CENTER FOR DIGITAL GOVERNMENT

The New Data Paradigm

Revamping Government Data Infrastructure for the Future

Steve Towns
Center for Digital Government

The **Imperative**: Why Modernize Now?

Exploding Data Volumes

Government agencies face unprecedented growth in data from diverse sources.

Rising Citizen Expectations

Residents demand seamless digital services and personalized experiences.

Escalating Cyber Threats

The threat landscape grows more sophisticated and dangerous daily.

Budget & Workforce Pressures

Agencies must do more with constrained resources and staffing challenges.

Session Objectives

Analyze Today's Landscape

Understand current government data challenges and opportunities across agencies.

Explore AI Intersection

Examine the critical relationship between data infrastructure and AI applications.

Envision Future Trends

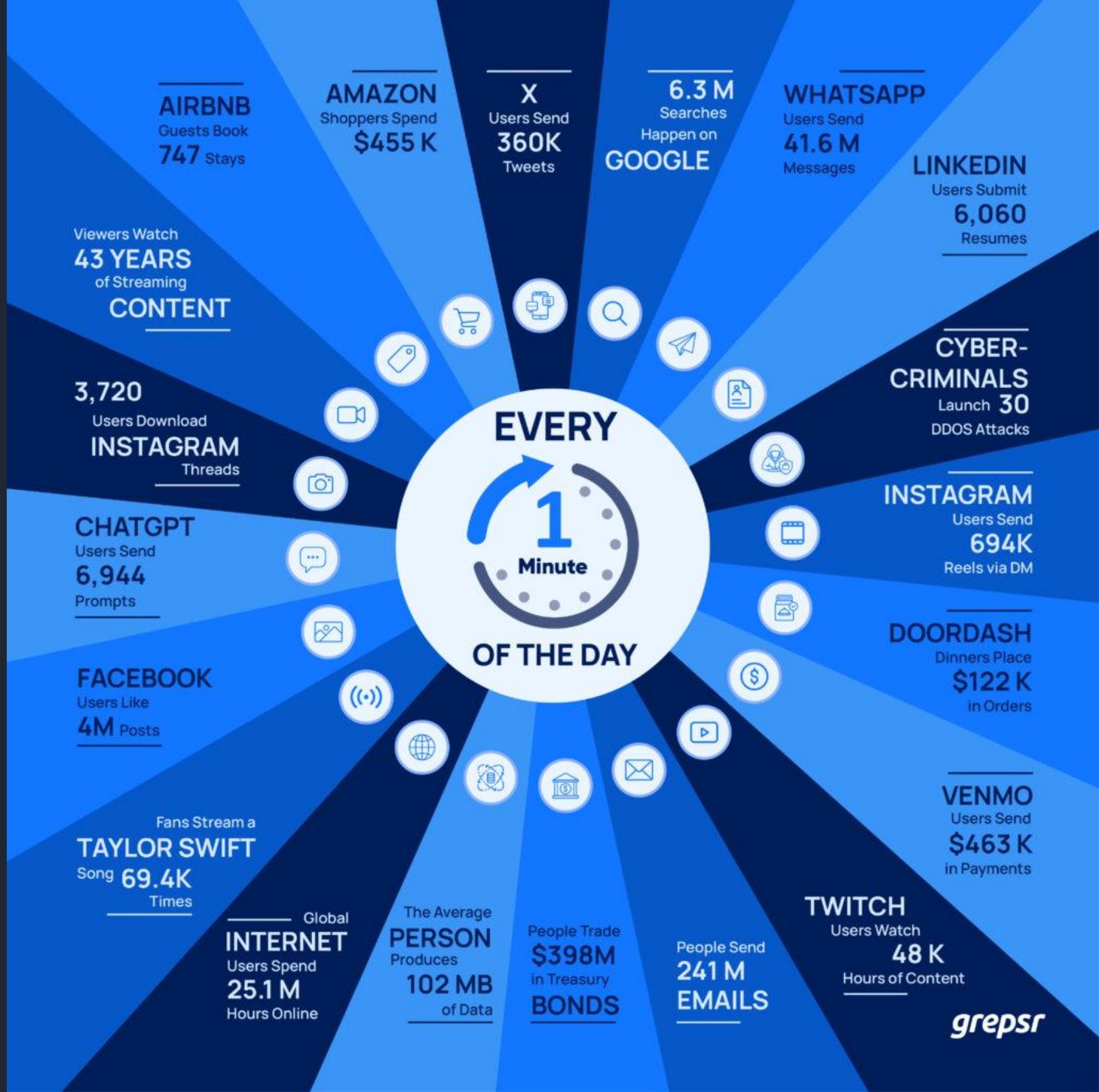
Identify key developments that will shape government data environments.

Outline Modernization Tactics

Provide actionable strategies for transforming your data infrastructure.

The Current Reality: Data Overload

We Create Lots of Data, Every 60 Seconds.



We Estimate That State & Local Governments Create 4.5-5 Zettabytes of Data Per Year.

21 zeroes!

1 zettabyte = 1,000,000,000,000,000,000,000 bytes

To Put That Into Perspective, Let's Look At The Library of Congress.

- Total Items: Over 178 million items.
- Cataloged Books & Print Materials: Over 38.6 million.
- Manuscripts: More than 70 million.
- Rare Books & Special Collections: Over 800,000 items.
- Maps: More than 5.5 million.
- Sheet Music: More than 8.2 million.
- Sound Recordings: More than 3 million items.
- Moving Images: 1.8 million items.

State and Local Governments Create 4,400 to 5,000 Libraries of Congress' Worth of Data Per Year (4.5-5 Zettabytes)

But What Are Agencies Doing With This Data?

Most Common Use Cases

Top Use Case Category	Description	Key Applications
Fraud Detection & Prevention	Applying advanced analytics and AI/ML to identify, predict, and prevent fraud, waste, and abuse in government programs and services.	Benefits programs (unemployment insurance), payment integrity, identity verification, general government fraud/waste/abuse reduction.
Enhanced Decision Making & Planning	Leveraging predictive analytics, simulations, and decision intelligence platforms for better policy, resource allocation, and operational planning.	Public health response, supply chain resilience, procurement optimization, urban planning (spatial simulation), environmental monitoring, resource allocation, policy impact analysis.
Improved Citizen Experience (CX)	Using data analytics and AI to understand citizen needs, personalize interactions, improve service delivery, and provide proactive information.	Understanding citizen journeys, personalized portals/communication, AI chatbots for service, proactive information dissemination, analyzing CX quality/feedback.

Most Common Use-Cases

Top Use Case Category	Description	Key Applications
Data Governance, Security & Privacy	Establishing robust frameworks, policies, and tools for data quality, security, ethical AI use, regulatory compliance, metadata management, and privacy enhancement.	Data quality frameworks, security protocols, ethical AI guidelines, regulatory compliance (e.g., privacy laws), metadata catalogs/lineage, synthetic data for privacy, quantum encryption preparation.
Real-time Monitoring & Response	Employing data dashboards and analytics for real-time insights, situational awareness, and rapid response in various government functions.	Public health surveillance, emergency management coordination, infrastructure monitoring (e.g., traffic, utilities), operational performance tracking, environmental monitoring (real-time).

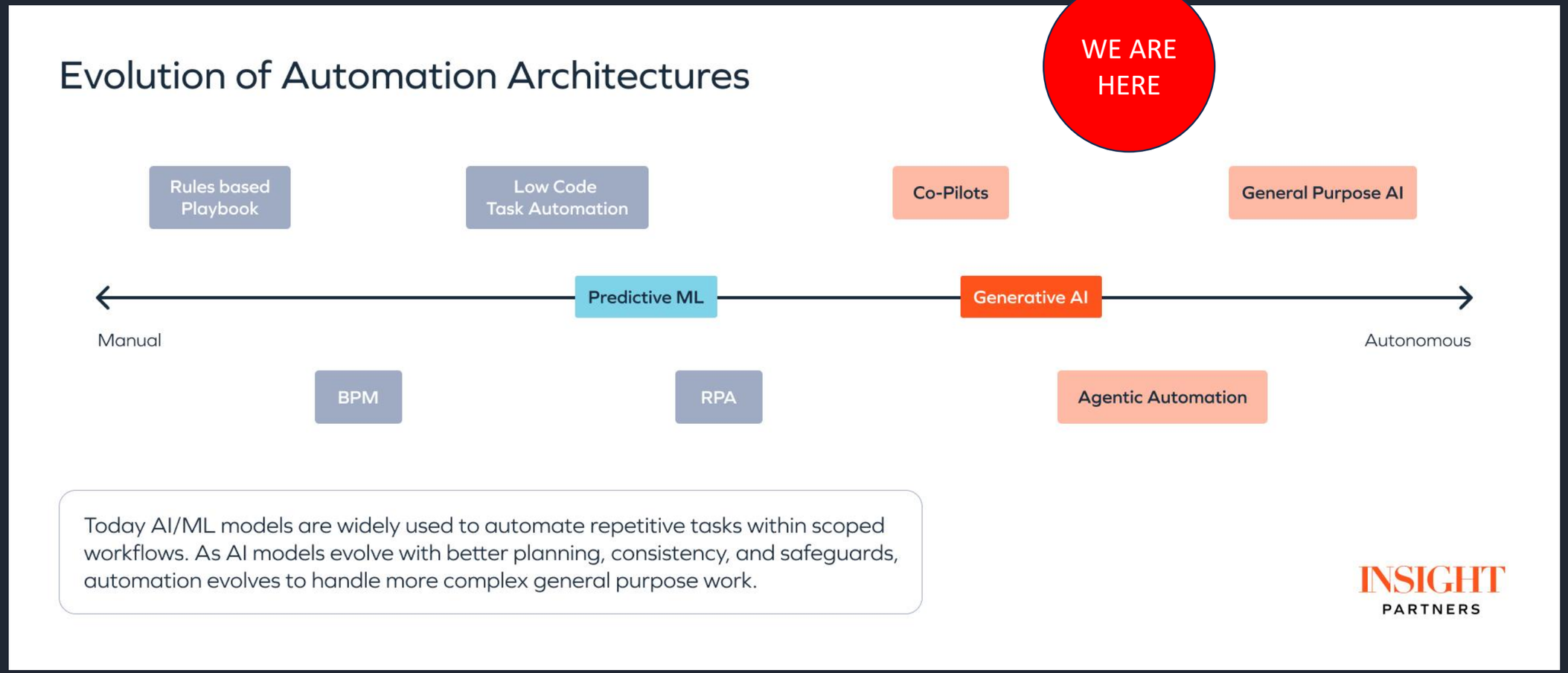
But What About AI?

Welcome to the AI Era

AI-Driven Automation & Efficiency

Utilizing AI/ML and agents to automate routine tasks, constituent services, and IT operations, improving overall efficiency.

Administrative task automation (reports, forms, grants), automated constituent services (chatbots, NLP), IT operations (code generation, testing), core system task automation.



Where Data and AI Intersect

Data Feeds AI

High-quality, accessible data is the foundation for AI success.

AI Enhances Data

AI analyzes and extracts value from complex datasets.

Governance Enables Both

Strong frameworks ensure trustworthy AI and data.

Continuous Improvement

The relationship creates a virtuous cycle of value.

Bad Data In = Bad Data Out

But This Data Isn’t Always Clean

Disparate Systems

Government data exists in different formats and on disconnected platforms.

Integration Obstacles

Existing systems are not easy to integrate or built with the same language.

Limited Visibility

Fragmentation hinders comprehensive view of bigger data picture.

Workforce Barriers

Lack of available staff makes it difficult to navigate complex data decisions.

By 2027, 60% of organizations will fail to realize the anticipated value of their AI use cases due to ***incohesive data governance frameworks.***

Source: Gartner / 2024

It All Starts With Governance

Control & Sovereignty

Maintain agency oversight of AI systems and data

Ethics & Compliance

Address ethical concerns and regulatory requirements

Traceability & Transparency

Ensure decisions can be explained and verified

Data Quality & Integrity

Provide clean, accurate data for AI systems

This Forces A New Conversation On Infrastructure: The GenAI Tech Stack

GenAI Public Sector Tech Stack

The Existing Government Tech Stack

Application Layer

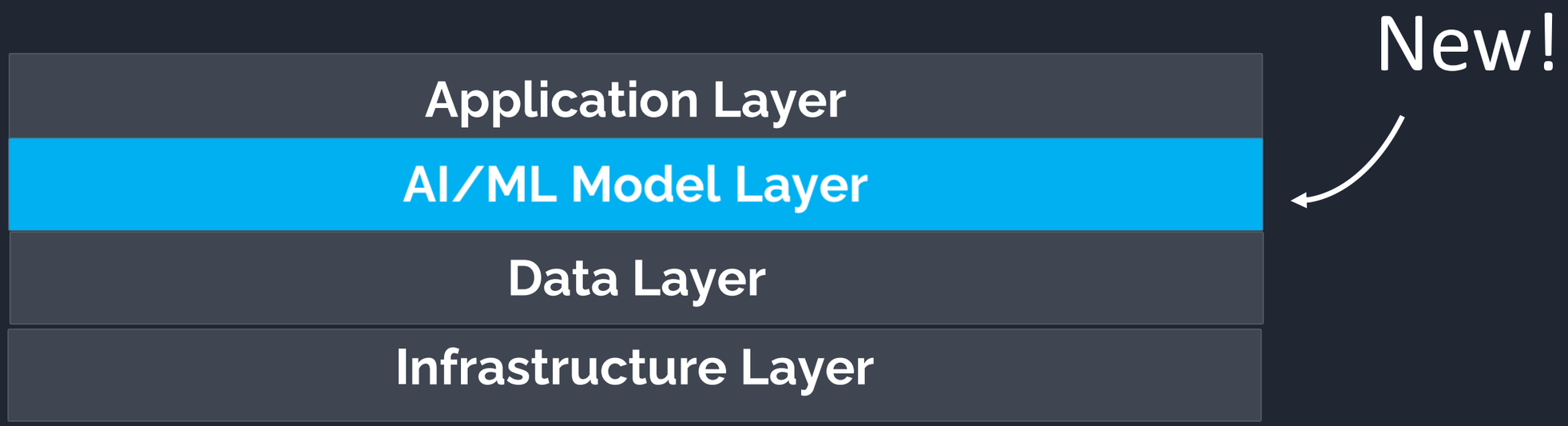
Data Layer

Infrastructure Layer

Source: CPSAI GenAI Tech Stack v1

GenAI Public Sector Tech Stack

The Evolved Government Tech Stack



Source: CPSAI GenAI Tech Stack v1

The Evolved **Government AI** Tech Stack – *Simplified*

Layer	What It Does	Government Example
Application Layer	End-user tools using AI	City 311 Chatbot powered by ChatGPT or Gemini
AI/ML Model Layer	Pre-trained or custom AI models	State DMV using GPT-4 to summarize form submissions
Data Layer	Stores, catalogs, and prepares data	Health Dept. building a catalog of public health datasets
Infrastructure Layer	Provides compute/storage	Agency using AWS or Azure for cloud-based AI workflows

Infrastructure

Layer	Description	Examples of Companies
Compute Layer	Computational resources and hardware for AI operations	AWS, Google Cloud, Microsoft Azure, NVIDIA, AMD, Intel, Edge Impulse, FogHorn, Oracle's OCI GenAI, CoreWeave, watsonx
Frameworks and Libraries	Software tools providing functions for building AI models	TensorFlow, PyTorch, MXNet, Scikit-learn, Keras, OpenCV
Deployment and Monitoring	Tools for deploying and monitoring AI models	Kubernetes, Docker, TensorFlow Serving, MLflow, Seldon, Tecton, MLOps tools
Security and Compliance	Ensuring AI models and data comply with regulations and are secure	Snyk, Dataminr, Aqua Security, Immuta, OneTrust, BigID, TrustArc

Source: CPSAI GenAI Tech Stack v1

Data (Collection/Management) Layer

Layer	Description	Examples of Companies
Data Management and Annotation	Tools for collecting, managing, and annotating data	Snowflake, Datadog, Splunk, Labelbox, Scale AI, Appen, DataRobot, Dataiku

New! Model Layer

Layer	Description	Examples of Companies
Pre-trained Models and APIs	Pre-built models and APIs for integrating AI capabilities	OpenAI (GPT-4o), Hugging Face, IBM Watson AI, Google Cloud AI, Microsoft Cognitive Services, Watsonx (IBM), Cohere, AI21 Labs, Anthropic
Model Development and Training Platforms	Platforms for creating, training, and fine-tuning AI models	Google Colab, Jupyter Notebooks, Kaggle, H2O.ai, Databricks, Paperspace, Watsonx (IBM), AWS Bedrock
Hyper Local Models	AI models developed for localized applications	Amperity, Synthesia, Farmers Edge

Source: CPSAI GenAI Tech Stack v1

New! Model Layer

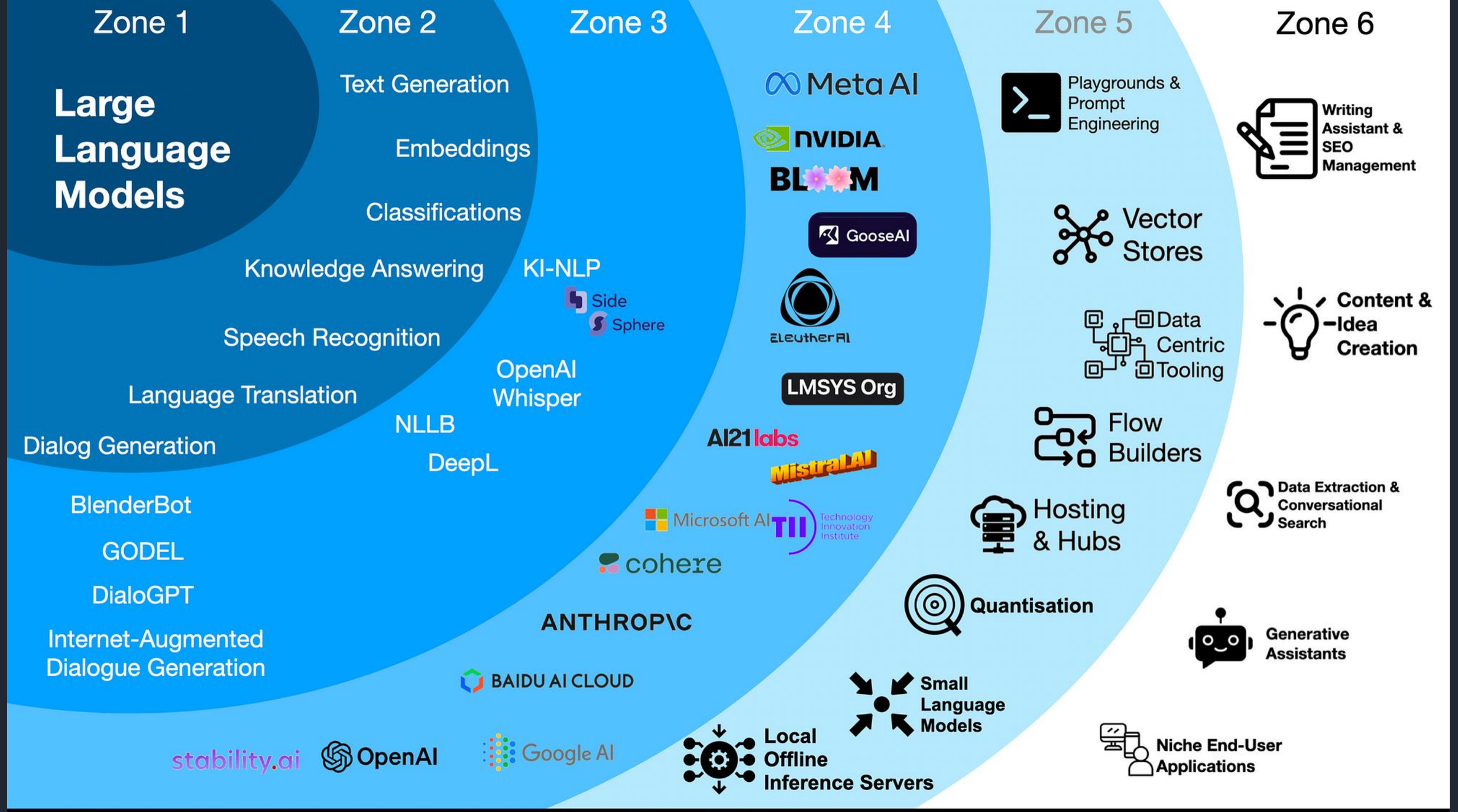
Layer	Description	Examples of Companies
Fine-Tuned Models	Customized versions of foundational models	(Customized versions of foundational models)
Foundational Models	Large-scale models serving as the base for further training	GPT-4o (OpenAI), BERT (Google), T5 (Google), Megatron (NVIDIA), Watsonx (IBM), Oracle's GenAI, AWS Titan, Anthropic, AI21labs, LLaMA, PaLM, Huggingface

Source: CPSAI GenAI Tech Stack v1

Application Layer

Layer	Description	Examples of Standalone Technologies
Application Layer	End-user applications leveraging generative AI	ChatGPT (OpenAI), IBM Watson Assistant, Google Gemini, Jasper.ai, Copy.ai, Grammarly, DALL-E (OpenAI), RunwayML, Synthesia, MidJourney, GitHub Copilot, Anthropic Claude

Source: CPSAI GenAI Tech Stack v1



Version 5

www.cobusgreyling.com

What's Coming Next?

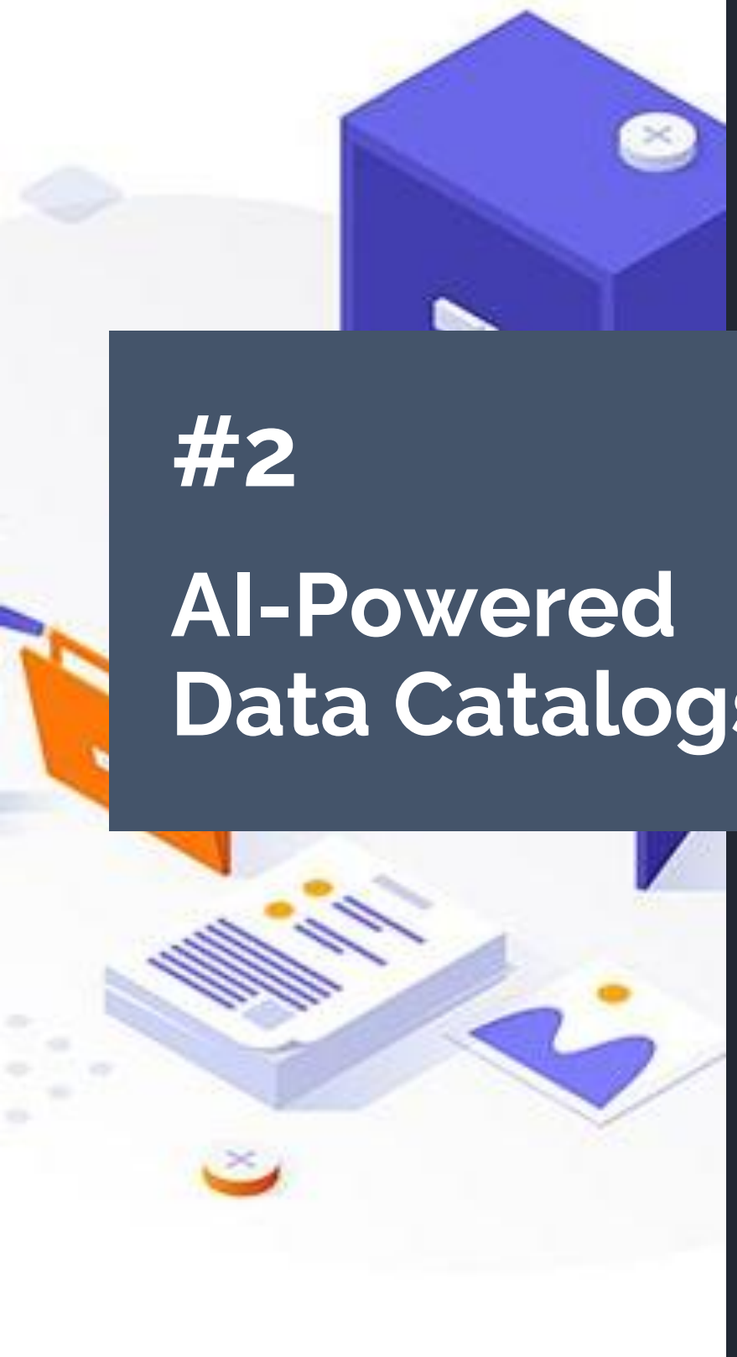
Top 10 Data Predictions
For State & Local Government

#1

Data Processing & Intelligence Will Grow at the Edge

- **Future Prediction:** Localized AI and smarter edge IoT devices (e.g., traffic cameras, utility infrastructure, etc.) will transform edge nodes into critical endpoints. These will require localized governance frameworks to manage real-time processing, privacy, and security at the source.

- **Approaches:** Develop localized governance policies for edge devices, emphasizing secure data processing and consistent integration with centralized systems.



#2

AI-Powered Data Catalogs

- **Future Prediction:** AI-driven data catalogs will increasingly automate metadata management, helping agencies keep pace with ever-expanding datasets.
- **Approaches:** Build agency-wide cataloging systems that auto-classify data and track lineage for improved data accuracy and compliance.

#3

Increasingly Fragmented Regulations

- **Future Prediction:** As regulatory environments continue to evolve across states and sectors, public agencies will need to navigate a complex set of regulations impacting data, privacy, and specific AI use cases.

- **Approaches:** Implement adaptive compliance monitoring systems that align with evolving data privacy and AI regulations.

#4

The Rise of New Privacy-Enhancing Technologies (PETs)

- **Future Prediction:** Confidential computing and similar technologies will become essential (and required) to secure sensitive data, enabling new forms of collaboration across agencies and sectors.
- **Approaches:** Explore PET technologies like confidential computing to process sensitive data securely in shared environments, such as joint public-private initiatives.

#5

AI Literacy & Change Management

- **Future Prediction:** Workforce constraints will persist, but AI will enable hyper-personalized reskilling and upskilling, democratizing access to essential knowledge.

- **Approaches:** Integrate AI-driven learning platforms to tailor training programs for staff based on their specific roles and gaps. Incorporate data/AI literacy and change management components into procurements for industry partners.

#6

API Governance Is the New Platform Approach

- **Future Prediction:** AI Agents and Agentic AI will revolutionize the way public sector services are delivered and create the need for new levels of assurance around identity.
- **Approaches:** Create governance frameworks that oversee API usage, access controls, and performance metrics for AI agents and third-party integrations.

#7

Data Ownership Becomes More Gray

- **Future Prediction:** As new data sources emerge and consumer privacy laws tighten, determining ownership at various data layers will become a critical challenge. States are expected to enact stricter privacy laws, particularly for AI-generated data.

- **Approaches:** Establish clear data stewardship and rights frameworks, focusing on transparency and individual ownership.

#8

Self-Orchestrating Data Governance

- **Future Prediction:** Shadow AI and shadow applications will proliferate, requiring adaptive governance that can respond dynamically to unauthorized data flows and applications.
- **Approaches:** Deploy automated governance systems capable of detecting and managing shadow applications in real-time.

#9

Adapted Governance for Synthetic Data & AI Model Finetuning

- **Future Prediction:** Agencies will be required to take proactive steps to remove bias and enforce ethical standards in their data and the vendors they work with. Agencies will also need to address the complexities of where and how synthetic data are used in operations.

- **Approaches:** Embed and enforce ethical standards in your own data and across partners that also address the complexities of synthetic data. Leverage industry models from NIST, International Best Practices, etc.

#10

Everyone Believes They Are A Data Scientist & Data Expert

- **Future Prediction:** As Generative AI democratizes data science, public sector employees at all levels will become more confident in its outputs.

- **Approaches:** Remind staff that just because they have spelling and grammar checks in Word, does not mean they have a degree in linguistics or journalism.

Building The Right Foundation

AI-Era Data Procurement Considerations

- Update procurement for fast-moving AI capabilities
- Require AI transparency: data sources, models, bias handling
- Use agile, modular contracts that evolve with tech
- Ensure alignment with privacy laws and ethics

The Vision: A Future-Proof Government Data Foundation

Interoperable	Scalable	Secure	Intelligent
Data flows seamlessly and securely across systems	Infrastructure adapts to changing demands	Resilient against evolving threats	Leverages data and AI for insights
• Standard APIs and data formats	• Cloud-based resources	• AI-powered defense	• Predictive analytics
• Unified access controls	• Elastic computing capacity	• Zero-trust architecture	• Automated decision support
• Cross-agency sharing capabilities	• Flexible storage solutions	• Automated compliance checks	• Continuous process optimization

With Embedded Modern Data Governance

Strategic Value
Data as a strategic asset

Roles & Responsibilities
Clear ownership and stewardship

Policies & Standards
Guidelines for quality and security

Tools & Technology
Catalogs, quality, and integration tools

Data Literacy
Education programs for all staff

What About The Legacy Stuff?

- **API Bridges** - Connect legacy systems to modern apps without full replacement
- **Data Lakes & Lakehouses** - Consolidate data to support AI and analytics readiness
- **Modular Modernization (Hybrid Cloud)** - Modernize in parts using flexible cloud models alongside legacy
- **Incremental Modernization Examples** - Wrap legacy apps with APIs
- **Migrate workloads to containers** - Introduce low-code platforms

Where To Go From Here

Acknowledge Challenges

Recognize current data infrastructure limitations and their impact.

Envision the Future

Build a clear vision for your modernized data ecosystem.

Take Strategic Action

Implement practical modernization tactics with a phased approach.

Collaborate & Learn

Share experiences and best practices across agencies.

CENTER FOR DIGITAL GOVERNMENT

Thank You!

Government Technology: govtech.com
Center for Public Sector AI: govtech.com/cpsai
Center for Digital Government: govtech.com/cdg

Steve Towns
Center for Digital Government