Artificial Intelligence (AI) for Intelligent Transportation Systems (ITS)

Challenges and Potential Solutions, Insights, and Lessons Learned

www.its.dot.gov/index.htm

Final Report – October 2022

FHWA-JPO-22-968

U.S. Department of Transportation

Produced by Noblis
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
Intelligent Transportation Systems Joint Program Office

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The U.S. Government is not endorsing any manufacturers, products, or services cited herein and any trade name that may appear in the work has been included only because it is essential to the contents of the work.

Technical Report Documentation Page

1. Report No.

FHWA-JPO-22-968

2. Government Accession No.

(Delete and insert information here or leave blank)

3. Recipient's Catalog No.

(Delete and insert information here or leave blank)

4. Title and Subtitle

Artificial Intelligence (AI) for Intelligence Transportation Systems (ITS) Challenges and Potential Solutions, Insights, and Lessons Learned

5. Report Date

October 2022

6. Performing Organization Code

(Delete and insert information here or leave blank)

7. Author(s)

Meenakshy Vasudevan, Haley Townsend, Matt Samach, Atizaz Ali, Patrick Walsh, Peiwei Wang, Anand Seshadri, and Ian McManus

8. Performing Organization Report No.

(Delete and insert information here or leave blank)

9. Performing Organization Name and Address

Noblis Inc.
500 L'Enfant Plaza, S.W., Suite 900
Washington, D.C. 20024

10. Work Unit No. (TRAIS)

(Delete and insert information here or leave blank)

11. Contract or Grant No.

(Delete and insert information here or leave blank)

12. Sponsoring Agency Name and Address

Intelligent Transportation Systems (ITS) Joint Program Office (JPO)
1200 New Jersey Avenue, S.E.,
Washington, DC 20590

13. Type of Report and Period Covered

Final

14. Sponsoring Agency Code

ITS JPO

15. Supplementary Notes

Work Performed for: Robert Sheehan (ITS JPO; Task Order Manager) and Dr. Jonathan Walker (ITS JPO)

16. Abstract

Artificial Intelligence (AI), including Machine Learning (ML), offers the opportunity to make transportation systems safer, and more equitable, reliable, accessible, secure, efficient, and resilient. However, several challenges exist that could impede the successful adoption of AI for Intelligent Transportation Systems (ITS) and the potential realization of these benefits.

Throughout the last few years of the ITS JPO's research and market engagement efforts to assess the potential of AI for ITS, a variety of challenges to AI adoption in ITS have been surfaced. These challenges include, but are not limited to, issues surrounding data, supporting technology, bias, security, privacy, ethics and equity, generalization, model drift, explainability, liability, talent/workforce availability, and stakeholder perception. While these challenges to AI adoption and implementation cut across domains, the purpose of this report is to focus on their implications for ITS as well as insights that agencies could consider in helping to mitigate them.

17. Keywords

Artificial Intelligence, Machine Learning, bias, security, privacy, ethics and equity, generalization, model drift, explainability, liability

18. Distribution Statement

FINAL

19. Security Classif. (of this report)

(Delete and insert information here or leave blank)

20. Security Classif. (of this page)

(Delete and insert information here or leave blank)

21. No. of Pages

137

22. Price

Acknowledgements

The authors would like to thank the U.S. Department of Transportation (USDOT) Intelligent Transportation Systems (ITS) Joint Program Office (JPO) for sponsoring this work. Specifically, the authors would like to thank Robert Sheehan (ITS JPO, Government Task Order Manager), and Dr. Jonathan Walker (ITS JPO).

The authors would also like to thank Joe Gregory (FHWA Office of Operations), David Kuehn (FHWA Office of Corporate Research, Technology, and Innovation Management), and James Pol (FHWA Safety R&D) for providing valuable feedback.

The authors would also like to thank Aaron Sessions, Justin Anderson, and Dr. Karl Wunderlich from Noblis, for their inputs.

Executive Summary

Artificial Intelligence (AI), including Machine Learning (ML), offers the opportunity to make transportation systems safer, and more equitable, reliable, accessible, secure, efficient, and resilient. However, several challenges exist that could impede the successful adoption of AI for Intelligent Transportation Systems (ITS) and the potential realization of these benefits. These challenges include, but are not limited to, issues surrounding data, supporting technology, bias, security, privacy, ethics and equity, generalization, model drift, explainability, liability, talent/workforce availability, and stakeholder perception. While these challenges to AI adoption and implementation cut across domains, this report focuses on their implications for ITS as well as insights that agencies could consider in helping to mitigate them. Table 1 summarizes these 12 challenges, their implications for ITS, and insights and lessons learned that agencies could consider.

Table 1. Summary of Challenges to AI Adoption, Implications for ITS, and Lessons Learned

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
1	**Data** Lack of sufficient, high-quality, and relevant data	• Collecting sufficient high-quality data can be difficult when infrastructure-based sensors are sparse, or funding is limited. Labeling large sets of unstructured data can be costly. Data fusion from diverse sources can be complicated and costly. • Data manipulation and feature extraction can have major implications on the performance and capabilities of AI systems because they can allow the AI system to learn from more generalizable data points or capture complex relationships between the features. • AI systems can learn undesired behaviors and relationships in the training phase if the data are not thoughtfully prepared. • A lack of standards for data access and sharing are an impediment to accelerating the maturity of AI-enabled ITS. • Collection of massive datasets presents challenges related to supporting technology (see #2) and data privacy (see #5).	• Using/sharing data with sufficient metadata and documentation so that practitioners are aware of the nuances, potential pitfalls, and recommended uses of the data. • Developing a comprehensive data management strategy to ensure organizational alignment in data governance. • Reusing data as much as possible to reduce duplicated effort for similar use cases. However, just because certain data are available does not mean that they will be useful for the task at hand, so consider use case relevance. • Considering synthetic, imputed, and human/AI collaborative approaches to creating data and performing extensive validation. • Availability does not imply data relevancy or usefulness.

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
2	**Supporting Technology** Inability of legacy systems to support the addition and integration of new AI-based functionalities, due to software-hardware integration issues, limited data storage capacities, and restricted computational power of legacy systems	• Infrastructure costs associated with AI systems can be high; lack of available funding may require prioritization of projects. • Lack of documentation for legacy systems can make them less adaptable at supporting new equipment. • Unlocking the full potential of AI-enabled ITS applications requires resolving integration and compatibility issues, as well as data storage and computational power problems. • ITS equipment requires continuous power supply which can be costly and challenging specifically for rural AI for ITS applications. • AI/ML algorithms have carbon costs associated with them.	• Launching pilot deployments to uncover potential barriers and demonstrate benefits • Leveraging existing ITS infrastructure where applicable to minimize costs • Using cloud computing to increase the computational speed • Using edge computing to overcome bandwidth and latency issues • Leveraging clustered computing to augment processing power • Adopting systems engineering best practices to build more adaptable and resilient systems • Building sustainable AI solutions to reduce negative environment impacts

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
3	**Bias** When AI systems lead to unfair and unequitable outcomes due to underrepresentation of subpopulations, or due to human or systemic/institutional biases	• Groups of roadways users may be underrepresented in training datasets used for machine vision applications (e.g., for TSMO applications such as Smart Intersections). • Data are sparse for work zones despite being one of the most vulnerable areas of roadways for fatalities. • When data are collected from users of mobile applications for transportation, the samples may be biased. • Data may be biased due to unequitable distribution of sensors along roadways.	• Adopting a socio-technical systems approach to mitigating bias in AI systems • Bringing together diverse teams for AI systems development • The most accurate model is not always the one with the least harmful impact • Collecting sufficient data to measure error statistics across demographic groups • Choosing fairness metrics that reflect the values of the organization and the groups for whom the AI system has the highest risk of harm • Monitoring bias mitigation is an ongoing process that extends throughout the AI system lifecycle

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
4	**Security** When an attack, via cyberspace, targeting an enterprise's use of cyberspace for the purpose of disrupting, disabling, destroying, or maliciously controlling a computing environment/infrastructure; or destroying the integrity of the data or stealing controlled information (NIST, 2022)	• Insufficient attention has been paid to the ways in which AI can be used maliciously. • Malicious entities could compromise the integrity of the decision-making process (e.g., data poisoning, model evasion). • Hacking cyber-physical infrastructure (e.g., DMS, Colonial Gas Pipeline) poses a threat. • Connected vehicle adoption may increase vulnerabilities. • Automated vehicles, which rely heavily on AI algorithms, make safety-impacting driving decisions. • Agencies and the public may mistrust AI applications.	• Understanding potential security threats from misuse of AI-based applications to better forecast, prevent and mitigate the threats • Following cybersecurity best practices • Collaboration among various stakeholders to identify transportation cybersecurity best practices • Developing workforce and domain expertise to curtail security issues • Strengthening the security of AI systems by addressing vulnerabilities • Utilizing intrusion and misbehavior detection systems to enhance safety of AI-based ITS systems • Retraining ML models at regular intervals to retain the quality of ML predictions • Securing physical infrastructure to block potential physical intrusion into the system

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
5	**Privacy** Inability of an AI system to protect individual privacy, including personally identifiable information (PII) and other sensitive information	• New AI applications in ITS could heighten identity, behavioral, and location privacy concerns. • Privacy leakage could lead to liability issues for the agency and reduced trust from system users. • Agencies may have to consider tradeoffs between privacy and utility in data. • AI applications in ITS that may rely on or capture sensitive information, such as pedestrian detection, automated license plate readers, personalized traveler information, and driver monitoring, could pose higher privacy risks.	• Obscuring/encrypting sensitive data • Collecting non-sensitive data • Using synthetic data • Applying differential privacy • Using a distributed protection technique • Using edge computing to limit PII collection • Establishing data sharing techniques • Developing privacy policies

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
6	**Ethics and Equity** When AI applications, whether intentionally or unintentionally, profile and discriminate against individuals/populations based on unfair or unclear criteria or lead to unethical or inequitable outcomes	• AI-enabled ITS systems can inform, or even make, decisions that greatly impact human lives. For example, Automated Vehicles (AVs) could encounter major ethical dilemmas in their driving decision making. • Disadvantaged populations could be unfairly discriminated against via AI-enabled ITS, negatively impacting equity. For example, discrimination could occur in infrastructure and asset management decisions or in language processing. • Inequitable outcomes could occur from competing objectives or biased data collection. • Seemingly negligible development choices, such as spatial resolution or sensor placement, could lead to unintentional consequences.	• Creating AI systems with ethics, equity, and transparency at the forefront • Translating ethical frameworks into engineering • Supporting workforce training and education to meet future AI needs • Including diverse stakeholders throughout AI development • Promoting ethical, trustworthy AI use and development • Applying guidelines to promote responsible AI • Documenting processes, success metrics, and expectations • Including a human-in-the-loop for critical decisions

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
7	**Generalization** When a trained ML model does not adapt well to unseen data, it may have underfit or overfit its training data, which could lead to poor performance	• Vendors may promote their AI solutions as being able to work anywhere, but AI solutions are not necessarily designed to work everywhere. • Since non-recurring conditions are far less common than recurring conditions, they present a challenge in terms of having enough data to train an ML model to detect and classify them correctly. • ML models require large quantities of representative data to generalize well, but real-world data can be expensive to acquire, and simulated data may not be fully representative.	• Having representative data for training • Making the training data more robust • Handling edge cases • Limiting overfitting • Combining ML techniques or models • Using model testbeds • Re-training for new locations • Considering transfer learning • Developing and using standards to enable interoperability

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
8	**Model Drift** When a trained model's input data, output data, or relationship between the two changes over time leading to system performance degradation	• Model drift could lead to AI system performance degradation, which in turn, could reduce ITS performance and user trust. • Sensor malfunctions or hardware/software updates could lead to incorrect predictions if an AI system has not been trained on these occurrences. • If the input data in an operational setting starts to drift away from the data used to train the model, the performance of the AI system might start to degrade. For example, the introduction of a new ridesharing service that was not captured in the training data could lead to an AI-enabled traveler information system no longer offering the most relevant options to travelers. • Policy changes that impact the target variable could lead to concept drift. For example, if a freeway agency changes one of its lanes from an HOV-2 to an HOV-3, an AI-based vehicle occupancy detection and tolling enforcement application would need to be updated to learn this change or else it could incorrectly enforce toll rates.	• Having a plan in place for model drift assessment and mitigation • Establishing appropriate ranges of data and model drift • Regularly monitoring and improving the system • Retraining the model • Considering online learning

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
9	**Explainability** Inability of an AI system's process and decisions to be understood by humans	• Explainability is especially important for safety-critical and other high-stakes decisions with greater risk and liability concerns for the agency. • The level of explanation required for an AI-enabled decision support system depends on the task at hand and level of supervision from the ITS decision maker. • If a vendor's AI solution is not transparent and explainable, this could reduce agency and user trust in the overall procured system. • Even simple explanations for AV decisions could improve driver and pedestrian interaction with and trust of the AV.	• Understanding potential tradeoffs between interpretability and performance • Balancing explainability with security and privacy • Improving transparency through documentation • Using interpretable models • Engineering interpretable features • Outputting multiple performance metrics • Visualizing results • Exploring post-hoc explainable AI (XAI) methods • Using explainable AI (XAI) analysis for validation of model strategies and to improve trust in AI outcomes • Considering non-AI alternatives

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
10	**Liability** Lack of clear definition of who is liable when a vehicle, device, equipment, or system that uses AI is involved in a crash, is hacked, or produces erroneous results	• If the AI application fails due to bias in the data, it is currently unclear whether the liable party for the failure is the application developer or the data provider • Liability is unclear when a vehicle, device, equipment, or system that is powered by an AI application is involved in a crash or results in fatalities. • Lack of clarity of safety expectations may regarding the damages that results from cybersecurity breaches in an AI product. • If an AI-enabled application has poor performance resulting in significant productivity losses, it is unclear who should be held accountable.	• Partnering closely with agency risk management teams to consider legal and compliance issues from the perspective of organizational experts. • Assessing legal restrictions for the data to establish contracts and agreements in ways the data should be collected and used. • Assessing legal restrictions for the AI algorithm to establish contracts and agreements on all aspects of algorithm use and ownership. • Identifying possible risks throughout the AI pipeline, including considering downstream uses of AI system outputs. • Maintaining human accountability by assigning responsibility for AI system outcomes on specific individuals and organizations.

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
11	**Talent/Workforce Availability** When there is lack of talent/expertise in building trustworthy, ethical AI algorithms, or integrating, operating, and maintaining real-world AI-based systems	• Workforce talent and education are key bottlenecks to successful deployment and integration of AI systems into the operations of government agencies. • Domain experts in the transportation industry often do not have sufficient AI knowledge to work alongside data scientists in building models that are relevant and operationally useful. • Due to budget limitations, agencies have limited staff to operate and maintain AI-based systems. Therefore, balancing hiring decisions between ML/AI expertise and domain expertise can be a challenge.	• Improving diversity in the workforce, and balancing AI talent and domain expertise to overcome challenges related to limited resources • Collaborating with partners for AI expertise • Providing client training to make deployment smoother, leading not only to improved technical proficiency of personnel but also buy-in for AI-enabled systems • Conducting periodic education and training for current staff, new hires, and domain experts, so they can keep up with advances in AI

ID	Challenge and Description	Implications for ITS	Insights and Lessons Learned
12	**Stakeholder Perception** When stakeholders are skeptical or mistrustful of AI systems or have exaggerated expectations of AI systems' capabilities	• If stakeholders lack a clear understanding of the capabilities of AI, this can lead to skepticism and mistrust or to blind belief in AI as a solution for all problems, both of which could impede the successful implementation of AI. • Due to perceived high costs and risk aversion, agencies may prefer to deploy traditional ITS systems rather than AI-based systems. • Ethics, liability, and privacy issues could also affect stakeholder perception of AI. Agencies may have to contend with these institutional challenges when implementing AI solutions.	• Conducting stakeholder analysis to identify stakeholders and their needs • Building trustworthy and ethical AI systems • Engaging with the user community early and often to gain buy-in and understand stakeholder needs • Demonstrating the value of AI to keep stakeholders on board with the project • Exchanging information with other deployers to share insights, lessons learned, and preliminary results • Ensuring leadership buy-in of AI techniques for initial and continued support • Setting stakeholder expectations, including on the implementation timeline • Promoting public understanding of AI to clarify what it is and how it could play a role

Key Takeaways

Overarching key takeaways are summarized below.

- **The twelve challenges for AI adoption and successful implementation are not unique to ITS.** They are broad technical and institutional challenges that impact a wide variety of sectors. Many of the insights and lessons learned in this report are gleaned from other sectors and could potentially be applied to ITS.

- **There may be tradeoffs between addressing different challenges.** For example, greater explainability could provide more information for malicious actors to manipulate, potentially breaching security and/or privacy. Adding robust, large scale data sources may boost AI performance but could be costly to store and implement.

- **Addressing these challenges is an ongoing exercise.** These challenges are dynamic and, like AI itself, will evolve over time. For example, cybersecurity concerns today may look different than cybersecurity concerns next year as malicious actors find new ways to hack into systems. Additionally, stakeholder buy-in is important not only at the onset of a project but also throughout the project to support its continued success. The deployment of new AI techniques may require new staff expertise. Overall, challenges and risks are dynamic and addressing them is an ongoing exercise.

- **Maintaining a human-in-the-loop is helpful in identifying and mitigating these challenges.** Ongoing human oversight of AI/ML applications in ITS can help in identifying and mitigating potential issues, particularly those that the machine may not catch and those that may require making tradeoffs in how they are addressed. Having both domain and AI/ML expertise on staff is useful for not only the initial implementation but for ongoing operations and maintenance of the system and its AI/ML applications.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

The U.S. Department of Transportation (USDOT) has long been a leader in research, development, and evaluation of technologies for transportation and a strong supporter of the adoption and use of Intelligent Transportation Systems (ITS). ITS includes multiple components of transportation infrastructure, vehicles, back offices, services, and other tools and mechanisms that serve all transportation users, including underserved communities and groups, private-sector vendors of technology equipment and applications, and operators and implementers of ITS whether privately or publicly owned (e.g., state, local, and tribal governments). ITS can improve the capabilities of the transportation system by integrating advanced information and communications-based technologies (ICT) into transportation infrastructure and vehicles.

The ITS Joint Program Office (JPO)'s mission, according to Strategic Plan 2020-2025 (Chan-Edmiston et al., 2020), is to lead collaborative and innovative research, development, and pilot deployment, and facilitate implementation of ITS to improve the safety and mobility of people and goods. In line with this mission, the ITS JPO and its modal partners have been leading the way in tackling fundamental problems in transportation by leveraging emerging technologies and strategies, including spectrum utilization, cybersecurity, computer processing, connected and automated vehicles, shared mobility services, accessible transportation technologies, and artificial intelligence (AI).

USDOT recognizes the promise AI offers for achieving considerable benefits in safety, mobility, equity, efficiency, accessibility, productivity, resilience, and reduction of individual and societal costs, emissions, and other negative environmental impacts. In the last few years, explorations into AI have grown tremendously within the USDOT (Thompson, 2019). Some of the USDOT's modal administrations, including the Federal Highway Administration (FHWA), Federal Railroad Administration (FRA), and the Federal Aviation Administration (FAA), have been at the forefront of adopting AI solutions for transportation mission delivery. AI-enabled applications are being explored and implemented for video analytics, safety analysis, and data fusion, among others. For example, the FHWA's Exploratory Advanced Research Program supports research projects on AI for making sense of big data and using video analytics to help analyze driver behavior (FHWA EAR Program, 2022). Additionally, the FHWA's Traffic Analysis Tools (TAT) Program is investigating the use of AI for developing prediction techniques and evaluation tools (FHWA Office of Operations, 2022). The FHWA's Advanced Transportation and Congestion Management Technologies Deployment (ATCMTD)

Program has awarded 48 grants (so far) to develop and deploy cutting-edge technologies, including AI, to improve safety and mobility (FHWA, 2020). The FRA is developing a suite of technologies for predictive analytics and intruder detection using AI and unmanned aircraft systems (UAS) (Baillargeon, 2019). Other agencies, such as the Federal Transit Administration (FTA), Federal Motor Carrier Safety Administration (FMCSA), and the Pipeline and Hazardous Materials Safety Administration (PHMSA), are exploring the promise that AI has to offer in citizen-facing services.

In 2019, the ITS JPO initiated the AI for ITS Program with the vision to "advance next generation transportation systems and services by leveraging trustworthy, ethical AI (including machine learning) for safer, more efficient, and accessible movement of people and goods" (Walker, 2021). As part of this effort, the ITS JPO developed the following definition of AI in the context of ITS: *Artificial Intelligence (AI) refers to processes that make it possible for systems to augment routine human tasks or enable new capabilities that humans cannot perform. AI enables systems to: (1) sense and perceive the environment, (2) reason and analyze information, (3) learn from experience and adapt to new situations, potentially without human interaction, and (4) make decisions, communicate, and take actions.*

Over the last few years, the AI for ITS Program conducted a series of research and market engagement to assess the potential of AI for ITS. A comprehensive review of literature was conducted to understand how AI is being leveraged to address ITS needs, specifically to improve transportation system and users' safety, mobility, accessibility, productivity, efficiency, and environmental impacts (Vasudevan, Townsend, et al., 2020). This investigation focused on AI-enabled ITS applications that the USDOT and its state and local partners might apply to the planning, operation, and maintenance of the multimodal surface transportation system, as well as applications developed by the private sector that USDOT may have a role in enabling.

Building on this preliminary investigation into promising applications of AI for ITS, the ITS JPO decided to conduct market research to get feedback from public sector agencies, industry, research laboratories, academia, and other stakeholders on deployment-ready applications that leverage AI to address ITS needs, existing capabilities in developing and deploying AI-enabled ITS applications, and USDOT role and investment areas to facilitate next generation ITS leveraging AI.

A variety of challenges to AI adoption in ITS have been surfaced through the review of literature, interviews with subject matter experts, strategic sessions with USDOT modal experts, AI for ITS Program webinars and market research, and other AI-related conferences, webinars, and workshops organized by agencies such as the National Institute of Standards and Technology (NIST), SAE, American Society of Civil Engineers (ASCE), and USDOT's University Transportation Center (UTC) Program. Some of these challenges are technical while others are more organizational, ranging from data requirements, bias, and privacy concerns to ethics, liability, and stakeholder perception. While these challenges to AI implementation cut across domains, this report attempts to focus on their applicability to ITS and what agencies could consider doing to help mitigate them.

1.2 Purpose of this Report

The purpose of this report is to provide USDOT and transportation agency staff with awareness of potential challenges to the use of AI for ITS as well as potential solutions, insights, and lessons learned to help overcome them. Recognizing that these challenges are broad, this report does not attempt to be comprehensive in its discussion. Instead, this report attempts to highlight pertinent available information with a focus on insights that could be most relevant to ITS and the transportation agency staff.

1.3 Glossary

This report assumes some foundational knowledge of AI and ML concepts, and therefore, uses a variety of common terms throughout. Practical definitions of these terms are summarized in the glossary in Table 2.

Table 2. Glossary of Terms Used in This Report Related to AI/ML

Term	Definition
Adversarial Machine Learning	Using malicious inputs designed to fool machine learning models. Technique to find a perturbation that changes the prediction of a machine learning model (*Papers with Code - The Latest in Machine Learning*, 2022). Adversarial Machine Learning is a collection of techniques to train neural networks on how to spot intentionally misleading data or behaviors. This differs from the standard classification problem in machine learning, since the goal is not just to spot “bad” inputs, but preemptively locate vulnerabilities and craft more flexible learning algorithms (DeepAI, 2019).
Black box Models	When users are able to provide inputs and view outputs of the target ML model but are unaware of the architecture/structure of model.
Classification	ML model that distinguishes among two or more discrete classes (Google, 2021).
Clustered computing	A distributed computing concept where multiple machines on the same network act as a single entity, providing increased computational power (Baker, 2000).

Term	Definition
Cross validation	Testing the target ML model with a dataset unknown to it (dataset not used to train model) (Google, 2021).
Cyberattack	An attack, via cyberspace, targeting an enterprise's use of cyberspace for the purpose of disrupting, disabling, destroying, or maliciously controlling a computing environment/infrastructure; or destroying the integrity of the data or stealing controlled information (NIST Glossary, 2022).
Cybersecurity	The process of protecting information by preventing, detecting, and responding to attacks (NIST Glossary, 2022).
Data augmentation	Artificially boosting the range and number of training examples by transforming existing examples to create additional examples (Google, 2021).
Data poisoning	Adversarial attack that tries to manipulate the training dataset in order to control the prediction behavior of a trained model such that the model will label malicious examples into a desired classes (e.g., labeling spam e-mails as safe) (*Papers with Code - The Latest in Machine Learning*, 2022).
Edge computing	A distributed computing concept where servers are placed closer to devices to reduce latency in communication between devices (Arabi, 2020).
Feature engineering	The process of determining which features might be useful in training an ML model, and then converting raw data from log files and other sources into said features (Google, 2021).
Feature extraction	Retrieving intermediate feature representations calculated by an unsupervised or pretrained ML model for use in another model as input (Google, 2021).
Features	Also referred to as "predictors" (Google, 2021).
Hardware acceleration	The process of offloading data-intensive tasks and functions to hardware (GPUs, FPGAs, ASICs) to speed up computational performance and timing (Zhao et al., 2017).
Human-in-the-loop	Active human oversight of the AI system, "with the human retaining full control and the AI only providing recommendations or input" (GAO, 2021).

Term	Definition
Human-on-the-loop	Human supervision in which "the human is in a monitoring or supervisory role, with the ability to take over control when the AI model encounters unexpected or undesirable events" (GAO, 2021).
Human-out-of-the-loop	The lack of human supervision of the execution of decisions, as in the AI system has full control without the option of human override (GAO, 2021).
Imbalanced dataset	When certain conditions, subpopulations, or classes are overrepresented in the data set, while others are underrepresented.
Inference	Process of making predictions by applying the trained model to unlabeled examples.
Labels	Also referred to as "ground truths" (Google, 2021).
Localized models	ML models trained to the specific characteristics of a network.
Machine Learning (ML)	A broad subfield of AI in which computers learn from data, discover patterns and make decisions without human intervention. The ML field is broadly categorized into supervised, semi-supervised, unsupervised and reinforcement learning (Vasudevan et al., 2020).
Model evasion	network is fed an "adversarial example" — a carefully perturbed input that looks and feels the same as its untampered copy to a human — but that completely throws off the classifier (Ilmoi, 2019).
Model overfitting	Creating a model that matches the training data so closely that the model fails to make correct predictions on new data (Google, 2021).
Natural Language Processing (NLP)	An AI technique for parsing, processing, and analyzing natural human language (Google, 2021).
Parameter	A variable of a model that the machine learning system trains on its own
Pipeline	The infrastructure surrounding a machine learning algorithm. A pipeline includes gathering the data, using the data to create training data files, training one or more models, and exporting the models to production (Google, 2021).
Predictions	The output of a machine learning model, e.g., confidence scores for classifiers (Google, 2021).

Term	Definition
Ransomware	Type of malicious attack where attackers encrypt an organization's data and demand payment to restore access (NIST, 2022).
Regularization	The penalty on a model's complexity. Regularization helps prevent overfitting. Different kinds of regularization include (Google, 2021): • L_1 regularization • L_2 regularization • Dropout regularization
Security	Protection against intentional subversion or forced failure. A composite of four attributes – confidentiality, integrity, availability, and accountability – plus aspects of a fifth, usability, all of which have the related issue of their assurance (NIST, 2022).
Spear phishing	An attempt to acquire sensitive information or access to a computer system by sending counterfeit messages that appear to be legitimate (*Counterintelligence_Tips_Spearphishing*, n.d.)
Spyware	Software that is secretly or surreptitiously installed into an information system to gather information on individuals or organizations without their knowledge (NIST, 2022).
Supervised learning	Training of a model on datapoints that include labels (Google, 2021).
Testing data set	Also referred to as "validation data set" depending on the context (Google, 2021)
Training data set	The set of examples used to fit the target model and depends on the learning method (Google, 2021)
Unsupervised learning	Training of a model on datapoints that do not include labels (Google, 2021)
White box Models	The structure and parameters of the model are known to all users that have access to the model

1.4 Organization of this Report

This document is organized into the following chapters:

- **Chapter 2 Challenges to Adoption for ITS** – describes 12 challenges to AI adoption and successful implementation in ITS, including a summary table, description, implications for ITS, and insights and lessons learned for each challenge. The tables are meant to serve as convenient summaries for each challenge, particularly for USDOT and agency decision makers seeking quick, high-level views of the material. All 12 individual summary tables are combined in the Executive Summary in Table 1.

- **Chapter 3 Key Takeaways** – summarizes high-level key takeaways across all 12 challenges.
- **References** – includes a list of references in this report ordered by author-date with links where possible.

2 Challenges to AI Adoption for ITS

AI offers the promise to improve the safety, mobility, accessibility, equity, productivity, and efficiency of transportation systems. AI has many practical applications in the transportation domain that could promote these goals. However, there are significant challenges to the adoption and successful implementation of AI for ITS. Given below are cross-cutting challenges that agencies could face in implementing AI-enabled solutions to address problems seen on their transportation networks, corridors, and systems. Figure 1 illustrates the 12 major challenges summarized in this chapter.

This chapter describes these 12 challenges, and corresponding key implications for ITS, and potential solutions, insights, and lessons learned to help overcome them.

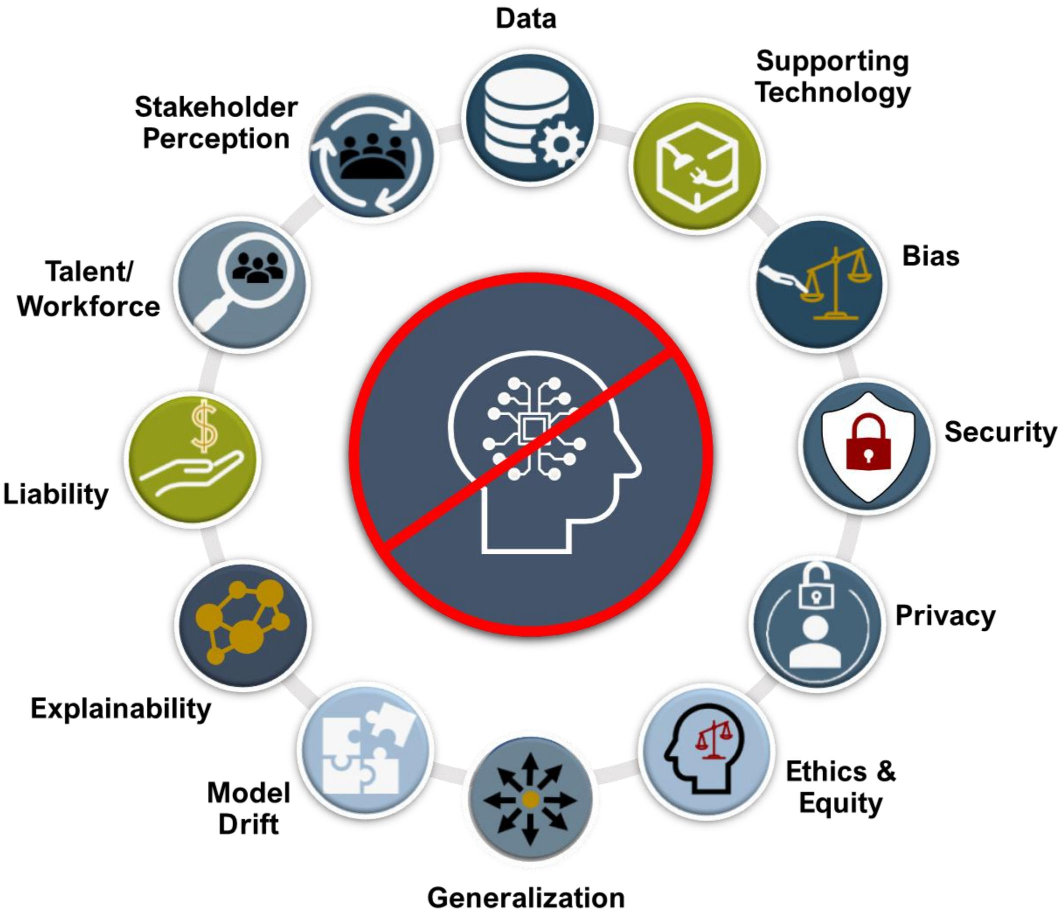

Figure 1. Challenges to AI Adoption for ITS

2.1 Data

Table 3. Summary of the Data Challenge and Potential Strategies to Address It

Summary of Data	
What is it?	Lack of sufficient, high-quality, and relevant data
Why does it matter for ITS?	• Collecting sufficient high-quality data can be difficult when infrastructure-based sensors are sparse, or funding is limited. Labeling large sets of unstructured data can be costly. Data fusion from diverse sources can be complicated and costly. • Data manipulation and feature extraction can have major implications on the performance and capabilities of AI systems because they can allow the AI system to learn from more generalizable data points or capture complex relationships between the features. • AI systems can learn undesired behaviors and relationships in the training phase if the data are not thoughtfully prepared. • A lack of standards for data access and sharing are an impediment to accelerating the maturity of AI-enabled ITS. • Collection of massive datasets presents challenges related to supporting technology and data privacy.
How can it be addressed?	• Using/sharing data with sufficient metadata and documentation so that practitioners are aware of the nuances, potential pitfalls, and recommended uses of the data. • Developing a comprehensive data management strategy to ensure organizational alignment in data governance. • Reusing data as much as possible to reduce duplicated effort for similar use cases. However, just because certain data are available does not mean that they will be useful for the task at hand, so consider use case relevance. • Considering synthetic, imputed, and human/AI collaborative approaches to creating data and performing extensive validation. • Availability does not imply data relevancy or usefulness.

Description of Challenge

Most modern AI algorithms ingest large amounts of data to make inferences or predictions. AI algorithms that are trained on data fall under the category of machine learning (ML). These ML systems begin with data and then infer rules or decision procedures to predict outcomes (GAO, 2021). The most powerful category of ML algorithm in the modern era has been artificial neural networks, which are inspired by functions of the human brain. They have surpassed other ML algorithms in tasks such as machine vision and object detection. Typically, neural networks, especially deep (many layers of neurons) neural networks, require huge amounts of training data to sufficiently tune the many parameters of these large models. In speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment, the Delaware DOT mentioned that a big issue they are finding is the quantity of useable data. For example, they said the penetration rate of equipped vehicles relative to the total number of vehicles on the roadway is low from one of their data providers.

Lack of sufficient high-quality data is a common barrier encountered when deploying AI systems. It would be helpful for agencies to have processes for obtaining, storing, and validating high-quality data to ensure that their AI systems produce consistent and accurate results. Lack of quality control in the data collection process can lead to systems being trained on poor quality data that does not reflect real world deployment use cases. This can appear in many forms. Incorporation bias, i.e., bias introduced at the point a data set is labeled, means that data may be systematically labeled incorrectly or inconsistently based on the viewpoints of the human who labeled the dataset (Heaven, 2021). Another quality control concern is that if data are spliced together sloppily from multiple sources, they might contain duplicates. This can introduce two possible problematic outcomes. If the duplicates appear only in the training data, the model will incorrectly over represent that data point. If the duplicate appears in both the training and the test data, then the model will be evaluated on the same data that was used for training, leading to falsely over-optimistic model performance (Heaven, 2021).

Sparsity of labeled data can be a limiting factor. Many AI systems are built for "supervised learning" tasks – where the system trains by associating feature variables (inputs) with labels (outputs), and then attempts to predict the labels of new unseen data based on the features that are fed in. Labeling datasets is often a costly and time-consuming activity. Another issue pertaining to the data is not having a diversified and appropriately balanced data which can lead to model bias and discrimination.

If sufficient documentation about the origins, nuances, and metadata of datasets does not accompany shared data, practitioners may incorrectly interpret what the variables are measuring. When data are shared without accompanying supporting documentation, practitioners often make incorrect assumptions about what the data are representing, leading to AI applications not suited to their intended use cases (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022).

Data transformation is an important part of the AI development process. Typically, data scientists will transform data to be most suitable to the needs of their use case. This often takes the form of feature extraction, the process of determining what features could be useful for training a model and converting raw data into said features (Google, 2021). These might include data manipulations such as one-hot encodings of categorical data, n-gram transformations for natural text data, Z-score transformations for numerical data, or merging of different data sets. Decisions about the data during development might impact system capabilities and outcomes. Data transformations may also introduce bias, privacy risks, and equity issues, which are addressed in more detail in their respective sections of this report.

AI systems may sometimes learn undesired behaviors and relationships based on features of the data that do not help in their intended real world operational setting. A recent pertinent example is of a ML screening tool for detecting COVID in scans of hospital patients. Because patients who were scanned while lying down, as opposed to sitting, were more likely to be seriously ill, the AI learned incorrectly to predict COVID risk based on a person's position (Heaven, 2021).

Implications for ITS

Some of the ways that data-related challenges could impact AI-enabled ITS applications are summarized below.

- **Sufficient breadth of high-quality, relevant data:** AI applications in ITS will require a wide spectrum of large amounts of relevant data. For transportation agencies, collecting sufficient data can be a major challenge, especially when infrastructure-based sensors are sparse or when funding is limited. Sometimes practitioners may be tempted to use easily accessible datasets that may not be sufficiently relevant or of high enough quality for the intended ITS use case. Staff supporting the Delaware DOT's ATCMTD deployment mentioned in a presentation that, in reality, one has to make tradeoffs between data availability and data quality (Donaldson et al., 2022). Fusing and integrating data from multiple sources and sensors presents another layer of complexity. Depending on the ITS application, practitioners may need to incorporate data sets that account for socio-demographics, weather, traffic counts, travel time/speed, public transportation vehicles and ridership, parking, incidents, etc. (Qian, 2021). Often projects start with vast quantities of unstructured and unlabeled data, which can be especially challenging to manage. AI practitioners may need to label the data or impose structure to make them usable for decision making (Steier, 2021). When the data reflect heterogeneity and social inequalities in populations, it is also possible that the AI use case leads to disparate and possibly harmful outcomes. For more discussion on this topic, see the Bias Section of this report.

- **Making large, disparate data sources useable:** Before getting to ML development, agencies may have to figure out how to integrate disparate data sources. For example, in speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (M. Haselkorn et al., interview, April 2022), the Washington State DOT (WSDOT) explained how the first wave of their project was focused on shared situational awareness across agencies in Seattle. To get there, they built an integrated dispatch feed and an inter-agency action log, which will serve as the data infrastructure backbone for the next wave of ML algorithm development to identify cross-agency incidents. In other cases, agencies may already have access to a sufficient quantity of high-quality and relevant data to use for AI but may struggle to make it useful for AI/ML applications. For example, in speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (E. Kopinski et al., interview, April 2022), the Missouri DOT (MoDOT) mentioned that access to data is not always the challenge. Instead, knowing how to use available data for AI/ML can be difficult, especially when the data come from a variety of different sources.

- **Data manipulation and feature extraction:** Given the potential variety among data sources and formats, ingestion of diverse data into an AI system can be a major challenge. Typically, a large part of the machine learning pipeline are data manipulation and feature extraction. This can have a major impact on the system's performance and capabilities, so practitioners may have to balance mission goal tradeoffs at this phase (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022). For instance, take an application that uses machine vision and satellite imagery to predict traffic density throughout a city. The AI engineer chooses at what resolution the data should be fed to the system. Higher resolution data might help identify small extreme values, like neighborhoods or even blocks that are very traffic heavy, but they also might introduce more noise and require more compute resources. Lower resolution data settings might take fewer computing resources to train the model and represent the system from a high-level view better but might not detect extreme cases that might be indicative of serious inequalities, such as neighborhoods situated in areas of abnormally high congestion. In other cases, an agency may have little control over the data's resolution. Staff supporting Delaware DOT's ATCMTD deployment mentioned in a presentation that having access to higher resolution data would be helpful (Donaldson et al., 2022). For example, the TMC receives travel time information from Bluetooth data at a 5-minute refreshment rate but having that data at a higher resolution (e.g., every minute) would give them more information to detect incidents earlier.

- **Learning undesired behaviors and relationships:** It is possible for AI systems to learn undesired behaviors and relationships at the training phase, which will cause it to be suboptimal or even fail in operational settings, based on oversights by the data scientist. An ITS example might be an AI system trained to identify highway crashes based on images of crash sites taken by traffic cameras (Vasudevan, Townsend, Dang, O'Hara, et al., 2020). If some of the images ingested by the system are from crashes after emergency vehicles have already arrived, it might learn to predict that a crash has occurred by identifying an emergency vehicle, a strategy that would be useless in an operational setting.

- **Standards for data access and sharing:** Recent market research conducted by the USDOT revealed that ITS practitioners believe that a major role for USDOT is in supporting the development of standards to ensure that data can be accessed and shared for execution of AI-enabled ITS applications. Maturity of AI for ITS applications may be accelerated if federal and state agencies effectively establish ground rules for ownership and exchange of ITS data as well as set protocols for secure exchange of data (See Appendix A). In speaking with the ITS JPO's AI for ITS Program about their ATCMTD deployment (G. Donaldson & M. Rosica, interview, June 2022), the Delaware DOT (DelDOT) emphasized the importance of easily accessing data from providers and being able to use it the way they want. A key question that DelDOT's software development team asks data providers is "do you have an API?" because they do not want to have to create the API themselves.

- **Data privacy:** Collection of massive data sets and advances in technology that facilitate correlation of data about individuals may create new risks for agencies that collect data. It might be helpful for agencies to improve their efforts to protect sensitive data and appropriately limit collection of personally identifiable information (PII). Maintaining some data may also be perceived as a potential liability for agencies especially if the video feeds or images from traffic cameras could be used for criminal/civil proceedings upon request. Strategies for anonymization of data may be employed, but they run the risk of reducing the usability of the data for building AI-enabled ITS applications. Therefore, in some cases there may exist a tradeoff between privacy preservation and usability of data. For more information, please see the Privacy Section.

Insights and Lessons Learned

Some potential strategies and lessons learned to address data-related challenges are summarized below.

- **Using/sharing data with sufficient metadata and documentation:** When sharing data either through open-source platforms or proprietary agreements, agencies and private sector data providers might want to consider including sufficient supporting documentation so that users will know how to correctly interpret the phenomena being measured. Practitioners similarly might want to only utilize data procured outside of the organization if there is enough supporting documentation to thoroughly understand the data. Documentation can include metadata, data dictionaries, and a description of the data collection methodology, including possible issues or statistical biases that may be present within the data (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022). Supporting data documentation can outline contexts and contents of datasets, including their motivation, composition, collection process, and recommended uses (Crawford et al., 2021).

- **Supporting data standards development and implementation:** The USDOT and other agencies could participate in and show leadership in the broader industry standards development for data access and sharing. Additionally, the government could have a major role in implementing such standards for government-funded projects or government-controlled data.

- **Developing a comprehensive data management strategy:** Areas covered by data management strategies include enterprise data organization, cross-organization data vision, data governance, metadata management, analytics and regulatory data, and data quality management (Steier, 2021). High-quality data management processes aim to successfully account for changes to the operational environment, possible biases, and the potential for adversarial exploitation throughout the life of a system. Although automation of data functions can create more efficient systems, practitioners may want to be aware of possible tradeoffs between automation and observability (Horneman et al., 2019).

- **Reusing data as much as possible to reduce duplicated effort:** Data wrangling, structuring, and labeling are very costly and time-consuming phases of the AI development cycle. It is therefore important to make labeled data as widely available throughout the organization as possible. Reusing data facilitates teams in rapid-prototype and experimentation with new models. Data management strategies account for the changing nature of data sources and formats. Practices such as rigorous data versioning and mapping versions of datasets to deployed models are recommended (Amershi et al., 2019).

- **Considering synthetic, imputed, and human-in-the-loop approaches to data:** Synthetic data are artificially produced data that are intended to mirror the features of real data (GAO, 2021). There are various reasons an organization might use synthetic approaches: ethical issues associated with collection of real-world data, protection of privacy and personally identifiable information, or because of some dimension underrepresented in a training sample. Extensive validation is required in high impact/risk use cases because practitioners are often hesitant to trust the accuracy of synthetic data. Human bias can be introduced into systems if synthetic data are not thoroughly validated because the data are synthesized based on a set of rules that reflect the perspective of the human who created them (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022). Imputed data are substitute values for missing data meant to maintain the usability of the dataset. There are many imputation methods that carry their own risks in terms of model performance and behavior, especially in cases when the missing values show a pattern (i.e., the data are not missing at random). Finally, a human-in-the-loop approach can create more robust datasets. One USDOT University Transportation Center research project found that by having an iterative approach with human experts labeling samples of data for the AI, they significantly improved their model's performance compared to baseline models. The research team has experts label a small seed dataset, after which a transfer learning model trains itself to label new data, which is then re-reviewed and validated by the experts (Banaei-Kashani & Rens, 2021).

- **Availability does not imply data relevancy or usefulness:** Using data simply because it is inexpensive or easy to find might not be the best approach for building ML models. This often leads to messy and noisy data being used. Agencies might have to actively choose to build better data sets, specifically for building ML models, rather than creating cheap data that are ungoverned and unconsented, which can result in unexpected and highly biased behaviors (NIST, 2022). For example, researchers from Vanderbilt University supporting the Tennessee DOT's ATCMTD deployment decided to manually annotate 350,000 3D boxes with 8 points each to track vehicles in video frames when they could not find sufficient, high-quality existing 3D data to use (Work, 2022). In another example, the Washington State DOT (WSDOT) is developing a Virtual Coordination Center (VCC), which is a dashboard platform that combines data in near-real time from agencies across Seattle to help users identify incidents requiring collaboration from multiple agencies (e.g., large crashes, fires, fatalities, crime, etc.), as part of their ATCMTD deployment (Haselkorn & Webster Heublein, 2022). While their goal is to eventually train a supervised ML algorithm to automatically flag VCC-level incidents, at present, they lack sufficient labeled data to develop it. Therefore, they hope to gather enough labeled data from users manually flagging VCC-level incidents in the platform during the first year of operation to then develop the ML algorithm the next year.

2.2 Supporting Technology

Table 4. Summary of the Supporting Technology Challenge and Potential Strategies to Address It

Summary of Supporting Technology	
What is it?	Inability of legacy systems to support the addition and integration of new AI-based functionalities, due to software-hardware integration issues, limited data storage capacities, and restricted computational power of legacy systems

Summary of Supporting Technology	
Why does it matter for ITS?	• Infrastructure costs associated with AI systems can be high. • Lack of available funding may require prioritization of projects. • Lack of documentation for legacy systems can make them less adaptable at supporting new equipment. • Unlocking the full potential of AI-enabled ITS applications requires resolving integration and compatibility issues, as well as data storage and computational power problems. • ITS equipment requires continuous power supply which can be costly and challenging specifically for rural AI for ITS applications. • AI/ML algorithms have carbon costs associated with them.
How can it be addressed?	• Launching pilot deployments to uncover potential barriers and demonstrate benefits • Leveraging existing ITS infrastructure where applicable to minimize costs • Using cloud computing to increase the computational speed • Using edge computing to overcome bandwidth and latency issues • Leveraging clustered computing to augment processing power • Adopting systems engineering best practices to build more adaptable and resilient systems • Building sustainable AI solutions to reduce negative environment impacts

Description of Challenge

Much of existing ITS infrastructure and supporting technology is based on legacy systems with custom software developments for specific applications and often do not support new technology integration (Systems Engineering Guidebook for ITS, 2009). Legacy systems might be composed of various hardware and software components that are no longer updated or improved. These legacy systems are unable to process and store large quantities of complex data generated by AI applications and have compatibility issues with new interfaces and software integration leading to possible bandwidth, latency, timeout, storage, and communication issues (Vasudevan, Townsend, Dang, et al., 2020; Vasudevan, Townsend, Schweikert, et al., 2020).

The foundation of modern AI applications lies in the quantity and quality of large datasets used to train AI/ML models. Handling, processing, and storing large datasets pose a challenge for agencies which often lack resources to warehouse large datasets, thus inhibiting the potential of AI applications. Furthermore, AI algorithms can be computationally expensive and may require cutting edge technology to process large datasets to draw insights. Legacy systems have limited computational power and are one of the potential barriers for AI adoption. Many organizations and agencies continue to use legacy systems due to higher capital costs of new systems, prior investments and commitments, and challenges/risks posed by migrating to a new system (Problems with Legacy Systems, 2021).

Another challenge with existing systems is adding new functionality, such as combining new software with existing hardware. This can lead to compatibility and integration issues. Integration is the process of combining hardware and software components, sub-components, and interfaces. Legacy systems may have been built to serve a specific purpose and the existing hardware/software may not be able to support new functionalities. In some cases, a complete overhaul of existing systems is necessary, which can be difficult for agencies with budget constraints.

Implications for ITS

Some of the ways that challenges related to existing systems and supporting technology could impact AI-enabled ITS applications are summarized below.

- **Higher cost of ITS support infrastructure and limited funding:** A survey of municipal executives and city officials in North America on the readiness of cities to undertake smart cities initiatives revealed that funding is the top barrier and cities are struggling to find innovative funding alternatives (Learn, 2014). Planning, deployment, and maintenance of systems utilizing AI technologies in ITS contexts involve cost components which include but are not limited to hardware manufacturing, custom software and interface development, data storage, data servers, communications, and power networks. For example, the deployment of Automated Traffic Signal Performance Measures (ATSPM) is typically comprised of the following components which increases the overall cost of implementation (Lattimer, 2020).
    - Advanced traffic signal controllers
    - Power and communication network (Cellular, Fiber, etc.)
    - Central server for data storage and processing data and video feeds
    - Special ATSPM software
    - Detection system (detectors, cameras, etc.)
    - Supporting infrastructure (light poles, ground cabinets, etc.)

- **Compatibility, data storage and processing issues with existing ITS systems:** With a limited budget and lack of well documented and proven benefits for new AI technologies, agencies are typically reluctant to adopt new technologies. Many existing traffic signal systems are not able to support or are not compatible with advanced traffic controllers. Such hardware-software compatibility issues could result in unanticipated delays from integrating systems. Furthermore, traffic management centers often do not have the capacity to store large amounts of data for analytics purposes because AI-based technologies involve massive datasets, such as video feeds from traffic cameras.

- **Continuous power and communication supply needed:** AI based technologies and applications require continuous power supply, communication networks and advanced servers to transmit and receive large amount of data for real-time analytics. This is of crucial importance for rural AI applications in transportation where it may be difficult and extremely costly to bring power and communications. Inclement weather conditions may also cause power outages, disruption in roadway network, damage to the field ITS devices and other supporting infrastructure thus making it as a potential barrier for AI adoption.

Insights and Lessons Learned

Despite the challenges and barriers discussed above, several AI based technologies have made their way into the ITS industry. They have been deployed in small-scale controlled environments such as autonomous shuttles and intelligent traffic signal controllers updating traffic control/signal plans in real-time based on the demand (Lopez Conde & Twinn, 2019). The autonomous shuttles such as Olli (Local Motors, 2021) can analyze traffic conditions and make real-time routing decisions accordingly. Furthermore, some states like Utah and Georgia are among the early adopters of innovative ATSPM systems and have well documented benefits-cost analysis. Many other states are assessing the relevance and benefits of AI technology, with some states in pilot phases. The technology is relatively mature and there are numerous vendors in the market offering advanced signal controllers capable of real-time traffic signal optimization such as adaptive signal controllers (Day et al., 2020; Lattimer, 2020).

Some of the potential solutions are documented below to help accelerate the adoption of AI applications in ITS industry.

- **Launching pilot deployments to uncover potential barriers and demonstrate benefits:** Conducting pilot studies for new ITS applications leveraging AI technology can help overcome some of the barriers for AI adoption. It can help demonstrate the potential benefits of AI technology and help understand the system requirements such as hardware-software integration, data storage, communication flow before large-scale deployment. The concept of conducting pilot studies is not new. The U.S. Department of Transportation is currently supporting the advancement of connected vehicle technology to uncover potential barriers and documenting the lessons learned (USDOT, 2022). An AI based traffic management pilot study conducted by Nevada Department of Transportation in partnership with Nevada Highway Patrol and Regional Transportation Commission demonstrated safety benefits in terms of crash reduction and increase in emergency response times (*AASHTO Journal, 2019*).

- **Leveraging existing ITS infrastructure where applicable to minimize costs:** The deployment of ITS systems such as toll collection systems or express lanes require construction of costly infrastructure, such as fiber optic communication networks, ITS devices (CCTV cameras, DMS, speed detectors, etc.), ITS/light poles, cabinet controllers and switches, and transportation management center (TMC) support. Much of this existing infrastructure components can still be used for various applications and purposes. For example, in speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (G. Donaldson & M. Rosica, interview, June 2022), the Delaware DOT emphasized the importance of their telecommunications system to support AI and other technologies. They have invested over the past 25 years in enhancing and expanding the state's telecom system. Even though technology has changed over the years, they have been able to adapt the telecom, central office system, and field requirements accordingly, which they attribute as a big part of their success. In terms of sensor infrastructure, a camera pole may be used to mount multiple CCTV cameras as well as detectors. For communication networks, spare fiber optic cables (strands) also known as "dark fiber" may be used for new ITS applications by multiple stakeholders. Fiber resource sharing is quite common in the State DOT practice where state-owned right-of-way and fiber network is leased to various private entities through resource sharing agreements (*MDOT SHA*, 2022). VDOT has access to 3,700 miles of fiber with spare capacity of around 2,500 miles which could be used to accommodate communication needs of newer AI-focused ITS applications (Farajian, 2019). Many AI applications rely on data feeds from CCTV cameras for video analytics purposes. Rather than installing new cameras for AI applications, feeds from existing CCTV cameras may be used, thus lowering the cost of AI applications and alleviating budget constraints (ITS JPO, 2022).

- **Using cloud and edge computing to overcome computational, bandwidth and latency issues:** Legacy systems used by many state and local transportation departments often have limited data storage and computational power. Video feeds from a few dozen city cameras may easily impact the capacity of data servers housed in the TMCs. One possible solution to tackle this challenge is by exploring cloud computing platforms such as Amazon, Microsoft, and Google (*Statistica*, 2021). Utilizing cloud computing can alleviate data storage and computing requirements from the TMCs. However, with a large amount of data being collected and transmitted, issues related to bandwidth and latency may still arise. A solution might be to explore edge computing options, which bring computing as close to the source of data as possible to reduce latency and bandwidth use. A research study was conducted to augment driving behavior analytics by incorporating AI/ML algorithms on edge computing platforms (Qi, 2020). Because the majority of the data is collected and analyzed closer to the source rather than being stored on the cloud or TMC, this may help overcome some of the privacy concerns with AI applications related to distracted driver behavior which involve audio-visual data collection. According to one estimate, by 2025 around 75% of the data will be processed outside the traditional data center or cloud (*Gartner*, 2018).

Another possible solution is the use of Field Programmable Gate Arrays (FPGA) and Application Specific Integrated Circuits (ASIC) as hardware accelerators to offload data intensive computation from the CPU and increase bandwidth and reduce latency (Possa 2011). Hardware accelerators have become more prevalent in dealing with large datasets due to their massive parallelism and reconfigurability on the bit level (Danopoulos et al. 2020). Current FPGAs on the market that can deal with these large datasets tend to be very expensive, but cloud platforms such as AWS have made this solution more feasible by hosting these high-powered FPGAs on their rentable instances. These hardware accelerators can be used similarly to an AWS instance and can be significantly easier to integrate for consumers (Lesser et al. 2021).

- **Leveraging clustered computing to augment processing power:** With the lack of processing power in current legacy systems, an alternative solution is the adoption and utilization of multiple computing nodes working as a singular entity. Some of the benefits of using a “cluster” of computers is higher availability in the event of device failure, load balancing to ensure even workload across nodes, and higher performance through the parallelization of tasks across devices to process large datasets quickly and efficiently. Additionally, these computing clusters can be moved closer to the local space to enable lower latency of communication for time dependent tasks (Sharma et al. 2009).

- **Adopting systems engineering best practices to build more adaptable and resilient systems:** With the lack of proper documentation for legacy systems it is challenging to add newer functionalities without a proper understanding of the critical decisions of when, why, and how decisions were made regarding system components and sub-components design. Application of the systems engineering process to ITS projects helps keep better documentation as well as building more adaptable and resilient systems (TxDOT, 2021). Furthermore, FHWA has also developed a guide document to apply scrum and agile methods into ITS project development (Staples et al., 2017). Such requirement-led design of systems coupled with software-hardware integration, verification and validation can help resolve compatibility issues and enable ITS systems to incorporate newer AI based functionalities.

- **Building sustainable AI solutions to reduce negative environment impacts:** As data becomes more available than ever at an unprecedented rate, AI/ML algorithms continue to develop and mature. However, processing such large-scale data and training AI/ML algorithms require very high computational power leading to higher energy consumption. The environmental impacts of AI systems have been studied and documented by researchers more recently and studies have shown that environmental impact of training a single, large deep learning algorithm such as NLP could approach that of carbon emissions of five vehicles over their life span (Lacoste et al., 2019; Strubell et al., 2019). Sustainable AI approaches can help lessen the negative environmental impacts of AI systems by training AI/ML models in a faster and efficient manner, shrinking down the size of models, utilizing fewer compute cycles, obtaining higher utilization rates of existing hardware, and preventing idle power consumption (The Imperative for Sustainable AI Systems, 2021).

2.3 Bias

Table 5. Summary of the Bias Challenge and Potential Strategies to Address It

Summary of Bias	
What is it?	When AI systems lead to unfair and unequitable outcomes due to underrepresentation of subpopulations, or due to human or systemic/institutional biases
Why does it matter for ITS?	• Groups of roadways users may be underrepresented in training datasets used for machine vision applications (e.g., for TSMO applications such as Smart Intersections). • Data are sparse for work zones despite being one of the most vulnerable areas of roadways for fatalities. • When data are collected from users of mobile applications for transportation, the samples may be biased. • Data may be biased due to unequitable distribution of sensors along roadways.
How can it be addressed?	• Adopting a socio-technical systems approach to mitigating bias in AI systems • Bringing together diverse teams for AI systems development • The most accurate model is not always the one with the least harmful impact • Collecting sufficient data to measure error statistics across demographic groups • Choosing fairness metrics that reflect the values of the organization and the groups for whom the AI system has the highest risk of harm • Monitoring bias mitigation is an ongoing process that extends throughout the AI system lifecycle

Description of Challenge

Once just the supporting technology of online advertisement and spam filters, ML is now proliferating through more foundational and high-stakes industries. As it spreads, so does its potential to perpetuate discriminatory practices and adverse social outcomes

(Chouldechova & Roth, 2020). NIST categorizes potential biased outcomes resulting from AI systems in the three following categories (Schwartz et al., 2022):

1. **Systemic biases** result from decisions and practices that organizations undertake that result in some social groups being advantaged or favored over others. They can be a result of institutional biases like institutional racism, sexism, and ageism or lack of consideration of universal design principles. These biases can be present in datasets used to train AI, but also in the processes, norms, and decisions made across the AI lifecycle. Systemic bias can also stem from biases in the hardware used to collect data. Bias maybe introduced through sensors when sensors are tuned and calibrated to be more sensitive for some population than others. For example, a vehicle classification sensor whose parameters are adjusted to classify the distinction between a sedan and sports utility vehicle (SUV) may not accurately classify other vehicle types such as pickup trucks. Choices of location and density of sensor placement may also result in biased outcomes.

2. **Statistical and computational biases** appear from sampling or representational errors that mis-map training data samples to the operational contexts the AI system is deployed in. They can arise when data is systematically (non-randomly) missing about certain social groups and cannot effectively extrapolate when applied in operational settings to groups underrepresented in its training data. The error can also arise because of poor data collection processes, misinterpreted data, mislabeled data, poor encodings of complex social phenomena into simpler mathematical representations, wrong data, treatment of outliers, and imputation factors. Once the AI system is put into operational practice, this can lead to unexpected disparate and unfair outcomes because of the differences between the training data and the real-world data it is being fed.

3. **Human biases** emerge from faulty heuristics in human thought and judgement. They are not unique to human interactions with AI and can be present in any setting, but it is useful to consider them in socio-technical systems that involve AI. They influence how an individual or group of individuals perceive information, such as the outputs of an AI system, and then make decisions based on those perceptions. Humans often possess hidden or unknown biases which can impart bias on the AI system if they are not identified and are passed along during data collection, feature selection, and model building (*Artificial Intelligence and the Economy*, 2022).

NIST recommends addressing AI bias through examining three challenge areas: issues having to do with datasets, issues of measurement and metrics to support testing, evaluation, validation, and verification (TEVV), and issues related to human factors (Schwartz et al., 2022).

- **Challenges related to bias in data** come in many forms. Sampling bias occurs when datasets are based on data samples that are not randomized or are not representative of a population for which the algorithm is making predictions. Often data for ML are scraped from social media or mobile applications and used to build models that are assumed to work on a general population, when they are only representative of users of those platforms (NIST, 2022). Many popular ML models work under assumptions on the data, such unimodal distributions with low multicollinearity; however, data often are multimodal when not disaggregated by demographic features (Schwartz et al., 2022). A naïve assumption of some practitioners is that by removing sensitive attributes from the data, such as race or gender, the system will not be able to produce biased results. However, quite often these models still have the potential to discriminate because of collinearity in variables (e.g., zip code often being a strong predictor of race) (Ghani et al., 2021).

- **Challenges related to TEVV bias** require a holistic view of algorithms, data, and fairness metrics. One risk is that practitioners use a faulty proxy variable in place of a variable of interest that they cannot measure. For instance, a ML system that predicts criminal activity may use arrest data as a proxy for crime; however, there is reason to believe that minority populations are policed at higher rates, making arrests a biased proxy for criminality (Chouldechova & Roth, 2020). Another consideration is sacrificing bias reduction for system accuracy. With heterogeneous data, a system may optimize overall accuracy by performing well on the majority group within the dataset and sacrificing performance on minority groups (Chouldechova & Roth, 2020). It is possible and common that a variable is positively correlated with the target variable with the majority group but negatively with the other groups (Ghani et al., 2021). Finally, human designers often make decisions about what variables to exclude or include in a model, another avenue for potential bias (Schwartz et al., 2022).

- **Challenges related to human factors** often surface when AI systems are deployed in real world settings. Frequently, as time passes after an application is deployed, the users repurpose or use it in unforeseen ways (Schwartz et al., 2022). When the system is informing some intervention, organizations run the risk of creating feedback loops, scenarios where the system is informing decisions that will impact future data and therefore system behavior. The example of an AI system for crime prediction is instructive. If the system is being used by police to decide where to patrol, it is likely that more arrest data will come out of the areas the system is already suggesting, potentially accumulating more bias over the lifetime of system operation (Chouldechova & Roth, 2020).

Implications for ITS

Some of the ways that AI bias could impact ITS are summarized below.

- **Groups of roadway users underrepresented in training datasets:** Many ITS systems that leverage AI will have primary functionality based on machine vision algorithms. For instance, agencies are already deploying applications that use object detection on CCTV video camera feeds to understand traffic and pedestrian densities and flows in real-time (Ozbay, 2022). For these applications to function robustly, they have to be trained on sufficient images of road-user types. For instance, vehicle types with limited market penetration, such as 3-wheel motorcycles or dirt bikes, might not be sufficiently represented in training data.

- **Sparsity of work zone data:** According to FHWA, there were 857 total work zone traffic fatalities in the year 2020. This represents up to 3% of all workplace fatalities every year. Because of the elevated risk of traffic incidents in work zones, it is vital that AI ITS systems be robust to work zone conditions which generally do not resemble other road segments. For instance, automated vehicles could struggle to recognize construction workers, who often do not resemble other pedestrians due to equipment such as orange vests, as humans. One project addressing this is NYU C2SMART University Transportation Center, who have developed a VR testbed to collect higher quantities of work zone data (Ergan et al., 2021).

- **Biased sampling due to mobile applications:** Many transportation services are now tied to mobile phone applications. The companies that run these services collect data on their user bases and at times share that data with agencies for partnerships and planning. This data will likely be non-representative if carelessly used as representative of the entire ecosystem of transportation users or as a proxy for transportation services demand. For example, people living in transit deserts are often dependent on ride-hailing services (Schwartz et al., 2022). If an AI model uses those data for predicting demand of transportation service types, it might come to the conclusions that demographics from those neighborhoods prefer ride hailing, when they actually exhibit that behavior because of lack of feasible alternatives.

- **Biased sampling based on sensor locations:** Almost always, sensors of different types are not distributed uniformly across geographic areas. Usually, sensors are most dense in areas where most humans live (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022). There can also be unequal distribution of sensors across geographies that correlate with demographics, such as having more robust sensing units in higher income areas. This can impact and bias AI system performance in a myriad of ways. For instance, a state agency might implement an AI system that uses roadside sensors to predict road weather conditions. In less populated rural areas, there might be less density of robust sensors, leading to worse predictions about snowfall or heavy rain in these areas.

Insights and Lessons Learned

Some potential strategies and lessons learned to identify, mitigate, and address bias in AI are summarized below.

- **Adopting a socio-technical systems approach to mitigating bias in AI systems.** An organization’s goal is not only to verify that the AI model is fair, but also that the overall system and its outcomes are fair (Ghani et al., 2021). The organization could consider adopting processes that include involving stakeholders, examining cultural dynamics and norms, and assessing societal impacts (Schwartz et al., 2022). Different stakeholder groups might have different ideas of what constitutes bias in an operational setting, so organizations could consider working across these groups to come to a reasonable consensus (GAO, 2021).

- **Bringing together diverse teams for AI systems development.** Socio-technical approaches require intimate understanding not just of engineering, mathematics, and computer science, but also human and social factors that will determine outcomes in operational settings. As such, teams developing AI systems may want to include experts on cognitive science, social science, human factors engineering, design, and more. They can also include people of different demographics and experiences who can identify a wide variety of concerns. Red teams (teams that identify vulnerabilities, risks, and failure modes of systems), in particular, can benefit from diverse perspectives to be able to identify a wide range of vulnerabilities (NIST, 2022).

- **The most accurate model is not always the one with the least harmful impact.** Selecting models solely on accuracy is usually not the best approach to mitigating bias. The choice of the model’s objective function itself might encode bias. Sub-populations may be harmed because the algorithm over-indexes on the majority (Schwartz et al., 2022). One technique to tackle this issue is by resampling and/or reweighting data points corresponding to protected groups in the algorithm (Ghani et al., 2021).

- **Collecting sufficient data to measure error statistics across demographic groups.** Models that are “blind” to protected status such as race or gender can still discriminate. Because of correlations among covariates, simply not including protected status categories as features, such as race or gender, in the ML algorithm does not guarantee that the algorithm will not lead to discriminatory outcomes (Ghani et al., 2021). Instead, practitioners may want to collect sufficient data that is more diverse and appropriately balanced to measure disparate impacts across relevant groups so that bias can be identified and remediated.

- **Choosing fairness metrics that reflect the values of the organization and the groups for whom the AI system has the highest risk of harm.** When attempting to quantify fairness, there are many different metrics that AI teams may consider. They may choose to look at the following metrics and many more across groups: false positive rate, prevalence, false discovery rate, and false negative rate. However, it has been proven that except for in trivial settings, it is impossible to simultaneously equalize all of these metrics (Chouldechova & Roth, 2020). Therefore, it is helpful for AI teams to understand what metrics members of affected groups consider most fair in a given operational setting (Ghani et al., 2021). For instance, for an ML algorithm that decides whether to grant bail based on predicted risk of crime recidivism, is it more important that the chances that a given Black or white person will be wrongly denied bail is equal, or that for people who should be released, the chances that a given Black or white person will be denied bail is equal (Ghani et al., 2021)?

- **Monitoring bias mitigation is an ongoing process that extends throughout the AI system lifecycle.** Risk measurement is not a static activity. AI teams may want to consider not only continuous data collection on fairness metrics after system deployment, but also be willing to reevaluate the metrics decided on pre-deployment as new information about the operational context is understood (NIST, 2022). There is no one set list of fairness metrics that works in every context, and it is inevitable that the framework of measures a team decides on originally will not be entirely correct. As the team learns more about the adverse and disparate social effects that the system might cause post-deployment, they can reevaluate their metric collection based on core organizational values (NIST, 2022).

2.4 Security

Table 6. Summary of the Security Challenge and Potential Strategies to Address It

Summary of Security	
What is it?	When an attack, via cyberspace, targeting an enterprise's use of cyberspace for the purpose of disrupting, disabling, destroying, or maliciously controlling a computing environment/infrastructure; or destroying the integrity of the data or stealing controlled information

Summary of Security	
Why does it matter for ITS?	• Insufficient attention has been paid to the ways in which AI can be used maliciously. • Malicious entities could compromise the integrity of the decision-making process (e.g., data poisoning, model evasion). • Hacking cyber-physical infrastructure (e.g., DMS, Colonial Gas Pipeline) poses a threat. • Connected vehicle adoption may increase vulnerabilities. • Automated vehicles, which rely heavily on AI algorithms, make safety-impacting driving decisions. • Agencies and the public may mistrust AI applications.
How can it be addressed?	• Understanding potential security threats from misuse of AI-based applications to better forecast, prevent and mitigate the threats • Following cybersecurity best practices • Collaboration among various stakeholders to identify transportation cybersecurity best practices • Developing workforce and domain expertise to curtail security issues • Strengthening the security of AI systems by addressing vulnerabilities • Utilizing intrusion and misbehavior detection systems to enhance safety of AI-based ITS systems • Retraining ML models at regular intervals to retain the quality of ML predictions • Securing physical infrastructure to block potential physical intrusion into the system

Description of Challenge

While AI/ML applications continue to develop and mature at an unprecedented rate as data becomes more available, **their wide-scale adoptability in the transportation sector is still limited primarily due to safety and security concerns**, as well as lack of comprehensive standards related to automated vehicles (Koopman et al., 2019; Koopman & Wagner, 2018). Quantifying the benefits of adopting AI-based technology is well documented and has been a topic of interest in the industry, academia, and government bodies. At the same time, less attention has historically been paid to the ways in which AI/ML applications can be exploited by bad actors to compromise the decision-making of AI systems (Brundage et al., 2018). Security and safety of cyber-physical infrastructure is of paramount importance and a potential barrier for AI applications.

Cyberattacks continue to be on the rise. The automotive industry is one of the most critical industries vulnerable to cyberattacks (The Road to Secure and Trusted AI, 2021).

Cyberattacks can be broadly classified into two categories. In the first category, attacks are launched through malicious use of AI to steal sensitive information through impersonation. In the second category, adversaries exploit the inherent weakness in AI to steal information, distort the data, and corrupt the ML models.

Cyberattacks through Malicious Use of AI for Impersonation. According to Statista, a survey of 309 business leaders conducted in 2021 documented potential scenarios of AI-enabled cyberattacks worldwide (Types of AI-Enabled Cyberattacks 2021, 2022). Nearly 68% of the respondents indicated that AI can be used for impersonation and spear phishing-attacks. Spear phishing attacks attempt to acquire sensitive information or access to a computer system by sending counterfeit messages that appear to be legitimate. Other types of attacks mentioned included, ransomware, misinformation, undermining data integrity, targeting networks, and deepfakes.

Cyberattacks through Exploitation of AI Weakness. While machine learning algorithms, such as deep neural networks, have achieved impressive results in terms of image classification, they can be very unstable to small perturbations of the images (Steinhardt et al., 2017). AI specific cyberattacks can occur during various stages of ML model development i.e., during training and/or production (predictions). Adversarial machine learning attacks use malicious inputs designed to fool machine learning models. These attacks can include data poisoning, where malicious actors inject false training data with the aim of corrupting the trained model and compromising the classification and/or detection results (Steinhardt et al., 2017). Another form of adversarial machine learning attack is evasion, which happens at the inference or model prediction stage. In this type of attack, the data is distorted and manipulated (i.e., changes in pixels of an image) so that the model fails to correctly classify the results accurately (Suciu et al., 2018).

Implications for ITS

- **Insufficient attention has been paid to the ways in which AI can be used maliciously.** The benefits of AI systems are well understood and documented; however, less attention has been paid to the ways in which AI/ML algorithms can be exploited to compromise the overall accuracy of the models as well as compromising their decision-making process, especially in transportation sector.

- **Malicious entities could compromise the integrity of the decision-making process (e.g., data poisoning, model evasion).** There are several ways in which malicious actors can exploit AI/ML algorithms and impact decision outcomes. For self-driving cars, an image of a stop sign can be tampered with by changing a few pixels in specific ways through a data poisoning attack. In this scenario, although humans can still recognize as a stop sign, it nevertheless could be misclassified by an ML model (Brundage et al., 2018). Speed limit signs can also be tampered with in a way that a self-driving vehicle would misclassify the speed limits and pose danger to the operation of the transportation network.

Malicious entities could also take direct control of a vehicle, manipulate the operation of vehicle, and exploit the vulnerabilities in the systems using adversarial ML, which may have dire consequences on safety.

- **Hacking of cyber-physical infrastructure poses a threat.** Historically, the ITS industry has experienced fewer cyber related attacks compared to other industries like the internet or biometrics; however, it is highly vulnerable to such attacks. There have been some cyberattacks on critical infrastructure and transportation systems recently. Dynamic message signs (DMS) which serve as a bridge between transportation agencies (TMCs) and traveling public have historically been hacked several times (Kelarestaghi et al., 2018). A more recent ransomware attack halted the operation of the Colonial Gas Pipeline, for which state of emergency was declared calling for a coordinated response to the threat. Vasudevan et al., 2020 highlighted other potential cyberattacks against ITS systems including hacking into and changing traffic signs, hacking automated fare payment systems, and manipulating communication between first responders.

- **Connected vehicle adoption may increase vulnerabilities.** According to researchers, as the connected vehicle market adoption rate increases, cyber threats will increase over time. Connected and automated vehicles utilize numerous communication technologies and are very susceptible to cyber attacks. These communication technologies include but are not limited to Dedicated Short Range Communication (DSRC), Satellite Communication, 3G, 4G and 5G cellular, Wi-Fi and LiDAR. Cyber attacks against connected vehicles and ITS systems will be of high impact as the cyber attacks might potentially lead to multiple vehicular crashes, create traffic jams, impact traffic operation leading to a ripple effect causing huge financial losses (Huq et al., 2017).

- **Automated vehicles, which rely heavily on AI algorithms, make safety-impacting driving decisions.** SAE level 5 autonomy requires that the AVs take complete control of driving conditions and make decisions (SAE J3016C, 2021) such as steering control for movements, emergency braking when the hazards are detected on the road, acceleration and deceleration, stopping at a stop sign, etc. AVs will make such decisions based on detection systems by incorporating data and communication from various sources such as sensors, LIDAR, Roadside Units (RSUs), and navigation systems. For level 5 autonomy, the AI/ML algorithms will need to accurately classify objects such as regulatory speed limits, stop signs, and roadway hazards under all conditions. Furthermore, AI/ML algorithms used in AVs will also need to be trained to identify tampered roadway signs to avoid adverse outcomes.

- **Agencies and the public may mistrust AI applications.** These types of attacks on cyber-physical systems are of significant concern for public agencies like state and local departments of transportation and can also have a negative impact on public perception about safety of transportation systems leveraging AI technology. A wide-scale implementation and adoption of AI systems in the ITS sector will require enhanced safety and security throughout the operational lifetime of the systems as identified in the Asilomar principles (AI Principles, 2017). Please see the Stakeholder Perception Section for more information.

Insights and Lessons Learned

The growing number of cyberattacks and exploitation of AI/ML by malicious actors in various industries have raised security concerns to consider. While the number of cyberattacks in the ITS domain have been less than other industries, the threat still exists. As higher levels of vehicle autonomy are achieved and automated vehicles penetrate the market, the number of cyber-physical attacks are also expected to rise. Some of the potential solutions to help address the security concerns of cyber-physical systems are discussed below.

- **Understanding potential security threats from the misuse of AI-based applications to better forecast, prevent and mitigate the threats:** As discussed earlier, insufficient attention has been paid to the ways in which AI/ML systems can be manipulated by bad actors. This manipulation can take several forms like malicious entities hacking in to and taking control of automated vehicles or exploiting the AI/ML algorithms to compromise the decision-making by data poisoning and adversarial machine learning. To mitigate these potential security threats, the ITS AI community, agencies and academia will need to understand the possible scenarios in which systems can be exploited in order to help forecast, prevent and mitigate the threats (Brundage et al., 2018).

- **Following cybersecurity best practices:** Agencies can improve the overall security of their systems, including AI components, by following general cybersecurity best practices. The USDOT's *Cybersecurity and ITS Best Practice Guide* (Krause et al., 2019) highlights security best practices, such as changing default usernames and passwords, encrypting communications between the TMC and ITS infrastructure, turning off unused ports and protocols, and conducting penetration testing.

- **Collaborating among various stakeholders to identify transportation cybersecurity best practices:** Collaboration between federal and state agencies, academia, and practitioners as well as public-private partnerships can help reduce cyber risks to the nation's critical infrastructure (NIST, 2019). Close collaboration between agencies, industry, and academia can help bring potential solutions and best practices to the industry and/or expand the existing capabilities of algorithms to secure AI systems. For example, DARPA's 2016 Grand Challenge called "the Spectrum Collaboration Challenge (SC2)" encouraged researchers to use AI to optimize use of the wireless spectrum (DARPA, n.d.). Additionally, the USDOT created the "Inclusive Design Challenge" in 2020 to seek solutions from industry and academia to enable people with disabilities to use AVs to access jobs, healthcare, and other critical destinations researchers (USDOT, 2022b). A similar challenge or other outlet for cross-sector collaboration could benefit the security of AI systems.

- **Developing workforce and domain expertise to curtail security issues:** State DOT staff has traditionally been composed of transportation planners, engineers and designers with a civil/architecture engineering/planning background, and often lack resources and skills to implement and maintain AI-based ITS systems. With the advent of ITS applications in transportation sector, domain expertise has evolved. However, the agencies will need to develop the skillsets of existing staff, expand, and diversify the workforce across cybersecurity domains, as well as collaborate with other industries to implement the best practices to curtail security issues.

- **Strengthening the security of AI systems by addressing vulnerabilities:** While ML techniques such as deep neural networks offer predictions and accuracy at a great level, they are still vulnerable to small perturbations in the input data. As discussed earlier on, bad actors can distort and manipulate the images which may seem normal to a human eye, but which would be misclassified by ML algorithms. Some robust algorithms like DeepFool have been introduced recently that detect the perturbations that fool deep networks, enhancing the reliability and robustness of ML classifiers against adversarial attacks (Heaven, 2019; Madry et al., 2019; Moosavi-Dezfooli et al., 2016).

- **Utilizing intrusion and misbehavior detection systems to enhance safety of AI-based ITS systems:** Intrusion detection systems have received much attention in the computer science domain. They train AI/ML algorithms to detect network intrusions, behavior-based anomalies, and misuse (John et al., 2016; Liao et al., 2013). These systems are very common in daily use applications such as spam/phishing email classification or real-time credit card fraud alert systems. In the ITS domain, recent similar work was conducted at the University of Virginia, in which a proposed rule-based misbehavior detection system detects and classifies false information through vehicle communication technologies (Gyawali & Qian, 2019). Although these systems can detect anomalies with a high confidence and accuracy, more work is needed to enhance the safety and security of AI systems as it has direct impact on public safety and critical operations.

- **Retraining ML models at regular intervals to retain the quality of ML predictions:** A way to retain the quality of ML predictions and to remove data poisoning, noise and bias issues is to retrain the ML models at regular intervals. Maintaining the predictive power of deployed ML models is difficult and may decline over time. With the possible addition or deletion of data points, retraining the models may be necessary as real-world data keeps evolving and older training data may no longer be a good representation of the real-world.

- **Securing physical infrastructure to block potential physical intrusion to the system:** Another possibility to block intrusion into AI-based ITS systems is to secure the physical infrastructure. ITS data from field devices is sent to the TMC via fiber optics or cellular communications. The cabinet control devices for CCTV cameras, DMS signs, speed detectors, weather sensors and incident detection cameras may be vulnerable to possible intrusion. The device power and communication control come from the cabinets housing device controllers. Physically securing the field devices, such as by installing attack resistant cabinet door locks, can block potential physical intrusion to the system. Additionally, limiting who has control access of physical infrastructure could help with security. For example, in speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (E. Kopinski et al., interview, April 2022), the Missouri DOT (MoDOT) explained that for the traffic vision camera feed, their vendor can only look at the view set by MoDOT. The vendor cannot physically control the camera (e.g., zoom).

2.5 Privacy

Table 7. Summary of the Privacy Challenge and Potential Strategies to Address It

Summary of Privacy	
What is it?	Inability of an AI system to protect individual privacy, including personally identifiable information (PII) and other sensitive information
Why does it matter for ITS?	• New AI applications in ITS could heighten identity, behavioral, and location privacy concerns. • Privacy leakage could lead to liability issues for the agency and reduced trust from system users. • Agencies may have to consider tradeoffs between privacy and utility in data. • AI applications in ITS that may rely on or capture sensitive information, such as pedestrian detection, automated license plate readers, personalized traveler information, and driver monitoring, could pose higher privacy risks.

Summary of Privacy	
How can it be addressed?	• Obscuring/encrypting sensitive data • Collecting non-sensitive data • Using synthetic data • Applying differential privacy • Using a distributed protection technique • Using edge computing to limit PII collection • Establishing data sharing techniques • Developing privacy policies

Description of Challenge

According to NIST (NIST Glossary, n.d.), sensitive information is information where the loss, misuse, or unauthorized access or modification could adversely affect the national interest or the conduct of federal programs, or the privacy to which individuals are entitled under 5 U.S.C. § 552a (the Privacy Act) (GAO, 2021). AI uses massive amounts of data that could impact the privacy of individuals and/or institutions through data manipulation, speech, face or image recognition, and tracking (Vasudevan, Townsend, Dang, et al., 2020).

Even if sensitive data are secured, ML models trained on sensitive data could be exploited to reveal it. For example, ML models could leak information about the individual data records on which they were trained via a membership inference attack. In a membership inference attack, given a data record and black-box access to a model, a attacker can determine if the record was in the model's training dataset (Shokri et al., 2017). An attacker could use a membership inference attack to reveal or reconstruct data used to train the original ML model, which could reveal sensitive data (if it was used in model training). Membership inference attacks are possible based on the concept of overfitting. A trained ML model will likely perform better on training examples than those it has never seen before.

Implications for ITS

As the scope and breadth ITS applications continue to advance, a primary concern regarding the tracking of both people and goods within the transportation system is that of privacy (Fries et al., 2011). Researchers classified privacy issues in ITS into three broad categories: (1) identity privacy, (2) behavioral privacy, and (3) location privacy (Hahn et al., 2019). Identity privacy refers to the privacy of a driver, passenger, traveler, pedestrian, or other user's real-world identity, which could be identified via their driver's license number, name, biometric data, or other personal data. Behavioral privacy refers to the privacy of data that describes various aspects of a group or individuals and their actions within ITS, such as travel patterns. Finally, location privacy within ITS refers to the right of a user to travel or move about the system without concern of their location being exposed.

While privacy concerns in ITS are not new, new AI/ML applications in ITS could heighten them. If sensitive data are communicated to and processed by AI systems, then there are new opportunities for those data to be leaked or hacked. Therefore, agencies may have to consider tradeoffs between privacy and utility in data. Deep learning in particular, such as computer vision applications, may bring tension between needing large amounts of labeled data and its usage that can reveal PII. While anonymization techniques, such as generalization and bucketization, are designed to protect the privacy, they may reduce the utility of the data (Li and Li, 2009).

A few example AI applications in ITS where privacy could be a challenge, since they may rely on or capture sensitive information, are summarized below.

- **Pedestrian detection:** Computer vision applications that detect individual pedestrians could run the risk of leaking identity, behavioral, and/or location privacy information. For example, behavioral privacy could be a concern if pedestrians are tracked over time (e.g., the same pedestrians wait at the same transit stops on weekdays).

- **Automated License Plate Readers (ALPRs):** ALPRs are AI-enabled camera systems that automatically capture license plate numbers that come into view. They may also capture additional data, such as the location, date, time, and vehicle attributes (e.g., color, make, model). Since a license plate number is connected to an individual vehicle registered in one state, it could be linked to even more sensitive personal information. A report by NHTSA identified in their literature review that some watch groups have expressed concerns regarding the potential that law enforcement could use ALPR data to track people for illegitimate purposes and target communities based on race, religion, or ethnicity (Zmud et al., 2021). If an ALPR system is abused, it could lead to the agency losing access to its ALPR program or worse. To mitigate this issue, many law enforcement agencies have developed policies that protect the privacy of the data collected through their ALPR programs (Zmud et al., 2021).

- **Personalized traveler information:** To benefit individual users, AI-enabled personalized traveler information applications require individual data (e.g., GPS and trajectory data, type of locations frequented, etc.). For example, many route planning applications use AI to learn individual user preferences and driving patterns to make route recommendations (Vasudevan, Townsend, Dang, O’Hara, et al., 2020). However, there is a risk of location, behavioral, and possibly identity privacy leakage if the data and model are not fully protected.

- **Driver monitoring:** AI can be used to detect whether a driver is in the vehicle and even recognize who is operating the vehicle using facial recognition. AI can also detect distracted, drowsy, and impaired driving by monitoring head position, eye openness, posture, and other features (Vasudevan, Townsend, Dang, et al., 2020). These driver monitoring applications could impact identity and behavioral privacy.

Insights and Lessons Learned

Researchers and deployers have used a variety of strategies to protect privacy, some of which are summarized here. Leveraging a combination of strategies is more likely to preserve privacy than using a single strategy. Some privacy preservation strategies that may be useful for AI applications in ITS are summarize below.

- **Obscuring/encrypting sensitive data:** Obscuring the data used for AI applications can help protect privacy. For example, the sensitive data could be removed, masked, or coarsened. Removing sensitive data could include hiding sensitive columns or replacing sensitive strings. Masking sensitive data could include a variety of techniques such as using a substitution cipher (i.e., replacing sensitive values with encrypted ones) or tokenization (i.e., substituting sensitive values with non-sensitive dummy values or "tokens") (*Considerations for Sensitive Data within Machine Learning Datasets | Cloud Architecture Center*, n.d.). Another masking strategy is homomorphic encryption, which is a form of encryption that allows users to perform computations on the data without decrypting it (i.e., perform "encrypted ML") (Thaine, 2020). Coarsening sensitive data could include rounding or binning specific values. For example, as part of the USDOT Connected Vehicle Pilot Deployment Program, the New York City pilot site released event records that obfuscated time, location, and vehicle-specific identifying data elements originally contained within the raw field collected records (*ITS Data Sandbox - NYCDOT - DataProcessing.Txt*, 2017/2021).

- **Collecting non-sensitive data:** In some cases, non-sensitive data may act as a viable alternative to sensitive data. For example, researchers have explored the use of thermal imagery as a privacy-preserving alternative to video imagery for pedestrian detection (Kieu et al., 2019). In another example, researchers from the Connected Cities for Smart Mobility towards Accessible and Reliable Transportation (C2SMART) University Transportation Center (UTC) developed a continuous, real-time pedestrian detection framework that uses public CCTV feeds and deep learning-based video processing to analyze sidewalk and roadway density (USDOT Office of the Assistant Secretary for Research and Technology, 2021). This approach preserves privacy due to both the low-resolution nature of the camera feeds and the conversion of vehicles and pedestrians into untraceable objects. In speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (G. Donaldson & M. Rosica, interview, June 2022), the Delaware DOT expressed that they are very careful about the type of data they collect and who has access to it. Most of the data they use is non-sensitive, public data.

- **Using synthetic data:** According to the GAO, synthetic data are artificially produced data that are intended to mirror the features of real data. They provide an approach to preserve privacy when systems use sensitive or personally identifiable information. Synthetic data can serve as a practical replacement for the original sensitive data (GAO, 2021). However, to be of value, it is helpful for the synthetic data to be representative of the intended population, and it is also helpful to document and explain any assumptions made.

- **Minimizing model overfitting:** Model overfitting is the key weakness that allows membership inference attacks to be successful. Therefore, minimizing overfitting is one way to help prevent them. Regularization is one strategy that could help minimize overfitting. Regularization is a class of methods (e.g., Lasso, ridge regression) that fit a model using all predictors but constrain/regularize the coefficient estimates to reduce overfitting (Chouldechova, 2017).

- **Applying differential privacy:** Differential privacy is a “formal mathematical framework for quantifying and managing privacy risks” and can provide provable protection against a wide range of potential attacks, including those in machine learning (Wood et al., 2018). Random noise is added during processing to generate ambiguity downstream so that privacy-impacting inferences cannot be made based on ML model predictions (Prabhu, 2020). There are a number of opensource privacy protection libraries available that include implementations of differential privacy.

- **Using a distributed protection technique:** In some cases, using a distributed or decentralized privacy protection technique, such as federated learning or secure multi-party computation (MPC), may make sense. Federated learning is “on-device” (Thaine, 2020) or “collaborative” machine learning that enables a shared model to be learned while keeping all training data on the separate devices or servers (“Federated Learning,” 2017). Secure MPC or “privacy-preserving computation” is technique that allows for the distributed computation of a function over distributed inputs without revealing additional information about the inputs, such as sensitive data (Frikken, 2011). For example, in healthcare, secure MPC has been applied to enable pharmacological collaboration and genome-wide association analysis while keeping individual patient records private (Telenti & Jiang, 2020).

- **Using edge computing to limit PII collection:** Edge computing where the data are collected and processed as close to the source as possible without storing it on servers could help mitigate some privacy issues. For example, an AI application for pedestrian detection that processes video feeds could use edge computing to limit collection and storage of PII by only transmitting derived pedestrian counts.

- **Establishing data sharing agreements:** According to the US Geological Survey (USGS), data sharing agreements are formal contracts that detail what data are being shared and the appropriate use for the data (*Data Sharing Agreements | U.S. Geological Survey*, n.d.). These agreements may be beneficial when proprietary and/or sensitive data are being shared across multiple entities. The type of agreement or agreements to consider will depend on the relationship between entities sharing the data. For example, according to recent market research on AI for ITS (See Appendix A), a majority of respondents mentioned using at least some form of proprietary data for their AI applications, which were either purchased from agencies for whom the AI applications were being developed or were provided by the agencies at no cost through data sharing agreements.

- **Developing privacy policies:** Clearly defined policies and guidelines on what type of user data may be tracked, when, and for what purpose are helpful (Vasudevan, Townsend, Schweikert, 2020). For example, in one case study identified by NHTSA in its state of the practice report on ALPRs (Zmud et al., 2021), significant negative public sentiment regarding potential privacy concerns over law enforcement's use of ALPRs prompted the state legislature to hold hearings, which resulted in the development of a statewide policy with clear guidelines on data protection, access, collection, and retention. Interviewees from this case study indicated that the existence of this policy to protect resident privacy resulted in the agency receiving no privacy complaints about their agency's use of ALPRs. In another example, in speaking to the ITS JPO's AI for ITS Program (T. Geara et al., interview, April 2022), the City of Detroit mentioned that their city council is adamant about protecting privacy. An ordinance came out a few years ago on how cameras can be used, including how long the video recording can be stored. The City of Detroit abides by that in their ATCMTD deployment and emphasized that all the data falls under the Freedom of Information Act. Similarly, the Missouri DOT (MoDOT) shared with the ITS JPO's AI for ITS Program that per MoDOT policy, videos will be deleted immediately to protect privacy (E. Kopinski et al., interview, April 2022). Additionally, they use high-mounted freeway monitoring cameras that do not clearly show license plate numbers, which also helps to protect behavioral and location privacy.

2.6 Ethics and Equity

Table 8. Summary of the Ethics and Equity Challenge and Potential Strategies to Address It

Summary of Ethics and Equity	
What is it?	When AI applications, whether intentionally or unintentionally, profile and discriminate against individuals/populations based on unfair or unclear criteria or lead to unethical or inequitable outcomes.
Why does it matter for ITS?	• AI-enabled ITS systems can inform, or even make, decisions that greatly impact human lives. For example, Automated Vehicles (AVs) could encounter major ethical dilemmas in their driving decision making. • Disadvantaged populations could be unfairly discriminated against via AI-enabled ITS, negatively impacting equity. For example, discrimination could occur in infrastructure and asset management decisions or in language processing. • Inequitable outcomes could occur from competing objectives or biased data collection. • Seemingly negligible development choices, such as spatial resolution or sensor placement, could lead to unintentional consequences.
How can it be addressed?	• Creating AI systems with ethics, equity, and transparency at the forefront • Translating ethical frameworks into engineering • Supporting workforce training and education to meet future AI needs • Including diverse stakeholders throughout AI development • Promoting ethical, trustworthy AI development and use • Applying guidelines to promote responsible AI • Documenting processes, success metrics, and expectations • Including a human-in-the-loop for critical decisions

Description of Challenge

Ensuring ethics and equity in AI systems is a major challenge and focus for its successful implementation. This topic is wide-ranging and different aspects of it are covered in this report in the Bias, Security, Privacy, Explainability and Liability sections. Focusing on ethics and equity is important to ensure that AI systems can reach their maximum potential and gain public trust without inflicting harm on vulnerable populations, the workforce, and the public in general (Vasudevan, Townsend, et al., 2022).

AI systems can inform, or even make, decisions that greatly impact human lives. The data that many AI systems train on can include sensitive and confidential information about individuals. Because of these aspects of AI, having a robust and integrated approach to ethics is necessary (AI World, 2021) as hackers can misuse the information or poison the data for making unethical decisions. Additionally, even if systems are not hacked, the AI applications can lead to unethical decisions. Disadvantaged populations could be unfairly discriminated against via AI-enabled ITS, negatively impacting equity.

Jobs are already being replaced due to AI implementation. According to the IEEE European Public Policy Initiative 2017 statement on AI (IEEE European Public Policy Initiative, 2017), AI applications will continue to substitute humans in repetitive, less skillful work or critical tasks (such as in medicine). AI adoption can replace humans in tasks that are susceptible to human-error such as video monitoring for incident detection where a person may become tired and inattentive after hours of monitoring. AI could also upset existing system industries (i.e., manufacturing, energy, medical systems, etc.), with potential consequences in terms of jobs or economic strength in these industries (IEEE European Public Policy Initiative, 2017; IEEE Advancing Technology for Humanity, 2019).

Many organizations do not appear to have a systematic approach for cultivating trust among stakeholders in AI-enabled applications. Recent market research on AI for ITS (See Appendix A) revealed that a majority of respondents to the AI for ITS Sources Sought Notice (SSN) demonstrated an ability to make their systems transparent, while considering the legal and ethical implications, but lacked a systematic approach for cultivating trust among stakeholders. As one respondent noted, for gaining trust, it is critical for the AI system to consistently produce outputs that the system’s operators consider reasonable. Conversely, a single error could “foul that trust for a long time.” For the ethical operation of AI, the respondent noted that the AI decisions should minimize bias and be fair, transparent, responsible, and interpretable (Vasudevan, Townsend, et al., 2022). AI systems need to be explainable, with their input and process able to be examined to prevent the perception of a confusing and mysterious “black box.” Overall, most of the respondents recognized the need for building trustworthy and ethical AI systems but lacked systematic approaches for achieving these aims.

The public has expressed concerns regarding fairness and acceptability with using AI-enabled ITS applications for critical decisions. A 2018 Pew Research Center survey on “Public Attitudes Toward Computer Algorithms” (Smith, Aaron, 2018) found that the public has expressed broad concerns about the fairness and acceptability of using computers for decision-making in situations with important significant real-world consequences. There are several themes driving these concerns. Some of the more prominent concerns mentioned in response to open-ended questions include the following:

- **Privacy violation:** This was the top concern for 26% of the respondents. For example, respondents found the use of personal finance scores (such as a FICO or credit score) unacceptable.
- **Lack of fairness:** There was concern regarding the fairness of decisions processes. For example, there was concern regarding the fairness of automated screening of job applications.
- **Lack of human-in-the-loop for important decisions:** This is the top concern of those who find the automated resume screening concept unacceptable (36%), and it is a prominent concern among those who are worried about the use of video job interview analysis (16%).
- **Inability of machines to capture nuances in human decision-making:** Humans are complex, and AI-enabled systems are incapable of capturing nuances. For example, 26% of these respondents argue that every individual or circumstance is different and that a computer program would have a hard time capturing human nuances. Roughly half of these respondents mention concerns related to the fact that all individuals are different, or that a system such as this leaves no room for personal growth or development (Smith, 2018). Please also see the Stakeholder Perception Section for more information on this challenge.

Given concerns regarding ethics and equity of AI-enabled systems and the complexity in assessing these aims, some agencies are working to define what it means for AI to be ethical. For example, federal defense agencies are beginning to define AI ethical principles. The U.S. Department of Defense (DOD) established AI ethical principles that encompass five major areas, and the U.S. Intelligence Community (IC) has committed to designing and developing AI based on three core principles (Defense, 2022; Office of the Director of National Intelligence (ODNI), 2020). The two agencies' core areas of focus overlap in some respects and focus on creating ethical and equitable AI with the following traits:

- **Responsible:** Exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.
- **Objective and equitable:** AI will provide objective intelligence while taking deliberate steps to identify and mitigate unintended biases in AI.
- **Traceable, transparent, and accountable:** Develop and deploy AI systems such that personnel possess an understanding of the technology, deployment process, and operational methods. This includes transparent and auditable methodologies, data sources, design procedures and documentation and the ability to identify responsibility and hold developers and users accountable for the system's outcomes.
- **Human-centered development and use:** Tempering technological guidance with human judgement, especially when an action may infringe on an individual's civil rights and freedoms.
- **Reliable:** AI will have explicit, well-defined uses and the safety, security and effectiveness of AI will be subject to testing and assurance across its life cycle.

- **Governable:** Design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.

Implications for ITS

Some of the ways that challenges related to ethics and equity could impact AI-enabled ITS applications are summarized below.

- **Inequitable outcomes from competing objectives:** AI equity impacts for ITS potentially include impacts to modes and traffic flows (e.g., prioritizing transit versus vulnerable road users [VRUs] versus vehicles for an adaptive signal controller) to balance both throughput and various other needs.

- **Inequitable outcomes from biased data collection:** Additionally, ethics and equity factor into ITS systems in terms of the data that is being fed into an AI system. For example, if an AI system for emergency response relies most heavily on connected vehicle data, then non-connected vehicles could be overlooked. Additionally, AI-enabled traffic signal applications pose ethics concerns if they are collecting data from a subset of the population that is not representative to build AI prediction algorithms. For smart intersection crossing, potential issues arise around omitting pedestrians that are visually impaired or that have cognitive or physical disabilities and who may take longer to cross a given street. If data do not account for all VRUs, then the prediction algorithm could end up discriminating against them. Using AI for workforce selection and scheduling requires equitable practices to ensure that the data do not contain underlying prejudices and can distribute workloads equitably (Vasudevan, Townsend, Schweikert, Wunderlich, et al., 2020).

- **Discrimination in infrastructure and asset management decisions:** Regarding infrastructure, discrimination could occur if AI is used in making investment decisions regarding where to install equipment/devices for collecting data, for sharing information (e.g., via variable message signs), lighting, or where to build a road or expand the number of lanes. Infrastructure changes, such as equipment maintenance and roadway construction, may be unequal when looking at underdeveloped, underserved, or physically divided communities. Road-related discrimination that can be dated back to the mid-1900s is one example of this inequality of service, namely, the construction of expressways that divided and displaced minority communities in New York City (Ploschnitzki, 2017). These communities may also be deprived of services due to lack or ease of accessibility compared to other parts of the area (Bornstein, 2017). Another concern is asset management—discrimination could occur in prioritization of decision-making for maintenance and repair of pavements, bridges, tunnels, and other traffic control devices; underserved populations and rural communities may suffer.

- **Ethics concerns in AV driving decisions:** Regarding AI and automated vehicles (AVs), potential ethics and equity issues arise around predicting behaviors of other vehicles, drivers, and VRUs in the surrounding environment. AI-enabled AVs may be forced to decide who is "more dispensable" when a crash is imminent. Cybersecurity is also a large concern for AI in automated vehicles and ITS. An attack on AI could result in loss of human life, loss or misuse of personal data, and overall degradation of the transportation system and communications network.

- **Discrimination in language processing:** AI ethics issues could arise in language processing applications. For example, discrimination could occur if AI is not trained on different accents or cultural differences, domestically as well as internationally.

- **Unintentional consequences from development choices:** Choices made during AI model development can greatly impact results. An example is given of trying to predict urban heat islands and choosing the spatial resolution of the model. Too low a resolution may overlook extreme values in small neighborhoods, but too high a resolution may introduce noise (Board on Atmospheric Sciences and Climate , 2022). Sensors used for AI data input often do not cover all populations equally, and many of them require daylight (Board on Atmospheric Sciences and Climate , 2022). This could also apply in the case of sensor-based data collection for AI-enabled ITS applications.

Insights and Lessons Learned

Some strategies for considering ethics and equity in AI development and deployment are summarized below.

- **Creating AI systems with ethics, equity, and transparency at the forefront:** AI systems may be developed to minimize bias and act without prejudice. This goal starts with ensuring the data that a system is built on does not contain implicit or explicit biases that could lead to model bias. During development, the model may be tested consistently to detect and remove bias during model building and include input from different communities who may not always have a seat at the table. Prior to implementation, new AI systems may be tested and demonstrated in a controlled environment to gain trust, demonstrate use, and define the system’s purpose and boundaries. To ensure ongoing ethical use and user and public trust after an AI system is implemented, transparency is key. Additionally, humans can consistently verify the system to ensure performance, explainability, and other outcomes are maintained while minimizing model drift (Vasudevan, Townsend, et al., 2022).

- **Translating ethical frameworks into engineering:** Ethical and equitable AI systems are both an engineering problem and a policy concern. From the initial design, ethical and equitable AI systems center on a diverse group of equal stakeholders to avoid power centralization and data practices that may exploit vulnerable populations (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022). Established standards (e.g., IEEE P7000/P7010 "Wellbeing Metrics Standard for Ethical AI and Autonomous Systems") can be used from a policy perspective to meet ethics goals while keeping a human-in-the-loop to monitor the performance of AI decision making (*AI World Government In-Person & Virtual | October 18-19, 2021*, n.d.). It is helpful for engineers to understand ethics and equity when building AI models to retain valuable context that can be used to make ethical decisions (NIST, 2022). Losing this context while developing a model can create abstraction and governance issues. In addition, engineers are likely responsible for ensuring that the most high-risk applications are paired with low-risk technology to ensure maximum success in implementation and equity (Dunnmon et al., 2021).

- **Supporting workforce training and education to meet future AI needs:** Supporting education to advance technical expertise and to retrain individuals in the workforce could help those whose jobs are disrupted by AI adoption. Designing strategies that prepare the labor force for the transition, create jobs that take advantage of both AI and human skills, and supplement the workforce with educational programs that increase technological literacy could help in facilitating the relationship between humans and AI (IEEE Advancing Technology for Humanity, 2019). Additionally, encouraging credentials for AI creators and operators could help to ensure they can demonstrate an appropriate level of knowledge in an AI system (IEEE Advancing Technology for Humanity, 2019).

- **Including diverse stakeholders throughout AI development:** Involving multidisciplinary stakeholders, both internal and external to the entity, throughout an AI system's development could help ensure that societal concerns are considered. In September 2020, the Comptroller General of the United States (CG) convened a forum of experts in industry, government, nonprofits, and academia to discuss factors affecting oversight of AI, including governance, sources of evidence, methods to assess implementation of AI systems, and identifying and mitigating potential bias and inequities. Forum participants discussed how to operationalize recent principles and frameworks on the use of AI into practices for managers and supervisors of these systems, as well as mitigation strategies to address challenges in implementing AI in the public sector. A participant in the forum stated: "We [built] an AI ethics board that has representatives from all divisions [who] can add ... decision power" (GAO, 2021).

- **Promoting ethical, trustworthy AI use and development:** Inspiring trust and confidence in AI is important for its successful adoption. For example, the U.S.

Department of Health and Human Services (HHS) published its *AI Strategy* in 2021 to outline its approach to promote trustworthy AI (U.S. Department of HHS, 2021), aligned with federal directives (i.e., Executive Order 13960 and Executive Order 13859). According to their AI strategy, HHS aims to support divisions in deploying reliable, explainable, non-biased, and secure AI systems that respect citizens' privacy and data security. HHS will also support divisions in developing policies that ensure transparency and accountability in AI use by communicating department-specific principles for effective, equitable, safe, secure, and ethical AI and data used to create and operate AI. HHS will also promote and support the application of existing cybersecurity frameworks to AI use cases and promote the evaluation of applied AI for accuracy, effectiveness, and equity.

- **Applying guidelines to promote responsible AI:** Applying guidelines for what it means for AI to be responsible to projects and programs could help to ensure AI starts and continues to be responsible, ethical, and equitable. For example, the DoD's Defense Innovation Unit (DIU) has *Responsible AI Guidelines* to operationalize the DoD's Ethical Principles for AI (Dunnmon et al., 2021). From applying these guidelines, the DIU has identified a number of lessons learned for each phase of the AI development lifecycle. A few lessons during the planning phase include defining the task and success metric and prescribing processes to safely address system errors and revert malfunctioning systems back to a previously functioning version. A few lessons during the development phase include assigning the authority to make changes to the capability to a specific, accountable party and designing the system interface to give users the ability to understand how outputs are produced. Finally, a few lessons during the deployment phase include conducting continuous task and data validation and confirming that new data do not degrade system performance.

- **Documenting processes, success metrics, and expectations:** AI capabilities cannot be used confidently without comprehensive, even-handed documentation (Dunnmon et al., 2021). In addition to documentation, defining success through key performance metrics is important to benchmark ethics and equity in AI systems (Dunnmon et al., 2021). AI customers and vendors could set and document expectations, since AI is not magic and developers should not act like magicians (Dunnmon et al., 2021). Additionally, different stakeholder groups involved in the AI project may have different working assumptions. Documenting these assumptions as well as expectations may help to ensure everyone is on the same page. Agencies may want to consider investing time and resources in documentation and best practices.

- **Including a human-in-the-loop for critical decisions:** Bringing together AI expertise and human intelligence can mitigate public concerns regarding fairness and acceptability of AI-enabled ITS applications for critical decisions. The extent of human involvement in AI decision making depends on the complexity and exigency of automation and of the industry. For example, certain industries like healthcare, will always require human domain expertise especially for medical diagnoses. It is a good practice to complement AI decision making with human intelligence.

2.7 Generalization

Table 9. Summary of the Generalization Challenge and Potential Strategies to Address It

Summary of Generalization	
What is it?	When a trained ML model does not adapt well to unseen data, it may have underfit or overfit its training data, which could lead to poor performance
Why does it matter for ITS?	• Vendors may promote their AI solutions as being able to work anywhere, but AI solutions are not necessarily designed to work everywhere. • Since non-recurring conditions are far less common than recurring conditions, they present a challenge in terms of having enough data to train an ML model to detect and classify them correctly. • ML models require large quantities of representative data to generalize well, but real-world data can be expensive to acquire, and simulated data may not be fully representative.
How can it be addressed?	• Having representative data for training • Making the training data more robust • Handling edge cases • Limiting overfitting • Combining ML techniques or models • Using model testbeds • Re-training for new locations • Considering transfer learning • Developing and using standards to enable interoperability

Description of Challenge

Generalization refers to an ML model's ability to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model (*Generalization | Machine Learning Crash Course*, n.d.). One of the main goals of developing an ML model is for it to generalize well to unseen data. Therefore, it aims to capture useful trends (i.e., not underfit) while ignoring meaningless, random fluctuations in the data (i.e., not overfit) (Chouldechova, 2017). However, it can be difficult to strike the right balance during development. Additionally, even if a trained model performs well against test data, it could still unexpectedly encounter completely new data after it is deployed. For example, if the trained model encounters new traffic patterns or conditions (e.g., a new work zone or lane closure; new traffic patterns from a major sporting event, etc.) that are not representative of data that were used to develop the model, this could lead to poor performance. Overall, it can be challenging to transfer pre-trained models to new situations with different data that might fall outside the distribution used for development.

Implications for ITS

Some of the ways in which AI/ML generalization could come into play for ITS are summarized below.

- **Scalability and transferability of vendor AI solutions:** Agencies may have to deal with scalability and transferability issues if investing in certain vendor products that are not necessarily designed to work everywhere (Vasudevan, Townsend, Schweikert, 2020). For example, a decision maker at a state DOT may want to exercise caution prior to buying an AI solution marketed by a vendor if the AI solution was developed for a small area within a dense urban arterial network. It could have challenges generalizing well to different situations. In general, a trained ML model will perform best on the data it was trained on. If new, real-world data do not follow a similar pattern as the training data, then the model will not perform as well, and therefore, will have less utility for the agency.

- **Inadequate data for non-recurring conditions:** It can be difficult to have adequate data on non-recurring traffic conditions (e.g., crashes) for training an ML model. Even if some data are available, since crashes are rare events amidst regular traffic flows, the data are likely to be highly imbalanced (i.e., there are far fewer instances of non-recurring event records in the data set when compared to recurring records). Additionally, even if an ML model has seen many different incidents during development, each new incident is likely to look a little different and could have different impacts on non-recurrent traffic patterns. This could make it difficult for a trained model focused on identifying and classifying non-recurring patterns to generalize.

- **Representative simulated data:** ML models need large quantities of data to learn from, but real-world data can be expensive to collect or acquire. Simulated data, on the other hand, may offer a cheaper and easier alternative to work with. However, while simulated data tend to be relatively clean and error-free, real-world data may not always behave so nicely. Therefore, if the simulated data do not fully represent possible real-world conditions, then an ML model trained on them may have challenges generalizing. Agencies may want to use caution when procuring vendor solutions that have been built primarily using simulated data.

Insights and Lessons Learned

Researchers and deployers have used different strategies to improve model generalizability, some of which are summarized below.

- **Having representative data for training:** If fewer training examples are provided than parameters that exist in the ML algorithm, then generalization (and therefore performance) may be an issue, since in this case, the algorithm typically memorizes the training data rather than learns useful trends (Zhang & van der Baan, 2021). Generalization is improved if the provided samples are representative, in that they describe all features of interest well. In addition to being representative, it is helpful for the data to be complete to prevent poor performance, meaning the examples span the full solution space (Zhang & van der Baan, 2021). To frame it another way, since an ML model is trained within the distribution of its training data, it cannot extrapolate beyond the bounds of those training data labels. For example, if an ML model is trained to classify images of cows, goats, and horses, then it will not be able to correctly classify an image of a bird. Ensuring completeness during training is challenging unless the target application is well understood (Zhang & van der Baan, 2021). Therefore, understanding the context in which the ML model will operate is key for ensuring the training data are representative of all possible patterns and outcomes.

- **Making the training data more robust:** In some cases, the availability of robust training data, in terms of both quantity and diversity, may be limited. The size of the data is a "deciding factor" for generalization performance (Xu & Goodacre, 2018). There are many strategies to increase the size and diversity of training data. For example, researchers tried a variety of up/over-sampling and data augmentation techniques, including Random Over Sampling, Synthetic Minority Oversampling Technique (SMOTE), and Adaptive Synthetic Sampling (ADASYN), to boost the robustness of the training data for classifying queue lengths at intersections, which was particularly helpful for the minority classes of queues with 15 or more vehicles (Vasudevan, O'Hara, Townsend, et al., 2022). However, to have a reliable estimation of model performance, there is a balance to be struck between having too few and too many samples in the training data and testing data (Xu & Goodacre, 2018). Additionally, if sampled or augmented too heavily, the training data could drift away from the actual data distribution, which is hopefully reflected in the held-out testing data. At a certain point, this could reduce model performance and generalizability.

- **Handling edge cases:** Even if the training data are robust and representative, the AI model may still encounter edge cases, such as extreme, unexpected, or adversarial data inputs. Part of the challenge with these cases is that they can lead AI to make "silent" errors that are difficult for developers to catch. While inputs that are obviously faulty (e.g., of the wrong data type) can be detected and flagged readily, more challenging model failures could happen because data preprocessing lacks context about what inputs are reasonable or expected (Harris, 2022). For example, a computer vision application at an intersection might miss a construction worker emerging out of a manhole that has not been cordoned off or marked as a construction zone, leading to fatalities. A stress testing-driven approach is one way to help manage these potential edge cases and reduce AI risk (Singer, 2022). By performing model stress testing during development, developers can observe model behavior and identify potential issues. In addition to designing tests, developing anomaly detection models that catch and flag outlier input values could help in dealing with potential silent model failures on edge cases (Harris, 2022).

- **Limiting overfitting:** Model overfitting can lead to issues in generalization. An overfit model refers to a model with low bias and high variance, meaning the model is too closely aligned to its training data points. Overfitting can lead to low performance on new, unseen data in the operational environment (GAO, 2021). There are many techniques to help avoid overfitting, such as regularization (e.g., early stopping), feature selection, and using cross-validation to identify models with low test error (Chouldechova, 2017).

- **Combining ML techniques or models:** When working with non-recurrent traffic conditions (e.g., incidents), it may help to leverage multiple learning techniques using data from multiple years (to ensure there is sufficient data) in the same well-defined, small area to make predictions (Qian, 2021). For example, unsupervised learning can be used to group/cluster non-recurrent traffic patterns from a wide variety of incidents so that within each group/cluster, traffic is similar (e.g., vehicle with flat tire on side of highway could lead to traffic shifting lanes and slowing down). Then, supervised learning could be used to predict traffic in advance within each group/cluster. Researchers at Carnegie Mellon University demonstrated an AI system that recommends contingency signal plans to accommodate non-recurrent traffic in Cranberry Township, PA by combining learning techniques (Yao & Qian, 2020). Specifically, they used two models for the plan recommendation task: real-time traffic prediction, which predicts future traffic flows up to 30 minutes in advance in 5 minute increments, and plan association, which selects and combines decision-making rules to recommend signal timing plans based on current and predicted traffic conditions (Yao & Qian, 2020).

- **Using model testbeds:** According to the Networking and Information Technology Research and Development (NITRD) Program, a formal federal program with 25 member agencies, testbeds provide environments to support development of real-world applications of AI that are robust and trustworthy, including enabling reproducibility testing (NTRD, 2022). Developers could use model testbeds to test the reliability and robustness of AI systems under different real-world conditions without the same potential real-world consequences (e.g., safety concerns). The NITRD Program has an "AI R&D Testbed Inventory" page that allows users to locate federally-supported testbed and testing resources that can support AI research (NTRD, 2021).

- **Re-training for new locations:** In some cases, model generalization to new contexts may be infeasible or undesirable given tradeoffs with performance. Researchers found that training localized models (e.g., for the specific roadway network) may be necessary to achieve adequate performance (Vasudevan, O'Hara, Townsend, et al., 2022). They found that models perform best when trained and tuned to the specific flow and characteristics of the network. For example, an ML model trained to detect an incident on a freeway corridor may not be able to successfully detect an incident on an arterial network without re-training the model or even considering additional features to achieve sufficient accuracy. They suggest that even if the two network types are similar (e.g., suburban freeways in two cities), that the models be re-trained to the local data for improved performance.

- **Considering transfer learning:** One promising approach to help address challenges with generalization is transfer learning, in which a model that is trained to accomplish a certain task applies that learning to a similar but different task (Vasudevan, Townsend, Schweikert, 2020). Using a general-purpose pre-trained ML model as the foundation upon which to build a more specialized ML model could help address generalization concerns since these models are often trained with very large quantities of diverse data. For example, researchers in collaboration with the North Carolina Department of Transportation (NCDOT) utilized transfer learning by adopting the Xception neural network architecture, which was developed by Google (Chollet, 2017), as the "feature extraction backbone" for their further customized roadside feature (e.g., guardrails, utility poles) detection solution (Yi et al., 2021). In a second example, to approximate pedestrian social distancing at intersections and bus stations, researchers in New York City used a pre-trained convolutional neural network to detect objects, filtered and focused it for their urban arterial context to classify objects in the camera view (e.g., pedestrian, bus, vehicle, bike), and then added a post processing filter for the distance calculations (Zuo et al., 2021).

- **Developing and using standards to enable interoperability:** Since the availability of representative data is a key driver of generalizability, data standards could help enable application interoperability, and therefore, generalization and transferability across vendors, locations, and contexts. According to respondents of the USDOT's AI for ITS Sources Sought Notice (*SAM.Gov*, 2021), standards are not sufficiently mature to allow for interoperability across locations or vendors. Additionally, respondents emphasized that common standards on data formats (both inputs and outputs) are needed; at present, significant calibration is required to cater an application to different sites (Vasudevan, Townsend, et al., 2022).

2.8 Model Drift

Table 10. Summary of the Model Drift Challenge and Potential Strategies to Address It

Summary of Model Drift	
What is it?	When a trained model's input data, output data, or relationship between the two changes over time leading to system performance degradation
Why does it matter for ITS?	• Model drift could lead to AI system performance degradation, which in turn, could reduce ITS performance and user trust. • Sensor malfunctions or hardware/software updates could lead to incorrect predictions if an AI system has not been trained on these occurrences. • If the input data in an operational setting starts to drift away from the data used to train the model, the performance of the AI system might start to degrade. For example, the introduction of a new ridesharing service that was not captured in the training data could lead to an AI-enabled traveler information system no longer offering the most relevant options to travelers. • Policy changes that impact the target variable could lead to concept drift. For example, if a freeway agency changes one of its lanes from an HOV-2 to an HOV-3, an AI-based vehicle occupancy detection and tolling enforcement application would need to be updated to learn this change or else it could incorrectly enforce toll rates.
How can it be addressed?	• Having a plan in place for model drift assessment and mitigation • Establishing appropriate ranges of data and model drift • Regularly monitoring and improving the system • Retraining the model • Considering online learning

Description of Challenge

Model drift (also referred to as model "decay") refers to the changes in the relationship between the data inputs and the prediction outputs (GAO, 2021), which could be driven by data and/or concept drift.

- **Data drift:** Data drift refers to the changes in the statistical properties of the input data in an operational environment, as compared to the training data (GAO, 2021). Data drift often refers to changes in the predictors or independent variables specifically. This drift could be driven by upstream process changes in data collection, data quality issues (e.g., malfunctioning sensors), or natural drift in the data (e.g., seasonal changes).

- **Concept drift:** Concept drift refers to the changes in the statistical properties of the output or target variable in an operational environment, as compared to the training environment (Shendre, 2020). For example, when classifying incidents that require coordinated human-in-the-loop action across multiple agencies, the definition of when coordinated action is needed could change over time.

Model drift, whether driven by data drift, concept drift, or some combination of the two, could lead to AI system performance degradation. If an AI system is not regularly monitored, model drift could occur undetected and lead to undesirable consequences.

Implications for ITS

Model drift could bring negative consequences to an ITS system. It could lead to incorrect predictions and AI system performance degradation. This performance degradation, in turn, could potentially lead to reduced trust from system users if the drift is not addressed quickly and appropriately.

A variety of sensor, data, and policy changes could potentially lead to model drift, and therefore, incorrect predictions in ITS. A few possibilities and their consequences are summarized below.

- **Incorrect predictions from sensor malfunctions or hardware/software updates:** A model's input data may drift as a result of sensor device break down or software updates that impact how measurements are recorded (Oleszak, 2021). This is also referred to as "covariate shift" since the distribution of the model inputs changes. For example, if a CCTV camera feed loses video from a malfunction or obstruction but continues to send footage to a computer vision system for queue length detection, this could lead to erroneous queue length predictions. The computer vision system may fail to detect any queues simply because it is failing to see the vehicles. Model drift may also occur because of sensor hardware upgrades, changes to a system, or use of hardware from different vendors. For example, if the model was trained on CCTV feeds from one vendor, incorrect predictions may be drawn if CCTVs feeds from a different vendor are used during the deployment phase due to potential shift in data.

- **Incorrect predictions from input data drift:** If the distribution or specific attributes of the input data to an AI system drift from the original training data, this could lead to model drift and performance issues. For example, if an agency decides to equip an additional 15% of its bus fleet with connected vehicle devices and send connected vehicle messages (e.g., Basic Safety Messages) from them to its AI-supported traffic prediction system, the system might incorrectly predict an increase in traffic overall if not properly calibrated to account for the increased connectivity (i.e., market penetration rate of connected vehicles). In another example, if there is a shift in travelers' preferred mode choice from the introduction of a new ridesharing service, then an AI-enabled traveler information system may no longer offer the most relevant options if it does not appropriately account for the new ridesharing mode.

- **Incorrect predictions from policy changes:** Policy changes that impact the outcome variable could lead to concept drift and incorrect predictions. For example, if a freeway management agency decides to change one of its High Occupancy Vehicle (HOV) lanes from an HOV-2 to an HOV-3 and it uses a computer vision system to classify vehicle occupancy, this could lead to erroneous tolling enforcement since what constitutes an HOV violation has drifted or changed (i.e., the meaning of the target variable has changed). The application would need to be updated to learn this change or else it could incorrectly enforce toll rates.

Insights and Lessons Learned

Some potential strategies and lessons learned to identify, mitigate, and address model drift are summarized below.

- **Having a plan in place for model drift assessment and mitigation:** The Department of Defense, Defense Innovation Unit's *Responsible AI Guidelines in Practice* emphasize the criticality of having a plan in place at the outset of the project for what to do when model drift or other issues occur. For example, if an automated task begins to fail or perform below a pre-determined threshold, then it may be necessary to revert to a manual process by trained personnel in the interim (Defense Innovation Unit (DIU) et al., 2021).

- **Establishing appropriate ranges of data and model drift:** The GAO's *AI Accountability Framework* suggests that entities establish the range of data and model drift that is acceptable to ensure the AI system produces desired results (GAO, 2021). The framework suggests that the ranges be established based on the nature, scope, and purpose of the components and the risks they pose.

- **Regularly monitoring and improving the system:** Once an application is deployed in ITS, it is considered a best practice to regularly monitor it, evaluate its performance, and make improvements. For example, the FHWA *Systems Engineering for ITS* "V" Diagram, which is adapted to the broader ITS project life cycle, includes "Operations and Maintenance" as well as "Changes and Upgrades" (National ITS Architecture Team, 2007). Additionally, Chapter 18 in the FHWA *Freeway Management and Operations Handbook* focuses on effective performance measurement, monitoring, and evaluation that includes regular monitoring of the data and system (Hatcher et al., 2017). Similar to other ITS and freeway projects, regular monitoring and improvement are necessary for AI applications deployed in ITS. Having a robust set of regularly scheduled tests can help with monitoring model drift and is considered a best practice by the DIU *Responsible AI Guidelines in Practice*. These tests can help inform whether drift has occurred, inform if the AI capability is performing sub-optimally, and help with diagnosing the problem (Defense Innovation Unit (DIU) et al., 2021). Upgrades may need to be made more frequently to an AI system whose data, scale, and/or role is expected to grow or change over time. Rather than launching a fully trained AI system immediately, the Tennessee DOT (TDOT) is taking an additive approach to their AI system development and deployment (L. Smith et al., interview, April 2022). To start, the AI system, which will suggest strategies for TMC operators to implement amidst incidents, heavy congestion, and other events, will use data from microsimulation modeling. Then, it will use real world data and results from operator feedback to try and improve its performance, while adhering to safety critical rules. TDOT plans to have their system continuously learn in-the-loop based on the initially deployed rule-based system and feedback on whether TMC operators accept or reject the AI system's recommendations (L. Smith et al., interview, April 2022).

- **Retraining the model:** Entities may need to retrain the components of the AI system if the data or model drift for each component exceeds the established acceptable range (GAO, 2021). It is considered good practice to regularly monitor the incoming data and retrain the model on newer data if the data distribution has deviated significantly from the original training data distribution (Amazon Web Services, Inc., n.d.). If there are high overhead costs associated with monitoring the data to detect a change in the distribution, then a simpler strategy may be to retrain the model periodically (e.g., weekly or monthly) (Amazon Web Services, Inc., n.d.).

- **Considering online learning:** In contrast to training a model on a complete set of data (i.e., batch learning), online learning is a form of machine learning for data arriving in a sequential order, where a learner aims to update the best predictor for future data at every step (Hoi et al., 2018). In the online learning case, the predictive model can be updated instantly for any new data instances. This could be an especially useful approach in cases with large data that arrive quickly (e.g., continuous streams of new data) (Hoi et al., 2018). While online learning may help to prevent against model drift, it may also bring new risks since the model is constantly changing. For example, neural networks have a tendency to forget what they learned in the past in an online learning setting since the weights are updated after each sample is received and then the sample is discarded (Lo & Ghiassian, 2019).

2.9 Explainability

Table 11. Summary of the Explainability Challenge and Potential Strategies to Address It

Summary of Explainability	
What is it?	Inability of an AI system's process and decisions to be understood by system operators or end users
Why does it matter for ITS?	• Explainability is especially important for safety-critical and other high-stakes decisions with greater risk and liability concerns for the agency. • The level of explanation required for an AI-enabled decision support system depends on the task at hand and level of supervision from the ITS decision maker. • If a vendor's AI solution is not transparent and explainable, this could reduce agency and user trust in the overall procured system. • Even simple explanations for AV decisions could improve driver and pedestrian interaction with and trust of the AV.
How can it be addressed?	• Understanding potential tradeoffs between interpretability and performance • Balancing explainability with security and privacy • Improving transparency through documentation • Using interpretable models • Engineering interpretable features • Outputting multiple performance metrics • Visualizing results • Exploring post-hoc explainable AI (XAI) methods • Using explainable AI (XAI) analysis for validation of model strategies and to improve trust in AI outcomes • Considering non-AI alternatives

Description of Challenge

Many ML models are complex and can be difficult to interpret (i.e., are often labeled as "black box" in nature) since their process for reaching a decision is not straightforwardly interpretable by system operators or end users, making it hard to understand *how* a decision was made. Reducing risk from AI relies in large part on designing algorithms that are "explainable" (Chenok, 2020). Explainability (also referred to as "interpretability") is a key component of trustworthy AI and is focused on answering the question: Can the system's outcome be justified with an explanation that a human can understand and/or that is meaningful to the end user (Wing, 2022)? If the answer is no, then this inscrutability can hamper users' trust in the system, especially in contexts with significant consequences, and therefore lead to rejection of the system overall (Rai, 2020). Additionally, this opaqueness can make it difficult to discover underlying algorithmic biases.

The National Institute of Standards and Technology (NIST) proposes four principles for explainable AI, which are centered around the humans that consume the explanations (Phillips et al., 2021):

- **Explanation:** A system delivers or contains accompanying evidence or reason(s) for outputs and/or process.
- **Meaningful:** A system provides explanations that are understandable to the intended consumer(s).
- **Explanation accuracy:** An explanation correctly reflects the reason for generating the output and/or accurately reflects the system's process.
- **Knowledge limits:** A system only operates under conditions for which it was designed and when it reaches sufficient confidence in its output.

Implications for ITS

Some of the ways in which AI/ML explainability could come into play for ITS are summarized below.

- **Safety-critical and other high-stakes decisions:** Explainability may be especially important for safety-critical decisions (e.g., wrong-way driving and pedestrian detection and response; emergency management) and other decisions in ITS with potentially severe consequences and/or liability concerns. For high stakes decisions, one might want to avoid a black-box model, unless it can be proven that an interpretable model does not exist with the same level of accuracy (Rudin, 2019). The Defense Innovation Unit's *Responsible AI Guidelines in Practice* emphasizes that "AI systems cannot be responsible for outcomes – humans must always bear responsibility," and humans, not machines, should make decisions that affect a person's quality of life (Defense Innovation Unit (DIU) et al., 2021).
- **Decision Support Systems:** As part of the Defense Advanced Research Projects Agency (DARPA) Program on Explainable AI (XAI), researchers conducted user studies and found that "different user types require different types of explanations" and "user cognitive load to interpret explanations can hinder user [task] performance" (Gunning et al., 2021). These insights could apply in the case of AI-enabled decision support systems in ITS. For example, AI could suggest Transportation Systems Management and Operations (TSMO) strategies during the planning stage, but a Traffic Management Center (TMC) operator or

decision-maker would still play a crucial role in vetting the AI-generated recommendations (Vasudevan, Townsend, Dang, et al., 2020). The Tennessee DOT (TDOT), as part of their ATCMTD deployment, is preparing to deploy an AI system that suggests strategies to TMC operators amidst incidents or heavy congestion. However, in speaking to the AI for ITS Program (L. Smith et al., interview, April 2022), one researcher supporting the project emphasized the difficulty in training human operators to use an AI system when that system cannot explain its decisions. Therefore, they are taking an incremental development and deployment approach to ensure operator awareness and understanding of the system, rather than doing a wholesale swap of the existing rule-based system with a fully trained AI system. There is likely a balance to be struck between too few and too many details, particularly when providing outputs (i.e., decisions and explanations) to decision makers. The type of explanation (and its level of detail) deemed appropriate for the situation will depend on the requirements of the given situation, the task at hand, the consumer, and the decision maker(s) (Phillips et al., 2021). If an explanation of the AI system's decision is not possible or is too difficult to make sense of, then the overall task could be hindered.

- **Vendor transparency:** As one speaker at AI World Government 2021 put it "AI is not magic, and vendors should not act like magicians. They must reveal their tricks" (*AI World Government In-Person & Virtual | October 18-19, 2021*). Transportation agencies might want to keep this in mind when looking to procure AI solutions from vendors. Agencies might want to understand how vendor systems are operating and how decisions are made. If the data, methods, and processes are not disclosed by the vendor and decisions are not explainable, then the agencies might find it difficult to explain the reasoning behind the decisions to decision makers and stakeholders, leading to mistrust in the AI-suggested decisions and system. In speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (T. Geara et al., interview, April 2022), the City of Detroit shared that they expect their computer vision vendor to validate its model's performance. However, the city makes sure the vendor conducts sufficient testing and validation. Additionally, the city asked for an explanation in cases where the model was behaving strangely or performing poorly. For example, in one case the computer vision system for traffic counts was showing a high error rate because it was counting cars on truck carriers as additional traffic near a facility that assembles vehicles.

- **Automated Vehicles:** According to NIST, the "Knowledge Limits" principle of explainable AI states that systems identify cases in which they were not designed or approved to operate, or in cases for which their answers are not reliable. This principle can increase trust in a system by preventing misleading, dangerous, or unjust output (Phillips et al., 2021). For example, according to SAE standard J3016 "Levels of Driving Automation" (*SAE Levels of Driving Automation^{TM} Refined for Clarity and International Audience*, n.d.), a Level 3 automated vehicle

(AV) uses AI to make dynamic driving decisions within its operational design domain unless it encounters a situation in which it requests the human to take control (e.g., system uncertainty or failure). Notifying the driver of *why* it is necessary to take control of the vehicle could aid in explainability. For example, if the Level 3 AV encounters a floating plastic bag on the road, it could audibly and/or visibly alert the driver with “unknown object ahead, please take control.” Not only could additional information be helpful to drivers and passengers, but it could also be helpful for pedestrians. For example, researchers in Germany compared the effectiveness of various vehicle-pedestrian communication implementations in AVs and found that AVs issuing a “high-content” message of “I’m stopping, you can cross” were rated as being the most trustworthy and reassuring by a focus group of pedestrians with vision impairments (Colley et al., 2020).

Insights and Lessons Learned

Explainability is a complex topic in AI that continues to garner increased attention. Some of the strategies and lessons learned to support explainability are summarized below.

- **Understanding potential tradeoffs between interpretability and performance:** There is generally a tradeoff between model interpretability and flexibility. While highly structured models, such as linear and logistic regressions and decision trees, tend to be easy to interpret, highly flexible (i.e., more complex) models, such as random forests and deep neural networks, tend to do a better job at prediction (Chouldechova, 2017). Depending on the use case, a more interpretable/explainable model may be preferred over a higher performing “black box” model. Sensitivity analysis might help decision makers choose their preferred approach. For example, a logistic regression model might only show marginally lower performance (e.g., accuracy) compared to a deep neural network to predict binary transportation mode choice (i.e., transit or not) but might offer a far simpler model and greater interpretability regarding *how* the model is categorizing the two. If the model is not too complex, then not only could it make predictions of mode choice, but also be useful in understanding which features are important for predicting mode choice (e.g., distance to nearest station, age, etc.).

- **Balancing explainability with security and privacy:** While explainable AI could bring new opportunities and potential benefits, it also introduces new threats to a system (Phillips et al., 2021). A potential negative consequence of having an explanation along with the AI output is the exposure of model details (Milli et al., 2018).

- **Improving transparency through documentation:** A simple and accessible approach to increasing transparency in ML lifecycles is through an improvement in both internal and external documentation. Documenting key decisions across the AI lifecycle can improve explainability and help users, auditors, and stakeholders understand the AI system (GAO, 2021). According to ABOUT ML (Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles) led by the Partnership on AI, this documentation process begins in the ML system design and set up stage, including system framing and high-level objective design (*ABOUT ML*, n.d.).

- **Considering system auditability:** Developing and deploying AI systems such that personnel possess an understanding of the technology, deployment process, and operational methods could help support explainability. This includes transparent and auditable methodologies, data sources, design procedures, and documentation (Defense, 2022). Additionally, planning for regular system auditing is one of the DIU's *Responsible AI Guidelines* for the development phase. Models can be audited in multiple ways, ranging from internal code and training process reviews to fuzzing (i.e., a form of negative testing to surface vulnerabilities) and deterministic testing (Defense Innovation Unit (DIU) et al., 2021). Overall, an AI system that is auditable by different entities, whether the government or a third party, is more likely to be transparent and explainable. Appendix V of the GAO's *AI Accountability Framework* describes key auditing standards that could be considered by auditors and audited entities to ensure accountability (GAO, 2021).

- **Using interpretable models:** Some models are inherently interpretable. These include decision trees and regression models, among others. According to NIST researchers, not only can these "self-interpretable models" explain the entire model globally, but they can also explain individual decisions locally (Phillips et al., 2021).

- **Engineering interpretable features:** Some features are more understandable by system operators or end users than others. For example, bus occupancy, trip duration, and time of day are human-interpretable features while a sparse matrix of text token counts often used in natural language processing (NLP) is not. Having interpretable features can be helpful even when the model itself is not inherently interpretable. For example, while a random forest classifier's decisions can be difficult to interpret, if the model uses human-interpretable features then the "feature importances" could help with model explainability (*Feature Importances with a Forest of Trees*, n.d.).

- **Outputting multiple performance metrics:** Different performance metrics can share different insights into how the model is correctly, and perhaps even more importantly, incorrectly determining its classifications or predictions. Panelists on the "Technical Characteristics of AI System Trustworthiness" panel at the NIST AI Risk Management Framework Workshop emphasized the importance of understanding and interpreting multiple AI performance metrics (e.g., accuracy, false positive rate, false negative rate, precision, recall) (*Kicking off NIST AI Risk*

Management Framework, 2021). For example, accuracy may not be the most informative performance metric, particularly in cases with imbalanced data since the model could classify most, if not all, observations in the majority class without reducing overall model accuracy by much. Additionally, the false positive and false negative rates provide insight into a model's misclassifications. In some cases, the risks of a false negative may be higher than a false positive, and vice versa. Being transparent about these metrics could aid in AI explainability.

- **Visualizing results:** According to some AI for ITS SSN respondents, visual displays of AI results (e.g., dashboards, GIS maps) could help enable interpretation (Vasudevan, Townsend, et al., 2022). For example, when predicting bottlenecks in the system, it could be helpful to display the results as dynamic congestion heatmaps. Additionally, simply outputting confidence scores along with the ML classifications could provide some insight into how the model is categorizing objects in a computer vision application (e.g., bus, bike, pedestrian).

- **Exploring post-hoc explainable AI (XAI) methods:** A body of ongoing work currently seeks to develop and validate explainable AI methods (Phillips et al., 2021). A few post-hoc explanation methods mentioned by the NIST Explainable AI authors include LIME (Local Interpretable Model-Agnostic Explainer), SHAP (Shapley Additive ex-Planations), ICE (Individual Conditional Expectation), and Partial Dependence Plots (PDPs). While established *decision* accuracy metrics exist, researchers are in the process of developing performance metrics for *explanation* accuracy (Phillips et al., 2021).

- **Using explainable AI (XAI) analysis for validation of model strategies and to improve trust in AI outcomes:** Although still a nascent field, explainable AI analysis can help alert practitioners if their systems are using data in a way contradictory to the proposed use case. For example, XAI software can produce heat maps on top of the images of objects they were meant to detect and label, showing exactly what parts of the image were most important to the classification algorithm (Turri, 2022). If, in the example of highway crash images shared above, the XAI showed that emergency vehicles were a pertinent feature in classification, the practitioner would identify the undesired strategy and begin the process of remediation of the dataset. Being able to explain outcomes of AI algorithms also can help improve public trust by reducing the extent to which systems are "black boxes" – programs whose logic cannot be seen or interpreted by system operators or end users. For more information on this topic, please see the Stakeholder Perception Section.

- **Considering non-AI alternatives:** In some cases, AI may not be the optimal approach for a given task, particularly if the task requires complete transparency, interpretability, and explainability (Defense Innovation Unit (DIU) et al., 2021).

2.10 Liability

Table 12. Summary of the Liability Challenge and Potential Strategies to Address It

Summary of Liability	
What is it?	Lack of clear definition of who is liable when a vehicle, device, equipment, or system that uses AI is involved in a crash, is hacked, or produces erroneous results (e.g., misclassifies vehicle occupancy)
Why does it matter for ITS?	• If the AI application fails due to bias in the data, it is currently unclear whether the liable party for the failure is the application developer or the data provider • Liability is unclear when a vehicle, device, equipment, or system that is powered by an AI application is involved in a crash or results in fatalities. • Lack of clarity of safety expectations may regarding the damages that results from cybersecurity breaches in an AI product. • If an AI-enabled application has poor performance resulting in significant productivity losses, it is unclear who should be held accountable.
How can it be addressed?	• Partnering closely with agency risk management teams to consider legal and compliance issues from the perspective of organizational experts. • Assessing legal restrictions for the data to establish contracts and agreements in ways the data should be collected and used. • Assessing legal restrictions for the AI algorithm to establish contracts and agreements on all aspects of algorithm use and ownership. • Identifying possible risks throughout the AI pipeline, including considering downstream uses of AI system outputs. • Maintaining human accountability by assigning responsibility for AI system outcomes on specific individuals and organizations.

Description of Challenge

The ability for AI to make predictions and produce results that inform decisions creates preconditions for potential damage caused by its actions. Thus, issues arise with respect to liability and compensation that may or may not be covered under existing legal provisions. A key issue is that under current national law, AI is not recognized as a subject of law. This means that any potential damages caused by an AI system cannot be held personally liable (Čerka et al., 2015). Because of this lack of clarity, the question of who is responsible for damaging outcomes of AI systems remains an open one. (Čerka et al., 2015).

The Federal Trade Commission (FTC) offers some discussion in how AI algorithms might be treated by US law (Smith, 2020). Firstly, consumer transparency is key to liability protection. The FTC warn if used to mislead consumers, AI that spoofs human interaction such as AI chatbots or deepfakes may be liable to face FTC enforcement action. Similarly, lack of transparency with consumers when collecting sensitive data can prompt FTC action. Secondly, decisions that are explainable to the consumer are another factor. If using algorithms to assign risk scores or allocate resources to consumers or communities, the AI system owner ought to be able to disclose key factors that affect that score. Finally, the fairness of decisions made by the algorithm is an important consideration. Federal equal opportunity laws, such as the ECOA and Title VII of the Civil Rights Act of 1964, could be relevant in cases where AI might result in discrimination against protected classes. This includes not just inputs to models, but also unequal outcomes.

The European Commission has been considering this question and other key questions surrounding liability for AI for a number of years. According to the most recent version of its “Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things, and Robotics” (European Commission, 2020), emerging digital technologies like AI, challenge aspects of existing liability frameworks and could reduce their effectiveness. Some key challenges regarding liability in AI applications that are mentioned in the European Commission’s report are summarized below.

- **Difficulty in tracing damage back to human behavior:** Due to the complexity of AI technologies, it can be very difficult to trace the damage back to a human behavior. Therefore, it can be difficult for victims to identify the liable person and prove all necessary conditions for a successful claim. This means that liability claims based on national laws may be difficult or overly costly to prove and consequently victims may not be adequately compensated (European Commission, 2020).
- **Difficulty in understanding the algorithm and data used by AI:** Understanding the algorithm and data used by AI requires analytical capacity and technical expertise that victims could find prohibitively costly. Additionally, access to the algorithm and data could be impossible without cooperation from the potentially liable party. Therefore, victims may struggle to make a liability claim.
- **Legal complexity from AI interacting with traditional technologies:** Products and services relying on AI will interact with traditional technologies, leading to added complexity in terms of liability. For example, automated vehicles will share the road with traditional ones for a certain time. Similar complexity of interacting actors will arise in some services sectors (such as traffic management and healthcare) where partially automated AI systems will support human decision-making (European Commission, 2020).
- **Legal complexity from the plurality of actors involved:** Combining different digital components in a complex ecosystem and the plurality of actors involved

can make it difficult to assess where a potential damage originates, and which person is liable for it. The costs for this assessment may be economically prohibitive and discourage victims from claiming compensation.

- **Legal complexity from AI autonomy:** Autonomy brings an additional level of legal uncertainty and complexity. It would be unclear how to demonstrate the fault of AI acting autonomously, or what would be considered the fault of a person relying on the use of AI. Autonomy may alter a product's characteristics substantially and affect its safety. It remains a question under what conditions self-learning features prolong liability of the producer and to what extent should the producer have foreseen certain changes (European Commission, 2020).

These and other characteristics of AI operating within existing legal frameworks make liability an ongoing challenge.

Implications for ITS

A recent issue of a popular computing journal asked which laws would apply if an AI-enabled self-driving car killed a pedestrian. The paper considers the question of legal liability for artificially intelligent computer systems. It discusses whether criminal liability could ever apply; to whom it might apply; and, under civil law, whether an AI program is a product that is subject to product design legislation or a service to which the tort of negligence applies. The issue of sales warranties is also considered. A discussion of some of the practical limitations that AI systems are subject to is also included (Čerka et al., 2015).

- **Unclear responsibility for bias:** If the AI application fails due to bias in the data, it is currently unclear whether the liable party for the failure is the application developer or the data provider (Vasudevan, Townsend, Schweikert, et al., 2020). For example, if a computer vision-based pedestrian detection system fails to recognize all pedestrians (e.g., pedestrians in wheelchairs), then the system could jeopardize the safety of these individuals and others on the roadway, leading to potential liability issues for the overseeing agency. The GAO's *AI Accountability Framework* mentions that according to a forum participant, some entities are discouraged from collecting protected class data (e.g., sex, race, age) or taking steps to mitigate bias, because doing so may raise risks associated with anti-discrimination liability. Instead, these entities prefer to remain unaware because they consider this the safest way of proceeding (GAO, 2021).

- **Unclear responsibility for fatalities and crashes:** Liability is unclear when a vehicle, device, equipment, or system that is powered by an AI application is involved in a crash or results in fatalities (Vasudevan, Townsend, Schweikert, et al., 2020). If an AI-enabled automated vehicle is involved in an accident, the liability is presently unclear. The vehicle manufacturer could be held liable or alternatively the user of an AI-based tool could be responsible for decisions made by the tool (NIST, 2022). Uncertainty surrounding fault for damage could discourage investment as well as increase information and insurance costs for producers and other businesses in the supply chain (European Commission, 2020).

- **Unclear responsibility for cybersecurity breaches:** It is currently not clear what the safety expectations may be regarding the damage that results from cybersecurity breaches in an AI product and whether such damage would be adequately compensated (European Commission, 2020). Because of the multitudes of data used in machine learning systems, agencies risk more exposure to higher impact cybersecurity breaches. Potentially sensitive information might be leaked about many people. For more discussion, please see the Security and Privacy Sections.

- **Unclear responsibility for poor performance:** If an AI-enabled application has poor performance resulting in significant productivity losses, it is unclear who should be held accountable. For instance, who would be responsible in the case of an AI algorithm implemented for freight routing optimization that made suboptimal decisions resulting in unreliable freight deliveries. The answer is debatable and likely context dependent (Vasudevan, Townsend, Schweikert, et al., 2020).

These and other questions regarding liability are important for security, privacy, and other considerations.

Insights and Lessons Learned

Some potential strategies and lessons learned to avoid and address AI liability concerns are summarized below.

- **Partnering closely with agency risk management teams:** Consider including legal, compliance, records management, classification, civil liberties, and privacy professionals to understand governing authorities, legal obligations, information management responsibilities, and risks associated with an AI project (Office of the Director of National Intelligence (ODNI), 2020).

- **Assessing legal restrictions for the data:** This includes establishing authority, agreements, contracts that govern the collection or acquisition of all sources of data related to an AI model. In addition to arranging the technical aspects of data governance, this process could determine what legal or policy restrictions exist on the data (Office of the Director of National Intelligence (ODNI), 2020).

- **Assessing legal restrictions for the AI algorithm:** This includes what authorities or agreements apply to the AI algorithm itself, including the use, modification, storage, retrieval, access, retention, and disposition of the AI algorithm. The agency also may want to consider determining if there are any proposed downstream applications of the AI that are legally restricted from using the underlying data (Office of the Director of National Intelligence (ODNI), 2020).

- **Identifying possible risks throughout the AI pipeline:** This includes determining whether combining data with other outputs from the AI application creates new legal, records management, or classification risks relating to how the information is maintained and protected. For example, data covered under the Privacy Act should only be used for a purpose compatible with the reason for which the data was collected (Office of the Director of National Intelligence (ODNI), 2020).

- **Maintaining human accountability:** Accountability in AI ensures the designers and developers are responsible for abiding by the goals and objectives laid out in any governance charter and enforces liability through a chain of command that makes certain the systems operator oversees algorithm decisions. Saying a specific decision was made because an algorithm recommended a certain course of action is not a satisfying answer for either the public or regulators. In the end, specific people and organizations need to be held accountable (David Sweenor, 2021). The DIU's *Responsible AI Guidelines in Practice* suggests assigning a "responsible mission owner" who is accountable for ensuring that the capability meets operational, organizational, and ethical requirements. In addition, it is suggested that the mission owner work in consultation with legal counsel to ensure an AI system is developed in compliance with all relevant laws and regulations (Defense Innovation Unit (DIU) et al., 2021).

2.11 Talent/Workforce Availability

Table 13. Summary of the Talent/Workforce Availability Challenge and Potential Strategies to Address It

Summary of Talent/Workforce Availability	
What is it?	When there is lack of talent/expertise in building trustworthy, ethical AI algorithms, or integrating, operating, and maintaining real-world AI-based systems

Summary of Talent/Workforce Availability	
Why does it matter for ITS?	• Workforce talent and education are key bottlenecks to successful deployment and integration of AI systems into the operations of government agencies. • Domain experts in the transportation industry often do not have sufficient AI knowledge to work alongside data scientists in building models that are relevant and operationally useful. • Due to budget limitations, agencies have limited staff to operate and maintain AI-based systems. Therefore, balancing hiring decisions between ML/AI expertise and domain expertise can be a challenge.
How can it be addressed?	• Improving diversity in the workforce, and balancing AI talent and domain expertise to overcome challenges related to limited resources • Collaborating with partners for AI expertise • Providing client training to make deployment smoother, leading not only to improved technical proficiency of personnel but also buy-in for AI-enabled systems • Conducting periodic education and training for current staff, new hires, and domain experts, so they can keep up with advances in AI

Description of Challenge

Many new jobs are requiring that employees interface with AI systems as part of their duties (IEEE European Public Policy Initiative, 2017). However, a lack of workforce talent and relevant education in AI are often cited as key bottlenecks to successful deployment and integration of AI systems into the operations of government agencies. Lack of talent was found to be the number one obstacle to deploying AI by government agencies (*AI World Government In-Person & Virtual* | *October 18-19, 2021*). Thus, **there exists a lack of workforce expertise, especially in the public sector, in understanding the strengths, weaknesses, and risks** (e.g., security, privacy, bias, liability risks) **of AI**. Technicians in the field today have to have a wide array of knowledge on not only transportation operations, but also on advanced IT systems, networking, security, etc., which likely requires additional training. Additionally, decision makers are often not trained to be able to understand the policy and ethical issues related to AI. Lack of expertise in AI could potentially result in agencies blindly accepting unrealistic claims by vendors of AI products (Vasudevan, Townsend, et al., 2022). Although more universities are offering degrees in analytics and even AI, many graduates are still finding themselves unprepared for working with AI systems at enterprise scales (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022).

In addition, **domain experts in transportation and complementary fields (e.g., data science, cybersecurity, computer science) are needed** to work alongside data scientists in building AI models that are relevant and operationally useful. Domain experts, having knowledge and understanding of the essential aspects of a specific field, are essential. This is because the process of discovery and evaluation for AI systems development is guided by an intuitive knowledge of what has value, both in terms of input and output, a deep understanding of the underlying theory for classes of algorithms, and of what makes contextual sense (Colleen McCue, 2007). While an automated system can analyze and forecast events on its own, humans are needed to interpret the results, provide the best courses of action, and shape the AI model (Jamie Butler, 2016). Speakers from the ML/AI to Advance Earth System Science workshop stated that: "Many times, practitioners are very strong in computer science but struggle with fundamental aspects of data organization. Many have trouble operationalizing/ automizing data at enterprise scale. Additionally, many employees have issues with streaming data because they usually study fixed data in university" (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022).

Due to budget limitations, agencies have limited staff to operate and maintain AI-based systems, therefore **balancing hiring decisions between ML/AI expertise and domain expertise can be a challenge.** Often, they need to decide which side is easier to train. The first option is hiring domain experts who might be able to use ML tools as an off-the-shelf product, but do not intimately know advantages, constraints, and interpretations of ML techniques. The other option is hiring professionals with computer science backgrounds who are well versed in the ML but lack domain knowledge (*Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop*, 2022).

Implications for ITS

Key implications of lacking AI-related talent and workforce availability for ITS are summarized below.

- **Traditional engineers and system operators lacking AI training:** Traditional ITS systems are typically managed and operated by engineers with a civil engineering background and limited awareness of AI techniques. Even as AI is becoming more prominent in many fields of study, civil engineering degree programs have not started including basic AI concepts. With a basic understanding of AI or ML concepts, engineers and system operators can provide guidance to data analysts and ML engineers of what parameters are relevant to an ITS scenario. This can help expedite ML application development. According to recent market research on AI for ITS (Vasudevan, Townsend, et al., 2022), training clients on data collection for AI model development was challenging. Agency capacity to implement and monitor AI-based systems is limited. There have been instances where the agency staff, due to a lack of familiarity with the requirements of automated AI applications, would make changes to the data without notifying the AI team, causing the automated data processing steps to abort (Vasudevan, Townsend, et al., 2022).

- **Data scientists lacking background in transportation and ITS:** Data scientists and those trained to develop AI-enabled applications do not traditionally have a background in transportation management. Development of new AI systems will require workforce training for developers, data scientists, and engineers. Data analysts may be able to pull insightful information from a dataset to be used to train an ML model. However, data analysts may not fully understand what features to include or the dynamic nature of transportation. For predicting queues at an intersection as an example, not including traffic flows on adjacent facilities due to a lack of knowledge of how traffic propagates could result in poor prediction accuracies. The data may be useful, but in the context of ITS, it is also important to understand the relevance of the parameters of the model being trained. For example, both weather and traffic flows are critical parameters for an incident prediction model. Without these relevant parameters, data analysts not knowing these operational conditions for the network would not train models correctly.

- **Budget and staff limitations:** Due to budget limitations, agencies have limited staff to operate and maintain AI-based systems. For example, a speaker at the 2021 Annual Meeting of the Transportation Research Board noted that there is a lack of sustainable funding to support professional staff with expertise in both the domains of traffic engineering and AI system development. Traffic engineering domain experts are needed to prune and develop training data that carry localized information, but the workforce is limited with 1 engineer per 250-500 signals (*TRB Annual Meeting*, 2021). Given these limitations, agencies often procure AI solutions from vendors. For example, in speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (T. Geara et al., interview, April 2022), the City of Detroit mentioned that they do not have the resources to hire their own technology staff, so they rely on the technology provider/vendor to develop and deploy their solution, validate the data, and update the model when inconsistencies are found. Additionally, the Missouri DOT (MoDOT) shared with the ITS JPO's AI for ITS Program that they have many vendors involved in the AI/ML components of their ATCMTD deployment (E. Kopinski et al., interview, April 2022). Due in large part to the many vendors involved in different parts of the project, it has taken MoDOT longer to deploy the technology than they expected.

Insights and Lessons Learned

Some of the approaches for potentially overcoming the challenge of lack of talent and workforce availability are summarized below.

- **Improving diversity in the workforce:** Too often when organizations seek to hire talent in the AI space, they assume they should focus on a small cohort of schools. AI systems learn from examples, so it helps to have a diverse team that can bring different lenses to a problem and identify appropriate datasets for training AI models. It naturally follows that assembling a team with different backgrounds that can speak to different aspects of the problem will result in a better selection of datasets (Dzombak, Rachel & Palat, Jay, 2021). Creating diverse environments could also encourage learning and opportunities for collaboration. This way, even with limited resources, organizations can overcome the challenge of balancing AI talent and domain expertise to a problem and identify appropriate datasets for training AI models. It naturally follows that assembling a team with different backgrounds that can speak to different aspects of the problem will result in a better selection of datasets (Dzombak, Rachel & Palat, Jay, 2021). For example, as mentioned in the Bias Section, including a diverse team to develop an AI system will help with minimizing or detecting intentional and unintentional biases while collecting data, for feature selecting, building the AI-model, and making decisions. Creating diverse environments could also encourage learning and opportunities for collaboration. This way, even with limited resources, organizations can overcome the challenge of balancing AI talent and domain expertise.

- **Collaborating with expert partners:** In some cases, particularly given budget and staff limitations, it may make sense for an agency to partner with others for AI expertise. For example, in speaking to the ITS JPO's AI for ITS Program, the City of Detroit mentioned that they have become more involved with university collaborations to help manage advanced AI tasks and analytics for various projects and evaluations (T. Geara et al., interview, April 2022). In one case, the city worked with Wayne State University to develop a method to predict COVID case rates a week in advance based on recent traffic volumes collected from the ML-based Automated Traffic Signal Performance Measure (ATSPM) data collection system deployed and weather data using a deep learning model with long short-term memory networks. In another case, as part of the U.S. DOT Automated Driving System (ADS) Demonstration Program, the city is collaborating with the University of Michigan to test a self-driving shuttle that uses AI to navigate driving scenarios within a neighborhood to help mobilize senior citizens and people with disabilities in reaching places of interest.

- **Providing staff training to make deployment smoother:** In addition to computer science and data experts that develop AI systems, there are many who increasingly use and work with AI systems in their regular activities. For example, staff in traffic management centers may be using AI tools to enhance sensing, predict congestion, or improve control measures. Even though these staff do not need to be experts in AI development, they could benefit from some forms of training. For example, in speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (G. Donaldson & M. Rosica, interview, June 2022), the Delaware DOT (DelDOT) mentioned that their existing engineering and maintenance staff has readily adapted to technology changes with training and experience. Their latest addition of AI and ML capabilities will require additional appropriate levels of training. Overall, DelDOT emphasized that knowledge, skills, and abilities (as well as data management requirements) need to be defined to plan, design, construct, implement, operate, and maintain advanced systems, especially as all systems associated with DelDOT's Integrated Transportation Management System (ITMS) are becoming more complex and sophisticated (e.g., increasing quantity and variety of detectors and data for machine vision). From the perspective of technologists developing and deploying AI systems, training of client staff is essential and cost effective. Training operators and users on the basics of AI systems (such as the need for consistent data formats) helps ensure AI-based systems can be operated and maintained smoothly. Other trainings can include data preparation, modeling, and quality assurance. For example, as mentioned in the Model Drift Section, an AI model may need to be trained with new data when the initial setting starts to drift away. Staff trainings not only improve the technical proficiency of personnel, they also can help build buy-in for AI-based systems (Vasudevan, Townsend, et al., 2022).

- **Conducting periodic education and training to keep up with advances in AI:** As AI technology changes rapidly, continuous training ensures that staff keep up with advances in the field. In addition to computer and data scientists, domain experts or junior staff can also benefit from ongoing training in advances in AI. Addressing workforce needs will help maintain technological competitiveness and ensure that the skills acquired by the workforce remain relevant in the future (IEEE Advancing Technology for Humanity, 2019). The IEEE European Public Policy Initiative Position Statement article urged that government, business, and educational institutions share the responsibility for investment in education and training in order to increase the AI skilled workforce.

- **Investing in future workforce talent:** Government could work with academia (i.e., K-12 as well as higher education), the general public, and the private sector to increase the future workforce capacity in AI for ITS. For example, coordination activities could take the form of internships, studio classes, workshops, and/or challenges, such as hackathons.

2.12 Stakeholder Perception

Table 14. Summary of the Stakeholder Perception Challenge and Potential Strategies to Address It

Summary of Stakeholder Perception	
What is it?	When stakeholders are skeptical or mistrustful of AI systems or have exaggerated expectations of AI systems' capabilities

Summary of Stakeholder Perception	
Why does it matter for ITS?	• If stakeholders lack a clear understanding of the capabilities of AI, this can lead to skepticism and mistrust or to blind belief in AI as a solution for all problems, both of which could impede the successful implementation of AI. • Due to perceived high costs and risk aversion, agencies may prefer to deploy traditional ITS systems rather than AI-based systems. • Ethics, liability, and privacy issues could also affect stakeholder perception of AI. Agencies may have to contend with these institutional challenges when implementing AI solutions.
How can it be addressed?	• Conducting stakeholder analysis to identify stakeholders and their needs • Building trustworthy and ethical AI systems • Engaging with the user community early and often to gain buy-in and understand stakeholder needs • Demonstrating the value of AI to keep stakeholders on board with the project • Exchanging information with other deployers to share insights, lessons learned, and preliminary results • Ensuring leadership buy-in of AI techniques for initial and continued support • Setting stakeholder expectations, including on the implementation timeline • Promoting public understanding of AI to clarify what it is and how it could play a role

Description of Challenge

One of the biggest hurdles that AI faces today is lack of public trust and acceptance. Stakeholders' views of AI for ITS could impede its adoption, successful implementation, or acceptance, due to risk aversion, exaggerated expectations, or mistrust. These and other factors contributing to stakeholder perception are summarized below.

- **Lack of trust in decisions:** While AI techniques can improve data analysis and support decision making, they are often seen as "black boxes." Users' inability to articulate the rationale for a decision can affect their level of trust in AI (Phillips et al., 2021). According to a 2018 Pew Research Center survey on "Public Attitudes Toward Computer Algorithms" (Smith, Aaron, 2018), most Americans find it unacceptable to use algorithms to make decisions with real-world consequences for humans. Fifty eight percent (58%) of Americans feel that computer programs will always reflect some level of human bias. Two-thirds of Americans (68%) find the personal finance score algorithm unacceptable because this violates privacy. Fifty-seven percent (57%) of Americans find automated resume screening unacceptable because of the notion that the human element is removed from important decisions. Another prominent concern mentioned is that “humans are complex, and these systems are incapable of capturing nuance,” especially when it comes to criminal risk scores. Additionally, a report from the British Computer Society (BCS) revealed that more than half of UK adults (53%) do not trust organizations that use algorithms to make decisions about them (Leprince-Ringuet, Dephne, 2020). This mistrust or inability to explain or interpret decisions can be a major barrier to the adoption of AI. Please also see the Explainability Section for more information on this challenge.

- **Lack of understanding of AI capabilities:** Due to a lack of AI knowledge, stakeholders may not have accurate expectations of the functionality and reliability of AI systems. According to recently conducted market research on AI for ITS (See Appendix A), AI product vendors have to provide detailed training to agency staff on the data preparation, analytic models, and quality assurance procedures to overcome their skepticism. While some stakeholders may be skeptical of AI, others may be overly enthusiastic about the capabilities of AI and view it as a solution for all problems. Neither blind faith nor skepticism of AI is helpful for its successful implementation. A blind believer may feel disillusioned if they do not fully understand the drawbacks of AI while a skeptic may miss out on AI’s benefits if they are not aware of the advantages of AI. Please also see the Talent/Workforce Availability Section for more information.

- **Fear of obsolescence and wasted investments:** While agencies may be open to exploring AI solutions, they are often concerned with how quickly technology can become obsolete. Dynamism in the field of AI can contribute to agencies’ perception that investing in long-term AI solutions is undesirable for fear of wasting their investments (Vasudevan, Townsend, Schweikert, et al., 2020).

- **Fear of unethical decisions and liability issues:** Ethics and liability concerns could also affect stakeholder perception of AI. According to the 2018 Pew Research Center survey (Smith, Aaron, 2018), the public is concerned about the fairness and acceptability of using computers for decision-making in situations with important real-world consequences, such as criminal risk assessment (56%), automated resume screening (57%), automated video analysis of job candidates (67%), and personal finance scoring (68%). Please see the Ethics and Equity, and Liability Sections for more information.

- **Fear of privacy loss:** AI uses various data to train the system, including sensitive data in some cases. Stakeholders may have concerns about how their personal information may be used and/or leaked to a 3rd party for other purposes. For example, according to the 2018 Pew Research Center survey (Smith, Aaron, 2018), privacy violation is a top concern mentioned by respondents, particularly for those who find personal finance scoring by algorithms unacceptable. Please see the Privacy Section for more information.

Implications for ITS

Some of the ways in which stakeholder perception of AI/ML could come into play for ITS are summarized below.

- **Risk aversion:** Agencies are often cautious when investing in untested or unproven technologies. Risk aversion is one of the primary reasons limiting an agency's inclination to experiment or deploy un-proven AI solutions (Vasudevan, Townsend, Schweikert, et al., 2020). For example, an agency may be hesitant to deploy an AI-enabled cyberattack prediction system that has only been tested using a simulation model, even if shown to accurately predict cyberattacks since the AI-enabled solution has not been tested in a real-world environment.

- **Budget constraints:** The cost of implementation of certain AI-enabled applications could prove to be significantly higher than the cost of a conventional system that already provides adequate performance. Due to limited resources, agencies are responsible for spending public funding responsibly and tend to avoid investing in innovative solutions that have not been tried before. Recent market research on AI for ITS (See Appendix A) has revealed that budget constraints and limited federal grant availability, particularly "short-term" funding, are the biggest barriers to widespread deployment of advanced safety systems; the need for funding exceeds availability. Additionally, many municipalities are often unaware of available funds or how to access them. These issues could limit or slow adoption.

- **Institutional challenges:** ITS technologies are often publicly accessed systems, so agencies will need to contend with institutional challenges related to privacy issues (such as PII), ethical issues, liability issues and other policy issues when implementing AI solutions (Vasudevan, Townsend, Schweikert, et al., 2020). For example, liability may be an issue in a case where an AI algorithm that determines automated vehicle behaviors leads to a crash. In another example, an AI-based travel time prediction system may need to collect vehicle trajectories to calculate travel times or recommend routes, which could potentially impact privacy. Additionally, AI-based systems may not be permitted to use certain data for new use cases without permission from those involved. For example, if an agency decides to install new camera equipment into signal controller cabinets for a procured computer-vision based system, they may need to make the public aware that their images will be captured at those locations.

Insights and Lessons Learned

Some approaches to potentially address negative stakeholder perception of AI are summarized below.

- **Conducting stakeholder analysis to identify stakeholders and their needs:** Before pursuing AI, agencies could conduct stakeholder analysis to identify all stakeholder groups who may affect or be affected by the project and their needs. While conducting a stakeholder analysis is considered a best practice for many project types, it could be especially valuable for projects involving emerging technology, such as AI, that may include new or different stakeholder groups. For example, relevant stakeholders in transportation may include advocates for particular positions (e.g., safety, environment, modes, etc.), associations of companies (e.g., OEMs, construction), constituent groups (e.g., commuters, tourists), elected officials, government agencies, and labor unions (Steier, 2021). In addition to identifying all stakeholders and their needs, it could be helpful to assess their level of interest and influence. For example, some stakeholders may have high influence/power but low interest in the project and its specifics. Agencies may want to keep these stakeholders satisfied but may not need to manage them too closely or keep them informed of all project details. Conversely, some stakeholders may have high interest but low influence/power, and would therefore, appreciate being kept in the loop of project happenings (i.e., regular updates) (Steier, 2021).

- **Building trustworthy and ethical AI systems:** For gaining trust, it is crucial for the AI system to consistently produce outputs that are reasonable, auditable, and explainable. This could start from making “data science a trusted profession – as trusted as the profession of doctor or lawyer,” suggested the director of policy at the British Computer Society (Leprince-Ringuet, Dephne, 2020). Documentation, visualizing results, and other insights and lessons learned from the Explainability, and Ethics and Equity Sections could be potential strategies in building trustworthy and ethical AI systems.

- **Engaging with the user community early and often:** Kicking off a project idea by engaging with expected stakeholders can improve not only stakeholder perception and buy-in but also streamline the development process. For example, the Tennessee DOT (TDOT) as part of their ATCMTD deployment, started off the process by hearing from the TMC, TDOT, and others, which made them “highly focused to deploy the right tools in the right spot” (L. Smith et al., interview, April 2022). Community engagement is not a one-time event. Instead, building and connecting with the user community is an ongoing effort. In speaking to the ITS JPO’s AI for ITS Program about their ATCMTD deployment (M. Haselkorn et al., interview, April 2022), the Washington State DOT (WSDOT) emphasized the importance of regularly engaging with the user community in a highly agile and participatory process when developing their ML use cases as part of their ATCMTD deployment. From the beginning, they worked to “win over the hearts of minds” of the user community and bring everyone to the table to discuss options. Additionally, the Delaware DOT (DelDOT) in speaking to the ITS JPO’s AI for ITS Program about their ATCMTD deployment (G. Donaldson & M. Rosica, interview, June 2022), emphasized the importance of developing relationships. For example, they have internal meetings every week with key user groups (e.g., state IT managers, DelDOT's IT group) to discuss how things are working and what needs changed.

- **Demonstrating the value of AI:** Deploying AI for the sake of deploying AI may not be the best strategy for winning over stakeholders. Instead, stakeholders will likely want to see how the AI deployment will address their needs. Once stakeholders initially buy-into the project concept with AI, the deployment will likely need to demonstrate the value of AI over time to keep them engaged. For example, in speaking to the ITS JPO's AI for ITS Program about their ATCMTD deployment (M. Haselkorn et al., interview, April 2022), WSDOT emphasized the need for their AI/ML application to demonstrate value to keep users and stakeholders onboard with the project.

- **Exchanging information with other deployers:** Providing opportunities for engagement across deployments could be a valuable way to assuage stakeholder hesitancies and allow AI deployers to learn from one another. Peer exchanges and cohorts (e.g., the Early Deployer ATCMTD Cohort) are two such outlets for engagement that the USDOT and other agencies have used to facilitate information sharing. In these informal settings, deployers can learn how others have implemented specific technologies, such as AI, as well as any challenges, lessons learned, and impacts they have seen. A benefit of this form of engagement is that deployers can share interim information about ongoing projects, which allows other deployers to learn and ask about the project more quickly than they would be able to otherwise (i.e., from waiting for a final report).

- **Ensuring leadership buy-in on AI techniques:** Leadership buy-in is a critical component to implement AI techniques into ITS. Leadership could include federal, state, and local agency leaders, among other key stakeholders. In speaking to the ITS JPO's AI for ITS Program (E. Kopinski et al., interview, April 2022), staff from the Missouri DOT (MoDOT) emphasized the importance of educating leadership when considering deploying AI/ML. While some decision makers may expect AI to be a near-term panacea, others may think AI is still 100 years away. Educating leaders on both the possibilities and potential challenges of AI could help set realistic expectations and foster buy-in. With leadership buy-in, agencies would have stronger support to develop standards, provide resources, educate the workforce, address institutional challenges, and resolve policy issues.

- **Setting stakeholder expectations:** According to the USDOT's 2020 *Plan for AI for ITS* report (Vasudevan, Townsend, Schweikert, et al., 2020), there is a need to develop a convincing narrative to set expectations and motivate stakeholders (e.g., agencies and their partners, original equipment manufacturers [OEMs], vendors, and developers). This could include a discussion of the evolutionary deployment and expected impacts, beginning with cost-effective, near-term deployments or prototype demonstrations that can evolve into complex and transformative long-term deployments. In speaking to the ITS JPO's AI for ITS Program (T. Geara et al., interview, April 2022), the City of Detroit mentioned the need for knowledge transfer as a key lesson learned as part of their ATCMTD deployment, which includes AI. Understanding AI's capabilities and constraints and being realistic in what it can and cannot do is important. Involved staff do not need to be experts but having some baseline knowledge is helpful.

- **Promoting public understanding of AI:** The success of AI technology depends on the ease with which people use and adapt AI applications. Therefore, promoting an understanding of AI and fostering trust with the public is beneficial for its successful implementation (IEEE European Public Policy Initiative, 2017). In addition, public opinion related to trust, safety, privacy, employment, society, and the economy tends to influence public policy (IEEE Advancing Technology for Humanity, 2019). According to a 2021 Study Panel Report on AI by Stanford University (Littman, etc., 2021), the AI community could facilitate a clearer public understanding that reduces confusion between AI and other information technologies. For example, a taxonomy of AI could serve as a useful frame of reference. Additionally, according to the report, participatory engagement and conversation *with* the public are considered more effective outreach mechanisms than educating or talking *to* the public after the fact. Many organizations are developing more deliberative and participatory models of AI public engagement, such as blogs and forums (Littman, etc., 2021). Such efforts could help boost public interest in and democratic involvement with AI.

3 Key Takeaways

The report summarized 12 challenges to AI adoption and implementation in ITS. The challenges and potential solutions were identified based on the AI for ITS Program's research and market engagement efforts to date. Please see Table 1 in the Executive Summary for a complete list of these 12 challenges, and their corresponding implications for ITS, and insights and lessons learned. Overarching key takeaways are summarized below.

- **The twelve challenges for AI adoption and successful implementation are not unique to ITS.** They are broad technical and institutional challenges that impact a wide variety of sectors. Many of the insights and lessons learned in this report are gleaned from other sectors and could potentially be applied to ITS.

- **There may be tradeoffs between addressing different challenges.** For example, greater explainability could provide more information for malicious actors to manipulate, potentially breaching security and/or privacy. Adding robust, large scale data sources may boost AI performance but could be costly to store and implement.

- **Addressing these challenges is an ongoing exercise.** These challenges are dynamic and, like AI itself, will evolve over time. For example, cybersecurity concerns today may look different than cybersecurity concerns next year as malicious actors find new ways to hack into systems. Additionally, stakeholder buy-in is important not only at the onset of a project but also throughout the project to support its continued success. The deployment of new AI techniques may require new staff expertise. Overall, challenges and risks are dynamic and addressing them is an ongoing exercise.

- **Maintaining a human-in-the-loop is helpful in identifying and mitigating these challenges.** Ongoing human oversight of AI/ML applications in ITS can help in identifying and mitigating potential issues, particularly those that the machine may not catch and those that may require making tradeoffs in how they are addressed. Having both domain and AI/ML expertise on staff is useful for not only the initial implementation but for ongoing operations and maintenance of the system and its AI/ML applications.

References

1. ABOUT ML. (n.d.). *Partnership on AI*. Retrieved March 21, 2022, from https://partnershiponai.org/workstream/about-ml/

2. Adversarial Machine Learning, 2019. DeepAI. May 17, 2019. https://deepai.org/machine-learning-glossary-and-terms/adversarial-machine-learning.

3. *AI World Government In-Person & Virtual* | October 18-19, 2021. (n.d.). AI World Gov – Virtual. Retrieved March 16, 2022, from https://www.aiworldgov.com

4. Amazon Web Services, Inc. (n.d.). *Amazon Machine Learning: Developer Guide*. https://docs.aws.amazon.com/machine-learning/latest/dg/retraining-models-on-new-data.html

5. Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software Engineering for Machine Learning: A Case Study. *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

6. *Artificial Intelligence and the Economy: Charting a Path for Responsible and Inclusive AI.* (2022, April 27). Stanford HAI. https://hai.stanford.edu/events/artificial-intelligence-and-economy-charting-path-responsible-and-inclusive-ai

7. Baillargeon, J. (2019, June 5). *Artificial Intelligence Research at FRA.* Presentation at the *2019 ITS America Annual Meeting*, Federal Railroad Administration. https://www.its.dot.gov/presentations/itsa_2019/Baillargeon_ITS19_20190523.pdf

8. Banaei-Kashani, F., & Rens, K. (2021). *Improving Deep Learning Models for Bridge Management Using Physics-Based Deep Learning.* https://www.mountain-plains.org/research/details.php?id=547

9. Čerka, P., Grigienė, J., & Sirbikytė, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, 31(3), 376–389. https://doi.org/10.1016/j.clsr.2015.03.008

10. Chan-Edmiston, S., Fischer, S., Sloan, S., & Wong, M. (2020). *Intelligent Transportation Systems Joint Program Office: Strategic Plan 2020-2025* (FHWA-JPO-18-746). USDOT ITS Joint Program Office. https://www.its.dot.gov/stratplan2020/

11. Chenok, D. (2020, May 18). *Risk-Based Decisionmaking for Applying AI and Other Emerging Technologies: Findings from Recent Research | IBM Center for The Business of Government*. https://www.businessofgovernment.org/blog/risk-based-decisionmaking-applying-ai-and-other-emerging-technologies-findings-recent-research

12. Chollet, F. (2017). *Xception: Deep Learning with Depthwise Separable Columns*. Google, Inc. http://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf

13. Chouldechova, A. (2017). *Introduction to Data Mining*. Carnegie Mellon University.

14. Chouldechova, A., & Roth, A. (2020). A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM*, 63(5), 82–89. https://doi.org/10.1145/3376898

15. Colleen McCue. (2007). *Data Mining and Predictive Analysis*. Science Direct. https://www.sciencedirect.com/topics/computer-science/domain-expertise

16. Colley, M., Walch, M., Gugenheimer, J., Askari, A., & Rukzio, E. (2020). Towards Inclusive External Communication of Autonomous Vehicles for Pedestrians with Vision Impairments. Proceedings of the *2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376472

17. *Considerations for Sensitive Data within Machine Learning Datasets | Cloud Architecture Center.* (n.d.). Google Cloud. Retrieved February 1, 2022, from https://cloud.google.com/architecture/sensitive-data-and-ml-datasets

18. Crawford, K., Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., & Daumé III, H. (2021, December). *Datasheets for Datasets*. https://cacm.acm.org/magazines/2021/12/256932-datasheets-for-datasets/fulltext

19. DARPA. (n.d.). *Spectrum Collaboration Challenge*. Retrieved May 2, 2022, from https://www.darpa.mil/about-us/timeline/spectrum-collaboration-challenge

20. *Data Sharing Agreements | U.S. Geological Survey.* (n.d.). Retrieved April 6, 2022, from https://www.usgs.gov/data-management/data-sharing-agreements

21. David Sweenor. (2021, October). *The Case for a Global Responsible AI Framework*. KD Nuggets. https://www.kdnuggets.com/2021/10/responsible-ai-framework.html

22. Day, C., O'Brien, P., Stevanovic, A., Hale, D., & Matout, N. (2020*). A Methodology and Case Study: Evaluating the Benefits and Costs of Implementing Automated Traffic Signal Performance* (FHWA-HOP-20-003). Article FHWA-HOP-20-003. https://trid.trb.org/view/1740631

23. U.S. Department of Defense (2022). *DOD Adopts Ethical Principles for Artificial Intelligence*. U.S. Department of Defense. https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/

24. Donaldson, G., & Rosica, M. (2022, June 13). *ATCMTD AI for ITS Interview with the Delaware DOT* [Virtual].

25. Donaldson, G., Trueman, C., & Zhao, G. (2022, January 26). *Artificial Intelligence Enhanced Integrated Transportation Management System (AI-ITMS): DelDOT's ATCMTD Project Update* [Virtual].

26. Dunnmon, J., Goodman, B., Kirechu, P., Smith, C., & Van Deusen, A. (2021). *Responsible AI Guidelines In Practice: Lessons Learned from the DIU AI Portfolio*. In U.S. Department of Defense. Defense Innovation Unit. https://www.diu.mil/responsible-ai-guidelines

27. Dzombak, Rachel & Palat, Jay. (2021, August 30). *5 Ways to Start Growing an AI-Ready Workforce*. https://insights.sei.cmu.edu/blog/5-ways-to-start-growing-an-ai-ready-workforce/

28. Ergan, S., Zou, Z., Khan, J., Bernardes, S. D., Lu, D., & Shen, Y. (2021). *Work Zone Safety: Behavioral Analysis with Integration of VR and HIL*. 40.

29. European Commission. (2020). *Commission Report on safety and liability implications of AI, the Internet of Things and Robotics*. https://ec.europa.eu/info/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en

30. Farajian, M. (2019). *FIBER OPTICS OPPORTUNITIES INITIATIVE PHASE 1 FINDINGS*. 19.

31. *Feature importances with a forest of trees.* (n.d.). Scikit-Learn. Retrieved March 24, 2022, from https://scikit-learn/stable/auto_examples/ensemble/plot_forest_importances.html

32. *Federated Learning: Collaborative Machine Learning without Centralized Training Data.* (2017, April 6). Google AI Blog. http://ai.googleblog.com/2017/04/federated-learning-collaborative.html

33. FHWA. (2020, June 16). *U.S. Department of Transportation Awards $43.3 Million in Advanced Transportation and Congestion Management Technologies Grants.* https://cms8.fhwa.dot.gov/newsroom/us-department-transportation-awards-433-million-advanced-transportation-and-congestion

34. FHWA EAR Program. (2022). *The Role of Artificial Intelligence and Machine Learning in Federally Supported Surface Transportation: 2022 Updates* (FHWA-HRT-22-026). https://rosap.ntl.bts.gov/view/dot/61039

35. FHWA Office of Operations. (2022, January 31). *Tools for Tactical Decision-Making/Advancing Methods for Prediction Performance.* Research: Active Transportation and Demand Management. https://ops.fhwa.dot.gov/atdm/research/index.htm#ttdm

36. FHWA, Technology Readiness Level Guidebook, Publication No. FHWA-HRT-17-047, September 2017. Report Link: https://www.fhwa.dot.gov/publications/research/ear/17047/17047.pdf

37. Fries, R., Chowdhury, M. A., & Reisi Gahrooei, M. (2011). Maintaining Privacy While Advancing Intelligent Transportation Systems Applications—An Analysis (No. 11–1490). Article 11–1490. *Transportation Research Board 90th Annual Meeting*. https://trid.trb.org/view/1091966

38. Frikken, K. B. (2011). Secure Multiparty Computation (SMC). In H. C. A. van Tilborg & S. Jajodia (Eds.), *Encyclopedia of Cryptography and Security* (pp. 1121–1123). Springer US. https://doi.org/10.1007/978-1-4419-5906-5_766

39. GAO. (2021). *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities* (GAO-21-519SP). https://www.gao.gov/products/gao-21-519sp

40. Gartner. (2018). https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders

41. Geara, T., Chee, J., Fisher, T., Jacob, S., & Adegbaju, O. (2022, April 27). *ATCMTD AI for ITS Interview with the City of Detroit* [Virtual].

42. *Generalization | Machine Learning Crash Course.* (n.d.). Google Developers. Retrieved March 18, 2022, from https://developers.google.com/machine-learning/crash-course/generalization/video-lecture

43. Ghani, R., Rodolfa, K. T., & Saleiro, P. (2021). Dealing with Bias and Fairness in AI/ML/Data Science Systems. *International Conference on Computational Social Science.*

44. Google. (2021, September). *Machine Learning Glossary | Google Developers*. https://developers.google.com/machine-learning/glossary

45. Gunning, D., Vorm, E., Wang, J. Y., & Turek, M. (2021). DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters*, 2(4), 11. https://doi.org/10.1002/ail2.61

46. Hahn, D., Munir, A., & Behzadan, V. (2019). Security and Privacy Issues in Intelligent Transportation Systems: Classification and Challenges. *IEEE Intelligent Transportation Systems Magazine*, PP, 1–1. https://doi.org/10.1109/MITS.2019.2898973

47. Harris, J. (2022, February 9). Catching edge cases in AI: Yaron Singer on building more robust, fault-tolerant AI systems. *Medium*. https://towardsdatascience.com/catching-edge-cases-in-ai-b9860589ece4

48. Haselkorn, M., Aragon, C., Kendrick, J., Cornell, B., & Phelps, T. (2022, April 12). *ATCMTD AI for ITS Interview with the Washington State DOT* [Virtual].

49. Haselkorn, M., & Webster Heublein, H. (2022, March 28). *Demonstration of the Virtual Coordination Center (VCC) by Washington State DOT*. Early Deployer (ATCMTD) Cohort Monthly Meeting, Virtual.

50. Hatcher, S. G., McGurrin, M. F., Vasudevan, M., Burgess, L., Haase, D., Levine, S., & Havinoviski, G. (2017). *Freeway Management and Operations Handbook* (No. FHWA-HOP-17-031). Federal Highway Administration Office of Operations.

51. Heaven, W. (2021, July 30). Hundreds of AI tools have been built to catch covid. None of them helped. *MIT Technology Review.* https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/

52. Hoi, S. C. H., Sahoo, D., Lu, J., & Zhao, P. (2018). Online Learning: A Comprehensive Survey. *ArXiv:1802.02871* [Cs]. http://arxiv.org/abs/1802.02871

53. Horneman, A., Mellinger, A., & Ozkaya, I. (2019). *AI Engineering: 11 Foundational Practices.* Carnegie Mellon University: Software Engineering Institute, 4.

54. IEEE Advancing Technology for Humanity. (2019). *IEEE Position Statement Artificial Intelligence.* https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE18029.pdf

55. IEEE European Public Policy Initiative. (2017). Artificial Intelligence: Calling on Policy Makers to Take a Leading Role in Setting a Long-Term AI Strategy. *IEEE European Public Policy Initiative.* http://globalpolicy.ieee.org/wp-content/uploads/2017/10/IEEE17021.pdf

56. ITS Data Sandbox—NYCDOT – DataProcessing.txt. (2021). [Jupyter Notebook]. Usdot-its-jpo-data-portal. https://github.com/usdot-its-jpo-data-portal/sandbox/blob/master/doc/nycdot/DataProcessing.txt (Original work published 2017)

57. ITS JPO. (2022). *Leveraging Existing Infrastructure and Computer Vision for Pedestrian Detection.* ITS JPO ITS Deployment Evaluation Program. https://www.itskrs.its.dot.gov/decision-support/case-study/leveraging-existing-infrastructure-and-computer-vision-pedestrian

58. Jamie Butler. (2016, September 9). Domain Expertise And AI: Conquering The Next Generation Of Cyber Threats. *Forbes Magazine.* https://www.forbes.com/sites/forbestechcouncil/2016/09/09/domain-expertise-and-ai-conquering-the-next-generation-of-cyber-threats/?sh=14f0a62a6cd1

59. *Kicking off NIST AI Risk Management Framework Workshop.* (October 19-21, 2021). NIST. https://www.nist.gov/news-events/events/2021/10/kicking-nist-ai-risk-management-framework

60. Kieu, M., Bagdanov, A., Bertini, M., & Bimbo, A. (2019). Domain Adaptation for Privacy-Preserving Pedestrian Detection in *Thermal Imagery* (pp. 203–213). https://doi.org/10.1007/978-3-030-30645-8_19

61. Kopinski, E., Intaratip, P., Wassman, A., & Dolde, M. (2022, April 12). *ATCMTD AI for ITS Interview with the Missouri DOT* [Virtual].

62. Krause, C., Anderson, J., Shain, K., Nana, L., Mazzone, T., McNaught, S., & Jackson, M. (2019). *Cybersecurity and Intelligent Transportation Systems: A Best Practice Guide* (FHWA-JPO-19-763). USDOT ITS Joint Program Office. https://rosap.ntl.bts.gov/view/dot/42461

63. Lattimer, C. R. & Atkins North America. (2020). *Automated Traffic Signals Performance Measures* (FWHA-HOP-20-002). https://rosap.ntl.bts.gov/view/dot/54065

64. Learn, W. (2014). *Smart City Readiness: Understand the Issues to Accelerate the Journey.* 8. https://www.cisco.com/c/dam/m/en_in/innovation/smartcities/assets/white-paper-c11-732985.pdf

65. Leprince-Ringuet, Dephne. (2020, September 8). *Big bad data: We don't trust AI to make good decisions.* https://www.zdnet.com/article/big-bad-data-we-dont-trust-ai-to-make-good-decisions/

66. Li, Tiancheng, and Ninghui Li. On the tradeoff between privacy and utility in data publishing. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, pp. 517-526. 2009.

67. Littman, etc. (2021). Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. https://ai100.stanford.edu/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence-ai100-2021-study

68. Lo, Y. L., & Ghiassian, S. (2019). Overcoming Catastrophic Interference in Online Reinforcement Learning with Dynamic Self-Organizing Maps. *ArXiv:1910.13213* [Cs]. http://arxiv.org/abs/1910.13213

69. Lopez Conde, M., & Twinn, I. (2019). How Artificial Intelligence is Making Transport Safer, Cleaner, More Reliable and Efficient in Emerging Markets. International Finance Corporation, Washington, DC. https://doi.org/10.1596/33387

70. *Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges Workshop.* (2022, February 10). National Academies of Sciences, Engineering and Medicine. https://www.nationalacademies.org/our-work/machine-learning-and-artificial-intelligence-to-advance-earth-system-science-opportunities-and-challenges---a-workshop

71. MDOT SHA. (2022). https://roads.maryland.gov/mdotsha/pages/Index.aspx?PageId=872

72. Milli, S., Schmidt, L., Dragan, A. D., & Hardt, M. (2018). Model Reconstruction from Model Explanations. *ArXiv:1807.05185* [Cs, Stat]. http://arxiv.org/abs/1807.05185

73. National ITS Architecture Team. (2007). *Systems Engineering for Intelligent Transportation Systems: An Introduction for Transportation Professionals* (FHWA-HOP-07-069). https://ops.fhwa.dot.gov/publications/seitsguide/index.htm

74. Nevada DOT Shares Grant for AI-based Traffic Project. (2019, October 18). *AASHTO Journal.* https://aashtojournal.org/2019/10/18/nevada-dot-shares-grant-for-ai-based-traffic-management-project/

75. NIST. (2022). *AI Risk Management Framework (Draft).* NIST. https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf

76. NIST. (2022, March). *Building the NIST AI Risk Management Framework: Workshop #2.* NIST. https://www.nist.gov/news-events/events/2022/03/building-nist-ai-risk-management-framework-workshop-2

77. NIST Computer Security Resource Center. (n.d.). Glossary: sensitive information. Retrieved March 29, 2022, from https://csrc.nist.gov/glossary/term/sensitive_information

78. NTRD (2022). "About the Networking and Information Technology Research and Development (NITRD) Program." https://www.nitrd.gov/about/.

79. NTRD (2021). "AI R&D Testbed Inventory." https://www.nitrd.gov/apps/ai-rd-testbed-inventory/.

80. Office of the Director of National Intelligence (ODNI). (2020). *Artificial Intelligence Ethics Framework for the Intelligence Community, v. 1.0.* https://www.dni.gov/index.php/features/2763-principles-of-artificial-intelligence-ethics-for-the-intelligence-community

81. Oleszak, M. (2021, September 21). Don't let your model's quality drift away. *Medium*. https://towardsdatascience.com/dont-let-your-model-s-quality-drift-away-53d2f7899c09

82. Ozbay, K. (2022). COVID-19's Effect on Transportation: Developing a Public COVID-19 Data Dashboard – C2SMART Home. https://c2smart.engineering.nyu.edu/covid-19-sociability-board/

83. Phillips, P. J., Hahn, C. A., Fontana, P. C., Yates, A. N., Greene, K., Broniatowski, D. A., & Przybocki, M. A. (2021). *Four Principles of Explainable Artificial Intelligence* (NIST Interagency/Internal Report (NISTIR)-8312). https://www.nist.gov/publications/four-principles-explainable-artificial-intelligence

84. Prabhu, M. (2020, September 22). Security & Privacy in Artificial Intelligence & Machine Learning — Part-6: Up close with Privacy. *Medium*. https://towardsdatascience.com/security-privacy-in-artificial-intelligence-machine-learning-part-6-up-close-with-privacy-3ae5334d4d4b

85. Qi, B. (2020). *IMPROVING SMART TRANSPORTATION APPLICATIONS WITH VEHICULAR EDGE COMPUTING*. 139.

86. Qian, S. (2021). AI and Predictive Analytics: How to Make Better Decisions with Transportation Data? As part of the Carnegie Mellon University *Managing AI in Transportation* course.

87. Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

88. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

89. SAE Levels of Driving Automation Refined for Clarity and International Audience. (n.d.). Retrieved March 16, 2022, from https://www.sae.org/blog/sae-j3016-update

90. Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence.* National Institute of Standards and Technology (NIST). https://doi.org/10.6028/NIST.SP.1270

91. Shendre, S. (2020, May 14). Model Drift in Machine Learning models. *Medium*. https://towardsdatascience.com/model-drift-in-machine-learning-models-8f7e7413b563

92. Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. *2017 IEEE Symposium on Security and Privacy (SP),* 3–18.

93. Singer, Y. (2022, June 28). Eliminating AI Risk – One Model Failure at a Time. Presentation at the *Data+AI Summit 2022* organized by databricks. https://databricks.com/dataaisummit/session/eliminating-ai-risk-one-model-failure-time

94. Smith, A. (2020, April 8). *Using Artificial Intelligence and Algorithms*. Federal Trade Commission. http://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms

95. Smith, A. (2018, November 16). *Public Attitudes Toward Computer Algorithms.* Pew Research Center. https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algorithms/

96. Smith, L., Nickerson, M., ElSaid, S., Work, D., Weston, C., Davis, M., & Heimsness, P. (2022, April 20). *ATCMTD AI for ITS Interview with the Tennessee DOT* [Virtual].

97. Statistica. (2021). Statista. https://www.statista.com/statistics/967365/worldwide-cloud-infrastructure-services-market-share-vendor/

98. Steier, D. (2021, May 7). Developing and Implementing an Enterprise AI Strategy. As part of the Carnegie Mellon *University Managing AI in Transportation* course.

99. Systems Engineering Guidebook for Intelligent Transportation Systems, Version 3.0 (November 2009). https://www.fhwa.dot.gov/cadiv/segb/

100. Telenti, A., & Jiang, X. (2020). Treating medical data as a durable asset. *Nature Genetics*, 52(10), 1005–1010.

101. Thaine, P. (2020, October 8). Perfectly Privacy-Preserving AI. *Medium*. https://towardsdatascience.com/perfectly-privacy-preserving-ai-c14698f322f5

102. Thompson, D. (2019, June). *Driving Innovation: Early Impact of AI on Highway Transportation*. Presentation at the *2019 ITS America Annual Meeting*, FHWA. https://www.its.dot.gov/presentations/itsa_2019/ITS_AmericaAI_Thompson.pdf

103. TRB Annual Meeting. (2021). *Transportation Research Board Annual Meeting* 2021.

104. Turri, V. (2022, January 17). *What is Explainable AI?* SEI Blog. https://insights.sei.cmu.edu/blog/what-is-explainable-ai/

105. TxDOT. (2021, January). *TxDOT Statewide Systems Engineering Version 3.1.* https://ftp.txdot.gov/pub/txdot-info/trf/tsmo/systems-engineering.pdf

106. U.S. Department of HHS. (2021). *Artificial Intelligence Strategy.* https://www.hhs.gov/sites/default/files/final-hhs-ai-strategy.pdf

107. USDOT. (2022a). Using Connected Vehicle Technologies to Solve Real-World Operational Problems. https://www.its.dot.gov/pilots/

108. USDOT. (2022b, January 31). DOT Inclusive Design Challenge. https://www.transportation.gov/accessibility/inclusivedesign

109. USDOT FHWA. (2021, July 30). Artificial Intelligence for Intelligent Transportation Systems: Sources Sought. SAM.Gov. https://sam.gov/opp/10cf7c2f22a04c098e767e25eca2fdee/view

110. USDOT Office of the Assistant Secretary for Research and Technology. (2021, July). Using Video Feeds from Public Traffic Cameras and Computer Vision to Analyze Social Distancing and Travel Patterns during the COVID-19 Pandemic. *UTC Spotlight, Quarterly Issue 3.* https://www.transportation.gov/utc/using-video-feeds-public-traffic-cameras-and-computer-vision-analyze-social-distancing-and

111. Vasudevan, M., O’Hara, J., Townsend, H., Asare, S., Muhammad, S., Ozbay, K., Yang, D., Gao, J., Kurkcu, A., & Zuo, F. (2022). *NCHRP Research Report 997: Algorithms to Convert Basic Safety Messages into Traffic Measures* (National Cooperative Highway Research Program, Transportation Research Board NCHRP 03-137).

112. Vasudevan, M., Townsend, H., Dang, T. N., O’Hara, A., Burnier, C., & Ozbay, K. (2020). *Identifying Real-World Transportation Applications Using Artificial Intelligence (AI): Summary of Potential Application of AI in Transportation* (FHWA-JPO-20-787). https://rosap.ntl.bts.gov/view/dot/50651

113. Vasudevan, M., Townsend, H., & Schweikert, E. (2020). *Identifying Real-World Transportation Applications Using Artificial Intelligence (AI): Plan for Artificial Intelligence for Intelligent Transportation Systems* (FHWA-JPO-20-813). https://rosap.ntl.bts.gov/view/dot/53932

114. Vasudevan, M., Townsend, H., Schweikert, E., Wunderlich, K. E., Burnier, C., Hammit, B. E., Gettman, D., & Ozbay, K. (2020). *Identifying Real-World Transportation Applications Using Artificial Intelligence (AI)- Real-World AI Scenarios in Transportation for Possible Deployment* (FHWA-JPO-20-810). https://rosap.ntl.bts.gov/view/dot/50752

115. Vasudevan, M., Townsend, H., Wang, P., Samach, M., & Kuruvilla, E. (2022a). *DRAFT Artificial Intelligence (AI) for Intelligent Transportation Systems (ITS) Program: Market Research Report.* ITS Joint Program Office. TBD.

116. Walker, J. (2021). Artificial Intelligence (AI) for Intelligent Transportation Systems (ITS) Program: Emerging and Enabling Technologies Research Factsheet. USDOT ITS Joint Program Office. https://www.its.dot.gov/research_areas/emerging_tech/pdf/ITSJPO_AIforITS_Program.pdf

117. Wing, J. (2022, February 24). Trustworthy Artificial Intelligence—Center for *Connected and Automated Transportation (CCAT) Distinguished Lecture Series.* https://www.youtube.com/watch?v=cUf01ypQI9U

118. Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., Obrien, D. R., Steinke, T., & Vadhan, S. (2018). Differential privacy: A primer for a non-technical audience. *Vanderbilt Journal of Entertainment & Technology Law*, 21(1), 209–275. http://www.jetlaw.org/journal-archives/volume-21/volume-21-issue-1/differential-privacy-a-primer-for-a-non-technical-audience/

119. Work, D. (2022, June 2). *Improving ICM through AI*. ASCE International Conference on Transportation & Development, Seattle, Washington.

120. Xu, Y., & Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, 2(3), 249–262. https://doi.org/10.1007/s41664-018-0068-2

121. Yao, W., & Qian, S. (2020). Learning to Recommend Signal Plans under Incidents with Real-Time Traffic Prediction. *Transportation Research Record*, 2674(6), 45–59. https://doi.org/10.1177/0361198120917668

122. Yi, H., Bizon, C., Borland, D., Watson, M., Satusky, M., Rittmuller, R., Radwan, R., Srinivasan, R., & Krishnamurthy, A. (2021). AI Tool with Active Learning for Detection of Rural Roadside Safety Features. 2021 IEEE International Conference on Big Data (Big Data), 5317–5326. https://doi.org/10.1109/BigData52589.2021.9671360

123. Zhang, C., & van der Baan, M. (2021). Complete and representative training of neural networks: A generalization study using double noise injection and natural images. *GEOPHYSICS*, 86(3), V197–V206. https://doi.org/10.1190/geo2020-0193.1

124. Zmud, J., Walden, T., Ettelman, B., Higgins, L. L., Graber, J., Gilbert, R., & Hodges, D. (2021). *State of Knowledge and Practice for Using Automated License Plate Readers for Traffic Safety Purposes* (Report No. DOT HS 813 051). https://rosap.ntl.bts.gov/view/dot/55586

125. Zuo, F., Gao, J., Kurkcu, A., Yang, H., Ozbay, K., & Ma, Q. (2021). Reference-free video-to-real distance approximation-based urban social distancing analytics amid COVID-19 pandemic. *Journal of Transport & Health*, 21, 101032. https://doi.org/10.1016/j.jth.2021.101032

APPENDIX A: AI for ITS Sources Sought Notice Summary of Responses

The appendix provides an overview of the AI for ITS Sources Sought Notice (SSN), summarizes the SSN review methodology, and summarizes responses to the SSN.

A.1 Overview of Sources Sought Notice

The purpose of the Sources Sought Notice (SSN) was to solicit feedback from public sector agencies, industry, research laboratories, academia, and other stakeholders on “deployment-ready” applications that leverage AI to address ITS needs, specifically to improve the transportation system and users’ safety, mobility, equity, accessibility, productivity, efficiency, and environmental impacts. Deployment-ready AI-enabled ITS applications are those that have been successfully prototyped and validated to address specific ITS challenges and are sufficiently mature for integration into existing ITS operations within six to nine months. This includes data acquisition, processing, re-training, testing, and validating the application, and integrating with existing ITS. However, USDOT recognizes that ongoing development, test, validation, and maintenance, are necessary after the application is initially deployed in the field to accommodate new and/or changing data sources, prevent performance degradation, ensure continued security and privacy. The DOT sought responses to a series of questions (listed in the SSN Appendix), to help shape potential investments towards pilot deployments of ITS applications that utilize AI.

The AI for ITS SSN was issued on 30 July 2021 and closed on 10 September 2021. On 12 August 2021, the ITS JPO AI for ITS Team held a webinar on the SSN. The deadline for questions on the SSN was 16 August 2021. The questions and answers on the SSN were posted on 23 August 2021. All SSN documents, including the SSN, the Appendix with questions, and the questions and answers can be found at SAM.gov.

Although AI can power applications throughout the entire transportation system, the ITS-JPO is interested in existing capabilities in developing and deploying AI-enabled ITS applications that fall under the seven categories shown in Figure A-1 - Transportation Systems Management and Operations (TSMO), Asset Management and Roadway Construction and Maintenance, Commercial Vehicle and Freight Operations, Transit Operations and Management, Accessible Transportation, Emergency Management, and Traveler Decision Support Tools.

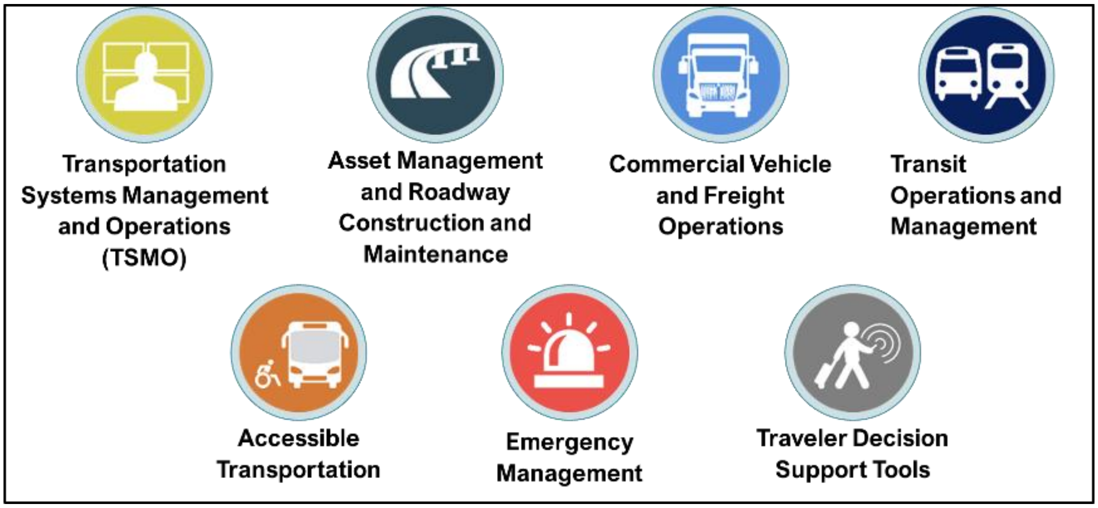
Transportation Systems Management and Operations (TSMO)
Asset Management and Roadway Construction and Maintenance
Commercial Vehicle and Freight Operations
Transit Operations and Management
Accessible Transportation
Emergency Management
Traveler Decision Support Tools

Figure A-1 AI for ITS Application Categories

The potential for AI in ITS is both broad and substantial in these seven high-value categories ranging from ensuring safety, improving situational awareness and systems management, optimizing fleet operations, improving accessibility, and maximizing infrastructure investments through proactive asset management. AI-enabled ITS applications within these seven categories can be used to improve safety, mobility, accessibility, productivity, and efficiency and reduce climate change impacts. The SSN included a preliminary list of 31 “deployment-ready” AI for ITS applications that fall within the seven categories, recognizing the applications may not be comprehensive.

The “AI for ITS Sources Sought Notice Appendix: Questions for Respondents” included nine questions that sought to gain insights from operational testing and early deployments of AI for ITS as well as information on existing technical capabilities of public, private, and academic sectors in developing and deploying AI-enabled ITS applications within the seven categories. ITS JPO also sought to gather similar information and insights on additional suitably mature applications not covered in the preliminary list of 31 “deployment-ready” AI for ITS applications in the SSN, that leverage the data collection, processing, and analysis potential of AI to enable safer, more equitable, more efficient, and more reliable surface transportation system planning, operation, and maintenance, particularly among public-sector transportation agencies.

The first question in the SSN Appendix asks respondents whether they agree with the characterization of the maturity of the 31 AI-enabled applications listed in the SSN. The second question asks if the respondent is aware of deployment-ready AI-enabled ITS applications that are not covered in the SSN. The third question asks if the respondent has leveraged AI techniques for data collection, processing, or analysis to address ITS needs. The third question includes 14 sub questions asking for more detail about the problem addressed, deployment in the field, the application concept, the main users/beneficiaries, the system or application interpretability, transparency and ethics considerations, collaboration, data requirements, cybersecurity, benefits and evaluation metrics, cost estimates, challenges and lessons learned, scaling, and ongoing operations and maintenance. The fourth question asks if the respondent is aware of proven AI-enabled applications from other domains that could be rapidly adapted and integrated within the ITS ecosystem. The fifth question asks about interoperability across vendors and locations. The sixth question asks whether AI operations in transportation are trustworthy and ethical. The seventh question asks about which ITS challenges might benefit most from targeted investments. The eighth question asks respondents to identify the top three roles for USDOT to support agencies in leveraging AI. Finally, the ninth question asks about the future of AI for ITS.

A.2 Summary of Responses

Out of the total number of SSN responses received, 50% were from small businesses, 38% were from large businesses, 8% from academic centers, and 4% from federally funded research and development centers (FFRDC). Among these, 13% were identified as minority-owned businesses. The most common domain by far is technology (67%), with transportation (25%) as a distant second, followed by energy (4%).

A.2.1 Summary of Responses to Each Question in the SSN

Question 1: Do you generally agree with the characterization of the maturity of the AI-enabled applications listed in Table 2 of the SSN? To support the reasoning behind your answer, can you provide references and/or evidence?

Response: Majority of the respondents generally agreed with the characterization of the maturity of the AI-enabled applications in Table 2 of the SSN.

Applications that were noted to be of lower maturity than what was indicated in Table 2 of the SSN were Proactive Incident Management and AI for Multimodal Trip Planning that accounted for emissions and traveler preferences. One respondent noted that AI for data fusion in the TMC, AI for Work Zone Safety and Information Dissemination, Comprehensive traffic modeling using prediction, AI for Weather Prediction and Response, AI for road weather management, and Proactive Incident Management specifically for transit event logging were of higher maturity than what was indicated in Table 2. These appear to be ones that the entity had developed in their lab. Reviewers found the justifications provided by the respondents for the deviations in maturity characterizations to be reasonable.

One of the respondents mentioned that even though majority of the applications may be technologically mature, institutional barriers (e.g., existing agency practices, cross-agency collaboration, public-private collaboration) may prevent or hinder their adoption.

Other barriers include lack of ground truth data for validation, data processing and computing bottlenecks, interoperability, and independent validation of tool scalability. Another respondent indicated that some of the applications listed in Table 2 needed to "go through a rigorous evaluation process in terms of accuracy and effectiveness" prior to deployment in the field.

Question 2: Are you aware of deployment-ready AI-enabled ITS applications that are not covered in Table 2 of the SSN? If so, please provide a summary of the application concept, its categorization according to Table 1 of the SSN, and supporting references and/or evidence.

Response: Nearly half of the respondents provided AI-enabled ITS applications that were deployment-ready but not covered in Table 2 of the SSN. Based on the information provided, reviewers were generally in agreement with the respondents' assessments.

Most of the applications fell under the TSMO category. Examples of applications that fell under the TSMO category include AI-enabled applications for: data fusion; multi-modal people movement analytics; congestion causality analytics; traffic predictions; dynamic pricing; vehicle re-identification; real-time driver assistance using on-board barcode reader and edge computing; transportation planning and disaster recovery using large volumes of GIS data; Backoffice automation; predictive energy analytics; environmental sustainability (e.g., air quality-based vehicle routing); customer support (e.g., Conversational AI, agent decision support, agent training); and trespass detection and prediction along railroad right-of-way.

Examples of applications that fell under Commercial Vehicle and Freight Operations category include port analytics; automated vehicle inspections; driver safety scoring; and container expected time of arrival.

Examples of applications that fell under Traveler Decision Support Tools category include AI-enabled applications for on-demand transportation services (e.g., macro-transit); occupancy detection and notification; and real-time parking availability, location, and fares.

Examples of applications that fell under Transit category include AI-enabled applications for Transit Event Logging System (TELS), and on-demand transportation services.

An example of an application that fell under Remote Sensing category includes the use of AI and computer vision to analyze satellite, balloon, and mobile geolocation data, to study a range of human activities to provide strategic insights.

Question 3a: Have you leveraged AI techniques for data collection, processing, integration, or analysis to address ITS needs? What was the problem that was addressed (e.g., crash / risk identification; recurrent and non-recurrent congestion)?

Response: Most respondents have leveraged AI techniques to address ITS needs. Most AI-enabled applications proposed by the respondents addressed the following problems: crash prevention and risk identification, recurrent and non-recurrent congestion, traffic prediction and operations, emergency management, and large-scale asset management. Two applications are freight-related applications focusing on port cargo prediction and freight route optimization. One application focuses on flight delay prediction. Note that most applications proposed by the respondents could help address problems in the Administration and USDOT priority areas as shown in Figure A-2, with Safety and Infrastructure as the two most often cited by reviewers. Fewer reviewers cited applications as being able to help address challenges in Equity, which could indicate a gap.

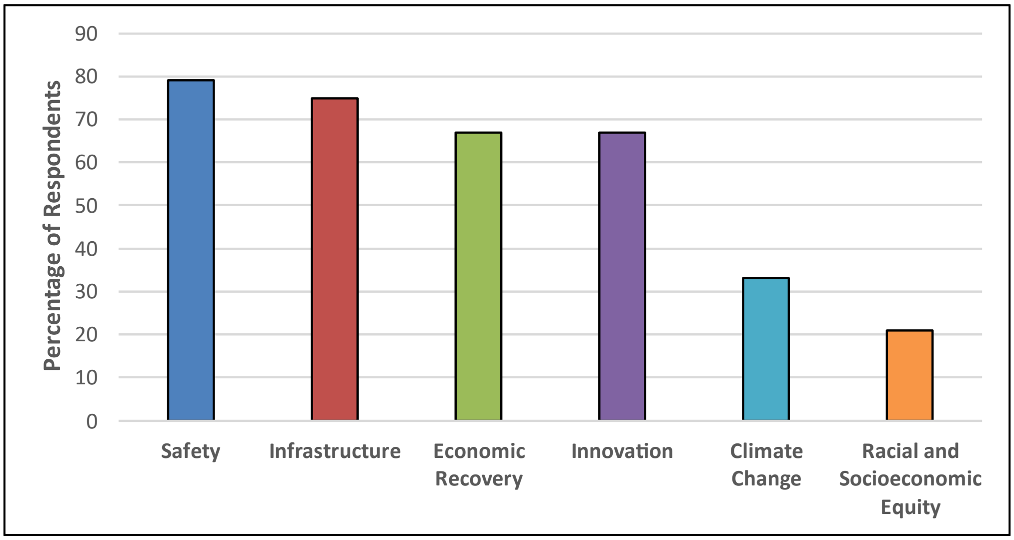

Figure A-2 Percentage of Respondents Addressing Problems in Administration/USDOT Priority Areas

Question 3b: Has this been implemented or deployed in the field? If so, where? Is the implementation/deployment still ongoing?

Response: According to the rating of the technology maturity of the proposed AI-enabled applications, most entities proposed applications that have been deployed in the field. Some entities proposed applications that have been deployed in more than one location and these applications have a TRL of 7 or higher. For more detailed information on the application maturity, please go to Section A.2.2.

Question 3c: Please provide a summary of the AI-enabled application concept. Did you make use of edge computing? Please also include its categorization according to Table 1 of the SSN and the AI techniques (e.g., machine learning, natural language processing, object recognition) that were applied.

Response: More than half of the respondents made use of edge computing in their AI-enabled applications. They leveraged AI techniques such as, deep learning, neural networks, reinforcement learning, computer vision, convolutional neural network (CNN), etc., to address ITS needs. Figure A-3 plots the number of applications in each application category. Most applications proposed by the respondents fall into the TSMO category. Note that one application may fall into multiple application categories. For example, road weather prediction may fall under both TSMO and Traveler Decision Support Tools since the respondent's application can help a TMC operator manage the transportation system as well as provide travelers with situational awareness of the road weather conditions.

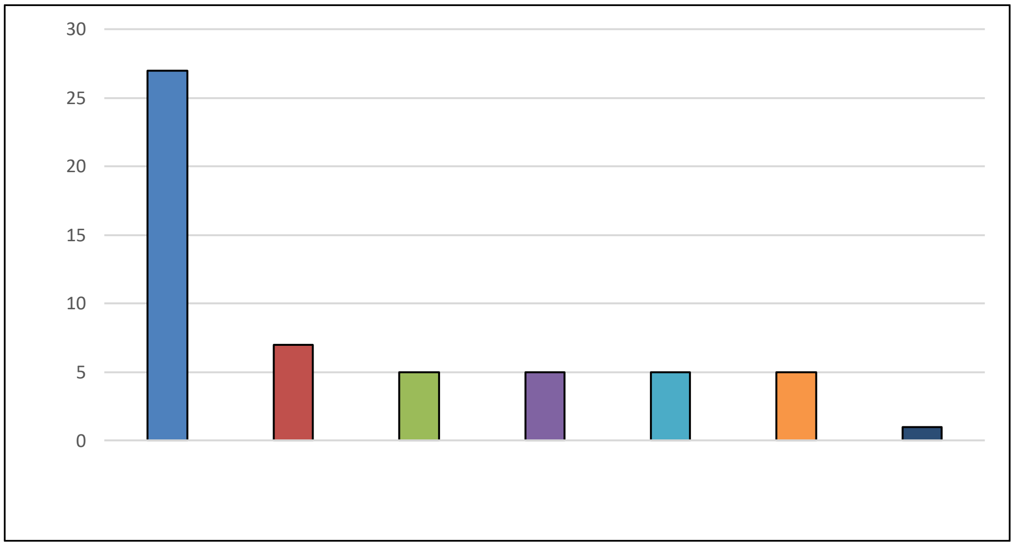

Figure A-3 Number of AI-Enabled Applications by Category

Question 3d: Who were the primary users and beneficiaries of this AI-enabled application (e.g., Infrastructure Owner Operators, travelers, TMC staff, etc.)?

Response: The primary users and beneficiaries of the AI-enabled applications proposed by the respondents are the Infrastructure Owner Operators (IOO), followed by travelers, pedestrians, and micro-mobility users (see Figure A-4). Most AI-enabled applications developed by the respondents fall into the TSMO category. Therefore, most respondents said they collaborate with the public sectors and the primary users and beneficiaries are IOO.

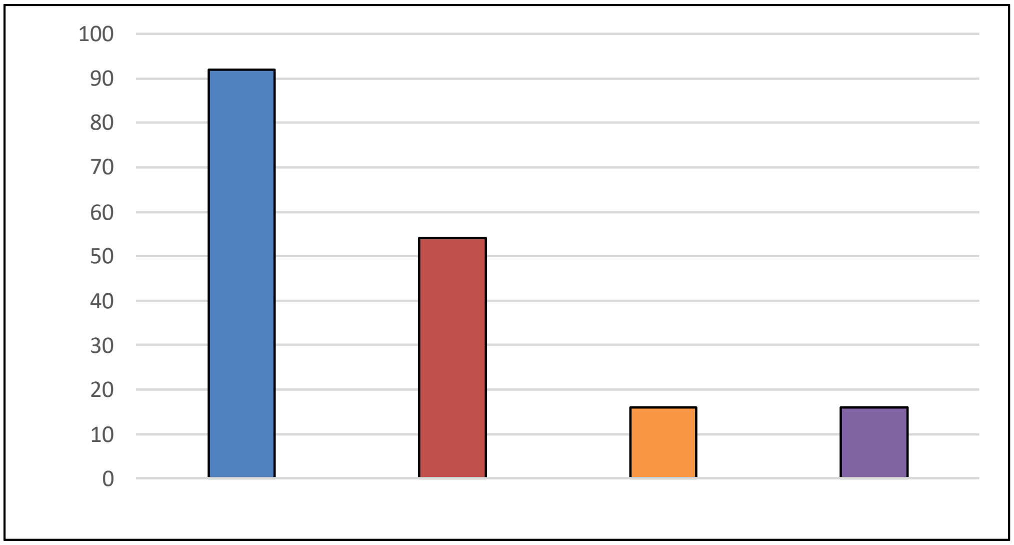

Figure A-4 Primary Users and Beneficiaries of AI-Enabled Applications

Question 3e: Are decisions made by the AI-enabled application easily interpretable by the primary users and beneficiaries?

Response: Most of the entities failed to provide an adequate response on the interpretability of their AI-enabled applications. Those who did, mentioned that visual display (e.g., through dashboards, satellite images, Geographic Information System [GIS] maps) of results enables easy interpretation. It was noted that some applications are simple to understand because the outputs are traffic engineering performance measures. Some entities touched on the interpretability of the reasoning behind the decisions. It was mentioned that prior to system deployment, all impacts are explained and agreed to by the traffic partner agencies and transit operators. An initial training was conducted for the users of an application to allow for independent operation and analysis of results. It was noted that an AI technique (Decision Trees) is "naturally explainable" and that in addition uses "plain language text to explain the AI-based decisions." It appears that few entities have a clear strategy in place for developing interpretable AI models.

Question 3f: How transparent is the system to the public? Is there a strategy in place for cultivating trust in the AI-enabled application? What legal and ethical considerations have been made regarding system transparency?

Response: Majority of the respondents demonstrated ability to make their systems transparent, while taking into consideration the legal and ethical implications. However, they do not appear to have a systematic approach for cultivating trust among stakeholders.

Some mechanisms that the respondents used for making their systems transparent, while protecting privacy include: sharing sensitive information only with agencies and not with the public; removing PII from data and not sharing data with unauthorized users; not using or recording any satellite imagery data that reveal identities of vehicles and pedestrians; blurring images to preserve privacy; securely archiving video footage; immediately reducing video imaging data to positional information and discarding raw data where personal or vehicle identification can be done; receiving consent from drivers prior to using their information; and informing the public about the technology being used and why. One respondent noted that they had an ethics plan.

Question 3g: Did you collaborate with any organizations or agencies for developing and/or implementing the AI-enabled application? If so, who (e.g., academia, private sector) did you collaborate with? What type of collaboration did you have (e.g., signed memorandum of agreement to share resources, informal collaboration)?

Response: Most respondents collaborated with public sector agencies for developing and implementing AI-enabled applications due to the funding and deployment effort (see Figure A-5). Private sectors are the second frequently mentioned sectors for product development effort, followed by academia for research purpose. Most AI-enabled applications developed by the respondents fall into the TSMO category. Therefore, similar to Question 3d, most respondents said they collaborate with the public sectors.

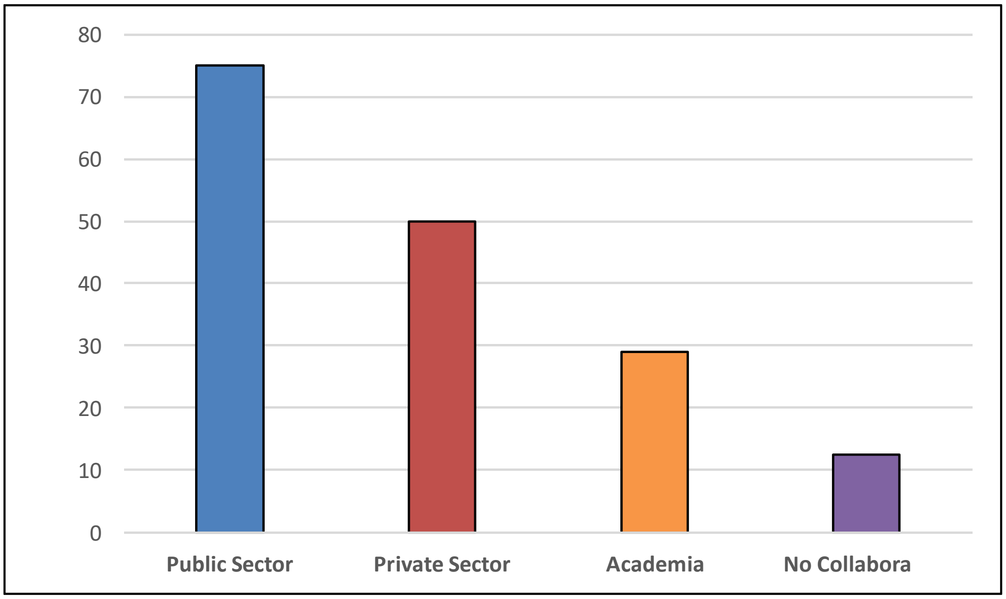

Figure A-5 Sectors Frequently Collaborated with for Developing AI-Enabled ITS Applications

Question 3h: What are the data requirements (i.e., type, quantity, quality, frequency, and latency of data) for the application? Are the data readily accessible to you? Did you use or create any proprietary data? Did you have an established data sharing agreement with other players? Do sufficient, affordable data sources exist to effectively operate AI for ITS? Did you assess for data imbalance? Did you check for possible bias in the data? Did you need labeled data?

Response: Majority of the respondents used proprietary data, which were either purchased from agencies, for whom the applications were being developed, or provided by the agencies at no cost through data sharing agreements. Some respondents used publicly available data (e.g., state DOT data, General Transit Feed Specification [GTFS] data) or data from simulation models. Given below are some of the key takeaways:

- Public Domain Data: Public domain data (specifically, imagery data) may suffer from a lack of quality control. One of the respondents noted that public domain imagery data are poorly annotated and suffer from data inconsistencies, and so they had to use proprietary data.

- Labeled Data: Automated tools/methods can be used to augment hand-labeling for training data. One of the respondents was able to increase the number of images by ten-fold by using a combination of hand-labeling and automated tools/methods, compared to just hand-labeling.

- Data Accuracy/Completeness: Street and road data, and public transit network data are widely available with different levels of completeness, accuracy, and granularity, but enriched data come at a cost. One of the respondents noted that enriched data, such as driving directions come at cost; According to the respondent, the hardest data to acquire are non-geographical datasets like fares, emissions, or public preferences.

- Data Latency: Use of on-demand data can increase latency if the data are coming from a data provider. One of the respondents noted that if on-demand data are used, it can increase the lag time for processing results since the data has to come from the data provider and is not situated in the developer's servers (on-prem or cloud)

- Bias/False Positives: Rigorous and thoughtful feature engineering is crucial for avoiding false positives, especially in near-incident events. One of the respondents extracted more than 25 features from computer vision metadata, and subsequent projects also benefited from this process, which is a bonus for most ML-related efforts. Another respondent noted that “balancing the data against bias and noise is a continuous process,” needing “regular, independent, and automated tests.”

Question 3i: Have you conducted security analysis of your AI-enabled ITS application (e.g., Confidentiality, Integrity, Availability (CIA))? Are there any medium/high security risks associated with your application? If yes, what are these risks? What Cybersecurity concerns must be addressed for an operational AI ITS system? What Cybersecurity concerns must be addressed for ensuring personal security?

Response: Majority of the respondents either did not respond to the question specifically or indicated that they have not conducted a security analysis of their AI-enabled ITS application. Those who responded, mentioned a few best practices that their organizations follow. These include: (i) minimizing data generation, transmission, and access by different people, (ii) transmission of only anonymized and encrypted data to the cloud; (iii) preserving privacy by avoiding face recognition technologies or blurring faces; (iv) processing data on the edge and confirming security patching on edge devices; (v) using internal Wide Area Network for the ITS infrastructure; (vi) frequent penetration testing; (vii) following Open Web Application Security Project (OWASP) development standard or security best practices of commercial services (e.g., Google Cloud, Amazon Web Services; and (viii) using multi-layered cybersecurity (e.g., VPNs, firewalls, encryption, etc.).

Some of the respondents indicated that they recognize cybersecurity to be a top concern but did not discuss the specific processes and procedures in place.

Loss of PII was identified as the main cybersecurity concern. Data poisoning was also mentioned as a threat. Data poisoning is the act of tampering with the input data to influence the output. system.

Question 3j: Have you measured the benefits? If yes, what improvements did you see? What evaluation metrics were used and why? Do sufficient, affordable data exist for measuring the performance of your AI-enabled system? Did leveraging AI techniques help address the problem? How did you benchmark the performance of your system?

Response: Nearly half of the respondents have evaluated their systems to measure benefits, including receiving qualitative feedback from clients (e.g., AI solution is faster, more accurate, or less expensive).

Some benefits reported by the respondents include, improved application performance, decreased congestion, improved speed and accuracy of incident detection, reduced wrong-way related incidents, improved detection of lane violations, improved bus speed and on-time performance, reduced hard braking by trucks due to queue warning at work zones, increased reliability and reduced breakdowns of fleet vehicles, increased profit and time savings to freight agencies, improved access to systems, improved response time and reduced human error due to automated actions, rapid pavement assessment (50% reduction in parts and labor cost), and consolidated workflow.

The respondents reported the use of the following evaluation metrics: precision, recall, F1 score, receiver operating characteristic (ROC), area under the curve (AUC), and mean average percentage error (MAPE). One respondent noted that in addition to industry performance metrics for ML models, metrics specific to the function of the system should also be considered. However, no further information on examples of such metrics was provided.

Most of the respondents either did not respond to the question or indicated that they had not evaluated their systems to measure the benefits. A few of the respondents have tested their systems, prior to deployment, and noted improved accuracy.

Question 3k: What are the rough order-of-magnitude cost estimates for the AI-enabled application (including the cost of developing and running the AI functionality as well as cross-cutting costs such as data, sensors, hardware, etc.)?

Response: A rough order of magnitude cost estimates were provided for respondent applications, which are summarized below:

- For incident detection, software development and deployment cost for an entire state-wide sensor network is approximately $5M. The cost of sensor deployment is dependent on the coverage area and number of sensors.
- For wrong-way detections, the cost is $1M per year.
- For traffic volume estimation at a single intersection, the cost is approximately $3M-$5M for development, testing, hardware procurement and installation, and resources to monitor, update, and report on system performance for five years.
- For vehicle tracking, non-recurring costs are approximately $1.2M; each intersection will cost an additional $25,000.
- For object detection/geolocation, which is designed to run in a cloud environment, the monthly compute costs start at $10,000/month and grow based on data storage and compute requests.
- For pavement assessment, the cost is approximately $100/mile, but there are other client specific costs, which were not reported.
- For predictive asset maintenance, the ongoing cost is between $20 and $50 per vehicle per month, with a one-time $400 hardware setup fee.

- For machine vision traffic monitoring, software development cost is in the order of $2M-3M. Hardware and deployment costs are dependent on the density of the camera network and the connectivity.
- For port cargo predictions, the minimum cost is $1.2M.
- For freight route optimization, there is a one-time cost of $300,000 to $500,000, and an annual cost of $50,000.

A majority of respondents either provided inadequate (e.g., $0 to $10M) or no response.

Question 3l: What challenges or issues (e.g., institutional, legal, technical, operational, etc.) did you face while implementing the AI-enabled application? What are key lessons learned?

Response: The respondents noted a series of technical, and institutional challenges, including:

- Heterogeneity of data format, frequency, and quality
- Inadequate environment and lighting conditions, for computer vision-related applications
- Data availability, data sharing, communications, and cooperation between agencies and jurisdictions
- Lack of ground truth data
- Real-time processing of large amounts of data from edge devices
- Integrating the various layers of technology for the first time
- Software and hardware incompatibilities
- Keeping system up to date
- Vetting root cause assumptions and identifying true problems (as the client may not know)
- Selection of relevant features for building ML models (helps to leverage subject matter experts)
- Communication between technical and non-technical groups
- Lack of best practices

- Insufficient information on benefits of applications
- Reluctance among technicians to fully adopt any new technology; difficulty getting buy-in from prospective users
- Skepticism among customers when AI model outperformed expectations, due to lack of familiarity with AI capabilities
- Difficulty with developing contract vehicles for local governments and developing tools for larger organizations
- Lack of funding

Key lessons learned, as reported by the respondents, include the following:

- Security, privacy, fairness, and the associated legal framework to codify and represent those aspects must be central to AI-related projects.
- Ensure metadata is available, and confirm timestamp (intervals, frame rate) on data from multiple sources are identical before data fusion.Re-configure parameters after installation.
- Communications technologies, edge computing infrastructure, virtualization technologies, and sensors need to be integrated and require automation capabilities to respond to demand in real time.
- Considering the existing IT challenges states/localities face now, moving to more complex and distributed approaches will be a challenge technically and institutionally.
- Demonstrate benefits to encourage adoption of any new technology.
- Develop best practice training materials prior to deployment, based on experiences gained throughout the process.
- Train operators/users on the basics of AI systems, such as the need for consistent data formats.
- Educate clients about details of the AI pipeline, including data preparation, modeling, and quality assurance, to help build buy in.
- Conduct research and gain an understanding of client data and operations to find the real problem(s) before diving in. Work with domain experts to ensure the technical solution solves the business problems.
- Focus on continuously improving the AI models.

- Focus on areas of strengths or expertise (e.g., develop ultra-efficient AI software at the edge), and leverage deployment partners to scale the solutions.

Question 3m: What challenges or issues do you foresee if the application were to be applied at a larger scale?

Response: Key challenges to deploying AI-enabled solutions at scale, as noted by the respondents, include:

- Logistical maintenance of the hardware and software
- Lack of sufficient computational resources, such as GPU and memory
- Ability to handle the vast amounts of data ingested when scaled
- Maintaining data integrity, as unforeseen data issues and irregularities may be introduced at scale, requiring enhancements to data quality control processes
- Cybersecurity
- Deploying sufficient cameras, to match road network density, for computer-vision related applications
- Interoperability and integration, as standards frameworks are not mature enough
- System adoption, as there may be policy changes within the client organization
- Unifying the knowledge learned across multiple deployments while maintaining cybersecurity and privacy
- Limited professional and agency capacity to implement and monitor AI-based systems
- Budget allocation and federal grant availability, for widespread deployment of advanced safety systems, as the need greatly exceeds available funding
- Lack of awareness, especially among municipalities, of available funds or methods to access them

Question 3n: How is the AI-enabled application maintained? Can it incorporate new data? Can it respond to new conditions?

Response: Majority of the respondents indicated that they periodically updated their hardware and software, including incorporating new data. Given below are some best practices noted by the respondents for AI-enabled system maintenance:

- Regular monitoring of model performance, especially during deployments; significant deviations possibly due to new data/conditions can lead to model degradation
- Constant recording of new data (e.g., vehicle data, signal data) for continuous re-training and refinement of models and improved performance
- Continuous monitoring of data to maintain sufficient quality of data
- Periodic calibration and validation of sensor data, camera views, and other data
- Periodic back up of database
- Calibrate application to each site (even if previously deployed elsewhere)
- Consider built-in health monitoring feature to track sensor's condition for easy maintenance
- Continuous refinement of AI models (e.g., using regularization techniques for model generalization)

Question 4: Are you aware of proven AI-enabled applications from other domains that can be rapidly adapted and integrated within the ITS ecosystem? If so, can the AI-enabled application, from another domain, be deployed within the ITS ecosystem in 12 months?

Response: According to the respondents, examples of AI-enabled applications from other domains that can be rapidly adapted and integrated within the ITS ecosystem, include:

- Computer vision techniques from the medical domain (e.g., medical imaging)

- AI/ML-based packet routing in the communications domain could be applied to a transportation network for re-routing traffic due to non-recurrent congestion
- Reinforcement learning for optimal control could be applied to traffic signal control.
- Data processing and modeling technologies for energy efficiency could conceptually be used for vehicle routing
- Fraudulent transaction detection could be applied wherever payments are accepted (e.g., train, bus, or tollway)
- Risk management applications and social engagement applications
- Application for extracting, transforming, and loading acquisition data, could be useful for state and local agencies

A few respondents noted that while there are proven applications in other domains, it is difficult to estimate whether these applications can be rapidly deployed in the ITS ecosystem within 12 months due to the complexity of implementation. One respondent noted that while AI image processing systems deployed in autonomous automobiles, smart systems in space vehicle design (e.g., communication, navigation), and AI systems integrated into jet aircraft for safety can all be integrated into the ITS ecosystem, the feasibility of rapidly deploying these applications for ITS is unknown since the data or sensor requirements are unknown.

Question 5: Are multi-vendor AI for ITS solutions for the same or different transportation operations interoperable within a single location or across locations regionally or nationally?

Response: Majority of the respondents did not provide a clear response to the question. A few seemed to confuse interoperability with interpretability. Others discussed interoperability more generally, rather than for their application. One respondent noted that for interoperability across vendors, common standard for data formats (input and output) is needed. For interoperability across locations, the interoperability will also depend on the system design and software architecture to enable transferability to other regions. This requires significant calibration of the application to each site. Another respondent noted that standards frameworks and protocols are not yet mature enough to allow seamless integration of applications between multiple vendors. The standards and policies are not yet mature enough for allowing interoperability across locations.

A few respondents specifically mentioned that their systems were interoperable across locations. Methods used by these respondents include the use of open architecture for maximum flexibility; use of a containerized solution; use of industry-wide standard formats; and use of open platforms with open data formats.

Question 6: Are AI operations in transportation trustworthy and ethical? How are you determining them to be trustworthy and ethical?

Response: Majority of the respondents recognize the need for building trustworthy and ethical AI systems. As one respondent noted, for gaining trust, it is critical for the AI system to consistently produce output that system's operators consider reasonable. Conversely, a single error could "foul that trust for a long time." For ethical operation of AI, the respondent noted that the AI decisions should minimize bias and be fair, transparent, responsible, and interpretable. Example best practices employed by respondents for trustworthy and ethical AI include:

- Operate within a data governance structure that is ethical, and constantly test for bias
- Conduct regular human verification for quality control
- Provide a mechanism to request proof upon demand (e.g., a picture that was used to make an AI decision so a person can confirm upon request)
- Ensure the system performs well for all people regardless of gender, ethnicity, etc. (minimize bias and act without prejudice)
- Test system in a controlled environment to gain system operators' trust
- Ensure that the traffic partner agencies remain owners of the data to build trust
- Ensure results are auditable and explainable, and are limited to actions for only their designed purpose
- Build systems that adhere to federal, state, and local requirements, policies, standards, regulations, and laws, and check for compliance
- Constantly monitor ethical standards as these do not remain fixed and transform in response to evolving situations (e.g., what was acceptable 10 years ago, may no longer be considered ethical)
- Use a committee of experts or the organization's ethics review board to evaluate work, including bias assessment and mitigation, model interpretability, governance, and model lineage

Question 7: What ITS challenges might benefit most from targeted investments by DOT through pilot deployments of applications leveraging AI?

Response: According to the respondents, challenge areas that would benefit the most from targeted DOT investments include:

- Safety: Reduce fatalities; Stop-collision prevention; emergency vehicle preemption; roadway worker and fleet death reductions; early incident detection and confirmation; automated incident response plan generation and implementation; safety-affirmative intersection messaging; and pedestrian detection
- Mobility: Reduce congestion; improve traffic flow; removal of queue buildups; promote high occupancy vehicle (HOV) lane usage; enforce HOV and Express Lane usage; transit signal priority; freight mobility (optimize load and truck movements); real-time signal optimization; traffic signal coordination; and comprehensive real-time traffic modeling and prediction
- Accessibility: Optimize curb usage; and maximize mixed-use (cars, trucks, pedestrians, cyclists, buses, and ridesharing operators) accessibility of the curb at busy intersections
- Environment: Reduce pollution
- Efficiency: Handle large volumes of AV, CV, and CAV data to support infrastructure and TSMO operations; explore alternative data sources; and data fusion from multiple real-time data sources
- Infrastructure: Predictive asset management; standardized assessment of pavement and other assets
- Funding: Direct award of funds to municipalities' emergency services

Question 8: What, in your opinion, are the top three roles for DOT to support agencies in leveraging AI for safer, equitable, more accessible, and more efficient operations and management of multimodal transportation systems? Please select from the choices below and provide reasoning.

a. Focus predominantly on mature (i.e., having a maturity rating of 6 and higher on the Technology Readiness Level (TRL) scale1) AI-enabled applications and how they can support ITS.

b. Develop labeled data and other resources.

c. Conduct advanced research and testing of AI-enabled applications.

d. Conduct prototype testing/demonstrations of AI-enabled applications.

e. Develop standards to ensure that data can be easily accessed and shared for execution of AI-enabled ITS applications.

f. Resolve AI-related policy issues (e.g., data governance and data sharing policies).

g. Evaluate AI-enabled ITS deployments.

h. Other.

Response: According to the respondents, the **top three** USDOT roles should be to *resolve AI-related policy issues*, *develop standards* to ensure that data can be easily accessed and shared for execution of AI-enabled ITS applications, and *conduct prototype testing/demonstrations* of AI-enabled applications. Focusing on predominantly mature applications and developing labeled data and other resources were other top roles selected by the respondents. One of the respondents noted that a critical roadblock to successful AI deployments is the lack of guidelines for data sharing and governance, which cannot be resolved by individual agencies and companies. Entities tend not to save data, let alone share it. Another respondent noted that availability of labeled data will result in advanced AI-enabled ITS applications by “attracting researchers from other domains to conduct research and development of ITS applications,” and “giving researchers a common set of data to benchmark their AI algorithms.” A respondent noted that agencies are less likely to deploy AI-enabled solutions if the technology has not been tested by peers. According to the respondent, the “USDOT can accelerate the development of AI for ITS solutions if it can effectively establish the ground rules for ownership and exchange of ITS collected data, set the protocol for the secure exchange of data, provide education and assurances to agencies about cyber security risks, and fund AI for ITS demonstration projects that agencies can reference.”

Question 9: Where do you see AI headed in ITS and why? What challenges do you foresee?

Response: The respondents see a wide range of possibilities for AI in ITS in the future. These include:

- more autonomous, electric, and shared vehicle services (e.g., “moving a package from an autonomous aircraft via a robot to an autonomous delivery van to an unpiloted drone to the doorstep without any human interaction”)

- incorporation of physical dynamics into the AI models (e.g., physics informed neural networks (PINN))

- introduction of the digital twin concept to ITS

- application of deep reinforcement learning in traffic modeling and simulation

- limiting or removing the need for ML for real-time decisions, significantly improving ML, exploring generative models to allow ITS to have a "digital imagination" to allow "autonomous systems to not only adapt to a dynamic environment, but to anticipate a future unforeseen (and therefore untrained environmental factors) far beforehand"

- "a flexible enterprise AI platform that fits well into a government auditable approach"

According to the respondents, key challenges to the future of AI in ITS include:

- defining what is good in terms of metrics and outcomes

- lack of resources, including labeled and clean data, computation power, hardware

- lack of capability to integrate multiple platforms and fuse data from multiple sources with varying formats

- inability to create labeled data

- lack of ground rules for ownership and secure exchange of data

- lack of technological interoperability

- lack of standards to evaluate trustworthiness of AI applications

- deploying pilots in a disjointed or siloed manner, requiring expensive and rare talent to develop and maintain inhouse systems

- lack of workforce and expertise in AI, specifically in the public sector, to understand strengths, weaknesses, and risks (including cyber security risks), and recognize unrealistic vendor claims

- skeptical or apathetic attitude towards change and new technologies, requiring "technology stewards" to champion the new systems and provide support throughout the process of adoption

- lack of federal funding to support agencies for development and adoption

- potential proliferation of unethical AI systems

A.2.2 Summary of Applications Technology Readiness Level

Based on the information the respondents provided for their AI-enabled applications in response to Question 3, the reviewers rated the application TRL based on the FHWA definition (FHWA, 2017). Reviewers independently came to their assessment of the TRL of a respondent's application through a review of the evidence presented in the response. For example, if an application was successfully deployed in an operational environment and proven to be beneficial, the application was rated as "8 – proven in operational environments." If the application had been used in multiple locations with some customized features to meet the needs of different locations, it was rated as "9 – refined and adopted." It should be noted that a TRL of 8, "proven in operational environment" does not necessarily mean that the AI application performs better than traditional solutions, just that the service/product has been shown to work in an operational environment. Figure A-6 shows the number of the applications, provided by the respondents, by TRL. There is a total of 38 applications identified by the reviewers. Out of the 38 applications, four applications are at the prototype demonstration level, 12 applications have been demonstrated in operational environments, and 16 applications have been deployed and proven in operational environments. None of the applications were identified to be at the refined and adopted level (i.e., TRL of 9).

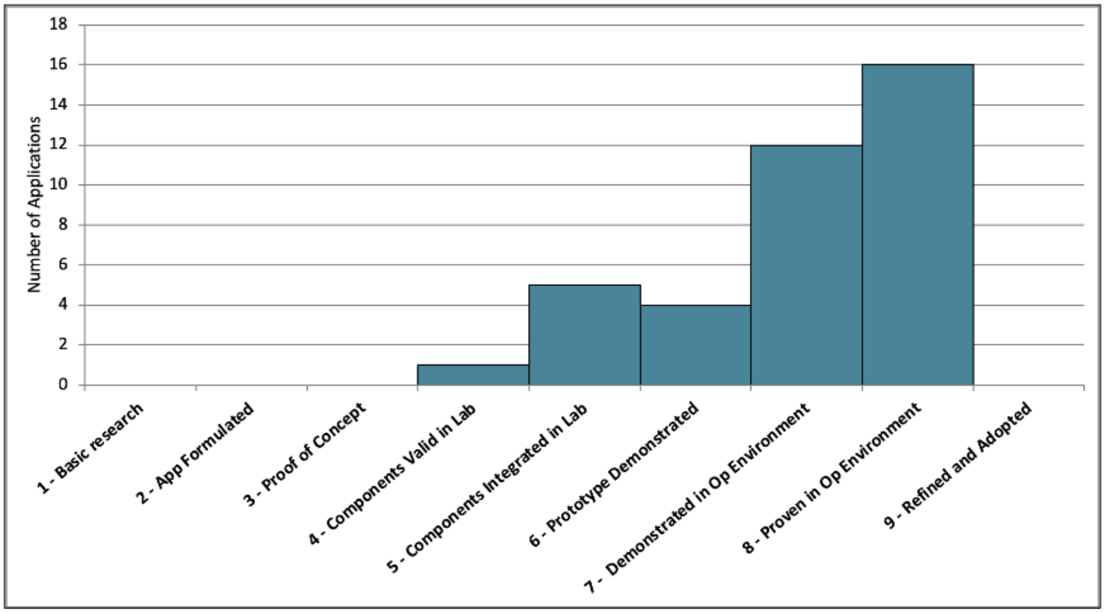

Figure A-6 Distribution of TRLs of Respondent Applications

A.2.3 Key Findings from Review of Responses

Key findings from the SSN responses review are summarized below:

- USDOT received a fair number of responses. The respondents provided a total of 38 AI-enabled applications. In some instances, only marketing materials were provided, which did not show any specific capabilities or previous project examples. Both large business entities and small business entities demonstrate the ability to deploy AI-enabled application for ITS.

- The respondents confirmed that USDOT accurately captured the characterization of the maturity of AI-enabled applications from the previous research.

- Most applications (32 out of 38) developed by the respondents are mature: prototype demonstrated (4), demonstrated in operational environments (12) and deployed and proven in operational environments (16).

- Most AI-enabled applications developed by the respondents fall into the TSMO category. Therefore, most respondents said they collaborate with the public sectors and the primary users and beneficiaries are IOOs.

- Many respondents made use of a variety open-source tools and packages for ML model development.

- A few respondents indicated that they did not conduct security analysis for their AI-enabled applications, but they did follow their organization's best practices, such as minimizing data generation, transmission, and access by different people, and ensuring transmission of only anonymized and encrypted data to the cloud.

- The respondents indicated that the top three USDOT roles are **resolve AI-related policy issues** (e.g., data governance and data sharing policies, ethical AI), **develop standards** to ensure that data can be easily accessed and shared for execution of AI-enabled ITS applications, and **conduct prototype testing/demonstrations** of AI-enabled applications.

U.S. Department of Transportation
ITS Joint Program Office-HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free "Help Line" 866-367-7487
www.its.dot.gov

U.S. Department of Transportation