Proposed Rollout Playbook

A pilot plan for testing one AI-assisted workflow with synthetic or approved data. The plan draws on nine years of operations experience and two portfolio prototypes. It does not claim that a team pilot has already happened.

CURRENT EVIDENCE	The n8n workflow ran connected tests with synthetic messages and fixed demo rules. About 40% is a coverage estimate based on the narrow FAQ scope, not a measured acceptance rate. SupportOps uses synthetic policies and a small live evaluation. No company-wide adoption or measured labor savings are claimed.

PHASE 1 · DEFINE THE PILOT

Choose one repetitive workflow with a named owner and a clear manual baseline. Agree on eligible items, quality criteria, data boundaries, and the decisions a person must keep. Record the targets before testing.

PHASE 2 · RUN IN SHADOW MODE

Use synthetic or approved work. The tool prepares a draft while the person completes the normal process. Compare quality, edits, human working time, and errors. Nothing reaches a customer or system of record without approval.

PHASE 3 · TEST WITH 2–3 AUTHORIZED USERS

Give each user a short working session and a task guide. Observe whether they can complete the workflow without prompting. Record hesitations, rejected outputs, and one change made from feedback.

PHASE 4 · DECIDE WHETHER TO EXPAND

Compare the results with the pre-set gates. Expand only if quality holds, approved-without-edit meets the target, and users choose to return. If a gate is missed, revise the workflow or stop the pilot and record why.

PROPOSED EXIT GATES

Measure	Proposed gate	How to record it
Quality	No worse than the agreed manual baseline	Score comparable tasks with the same criteria
Approved without edit	At least 50%, set before the pilot	Count outputs sent unchanged; track edited approvals separately
Voluntary repeat use	At least 60% of trained users return in week 4	Count users who choose the tool, not auto-processed volume
Safety	No error-escape rate above the manual baseline	Review approved outputs later found wrong

Proposed Measures

Define each measure before the pilot. Record observations separately from targets. Use comparable tasks and include review and correction time.

Measure	Definition	Source
Eligible use	Share of items users choose to put through the tool from the agreed eligible set	Tool log plus eligible-volume count
Approved without edit	Output approved and used with no text or routing change	Approval and edit record
Approved with edit	Output used only after a person changes it; report this separately	Approval and edit record
Human working time	Minutes a person actively spends reading, deciding, editing, and correcting	Timed task observation
Elapsed resolution time	Clock time from arrival to completion, including queue and wait time	System timestamps
Voluntary repeat use	A trained user chooses the tool again in a later week	Distinct users by week, excluding automatic processing

SAVINGS RULE

Estimate labor hours saved only from the change in human working time on comparable tasks. Do not use elapsed resolution time for labor savings. Report task count, baseline, review time, correction time, and limits beside the estimate.

COUNTER-METRICS

Track wrong outputs that passed review, unsafe or unsupported actions, and approval times too short for a real check. Pause the pilot if quality falls below the manual baseline or a data-boundary rule is breached.

WEEKLY REVIEW

Report the target, the observed result, one failure pattern, one user friction point, and one change. Keep proposed thresholds in one column and measured observations in another.

PILOT RECORD

Keep the task set, data type, user count, prompts, scoring criteria, edits, failures, and dates. Publish only results collected with permission. A guide is preparation; observed use is separate evidence.

Proposed Usage Policy

Rules to review with the workflow owner and IT or Security before a pilot. Adapt the details to the approved systems, data, and risk level.

1 · PEOPLE OWN DECISIONS

AI output is advisory. A named person reviews each draft, coaching note, route, or escalation before it affects a customer or employee. Nothing is sent automatically.

2 · EVIDENCE AND UNCERTAINTY STAY VISIBLE

Show the source material used, separate exact excerpts from generated commentary, and state what confidence labels mean. Weak evidence becomes a handoff, not a guess.

3 · USE THE LEAST ACCESS

Grant only the permissions needed for the pilot. Keep credentials in the platform store. New scopes need a written reason and review.

4 · SET DATA BOUNDARIES

Use synthetic or approved data for tests and training. Do not place secrets, payment data, or regulated data in prompts. Confirm provider terms and retention before live use.

5 · LOG THE WORK, NOT THE PERSON

Record the input reference, output, route, human decision, edit status, error, and timing needed to evaluate the workflow. Limit access and retention. Do not turn the log into employee surveillance.

6 · PLAN FOR FAILURE

Provide a one-step way to flag a bad output, a named owner for review, and an off switch. When the tool stops, work continues through the manual process.

IMPLEMENTATION CLAIMS CHECK

Portfolio project	What ran	Current limit
n8n email triage	Connected Gmail, Slack, Jira, and Sheets test workflow using synthetic messages and fixed demo rules; OpenAI path remains in the workflow	No production inbox, real customer reply, adoption measure, or measured time savings
SupportOps	Synthetic policy retrieval, cited answers, review handoffs, and a small live regression set	No live support system, customer data, or team pilot