DROPSOLID
ARTIFICIAL INTELLIGENCE

GUIDE 02

SOVEREIGN AI · IMPLEMENTATION

The 8-week AI launch playbook

You can plan and run an eight-week sovereign AI deployment. From preparation to production, week by week.

Prepare	Configure	Prove	Launch
Weeks 1–2	Weeks 3–4	Weeks 5–6	Weeks 7–8

GREEN LIGHT

Week 0
Four decisions to lock before Week 1

An eight-week window only works when the key decisions are already made. Week 0 is where you decide what success means.

Scope

- [ ] One use case
- [ ] One audience
- [ ] One launch language

Success criteria

- [ ] Contact reduction
- [ ] Answer accuracy
- [ ] Decline rate
- [ ] Cost per interaction

Ownership

- [ ] Editorial owner named
- [ ] Platform lead named
- [ ] Compliance approver named

Source readiness

- [ ] Sources to index are listed
- [ ] Team can access all of them
- [ ] No stale or contradictory pages

If any of these are unclear, don't start.

Week one is an expensive place to make decisions that belong to week zero. Most of the failure modes later in this guide can be traced back to one of the four points above being left open.

THE ROADMAP

Eight weeks, four stages

Every stage ends in something you can show to someone who was not in the room. The pages that follow take one stage each, with a week on either side of the spread.

Prepare 1–2

Week 1 Scope the question, set the baseline

Week 2 Index published content

Ends with a scoped use case, a real question set, a baseline and an indexed corpus.

Configure 3–4

Week 3 Retrieval, gateway and citations

Week 4 Calibrate confidence and refusal

Ends with a working answer path, citations on every answer and a documented threshold.

Prove 5–6

Week 5 Internal user pilot

Week 6 Correct at source, then measure

Ends with real internal use, scored answer quality and corrections made at source.

Launch 7–8

Week 7 Limited external exposure

Week 8 Widen, monitor, review

Ends with full exposure, monitoring in place, a trace export and a retrospective.

EVIDENCE

EU careers | European Personnel Selection Office

<8 wk
discovery to full public rollout

24
EU languages, source-grounded

70%
fewer questions to the support team

The timeline was short because the sovereign architecture already existed.

STAGE 1 · WEEKS 1–2

Prepare

Decide what the system has to answer, and record what today looks like. None of this work needs a model yet.

Week 1 Scope the question, set the baseline

Establish which questions the system has to answer, and what the situation looks like before you change anything.

- Write the use case in one sentence: who asks what, on which surface.
- Pull 50–100 real questions from support tickets, contact-centre logs or site search.
- Measure the baseline now: contact volume, handling time, or time to answer.
- Involve compliance and information governance in week one rather than week six.

OUTPUT A written question set, a success criterion with a number, and a recorded baseline.

☐ WEEK COMPLETE WHEN — you can state the use case and success metric in a sentence each, and the baseline is recorded in writing.

Week 2 Index published content, and nothing else

Build a corpus the organisation stands behind, with a named owner for every source in it.

- List the source systems and confirm which content is published and approved.
- Index it, then record what you deliberately excluded, and why.
- Fix contradictions and retire stale pages at source rather than in the AI layer.

WATCH OUT Indexing everything is quicker than agreeing what is accurate, and it is usually where the first confidently wrong answer comes from.

OUTPUT An indexed corpus, an exclusion list, and an owner per source.

☐ WEEK COMPLETE WHEN — every indexed source has a publication status and a named owner who can correct it.

STAGE 2 · WEEKS 3–4

Configure

Build the answer path, then decide how the system should behave when it does not know the answer.

Week 3 Retrieval, gateway and citations

A working answer path with the sovereignty posture your procurement requires.

- Choose embedding and language models through the model gateway, based on your requirements rather than on a brand.
- Decide where the vector store runs now, EU-hosted or self-hosted. Designing for the switch up front turns a later migration into a configuration change.
- Make citations part of the answer template rather than an optional field.
- Confirm which trace fields are recorded per interaction: question, retrieved passages, model, response and citations.

YOU NEED — Architecture sign-off, hosting decisions, the DPA and sub-processor position.

OUTPUT — A working answer path with citations, and a documented sovereignty posture.

☐ WEEK COMPLETE WHEN — every answer returns a clickable source, and you can state where inference, embeddings and vectors run.

Week 4 Calibrate confidence and refusal

Make the behaviour predictable when the corpus does not hold the answer.

- Run the week-one question set and review the answers with the content owner.
- Set the confidence threshold, then write the refusal wording and the escalation route.
- Tune retrieval before you touch prompts. Most weak answers turn out to be retrieval problems.

Refusal to answer is an answer.

For the architectural rationale behind traceability and steerability, see Whitepaper 03.

OUTPUT — A reviewed answer set, a documented threshold and a defined hand-off to a person.

☐ WEEK COMPLETE WHEN — questions below the threshold are declined and routed to a person instead of answered plausibly.

STAGE 3 · WEEKS 5–6

Prove

Roll out by exposure rather than by pilot: real users and real data, with a small audience that widens as confidence grows.

Week 5 Internal user pilot

Put the system in front of the team that answers these questions today. Real interactions, limited audience, with human fallback always visible.

- Score answers as they arrive: correct, incomplete, wrong or correctly refused.
- Capture feedback and production usage data.
- Log every gap as either a content or retrieval issue and route it to the right owner.
- Improve source content in parallel as issues surface.

Pilot is not a demo.

A demo is two colleagues asking questions they already know the answer to. A pilot is real production traffic with a limited audience and human fallback.

OUTPUT Production usage data, captured feedback and a prioritised set of source-content improvements.

☐ WEEK COMPLETE WHEN — human fallback works, and surfaced content gaps are being corrected at source.

Week 6 Correct at source, then measure

Turn what week five surfaced into content changes, and show the improvement with a number.

- Fix the content behind the wrong answers, in the source system.
- Re-run the question set and compare the result against week four.
- Review a handful of interactions end to end with compliance in the room.

OUTPUT A measured answer-quality figure and a corrected corpus.

☐ WEEK COMPLETE WHEN — the gaps that remain are content decisions rather than system defects.

STAGE 4 · WEEKS 7–8

Launch

Widen the audience in steps you can reverse, and hand the system over to the people who will run it.

Week 7 Limited external exposure

One page, one journey or one language: bounded, monitored and reversible.

- Turn on the public-endpoint protections before the audience arrives: rate limiting and abuse controls.
- Retain no conversation state unless the use case genuinely needs it.
- Publish what the system does and does not do, in the user's language.

OUTPUT — Live external traffic under rate limiting, and published usage guidance.

☐ **WEEK COMPLETE WHEN** — external traffic is running, monitored, and escalating cleanly to human support.

Week 8 Widen, monitor, review

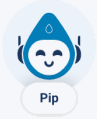
Open up to the full audience, with named operational owners and a result measured against the week-one baseline.

- Hand over operations: who watches the dashboard, who answers escalations, who owns content corrections.
- Export a trace for a sample of interactions and file it with information governance.
- Run the retrospective against the baseline rather than against expectations.

OUTPUT — A retrospective against the baseline, and a first draft of the next 90-day roadmap.

☐ **WEEK COMPLETE WHEN** — the service keeps running without the project team in the room.

WHAT GOOD LOOKS LIKE



EPSO went from discovery to full public rollout in under eight weeks across 24 EU languages, and reports a 70% reduction in incoming questions reaching its selection support team. **Pidpa** moved from internal evaluation to contact-centre use and then to external citizens, and retired its contact-centre scripts. Both deployments are documented end to end in Whitepaper 02.

WATCH OUT

What derails the plan

Six recurring problems derail otherwise workable timelines

Content nobody owns

Stale or contradictory pages produce confident, wrong answers, and no amount of prompt work will correct them.

Fix · an owner per source, in week 2.

Compliance arriving in week 6

Hosting, DPA and audit questions raised late tend to require a rebuild rather than a review.

Fix · compliance in the week-one room.

KPIs you cannot measure

Better service is hard to demonstrate in week eight. Deflection, handling time and time to answer are not.

Fix · one countable metric per goal.

No baseline

Measuring only after launch leaves the team arguing about whether anything actually improved.

Fix · measure in week 1, before anything changes.

Over-scoping

Every extra audience, language or source system added mid-flight moves week eight further away.

Fix · park additions for the next 90 days.

A demo instead of a pilot

Controlled demos hide the questions real users ask, and postpone every problem to after launch.

Fix · real traffic with a limited audience, from week 5.

NEXT STEPS

Where to go from here

NEXT MOVE

Check out our practical guides for implementation, by use case.

Guide 1 · Trusted Answers implementation
What to get right for external answers

Guide 2 · Knowledge Navigator implementation
What to get right for internal knowledge

UNDERSTAND MORE

For the thinking behind this playbook, see our whitepaper series.

Whitepaper 01 · A CIO's procurement guide for the EU AI Act

Whitepaper 02 · Sovereign AI in production

Whitepaper 03 · Glass box vs black box

Continue the conversation

Talk to Dropsolid about your use case, architecture and route to production.

Talk to Dropsolid AI

Whitepaper 02 · Sovereign AI in production →

DROPSOLID
ARTIFICIAL INTELLIGENCE

Read Whitepaper 02, Sovereign AI in production, for the deployment patterns and evidence behind this playbook.