

COMPANION WORKSHEET

Chapter 2 — First-Response Evidence

Use this card only after the Chapter 2 preflight and response lab. Copy observed values; do not fill gaps with assumptions.

Run identity

Date and time

Host label

Ollama CLI version

Ollama API version

Loopback API address

Listener address reported by ss

Cloud-disabled evidence

Selected model

Exact model name

SHA-256 digest from /api/tags

Stored bytes

Parameter-size label

Quantization label

remote_model

remote_host

Both remote fields must be empty for the local-model lane.

Response and placement

Prompt

Generation done

Completion reason

Wall duration

Ollama total duration

Loaded bytes from /api/ps

VRAM bytes from /api/ps

Context length from /api/ps

Post-unload model absent

Four-state result

Use only these complete states: Known and controlled, Known external dependency, Unknown, or Not applicable.

Question	State	Evidence or next check
Was the API endpoint loopback-only?		
Was the selected inventory entry local rather than remote/cloud-backed?		
Did the exact model digest appear before and after generation?		
Where did Ollama report the model allocation?		
Were cloud features effectively disabled?		
Did the selected model unload afterward?		
Was this run performed with external network paths disabled?		

Bounded conclusion

Proved

Not proved

A valid conclusion names the exact run. It does not turn one local response into a permanent claim about the entire machine, application stack, network, or future configuration.