

AI at your utility

Lessons for turning a first pilot
into dependable everyday work

16 lessons. A 90-day plan. Reusable worksheets.



READ THIS FIRST

A working guide for your utility

Practical lessons for electric, gas, and water teams choosing, testing, and operating AI. Use this guide with the companion AI journey PowerPoint template.

Who this is for

Executives, operational managers, engineering and regulatory teams, customer service leaders, IT, cybersecurity, and the people who will use the tools. A small utility may assign several responsibilities to one person. The responsibilities still need explicit owners.

How to use it

Read the roadmap and failure patterns first. Use the 16 lessons to design one bounded pilot, then complete the worksheets with your team. Bring the PowerPoint to a working session or an approval meeting. Every example and numerical threshold marked illustrative needs local validation.

Find what you need	Pages
Journey and recurring failure patterns	3-4
16 implementation lessons	5-20
Electric, gas, water, and InnPower examples	21-22
90-day plan and reusable worksheets	23-26
Glossary and source notes	27-28

What supports the guidance

This guide synthesizes Senpilot's 2025 and 2026 AI x Utilities handbooks and adds editorial implementation tools. The handbooks primarily describe electric utilities and reflect a vendor's experience. Gas and water examples here are illustrative adaptations. Anonymized stories are reported by Senpilot, not independently verified case studies.

Free to use

Download, share, and adapt this guide and the companion presentation for your team's planning and training. Please retain the aiforutilities.org attribution and sponsor credit when redistributing these resources. Sponsorship does not establish a procurement recommendation.

Source basis: Senpilot handbooks [H25, H26]. Full references and limitations on page 28.

THE JOURNEY

Five stages, five useful decisions

Progress comes from resolving a specific question at each stage. A pilot can move back to an earlier stage when the evidence changes.

Stage	Work to complete	Decision evidence
01 Understand	Observe one recurring workflow. Establish its burden and the consequences of error.	A named owner, a process map, and a measured baseline.
02 Choose	Compare a short list of tasks. Check whether rules, search, or process simplification would suffice.	A selected workflow with a credible benefit and manageable exposure.
03 Prepare	Approve data access. Define human review, test cases, and operational boundaries.	A pilot charter, representative data, and agreed acceptance criteria.
04 Prove	Test against current practice, including review time, difficult cases, and failures.	Evidence to proceed, revise, or stop. Record limitations alongside results.
05 Grow	Assign service ownership, monitor quality, and reuse controls in the next workflow.	A supported service with measured adoption and an expansion decision.

A foundation sized to the task

An asset planning assistant may need a reconciled asset view and approved health formulas. A document search pilot may need only a controlled library with current permissions. Define the minimum reliable foundation for the selected workflow. Expand it when the next decision requires more.

First working session

Ask each participant to bring one real task, one recent example, and one reason the work is difficult. Choose the workflow before discussing products. End the session with an owner and the next evidence to collect.

Adapted from H25, printed pp. 3, 6–10, and H26, pp. 12–14, 21–26.

FAILURE PATTERNS

Three ways a program gets stuck

The 2026 handbook describes three anonymized experiences. Treat them as useful warning patterns, not a representative survey of the sector.

Too much at once

In Senpilot's 'Utility Co' account, parallel experiments multiplied before teams had defined readiness and success. The program reportedly had 27 pilots and none in production. The practical lesson is to limit work in progress to what business owners, reviewers, and IT can support. An impressive backlog has little value if no workflow becomes dependable.

Waiting for certainty

The 'LDC Co' account describes a utility that kept waiting for a mature, guaranteed use case. A bounded experiment offers a more useful answer than another round of speculation. Make the uncertainty explicit, set a spending and time limit, and decide what result would justify another step.

Governance without delivery

The 'DSO Co' account describes an AI committee that produced a framework while operational work developed elsewhere. Policy and delivery need a shared backlog. Name the people who can approve a pilot, what evidence they need, and when they will decide.

Watch for	A practical correction
More pilots than available reviewers	Pause new starts until an owner can operate the existing work.
Repeated demos without access to your data	Require a bounded test on an approved representative sample.
No decision after several meetings	Assign a decision owner and an evidence deadline.
Benefits discussed only in percentages	Record the baseline, denominator, review effort, and actual use.

A useful portfolio rule

For a first program, one well-supported pilot is enough. Add a second only when the same team can sustain both. The right limit depends on capacity and risk, not utility size alone.

Reported examples: H26, pp. 12–14. Corrections and portfolio rule are editorial guidance.

LESSON 01

The workflow is the unit of change

A tool can make one step faster while leaving the overall process slower. Map the complete job, including handoffs and corrections.

What the handbooks teach

The 2026 handbook emphasizes process mapping because informal steps often determine whether AI fits daily work. The spreadsheet someone repairs, the engineer who checks an exception, and the email approval may matter more than the formal procedure.

What to do

- 01 Watch a practitioner complete several real cases. Record the trigger, inputs, systems, decisions, output, recipient, and waiting time.
- 02 Separate staff effort from elapsed time. A five-day wait for approval is different from five days of analysis.
- 03 Identify the step AI might assist and the steps that should remain deterministic or require professional judgment.
- 04 Map the exception route and manual fallback with the same care as the normal route.
- 05 Agree who accepts the finished output and what evidence they need to do so.

Utility example

For an asset condition report, the workflow might include obtaining GIS records, matching inspection IDs, applying approved formulas, drafting a narrative, engineering review, and document approval. Drafting takes less time only if the reviewer can quickly trace every material statement to an asset record or calculation.

Your next deliverable

Create a one-page map with columns for step, owner, source system, effort, failure mode, and required approval. Mark the exact beginning and end of the pilot. Keep 'report generation' out of scope until those boundaries are clear.

Evidence that the lesson is working

Reviewers agree on what a complete output looks like. The team can explain where time goes today, which step changes, and who handles work that AI cannot complete.

LESSON 02

A narrow first use case is easier to learn from

Choose a repeatable task with accessible evidence, a willing owner, and a meaningful but containable consequence of error.

A useful starting boundary

An internal draft or an assistant that only reads approved records can be easier to evaluate than autonomous customer decisions or control actions. 'Internal' does not automatically mean low risk: confidential data, engineering advice, or employment decisions can still carry serious consequences.

Criterion	Question to answer
Value	How frequently does this occur, and what burden or quality problem would improve?
Data	Can we access representative inputs with the right permissions now?
Verification	Can a qualified reviewer determine whether the output is acceptable?
Integration	Can we run a useful trial without changing critical systems?
Ownership	Will one team own adoption, exceptions, and the eventual service?
Consequence	Can we detect, contain, and reverse a mistake before it causes harm?

Separate gates from scores

A high value score cannot compensate for missing authority to use data or unacceptable safety exposure. Check those conditions first. Then rank eligible ideas. Use the scorecard on page 25 to make tradeoffs visible, and keep a written reason for the final choice.

Illustrative first pilot

Draft internal answers from a small, approved procedures library, with a source link for each substantive statement and staff approval before use. Define excluded topics and what happens when sources conflict. This is an example to assess, not a universal recommendation.

Handbook basis: H26, pp. 12–14. Selection framework is editorial guidance.

LESSON 03

A trusted view needs ownership and lineage

The 2025 handbook's 'AI Cake' connects a single asset view, utility-specific scoring, and AI applications. Its durable lesson is that context matters as much as access.

Make the source of each field clear

A shared view does not require moving all utility data into one physical database. It does require agreed identifiers, authoritative sources, definitions, and refresh expectations. A model cannot resolve disagreement about which system owns a field without an explicit rule.

What to establish

- 01 An asset ID crosswalk across GIS, work management, inspections, and finance. Keep original IDs for reconciliation.
- 02 A field owner for material attributes, with units, allowed values, and an effective date.
- 03 A record of source, extraction time, transformation, and any human correction.
- 04 Permission checks on both source access and the results returned to each user.
- 05 An explicit rule for stale, contradictory, or missing records. Surface the conflict instead of silently choosing a value.

Utility example

A transformer replacement recommendation can look precise while using an obsolete installation date or the wrong capacity unit. Before testing the recommendation, reconcile those inputs and show the source date. Keep the approved asset health equation version alongside the result.

Your next deliverable

Create a data contract for the chosen workflow: required fields, authoritative system, owner, refresh interval, permission boundary, and treatment of missing values. Begin with the smallest useful set of fields.

Evidence that the lesson is working

A reviewer can trace a material output back to its input records and identify who can correct them. The team knows whether the assistant is answering from a current feed or a dated snapshot.

LESSON 04

Missing data is a managed condition

AI can help find gaps and suggest repairs. It cannot turn an inferred value into an observed fact.

Keep evidence and inference separate

The 2025 handbook discusses incomplete, duplicate, noisy, and outdated data, and the need for feedback into field collection. Extend that into a clear policy: preserve the original value, label any proposed repair, and record who accepted it. A generated estimate should never silently overwrite the system of record.

Condition	Practical treatment
Missing asset attribute	Leave it unknown, request collection, or use a clearly labeled estimate if the workflow permits it.
Conflicting records	Route to the data owner. Retain both source values and the resolution.
Unmatched asset IDs	Hold the join for review. Test for false matches as well as missed matches.
Stale observations	Show age and refresh status. Define when the record becomes unusable.
AI-suggested correction	Keep the proposal, supporting evidence, reviewer, and approved change separately.

Prioritize collection by the decision

Do not delay a document search pilot to clean the entire asset estate. Do stop an asset decision that depends on an unverified critical attribute. Rank gaps by their effect on the selected decision, then assign the highest-impact items to a collection or validation process.

A useful quality measure

Track required-field completeness, duplicate rate, failed joins, age of records, and reviewer overrides. Report these by dataset or asset class. One aggregate 'quality score' can hide a serious local problem.

Handbook basis: H25, printed pp. 7–8; H26, p. 25. Data handling rules are editorial.

LESSON 05

Different tasks need different methods

The label 'AI' covers several kinds of systems. Select the approach that matches the job and the evidence you can obtain.

Work to do	Approach to evaluate	What to verify
Apply a tariff or an approved health formula	Deterministic rules and calculations	Rule version, units, boundary cases, and approval.
Find a policy or draft a source-based summary	Search with retrieval-assisted generation	Source support, date, permissions, and refusal when evidence is missing.
Forecast demand or flag likely asset failures	Statistical or machine learning models	Historical backtests, leakage, error by segment, and uncertainty.
Review images or inspection documents	Vision or extraction models	Missed defects, false alarms, image quality, and human review.
Carry out several steps across systems	An agent with approved tools	Action permissions, approvals, logs, retries, and recovery.

Use AI where uncertainty is acceptable

A language model can draft a narrative around a verified calculation. Keep the calculation itself in a tested function when exactness matters. A generated explanation is not proof that the underlying recommendation is correct or that it identifies a causal relationship.

A forecast needs its own evaluation

Compare to a credible existing baseline, such as the utility's current forecast or a simple seasonal method. Hold out time periods, test peak conditions separately, and examine errors across relevant asset or customer groups. A good average can hide a poor result when the utility most needs the system.

Your next deliverable

Write a short method choice: why the task needs AI, what simpler alternative you considered, and how a reviewer can verify the output. Include this in the procurement brief.

Handbook basis: H26, pp. 25–26. Method distinctions and evaluation advice are editorial.

LESSON 06

Integration is part of the product

A successful demo does not establish that an assistant can work with your identifiers, access rules, data refreshes, and outages.

Test the whole path

The 2026 handbook distinguishes ingestion, analytics, workflow automation, and knowledge retrieval. Each needs a different integration contract. Document where data enters, what the assistant can read or change, and how the final output reaches the team that needs it.

Questions to resolve early

- 01 Is the pilot using a file snapshot or a live connection? Who maintains it, and how will a user see freshness?
- 02 Does the integration preserve the source system's permissions? Test with ordinary and restricted users.
- 03 What happens when a dependency is unavailable, slow, or returns incomplete data?
- 04 Can a retry create a duplicate ticket or update the wrong record? How will the system prevent or detect that?
- 05 Can you export the inputs, approved outputs, logs, and configuration in usable formats?

Keep action permissions explicit

Begin with read access or draft creation when that is enough to prove value. Add write actions only after the business owner and technical reviewers approve the action scope. Each action should have a permitted target, required fields, an approval rule, and a recovery procedure.

Illustrative boundary

An assistant may draft an inspection follow-up in a work queue. The pilot excludes closing work orders, changing asset health scores, issuing switching instructions, or sending customer messages. A supervisor approves the draft through the existing workflow.

Evidence that the lesson is working

The team can demonstrate a normal run, stale data, a permissions denial, a dependency outage, and a duplicate retry. Every case has an understood outcome and an owner.

LESSON 07

Governance needs a decision owner

Effective governance connects operational accountability with IT, cybersecurity, privacy, and other relevant review functions.

Translate policy into a working route

The 2026 handbook warns that a committee can become detached from delivery. Its proposed allocation of governance to IT and cybersecurity is one organizational model. Each utility should assign responsibility according to its own authority, including safety, legal, records, labor, and regulatory review where relevant.

Responsibility	Accountable output
Executive sponsor	Business priority, funding limit, and authority for the next stage.
Business process owner	Scope, acceptable quality, human review, and benefit realization.
IT and cybersecurity	Architecture, identity, permitted integrations, and incident handling.
Data owner	Approved use, definitions, permissions, and correction process.
Relevant specialists	Applicable obligations, safety assessment, workforce impacts, and retention requirements.
Service owner	Monitoring, support, changes, training, and retirement of the service.

Separate assistance from control

Use a consequence-based review. A staff draft, a customer-facing answer, and a change to operational equipment have different failure costs. A planning pilot does not authorize live control of SCADA, switching, gas pressure, chemical dosing, protection settings, or other safety-critical functions.

Your next deliverable

A short approval record: intended use, prohibited use, affected people, responsible owners, open risks, evidence required, and decision date. Name who can pause the pilot immediately and who can authorize a restart.

Handbook basis: H26, pp. 12–14. Responsibility model and boundaries are editorial guidance.

LESSON 08

Security covers every copy of the data

Hosting location is one part of the review. Data may also appear in prompts, embeddings, logs, backups, support systems, and third-party services.

Map the complete data path

The 2026 handbook emphasizes residency and access controls. Before using sensitive information, document where each data type is stored and processed, who can access it, how long copies remain, and how deletion works. Confirm requirements with the utility's authorized reviewers rather than assuming one rule applies to every jurisdiction.

Evidence to request

- 01 A data-flow diagram covering model providers, subprocessors, support access, telemetry, and backups.
- 02 Contract terms and configuration evidence for training use, retention, processing locations, and deletion.
- 03 Single sign-on, role-based access, separation between customers, and permission-aware retrieval.
- 04 An inventory of components and versions, patch ownership, and a process for supplier or model changes.
- 05 An incident route, audit logs with appropriate access controls, and a tested ability to revoke access.

Open and closed models both need review

A locally hosted model can reduce some external data flows, but it still needs secure infrastructure, patching, access management, and evaluation. A hosted service can offer controls, but those controls must match the actual contracted service and configuration. Model availability alone does not establish security.

Treat retrieved content as untrusted input

A document or web page can contain instructions that attempt to redirect an agent. Keep retrieved text separate from authority to act, restrict tools, and test whether hostile content can expose information or cause an unauthorized action.

Before the first sensitive record

Verify the approved environment with a harmless test dataset. Confirm that a restricted user cannot retrieve protected material and that tool permissions match the signed pilot boundary.

LESSON 09

A citation must support the claim

Retrieval can make evidence easier to inspect. It does not guarantee a correct answer, a current source, or an appropriate interpretation.

Make verification a designed step

For regulatory research, procedures, and engineering narratives, require the source document, location within it, version or effective date, and the statement it supports. A real source linked to the wrong conclusion remains an error. Preserve the source version so a later reviewer can reconstruct the decision.

Review questions

- 01 Does the cited passage actually support this statement, including qualifications and exceptions?
- 02 Does the source apply to this jurisdiction, asset type, customer situation, and date?
- 03 Did the assistant omit conflicting evidence or confuse a proposal with an approved decision?
- 04 Are numerical results reproducible using an approved calculation and the recorded inputs?
- 05 Can the assistant say it lacks enough evidence, then route the question to the right person?

Regulatory example

A draft answer to an interrogatory should distinguish the current filing, prior decisions, internal analysis, and an analyst's interpretation. The reviewer checks the source passages and the final wording. Keep established legal and regulatory sign-off before submission. An AI-generated confidence statement is not a substitute for that review.

Your next deliverable

Build an evidence packet containing the request, source versions, retrieved passages, draft, reviewer corrections, approval, and final output. Record model and workflow versions. Do not rely on a generated account of the model's reasoning as the audit trail.

Evidence that the lesson is working

Reviewers can reproduce material calculations and verify claims without searching from scratch. Unsupported answers enter an exception route instead of receiving cosmetic source links.

LESSON 10

Evaluation must include difficult cases

Agree what acceptable performance means before the team sees the results. Measure the workflow people will actually use.

Build a representative test set

Collect typical examples and the conditions most likely to cause trouble: missing fields, poor scans, stale policies, similar asset IDs, conflicting sources, permission restrictions, unusual customer situations, and service outages. Use approved data. Keep a separate holdout set that the team does not use to tune prompts or configuration.

Dimension	Evidence to collect
Quality	Correctness, completeness, source support, and material errors by case type.
Consequence	Missed critical issues, unsafe advice, data exposure, or unauthorized actions.
Effort	Staff time for preparation, review, correction, and exceptions.
Service	Latency, failure rate, cost per completed task, and fallback success.
Adoption	Share of eligible work completed with the tool, with reasons for non-use.

Score consistently

Use a written rubric and reviewers who understand the work. Have a second reviewer check a subset and resolve disagreements. Report counts and denominators, not only percentages. Separate critical failures from minor wording corrections. A small test with no critical failures does not prove the failure rate is zero.

Illustrative acceptance criteria

For an internal document-drafting pilot: at least 90 of 100 held-out drafts meet the agreed quality rubric, median total staff effort improves by 20%, and no observed restricted-data exposure or unauthorized action occurs. These are example targets, not utility standards. Risk owners set the actual thresholds and test size.

Retest when the system changes

Repeat the relevant tests after changing the model, prompt, source library, connectors, permissions, or workflow. Track versions so the team knows which results support the current service.

Editorial evaluation framework, extending H25, printed p. 13, and H26, pp. 12–14, 24–26.

LESSON 11

Build or buy includes the years after launch

The 2026 handbook favors combining purchased foundations with utility-specific workflows. Evaluate that approach alongside other realistic options.

Option	What it can offer	What your team still owns
Buy a managed product	A packaged workflow and vendor support.	Fit, data approvals, evaluation, adoption, and supplier oversight.
Build on a platform	Reusable services with local business logic.	Integration, configuration, tests, and ownership of custom code.
Build and host	More control over deployment and implementation.	Security, operations, specialist capacity, model updates, and support.
Improve the existing process	A simpler solution using rules, search, or existing tools.	Process ownership, maintenance, and measurement.

Ask vendors to prove the difficult parts

Test on a utility-approved sample. Ask for evidence of permission enforcement, source traceability, recovery after a failed integration, usage-cost visibility, and model-change notification. Request customer references with a similar workflow and operating context, and distinguish demonstrations from supported production features.

Terms to resolve before commitment

- 01 Who owns data, configuration, derived records, prompts, evaluations, and approved outputs?
- 02 What support hours, response expectations, availability commitments, and remedies apply?
- 03 Can you export usable data and logs, disable access, and confirm deletion at exit?
- 04 What changes can the supplier make, and which require notice, evaluation, or approval?
- 05 What costs rise with volume, storage, usage, connectors, or additional departments?

Your next deliverable

Compare options over the same period and scope. Include internal effort, integration, evaluation, training, ongoing operation, and exit. Name the person who can keep each option working after its original champion leaves.

Handbook basis: H26, pp. 16–17, 20. Procurement questions are editorial guidance.

LESSON 12

Time returned and cash saved are different

Measure the net effort after review and correction. Show how freed capacity will change real work before calling it a benefit.

A consistent capacity model

Annual hours returned = monthly eligible tasks x 12 x adoption share x net minutes saved per task / 60. Net minutes saved includes the change in preparation, review, correction, and exception handling. Use actual adoption, not purchased seat count, when evaluating results.

Illustrative assumption	Value
Eligible tasks each month	500
Adoption share	60%
Gross time saved / added review per task	12 min / 4 min
Annual hours returned	$500 \times 12 \times 0.60 \times 8 / 60 = 480 \text{ h}$
Loaded labor value	$\$60/\text{h} \times 480 \text{ h} = \$28,800$
Annual running cost / one-time implementation	$\$18,000 / \$15,000$
Year-one net capacity value	$\$28,800 - \$18,000 - \$15,000 = -\$4,200$

Test the downside

At 30% adoption with the same effort and cost assumptions, annual capacity value falls to \$14,400 and year-one net capacity value becomes -\$18,600. At 90% adoption, those figures become \$43,200 and \$10,200. State the currency and whether costs include tax. These examples assume steady volume and adoption for a full year, so a phased rollout requires a monthly model.

A benefit needs a realization plan

Freed hours may reduce backlogs, improve turnaround, or avoid future overtime. Cash savings require a documented change in spending. Do not count the same hour as both labor capacity and contractor savings. Have finance confirm the treatment of costs, benefits, and any capitalization claim under the utility's own policies.

Editorial worked example. All figures are hypothetical. Context: H25, printed pp. 15, 18; H26, pp. 31–32.

LESSON 13

People need permission and time to learn

Adoption depends on whether staff can see a useful role for AI in their work and a credible role for themselves as it changes.

Create an approved place to experiment

The 2026 handbook links AI culture with access to tools and encouragement. Pair that access with clear rules about approved environments, permitted data, prohibited uses, and review obligations. Staff should know where to ask questions and report a mistake without hiding it.

What leaders can do

- 01 Explain the specific problem the pilot addresses and how decisions about roles or workloads will be made.
- 02 Reserve training and feedback time in the work plan. Do not assume people can absorb a new workflow on top of a full workload.
- 03 Involve frontline staff, supervisors, HR, and relevant employee representatives early.
- 04 Teach source verification, exception handling, and when to stop. Prompt writing alone is incomplete training.
- 05 Keep a route for staff to challenge an output or opt for the approved manual process when needed.

Preserve expertise deliberately

Ask experienced staff to explain the exceptions, assumptions, and local knowledge behind decisions. Turn that knowledge into reviewed procedures and labeled examples. Record the author, reviewer, and review date. A transcript or model memory alone is not a maintained knowledge base.

A useful learning session

Choose one typical case, one difficult case, and one case the system should refuse or escalate. Have a practitioner explain the expected result before the team uses the tool. Compare outputs and write down the corrections.

Evidence that the lesson is working

Staff can explain what the tool is allowed to do, verify its output, and report a problem. Newer colleagues learn the underlying work rather than only learning to accept suggestions.

LESSON 14

Usage needs a place in the daily workflow

Training attendance and login counts do not show that AI helps people finish the work.

Design the handoff into normal practice

Make the output available where staff already review, approve, and record the work. Define what a user should do when a suggestion is wrong, partially useful, or missing. If people copy information between several tools to use the assistant, measure that extra effort.

Signal	What to investigate
People try it once and stop	Is the task useful, the result trustworthy, and the extra effort justified?
Many outputs need rewriting	Is the requested format clear? Are sources, rules, or examples missing?
Only the champion uses it	Does it fit other roles, permissions, devices, and levels of experience?
Reviewers become a bottleneck	Is the scope too broad, or does evidence take too long to inspect?
Users bypass required review	Is the workflow enforcing the approval boundary and making the right path usable?

Measure adoption with a denominator

Track eligible tasks, tasks attempted with AI, accepted outputs, exceptions, and returns to the manual route. Ask why people choose not to use it. Compare total effort and output quality for similar tasks. Participation alone is not a performance result.

Your next deliverable

A rollout plan with a small initial cohort, a training owner, a feedback channel, support coverage, and a date to review actual usage. Fix the common blockers before inviting the next group.

Evidence that the lesson is working

The workflow owner can explain which users benefit, which cases still cause difficulty, and what will change next. Adoption grows because the tool helps, and required review remains visible.

Handbook basis: H26, pp. 17–22, 26. Adoption measures are editorial guidance.

LESSON 15

Production requires a service owner

A pilot ends with an operating decision. A production service needs ongoing support, monitoring, and controlled change.

Before production

- 01 Name the service owner, technical support owner, business escalation contact, and backup for each.
- 02 Confirm actual usage limits, budget alerts, support hours, and the manual fallback's capacity.
- 03 Set monitoring for quality, permissions, source freshness, failures, latency, cost, and user complaints.
- 04 Define how model, prompt, configuration, data, and connector changes move through testing and approval.
- 05 Rehearse disabling tool access, returning work to the manual route, and notifying affected teams.

A usable incident sequence

Pause the affected function and revoke unnecessary action rights. Route work to the approved fallback. Preserve the relevant logs and source versions under the utility's records policy. Determine what outputs or records may be affected, assign remediation, and communicate through established incident procedures. Retest the failure and related cases before an authorized restart.

Monitor outcomes, not only uptime

A service can be available while answering from stale sources or producing plausible errors. Sample completed outputs and review exceptions. Look for changes in task mix, seasonal conditions, source documents, and user behavior. Use thresholds the owner can actually act on.

Your next deliverable

A concise runbook that another colleague can use: normal operation, dashboard links, alert thresholds, contacts, pause procedure, fallback, recovery checks, and restart authority. Test it with someone who did not build the pilot.

Evidence that the lesson is working

The service can survive the absence of its original champion. Users know where to report problems, and the utility can show what changed and why.

LESSON 16

Expansion should reuse evidence and controls

A successful first workflow creates useful assets. It does not automatically validate a new function, dataset, or degree of autonomy.

Reuse what is genuinely common

Carry forward approved identity controls, source connections, data definitions, evaluation templates, training materials, support routes, and procurement terms. Reassess permissions, consequences, and acceptance criteria for every new workflow. The same model can have different risk in a different setting.

Expansion	Evidence to refresh
More users	Permissions, training, support capacity, and performance under load.
More source systems	Data contracts, source quality, identifiers, and information access.
A new department	Workflow baseline, domain review, ownership, and actual demand.
Customer-facing use	Authentication, accessibility, language support, escalation, and communication review.
More autonomous actions	Explicit authority, consequence analysis, approval controls, recovery, and independent review.

Keep a portfolio with real status

For each use case, record its owner, stage, allowed actions, evidence date, benefits, costs, unresolved issues, and next decision. Include paused and retired workflows so the organization learns why they stopped. Retiring a weak pilot can be a sound outcome.

Scale, revise, or stop

Scale when agreed quality and control criteria hold and the service has an owner and credible value. Revise when the task remains useful but a specific fix could improve results. Stop when consequences cannot be controlled, the foundation is unavailable, or the value does not justify ongoing effort.

Evidence that the lesson is working

The next team can start with reusable components and a clearer evaluation plan. Growth reduces repeated work without bypassing the decisions that make each use case safe and useful.

Handbook basis: H26, pp. 12–14, 20, 24–26. Expansion framework is editorial guidance.

UTILITY EXAMPLES

The lesson travels. The controls change.

These are illustrative pilot designs, not reported deployments or claims of achieved savings. Each needs a local domain owner and review.

Utility / workflow	A bounded pilot	Measures and boundaries
Electric: asset planning	Draft an asset condition narrative from approved records and a versioned health calculation.	Check asset joins, source support, review time, and material corrections. An engineer approves decisions. Exclude control actions.
Gas: inspection records	Extract inspection fields and flag missing documentation for staff review.	Measure extraction errors, missed exceptions, and correction time. Preserve originals. Exclude leak-response advice and pressure-control actions.
Water: maintenance history	Retrieve past work and draft a maintenance summary for a selected asset group.	Check equipment identity, dates, source support, and completeness. Exclude water-quality judgments, dosing changes, and process control.
Any utility: service support	Draft internal answers from an approved service library, with staff review before customer use.	Check policy applicability, accessibility, correct escalation, and handling time. Exclude unsupported restoration promises or account changes.

Match the data to the sector

Electric asset hierarchies, gas inspection records, and water process equipment have different identifiers, failure consequences, and operating practices. Keep domain experts involved in selecting examples, interpreting results, and defining prohibited outputs. A common software platform does not remove these differences.

Questions for the owner

What error would be most costly here? Who would detect it? What action must remain outside the pilot? What existing procedure applies when the assistant is unavailable or wrong?

Editorial adaptations. H25 and H26 primarily describe electric utility experience.

A REPORTED EXAMPLE

InnPower: a foundation, then broader use

Read the two handbook accounts as a progression in scope. Both are Senpilot-authored customer stories, not independent evaluations.

The 2025 account

The handbook describes a single asset view that connected GIS data with inspection records and spreadsheets. It then describes moving utility-specific health formulas into a shared system and using that foundation for analysis and draft asset condition reports. The transferable sequence is to connect relevant evidence, preserve the utility's calculation method, and make outputs easier for staff to use.

The 2026 account

The later handbook describes engineering, regulatory, and customer service agents. Engineering work includes asset search and project drafting. Regulatory work includes research and evidence preparation. Customer service work includes assistance with billing, requests, and periods of high demand. The broader scope makes access controls, review roles, and service ownership increasingly important.

What to learn	What to verify for your utility
Start with usable evidence	Which source systems and fields are essential to your chosen task?
Keep domain rules explicit	Which formulas and interpretations require approval and version control?
Expand across real workflows	Can each function demonstrate its own baseline, review process, and benefit?
Preserve accountability	Who owns outputs, changes, exceptions, and ongoing operation?

How to read the benefit claims

The editions report different benefit figures for different scopes and periods. This guide does not add them together or use them as a forecast. Before using a customer claim in a business case, request definitions, measurement period, baseline, implementation costs, and a separation of staff capacity from realized spending reductions.

The relevant question

Which parts of the implementation sequence are reproducible with your data, people, and approval process? That is a stronger starting point than importing another utility's savings figure.

Reported case: H25, printed pp. 14–18; H26, pp. 28–33. Interpretation is editorial.

IMPLEMENTATION PLAN

The first 90 days

This is an illustrative schedule for a bounded pilot. Procurement, data access, safety review, or integration work may require more time.

Period	Work and owner	Checkpoint
Days 1-15 Understand	Business owner observes the workflow, collects a baseline, and agrees the problem with practitioners.	Document the task, current effort and quality, affected users, and decision owner.
Days 16-30 Choose	Sponsor selects the pilot. IT, data owners, and relevant reviewers assess access and consequences.	Approve a bounded charter, cost ceiling, permitted environment, and excluded actions.
Days 31-45 Prepare	Delivery team configures sources and access. Domain reviewers define tests and acceptance criteria.	Confirm representative inputs, holdout cases, review rules, and manual fallback.
Days 46-60 Prove offline	Domain reviewers compare outputs with the baseline. Technical owners test failures and permissions.	Record quality, total effort, critical failures, cost, and required fixes.
Days 61-75 Prove in practice	Workflow owner supports a limited user cohort, with review before outputs affect real work.	Observe adoption, exceptions, support burden, and realistic benefit.
Days 76-90 Decide	Sponsor and owners review the evidence and unresolved risks.	Scale, revise with a deadline, or stop. Assign the production owner if proceeding.

If a gate is not met

Keep the next stage closed until the missing evidence or authority is in place. Narrow the scope, replace a dataset, or adjust the schedule. Do not reinterpret a fixed deadline as permission to skip a material review.

A small team's version

Use a single weekly working session and one shared decision log. Assign separate technical and business review even when a person holds multiple roles. Reserve practitioner time explicitly, and avoid starting another pilot until the current one has a supported outcome.

Editorial schedule informed by the implementation lessons in H25 and H26.

WORKSHEET 01

Pilot charter

Complete this with the business owner and the people who perform the work. Use additional space as needed. Transfer the agreed answers into the companion presentation.

Field	Your answer
Workflow and intended users	Task, department, trigger, and eligible case types:
Problem and baseline	Volume, total staff effort, elapsed time, quality, and measurement period:
Proposed assistance	Exactly what the system may produce or do:
Boundaries	Excluded data, decisions, actions, and case types:
Data and approvals	Sources, owners, permissions, and approved environment:
Acceptance and stop criteria	Required quality, effort, controls, and immediate pause triggers:
Ownership	Sponsor, workflow owner, technical owner, reviewers, and support contact:
Decision	Budget ceiling, deadline, decision owner, and next gate:

Approval record

Decision: _____ Date: _____

Conditions and unresolved issues: _____

Accountable owner: _____

WORKSHEET 02

Use-case selection and data readiness

Check hard gates first. Then compare eligible options using the same evidence and scoring scale.

Gate check

Confirm authorized data use, an accountable owner, an acceptable consequence boundary, and a feasible review/fallback process. Mark each gate yes, no, or unresolved. A no or unresolved answer requires action before a pilot begins.

Eligible candidate	Value	Data	Verification	Delivery	Total /20
[Workflow A]	__/5	__/5	__/5	__/5	--
[Workflow B]	__/5	__/5	__/5	__/5	--
[Workflow C]	__/5	__/5	__/5	__/5	--

Scoring anchors

Use 1 for weak evidence or major obstacles, 3 for credible potential with resolvable gaps, and 5 for strong evidence and readiness. Value concerns frequency and benefit. Data concerns access and fitness. Verification concerns the ability to judge results. Delivery concerns integration and team capacity. Higher scores are better. A total helps discussion but does not override a failed gate.

Required source or field	Owner / authority	Gap and action
[Dataset / required fields]	[Owner / access approval]	[Missingness / refresh / ID issue]
[Dataset / required fields]	[Owner / access approval]	[Action / owner / due date]
[Dataset / required fields]	[Owner / access approval]	[Acceptance evidence]

Selection record

Chosen workflow: _____
Why this one, given the tradeoffs: _____
Unresolved gate / owner / due date: _____

WORKSHEET 03

Evaluation and decision record

Set thresholds before testing. Record the results that would support stopping as clearly as those that would support expansion.

Measure	Baseline / target	Observed result
Quality on holdout cases	[Rubric / count / threshold]	[Passes / total / error types]
Critical failures	[Defined severity / stop rule]	[Count / cases / remediation]
Total staff effort	[Minutes / case distribution]	[Preparation + review + correction]
Adoption	[Eligible tasks / expected share]	[Used / eligible / reasons]
Operating cost	[Budget / volume assumptions]	[Usage + support + other costs]
Fallback and access	[Required scenarios]	[Evidence / failures / owner]

A complete evidence packet

Attach the dated baseline, dataset description, test cases and rubric, system versions, source and permission checks, reviewer results, cost model, user feedback, unresolved issues, and proposed service runbook. Use a link to a controlled record rather than copying sensitive material into the presentation.

Decision	When it fits
Proceed	Acceptance criteria hold, consequences are controlled, and ownership and value are credible.
Revise	A specific, bounded change could close a material gap. Assign an owner and retest date.
Stop	The task lacks value, data access is unavailable, or the consequences cannot be acceptably controlled.

Decision log

Decision and scope: _____

Conditions / owner / due date: _____

Approver / date / next review: _____

Reusable editorial worksheet. A passed pilot supports only the tested scope and conditions.

REFERENCE

A short working glossary

Use these terms to make decisions more precise. Agree local definitions when the same word means different things across teams.

Term	Meaning in this guide
Agent	An AI system that can select and use tools to carry out steps. Its permitted actions require explicit limits.
Baseline	Measured performance of the current workflow, including effort, quality, and relevant conditions.
Capacity value	A monetary estimate of staff time available for other work. It is not automatically a cash saving.
Data lineage	The trace from an output through transformations to original records and versions.
Drift	A change in data, usage, or performance that can make earlier evaluation less representative.
Evaluation / holdout	A structured test of performance. A holdout contains cases kept separate from development and tuning.
Grounding / retrieval	Providing relevant source material to support an answer. Source verification is still necessary.
Hallucination	An output that is false, invented, or unsupported, even when expressed confidently.
Human approval	An accountable person's review before an output or action takes effect, with authority to reject it.
Prompt injection	Untrusted content that attempts to redirect a system's behavior or misuse its access and tools.
System of record	The agreed authoritative source for a particular record or field.
OT / SCADA	Operational technology / supervisory control and data acquisition. These connect to physical operations and need their own controls.
GIS / EAM / OMS	Geographic information system / enterprise asset management / outage management system.
AMI / ERP	Advanced metering infrastructure / enterprise resource planning.

Editorial definitions for this resource. Terminology may vary between utilities and products.

SOURCES & USE

Where the lessons come from

September 2026 edition. This is an educational resource from aiforutilities.org, sponsored by Senpilot.

[H25] Senpilot, AI x Utility Handbook 2025

Source file: Senpilot – AI for Utilities Handbook – 2025.pdf. The 21-page PDF contains a cover, contents page, and printed pages 1–19. References in this guide use the printed page numbers. Main material: AI Cake (p. 3), data and implementation (pp. 6–10), auditability (p. 13), and InnPower (pp. 14–18).

[H26] Senpilot, AI x Utilities: Practical lessons from utilities around the world

Source file: Senpilot Handbook AI x Utilities 2026.pdf, 34 pages. Main material: anonymized failures (pp. 12–14), implementation lessons (pp. 16–24), integration models (pp. 25–26), and InnPower (pp. 28–33). Page references match the PDF page count.

[N1] NIST, Generative Artificial Intelligence Profile

NIST AI 600–1, July 2024, is a voluntary companion to the AI Risk Management Framework. It discusses risks such as false outputs, privacy, information security, and human–AI interaction, and offers actions for governance, provenance, testing, and incident disclosure. It is a reference for risk discussions, not a certification or a utility-specific legal requirement. [Read the NIST profile.](#)

[E1] U.S. Department of Energy, AI risk assessment

DOE's April 2024 initial assessment considers benefits and risks of AI for critical energy infrastructure. It provides context for separating administrative assistance from applications with operational consequences. It does not validate any particular product or workflow. [Read DOE's assessment announcement.](#)

Editorial approach

The worksheets, thresholds, schedule, worked financial example, and gas/water scenarios are new editorial guidance. We have not independently audited the handbook cases. We do not repeat their adoption forecasts, broad security assurances, universal deployment–cost claims, or benefit percentages as general sector facts. Local obligations and engineering decisions remain with the utility's authorized professionals.

Continue your work

Use the companion AI journey PowerPoint and the free tools at aiforutilities.org/toolkit. Senpilot sponsors this educational initiative. Learn about the sponsor at senpilot.com.