AI Implementation Execution Playbook

A practical, no-hype system to help an operations-led business choose the right AI initiative, select the right technical approach, execute with control, and prove business value quickly.

Version: March 9, 2026 | Use this with your leadership, operations, and technical teams.

How to Use This Playbook

1. Run Sections 1-3 in week one to choose the right first workflow.
2. Use Sections 4-8 to design an implementation that can survive production reality.
3. Use Sections 9-12 to manage execution, adoption, and executive reporting.
4. Use Section 13 before launch to reduce avoidable failures.

> Most failed AI projects do not fail because the model is bad. They fail because workflow scope, ownership, baseline metrics, and launch criteria are unclear.

What Good Looks Like in 90 Days

- One priority workflow implemented end-to-end (not a demo).
- Clear baseline vs. post-launch metrics.
- Human-in-loop controls for edge cases.
- Documented runbook, owner, and escalation path.
- Decision made on scale, iterate, or stop based on measured outcomes.

Created by Phil Maher - AIAdvocate.org

1) Opportunity Prioritization Scorecard

Score each candidate workflow from 1 (low) to 5 (high). Pick the top item by weighted score.

Criterion	How to Score	Weight	Score (1-5)
Business impact	Revenue, cost, risk, or cycle-time effect if improved.	30%	
Volume/frequency	How often the workflow runs and how many staff touch it.	20%	
Data readiness	Input quality, structure, and access today.	15%	
Process maturity	Documented steps, known exceptions, stable ownership.	15%	
Adoption likelihood	User willingness and manager support to change behavior.	10%	
Implementation effort	Technical complexity and integration burden (reverse score).	10%	

```
Weighted Score = SUM(criterion score x weight). Target first pilot: >= 3.7 / 5 with clear owner.
```

Candidate Workflow List

Workflow	Owner	Current pain	Weighted score	Decision

Example: NovaPay (25-person fintech)

Here is how NovaPay, a fictional 25-person fintech company, filled in the scorecard to choose their first AI pilot.

Criterion	Weight	Invoice Reconciliation	Client Onboarding Docs	Compliance Report Generation
Business impact	30%	4	3	4

Criterion	Weight	Invoice Reconciliation	Client Onboarding Docs	Compliance Report Generation
Volume/frequency	20%	5	3	4
Data readiness	15%	4	2	3
Process maturity	15%	3	2	4
Adoption likelihood	10%	4	3	3
Implementation effort (reverse)	10%	3 (reverse = 3)	2 (reverse = 4)	3 (reverse = 3)
Weighted Score		**3.95**	**2.75**	**3.60**
Decision		Pilot — Q2	Defer — data not ready	Phase 2 — after invoice pilot

> NovaPay chose invoice reconciliation: highest weighted score, clear owner (Head of Finance), and clean input data. The compliance workflow scored close but depended on completing a data migration first.

2) 2-Hour Discovery Session Script

Participants

- Workflow owner (required)
- 2-3 frontline users
- Ops/finance stakeholder for KPI validation
- Technical lead or implementation partner

Agenda (120 minutes)

Time	Activity	Output
0-20 min	Map current state, step-by-step (real sequence, not SOP ideal state).	Current process map
20-45 min	Identify failure points, rework loops, decision bottlenecks.	Pain inventory
45-70 min	Define measurable success criteria and non-negotiable constraints.	KPI + guardrail draft
70-95 min	Clarify input data sources and exception patterns.	Data + exception map
95-120 min	Confirm pilot scope, owner, decision cadence, go/no-go date.	Pilot charter draft

Key Questions

- Which step consumes the most human hours today?
- Which failure causes the highest downstream cost?
- What decision still requires human judgment even after automation?
- What output quality threshold is acceptable at launch?

3) Baseline Metrics Worksheet

Capture baseline before any implementation. Without this, ROI claims are not credible.

Metric	Current baseline	Data source	Owner	Target (90 days)
Cycle time (start to completion)				
Touch time (human effort minutes/unit)				
Error/rework rate				
SLA miss rate				
Volume per week				
Cost per completed unit				

Measurement Rules

- Use at least 4 weeks of historical data where possible.
- Separate normal cases from exception-heavy cases.
- Define metric owner and update cadence (weekly recommended).
- Track both speed and quality; speed-only optimization usually breaks operations.

> If baseline data does not exist, spend one week instrumenting the workflow first. This pays for itself.

4) Pilot Charter Template

Pilot objective	
In-scope steps	
Out-of-scope steps	
Primary owner	
Decision maker	
Start date / end date	
Success criteria (must hit all)	
Stop criteria (if any occur)	

Quality Gate Design

Output type	Acceptance threshold	How reviewed	Escalation rule

Human-in-Loop Policy

- Low-risk cases: auto-process with random QA sampling.
- Medium-risk cases: queue for human confirmation.
- High-risk cases: mandatory human approval before release.

5) Data and Governance Checklist

Data Readiness

Control	Status (Yes/No)	Gap	Owner	Due date
Input sources identified and documented				
Field-level data quality checked				
PII/sensitive fields tagged and handled				
Retention/deletion policy defined				
Audit trail requirements documented				

Operational Controls

- Rollback path documented and tested.
- Alerting thresholds defined for quality drift and volume anomalies.
- Named incident owner and escalation chain.
- Change log maintained for prompts, models, and workflow rules.

6) Build-vs-Buy Decision Matrix

Decision factor	Build internally	Buy/vendor	Hybrid
Workflow uniqueness	High	Low to medium	Medium to high
Data sensitivity / residency	Strict control needed	Vendor acceptable	Sensitive core in-house
Time to value	Longer	Faster	Moderate
Internal technical depth	Required	Lower	Moderate
Long-term TCO predictability	Higher control	License risk	Balanced

Vendor Due Diligence Questions

- How is customer data used for training, if at all?
- Can we export all data and workflow history at any time?
- What are failure mode SLAs and support response guarantees?
- Can controls be tuned by workflow risk level?
- What is your roadmap for auditability and access controls?

> Pick the option that best fits your next 12 months of execution, not a theoretical 5-year architecture.

Model and Tool Selection Guide

Choosing the right model architecture early prevents costly rework. Use this table to match your workflow requirements to the right technical approach.

Factor	API-Based (GPT-4, Claude)	Open Source (Llama, Mistral)	Fine-Tuned / Custom
Data privacy	Data leaves your infra	Full control, on-premise	Full control
Setup time	Hours to days	Days to weeks	Weeks to months
Cost at low volume	Low (pay per call)	Higher (infra needed)	Highest (training + infra)
Cost at high volume	Can get expensive	Predictable, lower	Lowest per-unit
Customization	Prompt engineering	Moderate (RAG, adapters)	Highest
Best for	Prototyping, standard tasks	Sensitive data, predictable cost	Domain-specific accuracy

> **Decision Shortcut:** Start with an API-based model (Claude or GPT-4) for your first pilot unless data sensitivity requires on-premise. Most workflows don't need fine-tuning — RAG + prompt engineering handles 80% of use cases. Only invest in custom models after you've validated the workflow works.

Architecture Patterns to Know

- **Direct API call:** Best for classification, summarization, extraction on individual items.
- **RAG (Retrieval-Augmented Generation):** Best for Q&A over internal documents, knowledge bases, policies.
- **Agent workflows:** Best for multi-step processes requiring tool use, decisions, and integrations.
- **Batch pipeline:** Best for high-volume processing (reports, document classification, data enrichment).

7) Cost and ROI Model (Template)

```
Annual Savings = (Weekly hours reduced x loaded hourly rate x 52) + (rework cost reduction) + (SLA penalty reduction)
```

```
Payback (months) = Implementation cost / Monthly savings
```

```
12-Month ROI (%) = ((Annual savings - Implementation cost) / Implementation cost) x 100
```

Input	Value	Notes
Current manual hours/week		
Loaded labor rate (USD/hour)		
Estimated automation/augmentation %		
Current rework cost/month		
Implementation cost (one-time)		
Ongoing operating cost/month		

Sanity Checks

- Model conservative, expected, and aggressive scenarios.
- Exclude gains that depend on unrelated organizational changes.
- Track realized ROI monthly for the first quarter post-launch.

8) 12-Week Execution Plan

Week	Primary focus	Required output
1-2	Discovery + baseline instrumentation	Approved pilot charter and KPI baseline
3-4	Solution design + controls	Architecture, quality gates, incident flow
5-7	Build + internal test	Working pipeline and defect log
8-9	Controlled pilot rollout	Measured pilot results and adoption data
10	Hardening + runbook	Launch runbook and owner handoff plan
11	Launch decision	Scale/iterate/stop decision record
12	Scale prep	Next workflow backlog and staffing plan

Weekly Operating Cadence

- 30-minute delivery standup (execution blockers).
- 30-minute quality review (error themes and threshold drift).
- 30-minute stakeholder review (KPI trend and decisions).

Common Delays and How to Handle Them

Delay Pattern	Typical Impact	Prevention
Data access takes longer than expected	+1-2 weeks at weeks 3-4	Get IT/security sign-off in week 1, not week 3
Stakeholder availability for discovery	+1 week at weeks 1-2	Book all sessions before kickoff
Edge cases discovered during pilot	+1-2 weeks at weeks 8-9	Budget 30% of pilot time for exception handling
Integration complexity underestimated	+1-2 weeks at weeks 5-7	Prototype integration in week 3, not week 5

> Build 2 weeks of buffer into your executive timeline. A 12-week plan communicated as 14 weeks is better than a 12-week plan that slips to 16.

9) Executive Status Template (Weekly)

Week ending	
Overall status (R/A/G)	
Key KPI movement	
Top 3 risks	
Decisions needed this week	
Next week's objective	

Risk Register

Risk	Likelihood	Impact	Mitigation owner	Mitigation action

Decision Log

Date	Decision	Rationale	Owner	Review date

10) Adoption and Change Management Checklist

Adoption control	Status	Owner	Target date
Frontline users trained on new workflow path			
Escalation path documented for edge cases			
Old process retired or gated to avoid duplicate work			
Manager dashboard includes new KPIs			
Weekly feedback loop scheduled with users			

RACI (Who Owns What)

Workstream	Responsible	Accountable	Consulted	Informed
Workflow design				
Model/prompt quality				
Data controls				
Launch decision				

Common Resistance Scenarios

Scenario	What They're Thinking	How to Address
Frontline users ignore the new tool	"This will be replaced in 6 months anyway" or "I'm faster doing it my way"	Run a side-by-side demo with real cases. Show time savings on their actual work. Give them 2 weeks to try it before mandating.
Manager questions the ROI	"Prove this isn't just a tech experiment"	Share baseline metrics from Section 3. Show week-over-week KPI movement. Frame in dollars, not percentages.

Scenario	What They're Thinking	How to Address
Team fears job replacement	"Am I training my replacement?"	Be direct: "This handles the repetitive 60%. You handle the judgment calls, exceptions, and client relationships." Show how their role evolves, not disappears.

Role-Specific Messaging

- **Frontline users:** "This handles the repetitive parts so you can focus on the work that requires your judgment."
- **Managers:** "You'll see faster throughput and fewer errors. Here are the KPIs to watch."
- **Executives:** "This is a measurable operational improvement. Here's the ROI model and the 90-day proof point."

11) Launch Readiness Checklist

- Baseline and pilot metrics reviewed and approved.
- Known failure modes documented with fallback handling.
- On-call owner and escalation tree confirmed.
- Rollback tested in non-production environment.
- Compliance/security controls validated for the use case.
- User-facing change communication distributed.
- Week-1 post-launch monitoring plan scheduled.

Go / No-Go Form

Decision	GO / NO-GO
Reasoning summary	
Conditions for go-live	
Required post-launch actions	
Decision owner	

12) Common Failure Patterns and Fixes

Failure pattern	What it looks like	Fix
Wrong first use case	Low adoption, weak business impact	Re-run scorecard with hard KPI weights
No clear owner	Decisions stall, issues linger	Assign accountable owner with authority
No baseline metrics	Cannot prove value	Instrument workflow before scaling
Speed over quality	Error/rework spikes post-launch	Define quality gates and review thresholds
Tool-first architecture	Workflow mismatch and workarounds	Design from workflow constraints first
No rollback path	Incidents create operational chaos	Test rollback before launch

13) Do This Next Week

1. Pick 3 candidate workflows and run the prioritization scorecard.
2. Schedule a 2-hour discovery session with frontline users.
3. Capture baseline metrics for at least one workflow.
4. Create a pilot charter with explicit success and stop criteria.
5. Set a launch decision date now, not later.

> This playbook is designed so you can execute with your current team. Use it whether you work with an external implementation partner or not.

Validate Your Results

- After pilot launch, track realized ROI monthly for the first quarter.
- Compare actual time savings, error reduction, and cost impact against your baseline (Section 3).
- If ROI is below 50% of projections after 60 days, revisit scope, not the technology. The workflow design is usually the issue.
- Document what worked and what didn't — this becomes your playbook for the next workflow.

What to Do After This Playbook

- Take the free AI Readiness Assessment at **aiadvocate.org/ai-readiness-assessment** to benchmark where your organization stands.

- See real implementation case studies at **aiadvocate.org/work** to see how these frameworks translate into production systems.
- If you want help implementing: Book a strategy call at **aiadvocate.org/book-call**. Phil Maher has delivered 12+ AI systems using this methodology. The first call is free and focused on your specific situation.