
AI & Data
Strategic Considerations
FHWA CLAS Innovation Exchange
May 21, 2026
U.S. Department of Transportation
ITS Joint Program Office
Source: iStock.com/Vertigo3d/cherzoff

Disclaimer

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturer's names appear in this presentation only because they are considered essential to the objective of the presentation. They are included for informational purposes only and are not intended to reflect a preference, approval, or endorsement of any one product or entity.

Except for any statutes and regulations cited, the contents of this presentation do not have the force and effect of law and are not meant to bind the States or the public in any way. This presentation is intended only to provide information regarding existing requirements under the law or agency policies.

AI and Transportation

U.S. Department of Transportation
ITS Joint Program Office

Image Source (upper left): iStock/Leonardo Patrizi; iStock/metamorworks
Image Source (lower left): iStock/buzbuzzer; iStock/Shaxiaozi
Image Source (upper right): iStock/Tessa Munson; iStock/metamorworks
Image Source (lower right): iStock/TommL; iStock/metamorworks

3

AI presents a revolutionary solution to some transportation challenges by providing:

- more timely and robust processing of complex data;
- delivery of more precise insights and predictions; and
- advanced, cost-effective, and automated capabilities.

There are many uses for AI in transportation.

AI can **perceive** the environment and collect better data:
- Sensor data fusion
- Road asset and road user classification
- Pavement condition assessment
- Road weather condition identification
- Performance monitoring (speeds, volumes)

AI can **predict** future conditions:
- Traffic prediction (for short-term operations)
- Crash prediction

- Transit arrival time prediction
- Demand forecasting (for long-term planning)
- Maintenance prediction

AI can **act** or intervene directly to improve mobility and safety:

- Warnings and alerts for hazards
- Automated fleet dispatch
- Adaptive signal control
- Dynamic tolling
- Automated maintenance scheduling

- What is ITS?
    - ITS improves transportation by integrating advanced information and communications-based technologies (ICT) into transportation infrastructure and vehicles. ITS refers to a system of technologies and operational advancements that, when combined and managed, improve the capabilities of the overall transportation system.
- Mandated by United States Code Title 23 Chapter 5 and subject to provisions outlined in Sections 512-518, the Secretary of Transportation is directed to conduct an ongoing intelligent transportation system (ITS) program to research, develop and operationally test ITS and provide technical assistance in the nationwide application of ITS as a component of the surface transportation system (23 USC Sect 515).
- To meet that directive, the Intelligent Transportation Systems Joint Program Office (ITS JPO) **collaborates across modes** (highways, transit, freight, rail, etc.) to coordinate and plan the U.S. DOT's multimodal ITS technology research program, working toward improving transportation **safety, mobility, efficiency, and enhancing productivity** through the integration of innovative technologies into the

nation's transportation system.

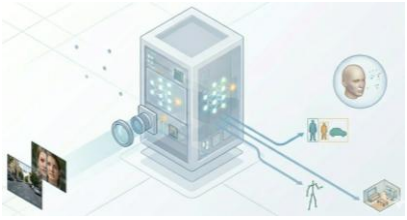
- What we do:
    - Support **adoption** of new technology
    - Encourage innovation through **testing** and **deployment**
    - Provide **technical assistance** in ITS

Other ITSJPO Program Areas (in addition to AI)

- Automation
- V2X , Next Generation Communications, and Interoperability
- AI for ITS
- Strategic Technology for Roadway Safety
- Professional Capacity Building (PCB)
- Deployment Evaluation
- ITS Architecture, Security, and Standards
- ITS4US Deployment

Check out the ITS JPO Website at: https://www.its.dot.gov/

What is AI?



Computer vision (IMAGES)

Machine learning (NUMBERS)

Generative AI (LANGUAGE)

Image Source: Gemini

U.S. Department of Transportation
ITS Joint Program Office

5

Modern AI is defined by its reliance on machine learning, the identification of patterns in data.

There are many credible definitions of AI floating around, including one that we touched on slide 3: when machines can learn, sense, reason, and act with human-like cognition. We won't address the definitions further here.

Practically speaking, it is helpful to think of AI applications in transportation being based on one or more of these three types of AI:
Machine learning – identifies patterns in numbers
Computer vision – identifies patterns in images
Generative AI – identifies and uses patterns in language.

AI Program

Transforming

Safety | Mobility | Efficiency | Resiliency

by

Developing Enabling Tech | Deployment Testing | Driving Adoption

Image Source: Gemini

AI Program: Resources

1 Get AI-Ready

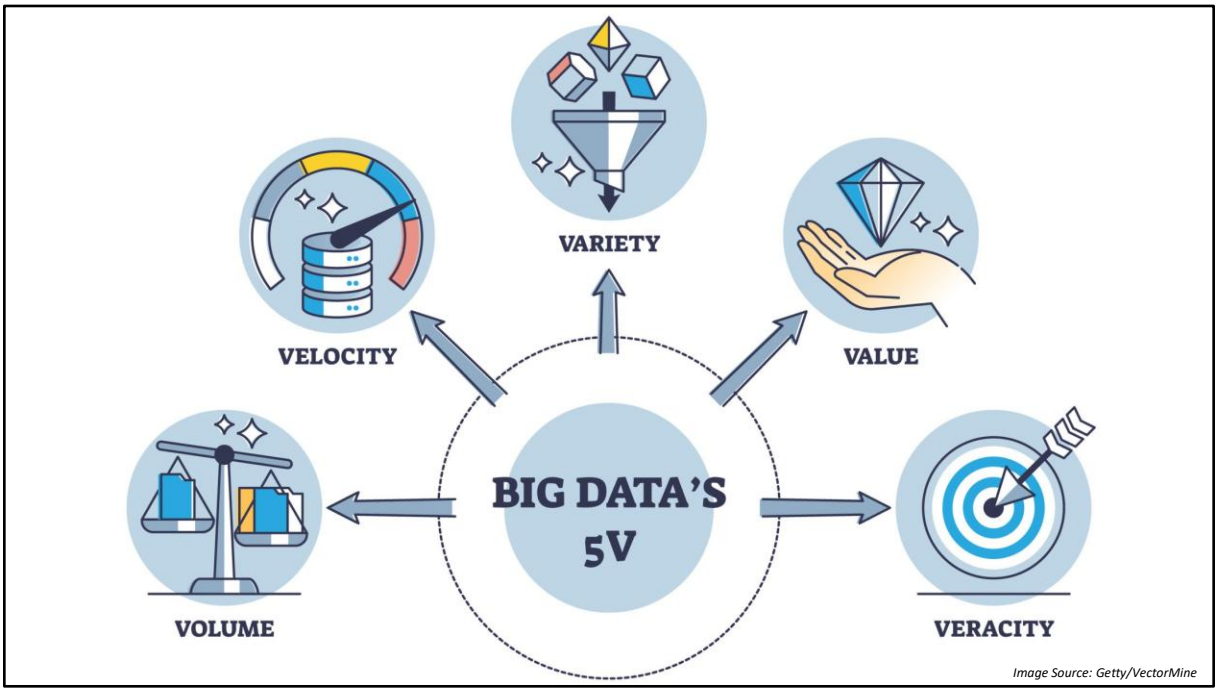
- AI Risks
- AI Evaluation
- AI Testing
- AI & Data Considerations

2 Put AI into Practice

- Funding Opportunities
- Project Highlights

3 Learn & Connect

- AI Organizational Assessments
- AI Case Studies
- AI Webinars, Peer Exchanges, Workshops
- AI “Insights” & Informational Materials



AI systems are fundamentally different from traditional rules-based systems, since they learn from data rather than follow pre-set instructions.

Consider the **5V's of big data** to ensure a solid data foundation for AI.

- **Volume** (or quantity)
    - Advanced AI models (e.g., deep learning) require vast amounts of training data covering a representative range of real-world conditions.
    - Additional data comes with additional costs, not just for sensor purchasing or collection, but also for transmission, processing, storage, and integration.
    - Greater volumes of data can slow down velocity, or the speed at which the data can be transmitted and processed, impacting real-time applications.
    - Agencies may need to assess whether performance gains justify data collection and storage costs.

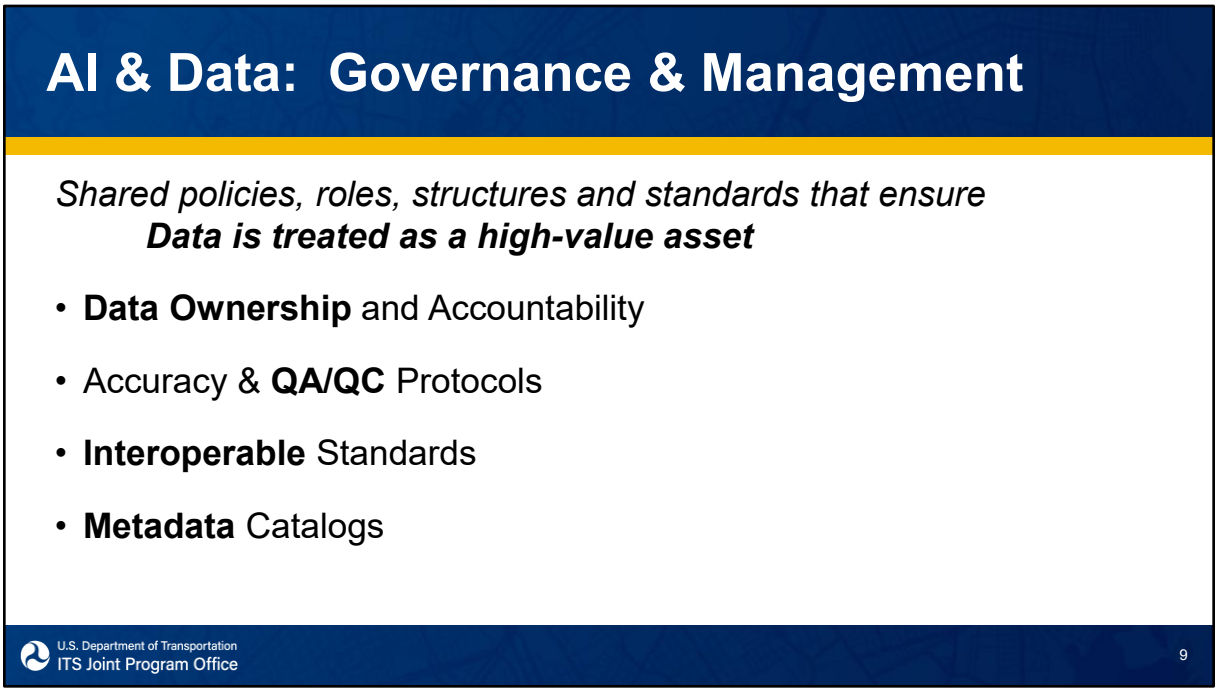
- **Velocity** (or data speed/frequency)
    - Data collection is trending towards real-time or near-real-time, e.g., traffic volumes every 30 seconds, and these data are increasingly granular.
    - Data transfer and storage costs can multiply with increased velocity (and volume).

    - Raw data may need to be aggregated, transformed, and further processed before it is input into AI.
    - Ensure the data units (e.g., every second, or every hour) align with agency objectives.

- **Variety**
    - Agencies collect data from a wide variety of sources.
    - Data fusion across a variety of sources may require working with vendors/Original Equipment Manufacturers (OEMs) to get object-track level, rather than aggregated, sensor data.

- **Value**
    - Just because data is available does not mean it is valuable for a given use case.
    - Excess or irrelevant data that is fed to AI can lead to noise, unreliable results, and wasted dollars.
    - Balance data volume, velocity, and value for the AI use case.

- **Veracity** (or quality)
    - Inaccurate, irrelevant, inconsistent, or biased data leads to unreliable AI behavior.
    - Labeled data (with ground truth) is important for AI model training and evaluation, but it requires domain expertise and can be a significant undertaking to create.
    - Data quality issues early in the data collection and processing pipeline (e.g., sensor misalignment) can compound downstream and lead to faulty AI outputs, so validation at the source is key.
    - Commercial off-the-shelf (COTS) AI solutions may not have been trained in specific local contexts (e.g., fail to account for local traffic laws, different weather conditions), making it important to test the AI under various conditions with an agency expert to verify outputs.
    - Quality transportation data is a long-term strategic asset that could be used by various AI applications.
    - Ensure a human-in-the-loop (HITL) to validate data veracity, especially during AI testing.



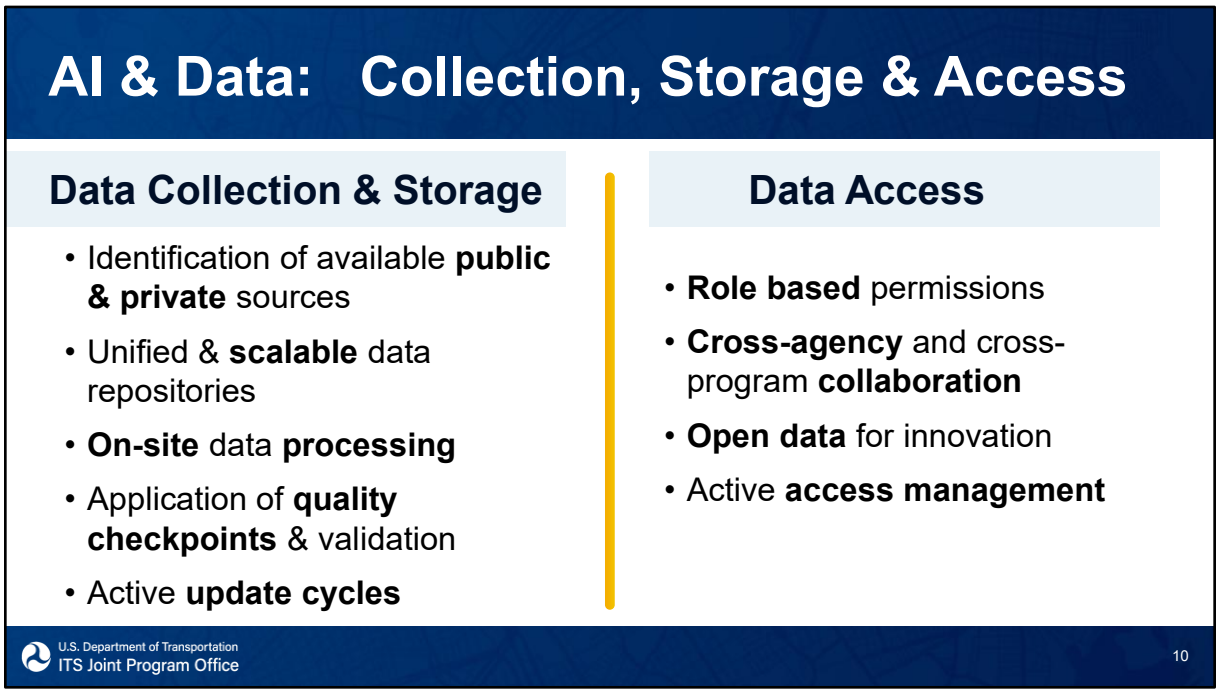
Data Governance & Management

High Level Summary: Serves as the strategic foundation for AI by ensuring that the data and information used is accurate, documented, and trustworthy. Through strong leadership and agency-wide policies, governance creates the formal structures needed to break down silos, ensuring that AI projects are coordinated across departments rather than developed in isolation. Cannot start by looking at only data collection without taking a more strategic view and that begins with governance discussions.

- **The "Rules of the Road":** Establishes the shared policies, roles, structures and standards that ensure data is treated as a high-value asset.
- **Data Ownership & Accountability:** Defines who "owns" and maintains either public or private source data within each program area to ensure information is kept reliable and up-to-date at the source.
- **Accuracy & QA/QC Protocols:** Focuses on using formal checkpoints to assess, clean and verify raw data so AI models don't produce errors or "hallucinations."
- **Interoperable Standards:** Adopting universal data formats so that

information from different programs or different agencies can "talk" to the same AI system seamlessly.

- **Metadata Catalogs:** Creating a "data dictionary" for critical data sources to set the foundation for data integration and fusion so that the AI and agency workforce understand data element meanings.



Data Collection & Storage

Summary: Serves as the foundational "fuel" for AI, providing the historical and real-time information necessary for algorithms to learn patterns and make accurate predictions. For transportation agencies, a robust storage strategy ensures this data is not only captured but remains organized, accessible, and high-quality, turning raw observations into a reliable asset.

Key Features

- **Identification of available public & private sources:** Much of the data that could support transportation system management is held by other entities, both private and public.
- **Unified & scalable data repositories**: Moving from "siloed" data to central digital data “warehouse” that evolve as needed and utilize cloud storage.
- **On-site data processing**: Using sensors that process data on-site to save on costs, while integrating other data inputs from disparate sources (e.g., dashcam videos, manual inspection reports, etc).

- **Application of data quality checkpoints & validation**: Establishing protocols for checking data accuracy and completeness before the AI is allowed to use it.
- **Active data update cycles**: Setting defined schedules to ensure AI uses latest available data, while automatically archiving "stale" info to prevent outdated decisions.

Data Access

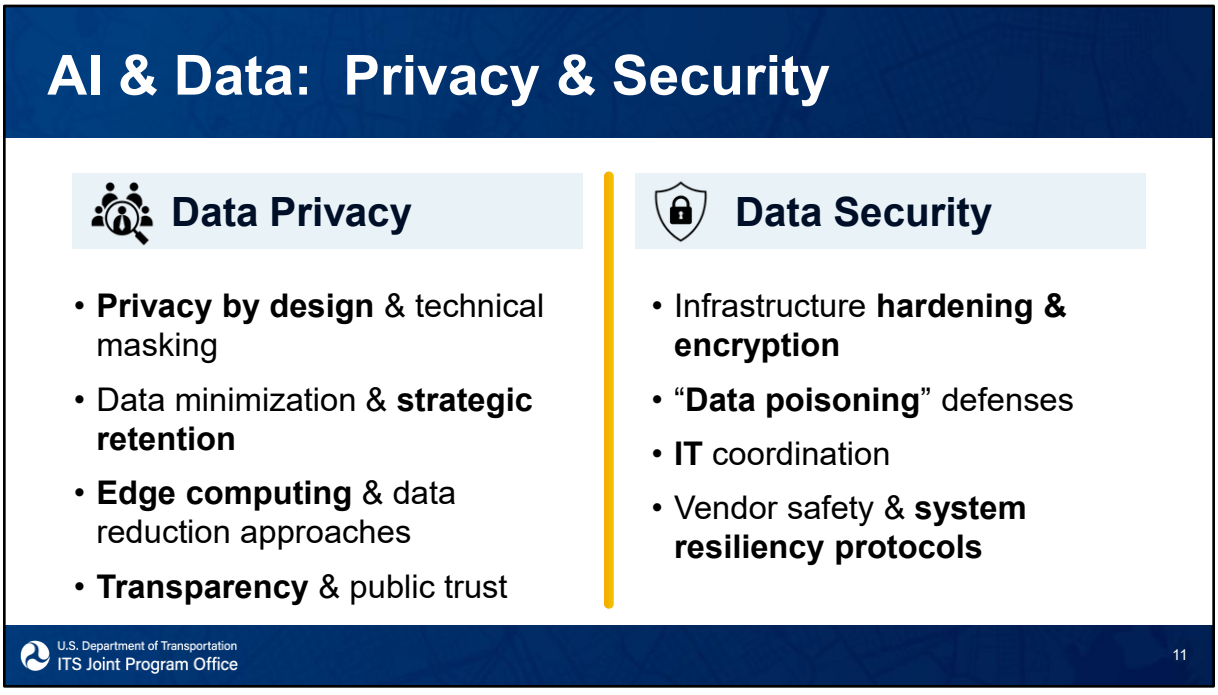
High level summary: Ensures the right information reaches the right person at the right time, breaking down internal agency or cross-agency silos to maximize the utility of available data. By balancing transparency with strict controls, agencies can foster regional innovation while ensuring that sensitive information remains protected from unauthorized use.

Role-Based Permissions: Ensuring that data access is tied to specific roles and levels of agency authority.

Cross-Agency & Cross- Program Collaboration: Across agencies--facilitating "Data Sharing Agreements" between a mix of agency types in a state or region.

Open Data for Innovation: Providing non-sensitive datasets to appropriate stakeholder groups (e.g., academia) to help spark innovative solutions.

Active Access Management: Regularly reviewing and enforcing who has the "digital keys" to the system to ensure that access is revoked when a staff member leaves or a project ends.



Privacy: Collecting (and storing) data for AI can bring new privacy concerns. Focus on serving as responsible stewards of public and citizen data through "masking" and "minimization," ensuring that individual identities remain protected as data moves through the system. For transportation agencies, this builds essential public trust by proving that technology is being used to deliver transportation benefits without compromising individual privacy rights.

Security*:* Ensuring that the data used for transportation programs and project is protected from tampering, theft, or unauthorized control. For transportation agencies, robust security is essential to maintain public safety, ensuring that every automated decision is based on verified, untampered information rather than malicious interference.

PRIVACY

- **Privacy by Design & Technical Masking:** Use automated tools to strip or obscure identifiable details (e.g., faces, license plates, or precise home addresses) before data is stored or used for AI training.
- **Data Minimization & Strategic Retention:** "Collect only what is needed" and set strict expiration dates (retention limits) tied to the specific project.

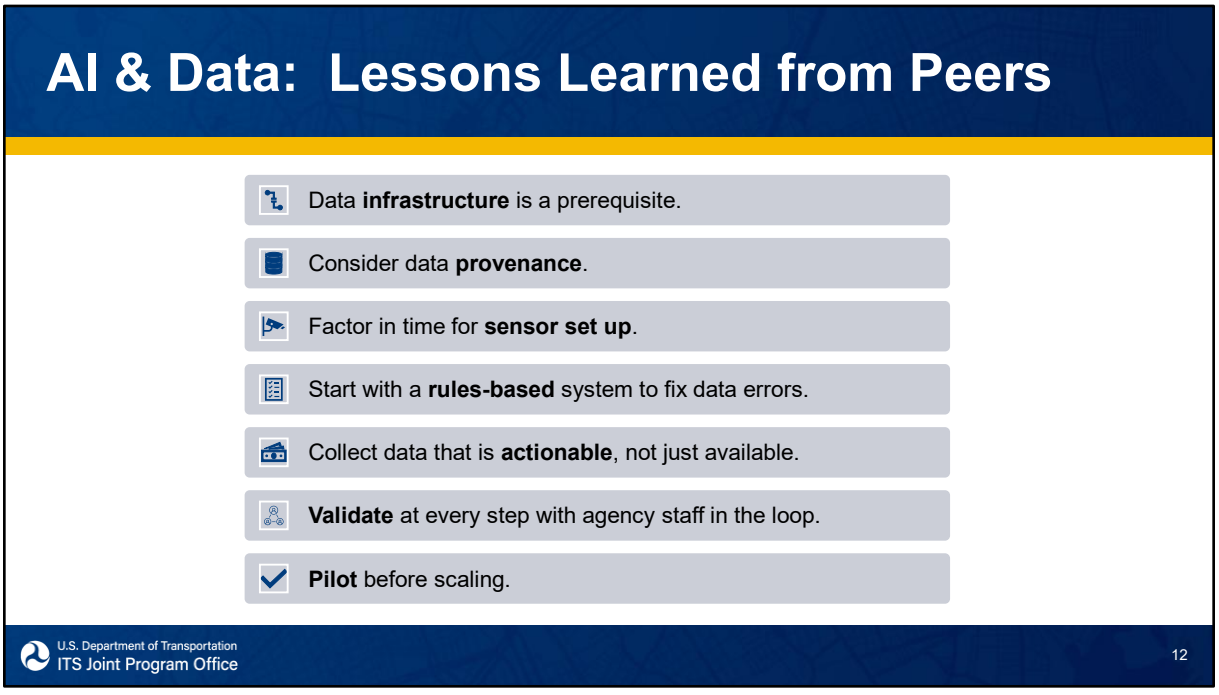
- **Edge Computing & Data Reduction Approaches:** Use "edge" processing to analyze data locally at the sensor and reduce sensor resolution to the minimum level required, ensuring sensitive raw details never leave the site or enter the cloud.
- **Transparency & Public Trust:** Maintain open communication with stakeholders regarding what data is collected, how it is protected (e.g., via encryption or synthetic data), and what specific steps are taken to avoid sensitive data collection.

Example strategies to **protect data privacy** for AI applications in transportation:

- Set a time limit for how long video data is stored.
- Obscure identifiable details from sensor data, e.g., mount sensors high enough to obscure identifiable details.
- Use synthetic data for training.
- Avoid using high resolution streaming video quality and/or use sensors other than visual cameras (e.g., LiDAR) that provide greater anonymity.
- Use encryption and set access controls to help protect sensor feeds.
- Avoid collecting and storing sensitive data when possible.
- Consider edge computing so sensitive data is not stored or transferred, just aggregated data.

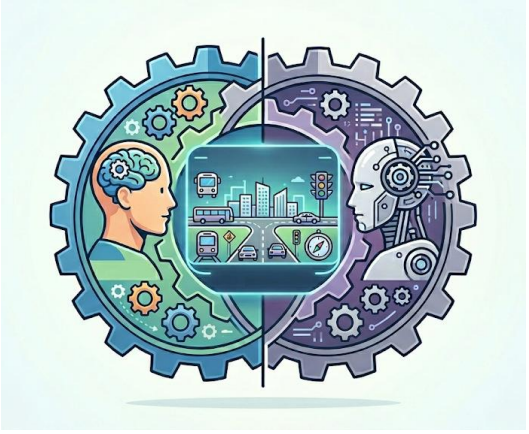
SECURITY

- **Infrastructure hardening & encryption:** Protecting physical equipment (e.g., smart traffic lights and bridge sensors) by "scrambling" (encrypting) the data feeds so hackers cannot take control of the signals or view the data.
- **"Data Poisoning" defenses:** Using digital filters, firewalls, and "human-in-the-loop" checks to catch and toss out fake or "untrusted" data before it reaches the AI, preventing bad actors from intentionally causing traffic jams or routing errors.
- **IT Coordination:** Partnering with the IT department from day one to set up secure data flows and monitor for "misbehavior," ensuring that any unusual activity in the AI system is flagged and stopped immediately.
- **Vendor safety & system resiliency protocols:** Requiring all outside software companies to meet strict safety standards and maintaining "fail-safe" manual overrides so that if the digital system is attacked, the physical roads and transit lines keep moving safely.



- Data infrastructure is a prerequisite as not all signals have fiber, cellular data can be unreliable, etc.
- No AI system can overcome poor data provenance.
- Sensor procurement, installation, and calibration takes longer than expected.
- Start with a rules-based system first to fix data processing errors and information flows before adding AI complexity.
- Make sure the data collected is actionable so costs don't add up for little value added.
- Validate at every step with a HITL, especially since many AI/tech vendors are new to transportation.
- Pilot before scaling to validate approaches and adequately test all components.

AI & Data: Workforce Integration



- In-house AI expertise
- Human validation
- Data-driven decision culture

Image Source: Gemini

To use data and AI well in your organization, you need to think of the system that delivers value as the marriage of the human and the AI-based system. AI on its own can't do anything. It needs a human to set it up, interact with it, validate it, and use its results.

In-house AI expertise: Preparing current staff to manage these data and AI intensive projects so the agency isn't 100% dependent on vendor solutions.
Human validation: Ensuring that agency staff have the skills to check AI outputs based on their real-world knowledge.
Data-driven decision culture: Shifting from "we've always done it this way" to an analytically driven decision process while integrating internal historical expertise.

We Want to Help You!

Lisa Randall
AI Program Manager
ITS JPO
Lisa.Randall@dot.gov

Matt Cuddy
JPO AI Program Team Manager
Volpe Center
Matt.Cuddy@dot.gov

How We Support You

- Development Tools
- Field Testing & Pilots
- Knowledge Sharing
- Guidance for Trustworthy AI
- Webinars & Outreach