The IGC AI Playbook

A practical guide to choosing, building and scaling AI in government.

DIRECTED BY LSE UNIVERSITY OF OXFORD

FUNDED BY UKaid from the British people

IGC AI PLAYBOOK FOR GOVERNMENT

STEP 1

Get Ready

Most AI projects fail because of missing buy-in, unclear ownership, or unaddressed concerns. This step helps you identify where to focus before building anything.

Readiness quiz

Answer honestly in order to properly identify focus areas to be AI-ready.

1. How do you feel about using AI in your organisation?

( ) Ready to start	( ) Curious but cautious	( ) Concerned about the risks	( ) Sceptical it will work

2. What is your biggest concern about AI?

( ) Data privacy and security	( ) Bias or unfair outcomes	( ) Staff losing jobs	( ) I don't have major concerns

3. Do you have leadership buy-in?

( ) Actively pushing for it	( ) Open to a concrete proposal	( ) Cautious, not opposed	( ) Not a priority right now

4. Are there people who would resist change?

( ) No, team is open	( ) A few individuals	( ) Significant resistance expected	( ) We haven't thought about it

5. Is there an internal champion who could own this?

( ) Yes, ready to lead	( ) Possibly, needs support	( ) No one obvious	( ) Not yet discussed

6. What is the level of overall digital and data literacy on your team?

( ) Staff are experienced with data analysis and familiar with tools such as Python, R, PowerBI or AI tools for data analysis.	( ) Staff have some familiarity with data analysis and work well in Excel.	( ) Staff do a lot of qualitative work but do not have much experience with data analysis, interpretation, or working with tools such as Excel.	( ) Staff mostly work with PDFs or paper files and do not have a strong understanding of digital tools.

7. Do you have engineers or technical staff?

( ) Yes, in-house team	( ) Some technical capacity	( ) We rely on external help	( ) No technical capacity

8. Have you built any AI projects before?

( ) Yes, successfully	( ) Tried with mixed results	( ) Small experiments only	( ) This would be our first

9. How good and available is your data?

( ) Clean, structured, and accessible	( ) Exists but needs cleaning or consolidation	( ) Incomplete or mostly undigitised	( ) We have limited relevant data

10. How clear are you on the problem you want to solve?

( ) Very clear	( ) A few ideas, not decided	( ) Something needs to change but not sure what	( ) Have not identified problems

Your focus areas

Based on your answers, identify which of the areas below apply to you. Most teams will have 1–2. Focus there before moving to Step 2.

Build buy-in first

Focus here if:

Q3 = Cautious / Not a priority, or Q4 = Significant resistance, or Q5 = No one obvious

Do this now:

-> Identify your champion and get their commitment before starting anything else

-> Book 30 minutes with your leadership to discuss one specific problem AI could solve

-> Share a relevant case study from the AI menu with leadership

Address concerns before starting

Focus here if:

Q1 = Concerned or Sceptical, or Q2 = Privacy / Bias / Jobs

Do this now:

-> Run a short "what worries us about AI" session with your team to bring concerns into the open

-> Map which data is sensitive vs. safe to use before choosing a solution

-> Start with a low-risk use case: text summarisation or document search

Find a technical partner

Focus here if:

Q7 = Rely on external / No capacity, or Q8 = First project

Do this now:

-> Be clear on the problem before approaching any technical partner. They should fit your needs.

-> Contact IGC to explore support options

-> Scope a small proof of concept with a partner before committing to a full build

Sort your data first

Focus here if:

Q9 = Mostly paper-based or Collect but rarely use

Do this now:

-> Spend one hour mapping where your data lives: who has it, in what format

-> Identify the one dataset you'd most want to use and find out how to access it

-> Consider data digitisation, text analysis or AI-powered search, which require less data to get started

You're ready	
Focus here if: *Most answers in the first two options across Q3–Q9*	Do this now: -> Run Step 2 with your team: identify your top 1–2 problems -> Set a 3-month goal: what do you want to have piloted? -> Connect with IGC to identify potential support

Building Buy-in

The biggest obstacle to AI adoption is rarely technical. It is almost always about who the current system works well for. Map your stakeholders before moving forward.

Question	Your answer
Who currently benefits from the way things are done?	
Who stands to gain from a change?	
Who might resist, and why?	
How will you bring them along?	

Identify your champion

Every successful AI project has an internal owner who cares whether it works and has the authority to drive it forward. If you cannot identify this person, stop here until you can.

Our AI champion is (name and role):

The champion needs:	+	+
Authority to make decisions	Time to dedicate to the project	Credibility with the team

Do this now

Get your champion's commitment in writing.

Get your leadership's buy-in on the specific problem you want to solve.

Run a short session with your team: what do we know about AI? What worries us?

STEP 2

Identify Your Problem

Start with a real problem your team faces every day. The technology should serve the problem.

Time	Best for	Output
30 minutes	Teams of 3–8	Your top 1–2 priority problems

02 Brainstorm

Use the questions below to draw out problems worth solving. Write everything down. You will prioritise in the next step.

What takes up the most time in your team? What's the most tedious part of that process?	What do you wish you didn't have to do?
Where do you spend money that feels excessive?	Where are citizens unsatisfied with services?
What is hard to predict?	Where is institutional knowledge stuck?
What can't you measure?	**What would you change first if you could?**

Problem 1

Problem 2

Problem 3

The smaller, the better

Look for a problem that is real, specific, and solvable. A small change to your existing processes that would save time or money. Think in terms of hours, days, weeks of effort. The goal is not to build a whole new system from the ground up.

02 Rate your problems

Score each problem on impact (how much it matters) and complexity (how hard it would be to solve). Circle the ones that have medium-high impact, lower complexity, and higher priority.

Scoring table

	Impact	Complexity
Low	Minor improvement to one process or team	Data exists, limited skills needed, low resistance
Medium	Meaningful time or cost saving, or better outcomes for a service	Some data prep, technical help, or alignment needed
High	Major improvement to a core service or large number of people	Significant data work, specialist skills, or political difficulty

Problem	Impact	Complexity	Priority
	Low / Medium / High	Low / Medium / High	Low / Medium / High
	Low / Medium / High	Low / Medium / High	Low / Medium / High
	Low / Medium / High	Low / Medium / High	Low / Medium / High
	Low / Medium / High	Low / Medium / High	Low / Medium / High
	Low / Medium / High	Low / Medium / High	Low / Medium / High

02 Classify your problem type

Match your circled problems to the types below. This will help you find the right approach in Step 3. If it does not align with any of the below, feel free to create a new problem type.

Manual workloads Data entry, filing, sending letters	**Data analysis** Analysing financial or admin data
Forecasting Predicting yields, revenue, demand	**Prioritisation** Where to focus limited resources
Policy evaluation Assessing intervention impacts	**Institutional knowledge** What does the statute say about X?
Citizen response Complaints, FAQs, call centres	**Data accessibility** Siloed, paper-based, or scattered data

Your priority problem	Problem type

Step 2 summary: your priority problems

Fill in the top two problems you identified here.

Priority Problem 1	Priority Problem 2
Problem description	**Problem description**
Problem type	**Problem type**
Impact score	**Impact score**

Next: Step 3

Take your priority problems to the AI menu. Then come back to Step 3 to record what you found and decide how to proceed.

IGC AI PLAYBOOK FOR GOVERNMENT

STEP 3

Defining the Solution

This step connects your problem to appropriate approaches. Use the IGC AI Menu alongside this workbook. It shows what is possible and what others have done in similar contexts.

Time	Also needed	Output
45 minutes	IGC AI Menu document	Chosen approach + risk assessment

Data requirements by approach

Before choosing an approach from the AI menu, check whether you have the data it requires. Use the table below as a quick reference.

Approach	Starting bar	Data you need	If you don't have it
Text analysis, search & chatbots	Low	Documents, policies, reports, or text records in any format, including scanned paper	Start with what you have. You may be able to create an impactful solution just by bringing together your documents and extracting insights.
Anomaly detection	Medium	A substantial sample of transaction or operational records, in consistent format, so that patterns and outliers can be identified.	Sort out the data layer first. This may mean data centralisation, cleaning, or dashboards before applying AI.
Prediction & classification	High	Structured historical data with past decisions or outcomes. More labelled examples improve accuracy.	Build a labelling process or pick a different approach.
Computer vision	High	Satellite, aerial, or imagery data. May require a specialist technical partner.	Consider remote-sensing partners and subject matter experts.

Your project can be smaller than the menu examples

Many of the case studies started small. Think about automating one repetitive task, building a single dashboard, or using an LLM to summarise one type of document.

Starting smaller is best to produce something with real, immediate impact. Given the scale governments work at, even a tiny change can result in large gains.

Match your problem to an approach

Review the AI menu, identify 1–2 candidate approaches for each priority problem, and record your thinking below. If you can think of other approaches, you may add them as well. The menu is a guide but is not necessarily exhaustive.

Priority problem	Candidate approach	Data gap to close?

Build vs. buy

This is a core decision governments have to make, and the answer depends on each given context. Choose the option that best fits your capacity and context.

Build	Buy / adapt	Hybrid
You have technical capacity and need something custom to your context	A proven tool already exists for this use case. Verify it works in your context.	Buy a base solution and customise it with local expertise, or partner with shared ownership

Internal	External
Advantages	**Advantages**
+ Builds in-house capacity	+ Faster to deliver
+ Cheaper over time	+ Access to specialist capabilities
+ Full ownership and control	+ Can be funded by donors
Watch out for	**Watch out for**
- Slower to deliver	- More expensive
- Requires existing or hired technical skill	- Risk of dependency on the vendor
- Team may be stretched across other priorities	- Donor funding may come with conditions that reduce your control

Question	Your response
Is there a proven off-the-shelf solution already working in a comparable context?	
If buying: What demonstration requirements and agency control requirements do you need?	
If building: Do you have realistic in-house or partner capacity (engineers, PMs)?	
How will you avoid vendor lock-in?	
What are the invisible costs of each option? (local context gaps, training, dependency)	
What are the visible costs of each option? (vendor prices, LLM use costs, engineer salaries)	

Your recommendation and rationale:

Risk assessment

Check each risk before committing to an approach.

Each of these risks has options for mitigation. Don't discard a project just because it may have some risks. Evaluating tradeoffs and being mindful of risks creates a strong and defensible approach when you build.

Risk	Key question	Level	Mitigation options
Ethics	Could this produce biased or unfair outcomes?	Low / Medium / High	Test outputs across different population groups before deploying Include affected communities in the design process Set a human review threshold: any decision above a certain impact level requires sign-off Document known limitations and make them visible to users of the system
Privacy	Does it involve personal or sensitive data?	Low / Medium / High	Use anonymised or aggregated data where the use case allows it Get legal sign-off on data use before building anything

			Apply data minimisation: only collect and use what the system actually needs Define a data retention and deletion policy before launch
Safety	What happens when it is wrong? Is there a human check?	Low / Medium / High	Build in a human review step for any high-stakes output Define what "wrong" looks like and integrate tests and alerts that flag it Start with the system as a recommendation, but a person makes the final decision Set a clear process for users to report errors Implement safeguards to prevent harmful outputs
Political	Who could block this, and how will you address that?	Low / Medium / High	Involve the teams most affected in the design process from the start Identify a senior sponsor who can clear blockages Frame the project around a shared problem, not around the technology
Dependency	Does this create reliance on a vendor we can't exit?	Low / Medium / High	Require access to your own data and model outputs, in writing, before signing any contract Consider open-source or standard options Build internal capacity to maintain the system as part of the project scope Define exit criteria: what would need to happen for you to switch providers

Your risk mitigation plan

Low risk options you can start with

Next: Step 4

You now have a problem and an approach. Next is building the smallest version that works.

IGC AI PLAYBOOK FOR GOVERNMENT

STEP 4

Pilot

Build the smallest version that proves it works. Treat every finding, including failures, as useful data to evaluate what works.

Time	Best for	Output
60 minutes	Champion + technical lead	Pilot plan + impact framework

Define what success looks like before you build

Agree on metrics and collect baseline data before the pilot starts.

The metrics should help you answer: How will we know if this worked?

Identify how you will measure it and determine impact.

J-PAL'S AI Evidence Playbook provides guidance on building evidence with AI programs at https://www.povertyactionlab.org/ai-evidence-playbook.

A key takeaway: an AI solution is not worth building unless you can show that it effected change. For example, you implemented a flashy new AI model, but did it lead to improved policy outcomes? Ideally, this is determined through randomized evaluations.

Metrics to consider	What it measures	How you will measure it	Baseline (before)	Target (at 6 months)
Time saved	Hours or days per month			
Cost saved	Estimated in local currency or staff time			
Quality improved	E.g. accuracy, speed, citizen satisfaction			
[Add your own]				

Minimum threshold to justify scaling (what result would make you confident to expand?)

Understand your user

Who are the end users? (be specific) What do they need the system to do?

Describe a specific person who will use this tool	
Name and role	*Give this person a name and describe their job*
A typical day	*What does their working day look like in relation to this problem?*
What this costs them	*E.g. time, errors, stress; what does the current situation take from them?*
What barriers there are to them using a new digital tool	*E.g. digital literacy, internet access*

Step	What they do now	Where they get stuck, wait, or make errors	With the AI tool
1			
2			
3			
4			
5			

Are there any changes you'd make to your MVP now that you've considered this?

Who is building this?

You don't necessarily need one person for each role, but the following responsibilities should be covered.

Role	Name or team	Time commitment
Internal owner	*Accountable for the project succeeding and defining the product output*	*Time commitment:*
Technical lead	*Responsible for building the actual tool*	*Time commitment:*
Product Manager	*Knows the problem and the context in depth. Responsible for the day-to-day management and acts as a bridge between the internal owner and the technical lead.*	*Time commitment:*
User representative	*Tests outputs and gives feedback during the pilot*	*Time commitment:*
Senior sponsor	*Clears institutional and political blockers*	*Time commitment:*

If you need external support

Strategies to consider when missing in-house expertise:

→ Leverage the IGC Evidence Lab

→ Look across government for cross-ministry support

→ Procure services from an outside firm

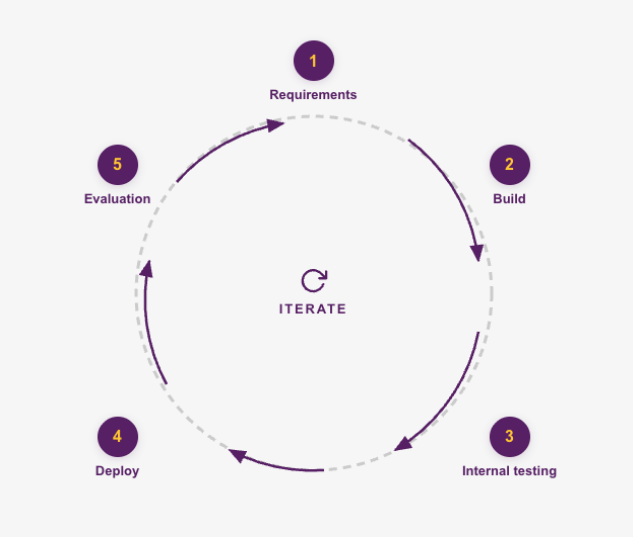
→ Ask for donor support

What we need from a technical partner:	What we will own ourselves:	How we will avoid long-term dependency

Development timeline

A note on agile

In the tech world, agile methodologies have become the standard. Going through each step sequentially and waiting for each to be complete before moving on - known as the waterfall methodology - creates bottlenecks and slows delivery. The faster approach is to iterate quickly through the full cycle: build a small piece, test it with real users, learn from what you find, and adjust. Then repeat the cycle. This is why we suggest a minimal version (the MVP) from the beginning - so you can get something out quickly to test and react to.



Set realistic dates. If you cannot fill in the "by when" column, the project is not ready to start.

Phase	What happens	Who leads	By when
Requirements	*Finalise what the tool needs to do and how success will be measured* *These need to be specific items listed out in a document which the developers can follow.*	*Internal Owner + Optional Analyst* *Incorporate feedback from technical lead*	
Build	*Develop the first version of the tool*	*Technical lead*	
Internal testing	*Check it works before showing it to end users*	*Internal Owner + Technical Lead*	
Pilot launch	*Deploy to a limited group of real users*	*Internal Owner, Technical Lead, User Representative*	
Evaluation	*Assess results against go / no-go criteria*	*Internal Owner*	

Define your MVP

Your minimum viable product is the simplest version that delivers the core value. Start here, then add features based on what you learn.

> **The bicycle principle**
> A bicycle solves the mobility problem immediately, even if a car is the eventual goal. Your MVP should work and deliver value as the first deployed version, which can then be iterated on.

What is the smallest version of this that would genuinely help? (one sentence)

Pilot scope: which users, geography, or function will be covered?

Pilot end date and decision point:

Budget and resources

Cost of building (staff time + any external costs)	
Cost of running for 6 months	
Cost of maintaining long-term	
Where the funding is coming from	

Pilot checklist

Go through this checklist before your pilot starts. If any item is not checked, address it first.

[ ]	Problem statement and success criteria defined before starting
[ ]	Impact measurement active, with baseline data collected
[ ]	Scope is limited: specific users, geography, or function
[ ]	Champion is briefed and committed
[ ]	Human review process in place for system errors
[ ]	Mechanism to capture user feedback during the pilot
[ ]	Go / no-go criteria agreed before the pilot begins
[ ]	Timeline is fixed: the pilot has an end date

Next: Step 5

Run your pilot and measure against your baseline. When you have evidence, take it to Step 5.

IGC AI PLAYBOOK FOR GOVERNMENT

STEP 5

Adapt and Scale

Pathways to scaling

Step #1 is moving the pilot into production. The product needs to go beyond provisional testing to implementing a real, maintained system that people actually use.

What needs to happen to move from pilot to production? (sign-offs, infrastructure, handover, training)

Who will continue to maintain it, and what is the process?

Criteria for scaling

Criterion	Status
Impact targets from Step 4 have been met or exceeded	Met / Partially met / Not yet
Users have adopted the system (not just piloted it)	Met / Partially met / Not yet
The technical system is stable and maintainable	Met / Partially met / Not yet
Political and institutional support is in place	Met / Partially met / Not yet
There is a clear owner for ongoing maintenance	Met / Partially met / Not yet

Identified risks or concerns have been addressed	Met / Partially met / Not yet

What changes or improvements will you make before scaling?

Once this is complete, scaling can take several forms.

Pattern 1 Deepen	Pattern 2 Replicate	Pattern 3 Adapt
Use the same use case and office, but address more cases. For instance, extend coverage from sample data to the full dataset, or expand it from in-office use for validation to on-the-ground use for real-time prediction.	Apply the same use case horizontally across more offices or locations. For instance, roll it out to additional districts or ministries.	Apply the same or a slightly adapted approach to a different use case. For instance, fraud detection methods for tax could be adapted to electricity theft.
Watch for: New edge cases not identified by the pilot. Plan a review at 3 months.	Watch for: Local context differences, training requirements, onboarding.	Watch for: Treating every case the same and copy-pasting the method. Each new domain needs to review the problem-definition step and adapt accordingly, using the steps in this playbook.

What does the first phase of scaled deployment look like?

Report your results

Document what happened, including what did not work. Honest reporting builds credibility, improves your next project, and helps other governments facing the same challenges.

Key results against your Step 4 metrics

What worked, and why?

What did not work, and why?

Key lessons to bring back into Step 2 for the next project

Share your results

Hold conversations to discuss outcomes and lessons learned with your team. **Maintain a repository** of AI projects so teams can easily find and access existing AI projects they can build on while avoiding repetitive work.

Where will you share and maintain documentation from your project?

Case Study

The problem: In Zambia, the national statistics agency needed a way to improve their national statistics through better classification of open-ended survey responses.

The context: The survey data was readily available, but occupation and industry classifications often suffered due to the overwhelming number of possible codes, poor and unstructured response quality, and lack of time and context.

The solution: An AI classification system was identified as a valuable approach to address this problem.

The pilot: An initial classification system was built and evaluated for accuracy, which showed promising results. It was refined and tested with various models, but not to perfection.

The deployment: The Zambia Evidence Lab scheduled a deployment with the statistics agency, involving a training with numerous public servants for hands-on testing and use of the tool. Some edge cases were identified, along with a post-hoc analysis of failure points which resulted in iterations for improving the prompt, testing different models, adjusting parameters, and improving input survey data.

Scaling: The first step was deploying it for use within the agency offices. Next was to deepen the tool's reach by implementing it on the ground during survey collection, involving change management with training for enumerators, a phased rollout, and randomized treatment groups to evaluate the real impact of the tool.