

The Golden Test Set

10-20 examples of perfect output you test the AI against before rollout.

If you cannot say what **good looks like, you cannot automate it.**

Keep 10 to 20 examples of perfect output. Score the AI against them and tune the prompt until it passes consistently. It cuts verification time by an estimated third.

Week two of a pilot, before launch.

Want this working in your team?

Book a Focus Call

humanspark.ai/call