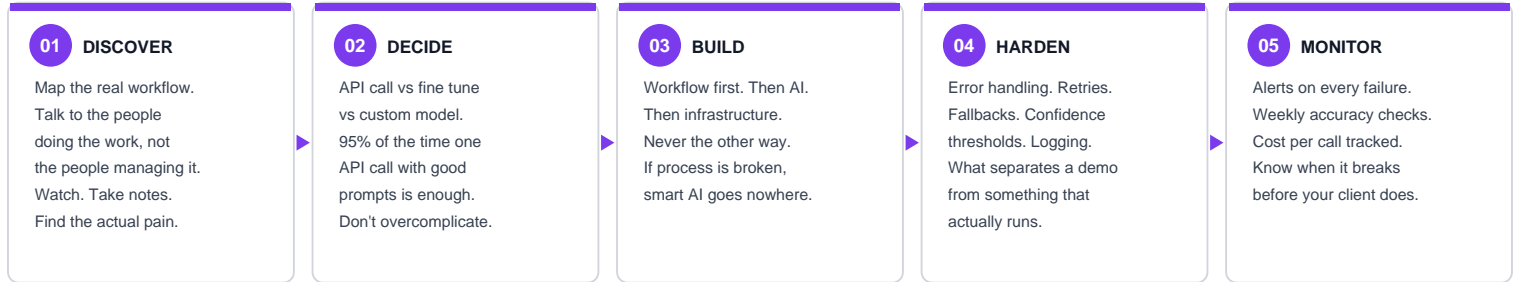




## The 5 Phases

Same order. Every project. No shortcuts.



## Stack Reference

What we use. 50+ builds.

LAYER	TOOLS	COST/MO	NOTE
AI / LLM	Claude, OpenAI, Gemini	\$50-200	One. Not three.
Orchestration	n8n, Make, Temporal	\$0-50	Glue layer.
UI / Dashboard	Retool, Bubble, custom	\$0-80	Team interface.
Database	Supabase or Postgres	\$0-25	Source of truth.
File storage	S3 or Cloudflare R2	\$0-10	Upload storage.
Monitoring	Logs + Slack alerts	\$0	Sleep at night.
<b>MONTHLY RUNNING</b>		<b>\$50-365</b>	Recurring.
<b>ONE TIME BUILD</b>		<b>\$1.5-5K</b>	Build + handoff.

## API or Custom Model?

**API call is enough** 95%

- ✓ Document reading / extraction
- ✓ Text or image classification
- ✓ Content generation / summaries
- ✓ Data matching, scoring, routing
- ✓ Email parsing and triage
- ✓ Image analysis or labeling

**Custom ML only if:** ~5% of projects

- ? Millions of items/hr throughput
- ? Zero training data for domain
- ? Sub 50ms inference needed
- ? Regulated full audit trail

### THE BUILD ORDER

Always this sequence.

- 1 WORKFLOW**  
Map process end to end.  
Fix what's broken first.
- 2 AI LAYER**  
Simplest AI that solves it.  
Usually one API call.
- 3 INFRA**  
Error handling, retries,  
monitoring, alerts, fallbacks.

### PRODUCTION CHECKLIST

- API returns garbage**  
Fallback to manual review
- Confidence below 70%**  
Flag for human check
- Request times out**  
Retry with backoff
- Unexpected format**  
Catch, log, alert team
- Accuracy drifts**  
Weekly sampling checks
- Cost spikes**  
Budget caps + alerts
- Latency degrades**  
Dashboard + threshold alerts

### COMMON MISTAKES

- X Building custom ML**  
API call works fine
- X Starting with the AI**  
Map workflow first
- X No error handling**  
Demo works, prod crashes
- X Using 3 AI providers**  
One is enough
- X Ship and forget**  
No monitoring or alerts
- X Automate broken process**  
Chaos just runs faster
- X Overengineering v1**  
Ship simple, iterate

### THE 90/10 RULE

# AI = 10%

The other 90% is error handling, edge cases, and infrastructure. That's the job.

AI INFRASTRUCTURE

### ASK ANY AI DEV THIS

## "What happens when the API returns garbage at 2am?"

*If they talk about the model, run.*

### BY THE NUMBERS

- 50+** AI apps shipped
- 95%** solved with one API call
- 11 days** avg build time
- \$1.5-5K** one time build cost

*Based on real client builds.*