PHOENIX

The UK Public Sector AI Adoption Gap

Why UK public sector AI stalls between pilot and value

Reading time: 12 minutes | Published November 2026



Microsoft Solutions Partner

Frontier Partner

Inside this report

Nearly every UK public sector organisation is using AI. Few can prove what it changed. This report brings together the published evidence on why pilots stall, and what closes the gap.

The report includes evidence from government trials, independent research, and industry analysis. It explains why so many pilots stop short of value, and what the organisations getting it right do differently.

00	Executive summary	P. 03
01	Data and sources	P. 05
02	The prize is real, and it has been measured	P. 06
03	Same tool, three departments, three answers	P. 08
04	Five reasons pilots stall	P. 10
05	The sector picture	P. 12
06	What good looks like	P. 15
07	Where to start	P. 17
—	Closing	P. 18
—	FAQ	P. 18
—	Source register	P. 19

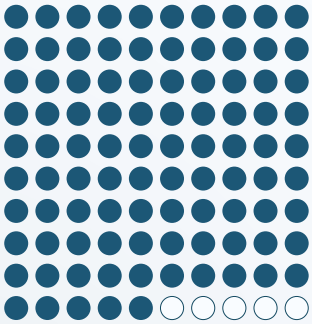
Every figure in this report is named, dated, and linked. Where the evidence is thin, we say so. The full source register is at the back.

Executive summary

Everyone is adopting. Almost nobody is scaling.

Two years ago the public sector question was whether to use AI at all. That question has been settled by staff, not by strategy. Adoption arrived from the bottom up, through browsers and free tools, faster than any technology before it.

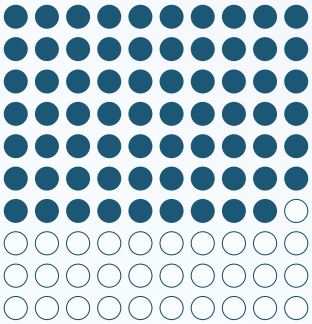
The numbers show how complete that shift is.



95%

of responding councils are using or exploring AI

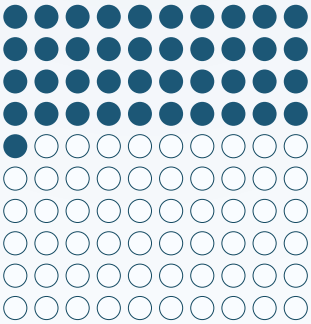
Source: LGA, 2025



79%

of charities use AI tools in day to day work

Source: Charity Digital Skills Report, 2026



41%

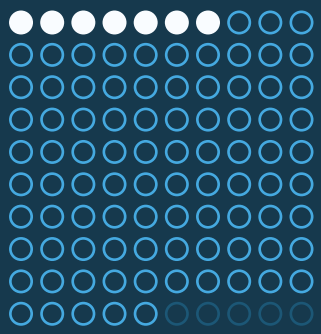
of public sector working time could be supported by generative AI

Source: Alan Turing Institute, ONS, and DfT, 2025

Next ⟶
Chapter 01. Data and sources

Book an AI Surveyor scoping call

Now look at the second set of numbers, from the same surveys.

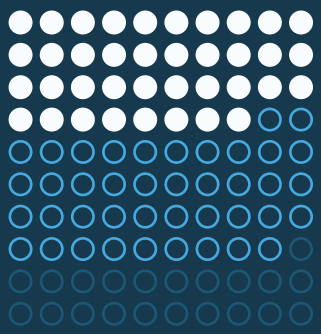


7%

of those councils rate themselves as leaders in their use of AI

- 7% rate themselves as leaders
- 95% are using or exploring AI

Source: LGA, 2025

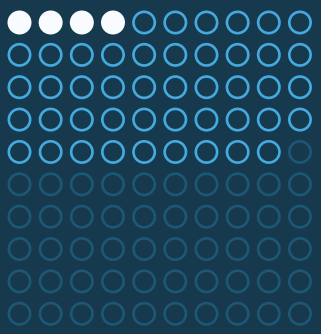


38%

of charities use AI actively or strategically

- 38% use AI actively or strategically
- 79% use AI tools day to day

Source: Charity Digital Skills Report, 2026



4%

of teachers report using generative AI for marking

- 4% of teachers use generative AI for marking
- 49% of education working time AI could support

Source: Alan Turing Institute, 2025

That is the adoption gap. Near-universal use sitting on top of very thin claims to maturity. The tools are in the building. The value is not yet on the balance sheet.

The gap is not caused by the technology. Three government departments ran trials of the same product and reached three different conclusions, which we cover in Chapter 03. What separated them was scope, use case selection, readiness, and how they chose to measure.

Closing the gap takes three things, in this order.

1 → 2 → 3

Direction

Decide what problem you are solving, and for whom, before you decide which technology solves it. Most stalled pilots reversed that order.

Proof

Build something real, with success measures agreed before the build starts. A pilot with no baseline cannot produce evidence, only opinion.

Trust

Design security, governance, and accountability in at the start. Added afterwards, they stop a working pilot from ever reaching production.

Next →
Chapter 01. Data and sources

Book an AI Surveyor scoping call

Chapter 01

Data and sources

01

This report contains no Phoenix survey and no Phoenix telemetry. It draws entirely on published evidence, and it names all of it. There are three groups of sources.



Government evaluations

Published findings from the Government Digital Service, the Department for Work and Pensions, the Department for Business and Trade, the Department for Science, Innovation and Technology, and the National Audit Office. These are the most useful sources available on public sector AI, because the organisations concerned had no product to sell.



Independent research

Work from the Alan Turing Institute with the Office for National Statistics and the Department for Transport, the Local Government Association, the Charity Digital Skills Report, and academic research into council AI readiness led by Heriot-Watt University.



Industry analysis

Forecasts and analysis from Gartner, and the widely cited MIT study on enterprise generative AI. We flag where that work is preliminary.



What we do not have

Phoenix does not yet publish outcome data for AI Surveyor engagements. We are not going to borrow someone else's numbers and imply they are ours. When we have our own results, with customer permission, they will appear here and be dated.

⟵ Previous
Executive summary

Next ⟶
Chapter 02. The prize is real

Book an AI Surveyor scoping call

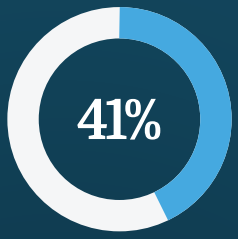
Chapter 02

The prize is real, and it has been measured

02

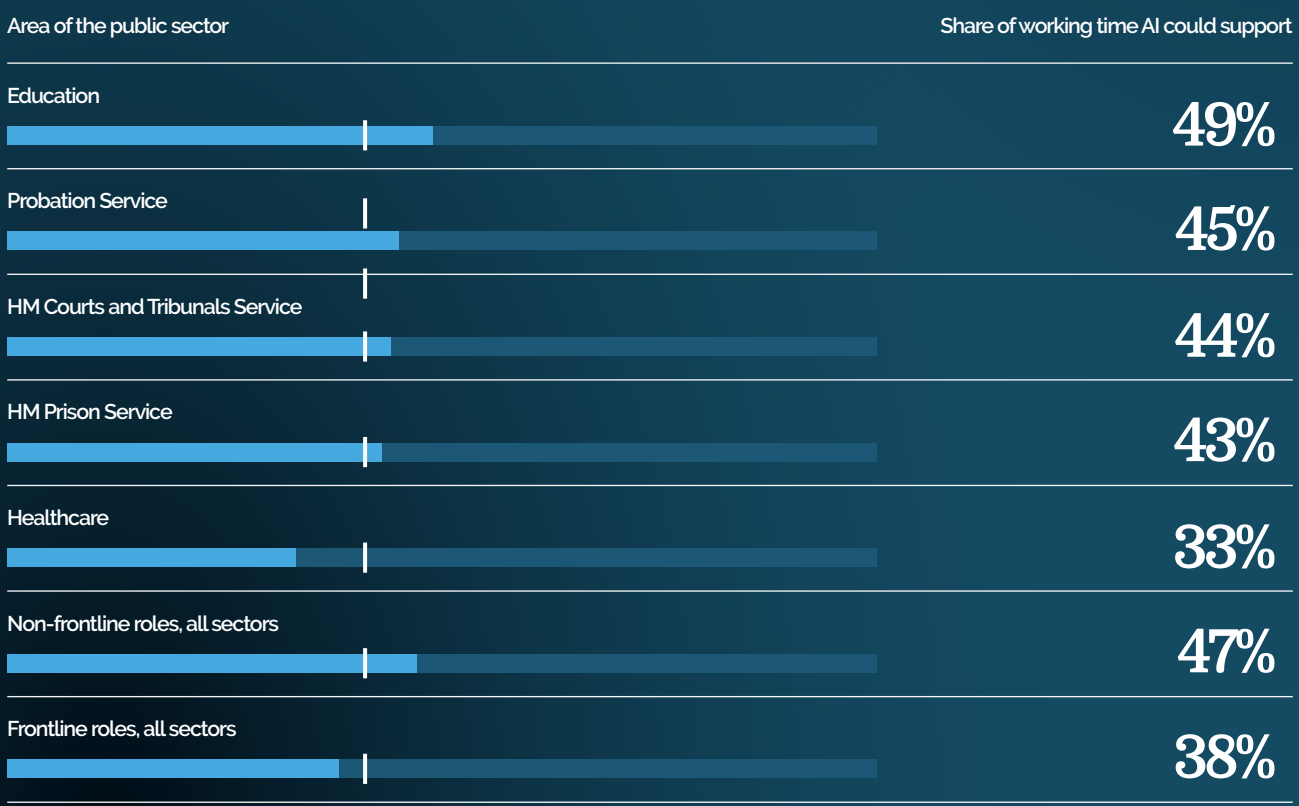
Start with the size of the opportunity, because it is genuinely large and it has been quantified by people with no commercial interest in the answer.

Researchers at the Alan Turing Institute, working with the Office for National Statistics and the Department for Transport, mapped what public sector staff spend their time doing. They then compared those activities against what generative AI can currently do.



Around 41% of public sector working time is spent on activities that generative AI could support. The figure varies by sector, and by how close to the front line the role sits.

Share of working time AI could support



Public sector average, 41%

Source: Mapping the Potential: Generative AI and Public Sector Work. The Alan Turing Institute, ONS, and DfT, June 2025.

← Previous
Chapter 01. Data and sources

Next →
Chapter 03. The prize is real

Book an AI Surveyor scoping call

The government has put a figure on the wider prize too. Its State of Digital Government Review found more than £45 billion a year in unrealised savings and productivity benefits, worth 4 to 7% of public sector spend, available through full digitisation of public services.

We should be straight about that number. The Commons Science, Innovation and Technology Committee later described it as worryingly optimistic, and the former head of the government's Incubator for AI called it a reasonable guess that is very hard to measure. Treat it as a direction of travel, not a business case.

Where trials have measured actual time, the results are more modest and more useful. The Government Digital Service ran the largest deployment of Microsoft 365 Copilot in any organisation to date: 20,000 staff across 12 government organisations, over three months.

26 minutes
saved per person, per day

13 days
a year, if sustained across a full working year

82%
of users would not want to go back to working without it

Source: Government Digital Service, M365 Copilot experiment findings, published June 2025.

More than 70% of those users agreed that Copilot cut the time they spent searching for information and doing mundane tasks. Document drafting saved 24 minutes on average. Building a presentation saved 19.

The question is no longer whether AI can help. It is whether your organisation can prove that it did.

⟵ Previous
Chapter 01. Data and sources

Next ⟶
Chapter 03. The prize is real

Book an AI Surveyor scoping call

Chapter 03

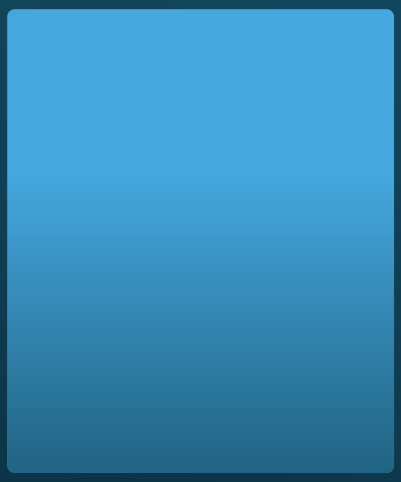
Same tool, three departments, three answers

03

This is the finding that should change how public sector organisations plan AI work.

Three UK government departments ran evaluations of the same product, Microsoft 365 Copilot, over overlapping periods. All three published. They did not agree.

Same licence. Three verdicts.



26 min

a day saved. Government Digital Service, 20,000 staff, 12 organisations

Source: GDS, published June 2025



19 min

a day saved, measured against a comparison group. DWP, 3,549 licences

Source: DWP, published January 2026

None

No robust evidence that time savings led to improved productivity. DBT, 1,000 licences

Source: DBT, published August 2025

The Department for Work and Pensions ran the most rigorous of the three. It compared 1,716 licence holders against 2,535 non-users, added interviews, and adjusted for differences between the groups. It found 19 minutes a day, plus small but statistically significant improvements in perceived work quality and job satisfaction. Neurodivergent staff reported particular benefit.

The Department for Business and Trade found something different. Staff liked the tool, with 72% satisfied or very satisfied. But the department could not find robust evidence that time savings had turned into productivity. Some tasks got slower. Slide production was around seven minutes faster but needed correcting for quality. Analysis in Excel took longer and was less accurate

⟵ Previous
Chapter 02. The prize is real

Next ⟶
Chapter 04. Five reasons pilots stall

Book an AI Surveyor scoping call

Usage was the most revealing number

1.14

Copilot actions per day, per average participant, across 63 working days.

Read that again. Roughly one action a day. This was not a technology that failed to deliver. It was a technology that was barely used.

The department was also honest about its own method. Productivity was not a key aim of the evaluation, so limited data was collected on it. The negative result is partly a measurement design story.

What differed

The licence was identical in all three cases. What varied was everything around it.

01 Whether specific use cases were identified before rollout, or licences were handed out and left to find their own purpose.

02 Whether a baseline existed, so that any change could be measured against something.

03 Whether staff were trained and supported, or expected to work it out.

04 Whether the evaluation was designed to detect productivity at all.

05 Whether security and data handling concerns were resolved up front. GDS found that unresolved concerns reduced the benefit for a minority of its users.

Nothing in these three evaluations is a technology failure. Every difference between them is a decision that was made, or not made, before the work began.

⟵ Previous
Chapter 02. The prize is real

Next ⟶
Chapter 04. Five reasons pilots stall

Book an AI Surveyor scoping call

Chapter 04

Five reasons pilots stall

The pattern is consistent across the evidence. Pilots stop for the same five reasons, and none of them is model performance.

04

No decision gate

The technology gets chosen before the problem is understood. Gartner reports that at least half of generative AI projects were abandoned after proof of concept by the end of 2025, citing poor data quality, inadequate risk controls, escalating costs, or unclear business value. Its earlier forecast had been 30%.

A decision gate is simply a point at which you stop and choose, with evidence in front of you. Copilot, Copilot Studio, low-code, pro-code, or machine learning are different answers to different problems. Picking first and discovering later is expensive.

Data foundations that are not ready

Councils responding to the Local Government Association named data availability, quality, and storage as their second largest area of unpreparedness, with 68% saying they were not very or not at all ready. Workforce skills came first at 71%.

The National Audit Office reached the same conclusion about central government. Updating legacy systems and improving data quality and access is fundamental to exploiting AI, and it takes time.

Governance arriving late

When the National Audit Office surveyed 87 government bodies, only 30% had risk and quality assurance processes that explicitly covered AI risks, though a further 46% planned to introduce them. Only 21% had an AI strategy at all.

Governance added after a pilot works is not a formality. It is the reason a working pilot never reaches production.

⟵ Previous
Chapter 03. Three answers

Next ⟶
Chapter 05. The sector picture

Book an AI Surveyor scoping call

A legacy estate in the way

The State of Digital Government Review found 28% of central government systems classed as legacy, up from 26% the year before. The picture is worse in parts of the sector that carry the most operational risk.

Reported legacy share of estate

Part of the public sector	Reported legacy share of estate
Central government	28%
NHS trusts	10% to 50%
Police forces	10% to 70%
Organisations unable to quantify their legacy estate at all	15%

Source: State of Digital Government Review, DSIT, January 2025.

Maintaining those systems often costs three to four times what a modern equivalent would. That is money not available for the thing you want to do next.

Feasible is not the same as adopted

This is the most overlooked reason of all. The Turing research found that 49% of education working time sits on activities generative AI could support, the highest of any sector. It also found that only 4% of teachers report using generative AI for marking.

Technical feasibility and successful adoption are different problems. One is solved by procurement. The other is solved by change management, training, permission, and trust.

The wider market signal

The most quoted figure in enterprise AI comes from MIT's NANDA initiative, which reported that around 95% of enterprise generative AI pilots produced no measurable impact on the bottom line. It is worth being careful with it. The work was preliminary, was not peer reviewed, and its method has been challenged. Treat it as directional.

The direction, though, matches everything above. Very few organisations are converting pilots into value, and the reasons are organisational rather than technical.

⟵ Previous
Chapter 03. Three answers

Next ⟶
Chapter 05. The sector picture

Book an AI Surveyor scoping call

Chapter 05

The sector picture

05

Public sector is not one market. Eleven sectors, eleven sets of pressures, eleven different first moves. Here is where each one sits, and where we would start.

Central Government

Where the sector sits

Around 28% of systems are legacy. Spend controls are tight and scrutiny is constant. Three departments have now published Copilot evaluations that disagree with each other.

Where we would start

Pick one process with a named owner and a measured baseline. Publish the result internally, whichever way it goes.

Local Government

Where the sector sits

95% of responding councils are using or exploring AI, but only 7% consider themselves leaders. Funding, staff skills, and staff capacity are the three barriers councils name most. Reorganisation is reshaping the map underneath all of it.

Where we would start

Fix the data foundations before building agents. A single view of the resident is worth more than a chatbot on top of fragmented records.

Policing

Where the sector sits

Legacy estate runs as high as 70% in some forces. Officer time is the currency that matters. OFFICIAL-SENSITIVE handling is not negotiable.

Where we would start

Target back office and administrative burden first, not frontline decision making. Prove time returned to the watch.

Fire and Rescue

Where the sector sits

Reliability, interoperability, and control room integration come before anything else. Capital funding is uneven and estates vary widely between services.

Where we would start

Start where failure is recoverable. Prevention, protection, and reporting workloads, not incident-critical systems.

⟵ Previous
Chapter 04. Five reasons pilots stall

Next ⟶
Chapter 06. What good looks like

Book an AI Surveyor scoping call

Defence

Where the sector sits

Every claim needs to be provable and every capability accredited. Classification boundaries shape what is possible before any technology conversation starts.

Where we would start

Define the governance and classification envelope first. Scope the use case inside it, not the other way round.

Healthcare

Where the sector sits

Around a third of working time sits on activities AI could support, the lowest of any sector, because so much of the work is human. Trusts report legacy estates of anywhere between 10% and 50%.

Where we would start

Administrative and documentation burden, with clinical safety and information governance involved from day one rather than at approval.

Charity and Social Impact

Where the sector sits

79% of charities use AI tools in day to day work, but only 38% use them actively or strategically. Accuracy concerns hold back 49%, and data privacy and security concerns hold back another 49%. 60% now have an AI policy or are writing one.

Where we would start

A policy staff can follow, then one funded use case that returns hours to the mission.

Housing

Where the sector sits

Regulatory scrutiny is rising alongside tenant expectations and ageing stock. Damp and mould, repairs, and complaint handling all generate high volumes of contact and evidence.

Where we would start

Tenant contact and repairs triage, with a complete audit trail for the regulator.

Schools and Trusts

Where the sector sits

Education has the highest AI potential of any sector at 49% of working time. Yet only 4% of teachers report using generative AI for marking. Safeguarding sits above everything.

Where we would start

Teacher workload relief, with safeguarding and data protection designed in and a trust-wide policy behind it.

Further Education

Where the sector sits

Multi-site estates, tight funding, and a national skills agenda to deliver. Colleges support learners from 16 to 60 with very different needs.

Where we would start

Timetabling and learner administration, measured in staff hours returned rather than in technology deployed.

Higher Education

Where the sector sits

Decisions are federated across IT, academics, finance, and senior leadership. Research data management, student experience, and cyber resilience compete for the same budget.

Where we would start

One faculty or one professional service function. Not an institution-wide programme.

Sector figures sourced in the source register. Where no published figure exists for a sector, the description is qualitative and no number is implied.

Chapter 06

What good looks like

06

Everything above points the same way. The organisations getting value from AI are not the ones with the best technology. They are the ones that decided what they were solving, proved it at small scale, and built the controls in from the start.

That is what AI Surveyor does.

Everything above points the same way. The organisations getting value from AI are not the ones with the best technology. They are the ones that decided what they were solving, proved it at small scale, and built the controls in from the start.

Advise. Find the right starting point.

Before any technology is chosen, Surveyor establishes where you are today. That means an AI strategy and readiness session, functional discovery of the process itself, technical discovery of the architecture around it, and a governance and security risk assessment run by our own GRC and cyber security teams.

The output is a document you can take to a board. Current readiness, immediate risks and blockers, the process as it works today, where AI and automation could intervene, an option set with indicative costs and licence implications, and our recommendation.

Build. Prove it with something real.

Phase two delivers a constrained pilot with success measures, evaluation criteria, and guardrails agreed before the build begins. Not a demonstration. A working proof of value, validated with the people who will use it, plus a scale plan and a benefits model.

This is the step that answers the Department for Business and Trade problem. If you agree the baseline first, you can prove what changed.

Security and governance. Inside both phases, not after them.

Security and governance sit inside both phases rather than arriving after them. The operating model, roles, and decision forums. Standards for data and for LLMOps. Preparation for scale

Where you need full platform readiness before adopting across the estate, standalone security and governance engagements can be added to the work package. That is a choice you make, not a fixed part of the engagement.

⟵ Previous
Chapter 05. The sector picture

Next ⟶
Chapter 07. Where to start

Book an AI Surveyor scoping call

Usage was the most revealing number

This is the single most important thing Surveyor does, and the thing most stalled pilots skipped.

Each phase includes a gate where the technology direction, the delivery model, and the specialists involved are all confirmed. Microsoft 365 Copilot, Copilot Studio, low-code, pro-code, or machine learning. The decision is made with evidence in front of you, before significant commitment.

One team, not a queue of suppliers

Deploying a secure AI agent is not an AI task. It takes six capabilities working together: Microsoft AI Workforce, AI App Development, Power Platform, Cyber Security, GRC, and Infrastructure. Surveyor brings all six as a single team with one point of accountability.



What you get

- Readiness assessment
- Pilot artefacts
- Use case shortlist and prioritisation
- Governance pack
- Scale roadmap

Who should be in the room

Surveyor sessions work best with the people who own the outcome and the risk. In public sector terms that usually means:

- Chief Digital and Information Officer, or Head of IT
- Chief Clinical Information Officer, in healthcare
- Director of Transformation, or the service director who owns the process
- Senior Information Risk Owner and Data Protection Officer
- Information governance or GRC lead
- The team leader whose people will use the thing

See how AI Surveyor works

← Previous
Chapter 05. The sector picture

Next →
Chapter 07. Where to start

Chapter 07

Where to start

07

Three recommendations, drawn from what the evidence above shows about the organisations that got this right.

1 → 2 → 3

Choose the process, not the product

Start with a process that has a named owner, a measurable volume, and a person who is fed up with it. Then work out which technology fits. Reversing that order is the most common and most expensive mistake in the evidence

Agree how you will measure it before you build it

The Department for Business and Trade could not prove its Copilot pilot worked, partly because it had not set out to measure productivity. The Department for Work and Pensions could, because it built in a comparison group. Decide your baseline first. It costs nothing and it changes everything about what happens next.

Design governance and security in at the start

Only 30% of government bodies had assurance processes covering AI risks when the National Audit Office looked. Governance is not what slows adoption down. It is what allows adoption to scale past the pilot.

Questions to ask before your next AI pilot

If you cannot answer these, the pilot is not ready.

01 Which specific process are we changing, and who owns it today?	**02** What does that process cost us now, in hours, in money, or in outcomes for the people we serve?	**03** How will we know if this worked, and what are we comparing it against?	**04** Is the data this depends on accurate, accessible, and lawful to use this way?
05 Who signs off the risk, and have they been in the room yet?	**06** What happens to the time we save, and who decides?	**07** If it works, what stops us scaling it next month?	**08** If it fails, will we know, and will we say so?

Talk to us about your first use case

← Previous Chapter 06. What good looks like

Next → Closing

Closing

The public sector has already adopted AI.

The staff did it before the strategies caught up. What remains is the harder part: turning tools people quite like into services that measurably work better, safely, at scale, and with evidence a board and an auditor will accept.

That is a direction problem, not a technology problem. It is also a solvable one.

Start a conversation with our AI Specialists.

Contact us now

FAQ

Why do so many public sector AI pilots fail?

Rarely because of the technology. The published evidence points to use cases chosen after the tool, data foundations that were not ready, governance added too late, and no agreed baseline to measure against. Gartner reports at least half of generative AI projects were abandoned after proof of concept by the end of 2025.

How much time can AI realistically save public sector staff?

Published UK government trials range from 26 minutes a day to no measurable gain, using the same product. The Alan Turing Institute estimates 41% of public sector working time sits on activities generative AI could support. The range is explained by approach, not by the software.

What is AI Surveyor?

A fixed-price Phoenix engagement that takes a public sector organisation from AI ambition to a working, governed pilot. It runs an Advise phase to establish readiness and choose the right approach, then a Build phase to prove value, with security and governance underneath both.

Do we need our data sorted before we start?

No, but you need to know the state of it. The Advise phase includes a data readiness assessment precisely because most organisations do not know. Finding a blocker in week two is far cheaper than finding it in month six.

Book an AI Surveyor scoping call

Source register

Every figure used in this report, with its origin, its publication date, and a link. Nothing is used in this document that is not in this table.

41% of public sector working time could be supported by generative AI. 49% education, 33% healthcare, 47% non-frontline, 38% frontline	**Mapping the Potential: Generative AI and Public Sector Work. The Alan Turing Institute, with ONS and DfT** **Published: June 2025 | turing.ac.uk**
Probation Service 45%, HM Courts and Tribunals Service 44%, HM Prison Service 43%	**Same report, as reported by PublicTechnology** **Published: June 2025 | publictechnology.net**
Only 4% of teachers report using generative AI for marking	**Cited within the Turing report above, from a then-forthcoming Turing paper** **Published: June 2025 | turing.ac.uk (PDF)**
26 minutes a day saved. 20,000 staff, 12 organisations. 13 days a year. 82% would not go back. Over 70% spent less time searching and on mundane tasks. Drafting 24 minutes, presentations 19 minutes	**Government Digital Service, M365 Copilot experiment. Trial ran September to December 2024** **Published: June 2025 | Microsoft UK Stories summary**
19 minutes a day saved. 3,549 licences. 1,716 users compared against 2,535 non-users. Perceived work quality up 0.49 and job satisfaction up 0.56 on seven-point scales	**Department for Work and Pensions evaluation. Pilot ran October 2024 to March 2025** **Published: January 2026 | The Register report**
No robust evidence that time savings led to improved productivity. 1,000 licences. 72% satisfied or very satisfied. 1.14 Copilot actions per user per day across 63 working days. Slides around seven minutes faster but needing correction. Excel slower and less accurate. 22% identified hallucinations	**Evaluation of the M365 Copilot Pilot in the Department for Business and Trade** **Published: August 2025 | assets.publishing.service.gov.uk (PDF)**
£45 billion a year in unrealised savings, 4 to 7% of public sector spend. 28% of central government systems legacy, up from 26%. NHS trusts 10 to 50%. Police forces 10 to 70%. 15% cannot quantify their legacy estate. Legacy maintenance often three to four times the cost of modern equivalents	**State of Digital Government Review, Department for Science, Innovation and Technology** **Published: January 2025 | assets.publishing.service.gov.uk (PDF)**
The £45 billion figure described as worryingly optimistic, and as a reasonable guess that is hard to measure	**Rewiring the state: Delivering digital government. Commons Science, Innovation and Technology Committee** **Published: 2026 | publications.parliament.uk**
21% of government bodies had an AI strategy, 61% planned one. 30% had risk and quality assurance covering AI risks, 46% planned to. 87 of 89 bodies responded. Legacy systems and data quality identified as fundamental	**Use of artificial intelligence in government, National Audit Office. Survey conducted autumn 2023** **Published: March 2024 | nao.org.uk (PDF)**
At least 50% of generative AI projects abandoned after proof of concept by the end of 2025, citing poor data quality, inadequate risk controls, escalating costs, or unclear business value. Earlier forecast was at least 30%	**Gartner. The 30% figure is the original July 2024 forecast; the 50% figure is Gartner's later reporting against it** **Published: 2024 and 2026 | gartner.com**

PHOENIX Microsoft

Book an AI Surveyor scoping call

Source register cont.

Around 95% of enterprise generative AI pilots produced no measurable impact on the bottom line. Based on 150 interviews and analysis of around 300 deployments	**The GenAI Divide: State of AI in Business 2025, MIT NANDA initiative. Preliminary and not peer reviewed. Method has been challenged. Use as directional only** **Published: 2025 | Fortune report**
95% of responding councils using or exploring AI. 7% considered leaders. 71% not ready on workforce skills, 68% not ready on data availability, quality, and storage. Top three barriers: funding, staff skills, staff capacity	**State of the sector: Artificial intelligence, 2025 update. Local Government Association. A third of English councils responded, so not representative of all councils** **Published: June 2025 | local.gov.uk**
79% of charities use AI tools in day to day work. 38% actively or strategically, up from 25%. 60% have or are developing an AI policy. 49% held back by accuracy concerns, 49% by data privacy and security	**Charity Digital Skills Report 2026** **Published: 2026 | charitydigitalskills.co.uk**
208 councils assessed for AI readiness across all four UK nations. Wide variation driven by leadership priorities, legacy systems, and pace of guidance	**AI Readiness of UK Local Authorities 2025. Heriot-Watt University with GoLLM and APSE** **Published: March 2026 | hw.ac.uk**