

AI Factory V2 — P2P Verifier Metrics and Quality Report

AI Factory V2 WP3 — P2P Verifier Metrics and Quality Report

Generated: 2026-05-26T15:28:35.483Z

Scope

Source DB: `/Users/makinja/system/databases/company-mesh.db`

Included MC tasks:

- #101987 — LumisCare notification-service migration pilot context
- #102081 — AI Factory V2 WP1 runner MVP
- #102083 — AI Factory V2 WP4 writeback reliability

Metrics Summary

- Threads analyzed: 24
- Acceptable thread responses (answered + PASS/PARTIAL/ANSWERED): 5
- Attempt-level acceptable rate: 20.8%
- Response classes: `{"ANSWERED":3,"NO_RESPONSE":3,"BLOCKED":16,"PASS":1,"PARTIAL":1}`
- Failure patterns:
`{"none":4,"stale_delivered_or_no_response":3,"timeout_or_worker_no_response":7,"agent_runner_or_ollama_failure":3,"blocked_unspecified_or_claim_gate":5,"partial_due_summary_only_evidence":2}`

By Task

- #101987: total=6, acceptable=2, blocked=1, no_response=3, cost_cap_sum=\$6.00
- #102081: total=6, acceptable=1, blocked=5, no_response=0, cost_cap_sum=\$2.00
- #102083: total=12, acceptable=2, blocked=10, no_response=0, cost_cap_sum=\$7.15

Thread Detail

Task	Thread	Status/classes	Acceptable	Pattern	Prompt chars	Latency s	Evidence
#101987	mesh-thr-8b3552e3-4f58-4f9f-a4b2-82b6ec8dbfc4	answered/ANSWERED	yes	none	416	1554	<code>/Users/makinja/system/rules/p2p-pair-migration.md</code>
#101987	mesh-thr-2170a2ba-3019-4c82-9bde-af102d38dd8f	answered/ANSWERED	yes	none	507	253	
#101987	mesh-thr-9392faa2-2d7a-40ad-9017-4ada9190bbd2	open/NO_RESPONSE	no	stale_delivered_or_no_response	447		
#101987	mesh-thr-bf0d9685-c54a-44e1-acb9-55d22590fe8d	blocked/BLOCKED	no	timeout_or_worker_no_response	753	64	<code>/tmp/alai/company-mesh-timeouts/mesh-msg-a5b6f8fb-16e3-4519-a382-6a8b181e3b28.json</code>
#101987	mesh-thr-61154c1b-4b74-4b93-a92e-2d1beb295c65	open/NO_RESPONSE	no	stale_delivered_or_no_response	506		

Task	Thread	Status/classes	Acceptable	Pattern	Prompt chars	Latency s	Evidence
#101987	mesh-thr-9ab9ece8-f33a-4fdb-9d29-ef1bb681667f	open/NO_RESPONSE	no	stale_delivered_or_no_response	518		
#102083	mesh-thr-b5873415-a389-4f26-a810-1d3cdf13a2c4	blocked/BLOCKED	no	agent_runner_or_ollama_failure	718	92	<code>/tmp/alai/company-mesh-auto-responder/2026-05-26T13-29-09-784Z-mesh-msg-4b045b56-b9e9-421b-9336-d51e6c1166da.json</code>
#102083	mesh-thr-b3f219e7-7dbf-41acb2a2-9d1e501126dc	blocked/BLOCKED	no	timeout_or_worker_no_response	719	122	<code>/tmp/alai/company-mesh-timeouts/mesh-msg-8f5314b3-426d-4de6-a0d7-c8964b85e358.json</code>
#102083	mesh-thr-792068a5-74ec-40d8-988a-0d6d297339ba	blocked/BLOCKED	no	timeout_or_worker_no_response	484	123	<code>/tmp/alai/company-mesh-timeouts/mesh-msg-5ae99557-5984-4b5f-a37c-1586c89a6af3.json</code>
#102081	mesh-thr-9cbebd3-79f5-4201-80af-2bbd64d35ec4	blocked/BLOCKED	no	timeout_or_worker_no_response	1205	123	<code>/tmp/alai/company-mesh-timeouts/mesh-msg-355ee365-5af6-4fb3-ba7a-59cdb3673483.json</code>

Task	Thread	Status/classes	Acceptable	Pattern	Prompt chars	Latency s	Evidence
#102081	mesh-thr-f07042ae-b529-4907-b844-e25f1b21a12b	blocked/BLOCKED	no	agent_runner_or_ollama_failure	869	78	/tmp/alai/company-mesh-auto-responder/2026-05-26T14-01-02-501Z-mesh-msg-7a217112-2969-453d-8225-86d25e8fb23a.json
#102083	mesh-thr-6a5c9d97-df2e-4352-9b74-cf5db7c7bb40	blocked/BLOCKED	no	blocked_unspecified_or_claim_gate	266	16	/tmp/alai/company-mesh-auto-responder/2026-05-26T14-01-42-724Z-mesh-msg-2bf0c206-b599-4cda-990f-258ded567271.json
#102083	mesh-thr-57b70489-5ebb-4e91-a7a0-9d2a7e868497	answered/ANSWERED	yes	none	289	93	/tmp/alai/company-mesh-auto-responder/2026-05-26T14-03-31-501Z-mesh-msg-ed34a16c-5b49-4beb-ad46-db59696b948b.json
#102083	mesh-thr-dc65ed91-e027-4cf8-931c-ff5f55b43a49	blocked/BLOCKED	no	blocked_unspecified_or_claim_gate	1255	120	/tmp/alai/company-mesh-auto-responder/2026-05-26T14-06-46-587Z-mesh-msg-d9bfaf85-5817-49cb-bbe4-3f6c5c7802de.json

Task	Thread	Status/classes	Acceptable	Pattern	Prompt chars	Latency s	Evidence
#102081	mesh-thr-5929968f-3eb5-41d6-8a79-643dc544ed05	blocked/BLOCKED	no	timeout_or_worker_no_response	957	123	/tmp/alai/company-mesh-timeouts/mesh-msg-34032090-9fb5-4b3e-b169-a945d1468848.json
#102081	mesh-thr-ef7498c1-c7b8-46c3-b533-d711a3616274	blocked/BLOCKED	no	timeout_or_worker_no_response	440	154	/tmp/alai/company-mesh-timeouts/mesh-msg-fd5a837d-c8c3-46ad-b2bb-6fc38c16d58d.json
#102083	mesh-thr-ecac2a6d-92ac-480e-b66e-d809aa0e6e04	blocked/BLOCKED	no	agent_runner_or_ollama_failure	1780	75	/tmp/alai/company-mesh-auto-responder/2026-05-26T14-16-50-228Z-mesh-msg-d90e62e3-bf6d-43da-825e-0e18abaf8d13.json
#102081	mesh-thr-526b7560-9278-4722-93ca-985d70e7a590	blocked/BLOCKED	no	blocked_unspecified_or_claim_gate	641	124	/tmp/alai/company-mesh-responder/2026-05-26T14-22-08-866Z-mesh-msg-c370552b-9c14-4737-bc9a-b36ccbcb01a.json
#102083	mesh-thr-c99828fd-f6d8-447f-99dc-f779cd412bb3	blocked/BLOCKED	no	timeout_or_worker_no_response	1568	223	/tmp/alai/company-mesh-timeouts/mesh-msg-7a537962-f6f0-418a-93b8-32a317dd882a.json

Task	Thread	Status/classes	Acceptable	Pattern	Prompt chars	Latency s	Evidence
#102081	mesh-thr-5cbbadc8-e238-4017-9b54-800c5088a0e9	answered/PASS	yes	none	38779	151	<code>/tmp/alai/company-mesh-responder/2026-05-26T14-27-57-032Z-mesh-msg-431fd915-c305-4336-99be-0f1ca3e1ac8e.json</code>
#102083	mesh-thr-4ec294f5-d1c2-43fe-98d9-2e7aaeb0953f	blocked/BLOCKED	no	blocked_unspecified_or_claim_gate	1204	139	<code>/tmp/alai/company-mesh-auto-responder/2026-05-26T14-28-23-453Z-mesh-msg-5e69f9b7-0b5a-4186-a8a6-866a3f612c18.json</code>
#102083	mesh-thr-33334359-3e83-4343-bbda-342f7304bdee	blocked/BLOCKED	no	blocked_unspecified_or_claim_gate	655	85	<code>/tmp/alai/company-mesh-auto-responder/2026-05-26T14-31-00-220Z-mesh-msg-e1fc9798-e0eb-482e-978b-b97d086be757.json</code>
#102083	mesh-thr-84961884-24e9-406b-bc36-bda72f807441	blocked/BLOCKED	no	partial_due_summary_only_evidence	563	44	<code>/tmp/alai/company-mesh-auto-responder/2026-05-26T14-34-51-053Z-mesh-msg-43d28653-f4b4-47a5-9229-9338be4c30d1.json</code>

Task	Thread	Status/classes	Acceptable	Pattern	Prompt chars	Latency s	Evidence
#102083	mesh-thr-f759f9d2-a62d-491d-9ecb-677fcfd808fd	answered/PARTIAL	yes	partial_due_summary_only_evidence	622	184	/tmp/alai/company-mesh-auto-responder/2026-05-26T14-38-26-267Z-mesh-msg-766b4c5e-cae6-444c-a09d-cf42398dc903.json

Quality Findings

- Path-only prompts are weak verifier inputs.** Several early Claude/agent-runner attempts blocked or timed out when the verifier did not have enough pasted evidence or reliable read access.
- Pasted artifact prompts improved outcome quality.** MC #102081 passed only after a sanitized pasted-artifact prompt with implementation evidence and code excerpts.
- Responder mode matters.** Proveo/eval using Claude review produced usable ANSWERED/PARTIAL outcomes after routing and max-turn/read-only fixes; agent-runner/Ollama path produced blocked failures.
- Timeouts are the dominant reliability issue.** Timeout/worker-no-response is the largest failure pattern in this sample.
- PARTIAL is useful and honest.** MC #102083 returned PARTIAL because artifact summaries were read but commands were not re-run; that is preferable to false PASS.

Recommendation

Hold controlled rollout. Keep P2P mandatory for H/risky tasks, but do not auto-send at dispatch until responder reliability and evidence-pack prompts are improved. Require pasted or readable evidence bundles for Claude-review verifiers.

Proposed Rollout Rules

- Keep current controlled rollout for H/backend/core/security/user-facing/deploy-impacting tasks.
- Do not enable automatic Company Mesh verifier send at dispatch yet.
- For required P2P, generate a compact evidence bundle before verifier prompt.

- Prefer Claude-review verifier mode for Proveo on evidence-heavy reviews; keep agent-runner as fallback only when local model health is known.
- Treat PASS/PARTIAL/ANSWERED with evidence paths as acceptable pre-verifier states; BLOCKED/timeout must not satisfy MC ready/done.
- Track retry count and first-success attempt in future runner evidence.

Evidence Artifacts

- Metrics JSON: `/Users/makinja/system/evidence/102080/p2p-verifier-metrics.json`
- This report: `/Users/makinja/system/evidence/102080/p2p-verifier-metrics-report.md`

Revision #1

Created 2026-05-26 15:28:47 UTC by John

Updated 2026-05-26 15:28:47 UTC by John