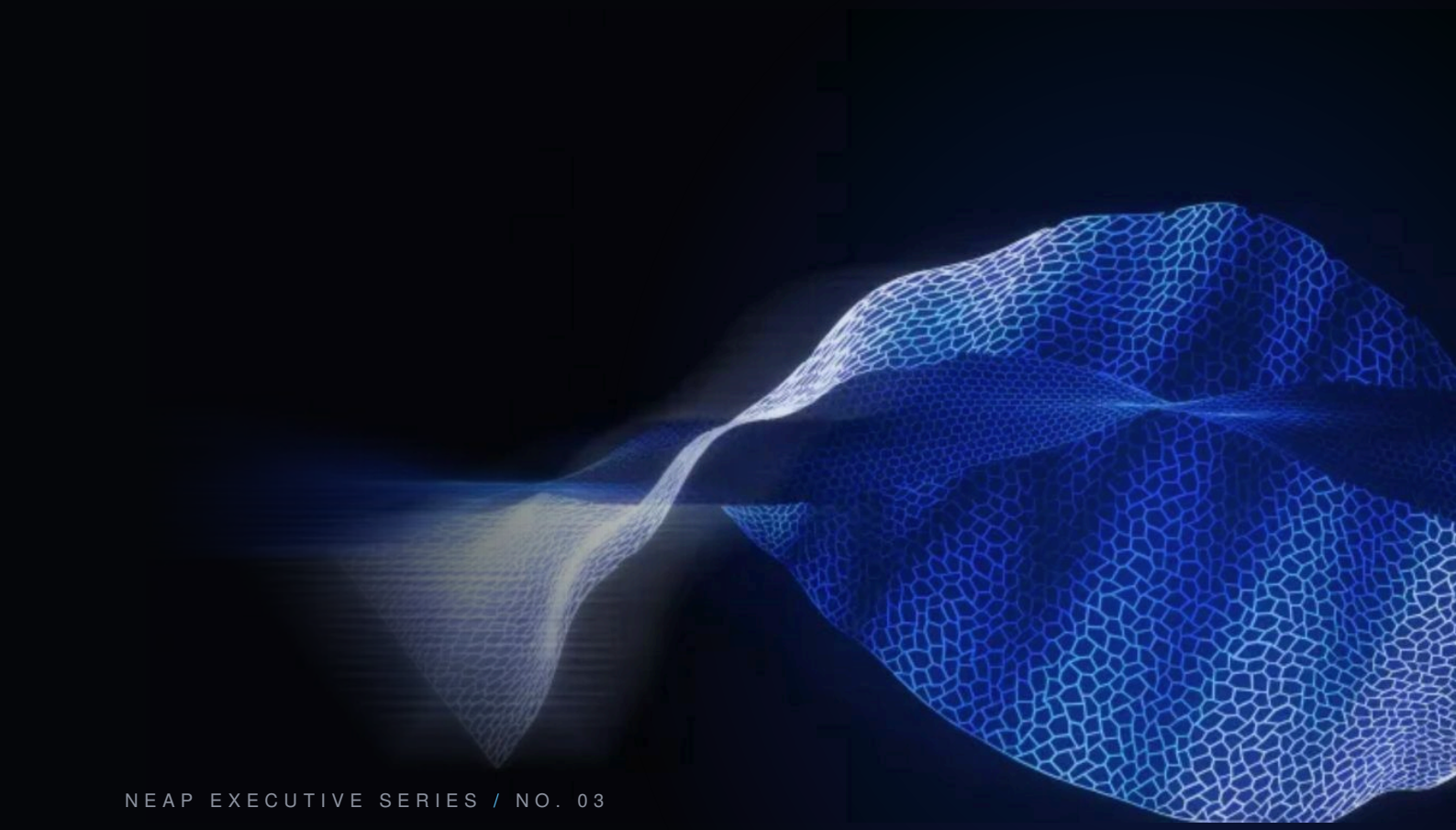
HTech / National Enterprise AI Program™



NEAP EXECUTIVE SERIES / NO. 03

From pilot to production

How to choose the right AI cases, build a business case that holds, run a pilot designed to go live, and scale what works.

ABOUT THIS GUIDE

Most AI pilots succeed. Most never reach the business. This guide is about the second part.

A pilot that works in a demonstration proves very little. What matters is whether the company chose a case worth solving, whether the pilot was built with production in mind, and whether anyone measured what changed. This guide gives a step-by-step method for each, drawn from global research and from HTech's delivery work.

01	Why pilots stall	The anatomy of pilot purgatory
02	The Operational Matrix	Six steps from a long list to one commitment
03	A business case that holds	Value, full cost, payback
04	Build, buy or partner	Choosing how to deliver
05	A pilot designed to go live	The pilot charter
06	Crossing into production	The checklist and the operating model
07	Measure, then scale	From one case to a portfolio
08	Your 90-day plan	With a commitment card to copy

READ THIS IF

You have AI ideas or pilots in motion and want them to show up in the accounts.

TIME TO READ

About 40 minutes. Chapters 02 and 05 are built to be used in a workshop.

BEFORE YOU START

No. 02, the Readiness Playbook, tells you whether you are ready for a pilot at all.

CHAPTER 01

Why pilots stall

"Pilot purgatory" is not bad luck. It is the predictable result of pilots designed to impress rather than to be used.

The numbers are stark. MIT's 2025 study of enterprise generative AI found that only about 5% of custom enterprise AI tools reached production, and that large enterprises took nine months or more to move from pilot to full implementation, against roughly 90 days for mid-market companies.[1] Gartner reported that at least half of generative AI projects were abandoned after the proof of concept by the end of 2025, because of poor data quality, inadequate risk controls, escalating costs or unclear business value.[2]

Six symptoms of a pilot that will not ship

SYMPTOM	WHAT IT USUALLY MEANS
It runs on a sample file, not live data	Integration, data quality and access have not been solved, only postponed.
The users are the project team	Nobody has tested whether the people who do the work will use it, or how their work changes.
Success is "the demo went well"	No baseline, no target, so no way to decide to invest or stop.
Nobody knows the cost per case	The economics at full volume are unknown, and may not work.
Security will "review it later"	A late review can block or reshape the whole design.
There is no named business owner	Nobody has a reason to fight for it after the pilot ends.

There is progress: in McKinsey's 2026 survey, 44% of respondents said AI was scaling across their enterprise, up from 38% a year earlier.[3]

THE CORE IDEA OF THIS GUIDE

Choose cases with the business, not for it. Design the pilot as the first release of a production system. Decide in advance what result would justify going live, and what would justify stopping.

CHAPTER 02

The Operational Matrix

Six steps take a leadership team from a long list of ideas to one committed case with an owner, a metric and a date. It is the method used in the NEAP Foundation Workshop.

STEP 1	STEP 2	STEP 3	STEP 4	STEP 5	STEP 6
List	**Filter**	**Score**	**Map**	**Break down**	**Commit**
Every place AI might help.	Remove what cannot work now.	Rate value and feasibility.	Place cases on the matrix.	Split the winner into steps.	Owner, metric, date.

Step 1 · List

Aim for 15 to 30 candidates drawn from real operations, not from vendor brochures. Ask each department head four questions: where do people **read, re-type or check documents**; where do **decisions wait** for someone to gather information; where do **customers or suppliers wait** for an answer; and where do **mistakes cost money**? Write each case as a sentence: "Use AI to [task] for [team] so that [result]."

Step 2 · Filter

Remove any case that fails one of four hard tests. Filtered cases are not rejected forever; they go on a list to revisit once the blocker is fixed.

- **Owner:** is there a business owner willing to be measured on the result?
- **Data:** does the data or document set exist, and can it be reached?
- **Rules:** is it legally and ethically acceptable, including under data protection law?
- **Volume:** does it happen often enough, or matter enough, to be worth automating?

Step 3 · Score

Score each remaining case from 1 to 5 on the criteria below. Keep the weights simple and agree them before scoring, so the result is not tuned to a favourite.

VALUE (50%)	ASK	FEASIBILITY (50%)	ASK
Financial impact	Hours, outsourcing, errors or revenue affected per year?	Data readiness	Is the data there, clean enough, accessible?
Strategic fit	Does it serve a top company priority?	Technical fit	Is this a proven use of AI, or research?
Speed to value	Will benefits show within a quarter of launch?	Integration	How many systems must it connect to?
Reach	How many people or customers benefit?	Risk	What happens when it is wrong? (5 = low risk)

Step 4 · Map

Plot each case by its value score and its feasibility score. The quadrant tells you what to do with it.

BUSINESS VALUE →

Strategic bets

High value, harder to deliver. Plan them, fix their blockers, start once a quick win has built confidence and capability.

Start here

High value and feasible now. Your first pilot should come from this quadrant, ideally the case with the clearest owner.

Avoid for now

Low value and hard to deliver. Park them, however exciting the technology looks.

Easy add-ons

Feasible but modest. Good for building skills, or to bundle into a larger case. Do not let them consume the budget.

FEASIBILITY →

Step 5 · Break down

Take the chosen case and write out today's process step by step. For each step, decide whether AI should do it, assist with it, or leave it to people, and where a person must approve. This is where "redesign the workflow", the habit that most separates high performers, becomes concrete.[3]

STEP TODAY	OWNER AFTER	NOTE
Receive claim by e-mail	AI	Classify, extract fields, check completeness.
Request missing papers	AI drafts	Officer approves the message before it is sent.
Check against policy	AI assists	Shows the relevant clauses with references; officer decides.
Approve or refer	Person	Unchanged; AI prepares a one-page summary.
Notify the customer	AI drafts	Template-based, in Arabic, French or English.

Illustrative breakdown for a claims process at an insurer.

Step 6 · Commit

End with a commitment card that the business owner signs. If no one will sign it, the case is not ready, whatever its score. A template is on the last page of this guide.

> **Workshop tip.** Run steps 1 to 4 with the leadership team in one session of about two hours. Steps 5 and 6 work best with the business owner and the people who do the work today.

CHAPTER 03

A business case that holds

A credible AI business case counts the full cost, converts saved time into money only when the time is actually redeployed, and states what would make it fail.

Value: be specific about the mechanism

Name the mechanism of value: fewer outsourced hours, fewer errors and write-offs, faster cash collection, higher conversion, avoided hires as volume grows. MIT found the clearest returns in back-office cases where outside spending can be cut, with companies reporting $2 million to $10 million a year from reduced outsourcing and 30% lower agency costs.[1] Productivity alone is not value: 80% of respondents in McKinsey's 2026 survey reported better individual productivity, but only 37% saw an effect on EBIT.[3]

Cost: count all of it

COST LINE	WHAT TO INCLUDE	OFTEN MISSED?
Build	Design, development, configuration, testing, licences for the first year.	Rarely
Integration	Connecting to the ERP, CRM, document stores, identity and logging.	Often
Data preparation	Digitising, cleaning and labelling the documents and records the case needs.	Very often
Running	Model usage (tokens), hosting, monitoring, vendor fees, at full volume.	Often
People and change	Training, process redesign, time of the business owner and reviewers.	Very often
Control	Security testing, legal and data protection review, periodic audits.	Often

Running costs deserve attention. Unit prices are falling fast, but usage grows with them, and about one in five organisations in McKinsey's survey says AI operating costs already constrain their use.[3,4] Estimate a **cost per completed case**, including retries and human review, and compare it with today's cost per case.

WORKED EXAMPLE: CLAIMS INTAKE AT A MID-SIZED INSURER

Volume	24,000 claims a year
Time saved per claim (measured in pilot)	12 minutes, of 30 today
Loaded cost per hour	$11
Share of time redeployed	60% (staff moved to other work; vacancies not refilled)
Gross annual value	24,000 × 12 ÷ 60 × $11 × 0.6 = **$31,680**
Annual running cost	$7,200 (model usage, hosting, monitoring)
One-off cost	$28,000 (build, integration, data preparation, training)
Payback	$28,000 ÷ ($31,680 − $7,200) ≈ **14 months**, before faster settlement and fewer errors

Illustrative figures. Replace each line with your own measured values.

Test the case before you trust it

Most AI business cases hinge on one or two assumptions. Show the result under less favourable values. In the example above, the share of saved time that is actually redeployed matters most:

SHARE OF SAVED TIME REDEPLOYED	NET ANNUAL VALUE	PAYBACK
30%	$15,840 − $7,200 = $8,640	about 39 months
60% (base case)	$31,680 − $7,200 = $24,480	about 14 months
80%	$42,240 − $7,200 = $35,040	about 10 months

RED FLAGS IN AN AI BUSINESS CASE

- Time saved is counted as money with no plan for how it will be used.
- Running costs are missing, or estimated at pilot volume rather than full volume.
- No baseline: today's time, cost and error rate were never measured.
- Accuracy is assumed from a vendor demo, not measured on your own documents.
- Integration and data preparation appear as a single small line, or not at all.
- Every benefit lands in year one, with no allowance for adoption taking time.

CHAPTER 04

Build, buy or partner

How you source an AI case affects its chance of reaching production as much as what the case is.

MIT's research found that AI tools bought from specialised vendors or built with external partners reached deployment about 67% of the time, against roughly 33% for purely internal builds.[1] At the same time, building has become cheaper: 32% of McKinsey's 2026 respondents said they had decided against buying at least one software product because they could build it in-house with AI coding tools.[3] The right answer depends on the case.

ROUTE	BEST WHEN	WATCH FOR
Buy	The need is common across companies (meeting notes, translation, general assistants) and a mature product fits with little change.	Data location and use for training; Arabic quality; per-user costs at scale; lock-in.
Partner	The case is specific to your process and systems, and you lack experience integrating, evaluating and running AI in production.	Make knowledge transfer and documentation part of the contract, so you can own it later.
Build	The case is core to how you compete, you have engineering capacity, and you will run and improve it for years.	Underestimating evaluation, security, monitoring and support; key-person dependency.

Questions that decide it

- Is this case a **differentiator** or a **utility**? Utilities should usually be bought.
- How much of the effort is **integration with our systems** rather than AI itself? The more there is, the more a partner who knows integration helps.
- Who will **run and improve** it in two years? If the answer is "us", insist on a path to ownership from the start.
- Where must the **data stay**? Requirements to keep data in Lebanon or on your own servers narrow the options early.

HOW NEAP APPROACHES IT

NEAP pairs each implementation with NEAP Fellows, graduates from partner universities who work inside the delivery team. The company gets an experienced partner for the first projects and, at the same time, people who learn its systems and can stay on to run them.

Choosing a vendor: a short scorecard

CRITERION	WHAT GOOD LOOKS LIKE
Proof on your data	Willing to test on a sample of your real documents before contract, with agreed pass marks.
Production references	Clients running the solution live, not in pilots, ideally in your region and languages.
Data terms	Clear location, retention and no training on your data without consent, in writing.
Transparent unit cost	A price per case or per user at your expected volume, including model usage.
Exit and ownership	Your data, prompts, configurations and evaluation sets can be exported.

Clauses worth insisting on

- **Acceptance on quality:** payment milestones tied to measured accuracy on your evaluation set, not to delivery of software.
- **Data use:** no use of your data, prompts or outputs to train third-party models without written consent; stated data location and deletion on exit.
- **Cost ceilings:** a cap or alert threshold on usage-based charges, with monthly reporting.
- **Knowledge transfer:** documentation, runbooks and paired working so your team can operate the system.
- **Change of model:** notice before the underlying model is changed, and a re-run of the evaluation set when it is.

CHAPTER 05

A pilot designed to go live

Treat the pilot as the first release of a production system: real users, real data, a measured baseline, a target and a decision date.

The pilot charter

Agree these ten items in writing before any build starts. It takes an afternoon and prevents most of the failures in Chapter 01.

ITEM	WHAT TO WRITE DOWN
1 · Business owner	One named person accountable for the result, not the project manager.
2 · Scope	Which process steps, which team, which document types or customers. What is out of scope.
3 · Baseline	Today's volume, time per case, cost per case and error rate, measured before the build.
4 · Targets	Quality (accuracy on the evaluation set), adoption (share of cases handled) and value (time or cost saved).
5 · Evaluation set	100 to 300 real past cases with known right answers, including hard and messy ones, in every language used.
6 · Human review	Exactly where people approve, and what happens when the AI is unsure or fails.
7 · Data and security	What data is used, where it is processed, who can see it; sign-off from security and data protection before go-live.
8 · Integration path	How the pilot will connect to production systems, even if it starts with a simpler link.
9 · Cost per case	Measured during the pilot, projected to full volume.
10 · Decision gate	A fixed date, and the results that mean go live, extend, or stop.

A typical 10-week pilot

WEEKS 1–2
Prepare
Charter signed. Baseline measured. Evaluation set built. Data access agreed.

WEEKS 3–5
Build
First version tested against the evaluation set until it meets the quality bar.

WEEKS 6–9
Run
Real users on real cases, with review. Quality, adoption and cost tracked weekly.

WEEK 10
Decide
Results against targets. Go live, extend with changes, or stop and record why.

> **Stopping is a success too.** A pilot that shows a case does not pay, at a cost of ten weeks, has saved the company a much larger mistake. Record what was learned and move to the next case on the matrix.

CHAPTER 06

Crossing into production

Going live is when the real work begins: integration, security, monitoring, support and the change in how people work.

The production checklist

INTEGRATION AND OPERATIONS

- [ ] Connected to the source and target systems through supported interfaces, not manual exports.
- [ ] Handles volume peaks, outages and retries; a fallback exists if the AI service is unavailable.
- [ ] A named support owner, an escalation path and a documented runbook.

QUALITY AND MONITORING

- [ ] The evaluation set runs automatically before every change to prompts, models or data.
- [ ] Live quality is sampled and reviewed weekly at first, then monthly; drift triggers a review.
- [ ] Cost per case and total spend are tracked against budget, with alerts.

SECURITY AND COMPLIANCE

- [ ] Access follows the user's existing permissions; every action is logged.
- [ ] Tested against prompt injection and data leakage, the top risks in OWASP's list for AI applications. [5]
- [ ] Data protection obligations reviewed, including Law 81/2018 for personal data in Lebanon.

PEOPLE AND PROCESS

- [ ] Users trained on the new process, including when not to trust the output.
- [ ] Roles and targets updated to reflect the new way of working.
- [ ] The freed time has a planned use, agreed with the business owner.

Who runs it afterwards

Every production AI system needs three owners: a **business owner** accountable for its results, a **technical owner** who keeps it running and improves it, and a **risk owner** who checks it stays within the rules. In smaller companies one person may hold two of these roles, but all three must be named. BCG's 2026 research found that leading companies invest in governance and the workforce alongside the technology, and that seven in ten future-built and scaling companies are already retraining staff.[6]

CHAPTER 07

Measure, then scale

The companies that earn most from AI measure it in money, reinvest what they learn, and turn single cases into a portfolio built on shared foundations.

BCG reports that the most AI-mature companies measure AI's impact directly in the profit and loss and concentrate investment in a single coherent programme; they deliver 2.3 times the shareholder return of laggards.[6] McKinsey's high performers are twice as likely to have defined processes for measuring AI's impact.[3]

A simple measurement tree

LEVEL	EXAMPLE METRICS	REVIEWED
Financial	Cost per case, outsourced spend, revenue per customer, days sales outstanding.	Quarterly, by finance
Operational	Cases handled, cycle time, backlog, error and rework rates.	Monthly, by the owner
Adoption	Share of eligible cases going through the AI, overrides by users, user satisfaction.	Monthly
Quality and risk	Accuracy on the evaluation set, live sample accuracy, incidents, complaints.	Weekly at first

Four ways to scale

WIDER

More teams, same case

Roll the proven case to other branches, units or countries. Usually the fastest return.

DEEPER

More of the process

Move up the autonomy ladder within the same process once quality is proven over months.

ADJACENT

Similar cases

Reuse the components: the same document extraction can serve invoices, claims and contracts.

PLATFORM

Shared foundations

Common identity, logging, evaluation and cost tracking, so each new case starts further ahead.

THE SCALING TEST

Your third AI case should take noticeably less time and money than your first. If it does not, you are running projects, not building capability.

CHAPTER 08

Your 90-day plan

One quarter is enough to go from a list of ideas to a decision on your first production case.

DAYS 1–20

Choose

Run the Operational Matrix with leadership. Pick one case from the "start here" quadrant. Sign the commitment card.

DAYS 21–35

Prepare

Write the pilot charter. Measure the baseline. Build the evaluation set. Decide build, buy or partner.

DAYS 36–90

Pilot and decide

Build, run with real users, track quality, adoption and cost, and take the go or stop decision.

COMMITMENT CARD

The case, in one sentence	
Business owner	
Metric that will move	
Baseline today	
Target after the pilot	
What we will do with the freed time	
Budget approved	
Decision date	
Signed	

Continue with the series. No. 04 sets out the governance, data and regulatory groundwork to have in place before your first case goes live.

SOURCES AND NOTES

Where the numbers come from

1. Challapally, A., Pease, C., Raskar, R. and Chari, P., *The GenAI Divide: State of AI in Business 2025*, MIT Project NANDA, July 2025; as reported by Fortune (18 August 2025), Campus Technology (26 August 2025) and others.
2. Gartner, *Why 50% of GenAI Projects Fail — And How to Beat the Odds*, gartner.com/en/articles/genai-project-failure.
3. McKinsey & Company, *The state of AI in 2026: On the road to ROI*, 25 August 2026 (1,719 participants, 97 countries).
4. Stanford Institute for Human-Centered AI, *AI Index Report 2025* and *2026*; Epoch AI, *The plunging price of thought*, 22 September 2026.
5. OWASP Gen AI Security Project, *OWASP Top 10 for LLM Applications 2025* (LLM01 Prompt Injection; LLM02 Sensitive Information Disclosure).
6. Boston Consulting Group, *AI Is Starting to Pay Off* (Applied AI Index 2026), 30 September 2026; and *The Formula for Agentic AI Value*, 2026.

This guide is published by HTech S.A.R.L. as part of the National Enterprise AI Program (NEAP), an HTech program. It is general information for business leaders, not legal, financial or regulatory advice; check how any law or standard applies to your company with qualified counsel. Figures are quoted from the sources listed, as published by them, and were checked in November 2026. You are welcome to share this guide unchanged and free of charge. © 2026 HTech S.A.R.L.

YOUR NEXT STEP

Bring us your list of cases. Leave with the one to start.

The Operational Matrix in this guide is the method HTech uses with every client and every NEAP cohort. In a free 30-minute debrief we can run your top candidates through it with you and tell you honestly whether a roadmap is worth it yet.

FREE · 3 MINUTES

Take the Readiness Check

Ten questions across the five pillars of the NEAP Readiness Index. You get your level, your weakest pillar and the one next step that fits.

FREE · 30 MINUTES

Book a debrief with HTech

Walk through your result with an HTech consultant and leave with a short list of where AI can create value in your company first.

Book a debrief

NEXT NEAP COHORT

Join the Foundation Workshop

One day for your leadership team. You leave with a full readiness profile and a commitment card for your first AI case.

joinneap.com

HTech S.A.R.L. · Berytech, Mathaf, Beirut, Lebanon
info@htechco.com · +961 81 733 831 · htechco.com · joinneap.com

HTech / National Enterprise AI Program™