CODE *for* AMERICA

Government AI Landscape Analysis

The Government AI Journey: From Readiness to Public Value

April 2026

The government AI journey: From readiness to public value

Artificial Intelligence adoption and use in state government is not a single procurement decision or a one-time technology upgrade. It is an institutional journey across agencies and departments, as the public sector seeks to make sense and derive value from the fast-evolving technological changes in AI.

Looking for state-specific insights?

Explore detailed assessments, comparisons, and interactive data for each state: codeforamerica.org/ai-landscape

Most states are still early in this journey. Some are building governance frameworks. Others are piloting generative AI tools. A few are beginning to scale implementation. Very few have fully embedded AI into core operations with robust measurement and continuous improvement. The AI journey can be understood as a progression across four stages:

Readiness → Piloting → Implementation → Impact

Each stage builds upon the previous one. Readiness creates the necessary capabilities for experimentation. Piloting informs what will work and not work for full scale operational implementation. Implementation creates measurable outcomes. Impact strengthens governance and fuels the next cycle of innovation.

With that in mind, this assessment focuses most on generative AI adoption, use, and value, while also picking up on signals of previous implementations of predictive AI and intelligent automation, and new work in agentic AI. While the assessment took into account the general trajectory of the AI journey for states, we honed in on benefits delivery and similar use cases for the experimentation and deployment evaluation.

At its core, the AI journey reflects institutional growth:

- **Readiness** builds the foundation
- **Piloting** improves confidence
- **Implementation** delivers results
- **Impact** ensures accountability and improvement

States move at different speeds across these stages. Some may be established in governance but early in capacity building. Others may have strong infrastructure but limited executive coordination.

As such, this journey is not fully linear, but rather is iterative. As AI technology evolves rapidly, governments must remain flexible, adaptive, and willing to revisit earlier stages with new insights.

For example, a state could be deploying generative AI tools to staff and seeing significant productivity gains, but may not have engaged agentic AI for intelligent automation. A state seeking to adopt agentic AI will need to start the cycle again doing similar work on readiness and experimentation that it did for generative AI.

The goal is not perfection. The goal is progression—toward AI systems that are effective, ethical, transparent, and aligned with public service.

Core findings from the analysis

A nation transitioning from experimentation to operational integration, but not yet systematically measuring impact.

The analysis reveals a clear pattern: states are progressing quickly through governance and experimentation but more slowly through operational scaling and impact measurement.

Most states now recognize AI as a strategic public-sector capability rather than a niche technical tool. Executive orders, advisory councils, and AI task forces have proliferated since 2023, reflecting the rapid emergence of generative AI and its potential to reshape government operations. Governance structures have expanded particularly quickly as policymakers seek to establish ethical guardrails around AI systems.

Experimentation is also widespread. **Nearly every state has launched some form of AI pilot, frequently focused on generative AI tools that assist public employees with research, summarization, and administrative workflows.** These pilots have allowed governments to test AI capabilities while maintaining human oversight and minimizing risk.

However, the transition from experimentation to operational deployment remains uneven. **Only a small group of states have embedded AI into enterprise-scale workflows across multiple agencies.** In many jurisdictions, AI deployments remain confined to narrow use cases such as fraud detection, document processing, or customer service chat assistants.

The greatest maturity gap appears in the final stage of the AI journey: measuring impact and learning from results. **Few states have established systematic mechanisms to evaluate the public value of AI deployments.** While governance frameworks are widespread, robust performance measurement—such as tracking cost savings, service improvements, or outcomes—remains limited.

This report identifies a small group of leading states that are shaping the national trajectory of government AI adoption. These jurisdictions demonstrate several common characteristics:

- Strong executive leadership and cross-agency governance
- Enterprise data infrastructure that supports advanced analytics
- AI innovation labs or sandboxes for controlled experimentation
- Structured pilot programs with documented evaluation
- Early systems for measuring operational outcomes

These leading states—including Maryland, New Jersey, North Carolina, Pennsylvania, Texas, Utah, and Vermont—are not simply adopting AI tools. They are building the institutional capabilities required to govern AI as a long-term public-sector asset.

The national landscape can therefore be understood as a transition period. AI is moving rapidly from experimentation into operational government use. **The next phase will depend on whether states can convert early deployments into measurable public value through disciplined evaluation, transparent governance, and continuous learning.**

The coming decade will determine how quickly, deeply and effectively AI can become a transformative force in public administration or if it remains a collection of isolated technology experiments. States that invest early in governance, infrastructure and workforce capacity are positioning themselves to realize the full potential of AI while maintaining public trust.

The national pattern resembles a pyramid, where there is broad progress in readiness, rapid growth in piloting, limited scaling in implementation, and impact at scale remains rare.

Stages explained

Stage 1: Readiness

Creating the conditions for responsible adoption

Before AI can be used responsibly at scale, the right institutional conditions must exist. Foundational readiness focuses on three core pillars: leadership, capacity, and infrastructure.

Leadership

This category evaluates the organizational structure and leadership dedicated to AI initiatives within state government. A chief AI officer or designated executive leader is empowered to coordinate across agencies. Advisory groups or councils begin shaping policy direction. Governance structures define who approves AI tools, how risks are assessed, and how accountability is maintained. Without this clarity, AI efforts become fragmented.

Capacity

Governments must invest in people, not just tools. Foundational readiness includes AI literacy training and training on safeguards, workforce development partnerships, and internal communities of practice. This category evaluates the state's investments in developing AI literacy, skills, and expertise across its workforce. It examines formal training programs, partnerships with educational institutions and industry for knowledge transfer, and structured upskilling pathways for employees. Effective capacity building ensures that state employees have the necessary skills to leverage AI technologies appropriately, enables informed procurement decisions, and creates internal champions who can drive AI adoption within their agencies. Without workforce capability, AI adoption and use remains superficial or overly dependent on vendors.

Infrastructure

This category evaluates the technical foundation necessary to support quality data and advanced AI implementations. It examines data infrastructure and accessibility, computing resources and platforms, and partnerships with technical vendors and service providers. Robust technical infrastructure is essential for successful AI adoption, ensuring that agencies have access to high-quality data, sufficient processing capacity for complex models, and specialized expertise to address implementation challenges across diverse use cases. Foundational infrastructure allows governments to experiment safely and scale effectively later.

The outcome of Stage 1

The state is prepared to experiment responsibly. Governance is defined. Staff are learning. Infrastructure can support pilots. The environment is stable enough to move forward.

Stage 2: Piloting

Learning through structured innovation

Once foundational elements are in place, governments begin structured experimentation and piloting. In this stage, we see states building AI innovation labs with agencies or statewide, AI sandboxes for testing safely, pilot projects, and limited deployments with clear guardrails and timeframes for evaluation.

Such experimentation can reveal operational challenges like data or infrastructure limitations, ethical risks, workforce gaps, and procurement bottlenecks due to lack of established criteria or subject matter expertise. This stage builds confidence while identifying and addressing constraints. Importantly, safe experimentation prevents large-scale missteps. It allows states to adjust governance policies, refine procurement pathways, and better understand where AI creates real value.

The outcome of Stage 2

The state has validated promising use cases and developed operational insight. Leadership now has evidence to justify scaling specific tools and methods while discontinuing others.

Stage 3: Implementation

From pilots to production

At this stage, AI becomes embedded in government operations. It moves from isolated pilots to scaled systems. Implementation may include: AI-assisted case management, public-facing chat assistants, predictive tools for benefits administration, fraud detection models, and document automation.

At this stage, funding for AI implementation becomes more stable. Governance processes become routine as procurement pathways mature. Staff increasingly integrate AI into daily workflows. Essentially, this stage represents institutionalization.

AI is no longer experimental—it is operational. However, implementation introduces new complexity that requires new capabilities: ongoing model monitoring, bias mitigation, cybersecurity concerns, and, perhaps most significantly, change management in a space where there are few established platforms. Without disciplined evaluation, scaling can amplify risks as well as benefits.

The outcome of Stage 3

AI delivers tangible improvements in efficiency and service delivery. The state demonstrates measurable operational gains.

Stage 4: Impact

From deployment to public value

The final stage focuses on accountability and adaptation. AI systems must be monitored and measured for long-term impact and sustaining change. Government fully and continually assesses efficiency gains, cost savings, service quality improvements, and public trust.

Performance measurement becomes systematic. Feedback loops are embedded. Governance frameworks are updated based on lessons learned. Training programs evolve alongside new technologies. This stage transforms AI from a set of technology projects into a learning ecosystem that continues to incorporate evolving AI changes. It reinforces agility. Because AI evolves rapidly, states must continuously refine policies, infrastructure, workforce skills, and risk frameworks.

At the most mature version of this stage, measured AI impact will do much more: It will create meaningful, positive change, including better and easier access to benefits and an overall improved experience with government for the public.

The outcome of Stage 4

The state operates within a culture of responsible AI optimization. AI systems are continuously improved, evaluated, and aligned with public accountability.

Overall evaluation

Each state was evaluated against this rubric, which we cover in detail later. To develop a comprehensive view of state-level AI readiness and use, we conducted extensive desk research using publicly available materials with feedback loops from states and AI leaders. This included:

- Executive orders: Gubernatorial executive actions that established task forces, governance frameworks, or AI strategies
- Legislation and policies: Laws and bills related to artificial intelligence
- Agency guidance and reports: Strategic plans, policy documents, and technical guidance issued by state agencies, particularly IT departments
- Media and trade articles: Local and national news coverage, civic tech blogs, and industry reporting
- Direct state input: Opportunity for direct feedback and correction from states upon reviewing draft analysis
- Review of all framing and the rubric by an advisory council

State	Stage 1: Readiness	Stage 2: Piloting	Stage 3: Implementation	Stage 4: Impact
Alabama	Developing	Early	Early	Early
Alaska	Early	Early	Early	Early
Arizona	Established	Established	Developing	Early
Arkansas	Developing	Developing	Early	Early
California	Established	Established	Developing	Developing
Colorado	Advanced	Established	Developing	Developing
Connecticut	Established	Established	Early	Developing
Delaware	Developing	Early	Early	Early
Florida	Early	Early	Early	Early
Georgia	Advanced	Established	Developing	Developing
Hawaii	Early	Developing	Early	Early
Idaho	Developing	Developing	Early	Early
Illinois	Established	Developing	Early	Early
Indiana	Developing	Established	Developing	Developing
Iowa	Developing	Established	Early	Early
Kansas	Early	Early	Early	Early
Kentucky	Established	Developing	Early	Developing
Louisiana	Established	Early	Early	Developing
Maine	Developing	Developing	Early	Early
Maryland	Advanced	Established	Established	Established

Massachusetts	Established	Established	Established	Developing
Michigan	Established	Established	Developing	Early
Minnesota	Advanced	Established	Developing	Developing
Mississippi	Developing	Established	Early	Developing
Missouri	Developing	Developing	Early	Early
Montana	Established	Developing	Developing	Developing
Nebraska	Early	Early	Early	Early
Nevada	Developing	Developing	Developing	Developing
New Hampshire	Developing	Early	Early	Early
New Jersey	Advanced	Established	Established	Established
New Mexico	Developing	Developing	Developing	Early
New York	Advanced	Developing	Developing	Developing
North Carolina	Advanced	Established	Established	Established
North Dakota	Early	Developing	Early	Early
Ohio	Established	Advanced	Established	Early
Oklahoma	Developing	Developing	Early	Early
Oregon	Established	Established	Early	Early
Pennsylvania	Advanced	Established	Established	Established
Rhode Island	Developing	Developing	Early	Early
South Carolina	Established	Developing	Early	Early
South Dakota	Developing	Developing	Developing	Early
Tennessee	Established	Established	Established	Developing
Texas	Advanced	Advanced	Established	Established
Utah	Advanced	Advanced	Established	Established
Vermont	Established	Advanced	Established	Established
Virginia	Developing	Developing	Early	Early
Washington	Advanced	Established	Established	Developing
West Virginia	Early	Early	Early	Early
Wisconsin	Developing	Developing	Developing	Early
Wyoming	Early	Early	Early	Early
Washington, D.C.	Established	Established	Developing	Early

The distribution of the ratings include:

- **Stage 1 (Readiness):** Early 8, Developing 17, Established 15, Advanced 11
- **Stage 2 (Piloting):** Early 10, Developing 18, Established 19, Advanced 4
- **Stage 3 (Implementation):** Early 26, Developing 14, Established 11, Advanced 0
- **Stage 4 (Impact):** Early 29, Developing 15, Established 7, Advanced 0

In a correlation analysis, higher foundational readiness was strongly associated with stronger Stage 2 performance, and that relationship remains moderately strong into Stages 3 and 4. For the specific aspects of readiness, infrastructure is the strongest predictor of later-stage performance, especially for Stage 4 impact measurement/learning, with capacity building also very important.

Key lessons from the leading states

Across these leading, several consistent patterns emerge.

1. **Leadership matters:** States with strong executive leadership and dedicated AI governance bodies move faster.
2. **Data infrastructure determines scalability:** States with enterprise data platforms progress more quickly to operational AI.
3. **Structured experimentation accelerates learning:** Formal AI sandboxes and pilot programs generate better evidence.
4. **Workforce readiness is essential:** Training public employees in AI tools is a major driver of adoption.
5. **The risk tolerance for government is lower:** For government to responsibly move from piloting to implementation, there is rightfully a stronger line in the sand than for industry. Government technology has a real impact on people's lives, like access to benefits.
6. **Measurement remains the next frontier:** Even leading states are still developing systematic methods to measure public value from AI.

Stage 1: Readiness

Introduction

State governments across the U.S. are rapidly evolving in their readiness to adopt artificial intelligence in public services. In the broader AI journey, Stage 1 is not the most visible stage, but it is arguably the most consequential. It determines whether AI adoption will be coordinated or fragmented, accountable or opaque, strategic or reactive. The states that treat governance as a living, evolving discipline—rather than a one-time policy exercise—are positioning themselves for stronger outcomes in every subsequent stage.

Rubric: Readiness

Building on the 2025 baseline assessment, this 2026 update evaluates each of the 50 states and Washington, D.C. on three key domains of AI readiness: Leadership & Governance, AI Capacity Building, and Technical Infrastructure & Capabilities. We apply the same four-tier maturity rubric—Early, Developing, Established, Advanced—defined in the original framework.

- **Early:** Initial steps with only basic foundational elements emerging
- **Developing:** Core components in place with growing capabilities and some formalization
- **Established:** Mature implementations with systematic approaches and demonstrated effectiveness
- **Advanced:** Sophisticated, comprehensive frameworks and innovative, state-of-the-art approaches

Most states in 2025 were in the Developing stage across domains. Over the past year, many have made notable progress through new executive orders, legislation, task forces, and partnerships, while others still face challenges in moving beyond foundational steps.

AI readiness evaluation 2025 and 2026 comparison

State	Overall 2025	Overall 2026	Δ
Alabama	Developing	Developing	→
Alaska	Early	Early	→
Arizona	Established	Established	→
Arkansas	Developing	Developing	→
California	Established	Established	→
Colorado	Established	Advanced	↑
Connecticut	Established	Established	→
Delaware	Early	Developing	↑
Florida	Early	Early	→
Georgia	Established	Advanced	↑
Hawaii	Early	Early	→
Idaho	Early	Developing	↑
Illinois	Established	Established	→
Indiana	Developing	Developing	→
Iowa	Early	Developing	↑
Kansas	Early	Early	→
Kentucky	Developing	Established	↑
Louisiana	Developing	Established	↑
Maine	Developing	Developing	→
Maryland	Established	Advanced	↑
Massachusetts	Established	Established	→
Michigan	Developing	Established	↑
Minnesota	Established	Advanced	↑
Mississippi	Developing	Developing	→
Missouri	Early	Developing	↑
Montana	Early	Established	↑

Nebraska	Early	Early	→
Nevada	Developing	Developing	→
New Hampshire	Developing	Developing	→
New Jersey	Advanced	Advanced	→
New Mexico	Developing	Developing	→
New York	Established	Advanced	↑
North Carolina	Established	Advanced	↑
North Dakota	Early	Early	→
Ohio	Established	Established	→
Oklahoma	Developing	Developing	→
Oregon	Developing	Established	↑
Pennsylvania	Advanced	Advanced	→
Rhode Island	Developing	Developing	→
South Carolina	Established	Established	→
South Dakota	Early	Developing	↑
Tennessee	Established	Established	→
Texas	Established	Advanced	↑
Utah	Advanced	Advanced	→
Vermont	Established	Established	→
Virginia	Developing	Developing	→
Washington	Established	Advanced	↑
West Virginia	Early	Early	→
Wisconsin	Developing	Developing	→
Wyoming	Early	Early	→
Washington, D.C.	Established	Established	→

Overall trends of Stage 1

Overall, Stage 1 shows the strongest maturity distribution of the three foundational domains, suggesting that states have prioritized organizing governance structures before moving aggressively into experimentation and scaling.

In last year's analysis, the majority of states were in the Early or Developing stage for AI experimentation, with only a handful (e.g., New Jersey, Utah, Georgia) reaching Advanced practices. Virtually every state had made some progress since—many moving up one tier—as the urgency to explore AI grew. A few innovative states jumped into the Advanced tier by launching extensive pilot programs or formal sandboxes (notably, California, New York, Pennsylvania, Massachusetts, among others, joined the leaders). No state moved backward. Even traditionally slower adopters took cautious steps forward. Still, the gap between leaders and slower adopters remained pronounced: The top-tier

states had enterprise-wide AI labs or sandboxes and multiple pilots running, whereas the lowest-tier states were only beginning to coordinate pilots after initial task force studies.

Summary evaluation for Stage 1

Stage 1 represents the institutional awakening to artificial intelligence. It's the stage where state governments move from observing AI as an external technological trend to recognizing it as a governance responsibility. In this stage, the central question is not yet how to deploy AI at scale, but who is accountable for it, how it should be governed, and what structures must exist to guide its use responsibly.

Over the past two years, nearly every state has entered this foundational stage in some capacity. The rapid public emergence of generative AI tools accelerated executive attention. Governors, legislatures, and agency heads quickly realized that AI would affect procurement, cybersecurity, workforce policy, service delivery, and public trust. Leadership responses were often swift—executive orders, task forces, advisory councils, and policy memoranda proliferated across states beginning in 2023 and continues through today.

Nationally, most states now sit in the Developing or Established range for leadership and governance maturity. It is increasingly common to see AI responsibilities assigned to chief information officers, chief data officers, chief AI officers, senior advisers for AI, or newly created working groups. Formal advisory councils have become a dominant governance mechanism. These bodies typically include representatives from IT, legal, data governance, cybersecurity, and, in some cases, academia or civil society. Their purpose is to study AI risks and opportunities, recommend guardrails, and coordinate early policy responses.

However, true executive-level AI authority remains less common. AI governance is layered onto existing technology leadership roles. While this can be effective, it often reflects resource constraints and the early maturity of AI as a governmental discipline. The debate over whether a dedicated AI executive is necessary continues across states. What matters most at this stage is not the title itself, but whether there is clear ownership and accountability for AI strategy.

A defining feature of Stage 1 nationally has been the emphasis on guardrails. States have focused heavily on responsible AI principles, ethical frameworks, transparency considerations, and risk mitigation. The instinct has been cautious but constructive. Many executive orders and advisory reports emphasize human oversight, bias mitigation, privacy protection, and procurement controls. Rather than racing toward implementation, leadership structures have often prioritized understanding and containment.

At the same time, governance maturity varies widely. In some states, AI task forces exist primarily to study the issue and provide recommendations. In others, governance structures are operationalized—policies are issued, intake and review processes are established, and agencies are

required to submit AI use cases for approval. The latter model reflects a transition from exploratory governance to enforceable governance.

Another important national trend is legislative engagement. While executive action dominated the early response, legislatures are increasingly entering the space. Some states have enacted AI oversight laws, established statutory task forces, or mandated risk assessments. This growing legislative involvement signals that AI governance is moving from temporary executive attention to more durable institutional frameworks.

Despite this progress, Stage 1 governance remains uneven. In many states, coordination is still emerging. AI oversight may be decentralized, with individual agencies interpreting guidance independently. Cross-agency communication can be limited, and enterprise-wide standards are still forming. Moreover, governance structures often precede technical expertise, meaning advisory groups may lack deep in-house AI knowledge in early stages.

Still, the overall trajectory is clear. States have largely accepted that AI governance cannot be reactive. It must be proactive, structured, and ongoing. The creation of advisory councils, the issuance of statewide policies, and the assignment of executive responsibility all reflect an understanding that AI adoption without governance risks eroding public trust.

Stage 1, therefore, is less about technology and more about institutional alignment. It is about clarifying authority, setting expectations, and creating a foundation upon which experimentation and implementation can safely build. Without this stage, later stages of AI maturity become unstable. Governance and capability building establishes the conditions for responsible innovation.

Stage 2: Piloting

Introduction

Piloting refers to a state's capacity to conduct safe, controlled, and learning-oriented trials of AI technologies before full deployment. In this stage, state agencies run time-bound AI pilots or sandbox trials to explore specific use cases, require human-in-the-loop oversight for AI decisions, perform bias and risk assessments on AI systems, and document outcomes to inform scale-up or shutdown decisions.

Cross-functional teams (e.g., policy, IT, legal) are typically involved to ensure pilots align with ethical guidelines and public values. The leading states have institutionalized "safe sandbox" environments and learned to systematically pilot AI with clear go/no-go decision points, while lagging states were still organizing basic pilots.

Rubric: Piloting

Early: Ad hoc or no pilots

The state has no formal AI pilots or sandbox programs. Any experimentation is minimal, uncoordinated, or driven by individual agencies without central guidance. AI use cases, if any, are one-off and opportunistic (e.g., a single chatbot or analytics tool tried by an agency) with little oversight. Human-in-the-loop requirements or risk assessments are absent or purely reactive. There is no structured process to evaluate outcomes or document lessons learned. Teams involved are siloed (often just IT staff) and cross-functional review is nonexistent.

Developing: Initial pilots and guidelines

The state has launched a few time-bound AI pilot projects or a basic innovation sandbox”in a controlled setting. Pilots focus on narrow, internal-facing use cases to minimize risk (e.g., automating a routine back-office task). Some human oversight is introduced (e.g., requiring a person to approve AI-generated outputs in pilots) and basic bias or risk assessments might be done on an ad hoc basis for pilot projects. There are informal mechanisms to document pilot results, but they may not yet drive decision-making. Early norms are forming (e.g., an advisory group or task force reviews pilot proposals), but practices are not yet consistent across government. Cross-functional involvement is limited but growing, often triggered by issues encountered in pilots (like legal or privacy questions).

Established: Structured piloting and learning

The state has a formal AI piloting program or sandbox environment with clear entry/exit criteria and time frames for experiments. Multiple agencies are engaged in pilots, and there is an established process to vet proposals (often by a cross-agency AI committee). Pilots include human-in-the-loop requirements by design, ensuring officials can override or audit AI decisions. Bias and risk assessments are regularly conducted before and during pilots (e.g., reviewing algorithms for fairness). Each pilot is approached with a hypothesis and success metrics, and generates qualitative and quantitative learning outcomes. There is documentation of pilot results and a mechanism (e.g., pilot reports or dashboards) to decide whether to stop, adapt, or scale the solution. Cross-functional teams (e.g.; IT, program subject matter experts, legal, ethics officers) are routinely involved in designing and evaluating experiments. An example is a state establishing an AI enablement lab, where agencies safely test use cases with data privacy protections.

Advanced: Institutionalized experimentation culture

The state has a comprehensive, ongoing AI sandbox program codified in policy or law, often open to both internal projects and public-private collaborations. Experimentation is ingrained as a norm: Any new AI application goes through a pilot phase with oversight before full deployment. The state provides dedicated safe environments (e.g., secure test datasets, cloud sandboxes) for AI development and has possibly multiple sandboxes for different domains (e.g., health, transportation). Human-in-the-loop and ethical guardrails are mandated by formal policy (e.g., an executive order or statute on AI use) and AI never operates without appropriate human accountability. Bias audits and risk assessments are mandatory and results must meet defined standards before scaling. The state demonstrates a hypothesis-driven approach (pilots are essentially applied research projects with published findings). There is a strong ability to make go/no-go decisions: Leadership will shut down pilots that show poor results or high risk and rapidly scale those that succeed. Cross-functional oversight bodies (e.g., AI councils, ethics boards) regularly review pilot portfolios. The most advanced states even share their pilot results publicly to contribute to broader learning and set up partnerships (e.g., with universities or other states). In an advanced state, one might find that 20% or more of the state workforce is engaged in an AI pilot environment, reflecting widespread, responsible experimentation.

Overall trends for Stage 2

A common pattern is that the strongest Stage 2 evidence is found where states publish formal pilot case studies (Colorado), pilot evaluation reports (Pennsylvania), or structured sandbox experimentation environments with training and feedback channels (New Jersey). Another major signal is when pilot programs are paired with workforce readiness actions—Arizona's official GenAI training announcement is a clear example. A third signal is when experimentation is framed as a supervised "trial" with specific oversight—Utah's official announcement of a first-of-its-kind regulated trial illustrates this.

Summary evaluation for Stage 2

If the first stage of AI maturity is about organizing leadership and establishing guardrails, Stage 2 is about institutional learning. It's the stage where states move from policy conversations to hands-on experimentation. Here, AI becomes less theoretical and more practical—something to test, refine, question, and evaluate within real government environments.

Across the United States, this experimentation stage is now well underway. Nearly every state has engaged in some form of AI pilot or exploratory initiative. The rapid rise of generative AI in particular lowered the barriers to entry. Tools that once required significant technical investment became accessible through secure enterprise platforms, allowing agencies to begin structured trials without building complex infrastructure from scratch. This accessibility accelerated experimentation across jurisdictions, often faster than governance systems were prepared to absorb.

Most states fall within what would be considered a Developing and Established experimentation posture. They are piloting tools, issuing provisional guidance, and allowing controlled staff use, but are

still formalizing how experimentation should occur. In many cases, experimentation is decentralized. Individual agencies explore use cases independently, often prompted by immediate workflow challenges or vendor outreach. While this approach generates energy and creativity, it can also produce uneven documentation, inconsistent risk review, and fragmented learning.

A smaller group of states has moved beyond ad hoc experimentation to begin institutionalizing innovation. These states have established AI centers of excellence, innovation labs, or formal advisory councils that serve as hubs for testing and evaluation. In these environments, experimentation is not simply allowed—it is structured. Pilot proposals are reviewed, use cases are prioritized, and risk mitigation protocols are applied before testing begins. Some states are also developing sandbox environments where AI tools can be evaluated using de-identified or synthetic data, enabling agencies to test capabilities without compromising privacy or security. These more advanced experimentation frameworks tend to produce stronger transitions into operational implementation.

One of the defining features of this stage has been the dominance of generative AI as the testing ground. Rather than beginning with complex predictive modeling systems, many states started by experimenting with internal productivity tools. Staff used AI to summarize documents, draft communications, generate policy analyses, and support administrative workflows. These use cases were often framed as augmentation rather than automation—tools to assist human decision-making rather than replace it. This framing has been important. It has allowed experimentation to proceed within established oversight structures, reducing institutional resistance and aligning with public expectations around responsible AI.

Another important national trend is collaboration. States are increasingly learning from one another, sharing policy templates, pilot lessons, and procurement guidance. Cross-state networks, academic partnerships, and vendor engagements have created an informal but active ecosystem of experimentation. Rather than isolated innovation, we are seeing distributed institutional learning. States that participate actively in these networks tend to refine their experimentation processes more quickly.

Vendor relationships also play a prominent role in Stage 2. Many pilot programs begin with vendor-provided access to AI platforms. These partnerships can accelerate learning and provide technical expertise, but they also shape experimentation priorities. States with stronger governance and internal review processes tend to extract more durable learning from vendor-led pilots, ensuring that experimentation aligns with public sector values rather than simply technological opportunity.

What distinguishes stronger Stage 2 states is not the number of pilots underway—nearly all states are experimenting—but the degree to which experimentation is structured, documented, and coordinated. States that create centralized innovation bodies, establish sandbox environments, and require documented pilot objectives are building institutional memory. They are transforming experimentation from isolated curiosity into a repeatable capability.

Importantly, Stage 2 is the hinge in the AI journey. It determines whether states remain in perpetual pilot mode or transition successfully into operational implementation. Those that treat experimentation as disciplined learning—pairing innovation with governance and documentation—are positioned to move confidently into Stage 3. Those that experiment without structure risk fragmentation and stagnation.

The pressing question now is how experimentation is organized, evaluated, and converted into lasting public value.

Stage 3: Implementation

Introduction

The key question in this domain is ”Can the government deploy AI reliably and responsibly in real operations?” Stage 3 of the AI journey is all about moving from isolated pilots to embedding AI systems into everyday government workflows. The emphasis shifts from experimentation to trustworthy execution at scale, ensuring that AI solutions deliver consistent value in production environments without causing unintended harm. At this stage, agencies are not just trying out AI—they are relying on it for real services; therefore, reliability, ethics, and oversight become paramount.

Rubric: Implementation

Early: Ad hoc or limited deployment

The state may have one or two AI systems technically “in production,” but they are narrow, isolated, and often managed at the agency level without enterprise oversight. AI tools may support back-office automation or vendor-provided systems, yet there is no structured operational governance ensuring reliability over time. Ownership of AI systems is unclear; monitoring for model drift or unintended consequences is sporadic or reactive. Procurement processes do not meaningfully account for AI-specific risks, and workforce training is informal or optional. Governance policies may exist on paper, but they are not consistently embedded in daily workflows. At this stage, AI affects operations, but the state cannot confidently demonstrate disciplined oversight or lifecycle management.

Developing: Emerging integration

The state has moved beyond pilots and is deploying AI in real workflows, often in targeted domains such as fraud detection, document review, call center automation, or eligibility screening. Operational ownership structures are beginning to form—agencies may designate system leads or program managers responsible for AI tools—but practices vary across departments. Basic monitoring mechanisms exist, such as periodic reviews or informal performance checks. Procurement and

contracting processes are beginning to incorporate AI considerations, including vendor transparency or explainability clauses. Workforce enablement improves, with staff receiving training tied to the specific AI tools they use. However, implementation remains uneven. Some systems are well-governed; others rely heavily on vendor assurances or limited oversight. The state is operationalizing AI, but it has not yet institutionalized consistent enterprise-wide practices.

Established: Structured, accountable operations

AI is now embedded in multiple programs or services, and governance is actively practiced rather than merely documented. The state maintains an inventory of operational AI systems and assigns clear ownership and accountability for each. Regular review cycles assess system performance, fairness, and reliability. Human-in-the-loop requirements are designed into high-impact workflows, ensuring that AI augments rather than replaces professional judgment. Procurement standards include AI-specific risk provisions, and vendor lifecycle management is formalized. Workforce training is structured and aligned with operational deployment, equipping staff to interpret AI outputs responsibly. Importantly, the state can articulate how AI systems are monitored, updated, and evaluated over time. AI is no longer experimental; it is managed infrastructure.

Advanced: Institutionalized and optimized implementation

AI is treated as a strategic operational asset governed through enterprise-wide portfolio management. The state integrates AI oversight into budgeting, performance reporting, and strategic planning. Lifecycle management is continuous: models are regularly audited, retrained when necessary, and stress-tested against changing conditions. Cross-functional governance bodies—often codified in statute or executive policy—review AI deployments across agencies. Operational reliability, explainability, and accountability are demonstrably embedded in workflows. Public transparency mechanisms may exist, such as AI system inventories or performance dashboards. AI systems are scaled deliberately, and leadership can clearly explain both the benefits achieved and the safeguards in place. At this stage, AI implementation reflects institutional maturity rather than isolated innovation.

Overall trends for Stage 3

Public AI inventories are among the strongest signals of operational implementation, because they document AI tools in use, responsible agencies, and sometimes purpose/decision impact. GovTech notes that at least 13 states were creating inventories/repositories to track government AI use, and highlights Vermont and Connecticut as examples with public inventories.

Operational GenAI adoption often appears first as internal productivity tooling (employee pilots, secure sandboxes, training plus controlled access), and only later as resident-facing automation—evident in states that publicly describe employee-accessible tools and pilot findings (e.g., Pennsylvania and New Jersey).

A recurring limitation is weak public documentation of "ops fundamentals" like service level agreements, drift monitoring, retraining procedures, and incident response—even where inventories or pilots exist. This often keeps states in the Established ranking rather than Advanced.

Cross-agency operational scaling is most visible where states publish multi-agency pilot results or statewide inventories, such as Arizona's nine-agency pilot write-up and Connecticut and New York open-data inventories.

Where operational deployment is publicly documented, it is frequently supported by a centralized digital/IT body (state IT department, digital services agency), which can standardize procurement/use controls and baseline reporting.

Summary evaluation for Stage 3

Stage 3 represents a pivotal moment in the AI journey for state government. If Stage 2 is about experimentation and pilots, Stage 3 marks the transition from curiosity to commitment—from testing isolated use cases to embedding AI into the operational fabric of government. It is the stage where AI begins to affect real workflows, influence decision-making, and deliver measurable improvements in efficiency and service delivery.

Across the United States, most states are now somewhere within this operational transition. Very few remain purely in early experimentation, and only a small group have achieved fully scaled, enterprise-wide AI transformation. The majority are clustered between Developing and Established, meaning they have moved beyond pilots and are running production-level AI systems in select agencies, but have not yet institutionalized AI as a cross-government capability.

In practice, Stage 3 maturity often looks like this: Agencies are using machine learning models to detect fraud in benefits systems, generative AI tools are assisting staff with drafting and summarizing documents, predictive analytics is improving transportation planning or public health modeling, and call centers are deploying AI-powered triage tools. These systems are no longer hypothetical or proof-of-concept; they are operating in real environments with real consequences.

However, the degree of integration varies significantly. In many states, AI deployments remain concentrated in data-rich agencies, such as health and human services, labor, revenue, or transportation. Scaling across agencies—particularly into citizen-facing transformation—is less common. Operational AI today is still more often about improving internal efficiency than redesigning the public experience.

The operational transition: From pilot to production

One of the clearest national trends is the movement from pilot programs to production deployments. In 2023, much of the activity across states consisted of exploratory task forces, sandbox experiments, and narrowly scoped use cases. By 2025, many of those pilots had either matured into operational systems or informed structured governance and procurement frameworks.

Generative AI has played a particularly catalytic role in accelerating this shift. Unlike earlier waves of predictive analytics, generative AI tools are relatively easy to deploy and require less specialized infrastructure at the outset. As a result, states have been able to introduce enterprise versions of large language model platforms to assist staff with drafting, summarizing, coding, research, and administrative processing. These deployments often begin as productivity experiments but quickly evolve into embedded workflow tools.

Yet even as generative AI expands, most states remain cautious. Deployment is frequently accompanied by usage guidance, guardrails, and limited access tiers. Operationalization is happening—but deliberately.

Efficiency first, transformation later

Another consistent pattern across states is the prioritization of efficiency gains over full service redesign. The first operational AI systems tend to address backlog reduction, document processing, fraud detection, eligibility triage, and internal workflow automation. These use cases offer relatively clear return on investment and can be implemented without fundamentally altering public-facing systems.

More ambitious transformations—such as AI-enabled case management systems, personalized benefits navigation, or automated adjudication support—are emerging more slowly. These require stronger infrastructure, more robust governance, and greater public trust.

In other words, Stage 3 nationally is characterized more by operational optimization than systemic transformation.

Infrastructure as the great divider

A strong correlation has emerged between technical infrastructure maturity and the ability to scale AI operationally. States with consolidated data platforms, cloud-first strategies, centralized IT governance, and strong data interoperability have progressed more quickly into Established or Advanced operational maturity.

Where enterprise data environments exist, agencies can deploy models more safely, share tools across departments, and integrate AI into legacy modernization efforts. Where infrastructure remains siloed or outdated, AI tends to remain fragmented and agency-specific.

This infrastructure divide is widening the gap between leading and lagging states. Leaders are building reusable platforms and AI enablement centers; lagging states are still managing isolated pilots without systemic support.

Governance and centers of excellence as accelerators

States that have established AI centers of excellence, enablement labs, or structured cross-agency councils are scaling operational AI more effectively. These structures reduce duplication, provide standardized risk assessment processes, and create shared procurement pathways. They also provide a bridge between governance and implementation—ensuring that policy frameworks are not detached from operational reality.

In these states, AI deployment is no longer just an agency initiative; it is becoming an enterprise capability.

A nation in transition

Taken together, Stage 3 reflects a nation in operational transition. AI is no longer theoretical in state government. It is being used daily by public employees, embedded in fraud detection systems, influencing case workflows, and reshaping administrative processes. At the same time, full enterprise transformation remains rare.

The trajectory is clear: Over the next 12 to 24 months, we are likely to see more states formalize enterprise AI platforms, integrate AI into legacy modernization initiatives, and begin demanding stronger performance metrics. The leaders are already moving in this direction. Others are building the foundational governance and infrastructure necessary to follow.

Stage 3, then, is best understood as the bridge between experimentation and measurable transformation. It is where AI begins to matter operationally—but not yet systemically. The next stage will depend not only on technological capability, but on whether states can pair implementation with disciplined learning, accountability, and public value measurement.

Stage 4: Impact

Introduction

By Stage 4 of the AI Journey, a state has moved beyond readiness, experimentation, and operational implementation. AI systems are functioning within real government environments. The central question is no longer “Can we deploy AI?” but rather: “Is AI delivering measurable public value—and how do we continuously improve it?”

Stage 4 represents the shift from implementation to impact. This stage is defined not by the number of AI systems deployed, but by whether the state has built the structures, culture, and discipline necessary to measure outcomes, learn from results, and adapt over time.

Importantly, the publicly available data on the impact of AI projects is very limited. Therefore, this report uses capacity to evaluate impact as a proxy when this data doesn't exist.

Rubric: Impact

Early: Limited impact visibility

At the earliest stage of impact maturity, AI systems may be alive, but their broader effects are not systematically measured. Agencies may observe operational improvements—such as time savings or workflow efficiencies—but these insights are informal. There are no defined success criteria tied to public outcomes, equity considerations, or long-term value. Evaluation is reactive. Issues are addressed when they arise, but there is no structured review cycle. AI performance is not connected to budget decisions, procurement adjustments, or policy refinement. At this level, the state has implemented AI but cannot yet clearly demonstrate its impact.

Developing: Emerging impact measurement

As states mature, they begin to introduce structured reporting and performance tracking for AI systems. Agencies document metrics such as cost savings, accuracy rates, service improvements, or user satisfaction. Governance bodies may require periodic updates. AI inventories begin to connect with evaluation requirements. The focus shifts from “Does this system work?” to “How well does it perform, and who benefits?” Learning becomes more intentional. Pilot results inform scaling decisions. Some unintended consequences are documented and addressed. However, evaluation processes may still vary across agencies, and feedback loops are not yet fully embedded into strategic planning. Impact measurement exists—but it is still evolving toward consistency and depth.

Established: Institutionalized impact and accountability

At the Established level, impact measurement becomes systematic and recurring. AI systems are evaluated against defined public sector performance criteria, which may include:

- Operational efficiency
- Fiscal return on investment
- Service delivery improvements
- Accuracy and error rates
- Equity and bias mitigation
- Risk reduction

Evaluation is not optional. It is embedded in governance. Agencies are required to report on AI system performance at regular intervals. Findings influence decisions about scaling, redesigning, or discontinuing systems. Importantly, impact data begins to shape policy. Lessons learned inform procurement standards, risk frameworks, and workforce training. At this stage, the state can articulate not only where AI is deployed, but what measurable outcomes it is achieving.

Advanced: Adaptive learning and continuous optimization

At the most mature stage, the state treats AI systems as dynamic tools requiring continuous calibration. Impact measurement becomes ongoing rather than periodic. Monitoring mechanisms identify performance drift, emerging bias, or changing conditions. Feedback loops allow agencies to update models, refine processes, and improve outcomes iteratively. Evaluation expands beyond operational metrics to assess broader public value:

- Are communities experiencing improved access to services?
- Are disparities being reduced?
- Is public trust strengthened?
- Are unintended harms being prevented or mitigated?

Learning is shared across agencies. Insights from one deployment inform others. Impact findings are transparent and publicly communicated where appropriate. Governance adapts based on evidence. Policy frameworks evolve as lessons accumulate. Budget allocations reflect demonstrated outcomes.

The core shift in Stage 4

The defining feature of Stage 4 is not measurement alone—it is the integration of measurement into continuous learning. At this level, AI governance operates as a living system:

Deploy → Measure → Learn → Adapt → Improve → Repeat

The most advanced states understand that AI performance is not fixed at deployment. Models degrade. Context changes. Public expectations evolve. Stage 4 is where government moves from adopting AI to governing it intelligently over time.

Overall trends for Stage 4

Stage 4 of the AI journey examines whether states have moved beyond AI experimentation and deployment to systematically measuring outcomes, learning from results, and refining governance and implementation over time.

Across all 50 states and Washington, D.C., the evaluation shows that while AI governance activity is widespread, true continuous learning infrastructure remains concentrated in a small group of states. Most states now have one or more of the following:

- Executive orders or enterprise AI policies
- Task forces or advisory councils
- Acceptable-use guidance for generative AI
- Initial AI inventories or legislative reporting requirements

However, Stage 4 maturity requires more than policy adoption. It requires demonstrable evidence of the following metrics in a benefits delivery context:

- Quantified impact measurement (e.g., return on investment, cost savings, service improvements)
- Recurring reporting cycles
- Monitoring mechanisms (inventories, audits, dashboards)
- Feedback loops that adjust procurement, policy, or models based on findings
- Cross-agency learning and transparency

Summary evaluation for Stage 4

Several clear patterns emerge across the national landscape.

Governance is outpacing measurement

Nearly every state has initiated AI policy conversations, and many have adopted executive orders or acceptable-use frameworks. As noted in the AI Readiness metrics in Stage 1, governance scaffolding—especially around generative AI—has expanded rapidly since 2023.

However, performance measurement systems have not matured at the same pace. The majority of states are still focused on defining guardrails rather than building longitudinal evaluation mechanisms.

The result is a governance-heavy, measurement-light ecosystem in many jurisdictions.

Inventories are becoming the backbone of Stage 4

States that demonstrate higher Stage 4 maturity frequently rely on recurring automated decision system inventories or AI system registries.

Inventories serve three essential functions:

1. They create visibility into where AI is used.
2. They establish recurring reporting requirements.
3. They provide a platform for later performance and risk assessment.

Washington, Vermont, Connecticut, and New York illustrate how inventories can become the structural backbone for continuous learning.

Inventory-driven governance appears to be the most scalable pathway toward advanced Stage 4 maturity.

Pilots are measured, enterprise systems less so

Several states (e.g., Colorado, Pennsylvania) have produced high-quality pilot evaluation reports with quantifiable outcomes and structured feedback collection.

However, enterprise-wide dashboards or recurring statewide performance reporting remain rare. Pilot-level rigor does not always translate into systemic performance management.

This creates a gap between measured experimentation and measured implementation at scale.

Financial impact reporting is a differentiator

States in the Established tier frequently quantify:

- Cost savings
- Fraud detection recovery
- Efficiency gains
- Staff-hours saved
- Service-level improvements

North Carolina stands out for reporting measurable fiscal impact at scale and embedding audit and monitoring processes that reinforce learning over time.

Publicly articulated ROI appears to be a strong indicator of Stage 4 maturity.

Continuous monitoring is emerging but not yet standard

Only a handful of states demonstrate evidence of:

- Ongoing monitoring for system drift
- Automated audit or alert systems
- Regular legislative reporting tied to analytics outcomes
- Structured model or policy updates based on findings

Most states operate in annual or ad hoc review cycles rather than continuous performance tracking. The shift from periodic evaluation to adaptive, real-time learning remains the defining frontier of Stage 4.

Transparency and public trust are becoming central

Higher-maturity states increasingly treat transparency as part of performance management. Public inventories, annual reports, and published pilot findings signal a recognition that accountability is integral to sustained AI adoption.

The link between transparency and adaptive governance is becoming more visible nationally.

Strategic implications

The national landscape suggests that states are transitioning from AI adoption to AI institutionalization, but only a small cohort has reached the level where AI systems are continuously measured, refined, and aligned with measurable public value.
The pathway to advanced Stage 4 maturity appears to involve:

1. Formalized AI inventories
2. Recurring, statutory or executive reporting requirements
3. Quantified impact metrics tied to budget or policy decisions
4. Audit and monitoring mechanisms
5. Cross-agency knowledge sharing
6. Transparent publication of findings

States that combine these elements demonstrate the clearest evidence of adaptive learning.

The AI road ahead for state government

State governments are poised for significant advancements in AI readiness over the next year, influenced by both industry developments and internal policy shifts. Here are some of the driving trends, and our predictions for the next year.

- As states implement agentic AI tools in the near future, the baselines for readiness could shift dramatically, impacting subsequent stages of this rubric.
- More states will formalize governance structures, launch broader and often mandatory workforce training initiatives, and expand sandbox and testing environments.
- We'll see states move beyond isolated pilots toward coordinated experimentation portfolios. We will likely see a higher volume of successful pilots transition into operational programs, while agencies refine procurement strategies, technical standards, and governance practices based on lessons learned.

- We are likely to see more states formalize enterprise AI platforms, integrate AI into legacy modernization initiatives, and begin demanding stronger performance metrics.
- States that scale their AI implementations will be better able to articulate the impact of AI through clear evaluation frameworks that track efficiency gains, service improvements, and public value created by AI systems.
- Leading government agencies will start institutionalizing feedback loops that inform procurement, system design, and policy adjustments, enabling a cycle of continuous learning and responsible AI optimization.

Looking for state-specific insights?

Explore detailed assessments, comparisons, and interactive data for each state: **codeforamerica.org/ai-landscape**