ACCUFIND

The 5 decisions

Taking AI into production. A field card for anyone running or approving a government AI proof of concept. Free to reuse.

01 What decision changes, and who owns it

BEFORE THE POC STARTS

Name the decision or process the system will change, and the person accountable for that decision today. If no such person exists, there is nothing to hand the system to.

HOW TO CHECK

- Name the decision, determination, assessment or task the system changes. One sentence, in the language the agency already uses for it.
- Name the person accountable for that decision today, by role. If the role does not exist, stop here.
- Confirm that person controls, or can reach, the operating budget the service would sit in.
- Record what that person will do differently once the system works. If nothing changes, the benefit is not real.
- Check the agency's AI use case register for the same problem already solved elsewhere, as the checklist requires.

02 What the process does now

BEFORE THE POC STARTS

Measure the current process before anything is built: volume, time, cost and error rate. Without those numbers nothing measured afterwards means anything.

HOW TO CHECK

- Record volume, elapsed time, effort and error or rework rate for the current process, before the build starts.
- Take the measurement from records the agency already keeps where possible, rather than from asking people.
- Identify a comparison group, even an imperfect one. A team not using the system is worth more than a larger sample of users.
- Write down in advance what result would cause the agency to stop. An evaluation that cannot fail is not one.
- Say plainly what the saved time is expected to be used for, and check afterwards whether it was.

03 Whether the system is any good

END OF POC, REPEAT AT PILOT

Score the output against the source, and score whether it is worth acting on, and keep them apart. A demonstration is not an evaluation.

HOW TO CHECK

- Score two axes separately: whether the output is supported by the source, and whether it is worth acting on.
- Put the checkable axis on anyone who can read. Reserve subject matter experts for the judgement axis, because they are the scarce resource.
- Have two people score a sample independently and report how often they agreed. Report it even when it is poor.
- Evaluate the use case in front of you. Do not carry a score across from another task or another agency.
- Record the questions the system got wrong, not only the proportion.

04 Build, buy or tune

END OF POC

For almost every Australian government use case the answer is buy, or assemble from bought parts. Training or fine-tuning a model is warranted in named circumstances, and they are narrow.

HOW TO CHECK

- Establish what a frontier model with good retrieval and structured prompting achieves on the task first. That is the baseline every other option has to beat.
- If fine-tuning is proposed, name the capability gap it closes, in one sentence, that retrieval cannot.
- Price the recurring obligation, not the build: who redoes the work on each base model turn, and who maintains the evaluation harness that proves it is no worse.
- Check whether the training data includes personal information the agency holds for another purpose.
- Check the whole-of-government arrangements before building anything. The agency may already have access.

05 What production costs, and who pays

BEFORE THE PILOT STARTS

Identify the operating budget the service will sit in. Name the person who owns the benefit from a business-as-usual role, and the process change the benefit depends on. Do this before the build, not after it.

HOW TO CHECK

- Name the operating budget line the service will sit in, and the person who controls it.
- Name the benefit owner by role, and confirm the role is business-as-usual rather than project.
- Write down the process change the benefit depends on, and get the person who owns that process to agree to it in writing.
- Cost the recurring items: evaluation, monitoring, support, and migration to successor models.
- Check the consultation obligations that apply if the work changes, and start them on the right side of the decision.
- Confirm the production pathway before the proof of concept starts. The answer to who pays is cheapest to get wrong at the beginning.