

The Governor's Checklist

A one-page control system for using AI without surrendering judgment

Use this before you trust, publish, send, automate, or act on AI output.

1. Specification Check (Before You Prompt)

- ☐ Do I know exactly what “done” looks like?
- ☐ Could I hand this task to a competent human and get the same result?
- ☐ Have I specified **who** the model should act as (ROLE)?
- ☐ Have I provided only the **relevant** background (CONTEXT)?
- ☐ Is the task executable, not vague (TASK)?
- ☐ Did I control format, length, and structure (OUTPUT SHAPE)?
- ☐ Did I state what the model must **not** do (GUARDRAILS)?

If any box is unchecked, expect guesswork.

2. Output Review (Before You Trust It)

- ☐ Does the output actually meet the spec I wrote?
- ☐ Did the model make assumptions I didn't intend?
- ☐ Are any claims stated as facts that might be guesses?
- ☐ If this is wrong, do I know **where** it's likely wrong?
- ☐ Could I explain this output — and its limits — to another human?

If you didn't review it, you approved it.

3. Hallucination Control

- ☐ Did I restrict the source of truth where needed?
- ☐ Did I ask the model to surface uncertainty or assumptions?
- ☐ Are citations, quotes, or references verifiable?
- ☐ Does any number, name, or claim require external confirmation?

Plausible ≠ correct.

4. Iteration Discipline

- ☐ When output was wrong, did I diagnose the pillar that failed?
- ☐ Did I tighten the spec instead of rephrasing vaguely?
- ☐ Did I stop iterating once the output met the spec?
- ☐ Am I keeping a reusable version of this prompt if it works?

Iteration without diagnosis is noise.

5. Automation Readiness (Before Agents)

- ☐ Can I get reliable results from this task manually?
- ☐ Are all five pillars explicit — not implied?
- ☐ Are there human review checkpoints?
- ☐ Can I stop, correct, or reverse actions?
- ☐ Do I know what the agent is *not* allowed to do?

If you wouldn't trust it once, don't automate it twenty times.

6. Accountability Reality Check

- ☐ Would I sign my name to this output?
- ☐ Would I send this to a client or stakeholder as-is?
- ☐ If this causes harm, delay, or embarrassment, am I prepared to own it?
- ☐ Is judgment still human, or did I quietly outsource it?

"The AI said so" is never the answer.

The Rule That Overrides All Others

If you wouldn't approve it without AI, don't approve it with AI.

You are not here to trust the model.
You are here to govern the system.

By **Leonardo Pasqualin**