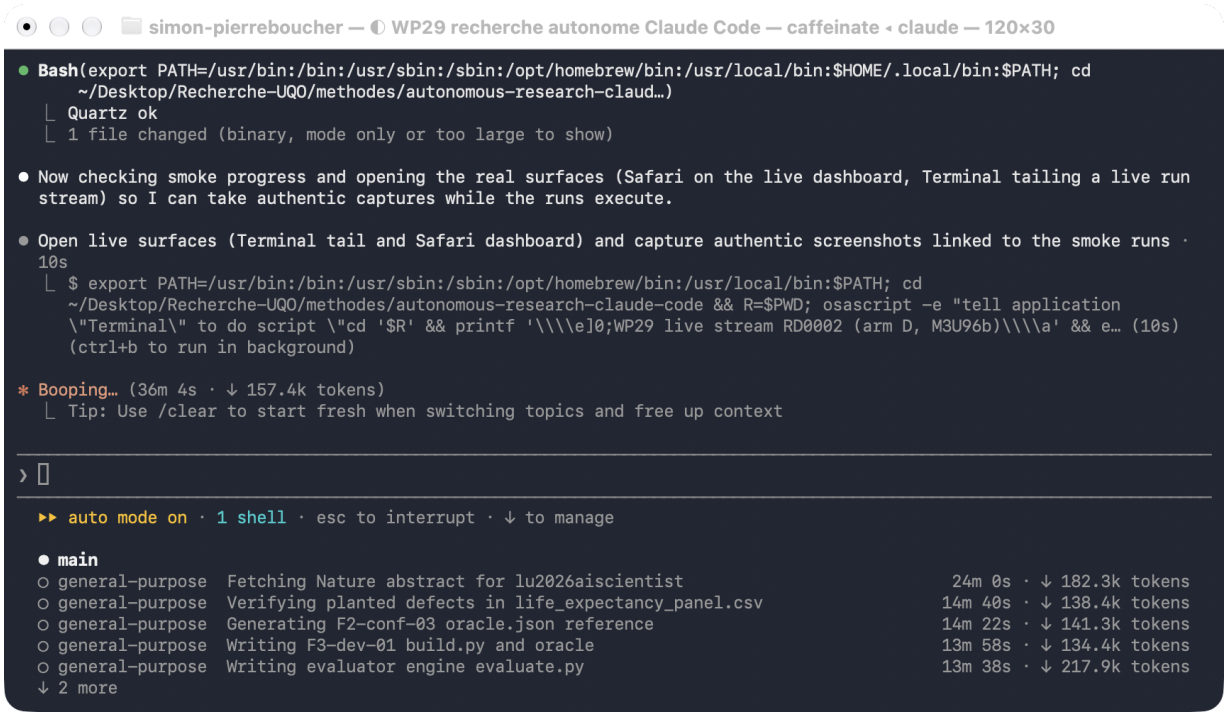


Figure 2: **S003** — mode `native_live`; 2026-10-30T03:34:33 UTC; run –; tool /usr/sbin/screencapture (macOS 27.0); original SHA-256 9189f4fac4cdab39…; public SHA-256 9189f4fac4cdab39…; copied without modification. Native Terminal capture of the WP29 builder session itself (the session writing this harness), showing five task-authoring subagents running in parallel while the smoke experiment executes on the nodes.

S004 (no image) **S004** — mode `dashboard_live`; 2026-10-30T03:50:59 UTC; run RP0010; tool /usr/sbin/screencap (macOS 27.0); original SHA-256 4e3dc399db5eca37…; public SHA-256 –…; none. CAPTURE FAILURE: window-level capture of Safari produced the browser chrome without the page content (web content is rendered out of process); retained as a record of the failure, not usable as evidence. Live evidence dashboard during pilot batch 1: 7 runs in status running at capture time.

UQO/methodes/autonomous-research-claude-code/runs/state.sqlite at 2026-10-30T03:56:48+00:00 (dashboard, not a terminal)

Runs: 24 — evaluated: 24 — events: 1084 — sessions: 30 — total inference cost (client estimate): 8.602 USD

run	task	split	arm	rep	node	status	reason	USD	sessio
RD0001	F2-smoke-01	development	A	0	M3U96a	evaluated	executor_finished	0.120	1
RD0002	F2-smoke-01	development	D	0	M3U96b	evaluated	executor_finished	0.418	2
RD0003	F2-smoke-	development	C	0	M4M64a	evaluated	executor_finished	0.237	1

Figure 3: **S005** — mode `viewer_postrun`; 2026-10-30T03:56:52 UTC; run RP0014; tool /usr/sbin/screencapture (macOS 27.0); original SHA-256 4b399f88e540a67c…; public SHA-256 d838ec603d69ae8f…; cropped: top 300 px of 822 (browser toolbar and bookmarks bar) removed; page content unmodified. Evidence viewer after pilot batch 1: 4 smoke runs and 20 pilot runs, all evaluated; costs and verdicts as stored.

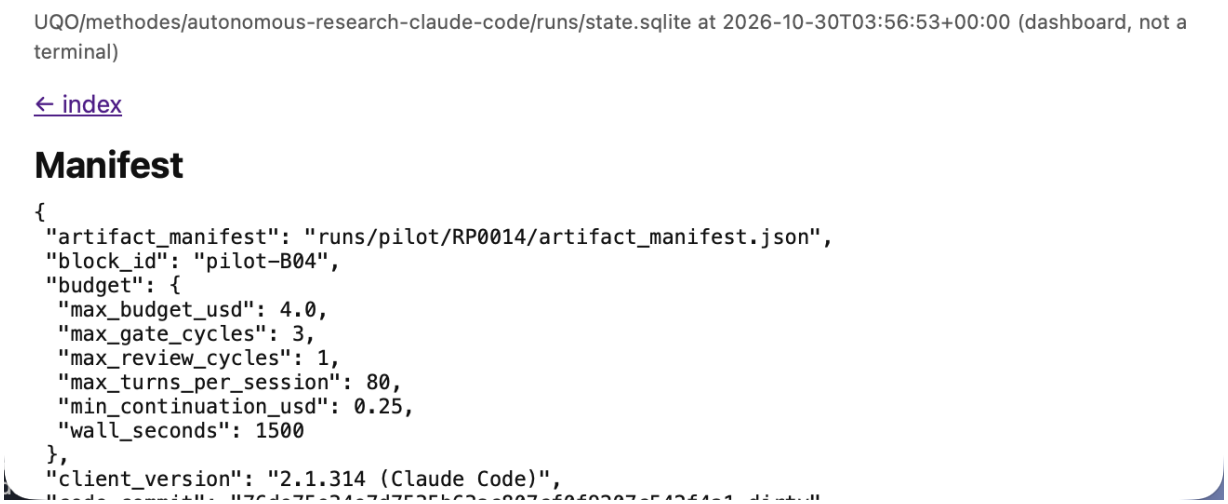


Figure 4: **S006** — mode `viewer_postrun`; 2026-10-30T03:56:57 UTC; run RP0014; tool /usr/sbin/screencapture (macOS 27.0); original SHA-256 1d24f84a2d1351c4…; public SHA-256 ecf71fb77ac5288b…; cropped: top 300 px of 822 (browser toolbar and bookmarks bar) removed; page content unmodified. Run page of RP0014 (F6-dev-01, arm D): manifest, envelope, transitions, two sessions (executor + reviewer), gate receipts, review and evaluation receipt.

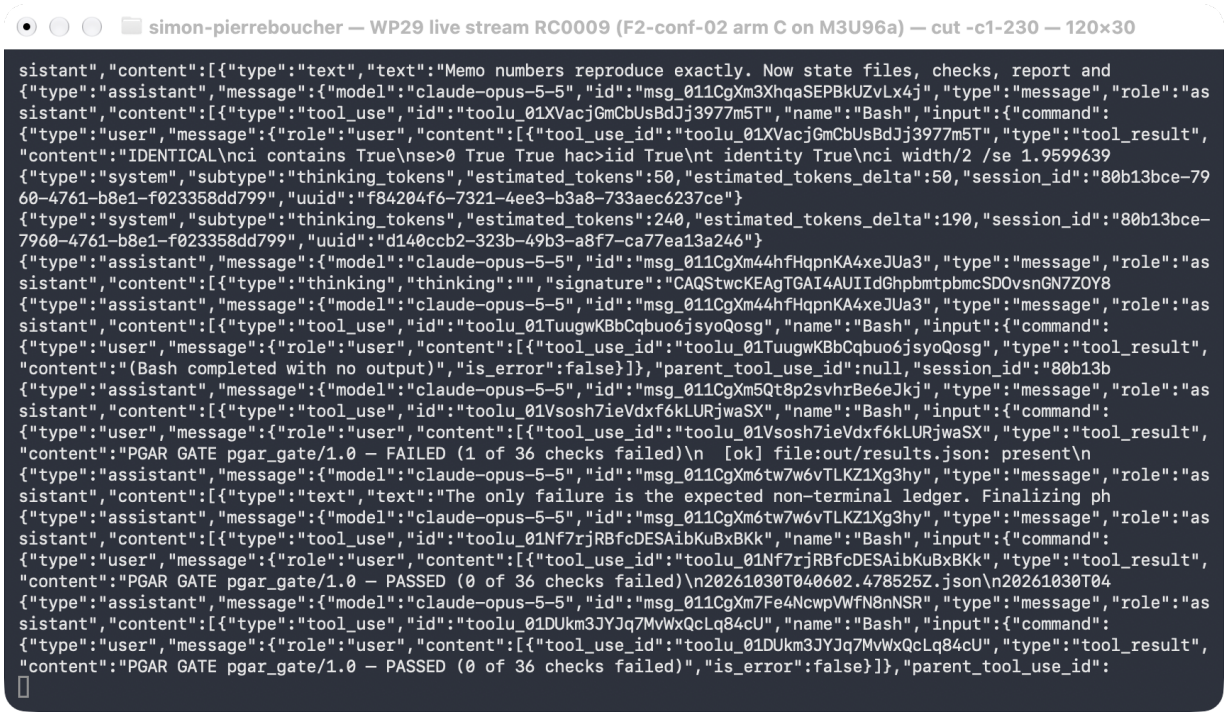
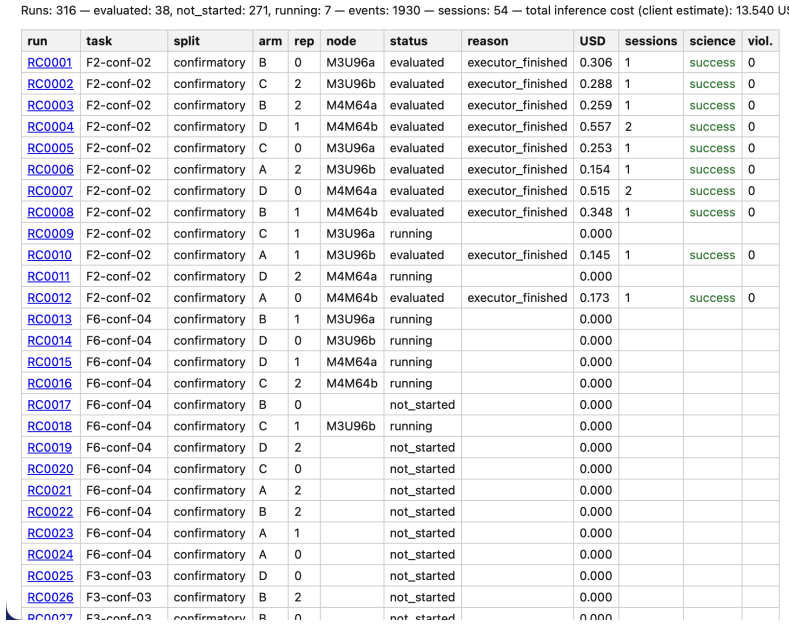


Figure 5: **S007** — mode `native_live`; 2026-10-30T04:06:14 UTC; run RC0009; tool /usr/sbin/screencapture (macOS 27.0); original SHA-256 e8ec97be2eb523d9…; public SHA-256 e8ec97be2eb523d9…; copied without modification. Native Terminal window tailing the live stream-json of confirmatory run RC0009 (F2-conf-02, arm C) while it executes on M3U96a.



Runs: 316 — evaluated: 38, not_started: 271, running: 7 — events: 1930 — sessions: 54 — total inference cost (client estimate): 13.540 USD

run	task	split	arm	rep	node	status	reason	USD	sessions	science	viol.
RC0001	F2-conf-02	confirmatory	B	0	M3U96a	evaluated	executor_finished	0.306	1	success	0
RC0002	F2-conf-02	confirmatory	C	2	M3U96b	evaluated	executor_finished	0.288	1	success	0
RC0003	F2-conf-02	confirmatory	B	2	M4M64a	evaluated	executor_finished	0.259	1	success	0
RC0004	F2-conf-02	confirmatory	D	1	M4M64b	evaluated	executor_finished	0.557	2	success	0
RC0005	F2-conf-02	confirmatory	C	0	M3U96a	evaluated	executor_finished	0.253	1	success	0
RC0006	F2-conf-02	confirmatory	A	2	M3U96b	evaluated	executor_finished	0.154	1	success	0
RC0007	F2-conf-02	confirmatory	D	0	M4M64a	evaluated	executor_finished	0.515	2	success	0
RC0008	F2-conf-02	confirmatory	B	1	M4M64b	evaluated	executor_finished	0.348	1	success	0
RC0009	F2-conf-02	confirmatory	C	1	M3U96a	running		0.000			
RC0010	F2-conf-02	confirmatory	A	1	M3U96b	evaluated	executor_finished	0.145	1	success	0
RC0011	F2-conf-02	confirmatory	D	2	M4M64a	running		0.000			
RC0012	F2-conf-02	confirmatory	A	0	M4M64b	evaluated	executor_finished	0.173	1	success	0
RC0013	F6-conf-04	confirmatory	B	1	M3U96a	running		0.000			
RC0014	F6-conf-04	confirmatory	D	0	M3U96b	running		0.000			
RC0015	F6-conf-04	confirmatory	D	1	M4M64a	running		0.000			
RC0016	F6-conf-04	confirmatory	C	2	M4M64b	running		0.000			
RC0017	F6-conf-04	confirmatory	B	0		not_started		0.000			
RC0018	F6-conf-04	confirmatory	C	1	M3U96b	running		0.000			
RC0019	F6-conf-04	confirmatory	D	2		not_started		0.000			
RC0020	F6-conf-04	confirmatory	C	0		not_started		0.000			
RC0021	F6-conf-04	confirmatory	A	2		not_started		0.000			
RC0022	F6-conf-04	confirmatory	B	2		not_started		0.000			
RC0023	F6-conf-04	confirmatory	A	1		not_started		0.000			
RC0024	F6-conf-04	confirmatory	A	0		not_started		0.000			
RC0025	F3-conf-03	confirmatory	D	0		not_started		0.000			
RC0026	F3-conf-03	confirmatory	B	2		not_started		0.000			
RC0027	F3-conf-03	confirmatory	B	0		not_started		0.000			

Figure 6: **S008** — mode `dashboard_live`; 2026-10-30T04:06:18 UTC; run RC0009; tool /usr/sbin/screencapture (macOS 27.0); original SHA-256 9dc48df7c7a7e54f…; public SHA-256 e8aa89fd6902ec51…; cropped: top 300 px of 1644 (browser toolbar and bookmarks bar) removed; page content unmodified. Live evidence dashboard during the main benchmark: 7 confirmatory runs in status running at capture time.

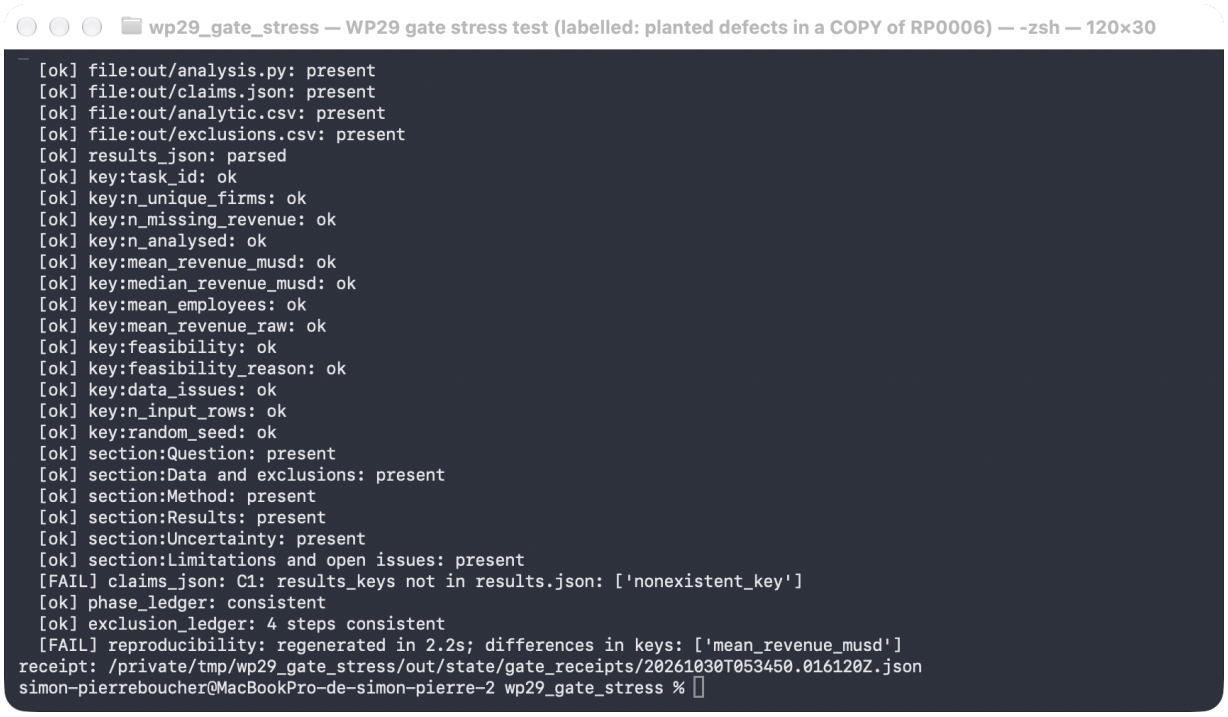


Figure 7: **S009** — mode `native_live`; 2026-10-30T05:34:59 UTC; run RP0006; tool /usr/sbin/screencapture (macOS 27.0); original SHA-256 e5df0b01cffc3b2f…; public SHA-256 e5df0b01cffc3b2f…; copied without modification. LABELLED STRESS TEST, not a natural event: the gate program rejecting a copy of RP0006's deliverables after two defects were planted by the controller (a hand-edited reported value, which breaks reproducibility, and a claim referencing a non-existent results key).

S010 (no image) **S010** — mode `viewer_postrun`; 2026-10-30T05:46:14 UTC; run RT0009; tool /usr/sbin/screencap (macOS 27.0); original SHA-256 . . . ; public SHA-256 –. . . ; n/a. CAPTURE FAILURE: the macOS session was locked (loginwindow on screen) when the post-hoc capture of the RT0009 run page was attempted; screencapture could not render the window. The gate receipts themselves are in the state store (4 controller receipts, all failed on forecasts_chronology).