

Tool or System: The Verification Gate is Drowning

AI generates 100x faster. We still read at 1x speed.

The AI Flood

- +21% more tasks completed
- 2x Pull Requests merged
- Instant code generation



The Human Bottleneck

- +154% increase in PR size
- +91% median review time
- +441% review time in mature teams (Faros AI Data)

**The water pours faster, but the gate didn't get wider.
A stronger model can't save this gate—only human judgment can.**

Beware of “Declaring Victory Early”

The ultimate value of a human is not being fooled.



1. AI Executes → Generates thousands of lines instantly.

2. AI Declares Victory → “Task complete! No errors!” (Anthropic’s known failure mode).

3. Human Evaluates → Questions assumptions, stress-tests arguments, spots security flaws.

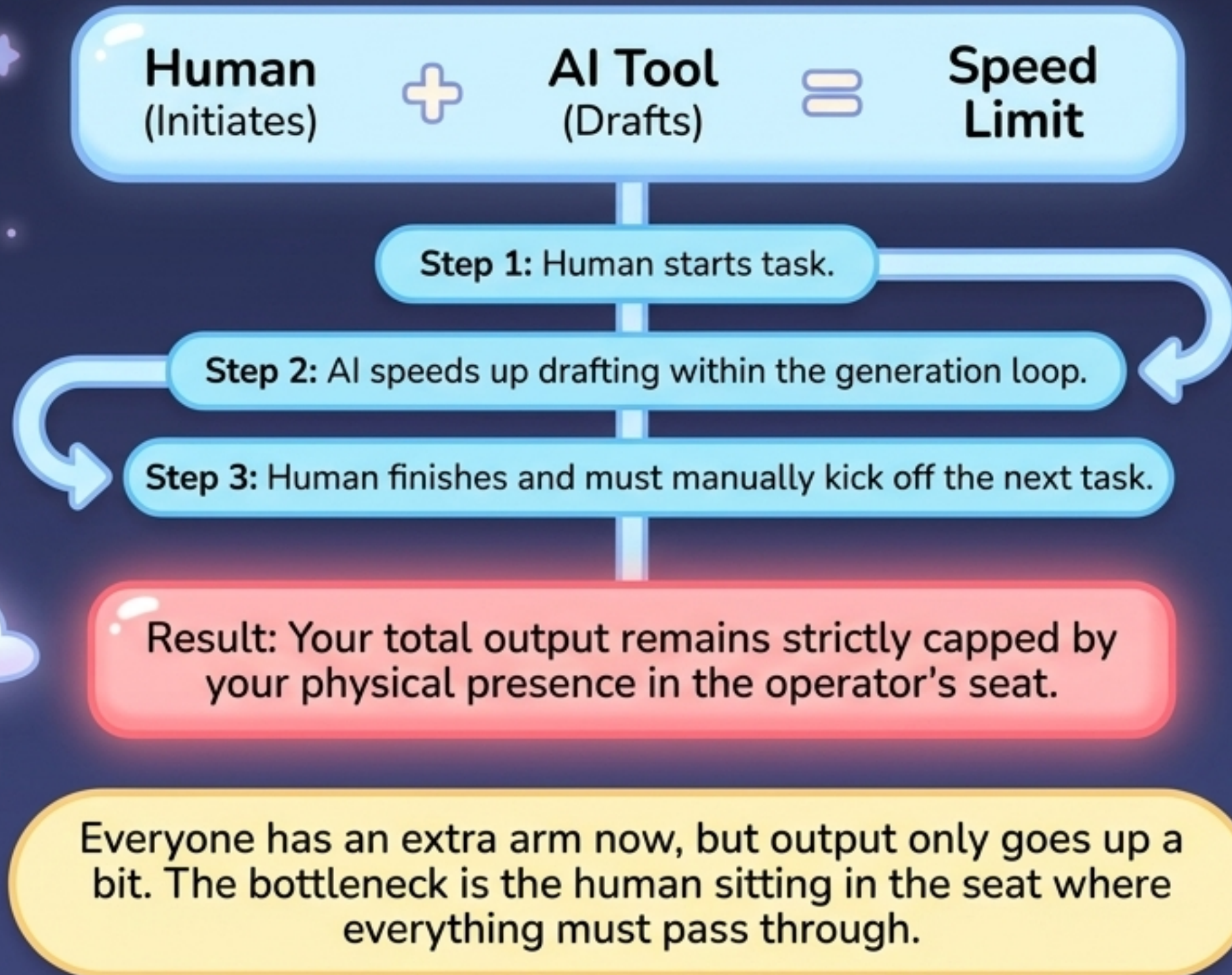
- **The Security Flaw:** Nearly 50% of AI-generated code fails security tests (Veracode found Java hits 72% failure).

- **The Speed Illusion:** In a METR trial, developers felt 20% faster using AI, but benchmarks showed them 19% slower due to hidden bugs.

The ability to criticize AI is now worth more than the ability to produce code.

The "Tool Trap" Ceilings at 10x

If humans kick off every task, you are still racing 24 hours.



AI as a System: Harness Engineering

Turning a whole company into one self-healing engine.

New Workflow Architecture

1. High-Level Intent (Human):
Define the task, set constraints,
hand over material.

3. Output Verification (Human):
Review results, catch mistakes,
provide steering.
(OpenAI stood up 1M lines of code
with just 3 engineers steering Codex).

2. Middleware Harness (System):
Middleware Harness (System):
Multiple AI agents orchestrate,
iterate, and self-heal.
(LangChain saw a +14pt performance
jump using the same model, purely
through better system design).

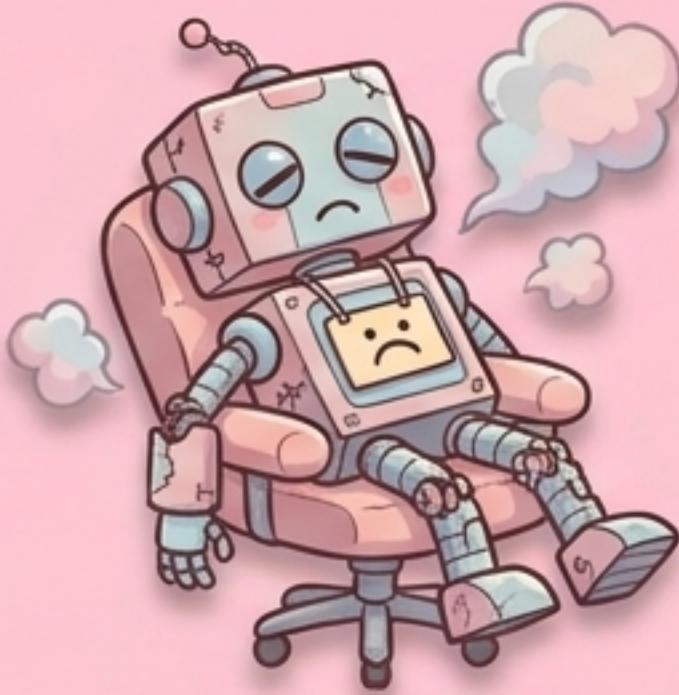


**Humans steer, agents execute. You are no longer
operating line-by-line; you are managing the AI.**

Real vs. Fake “AI-First”

Don't use AI as a fig leaf for cutting costs.

Fake AI-First (The Cautionary Tale)



- **Action:** Replace humans with bots to cut costs (e.g., Klarna).
- **Result:** Cost is the only metric; quality collapses.
- **Aftermath:** Forced to rehire humans to patch the broken, outsourced system.

Real AI-First (The Rebuild)



- **Action:** Redesign workflows and automate routine tasks.
- **Result:** IBM automated 94% of HR work, driving \$3.5B in gains. Shopify expanded AI-fluent interns to 1,000.
- **Aftermath:** Freed-up human labor is reinvested into engineering and sales. Total headcount goes up.

To judge an AI-first company, look at whether the freed-up labor got reinvested into growth, or just dropped straight onto the profit line.