

Case Study: An Autonomous AI System That Runs Like a Production Line

FirstPass Systems — Production-Grade AI for Manufacturing & Operations. Anthony Ford, founder. Built and operated by the founder: a fully autonomous AI research and reporting system in continuous production since June 2026. The domain is financial markets; the engineering problem — unattended AI doing scheduled work, reliably — is the same one facing every operations team.

The Challenge

Most AI pilots die within a month — not because the AI is bad, but because they are built like demos. Nobody plans for the API dropping mid-run, the subprocess that hangs, the OS permission that silently blocks a scheduled job, or the output format that drifts and crashes the report. In a plant, nobody would accept a machine that "usually works." AI systems deserve the same standard.

To prove that standard, FirstPass built a system that runs entirely unattended: twice every weekday, an AI agent drives a live desktop application, runs a multi-factor analysis, generates a complete report with defined decision levels, and emails formatted deliverables — with no human touching it. A third scheduled job reviews every open item each afternoon against systematic rules and escalates anything stale.

The Approach

Every failure mode was treated the way a manufacturing engineer treats one: find the root cause in the logs, design the countermeasure, verify it fires, and leave it instrumented.

Failure mode found in production	Countermeasure
Upstream API drops connection mid-run	Automatic retry with verified-delivery check before re-sending
AI subprocess hangs on a stuck UI step	Watchdog kills the process tree at 45 min and triggers retry
Two agents contending for one application	Schedule-level mutual exclusion — jobs deliberately spaced apart
OS permission trap silently blocks scheduled jobs	Root-caused, all automation paths relocated and documented
One malformed record crashes the whole report	Row-level isolation — bad records get flagged, the report always ships
AI output format drifts day to day	Tolerant parsers with documented fallbacks
Silent delivery failure	Every send logged and counted; a missed delivery fires an alert instead of disappearing

The system also grades its own output: every recommendation is logged and scored automatically by a quantitative model with conservative assumptions. It cannot hide a bad result — the scorecard is honest by construction.

The Results

- **In continuous production since June 2026** — roughly 15 scheduled autonomous runs per week on standard hardware, unattended.
- **100% eventual delivery across an audited 28-run window** — even though 29% of first attempts hit an upstream API failure. Every incident either self-recovered or raised an alert; none disappeared silently.

- **Every delivery verified:** all outbound reports are counted against a send log, so retries never duplicate and misses never go unnoticed.
- **Costs engineered, not guessed:** per-run cost measured across 43 production runs, three AI model tiers benchmarked, and the quality/cost optimum selected from data — plus a further ~50% token reduction identified.
- **The system improves from its own data:** analysis of 63 real outcomes produced a clear rule (disciplined short holds won at an 81% rate; rule-breaking long holds lost money), converted into an automated escalation in the daily review the same day.

Why This Matters for Your Plant

Swap "market screener" for your use case — quality-anomaly triage, supplier-change coordination, an agent that answers questions from your SOPs, downtime root-cause summaries — and the engineering problem is identical: an AI agent doing scheduled, unattended work where failure has to be detected, contained, and recovered, not discovered a week later.

It works because it was built the way plants build product: watchdogs are andon cords, verified delivery is poka-yoke, and the failure-mode table above is an FMEA. Not a dev shop learning how factories work — a manufacturing engineer who ships. That is what FirstPass means by production-grade.

Next step: a free 30-minute working session. Two or three specific ideas for your operation and a live demo of a working system — no pitch deck, no commitment. If one idea is worth proving, a fixed-scope two-week pilot puts a working tool on your real data, and you keep it either way.

Anthony Ford · FirstPass Systems · Thorntown, Indiana · anthony.ford@firstpasssystems.com