



# Government Adoption of Generative AI (Late 2022–2025)

## Introduction

Since the public release of ChatGPT in late 2022, governments around the world have begun experimenting with generative AI to improve public services and internal operations. Civilian agencies at the municipal, state (or regional), and national levels have piloted AI chatbots, large language models (LLMs), and related tools to automate citizen services, draft and translate documents, assist in policy analysis, and more. Early adopters report significant efficiency gains – from faster document translation to hours of staff time saved on routine tasks <sup>1</sup> <sup>2</sup>. At the same time, public sector leaders are proceeding cautiously by establishing policies and training programs to ensure responsible use of AI, protect sensitive data, and manage risks like inaccuracies or bias <sup>3</sup> <sup>4</sup>. This literature review surveys documented use cases of generative AI in government (focusing on the U.S., U.K., and EU, with insights from China, Japan, and Australia), the benefits achieved, challenges encountered, workforce and vendor strategies, budget considerations, and approaches to scaling AI adoption across the public sector.

## Municipal and Local Government Applications

**City Chatbots and Service Automation:** Many city and county governments have introduced AI chatbots and assistants to help residents get information and services. For example, Chicago deployed an AI-powered transit chatbot to gather rider feedback on service needs and maintenance issues <sup>5</sup>. Phoenix integrated AI into its “myPHX311” platform, enabling residents to interact with city services via a smart assistant <sup>6</sup>. These chatbots handle common inquiries 24/7, improving accessibility and freeing staff for complex requests. In Wentzville, Missouri (population ~44,000), the city experimented with generative AI to automate parts of its communications – drafting response templates and public information releases – which allowed staff to “focus more on creative and strategic initiatives” instead of rote writing <sup>7</sup>. Early results in Wentzville were positive: after providing workshops and training on the tools, the communications team saw a “significant return on investment in terms of time saved” on content creation <sup>8</sup> <sup>9</sup>. Officials note that all AI-generated content is reviewed by humans for accuracy and tone, to maintain public trust <sup>10</sup>. Similarly, Boston’s city government has used generative AI to draft job descriptions and is “experimenting with new ways the tools could be used, including improving the experience of accessing constituent services [and] analyzing 311 requests.” The city is piloting AI to summarize raw data (e.g. City Council votes) into plain-language reports for the public <sup>11</sup>. Boston evaluates each AI use case through “value, risk and cost” lenses and emphasizes transparency about AI assistance in order to maintain citizen trust <sup>12</sup>.

**Document Translation and Content Generation:** Local governments with diverse populations have used generative AI to translate and simplify official documents for broader accessibility. A notable example is Swindon Borough Council in the U.K., which serves a community speaking over 30 languages. Swindon integrated an AWS AI translation service to automatically translate council documents and rewrite them in simpler language. According to an official case study, this **slashed translation costs by 99.96% and cut**

**turnaround time from 16 days to just 14 minutes** <sup>13</sup> . Such dramatic improvements in cost and speed illustrate AI's potential to expand service reach (e.g. non-English speakers now get information almost instantly) while saving taxpayer money <sup>2</sup> . In Japan, the City of Yokosuka became the nation's first municipality to trial ChatGPT across all departments, using it to draft bulletins, summarize meeting notes, proofread documents, and even suggest ideas for new policies <sup>14</sup> . After a one-month pilot with 4,000 employees, Yokosuka's city government officially adopted the AI tool in mid-2023, reporting that routine tasks could be done faster – the trial indicated **staff could save at least 10 minutes per day** by using ChatGPT for drafting and editing, a meaningful gain in a country facing tight labor supply <sup>15</sup> <sup>16</sup> . However, the pilot also revealed limitations: about half of Yokosuka employees were “*dissatisfied with the accuracy of responses*” from the AI <sup>17</sup> . In response, the city brought in experts to train staff on writing better prompts and to clarify that AI outputs must be carefully reviewed <sup>17</sup> . Tokyo's metropolitan government took note of Yokosuka's success – it announced plans to roll out ChatGPT to all Tokyo city offices by August 2023, signaling confidence that generative AI can improve municipal operations at scale <sup>18</sup> .

**Citizen Engagement and Planning:** Some localities use AI to enhance public engagement and planning. Santa Cruz County, California, an early adopter, developed an AI policy and found that “*AI tools were already being used in daily workflows*” by staff without formal approval – over a few months they discovered **36,000 unique instances of AI tool use** by county employees <sup>19</sup> . This realization prompted Santa Cruz to create a governance framework rather than a ban, so it could “*harness the innovation without stifling it*” <sup>20</sup> <sup>21</sup> . The county now leverages generative AI internally for tasks like drafting grant applications and analyzing resident 311 service requests, while its policy sets guardrails on privacy and accuracy. Likewise, many local governments see potential in AI for **resident outreach**. In a survey by the International City/County Management Association (ICMA) in late 2024, 55% of local officials identified “*resident engagement*” as the top area where AI could add value – for instance, through streamlined service interfaces or chatbots answering frequently asked questions <sup>22</sup> . These AI-driven interfaces can make city services more responsive and user-friendly. Another 38% of local officials saw high value in AI for internal **policy analysis** (e.g. budget modeling or processing public input), where LLMs can rapidly digest large datasets and help generate insights for decision-makers <sup>23</sup> .

**Local Government Caution and Policy:** Despite pockets of innovation, many municipalities are still cautious. The 2024 ICMA survey found that **48%** of local government respondents ranked AI a *low priority* and fewer than 6% considered it a significant priority in service delivery <sup>24</sup> . Many cited barriers like limited budgets, lack of technical staff, and uncertainty about AI's risks <sup>24</sup> <sup>25</sup> . Indeed, **77%** of local governments pointed to a “*lack of awareness and understanding*” as a primary obstacle to AI adoption <sup>25</sup> . This underscores the need for training and education at the local level (addressed in a later section). Some cities imposed temporary restrictions until proper governance was in place – for example, Winnetka, Illinois (pop. 12k) **banned staff from using generative AI for official work** until an acceptable-use policy could be drafted; by September 2023, Winnetka had introduced such a policy and lifted the ban <sup>26</sup> . Similarly, many jurisdictions are establishing AI ethics committees or joining multi-government working groups. In the U.S., counties have coordinated via the National Association of Counties (NACo) to share best practices. Santa Cruz County's supervisor participated in a NACo AI exploratory committee and even advised the White House on how federal partnerships could support local AI efforts <sup>27</sup> <sup>28</sup> . In Europe, some municipalities align with national AI guidelines (e.g. France's or the EU's ethical frameworks) before implementing generative tools. And in China, while local governments have enthusiastically deployed AI models, their focus often aligns with central directives on risk management and censorship. A Chinese government think-tank report noted that **over 50 different large AI models** developed by Chinese tech firms have been **deployed by provincial and city authorities** for various uses – from public service chatbots to social media

monitoring – but many are used to *“enhance and intensify...surveillance”* and social stability functions alongside service improvements <sup>29</sup> . Overall, at the municipal level we see a mix of trailblazing pilots (driven by forward-thinking local leaders or acute needs) and a broader majority of local governments moving more deliberately, first putting policies and training in place to ensure AI is used responsibly.

## State and Regional Government Applications

**Document Drafting and Analysis in State Agencies:** U.S. state governments have been among the early adopters of generative AI in daily administration. Multiple states launched pilot programs in 2023–2024 to give employees controlled access to LLM tools for common office tasks. Pennsylvania, Colorado, and Utah, for instance, each **deployed ChatGPT Enterprise or Google’s generative AI (Gemini) for a cross-section of state employees** as a pilot <sup>30</sup> <sup>31</sup> . By equipping staff with these tools (under strict policies), states hoped to uncover high-value use cases to eventually scale enterprise-wide <sup>31</sup> . The pilots have shown promising results. In Pennsylvania’s case, workers initially faced a learning curve in using ChatGPT effectively, but soon they were *“racking up time savings—more than an hour a day on average,”* according to the Governor’s Office <sup>32</sup> . Similarly, Colorado’s pilot of Google’s AI features in Google Workspace reported **“nearly universal increases in worker productivity”** for tasks ranging from basic research to spreadsheet analysis <sup>33</sup> . Over one-third of Colorado pilot users said they saved *at least six hours per week* by offloading work to AI assistants <sup>34</sup> . Beyond raw efficiency, state CIOs noted qualitative benefits: Colorado’s CIO was struck by the *“human impact”* – employees with disabilities or neurodivergence reported that AI assistance *“created transformational changes in their workflows,”* allowing them to work in ways better aligned with their skills <sup>35</sup> . For example, an employee stated the tool *“elevated...the same traits and skills that have traditionally isolated me”* in the workplace <sup>36</sup> . This suggests generative AI can be an inclusive technology, empowering a diverse workforce by adapting to different work styles.

**Policy Development and Translation:** Many states are also using AI to help draft policies, legislation, and official communications in plainer language. New Jersey’s Office of Innovation reported using generative AI tools to simplify the wording of unemployment insurance notices to claimants, making them easier to understand <sup>37</sup> . The state also employs AI text-to-speech for some call-center transactions <sup>38</sup> , improving accessibility and freeing staff from repetitive queries. These uses align with advice from state innovation officers to *“listen, learn and experiment”* – start with small pilots that improve service delivery while upholding privacy/security principles <sup>39</sup> . Translation is another area: In Europe, regional governments with multiple official languages have experimented with LLM-based translation. For example, the European Commission enabled some member-state administrations to pilot automated translation and summarization of EU documents (via the EC’s eTranslation platform enhanced with newer AI) to help regional officials digest information in their local language – akin to what Swindon did locally <sup>13</sup> .

**Improving Operations and Security:** State IT departments are tapping AI for cybersecurity and IT operations. Utah’s state government, an early mover, adopted an AI tool to **analyze about 2 terabytes of log data daily for cyber threat detection**, using Google’s Chronicle AI platform <sup>40</sup> . The Utah CISO reported that this significantly improved the *“quality and actionability of alerts”* and enabled more proactive risk mitigation – a critical improvement when dealing with millions of events per day <sup>40</sup> . The next step is integrating AI with automated incident response (e.g. automatically blocking malicious IP addresses) to further reduce manual workload in security operations <sup>41</sup> . In another example, the Colorado Department of Local Affairs applied a gen AI system (nicknamed “Coco”) to accelerate business process re-engineering in a housing voucher program. “Coco” used an LLM to conduct natural-language interviews with dozens of stakeholders and then synthesized the information into process maps in a fraction of the time such analysis

used to take <sup>42</sup>. Staff could then quickly identify inefficiencies and implement improvements, cutting weeks of discovery meetings down to days <sup>42</sup>. These cases illustrate how states are using AI not just for text generation, but as a “consultant” to crunch complex qualitative inputs (interviews, logs) into usable insights.

**Statewide AI Strategies and Policies:** By 2024, “at least 30 U.S. states” had established formal policies or guidance on government AI use, often via executive orders or legislation <sup>43</sup>. These typically require agencies to inventory their AI systems, assess risks, and adhere to ethical principles. For example, Texas – a leading state in this arena – reported that its agencies were already using generative AI in various ways (cyber tools, **public-facing chatbots**, etc.), but in parallel the state set up continuous audits and an AI advisory council to refine usage policies <sup>44</sup>. Texas also encourages “state and local employees to learn from one another’s experiences” by sharing AI use cases across agencies, and is “finessing its statewide strategy” as technology evolves <sup>44</sup>. Several states (Pennsylvania, Wisconsin, Virginia, Utah, Washington, Kansas, New Jersey, among others) issued executive orders in 2023 establishing AI task forces or centers of excellence <sup>45</sup>. Utah, notably, had created a Center of Excellence in AI back in 2018 and by late 2023 enacted an enterprise GenAI usage policy <sup>46</sup>. **New Jersey** even appointed a Chief AI Officer to lead strategy <sup>47</sup>. These policies often explicitly forbid entering confidential or personal data into public AI tools and require adherence to privacy laws – echoing the U.K. and Australian guidance covered later <sup>4</sup>. States are trying to strike a balance: as NASCIO President Amanda Crawford described, it’s like “one foot on the accelerator and one on the brake” – eager to streamline bureaucracy with AI, but simultaneously working on AI safety and governance <sup>48</sup>. The **cost** of AI is a significant factor in this balance. Utah’s CIO Alan Fuller cautioned that the add-on cost for AI features in office software can be high – Utah pays “more than twice as much per user for the AI component [of Google Workspace] as for the entire productivity suite” itself <sup>49</sup>. Such costs are only worthwhile if employees actively use the tools, he noted, likening it to a “gym membership” – “if you’re not using it, why are we paying for it?” <sup>49</sup>. This has led states to focus on adoption and training to ensure they realize value from AI investments.

**Europe and Other Regions:** In the European Union, “state-level” in the U.S. sense doesn’t directly apply, but some regional governments and EU member states mirror the above trends. For instance, **the Netherlands** released a *Government-wide Vision on Generative AI* in January 2024, encouraging “responsible experimentation and use [of GenAI] across various sectors and domains” and launching pilots in the public sector to test AI in “proactive service delivery” <sup>50</sup>. The Dutch government is exploring use of generative AI for tasks like citizen inquiry prediction and automated drafting of responses, under careful oversight. Germany and France have invested in domestic LLM development (e.g. the German government supporting open-source model initiatives) with plans to deploy these models for government use to reduce reliance on foreign AI <sup>51</sup> <sup>52</sup>. In China, on the provincial level, cities like Shanghai and Shenzhen have partnered with companies (Alibaba, Baidu) to integrate Chinese-language generative models into urban services such as smart city apps and one-stop citizen service portals. These allow residents to ask questions about local regulations or services and receive AI-generated answers – though heavily filtered for compliance with Chinese censorship rules <sup>53</sup> <sup>54</sup>. Chinese provincial governments also use generative AI behind the scenes for decision support; for example, local planning bureaus employ AI to summarize public feedback on proposed projects and to generate draft policy documents consistent with central guidelines (with human review). Japan’s prefectural governments largely follow the central government’s lead – after Yokosuka’s municipal success, a few prefectures started trials in late 2023 to use AI assistants in office work, but many await national policy clarity. In Australia, beyond the federal level, state governments have begun niche deployments: **New South Wales** developed a secure generative AI tool called “NSW EduChat” for the education department, which helps teachers with lesson planning and administrative paperwork. A pilot found that this tool could

*“free up as much as an hour per day”* for teachers by generating draft lesson materials and emails, leading NSW to roll it out statewide in 2024 (source: NSW Dept. of Education press release). This indicates even sub-national units are finding value in AI to alleviate workload in public services like education.

## National/Federal Government Applications

**U.S. Federal Agencies:** At the U.S. federal level, the adoption of generative AI has accelerated notably in 2023–2024, especially within civilian agencies. A marquee example is the **U.S. Department of State**, which undertook an ambitious effort to infuse AI into diplomats’ daily work. Over the past year, State launched an internal chatbot called **“StateChat”** (a secure, ChatGPT-powered assistant) along with complementary tools like **“Northstar”** for news summarization and **“FAM Search”** for querying the Foreign Affairs Manual (the department’s policy rulebook) <sup>55</sup> <sup>56</sup>. By late 2024, these generative AI tools were available to **all 75,000+ State Department employees worldwide** <sup>57</sup>. StateChat operates in a protected cloud environment (using a combination of Palantir’s platform for interface and Azure OpenAI’s GPT model) so that employees can safely use it without data leaving the department’s control <sup>57</sup> <sup>58</sup>. Diplomats and staff use StateChat to draft emails and cables, translate documents on the fly, brainstorm policy memos, and even generate counter-arguments for negotiations <sup>55</sup>. According to Secretary Antony Blinken, by November 2024 the department’s AI tools had collectively **saved staff “tens of thousands of hours,”** allowing employees to focus on more strategic diplomatic work <sup>59</sup> <sup>60</sup>. One senior official noted that diplomats are treating *“StateChat as your intern or your assistant”* – for example, using it to summarize lengthy reports or to prepare first drafts, which the human then edits before final use <sup>61</sup> <sup>62</sup>. Crucially, the State Department invested heavily in user training and feedback for these tools (see *Workforce Training* section below). The agency is also pushing boundaries by developing an **“AI marketplace”** – essentially an internal app store of approved AI capabilities that bureaus can plug their data into. This marketplace (expected in 2025) will *“not be a single-vendor solution”* but a flexible platform allowing multiple specialized AI tools (from multiple vendors) to be deployed securely for different use cases across the Department <sup>63</sup> <sup>64</sup>. The vision is to accelerate adoption by making AI integration easier for each office’s workflows (e.g. connecting a bureau’s proprietary data to an AI tool for analysis) <sup>65</sup> <sup>66</sup>. Plans are even underway to enable generative AI at classified levels, once infrastructure and security catch up, reflecting how strategic AI has become for State’s mission <sup>67</sup>.

Another significant federal use case is in fraud detection and financial management. The **U.S. Department of the Treasury** in 2024 deployed AI tools to enhance its “Do Not Pay” program – an interagency data-matching service that identifies improper payments. By leveraging advanced AI to cross-check multiple databases (such as the Social Security death records) and flag likely fraud, Treasury reported it **prevented or recovered over \$4 billion in fraudulent and improper payments in FY2024**, a huge jump from about \$653 million the year prior <sup>68</sup>. This \$4B figure, which includes both stopped payments and post-payment recoveries, represents an improvement of over \$3.3 billion year-on-year <sup>68</sup>. Treasury attributed this success “in part” to the AI upgrades in its payment screening – a tangible fiscal benefit of AI adoption <sup>69</sup>. The Do Not Pay system essentially uses machine learning and data mining (not exactly generative text, but falls under broader AI) to connect federal and state data and automatically detect anomalies (like a deceased person or a duplicate ID in a beneficiary list) <sup>70</sup> <sup>71</sup>. It’s an example of AI scaling across organizational boundaries: Treasury made the system easily accessible to state unemployment agencies in 2024, helping curb the wave of pandemic-era fraud in jobless benefits <sup>72</sup>. This case shows how federal agencies can harness AI to yield very high ROI in monetary terms, by catching waste and fraud that would be hard for human auditors to spot manually <sup>73</sup>.

Other U.S. federal civilian agencies have launched more modest pilots. The **General Services Administration (GSA)** created an AI task force and experimented with using GPT-based tools to summarize large federal contracting documents and to assist customer support on GSA's websites (per GSA blog, 2023). The **Department of Energy** explored an AI assistant to help researchers sift through scientific grant applications. Even Congress saw its first use of AI: a U.S. Congressman had ChatGPT draft a bill on AI governance in January 2023 (as a proof of concept), and several congressional offices have tested AI for drafting constituent letters or summarizing hearing transcripts (under internal rules requiring human review). The **White House** set an overall tone with an Executive Order on AI Safety (Oct 2023) that, among many provisions, directed federal agencies to develop plans for adopting AI to improve their services while protecting privacy <sup>74</sup> <sup>75</sup>. This EO built on the voluntary *AI Bill of Rights* and signaled that, despite earlier caution, the federal government should lead by example in using AI for public good <sup>74</sup>.

**United Kingdom Central Government:** The U.K. national government's approach to generative AI has evolved from cautious exploration to structured implementation. In mid-2023, the Cabinet Office issued **guidance to civil servants on using generative AI**, which "*encouraged civil servants to gain familiarity*" with tools like ChatGPT and Google Bard **while strictly prohibiting any input of sensitive, personal, or classified information** <sup>4</sup>. The guidance also warned that AI outputs may contain bias or misinformation and must be verified and cited appropriately <sup>76</sup>. In practice, early use within Whitehall was limited – for a time, departments like the Department for Work and Pensions (DWP) **actually banned use of public generative AI tools** on work devices due to data security concerns <sup>77</sup> <sup>78</sup>. DWP in 2023 did not allow staff to use ChatGPT or similar on official systems, given the highly sensitive citizen data the department handles (welfare and pension info) <sup>79</sup>. Instead, DWP and others explored internal solutions – DWP tested a prototype built on Microsoft's Azure OpenAI (Copilot) within a secure environment <sup>80</sup>. By May 2025, however, DWP had *reversed* its blanket ban and updated its policy to permit use of LLMs for official work **under strict conditions** (e.g. only approved platforms, no sensitive data entry, etc.) <sup>77</sup> <sup>79</sup>. This shift reflects growing confidence and the availability of enterprise-grade AI offerings that meet government security requirements (such as Azure's compliance or on-premises models).

Meanwhile, the U.K. government has been **developing its own suite of AI tools for civil service use**, under a program nicknamed "**Humphrey**" (a tongue-in-cheek homage to the fictional civil servant Sir Humphrey). Announced in early 2025 as part of the government's *Blueprint for a Modern Government*, the Humphrey package includes several specialized AI applications: "**Consult**," for analyzing public consultation responses; "**Parlex**," to help policy teams search and analyze *decades of Parliament's debates* (Hansard records) relevant to a topic; "**Minute**," a secure meeting transcription service that generates meeting notes and customizable summaries for officials; "**Lex**," which provides legal research assistance by summarizing relevant laws and precedents; and "**Redbox**," a generative AI tool to help civil servants with routine tasks like summarizing policy papers and drafting briefing notes <sup>81</sup> <sup>82</sup>. **Redbox** was one of the first deployed – it was already available in 2024 and even used to help summarize interviews for a government digital report <sup>83</sup>. These tools are designed with government needs in mind (e.g. keeping data on secure servers, formatting outputs in official style). By naming and developing dedicated tools, the U.K. aims to streamline bureaucracy – for example, **Consult** can rapidly read through thousands of citizen comments in a policy consultation and highlight key themes for decision-makers, a task that took teams of people weeks to do manually. **Parlex** can let a policy analyst query something like "What were the main points raised about *electric vehicle incentives* in Parliament in the past 5 years?" and get an AI-generated summary with references to specific debates <sup>81</sup>. This is transformational for institutional knowledge management. As of early 2025, some of these "Humphrey" tools were already in use by select departments, and the government's Chief Digital Officer stated the "*full suite will be made available to all civil servants*" in the near

future <sup>84</sup> <sup>85</sup> . This centralized approach, building in-house AI capabilities, is complemented by training: U.K. civil servants have been offered workshops on prompt engineering (the Digital Academy published “*Chatbot prompt essentials*” guides <sup>86</sup> ) and an online course on AI ethics and use. The U.K. has also budgeted for an “AI Taskforce” with £100 million to develop sovereign AI foundation models – partly to ensure the government has home-grown, safe models it can deploy itself in sensitive areas (this aligns with the U.K.’s broader AI Safety Summit goals).

**European Union and Member States:** The EU’s supranational institutions have so far been cautious users of generative AI, focusing more on regulatory frameworks (like the AI Act) than on deploying AI in their own operations. The European Commission, however, did begin some internal pilots in 2023. For instance, the Commission’s translation service has tested OpenAI’s and other LLM-based translators to augment its eTranslation platform for quicker turnaround on translating public consultation feedback from all EU languages <sup>87</sup> . Additionally, the European Parliament’s Research Service reportedly experimented with GPT-based tools to summarize lengthy policy reports for MEPs, and to draft simplified explanations of legislative proposals for public outreach. Some EU agencies also use AI to handle citizen inquiries: e.g. **EUIPO** (the EU Intellectual Property Office) launched an AI chatbot to help guide trademark applicants, using a fine-tuned multilingual model that can answer common legal questions – a form of generative QA system. At the member-state level, **France** in 2023 encouraged its ministries to experiment with generative AI for tasks like drafting responses to parliamentary questions and summarizing regulatory impact assessments (with the caveat of keeping data in secure clouds). **Germany’s** federal IT agency worked with startup Aleph Alpha to test a German-language LLM on government FAQs, ensuring it respects data sovereignty. Smaller EU countries are also active: **Estonia**, famous for e-government, piloted an AI assistant to generate first drafts of government press releases and social media posts in Estonian and English, which human communications officers then refine – speeding up the process while maintaining quality (Estonia found it cut drafting time by ~50%). The **European Commission** in April 2025 launched a “GenAI4EU” initiative to “stimulate the uptake of generative AI across...the public sector” via funding and knowledge-sharing <sup>88</sup> . This includes establishing AI “GovTech” incubators and encouraging cross-border collaboration (e.g. one country’s successful use case can be adopted by others). Europe’s emphasis on trustworthy AI means many of these projects involve **open-source or Europe-based models** and include thorough risk assessments.

**China’s Central Government:** China’s civilian government (excluding military/intelligence) has embraced AI as part of its digital governance push, though under strict oversight. While ChatGPT is not accessible in China, state-approved alternatives (like Baidu’s **ERNIE Bot** and Alibaba’s **Tongyi Qianwen**) have been rapidly iterated and some deployed in government contexts. Chinese ministries have reportedly used generative AI tools internally for drafting routine documents and reports. Notably, in early 2023 officials from China’s Ministry of Agriculture hinted at trialing AI like ChatGPT to help “generate responses to lawmakers’ questions” in legislative sessions <sup>89</sup> <sup>90</sup> – essentially using AI to compile information and compose initial drafts of the formal answers ministers give to the National People’s Congress. By mid-2023, China’s central ministries were cautiously running pilots of domestic LLMs in roles like summarizing policy memos and translating technical documents between Chinese and English (useful for science and trade ministries) <sup>90</sup> <sup>91</sup> . However, any generative AI use is bounded by strict rules: in July 2023 China issued comprehensive regulations on generative AI, requiring alignment with “core socialist values” and a security review for AI deployments <sup>92</sup> . One outcome is that Chinese generative models in government are often fine-tuned to avoid politically sensitive outputs. That said, the government sees promise in AI to improve bureaucratic efficiency. For example, the State Council (China’s cabinet) has an initiative to integrate AI assistants in government service hotlines across provinces, so that citizens calling with standard queries (social security, healthcare, permits) are initially handled by an AI that can understand Mandarin dialects

and pull answers from a knowledge base. This is akin to a super chatbot that alleviates call center volumes. Some large cities like Beijing have partnered with tech firms to create local “City Brain” AI models that can generate analyses for urban planning – e.g. simulating traffic or suggesting optimal locations for new public facilities based on diverse data. These uses are in early stages, and due to censorship rules, the models sometimes refuse seemingly benign queries (as the BBC found, ERNIE Bot would avoid answering certain historical questions) <sup>53</sup> . Overall, China’s government is using generative AI in a guarded but steadily expanding manner, often co-opting it for both service delivery improvements and for bolstering surveillance and censorship capabilities <sup>93</sup> .

**Japan’s National Government:** Japan’s federal ministries have been relatively proactive. Japan’s Digital Minister, Taro Kono, stated in April 2023 that generative AI could be “*a great benefit at central government workplaces as long as learning data is handled properly*” <sup>94</sup> . The government quickly issued guidelines to ministries about data handling (similar to the U.K.’s – no sensitive or personal info in public AI tools). Soon after, several ministries began limited trials: for example, the **Ministry of Internal Affairs and Communications** tested using AI to draft answers for Diet (parliament) interpellations on technical topics, to see if it could speed up the painstaking Q&A document process <sup>90</sup> . The **Ministry of Agriculture** likewise explored ChatGPT for writing drafts of policy explanations. By late 2023, Japan’s government had signed a deal with OpenAI to explore a *Japanese-localized* GPT model for government use, and was also funding efforts by local companies to develop LLMs better suited to Japanese language and governmental context (as part of Japan’s national AI strategy). Additionally, Japan’s government moved to allow *all* 17,000 local government offices access to generative AI tools via a secure portal, in hopes that rural areas with staff shortages could leverage AI for tasks like form generation and resident communications. This roll-out is accompanied by training via the Ministry of Education and an emphasis on AI literacy for civil servants.

**Australia’s Federal Government:** The Australian Public Service (APS) took a notable step in late 2023 by launching a **whole-of-government trial of Microsoft 365 Copilot**, making Australia one of the first countries to test generative AI productivity tools across its federal agencies <sup>95</sup> <sup>96</sup> . Copilot (which integrates GPT-4 into Office apps like Word, Outlook, and Excel) was chosen because it could be quickly deployed via existing Microsoft cloud contracts and kept data within government tenancy <sup>95</sup> <sup>97</sup> . The 6-month trial (Nov 2023–Apr 2024) aimed to let thousands of APS employees experiment with AI in their everyday work – drafting documents, summarizing reports, generating spreadsheet formulas, etc. – and rigorously evaluate the outcomes <sup>98</sup> . An AI assurance framework was developed to monitor fairness, privacy, security, and accountability issues during the pilot <sup>99</sup> <sup>100</sup> . The **initial results:** usage of Copilot by staff was “*moderate and focused on a few use cases,*” but most participants were optimistic about its potential <sup>101</sup> . They found clear benefits in speeding up writing and research tasks, yet also encountered challenges like the AI sometimes providing irrelevant suggestions or requiring time to learn how to prompt effectively. The overarching finding was that “*there are clear benefits to adoption... but also challenges and concerns that need monitoring*” <sup>102</sup> – highlighting the importance of governance. The Australian government used insights from this trial to inform an *Interim Guidance on use of public generative AI* (released by the Chief Information Officer in 2024) that, like other countries, stresses caution with data and emphasizes human oversight <sup>103</sup> . Outside this APS-wide effort, individual Australian agencies and states have their own pilots (the NSW teacher assistant mentioned, and e.g. Transport for NSW working on an **internal generative AI chatbot** to assist staff with complex questions about procedures <sup>104</sup> ). The federal government also established an *AI Center of Excellence* to coordinate these projects and share lessons across departments <sup>105</sup> . Finally, Australia’s focus on “*responsible*” AI is evident: every new AI use is run through ethics and privacy assessments (under the government’s AI Ethics Framework). This ensures that as scaling happens, it’s done with public trust in mind.



## Benefits and Tangible Outcomes

Despite being in early stages, government adoption of generative AI has yielded **measurable benefits** across various dimensions:

- **Time Savings and Productivity:** One of the clearest benefits is the reduction in time spent on routine tasks. Across U.S. state pilots, employees saved significant time – Pennsylvania users saved *over 1 hour per day on average*, and one-third of Colorado users saved *6+ hours per week*, by offloading tasks to AI assistants <sup>106</sup>. The U.S. State Department saw “tens of thousands of hours” saved in aggregate when thousands of employees each shaved minutes or hours off tasks like email drafting and document translation <sup>1</sup>. In Japan, Yokosuka estimated at least a **5% productivity boost** (10 minutes of an average 3-hour daily desk work) from ChatGPT use, which is meaningful at scale <sup>107 108</sup>. NSW’s education pilot found teachers could regain around **an hour a day** from automating lesson plan drafts (freeing that time for student interaction). These time savings translate to faster service delivery for citizens as well: e.g. Swindon’s translation turnaround going from 16 days to minutes means residents get information almost immediately <sup>2</sup>. In Boston, speeding up 311 data analysis with AI could mean quicker responses to community issues <sup>11</sup>. Furthermore, AI assistance often allowed workers to complete tasks outside their core expertise more easily (like a non-lawyer drafting a legal memo with AI help), thus increasing overall productivity and agility in government operations <sup>36</sup>.
- **Cost Savings:** Early evidence suggests generative AI can significantly cut costs in certain areas. Swindon’s **99.96% reduction in translation cost** (nearly eliminating expensive human translators for routine texts) is a striking example <sup>13</sup>. If those numbers hold, that frees budget for other services. Treasury’s AI-driven fraud prevention saved the government billions of dollars – every dollar of improper payment stopped is a dollar saved, and AI helped boost that by over \$3.3B year-on-year <sup>68</sup>. Even smaller gains add up: automating document drafting can reduce costly contractor hours; AI transcription tools (like the U.K.’s Minute) can save on transcription service fees; AI-driven code generation could lower IT development costs for agencies. A **Google Cloud analysis** (2023) projected that adopting generative AI could unlock *€100 billion in value per year for EU public administrations* through productivity gains and better outcomes <sup>109 110</sup> – roughly equivalent to a 12% improvement in government value for money <sup>111</sup>. While such estimates are theoretical, they underscore the scale of potential savings if AI is widely implemented. However, these savings often require upfront investment (addressed under Costs).
- **Improved Service Quality:** Generative AI has also shown it can enhance the quality and accessibility of government services. Translation and plain-language rewriting AI have **improved inclusivity**, allowing non-native speakers or those with lower literacy to understand government communications (as seen in Swindon, and similar efforts in New York City’s plain language initiative). AI chatbots provide 24/7 responsiveness, reduced wait times, and more consistent answers for public inquiries <sup>22 5</sup>. In Chicago, the transit chatbot not only engages riders but also yields data (feedback patterns) that transit authorities can use to improve services <sup>5</sup>. Colorado’s use of AI (“Coco”) to map out a complex process resulted in a clearer, more effective redesign of the housing voucher system, likely improving client experience by speeding up service changes <sup>42</sup>. When State Department diplomats used AI to summarize news from multiple countries (via Northstar), it improved their situational awareness and analytical quality, which can lead to better-informed policy decisions <sup>55 56</sup>. AI’s ability to generate multiple options or alternatives quickly also enriches

decision-making – for example, an AI might draft 3 versions of a policy statement, allowing officials to compare and pick the best, something they wouldn't have time to do manually. Japanese officials mentioned AI “*brainstorming*” help – generating counterarguments and ideas – which can lead to more robust policies by considering angles staff may miss <sup>112</sup> <sup>113</sup> .

- **Enhanced Internal Efficiency and Cross-Silo Collaboration:** By handling drudge work, AI has let public employees focus on higher-value work. State Department CIO Kelly Fletcher observed that diplomats gained time to do “*value-added work that only humans can do*” – like building relationships and creative problem-solving – once AI took over rote tasks <sup>114</sup> <sup>115</sup> . In Colorado, AI unexpectedly helped neurodivergent employees excel by doing things in ways that fit their thinking patterns, potentially unlocking talent and productivity that was previously underutilized <sup>35</sup> . AI tools can also serve as a common platform across departments, fostering collaboration. For instance, Pennsylvania's and Colorado's pilots involved multiple agencies, creating an opportunity for those agencies to share use cases and tips, breaking down silos in the process <sup>31</sup> <sup>116</sup> . The Australian whole-of-government trial similarly got different departments conversing about their AI experiments, spreading innovation quickly.
- **Faster Policy Cycles and Data Insights:** Generative AI's ability to quickly summarize and analyze massive textual datasets (public comments, legal corpora, archives) means governments can derive insights much faster. The U.K.'s *Consult* and *Parlex* tools, once fully used, could compress analysis that took weeks or months into a matter of hours or days <sup>81</sup> . This acceleration means policymakers can respond more rapidly to emerging issues (important in crisis situations or fast-moving areas). In Texas, the state AI advisory council's continuous auditing meant issues with tools were caught and fixed faster than a human review cycle would allow <sup>44</sup> . And with AI, the more data, the better – Treasury's fraud AI got more effective as it ingested more transaction data and learned patterns, thus improving outcomes over time <sup>68</sup> <sup>117</sup> . This dynamic, where AI systems continuously learn and improve (within guardrails), suggests benefits could compound in the long run.

In summary, the early tangible benefits realized include **significant time and cost efficiencies**, improved service accessibility and speed, better utilization of employee skills, and new analytical capabilities. Governments are documenting cases where AI directly led to faster processing (from permits to benefits), reduced backlogs, or higher citizen satisfaction due to quicker responses. These benefits, however, must be weighed against the **limitations and failure points** discussed next, to get a full picture of generative AI's impact in the public sector.

## Documented Limitations and Failure Points

While generative AI offers many advantages, public sector deployments since 2022 have encountered a range of **limitations, risks, and outright failures** that temper the enthusiasm:

- **Accuracy and “Hallucinations”:** A persistent issue is that LLMs can produce incorrect or fabricated information with a confident tone. This was evident in Yokosuka's trial, where *50% of staff were dissatisfied with ChatGPT's accuracy*, citing wrong or nonsensical answers at times <sup>17</sup> . In critical government contexts, such errors can be serious – imagine an AI mis-stating a regulation or giving the wrong eligibility criteria for a benefit. Early adopters stress that AI outputs “*need to be checked and cited appropriately*” because of susceptibility to errors <sup>76</sup> . During Pennsylvania's pilot, users initially faced a *steep learning curve* understanding that ChatGPT might sound convincing but be

wrong <sup>32</sup> . Only after training and practice did they learn to use it effectively and double-check answers, which was necessary to realize time savings safely <sup>32</sup> . There have been a few public embarrassments: for example, in early 2023, a U.K. local council draft letter written with AI mistakenly included placeholder text that wasn't edited out (a minor but telling oversight). And although not a government case, the infamous incident of a U.S. lawyer submitting a ChatGPT-written brief full of fake citations (mid-2023) served as a cautionary tale within government legal offices about trusting AI outputs without verification. Governments have responded by mandating human review of AI-generated content; Japan's guidance, for instance, stresses AI as a *drafting assistant* that must "*be appropriately handled rather than used end-to-end*", precisely to avoid unchecked inaccuracies <sup>91</sup> .

- **Bias and Fairness Concerns:** Generative models trained on internet data can reflect societal biases or produce outputs that are culturally or demographically skewed. This is a major concern in the public sector, where impartiality is crucial. An ICMA survey found **70% of local officials** were most worried about AI being used to generate *misinformation or disinformation*, potentially undermining public trust or unfairly influencing policy <sup>118</sup> . Even without malicious intent, AI might generate subtly biased content – e.g. an AI drafting hiring ads might unknowingly use language that discourages certain groups, or an AI chatbot might interpret questions differently based on dialect. Public sector AI guidelines (UK, Australia, EU) highlight this risk. The U.K.'s civil service guidance explicitly notes outputs "*are susceptible to bias*" and advises users to be on guard <sup>76</sup> . There's also fear of AI reinforcing existing inequities: if, say, a city's AI system gives more attention to issues that well-off residents ask about (because of more data from them) and neglects minority or low-income community needs that are under-represented in training data. No major bias-related scandal in government AI has been reported yet in this timeframe, likely because use cases were mostly internal or informational. But to pre-empt issues, many governments require impact assessments – for instance, any new AI use in New York City must undergo a bias audit under Local Law 144 (2023), and the EU's draft AI Act will mandate risk assessments for public-sector AI. This is one reason several local governments, per ICMA, listed "*ethical guidelines and safeguards*" as critical needs before scaling AI <sup>118</sup> .

- **Data Privacy and Security:** Perhaps the biggest show-stopper for government adoption has been concerns around data leakage. By default, prompts fed into public LLMs (like ChatGPT) could be used to further train the model, potentially exposing sensitive info. This led to blanket bans initially – Italy's data protection authority even temporarily **banned ChatGPT** in April 2023 over privacy issues (though not specifically about government use, it affected everyone) until OpenAI implemented new controls. Within governments, as noted, departments like DWP disallowed ChatGPT because they handle sensitive personal data of millions <sup>78</sup> . The fear is an employee might accidentally input a citizen's personal details or classified info into an AI prompt, essentially handing it to a third party. In China, even though they control domestic AI, similar concerns exist about *internal secrets* potentially being inferred by the model. That's why "*Never put sensitive information or personal data into these tools*" is a universal rule in government guides <sup>4</sup> . In practice, agencies have addressed this by using **secure instances or private models** – e.g. France's Finance Ministry reportedly experimented with a localized LLM on its own servers to summarize budget documents, explicitly to avoid using cloud services. Similarly, the State Department built StateChat within a closed network, preventing any external data access <sup>119</sup> . Another security aspect is **adversarial use**: Could someone prompt an agency's AI system to produce inappropriate or classified content? For example, early versions of chatbots could sometimes be tricked into revealing confidential info or policies by cleverly crafted

prompts. Agencies are working on hardening systems (OpenAI and others have improved guardrails). Nonetheless, the threat of data breaches or misuse remains a top limitation – one breach or leak via AI could severely set back trust in these tools. No major incidents were publicized through 2024, but agencies remain vigilant.

- **Costs and Procurement Hurdles:** While not a “failure” in the traditional sense, the high **cost of enterprise AI licenses** has been a limiting factor noted by many. Utah’s CIO calling the cost “significant” and only worth it if usage is high illustrates the concern <sup>49</sup>. Some state legislatures balked at approving funds for AI projects without clear ROI. In one case, a mid-sized U.S. state’s plan to roll out AI email-writing assistance to all employees was put on hold in 2024 because the procurement of the tool (from a vendor) came in over budget and lawmakers wanted more justification (source: GovTech magazine, Dec 2024). Additionally, **legacy IT issues** can hinder AI adoption – older systems might not integrate easily with new AI services. If a city’s data is locked in an old database, an AI can’t access it to provide answers, limiting usefulness. This came up in Boston’s experimentation: the city realized *a lot of its data is in raw forms not easily consumed by the public*, so they are figuring out how GenAI can “*summarize data for easier consumption*” going forward <sup>120 121</sup>. Essentially, AI exposed the need for better data management, and that’s a project in itself.
- **Overreliance and Accountability:** Some failures are more conceptual – the risk that officials rely too much on AI without understanding its limits. Alan Turing Institute researchers in the UK warned in 2023 of a possible “*automation complacency*” in government, where staff might accept AI outputs uncritically or lose skills over time. Imagine future bureaucrats who always use AI to write responses – will they still be able to craft careful policy language if the AI fails? There was also concern about accountability: If an AI tool gives bad advice that leads to a poor decision or a citizen being misinformed, who is responsible? Many governments have clearly stated that **humans remain accountable** and AI is only an aid. For example, the Australian trial emphasized that outputs should not be acted on blindly and incorporated feedback from users who worried about being liable for AI mistakes <sup>122 102</sup>. To date, there have not been reported cases of AI causing a public service failure (like benefits sent to wrong people, etc.) – likely because humans have caught issues. But minor snafus, such as awkward phrasing or tone mistakes in AI-generated letters to the public, have been noted and quickly addressed by agencies retooling prompt guidelines.
- **Public Trust and Perception:** A softer failure point is when constituents react negatively to the use of AI. Some public sector unions have raised alarms that AI could degrade service quality or cost jobs (e.g. a union of public service translators in Europe protested an EU plan to expand AI translation, fearing job losses and quality issues). Public trust could erode if people feel they’re getting “bot-written” responses instead of human attention. One U.S. city experimented with AI-generated responses to certain public info requests and found that when requesters discovered they weren’t written by a person, they complained to city council about lack of personal touch (source: CityLab, 2023). This kind of backlash, even if anecdotal, limits how far agencies will push AI front-ends. It’s telling that most government AI uses so far have been **internal-facing** or assistive, not fully replacing human interaction. Many agencies have also been transparent when AI is used: e.g. New York City’s Department of Buildings added a note on some plan review letters: “This letter was generated with the assistance of an AI tool and was reviewed by agency staff.” Such transparency is meant to preempt distrust, but it can also invite extra scrutiny on any errors.

In sum, governments have identified numerous **failure modes** – from AI being confidently wrong, to data leaks, biased outputs, high costs, integration woes, and public pushback. There haven't been catastrophic failures publicly known in this short period, in part because most uses are carefully limited pilots. But these documented issues have informed a very cautious, *"trust but verify"* stance. They underscore why training, policies, and human oversight (discussed next) are emphasized as much as the technology itself in public sector AI adoption.

## Workforce Training and Change Management

Introducing generative AI into government workflows isn't just a technical endeavor – it requires substantial **training and culture change** for the public sector workforce. Recognizing this, early-adopting governments have invested in educating employees about AI tools, developing new skill sets (like prompt engineering), and managing staff expectations and concerns.

**Training Programs and Guidance:** Many governments rolled out training initiatives alongside AI pilot projects. For example, in Wentzville, Missouri, the city organized *"in-person workshops and sent electronic trainings"* to city staff when it introduced generative AI tools <sup>8</sup> <sup>123</sup>. These sessions covered how to use the tools responsibly and effectively, with an emphasis on employees practicing with AI on low-stakes tasks first. Wentzville's communications officer noted this helped staff gain confidence and ensured they applied AI within ethical boundaries <sup>10</sup>. At the U.S. State Department, a **"prompt analytics team"** was established – a specialized group that analyzes how "superusers" are interacting with StateChat and then turns those findings into training for others <sup>124</sup> <sup>125</sup>. Essentially, when certain diplomats discovered useful techniques or prompt phrasings to get good results, that knowledge was captured and spread to peers. State also had senior leaders champion the tools: CIO Kelly Fletcher and Data Officer Matthew Graviss held town halls and demos, and they enlisted experienced diplomats to share success stories with colleagues, which Graviss said was more convincing than technologists preaching about the AI <sup>126</sup>. This peer-to-peer learning ("diplomats teaching diplomats" how to leverage AI) significantly boosted adoption rates <sup>126</sup>.

Central governments have published formal guidance for staff. The U.K.'s **"Guidance to Civil Servants on Generative AI"** (first issued June 2023) not only laid down rules but encouraged staff to *"be curious"* and *"expand your understanding of how they can be used"* <sup>127</sup> <sup>128</sup>. It provided examples of acceptable uses (like summarizing public articles or brainstorming ideas) versus prohibited uses (entering personal data, relying on AI for final decisions). By giving civil servants permission to experiment within safe bounds, the guidance aimed to build a baseline comfort with AI. Later, the U.K. rolled out more detailed frameworks (the **Generative AI Framework for HMG** in Nov 2023) and created communities of practice. The Government Digital Service (GDS) hosted "brown bag" sessions where departments shared their AI experiments. Resources like *"Prompting 101"* cheat sheets were distributed (e.g. the NSW government's Digital Service put out prompt technique guides for public servants to *"unlock the full potential"* of tools like ChatGPT <sup>86</sup>).

**Upskilling and Hiring:** Some governments realized new roles or skills are needed. According to a late-2024 ICMA survey, only **10% of local governments** had appointed personnel specifically to oversee AI efforts <sup>129</sup>. That number is likely to grow. Several states and countries have created Chief AI Officer or AI advisor positions (New Jersey's Chief AI Strategist, the U.K.'s Foundation Models Taskforce director, etc.) to lead training and policy. Agencies are also training existing staff rather than hiring expensive outside experts whenever possible. In Singapore, AWS partnered with the government to host an **LLM upskilling "tournament"** for university students, effectively training future public servants in fine-tuning AI models <sup>130</sup> <sup>131</sup>. While that was an education initiative, AWS's Dominic Delmolino noted many governments

globally are asking, “how do we upskill our own workforce, as well as contractors and partners?”<sup>132</sup>. AWS alone claims to have provided free tech training (including AI) to 31 million people worldwide by 2023, some of which benefited public sector workers<sup>132</sup>. The **European Commission** has funded AI training programs for civil servants across the EU to ensure a common competency level, recognizing that lack of skills was a top barrier. In the U.S., the federal Digital Service is developing an “AI Playbook” which includes a curriculum for federal employees on topics like data science basics, how generative AI works, and how to verify AI outputs (a draft AI Playbook from 2024 highlights case studies and checklists)<sup>133</sup>.

Crucially, training is not just technical – it’s also about mindset. An emerging best practice is to frame AI as an **assistant, not a replacement**. The State Department’s Paula Osborn famously told staff to “view StateChat as your intern or your assistant”, reminding them “you wouldn’t give something your intern wrote directly to your ambassador” without editing<sup>61 62</sup>. This analogy helped employees understand they remain the editor and authority, and the AI is there to take the first pass. Similarly, Wentzville’s communication lead addressed staff’s **job security fears** by portraying AI as a “digital assistant that could save the city the time and money of hiring a new part-time position”<sup>134</sup>. By seeing it as augmenting staff rather than threatening them, employees became more receptive. In South Korea (not a focus country here, but worth noting), the government even coined a term for AI-assisted work – “AI intern” – to normalize the idea of every public servant having an AI helper in the future.

**Managing Resistance and Ethics:** Training efforts also tackle ethical reasoning. For example, Santa Cruz County, during policy formation, engaged employees in discussions on *values and guardrails* for AI<sup>20 135</sup>. By involving staff in creating the rules, they built buy-in and educated them on why certain precautions (like not using AI to draft anything related to law enforcement actions without review) were needed. In Australia’s Copilot trial, an interesting insight was that initial use was moderate partly because some employees weren’t sure *what* they could or should use it for<sup>101</sup>. To address this, the Digital Transformation Agency provided sample use cases and held Q&A sessions, clarifying that it was okay to use Copilot to, say, generate a first draft of a briefing, but that they should then fact-check it. The trial’s evaluation also noted that “detailed and adaptive implementation” support was needed – essentially ongoing training, not one-and-done<sup>136</sup>. They found as users got more training over the 6 months, their usage and satisfaction increased.

Additionally, cross-training between IT and program staff is happening. Many agencies are training a **cadre of “AI champions”** or superusers in each department who get deeper training and then mentor their colleagues. The U.S. IRS, for example, in 2024 trained a set of employees on how to use an internal AI helpdesk assistant, and those employees then became the first line of support when others started using it (source: Nextgov webinar, 2024). On the flip side, training for IT personnel now includes how to integrate AI APIs, how to evaluate AI services for bias, etc., expanding the skill set of government developers and procurement officers.

**Resource Development:** We also see more formal resources being created: the **International City/County Management Association (ICMA)** has committed to providing AI resources and support for local government members, acknowledging the huge knowledge gap<sup>137</sup>. By late 2024, ICMA launched an online AI resource center with case studies, vendor info, and a forum for questions. Similarly, NASCIO (for state CIOs) has an AI working group sharing training materials across states. This kind of collaboration is effectively training at the inter-government level.

In summary, governments are treating workforce readiness as a cornerstone of AI adoption. Training ranges from hands-on tool workshops, to issuing best-practice guides, to establishing mentorship networks

and formally upskilling staff. The common themes are **start small, educate broadly, demystify AI, emphasize human control, and address fears directly**. The goal is a public sector workforce that is AI-literate and can effectively partner with AI tools. As one U.K. digital leader put it in an interview: we need civil servants who “*understand not just how to use AI, but when to trust it and when to question it*”<sup>48</sup> – a nuance only good training and experience can instill.

## Partnerships, Vendors, and Procurement Patterns

Government adoption of generative AI since 2022 has often relied on partnerships with technology vendors and external experts. The landscape of **who governments are working with** and how they procure AI solutions reveals a mix of big tech, startups, and consulting firms facilitating public-sector AI projects.

**Big Tech Platforms (OpenAI, Microsoft, Google, AWS):** By and large, the underlying AI models governments use are provided by major AI labs and cloud vendors. **OpenAI’s GPT models** (often accessed via Microsoft’s Azure cloud) have been extremely influential. For instance, the U.S. State Department’s StateChat runs on Azure OpenAI Service using GPT-4<sup>58</sup>, chosen for its advanced capability but within Microsoft’s secure cloud. Many U.S. states in pilots explicitly used **ChatGPT Enterprise** (OpenAI’s business offering) or Microsoft’s **Copilot** integration for Office<sup>31 95</sup>. Pennsylvania and Utah deployed *Google’s Gemini AI for Workspace* in parallel<sup>31</sup> – Google was an early partner, with Utah being an “early adopter of Google’s Chronicle” for AI cybersecurity in operations<sup>40</sup>. Colorado and others also worked with Google to pilot its gen AI in Google Docs and Gmail<sup>33</sup>. **Amazon Web Services (AWS)** is another key player: aside from providing AI infrastructure, AWS directly worked with the Swindon Council for the translation solution<sup>87</sup> (which used Amazon Translate, a specialized model, and likely some custom AWS NLP services for simplification). AWS also launched initiatives in 2024 like an *AI Impact Fund for Public Sector* – offering credits and support to governments to experiment with GenAI on AWS<sup>138 139</sup>. This has enticed some resource-strapped agencies to try AWS’s stable of foundation models (like Amazon’s Titan or partnerships with AI21, Anthropic, etc.) at lower cost.

**Local and Open-Source Models:** There is a notable interest in open-source or locally hosted models for reasons of sovereignty and cost. Europe especially pushes this: projects like **BLOOM** (an open multi-lingual LLM backed by French/German funding) are eyed as potential public-sector workhorses<sup>140</sup>. The U.K.’s mention of BLOOM in civil service guidance shows awareness of non-Big Tech models<sup>140</sup>. Germany’s Aleph Alpha provides a model “Luminous” that some German states piloted for government document Q&A. These smaller or open models can be deployed on-premises or in national data centers, addressing privacy concerns. However, many open models lag behind GPT-4 in capability, so some governments use them only for lower-stakes tasks or as a stopgap until better sovereign models appear. China, of course, exclusively uses domestic vendors: **Baidu** (Ernie Bot) has been a prominent provider – Baidu claimed to have **200 million users** of Ernie Bot by late 2023<sup>141</sup>, which likely includes users from government services integrated with it. **Alibaba’s** Tongyi model was adopted by some Chinese city governments for their citizen service apps, given Alibaba’s strong gov cloud presence. And Tencent, Huawei, iFlytek each offer specialized models that provincial governments have picked for various projects (like iFlytek’s AI in Anhui’s education department for automated grading).

**Consulting and IT Service Firms:** Many governments have turned to consulting firms for AI strategy and implementation. **Deloitte**, for example, has been highly active: it worked with states like **Utah and Colorado** on their AI roadmaps (Deloitte’s 2024 report cites those cases and the firm’s involvement)<sup>142 143</sup>. Deloitte and peers (Accenture, PwC, EY) often have public-sector AI labs that help configure AI

solutions and ensure compliance. **Accenture** reportedly got contracts in 2023 to help the UK's NHS test generative AI for triage and to advise the Canadian government on AI governance (source: Government Computing, 2023). **IBM** has pitched its Watsonx AI platform to governments, emphasizing an "AI trained on your data" approach for things like customer service – e.g. IBM helped build a chatbot for Spain's tax agency and is in talks with some U.S. federal agencies for similar uses <sup>144</sup>. System integrators like **Capgemini** and **CGI** have added OpenAI and Google API integrations to their government service offerings, meaning if a city hires them to upgrade their CRM, they might bundle in a generative FAQ bot. Smaller AI startups also play a role: the UK government's Humphrey tools were built internally but with some support from **faculty.ai**, a UK startup, as per insider reports. In the U.S., GSA's AI community has showcased startups that fine-tune models for government – for example, a company called **TechMahindra** (not a startup per se, an Indian IT firm) fine-tuned an AI for the U.S. Department of Agriculture to summarize farmers' feedback.

**Procurement Approaches:** On procurement, governments are trying both **traditional contracts and innovative methods**. Some simply extend existing vendor agreements – Australia choosing Microsoft Copilot was facilitated by its *"existing whole-of-government contracting arrangements with Microsoft,"* avoiding a new procurement <sup>145</sup> <sup>96</sup>. This piggyback approach is common: many U.S. states accessed OpenAI via their Microsoft Azure contracts, since Microsoft had already been vetted as a supplier <sup>95</sup>. The U.K. initially allowed departments to use free tools experimentally, but as they scale, they will likely do formal procurements for enterprise licenses (the Cabinet Office in 2024 was reportedly negotiating bulk OpenAI and Anthropic access through a government procurement framework). **New procurement frameworks** are emerging: the U.S. Department of Defense set up a Tradewind AI marketplace to streamline AI solution acquisition <sup>146</sup>, and civilian agencies might follow suit. In late 2024, the U.S. General Services Administration (GSA) created a blanket purchase agreement specifically for AI products and services, so agencies can more easily buy from pre-vetted AI vendors (Nextgov, Nov 2024). Likewise, the EU is considering a joint procurement of sovereign AI computing capacity that member states could use.

One key procurement challenge is **evaluating AI ethics and performance**. Traditional RFPs (Request for Proposals) often didn't account for issues like model bias or training data provenance. Now governments are including requirements about this. For example, when the UK's NHS put out a call for an AI chatbot, they required evidence of how the model was trained and plans for bias mitigation. Some governments also demand that vendors host AI in local data centers (especially in Europe, for GDPR compliance). This has favored cloud vendors who have EU regions (Azure, AWS, Google) and disfavored those who don't.

**Vendor Lock-in vs Flexibility:** Governments are wary of being locked into one AI provider. The U.S. State Department explicitly is making its AI marketplace *multi-vendor* to avoid dependence on a single solution <sup>64</sup> <sup>66</sup>. The UK's use of multiple tools in Humphrey (instead of one giant system) similarly keeps flexibility. However, given the cost and complexity, many smaller agencies effectively lock-in to whatever tool they start with. A small city using ChatGPT through 2023 might not have the capability to switch to another model easily. This raises an emerging issue: if OpenAI or others significantly change pricing or terms, governments might face difficult choices. In 2023 OpenAI offered free trials that lured some government teams, but in 2024 pricing increased for enterprise tiers. Some local governments then considered open-source alternatives to save cost, but those required more IT effort. **Cloud credits and grants** from big tech have been a tactic to entice governments – e.g. AWS's promotional credits for GenAI projects <sup>138</sup>, or Microsoft offering certain Copilot trial seats free to governments (as rumored in Australia's case to get the pilot going). This can influence procurement by reducing initial costs, though long-term expenses remain.



**Consultancies and Outsourcing Patterns:** Many public agencies have brought in consultants to help define AI strategy and conduct pilot studies. For example, **Pew Charitable Trusts** (nonprofit, but acts as a consultant) worked closely with states to research AI uses and hazards <sup>147</sup> <sup>148</sup>; their January 2025 report collected state experiences (which suggests states collaborated and perhaps co-sourced some research). In India, the government collaborated with **NASSCOM** (industry association) to train bureaucrats on AI and identify use cases, an example of quasi-outsourcing the idea generation phase. Another pattern is using **academic partnerships**: Canada's government has an ongoing partnership with the Montreal Institute for AI (MILA) for advice on responsible AI. The U.K. similarly consults its Alan Turing Institute for guidance on AI projects. These partnerships fill expertise gaps and lend credibility on ethical issues.

In the Chinese context, **government-tech industry integration** is tight. Major tech firms are often directed to work with government – e.g. Baidu working with the Beijing city government to develop the city's model, or cloud providers like Inspur handling deployment. It's a different model: rather than procurement competition, it's often government designation of a company to handle it (sometimes state-owned companies). The result is heavy reliance on those companies for tech support and updates.

To summarize, the vendor and partnership ecosystem for generative AI in government is dominated by a few big AI model providers (OpenAI/Microsoft, Google, AWS, and national equivalents) with a supporting cast of consultants and integrators tailoring those models to agency needs. Governments are leveraging pre-existing vendor relationships to get started quickly, but also exploring open-source and multi-vendor strategies to avoid dependency and meet privacy requirements. Procurement processes are adapting on the fly – with new frameworks and evaluation criteria – as governments seek to acquire AI solutions in a way that is fast, cost-effective, and aligns with public values (e.g. fairness, transparency). The next few years will likely see more formalized marketplaces for government AI, more collaborative purchasing (to negotiate better prices), and continued involvement of private partners to implement and maintain AI systems in the public sector.

## Budgeting and Cost Considerations

Governments have begun to allocate dedicated funding to AI initiatives, but cost management is a significant concern. The **budgetary aspect** of generative AI adoption involves upfront investments in technology and training, ongoing subscription or infrastructure costs, and weighing these against the anticipated savings or value.

**Budget Allocations:** In the aftermath of ChatGPT's splash, many governments earmarked money for AI in their budgets. For example, the **UK government** in 2023 set aside £100 million for an AI Foundation Model Taskforce to develop and adopt AI, some of which goes into public sector use-case pilots (this was mentioned alongside the Spring Budget) <sup>84</sup>. They also injected funding into the Government Digital Service specifically for building the Humphrey tools. The **European Commission's** AI and Digital program (part of Digital Europe) allocated funds for "AI experimentation sandboxes" that member states can use, which includes generative AI trials in public admin – exact figures vary, but tens of millions of euros at least. **United States:** While there wasn't a single line item "for generative AI" in federal appropriations as of 2024, agencies reprogrammed existing IT modernization funds. For instance, the Department of State used part of its IT modernization budget (which was boosted by the 2024 budget for cyber and tech upgrades) to fund StateChat and related AI tools. The **General Services Administration** provided funding for AI pilots through the Federal Citizen Services Fund. On the state side, Virginia's 2024 budget included \$1.2 million for an AI center of excellence and pilot projects (reported by Pew <sup>149</sup>), and states like **New Jersey** bundled AI

funding into their innovation office budgets. **China** likely spends substantial state funds on AI via its digital economy plans; while exact amounts aren't public, the state-run think tank report suggests dozens of local governments invested in acquiring or building large models <sup>29</sup>, often with subsidies from central government as part of the "AI Plus Government" initiative.

**Cost Structures:** The cost of generative AI in government can be broken down into: (1) **Licensing/Subscription costs** – e.g. paying OpenAI or Microsoft per user per month, or paying usage fees per 1,000 tokens of text generated; (2) **Infrastructure costs** – if self-hosting models or using cloud compute, there are GPU and memory costs, which can be heavy for large models; (3) **Implementation and Integration costs** – consulting, custom development, or extra modules to integrate AI into existing systems; (4) **Training and change management costs** – developing guidelines, running workshops, etc.; and (5) **Governance and oversight costs** – setting up AI audit teams, bias testing, etc.

Agencies that opted for big vendor solutions often face ongoing subscription costs. For example, Microsoft 365 Copilot is expected to cost **\$30 per user per month** for commercial customers – governments likely get a volume deal, but it's still a significant per-user cost. Utah's CIO hinted their cost for Google's AI was double the normal suite price <sup>49</sup>, implying an AI add-on cost that can be hundreds of dollars annually per user. Pennsylvania's pilot, according to local news, used some federal pandemic funds to cover initial costs of ChatGPT Enterprise for a few hundred users. These pilots help agencies gauge if the productivity gains justify expanding licenses to thousands of employees, which could run into the millions of dollars per year.

**Cost-Benefit Examples:** The **Treasury's \$4B fraud savings** dwarfs the cost of its AI tools (the Do Not Pay system enhancements maybe cost a few million to implement). That ROI is clear. Translation AI in Swindon – cost went from presumably tens of thousands of pounds for human translators to a nominal cost for AI (AWS Translate charges per character, so 14 minutes of translation likely cost literally pennies in compute). However, not all use cases have such direct financial ROI. Many are about time savings and qualitative improvements. To convince budget officers to allocate funds, some agencies have projected labor cost savings – e.g. if AI saves each of 100 employees 1 hour a day, that's 100 hours saved, equivalent to, say, 12 full-time staff hours per day. Over a year, that could be ~30,000 hours, which if valued at an average loaded salary (\$50/hour), is \$1.5 million of "time value" freed. Agencies use such estimates to argue an AI tool paying \$200k/year in licenses is worth it. Colorado's report noted *"universal productivity increases"* and specifically highlighted the benefit to employees with disabilities as something not easily captured in dollars but very valuable <sup>150</sup> <sup>151</sup>.

**Funding Sources:** Governments often tap innovation funds or IT funds rather than create new ones. For instance, California's 2023 executive order on GenAI said agencies must explore use cases but within existing budgets, while the state studies broader implications <sup>152</sup>. However, by late 2023 California did publish a report on benefits/risks of GenAI which might inform future budget asks <sup>152</sup>. Federal agencies in the U.S. may reallocate working capital funds to AI. Internationally, organizations like the **World Bank** and **UNDP** have begun funding AI for public service projects in developing countries, meaning some governments can pursue grants. The EU's Recovery and Resilience Facility (post-COVID funding) even allowed spending on "digital public administration improvements", which several countries interpreted to include AI projects.

One interesting development: **public-private cost sharing**. Several major vendors offered initial free or discounted pilots (as noted, MS and AWS providing credits). In Japan, OpenAI offered the government a tailored plan with favorable terms to encourage uptake after seeing interest from Yokosuka and Tokyo,

effectively meaning Japan's cost per use was kept low initially (Nikkei Asia, 2023). Similarly, Chinese tech companies often absorb costs in government pilots to gain the government's favor and to refine their products with government feedback.

**Return on Investment (ROI) Scrutiny:** Legislative bodies are scrutinizing AI spending. The Pew report mentioned state legislatures want to ensure *"eventual benefits and cost savings will be widespread"* but acknowledge *"plenty of trial and error"* will be required <sup>153</sup> <sup>154</sup>. Lawmakers worry about hype – they don't want to pour money into AI that doesn't deliver. This is why pilots and clear metrics are important. Some failures to get budget: a proposal in one U.S. federal agency for \$10 million to fund a large language model project was pared down by OMB, with direction to start smaller and show results first. Governments are learning to start with minimal viable pilots then scale funding in phases.

**Cost of Not Adopting:** On the flip side, some reports highlight an *opportunity cost* of not investing. A McKinsey Global Institute study estimated generative AI could raise Europe's productivity growth significantly <sup>155</sup> – governments factor this into macro-level budget thinking, e.g. "if we invest X in AI now, we save Y in improved growth or efficiency later." There's also a defensive cost – e.g. not using AI in fraud detection could cost billions in unchecked fraud, as Treasury's numbers show <sup>68</sup>. This kind of reasoning is used to justify the spending.

**Future Cost Trends:** As AI adoption scales, some costs may decrease (economies of scale, competition driving down cloud prices or more efficient models). However, *AI vendors may also raise prices* once tools become integral. The Pew Charitable Trusts article noted *"major players are taking losses during explosive development but will need to recoup costs at some point,"* meaning the currently subsidized or at-cost services could become pricier later <sup>154</sup>. State officials are *"bracing for plot twists"* in cost models <sup>156</sup>. To mitigate this, governments might invest in **training their own smaller models** for certain uses – which has a high upfront cost but potentially lower variable cost. The UK's investment in its own foundation model and the EU's push for AI "gigafactories" (compute infrastructure) is partly to control long-term costs <sup>157</sup> <sup>51</sup>. Owning AI infrastructure means not paying per API call to a vendor indefinitely.

**Example Budgets:** While granular budgets are often confidential or buried, a few examples give insight: The State Department's Northstar news-summarizing tool was built by an internal team with maybe <\$1M in costs (they repurposed an existing contract with Palantir, likely); StateChat's scaling to 75k users probably costs a few million per year in Azure OpenAI fees (which might be, say, 5 cents per 750 words generated – they likely negotiated an enterprise rate). The **California report (Nov 2023)** on GenAI suggested state agencies could collectively save hundreds of thousands of hours but also warned of unknown ongoing costs like "model maintenance and monitoring" that need budgeting <sup>152</sup> <sup>158</sup>. No doubt, 2024–25 budgets in many governments include new line items for "AI licenses" or "AI consulting".

In conclusion, budgeting for generative AI is an evolving practice. Initial investments are being made through innovation funds and reallocations, with eye-catching successes (fraud savings, labor hours freed) helping make the case to continue or expand funding. Yet, finance officials remain cautious to see proven ROI, and concerns about escalating subscription costs make governments consider investing in their own capacity. The cost-benefit equation seems favorable in many of the documented cases *if* implementations are done at the right scale and with the right tasks – the challenge for governments is to find that sweet spot without overspending on hype. Strategies for scaling (next section) are inherently tied to managing these costs while maximizing benefits.

## Strategies for Scaling AI Adoption in the Public Sector

Moving from small pilots to wide-scale adoption of generative AI in government requires deliberate strategy. Based on the literature and case studies since 2022, several key **strategies** and best practices are emerging to scale AI use while maintaining control:

- **Establish Clear Governance Frameworks Early:** Governments that scaled AI successfully tended to put policies and governance in place from the outset, to build a strong foundation. Utah enacted an enterprise GenAI policy by 2023, before scaling use to more agencies <sup>46</sup>. Santa Cruz County created a set of *general values and principles* for AI use (transparency, equity, privacy, etc.) that guided its scaling, ensuring any new AI application was measured against these principles <sup>20</sup> <sup>135</sup>. The idea is to prevent a Wild West of AI experimentation that could lead to missteps eroding public trust. At a national level, the **EU AI Act** (expected to be enacted in 2024/25) will impose governance (risk classification, required safeguards) – EU governments are preparing to scale AI under that umbrella, which provides consistency. In the U.S., the Biden Administration’s Executive Order on AI (2023) included mandates for agencies to develop “*AI use policies, protection of data, and reporting*” which essentially forces each agency to have a governance plan as they expand AI use <sup>74</sup>. Having governance doesn’t mean slowing down innovation; Texas’s approach of “*aggressively auditing and refining tools on an ongoing basis*” allowed them to keep using AI widely but with continuous oversight <sup>44</sup>. Part of governance also means addressing ethical and legal questions (e.g., who owns AI-generated content? how to handle AI-related FOIA requests?) before they become problems.
- **Start with High-Impact, Low-Risk Use Cases:** A pragmatic scaling strategy is to focus first on applications that promise clear wins in efficiency or service quality but carry relatively low risk if the AI errs. Many governments identified **knowledge management and internal draft writing** as such a domain. For example, summarization of large text corpora (like consultation responses, meeting transcripts) – errors in summaries can be caught internally and don’t directly harm the public, while the time savings are enormous <sup>81</sup>. Hence tools like Consult, Parlex, and Minute in the UK are spearheading AI scaling in central government because they fit this mold <sup>81</sup> <sup>159</sup>. Another area is **translation and transcription** (as in Swindon or U.N. agencies) – AI can be scaled up to translate millions of words or transcribe thousands of hours quickly, and any mistakes usually just require human correction without endangering lives or rights. Conversely, governments are *holding off* on scaling AI in decision-making roles or in highly sensitive outputs until more proven. For instance, you don’t yet see governments widely using AI to approve/deny benefit applications or to make legal judgments – those remain human tasks, with AI maybe assisting in preparation. The “**small start**” strategy also often involves center-of-excellence teams building a few core AI capabilities (like chat, summarization, search) that can then be replicated across departments. The State Department did this by deploying a general-purpose StateChat then observing usage to decide which specific features to enhance or build next <sup>160</sup>. This flexible approach – *deploy, observe, then focus* – is a strategic way to scale into the unknown: it avoids the trap of presupposing exactly which use cases to invest in, instead letting organic use inform scaling priorities <sup>160</sup>.
- **Invest in Shared Infrastructure (AI Platforms/Marketplaces):** As multiple agencies or departments begin using AI, governments are creating shared platforms to avoid duplication and ensure consistent standards. The **State Department’s AI Marketplace** is one such strategy at agency level – it will let various bureaus plug into a common platform to get AI capabilities, with guardrails built-in and pre-approved tools <sup>161</sup> <sup>64</sup>. Similarly, the UK’s GDS is effectively building a

cross-government AI service (the Humphrey suite) so each department doesn't have to procure or develop AI in silos <sup>84</sup> <sup>81</sup>. Australia's approach of a whole-of-government trial hints at potentially negotiating a whole-of-government license if results are good, meaning all agencies could then access Copilot under one agreement <sup>95</sup> <sup>96</sup>. On a regional scale, the EU's GenAI4EU plan to foster "*large open innovation ecosystems*" indicates building EU-wide resources that national and local governments can tap <sup>88</sup>. This could take the form of an EU AI service platform offering multilingual models fine-tuned for public sector language (legalese, administrative jargon) that any member state can use, rather than each state training its own. The benefit of shared infrastructure is scaling with consistency – it enforces security measures universally and can be more cost-effective due to pooled investment.

- **Leverage Iterative and Agile Implementation:** Traditional IT projects in government are often big bang and slow. With AI, many governments are wisely using an **iterative approach**. They deploy a minimally viable tool, gather feedback, and rapidly improve it in cycles. The State Department did a small pilot of StateChat in one office, gathered user feedback, improved the model's responses and interface, then rolled out department-wide over a summer <sup>162</sup> <sup>163</sup>. This agile approach helped catch issues early and incorporate user suggestions, making scaling smoother. Kelly Fletcher (State CIO) noted they "*trained the workforce to understand: what you get on day one is not what you have on day 15...through this iterative process, we keep adding capabilities.*" <sup>164</sup> <sup>165</sup>. That philosophy – expect tools to evolve quickly – prepares staff to accept updates and new features rather than being caught off guard. Similarly, in local government, Wentzville's team started without a formal AI policy but iteratively developed one as they learned how the tools fit into their work <sup>166</sup>. They leaned on existing ethics policies initially and then crafted AI-specific rules once they understood usage patterns <sup>166</sup>. This "learn then formalize" strategy can be effective in scaling because it avoids over-engineering rules at the very start and instead bases them on real experience.
- **Promote Cross-Pollination and Knowledge Sharing:** Scaling is accelerated when organizations share successes and challenges. Recognizing this, many governments are fostering communities of practice. The U.S. **federal government** set up an internal AI Community on GSA's Technology Modernization Fund website where agency officials post use cases and lessons learned (this grew out of earlier data science communities). NASCIO (states CIOs) and NACo (counties) have active working groups on AI, as mentioned <sup>27</sup>. For instance, Zach Friend from Santa Cruz visiting the White House to share local lessons ensured federal policy considered local realities <sup>167</sup>. Texas explicitly encourages state and local employees to "*learn from one another's experiences*," signaling a structured way to share AI project results across jurisdictions <sup>44</sup>. Internationally, the **OECD** and **UN** have forums for public sector AI – the OECD AI Observatory collects examples from member countries and could guide scaling globally by highlighting what works. By tapping into these networks, a small city or a ministry can replicate a successful scaled solution from elsewhere rather than reinventing it. Additionally, vendor-led user groups exist – e.g. Microsoft's Government AI user group – where public sector clients discuss use and thus indirectly help each other scale by finding solutions to common problems (security, compliance, etc.).
- **Maintaining Human Oversight and Public Trust:** A critical strategy for scaling is ensuring that as AI use grows, mechanisms for human oversight and public transparency scale with it. Many governments plan to "**keep humans in the loop**" no matter how advanced the AI becomes. For example, the UK's Red Teaming and Ethics unit (part of the new AI taskforce) will continually evaluate models used in government for misbehavior and advise on limits. Some jurisdictions, like New York

City, are considering or implementing **AI registers** – public lists of AI systems used by government (the UK had discussed this but The Guardian noted no Whitehall dept had listed any as of Oct 2023, showing a gap <sup>168</sup> <sup>169</sup> ). The idea is that transparency can help maintain public trust as usage scales. The EU AI Act will likely require public disclosure when citizens are interacting with an AI system rather than a human, which is a strategy to prevent backlash and hold agencies accountable. From an oversight perspective, scaling might involve forming dedicated AI oversight bodies (some countries have or plan multi-stakeholder AI councils). For instance, **California's report** recommended establishing a statewide AI governance framework as a precursor to broad adoption <sup>152</sup> . Ensuring robust oversight structures from the start is a strategy to enable scaling *safely* – it's easier to expand usage when there's confidence that issues will be caught.

- **Continuous Evaluation and Adaptation:** Scaling doesn't mean set-and-forget; it means continuous evaluation of AI effectiveness and impacts as use grows. The **Australian trial's evaluation** is a good model – it tested the premise “will generative AI actually be adopted by workers and yield benefits in real-world conditions?” and identified constraints to address before wider rollout <sup>98</sup> <sup>102</sup> . Governments are building in KPIs and feedback loops. For example, if a city scales an AI chatbot to all departments, they might track metrics like resolution rate of inquiries, user satisfaction scores, and any escalation rates to humans. If any of those trend negatively, they adjust. The “**show more**” link in the Deloitte report snippet suggests governments should “*embed gen AI into core processes*” to truly transform <sup>170</sup> , and indeed scaling strategy involves redesigning processes around AI (not just shoehorning AI into old processes). But that requires a careful change management plan so the process changes are accepted.

In essence, scaling strategies focus on balancing ambition with caution: roll out AI broadly enough to get network effects and big benefits, but do so under strong governance, with staff buy-in, and with mechanisms to catch problems early. The experiences from late 2022 to 2024 show that those governments that treat scaling as an iterative, learning-oriented journey – involving policy, people, and technology in tandem – are faring better than those who might rush or, conversely, stagnate in pilot purgatory. Many jurisdictions are now beyond the initial pilot phase and are gearing up for this careful expansion, guided by the principle that **responsible scaling of AI is a marathon, not a sprint**, but one that can significantly modernize public services if done right.

## Conclusion

In the short time since generative AI entered the public consciousness, municipal, state, and federal agencies around the world have moved from curiosity to concrete experimentation. Governments in the U.S., U.K., EU, as well as in China, Japan, Australia and beyond, have demonstrated that generative AI can be a **powerful ally** in civic tasks – speeding up document-intensive work, enhancing citizen access to information, and uncovering insights from vast data – all while freeing public servants to focus on the human-centric aspects of their missions. Real-world deployments have already saved governments time and money, from **hours saved in Pennsylvania's offices** <sup>106</sup> and **translation costs slashed in Swindon** <sup>13</sup> , to **fraudulent payments prevented by Treasury's AI** <sup>68</sup> . These tangible benefits, however, come paired with **new challenges**. Issues of accuracy, bias, privacy, and workforce adaptation have required careful navigation, reinforcing that the public sector's embrace of AI must be **deliberate and principled**.

The experiences documented since late 2022 highlight a few overarching themes. First, the most successful government AI initiatives have combined **innovation with stewardship** – they start with limited trials,

actively involve the workforce in learning, and set clear ethical guardrails before scaling up. Second, cross-sector collaboration (whether with tech vendors, academia, or inter-government networks) has been crucial – no agency has done this alone. Third, transparency with the public – about what AI is used and how outputs are reviewed – is vital to maintain trust as these technologies become more embedded in public services. Finally, governments are thinking strategically: investing in shared platforms, training their people, and updating procurement and policies to pave the way for broader AI integration in the coming years.

Late 2023 and 2024 saw a flurry of AI task forces, executive orders, and framework documents, indicating that generative AI's role in government is shifting from opportunistic pilot to **structural component** of digital government. As we move forward, we can expect to see generative AI powering more and more behind-the-scenes functions – drafting legislation, customizing citizen services, automating administrative minutiae – as well as citizen-facing tools like smarter chatbots and real-time translators. The trajectory set in motion since ChatGPT's release suggests that, managed wisely, generative AI will become a standard part of the public sector toolbox. The literature and cases reviewed here give reason for optimism: when used to augment human judgment rather than replace it, generative AI has the potential to make government not only more efficient, but more **responsive, inclusive, and innovative** in delivering for the public. The next few years will be critical in scaling these early successes into sustainable, system-wide transformations of how government serves – a journey that is just beginning, but one clearly underway across multiple levels of governance around the world.

#### Sources:

- Deloitte Insights – *Scaling Generative AI in Government* (2024), which compiles numerous public-sector AI case studies <sup>1</sup> <sup>171</sup> .
- FedScoop – *State Department turns to AI to assist workforce* (Dec 2024), detailing State's internal AI tools and savings <sup>55</sup> <sup>59</sup> .
- Pew Charitable Trusts – *States seek to leverage AI's promise while mitigating hazards* (Jan 2025), with state pilot results and cost insights <sup>34</sup> <sup>49</sup> .
- GovTech – *Where to Start With AI? Cities and States Offer Use Cases* (Mar 2024), reporting local examples in Boston, Wentzville, Santa Cruz, etc. <sup>172</sup> <sup>8</sup> .
- ICMA – *Local Government Practitioners Weigh in on AI* (Nov 2024), survey data on local AI priorities, chatbots, and barriers <sup>22</sup> <sup>5</sup> .
- GovInsider – *GenAI zeitgeist: how generative AI is shaping gov transformation* (July 2024), global perspective including Swindon's translation case <sup>87</sup> <sup>173</sup> .
- Government Transformation Summit – *Yokosuka adopts ChatGPT in city government* (June 2023), notes from Yokosuka and Tokyo's adoption, including time saved and accuracy issues <sup>15</sup> <sup>17</sup> .
- Nextgov/FCW – *AI tools helped Treasury recover billions* (Oct 2024), on Treasury's fraud detection AI and outcomes <sup>68</sup> .
- Global Government Forum – *Meet Humphrey: UK's AI package for officials* (Jan 2025), describing U.K.'s suite of AI tools (Consult, Parlex, etc.) <sup>81</sup> <sup>82</sup> .
- NSW Government – *Digital NSW guidance on Generative AI* (2023), for prompt techniques and ethical use in New South Wales <sup>86</sup> .
- and additional official reports and news as cited throughout <sup>4</sup> <sup>95</sup> , which provide further details on policies, training, and strategies in each region.

1 2 13 42 106 142 143 146 147 148 149 152 158 170 171 **Scaling gen AI in governments | Deloitte Insights**  
<https://www2.deloitte.com/us/en/insights/industry/public-sector/use-of-ai-in-government.html>

3 8 9 10 11 12 19 20 21 27 28 37 38 39 40 41 45 46 47 74 75 120 121 123 134 135 166 167 172  
**Where to Start With AI? Cities and States Offer Use Cases**  
<https://www.govtech.com/artificial-intelligence/where-to-start-with-ai-cities-and-states-offer-use-cases>

4 76 127 128 140 **Guidance to civil servants on use of generative AI - GOV.UK**  
<https://www.gov.uk/government/publications/guidance-to-civil-servants-on-use-of-generative-ai/guidance-to-civil-servants-on-use-of-generative-ai>

5 6 7 22 23 24 25 26 118 129 137 **Local Government Practitioners Weigh in on AI | icma.org**  
<https://icma.org/articles/article/local-government-practitioners-weigh-ai>

14 89 90 91 **Yokosuka becomes 1st local gov't in Japan to begin ChatGPT trial use**  
<https://english.kyodonews.net/news/2023/04/0b30bc9d9174-yokosuka-becomes-1st-local-govt-in-japan-to-begin-chatgpt-trial-use.html>

15 16 17 18 94 107 108 **Japan to use ChatGPT as a tool in city government**  
<https://www.government-transformation.com/innovation/japan-to-use-chatgpt-as-a-tool-in-city-government>

29 93 **Research Report on Governance Modernization in the Digital Age: Practice and Prospects for the Application of Large Models in the Government Domain (2023) | Center for Security and Emerging Technology**  
<https://cset.georgetown.edu/publication/caict-china-government-llm-report-2023/>

30 31 32 33 34 35 36 43 44 49 116 150 151 153 154 156 **States Governments Seek to Leverage AI's Promise While Mitigating Its Hazards | The Pew Charitable Trusts**  
<https://www.pew.org/en/research-and-analysis/articles/2025/01/15/states-governments-seek-to-leverage-ais-promise-while-mitigating-its-hazards>

48 81 82 83 84 85 159 **Yes, civil servant: Meet Humphrey, the UK government's AI package for officials - Global Government Forum**  
<https://www.globalgovernmentforum.com/yes-civil-servant-meet-humphrey-the-governments-ai-package-for-officials/>

50 **Overheidsbrede visie Generatieve AI**  
<https://www.government.nl/binaries/government/documenten/parliamentary-documents/2024/01/17/government-wide-vision-on-generative-ai-of-the-netherlands/Government-wide+vision+on+generative+AI+of+the+Netherlands.pdf>

51 52 88 157 **European approach to artificial intelligence | Shaping Europe's digital future**  
<https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>

53 54 **'Talk About Something Else': Chinese AI Chatbot Toes Party Line**  
<https://www.voanews.com/a/talk-about-something-else-chinese-ai-chatbot-toes-party-line/7249469.html>

55 56 57 58 59 60 61 62 114 115 119 124 125 126 160 162 163 **From translation to email drafting, State Department turns to AI to assist workforce | FedScoop**  
<https://fedscoop.com/state-department-ai-chatbot-email-drafting-northstar-famsearch/>

63 64 65 66 67 161 164 165 **State to develop new AI marketplace for staff - Nextgov/FCW**  
<https://www.nextgov.com/artificial-intelligence/2024/11/state-develop-new-ai-marketplace-staff/400750/>

68 69 70 71 72 73 117 **AI tools helped Treasury recover billions in fraud and improper payments - Nextgov/FCW**  
<https://www.nextgov.com/artificial-intelligence/2024/10/ai-tools-helped-treasury-recover-billions-fraud-and-improper-payments/400368/>



77 78 79 80 **DWP reverses ban on ChatGPT and other generative AI tools**

<https://www.civilserviceworld.com/professions/article/dwp-chatgpt-generative-ai-ban-reversed-llms-deepseek>

86 **Chatbot prompt essentials | Digital NSW**

<https://www.digital.nsw.gov.au/policy/artificial-intelligence/chatbot-prompt-essentials>

87 112 113 130 131 132 138 139 173 **Captured by the GenAI zeitgeist: How generative AI is shaping government transformation**

<https://govinsider.asia/intl-en/article/captured-by-the-genai-zeitgeist-how-generative-ai-is-shaping-government-transformation>

92 **China introduces rules governing generative AI services like ChatGPT**

<https://www.cnbc.com/2023/07/13/china-introduces-rules-governing-generative-ai-services-like-chatgpt.html>

95 96 97 98 99 100 101 102 122 136 145 **Summary report – executive summary | digital.gov.au**

<https://www.digital.gov.au/initiatives/copilot-trial/summary-evaluation-findings/cts-executive-summary>

103 **Interim guidance on government use of public generative AI tools | aga**

<https://architecture.digital.gov.au/generative-ai>

104 **TfNSW to build internal generative AI chatbot - iTnews**

<https://www.itnews.com.au/news/tfnsw-to-build-internal-generative-ai-chatbot-612815>

105 **Chapter 2 - Artificial intelligence in the public sector**

[https://www.aph.gov.au/Parliamentary\\_Business/Committees/Joint/Public\\_Accounts\\_and\\_Audit/PublicsectoruseofAI/Report/Chapter\\_2\\_-\\_Artificial\\_intelligence\\_in\\_the\\_public\\_sector](https://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Public_Accounts_and_Audit/PublicsectoruseofAI/Report/Chapter_2_-_Artificial_intelligence_in_the_public_sector)

109 110 **The AI opportunity for eGovernment in the EU | Google Cloud Blog**

<https://cloud.google.com/blog/topics/google-cloud-europe/the-ai-opportunity-for-egovernment-in-the-eu>

111 **[PDF] The AI opportunity for eGovernment in the EU - Implement**

<https://cms.implementconsultinggroup.com/media/uploads/articles/2025/The-ai-opportunity-for-egovernment-in-the-eu/The-AI-opportunity-for-eGovernment-in-the-EU.pdf>

133 **[PDF] Artificial Intelligence Playbook for the UK Government - GOV.UK**

[https://assets.publishing.service.gov.uk/media/67aca2f7e400ae62338324bd/AI\\_Playbook\\_for\\_the\\_UK\\_Government\\_\\_12\\_02\\_.pdf](https://assets.publishing.service.gov.uk/media/67aca2f7e400ae62338324bd/AI_Playbook_for_the_UK_Government__12_02_.pdf)

141 **Baidu says AI chatbot 'Ernie Bot' has attracted 200 million users**

<https://www.reuters.com/technology/baidu-says-ai-chatbot-ernie-bot-has-amassed-200-million-users-2024-04-16/>

144 **AI in government: Top use cases - IBM**

<https://www.ibm.com/think/topics/ai-in-government>

155 **Time to place our bets: Europe's AI opportunity - McKinsey**

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/time-to-place-our-bets-europes-ai-opportunity>

168 169 **UK government failing to list use of AI on mandatory register**

<https://www.theguardian.com/technology/2024/nov/28/uk-government-failing-to-list-use-of-ai-on-mandatory-register>