Policy Analysis and Guidance to Support Secure Transportation Cyber-Physical Systems

Final Report

by

Trayce Hockstad
University of Alabama

Mizanur Rahman, University of Alabama
Steven Jones, University of Alabama
Latifur Khan, University of Texas at Dallas
Mashrur Chowdhury, Clemson University
M Sabbir Salek, Clemson University

May 2025

NATIONAL CENTER FOR TRANSPORTATION CYBERSECURITY AND RESILIENCY (TraCR)

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by the National Center for Transportation Cybersecurity and Resiliency (TraCR) under Grant No. 69A3552344812 and 69A3552348317 which is headquartered at Clemson University, South Carolina, USA, from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

Non-exclusive rights are retained by the U.S. DOT.

CONTACTS

For more information:

Trayce Hockstad
2031 Smart Communities and Innovation Building (SCIB)
Tuscaloosa, AL 35487
Phone: (810) 441-4291
Email: tahockstad@ua.edu

Latifur Khan
800 W. Campbell Road
Richardson, TX 75080
Phone: (214) 704-9634
Email: lkhan@utdallas.edu

TraCR
Clemson University
One Research Dr
Greenville, SC 29607
tracr@clemson.edu

ACKNOWLEDGMENT

This work is based upon the work supported by the National Center for Transportation Cybersecurity and Resiliency (TraCR), a U.S. Department of Transportation National University Transportation Center headquartered at Clemson University, Clemson, South Carolina, USA. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of TraCR. The U.S. Government assumes no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No. 8

2. Government Accession No. N/A

3. Recipient's Catalog No. N/A

4. Title and Subtitle
Policy Analysis and Guidance to Support Secure Transportation Cyber-Physical Systems

5. Report Date: May 2025

6. Performing Organization Code: N/A

7. Author(s)
Trayce A. Hockstad, https://orcid.org/0000-0002-1487-4564
Md. Mizanur Rahman, Ph.D.; https://orcid.org/0000-0001-8464-656X
Steven L. Jones, Ph.D.; https://orcid.org/0000-0002-0617-0604
Latifur Khan, Ph.D.; https://orcid.org/0000-0002-9300-1576
Mashrur "Ronnie" Chowdhury, Ph.D., https://orcid.org/0000-0002-3275-6983
M. Sabbir Salek, Ph.D., https://orcid.org/0000-0001-7326-3694

8. Performing Organization Report No. 8

9. Performing Organization Name and Address
National Center for Transportation Cybersecurity and Resiliency (TraCR),
Clemson University,
414 A One Research Dr, Greenville, SC 29607
Alabama Transportation Institute, University of Alabama
Smart Communities and Innovation Building, 28 Kirkbride Lane,
Tuscaloosa, AL 35401

10. Work Unit No. N/A

11. Contract or Grant No.
69A3552344812 and 69A3552348317

12. Sponsoring Agency Name and Address
U.S. Department of Transportation,
Office of the Assistant Secretary for Research and Technology,
1200 New Jersey Avenue, SE, Washington, DC 20590

13. Type of Report and Period Covered
Final Report, 01/01/2024 - 12/31/2024

14. Sponsoring Agency Code OST-R

15. Supplementary Notes
Conducted under the U.S. DOT Office of the Assistant Secretary for Research and Technology's (OST-R) University Transportation Centers (UTC) program.

16. Abstract

In a world of automated mobility, innovative but legally unprecedented technological advances are creating a host of policy issues for legislative and regulatory bodies. Although the need for regulatory and enforcement measures is dire, there is no singular federal law or federal regulatory framework that governs cybersecurity or data privacy focusing on transportation in the United States. The overarching goal of this project is to perform a nationwide survey of existing federal and state cybersecurity and privacy regulatory measures and analyze that legislative landscape in light of identified risks and threats to the transportation industry. The project attempts to answer: (i) what federal and/or state agencies are responsible for governing cybersecurity practices in the U.S., including risk assessment, preventative measures, detection of breaches, and remedial enforcement; and (ii) how do industry experts assess the greatest risks/threats to ensuring cybersecurity in the transportation sector? Key contributions include developing a novel prompt-based LLM model and a domain-specific question-answering system that will ensure the security of various systems in the transportation domain. The results of the above-discussed review and analysis could be used to construct a comprehensive transportation cybersecurity policy guidance document and/or toolkit.

17. Keywords
Computer Security, Data Privacy, Policy, Legislation and Regulation, Artificial Intelligence

18. Distribution Statement
No restrictions.

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
40

22. Price
N/A

Reproduction of completed page authorized

TABLE OF CONTENTS

List of Tables

List of Figures

EXECUTIVE SUMMARY

Although the need for regulatory and enforcement measures is dire, there is no singular federal law or federal regulatory framework that governs cybersecurity or data privacy focusing on transportation in the United States. Instead, policymakers have focused predominantly on *ex post* litigation-based remedies for consumers harmed through cybersecurity breaches. The overarching goal of this project was to perform a nationwide survey of existing federal and state cybersecurity and privacy regulatory measures and analyze that legislative landscape in light of identified risks and threats to the transportation industry. The project attempted to answer: (i) what federal and/or state agencies are responsible for governing cybersecurity practices in the U.S., including risk assessment, preventative measures, detection of breaches, and remedial enforcement; and (ii) how do industry experts assess the greatest risks/threats to ensuring cybersecurity in the transportation sector?

One of the contributions of this project was to gather, synthesize, and present most, if not all, of these existing regulations into a digestible form for industry partners to understand and reference. Researchers also sought to enable an LLM agent to answer questions based on current laws and practices, helping identify existing legal loopholes in the autonomous transportation industry. This approach streamlines said process by integrating current legislation, curating a relevant Q&A dataset, and designing pipelines to deliver factually accurate answers, ultimately supporting the identification of legal gaps in the field.

Over the first three quarters of the project year 2023-24, researchers developed an RAG-driven LLM pipeline to find legislative gaps in federal, state, and international level legislation. The venture aims to identify loopholes in federal and state legislation to address data privacy and cybersecurity challenges in autonomous vehicles, proposing necessary modifications to ensure future scenarios are effectively managed. It involves conducting a comprehensive national survey of transportation cybersecurity laws and regulations, identifying key concerns from industry stakeholders, and leveraging large language model (LLM)-based natural language processing (NLP) techniques to analyze legislative consistencies and gaps. The outcome will be a broadly applicable policy guidance document to assist researchers and policymakers in developing cybersecurity best practices, draft legislation, and regulatory frameworks.

Researchers also drafted questions to conduct a national survey with the input of several state departments of transportation. Based on the results of these initial research endeavors, researchers developed surveys to obtain additional input from other transportation industry participants regarding their most pressing concerns related to cybersecurity or data privacy issues. Survey questions were meticulously designed to facilitate a thorough analysis of regulatory gaps, and researchers leveraged connections with state DOT(s) to engage relevant individuals in the survey process, ensuring reliable outcomes and a high response rate.

While data security regulation is gaining traction among legislative bodies, existing laws remain fragmented in their approach to cybersecurity. Some industries, such as finance, healthcare, and insurance, have had sector-specific regulations for decades, while data protection laws outside these areas are relatively new and advancing slowly. At the same time, data collection, storage, and transfer have surged with the rise of connected and smart technologies, introducing unprecedented cybersecurity risks worldwide. Despite the growing volume of personal data collected daily, federal data security measures for critical infrastructure sectors have been slow to develop. Recognizing this gap, some states have begun implementing broad data privacy regulations.

Data breach notification laws are similarly inconsistent nationwide. States differ on whether data managers must notify affected individuals, state authorities, or consumer reporting agencies, and no clear regional patterns exist for analyzing these variations. Many states impose threshold requirements for breach notifications, but these thresholds vary widely. In states without consumer data protection laws, residents lack guaranteed rights to access, control, delete, or opt out of data collection altogether. However, data privacy should not be dictated by geography—an individual's personal information remains the same whether stored in Maine, Arizona, or elsewhere. Yet, differing security measures, collection practices, and breach notification laws create disparities in consumer protections and complicate compliance for data managers in critical infrastructure sectors. Given these challenges, further research is needed to assess the compatibility of emerging data security regulations and to establish a more unified and coordinated national approach to cybersecurity governance.

One way this might be accomplished is by aiding policymakers in the analysis of existing legislation to understand existing legal regulations as well as to draft future laws. These findings show that state-of-the-art LLMs struggle with specialized queries on recent legislation, leading to potential misidentification of legislative gaps. However, RAG-powered LLMs offer promise in identifying gaps in transportation cybersecurity by integrating recent legal advancements for factually accurate responses. This framework supports concept-based and state-specific analysis to uncover missing legislative elements but requires an additional module to compare responses across legislative bodies. Future work will enhance user guidance and seamless legal data integration. By mitigating hallucinations, this approach strengthens legislative analysis and supports evolving regulatory landscapes.

CHAPTER 1

Introduction

1.1 Need for Research

Evolving transportation systems are using computing and communications power to integrate and optimize systems for moving goods and people while focusing on equitably advancing society. Transformative technologies include autonomous vehicles (AVs), vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, and hardware and software that enable us to store, transfer, process, analyze, and act on huge amounts of data in real-time (Khan, M.A., et al., 2022) (Khayyam, H., et al., 2020) (Khan, Z., et al., 2022). All of this will reduce congestion, delays, crashes, fuel use, emissions, the monetary costs of transportation, social inequities, and more. However, the more individuals rely on automation and connectivity, the more malicious actors gain unprecedented opportunities to steal data, invade privacy, demand ransom, generate misinformation, and ultimately even shut down the systems on which lives, prosperity, and security depend (Chowdhury, A. et al., 2020) (Taeihagh, A. and Lim, 2019). Although the need for regulatory and enforcement measures is dire, there is no singular federal law or federal regulatory framework that governs cybersecurity or data privacy focusing on transportation in the United States. In a world of automated mobility, innovative but legally unprecedented technological advances are creating a host of policy issues for legislative and regulatory bodies. These include problems surrounding the amount, nature, and potential exploitation of data collected from connected transportation systems. Perhaps most concerning, current cybersecurity regulations overwhelmingly fail to require, or even encourage, the use of machine learning and predictive analysis to understand privacy threats, cyberattacks, and data theft. Instead, policymakers have focused predominantly on *ex post* litigation-based remedies for consumers harmed through cybersecurity breaches (Dempsey, 2021).

1.2 USDOT Relevance

This project directly focused on the TraCR's Research Thrust 3 "Society and Environment." In addition, this project addressed several of USDOT's strategic goals including (1) serving all citizens, particularly those from underserved backgrounds and rural areas ("Equity"); (2) improving the safety of urban, rural and underserved communities ("Safety"), and (3) ensuring safe and secure movements of people in rural and underserved communities ("Economic Strength and Global Competitiveness," "Organizational Excellence"). Moreover, in 2015, the Department of Homeland Security issued its Transportation Systems Sector Cybersecurity Framework Implementation Guidance, to encourage organizations to: (i) characterize current cybersecurity posture; (ii) identify opportunities for enhancing cyber risk management programs; (iii) find existing standards to support framework implementation; and (iv) communicate risk management issues to internal and external stakeholders. (Cybersecurity and Infrastructure Security Agency, 2020).

1.3 Research Goals and Objectives

The overarching goal of this project was to perform a nationwide survey of existing federal and state cybersecurity and privacy regulatory measures and analyze that legislative landscape in light of identified risks and threats to the transportation industry. Specifically, the objectives of this project were to:

Objective 1: Identify and analyze the gaps that exist in the United States' cybersecurity regulatory schematic as applied to transportation law and policy.

Objective 2: Develop a policy guidance document and/or toolkit to assist interested stakeholders in constructing and implementing effective transportation cybersecurity measures.

The project attempted to answer: (i) what federal and/or state agencies are responsible for governing cybersecurity practices in the U.S., including risk assessment, preventative measures, detection of breaches, and remedial enforcement; and (ii) how do industry experts assess the greatest risks/threats to ensuring cybersecurity in the transportation sector? The results of these two reviews were then analyzed using natural language processing methods to identify consistencies and gaps in what cybersecurity policy the nation *does have* and what the industry indicates it *should have*. Finally, this analysis will be used to develop a policy guidance document that can be shared with stakeholders who wish to develop and implement effective cybersecurity legislation and regulatory governance. This analytical process for the research question development is depicted in the figure below.

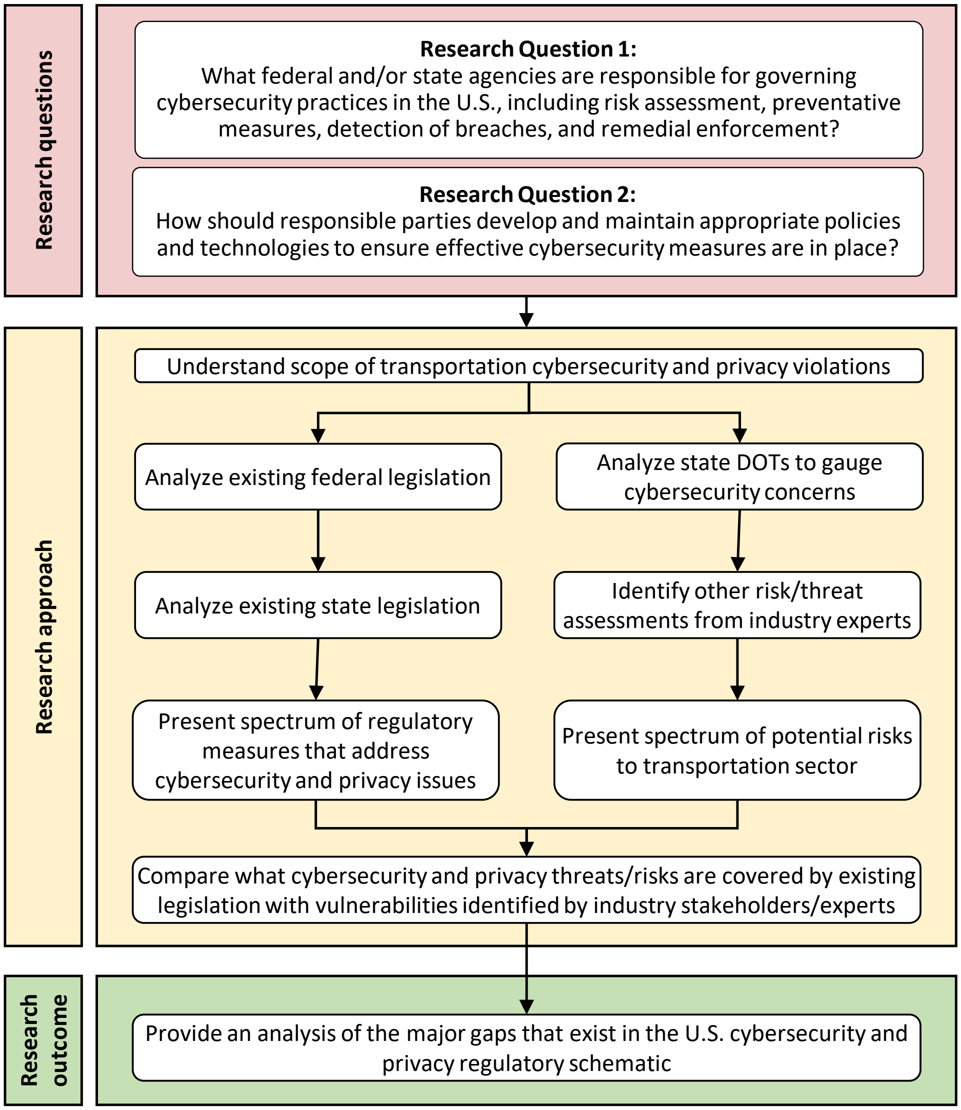
Research questions
Research Question 1:
What federal and/or state agencies are responsible for governing cybersecurity practices in the U.S., including risk assessment, preventative measures, detection of breaches, and remedial enforcement?
Research Question 2:
How should responsible parties develop and maintain appropriate policies and technologies to ensure effective cybersecurity measures are in place?
Research approach
Understand scope of transportation cybersecurity and privacy violations
Analyze existing federal legislation
Analyze state DOTs to gauge cybersecurity concerns
Analyze existing state legislation
Identify other risk/threat assessments from industry experts
Present spectrum of regulatory measures that address cybersecurity and privacy issues
Present spectrum of potential risks to transportation sector
Compare what cybersecurity and privacy threats/risks are covered by existing legislation with vulnerabilities identified by industry stakeholders/experts
Research outcome
Provide an analysis of the major gaps that exist in the U.S. cybersecurity and privacy regulatory schematic

Figure 1. Research Plan.

CHAPTER 2

Literature Review

2.1 Industry Survey Results

In 2022, a national survey of state departments of transportation (DOTs) was conducted, and 19 state DOTs responded to this survey. This was an effort to identify ways to assist with agencies' cybersecurity needs. After preparing the questionnaire, it was reviewed by the Institutional Review Board (IRB) at Clemson University to ensure that it was following all required guidelines. As a preliminary matter, every single entity either agreed or strongly agreed that cybersecurity was a concern for their respective agency. The responses more specifically indicated a widespread desire for: (1) workforce training and development; (2) guidance on identifying and implementing cybersecurity tools, standards, and best practices; and (3) advice on privacy issues (see Figures 2 and 3). The most pressing threats identified were cyberattacks on DOTs' communication systems, breaches of data storage, and assaults on physical systems and infrastructures (e.g., traffic light systems).

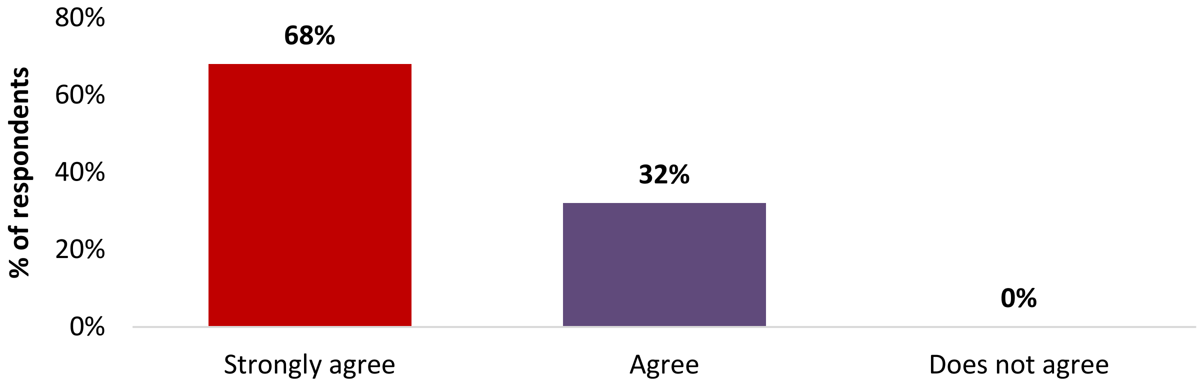

Figure 2. Is Cybersecurity a Concern for the Agency?

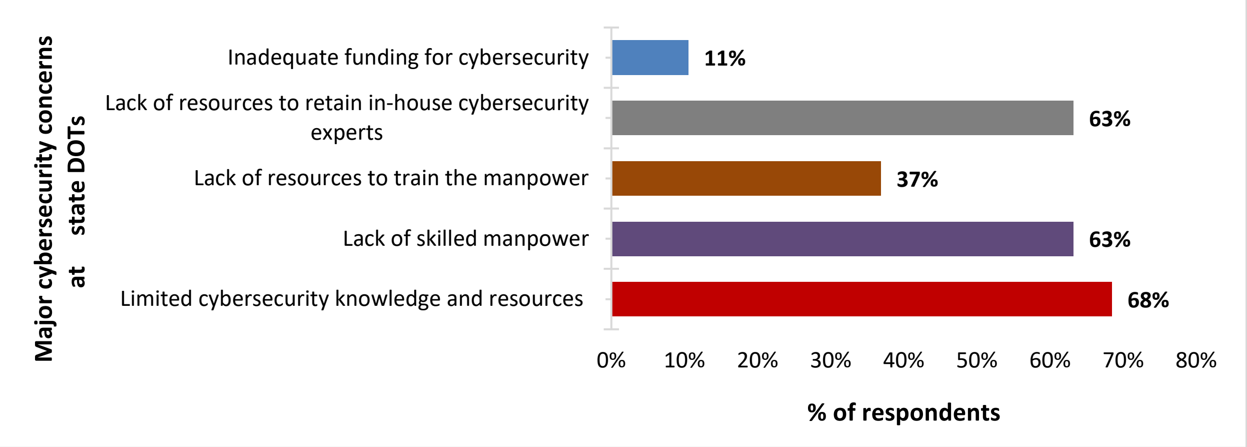

Figure 3. Major Cybersecurity Concerns at State DOTs.

2.2 Other Stakeholder Concerns

Cybersecurity has been a concern since the internet became widely adopted, but recent years have seen an unprecedented surge in privacy breaches and major cyberattacks. This trend largely stems from the shift to remote and flexible work arrangements introduced during the COVID-19 pandemic (DiFurio, D., 2023). In 2021, cyberattacks led to the shutdown of multiple hospitals, schools, and municipal governments in the U.S., along with the Colonial Pipeline breach, which caused a gas panic and financial strain on consumers. One survey found that ransomware attacks in the transportation sector increased by 186% between June 2020 and June 2021 (Bowcut, S., 2023). An annual IBM report indicated that in 2022, the average cost of a data breach in the U.S. was about $9.44 million.

Experts predict that the transportation sector will increasingly become a target for cybercriminals due to the industry's essential role. Millions rely on intermodal mobility networks for daily activities, including employment, education, healthcare, emergency services, and social or religious functions. The companies managing these systems are often profitable and influential, making them prime targets for financial exploitation. However, transportation firms have traditionally prioritized safety and physical security over cybersecurity (Bowcut, S., 2023). As a result, if the current rise in cyberattacks persists, the industry could face substantial risks.

Industry best practices have highlighted several key vulnerabilities in public transit that heighten cybersecurity risks. For example, public transit agencies increasingly subcontract certain system management responsibilities (Belcher, S., 2020). As emerging technologies and data sources are integrated through new vendor relationships, the potential for cybersecurity threats expands. Additionally, transit agencies adopting new technologies have not always implemented the necessary safeguards to address related risks. Consequently, the risk of operational network compromises is estimated to be growing faster than the technical ability to assess and resolve these vulnerabilities (Grzadkowska, A., 2018).

2.3 Regulatory Scheme

In 2015, the Department of Homeland Security released the *Transportation Systems Sector Cybersecurity Framework Implementation Guidance*, outlining how owners and operators could apply the National Institute of Standards and Technology's (NIST) Cybersecurity Framework to reduce cyber vulnerabilities (Cybersecurity and Infrastructure Security Agency, 2023). Organizations are encouraged to: (i) evaluate their current cybersecurity posture; (ii) identify opportunities to strengthen cyber risk management programs; (iii) support framework implementation; and (iv) communicate risk management concerns to stakeholders.

Cybersecurity law extends beyond data privacy to include legal tools for addressing cybercrime criminal activities conducted through networked technologies (Lukings, M.A., 2022). These laws aim to safeguard information and IT systems from unauthorized access while requiring institutions to protect online infrastructure from cyberattacks. As the internet has become a fundamental part of global society, the scale, motives, and tactics of cybercrime have also expanded (Dutta, N., et al., 2022).

Key federal cybersecurity regulations have historically included the *Privacy Act of 1974*, the *Health Insurance Portability and Accountability Act of 1996* (HIPAA), the *Gramm–Leach–Bliley Act of 1999*, and the *Homeland Security Act of 2002*, which incorporated the *Federal Information Security Management Act* (FISMA). Additionally, the *Computer Fraud and Abuse Act* grants individuals the right to seek compensation and injunctive relief for violations, while the *National Information Infrastructure Protection Act of 1996* later amended it by imposing harsher penalties for certain offenses.

In 2018, Congress amended the *Homeland Security Act of 2002* by passing the *Cybersecurity and Infrastructure Security Agency Act*, which created the Cybersecurity and Infrastructure Security Agency (CISA) within the Department of Homeland Security. CISA works with government and private sector partners to safeguard critical infrastructure by providing expertise on cybersecurity vulnerabilities, response strategies, and risk assessments. In March 2022, Congress enacted the *Cyber Incident Reporting for Critical Infrastructure Act of 2022* (CIRCIA), requiring certain critical infrastructure institutions to report major cyber and ransomware incidents to CISA. Additionally, the government has issued sector-specific cybersecurity guidance for industries such as nuclear, chemical, electrical, government contracting, and transportation. CIRCIA's final reporting rule is scheduled for release on September 15, 2025 (Public Law 117–103).

If a cyber incident raises national security or terrorism concerns, law enforcement can obtain information, facilities, and technical assistance under the *Foreign Intelligence Surveillance Act* (FISA). Similar investigative powers exist under the *Stored Communications Act*, *Federal Wiretap Act*, and the *Communications Assistance for Law Enforcement Act* (CALEA), all of which generally require oversight by a specialized federal court. However, there are no general mandates requiring organizations to create alternative access points in their IT systems or provide law enforcement with encryption keys upon discovering an incident (McNicholas, E., 2023). Instead, law enforcement must either rely on voluntary cooperation from private entities or seek access through judicial processes. In a related measure, the *Transportation Security Administration* issued *Pipeline Security Guidelines* in March 2018, directing operators to develop risk-based security plans that address various threats and vulnerabilities, including cybersecurity. These guidelines require operators to implement a cyber/Supervisory Control and Data Acquisition (SCADA) security plan (Dempsey, J., 2021).

Nearly all states have enacted laws imposing security requirements on institutions that collect, store, transmit, or manage personal information (PI), typically establishing a standard of reasonable security practices. Additionally, all states and four U.S. territories have breach notification laws requiring notification to state regulators and affected individuals in the event of a cybersecurity incident. While some states mandate notification of suspected access to PI, most require confirmation of acquisition. Notification deadlines vary, but 30 days is common. PI generally includes names or initials combined with other identifiers such as social security numbers, driver's license numbers, financial account details, and access credentials. Biometric data and login credentials for online accounts can also be included in these definitions.

Furthermore, nearly all automakers have adopted the *Automotive Privacy Principles* to guide privacy practices in the automotive industry (National Automobile Dealers Association, 2021).

These principles emphasize transparency, affirmative consent for sensitive data, and limited data sharing with government and law enforcement. Automakers commit to providing clear privacy policies, obtaining consumer consent before collecting sensitive data for marketing or third-party sharing, and restricting government access to specific circumstances. The *National Automobile Dealers Association* (NADA) defines sensitive data as geolocation information, biometric data, and driver behavior data.

2.4 Natural Language Processing Capabilities

Various states have diverse regulatory laws that result in a substantial volume of documents to sift through when identifying gaps. This can be a very time-consuming and resource-intensive process if these regulations need to be processed manually. Natural language processing (NLP) models offer an efficient solution by being equipped to extract valuable insights from these documents regarding the relationship between current policies and experts' concerns[1]. Key contributions of utilizing this approach included developing a novel prompt-based LLM model and a domain-specific question-answering system that will ensure the security of various systems in the transportation domain. More specifically, researchers used advanced question-answering or text summarization techniques involving transformer-based large language models (LLMs) to automatically extract relevant policies and laws from legal documents.

BERT (Devlin, J. et al., 2019) remains foundational in NLP for its bidirectional attention and pre-training, advancing language understanding. Yet, large LLMs are difficult to retrain and prone to hallucinations—fabricated or irrelevant outputs, especially with unfamiliar or complex inputs. To handle domain-specific needs, ConfliBERT (Hu, Y., et al., 2022) improves performance on political violence tasks using tailored training data. Still, hallucinations and unsafe outputs persist. Sun, H. et al. (2022) offered a taxonomy for unsafe conversational AI responses, addressing risks like harmful medical advice, and introduced the DiaSafety dataset and a classifier to detect and reduce such outputs.

Zhang, Y., et al. (2023) provide a comprehensive analysis of hallucinations in large language models (LLMs), examining their evaluation, underlying causes, and potential mitigation strategies. The study categorizes hallucinations into three types: input-conflicting hallucinations, which occur when generated content deviates from user instructions or input; context-conflicting hallucinations, which involve inconsistencies within model responses during multi-turn interactions; and fact-conflicting hallucinations, where the model produces information that contradicts established knowledge.

Fact-conflicting hallucinations are the most difficult to manage and require advanced retrieval-augmented strategies. Huang, L., et al. (2023) categorize hallucinations by source and impact, reviewing mitigation methods and outlining challenges and unresolved issues. Dziri, N., et al. (2022) found over 60% of hallucinations stem from dataset inconsistencies, worsened by model architecture during inference, highlighting the need for improved data curation and model design.

[1] Parolin ES, Hu Y, Khan L, Brandt PT, Osorio J, D'Orazio V. Confli-T5: An AutoPrompt Pipeline for Conflict Related Text Augmentation. In2022 IEEE International Conference on Big Data (Big Data) 2022 Dec 17 (pp. 1906-1913). IEEE.

Retrieval-based approaches show promise in addressing these issues. Lewis, P., et al. (2020) proposed Retrieval-Augmented Generation (RAG), combining parametric and non-parametric memory to boost factual accuracy in tasks like QA and fact verification using dense retrieval over external corpora. RAG set new benchmarks in several domains. Extending this, Zhang, T., et al. (2024) developed Retrieval-Augmented Fine-Tuning (RAFT), which improves citation and relevance via chain-of-thought reasoning, outperforming standard fine-tuning on datasets like PubMed and HotpotQA.

2.5 Gained Insights for LLM Development

As large language models (LLMs) continue to evolve, several critical insights have emerged that inform their ongoing development and deployment. Despite remarkable advances in generative capabilities, LLMs still exhibit fundamental limitations that must be addressed to ensure their safe and effective use, particularly in high-stakes domains. Key areas of concern include persistent hallucinations, unsafe outputs, and the influence of training data quality. At the same time, promising solutions are taking shape, especially in retrieval-augmented architectures that improve factual grounding. The following insights highlight both the challenges and the strategic directions guiding the next generation of LLM development:

1. **Hallucination in LLMs is a Persistent Challenge** – LLMs frequently generate factually incorrect or inconsistent responses, with fact-conflicting hallucinations being the most problematic, especially in high-stakes applications like healthcare and law.
2. **Unsafe Outputs Require Context-Sensitive Solutions** – LLMs still struggle with producing biased, harmful, or misleading responses, highlighting the need for better context-aware safety mechanisms rather than simple content filtering.
3. **Data Quality is a Key Factor** – A significant portion of hallucinations stems from poor-quality training data, emphasizing the need for better dataset curation and preprocessing to improve model reliability.
4. **Retrieval-Augmented Models Reduce Hallucinations** – Techniques like RAG and RAFT enhance factual accuracy by retrieving relevant external knowledge, making them superior to standard LLMs in tasks requiring precise information.
5. **The Future Lies in Hybrid Models** – The most promising approach involves combining generative LLMs with external retrieval systems, ensuring responses are both coherent and factually grounded**.**

CHAPTER 3

Methods

3.1 Legal Database Construction

Although there are numerous federal and state laws and regulations that govern transportation cybersecurity issues, no comprehensive overview of applicable standards exists to guide industry stakeholders. Best practices for risk identification and mitigation, civil and criminal sanctions for violations, and remedial enforcement measures are scattered across various statutes and administrative agency policies such that understanding and implementing effective and compliant cybersecurity governance is evasive for many transportation agencies and other participants. One of the contributions of this project will be to gather, synthesize, and present most, if not all, of these existing regulations into a digestible form for industry partners to understand and reference. To do so, the project will include a nationwide collection and synthesis of applicable existing federal, state, and administrative agency policies pertaining to transportation cybersecurity. This comprehensive approach ensures a more holistic understanding of the landscape of connected and autonomous vehicles and their cybersecurity and data privacy issues.

When constructing the database architecture, the following method of legislation retrieval was used:

Search Methodology: Results were filtered for each state jurisdiction individually. Category "statutes and legislation" was searched for the terms listed below (not in quotation marks). Results were limited to the category "Code of [State]". All available dates were included.	**Criteria for Inclusion:** Relevant code sections not duplicative were selected unless otherwise specified.
Search Term: cybersecurity	Relevant code sections selected.
Search Term: data breach	Relevant code sections *not duplicative* were selected.
Search Term: crim! use of computer	Relevant code sections *not duplicative* were selected.
Search Term: denial of service	Relevant code sections *not duplicative* were selected.
Search Term: ransomware	Relevant code sections *not duplicative* were selected.
Search Term: trade secrets	Each state's trade secrets act and/or relevant code sections was then selected.

Common search results identified as outside the scope: election security, cyberbullying, cyberstalking, taxation, CSE/C laws. After relevant legislation was compiled, the architecture for the LLM was organized so that the model could be trained in its analysis.

3.2 RAG-Driven Large Language Model Development

3.2.1 Objectives

Over the first three quarters of this project year 2023-24, researchers developed a RAG-driven LLM ("TraCR AI") pipeline to find legislative gaps in federal, state, and international level legislation. The venture aims to identify loopholes in federal and state legislation to address data privacy and cybersecurity challenges in autonomous vehicles, proposing necessary modifications to ensure future scenarios are effectively managed. It involves conducting a comprehensive national survey of transportation cybersecurity laws and regulations, identifying key concerns from industry stakeholders, and leveraging large language model (LLM)-based natural language processing (NLP) techniques to analyze legislative consistencies and gaps. The outcome will be a broadly applicable policy guidance document to assist researchers and policymakers in developing cybersecurity best practices, draft legislation, and regulatory frameworks.

Table 1. Task List & Summary.

Step #	Task Description	From	To Get
1	Response Validation and Quality Check	Answers generated by RAG-driven LLM	Current Scenarios of Current Cybersecurity Law Practices in Linguistic Format
2	Improving Quality of Responses	Answers generated by RAG-driven LLM	Handling Classical Hallucination Issues of LLMs and Generate Factually Correct Responses
3	Improving the Retriever's Performance	Context retrieved for generating answers by RAG-driven LLM	To Accumulate Appropriate Contexts from Data Sources for LLM Agents
4	Pinpointing Loopholes in Legislation	Answers Generated by RAG-driven LLM	To Identify Existing Legislative Gaps at Federal, State, and International Levels That Require Addressing

3.2.2 Methodology Overview and Completed Targets

Researchers began by creating retrieval indices for the entire dataset. The documents are divided into text chunks, and embeddings are generated from these chunks. When a question is asked, it is also converted into an embedding, and similar embeddings along with their corresponding text are retrieved from the indices as retrieved contexts. Next, researchers employ prompt engineering to construct a well-structured prompt for the LLM, incorporating the question, retrieved contexts, and necessary instructions. This prompt is then fed into LLM, which generates the final response.

The team implemented two approaches (Khandakar, A., et al., 2024) (Khandakar, A., et al., 2020):

Whole Data Index (WDI) – In this method, researchers created a single index for the entire dataset.

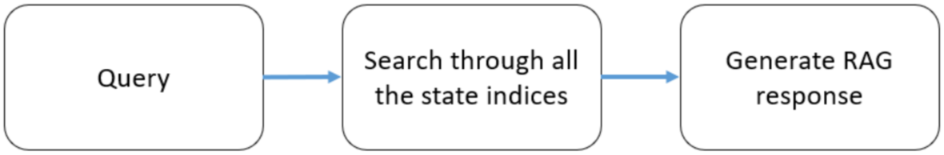

Figure 4. Flow of First Method (Whole Data Index).

State-Wise Index (SWI) – In this method, researchers segmented the dataset by state and created separate indices for each state.

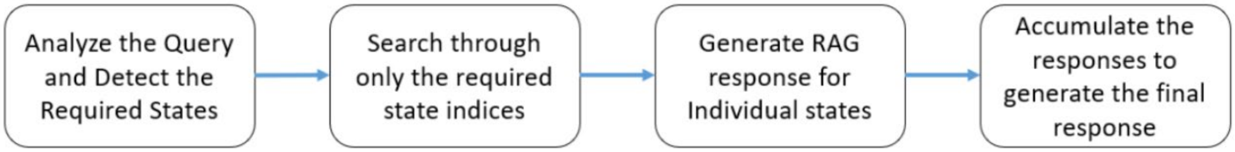
Analyze the Query and Detect the Required States
Search through only the required state indices
Generate RAG response for Individual states
Accumulate the responses to generate the final response

Figure 5. Flow of Second Method (State-Wise Index).

Next, researchers curated a comprehensive question dataset, one that is comprehensive in characteristics, to pinpoint the loopholes in the existing legislation. Some of the sample questions are provided below:

a. How do most states define critical infrastructure in the context of cybersecurity?
b. Can you provide examples of penalties or fines imposed for non-compliance with cybersecurity requirements?
c. How do states address data privacy concerns in the realm of cybersecurity?

Currently, researchers have 59 questions covering a lot of the important aspects of data privacy and cybersecurity in the transportation domain. The plan is to expand the Q&A dataset as researchers delve deeper into the project. The team further identified and completed the following:

- RAG pipeline development for retrieving relevant contexts/legislation and answering queries
- Website development for uploading newer versions of legislation and interactions via queries
- Response Validation and Quality Check
- Pinpointing Loopholes in Legislation

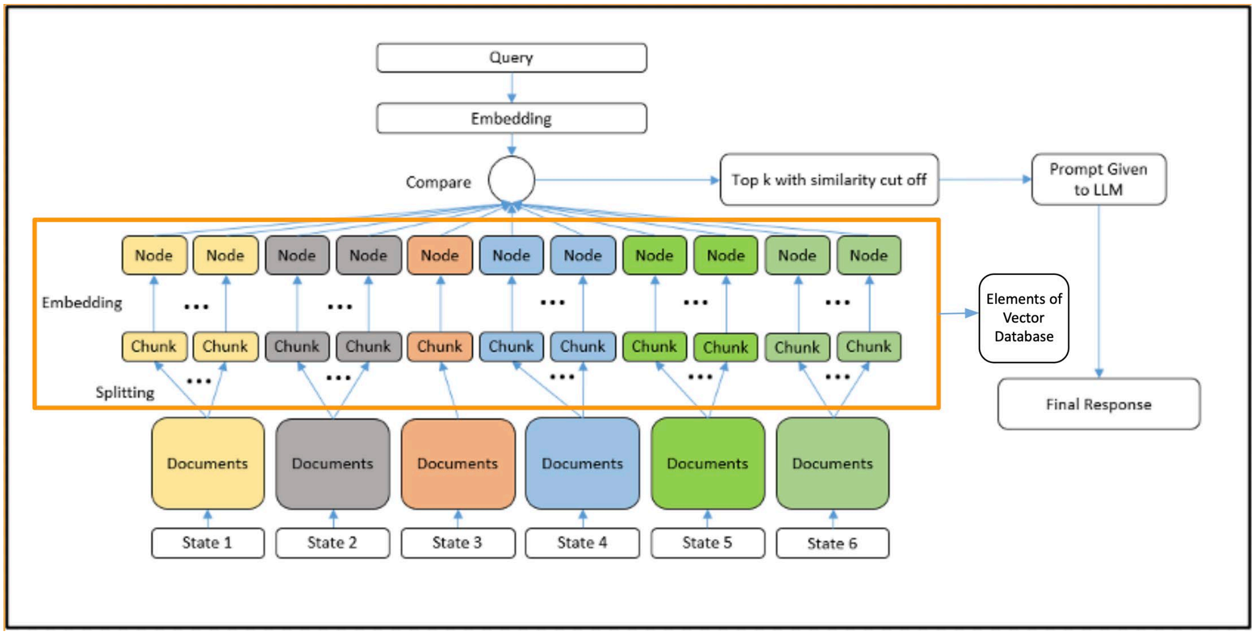

Figure 6. Complete Architecture for Index on Entire Data: Whole Data Index (WDI).

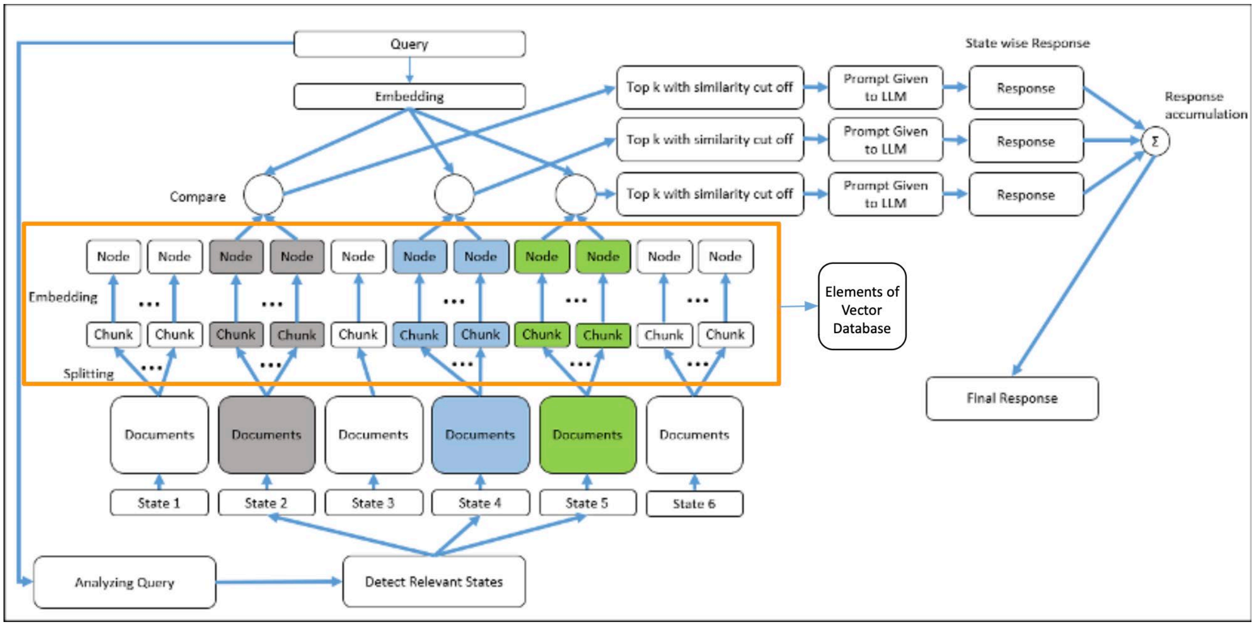

Figure 7. Complete Architecture for State-Wise Individual Retriever: State-Wise Index (SWI).

Through these processes, researchers were able to improve the quality of responses as well as the retriever's performance. This enabled us to:

- conduct a comprehensive evaluation using multiple metrics, including AlignScore, ParaScore, ROUGE Score, and BERTScore;
- assess the feasibility of the end-to-end approach, considering resource consumption and response fatigue;
- and evaluate the effectiveness of the generated answers in facilitating legislative gap analysis.

3.3 Industry Survey Design

Researchers also drafted questions to conduct a national survey with the input of several state departments of transportation. Based on the results of these initial research endeavors, the team developed surveys to obtain additional input from other transportation industry participants regarding their most pressing concerns related to cybersecurity or data privacy issues. The survey questions were meticulously designed to facilitate a thorough analysis of regulatory gaps, and researchers leveraged connections with state DOT(s) to engage relevant individuals in the survey process, ensuring reliable outcomes and a high response rate. This approach is vital to ensure the credibility of these responses. The focus of this survey was to identify the areas depicted in orange and yellow in the figure below: industry and government agency needs.

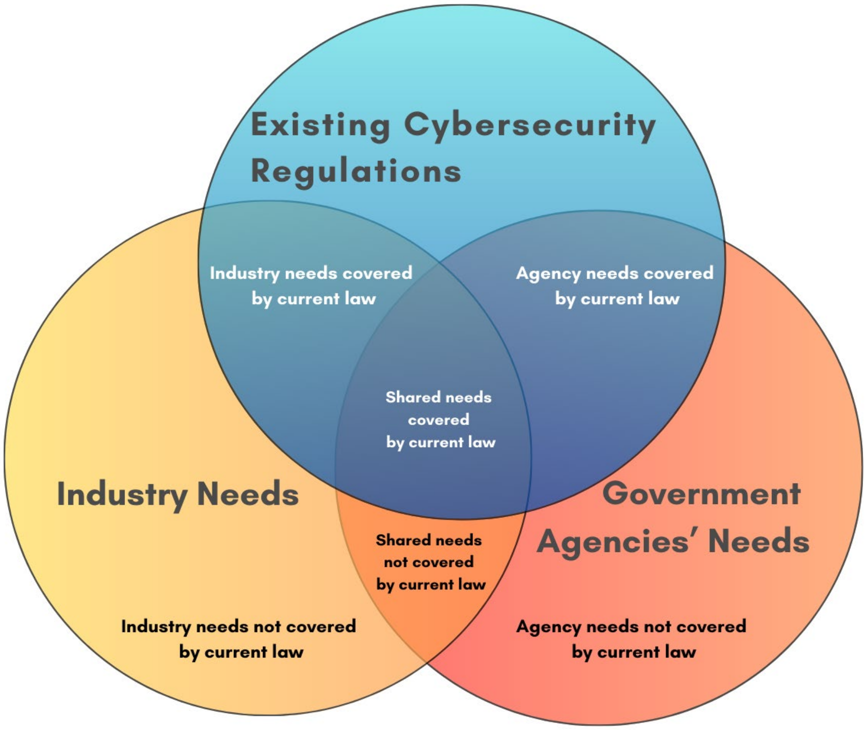
Existing Cybersecurity Regulations
Industry needs covered by current law
Agency needs covered by current law
Shared needs covered by current law
Industry Needs
Government Agencies' Needs
Shared needs not covered by current law
Industry needs not covered by current law
Agency needs not covered by current law

Figure 8. Targeted Subject of Research Results (Yellow and Orange Areas).

CHAPTER 4

Results

4.1 Validation Results and Comparative Analysis for LLM Functioning

Table 2 below presents the validation results of the open-ended Q&A dataset answers across different metrics

Table 2: Comparison of RAG-powered GPT with State-of-the-Art Commercial LLMs.

Model	**AlignScore**	**ParaScore**	**BERTScore**			**ROUGEL Score**		
			Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
ChatGPT3.5	0.64651	0.64087	0.83419	0.86653	0.84965	0.19754	0.42210	0.25063
ChatGPT4o	0.71040	0.62080	0.85262	0.86084	0.85644	0.24638	0.41660	0.29307
Gemini	0.66220	0.57926	0.83840	0.84516	0.84147	0.21843	0.55644	0.29974
Claude	0.66579	0.66195	0.84620	**0.89018**	0.86737	0.25837	0.32660	0.25692
RAG-powered GPT	**0.73201**	**0.70433**	**0.85312**	0.88861	**0.87033**	**0.32576**	**0.56292**	**0.37340**

Researchers were also able to conduct a comparative analysis of how various approaches perform across different settings and present case studies to evaluate the effectiveness of each approach in specific scenarios.

The findings presented in these tables highlight both the performance accuracy and response dynamics of different large language models (LLMs) and methods when handling open-ended Q&A tasks. Tables 2 and 3 present key findings on the performance and response characteristics of various large language models (LLMs) in open-ended Q&A tasks. Table 2 compares five models - ChatGPT-3.5, ChatGPT-4o, Gemini, Claude, and RAG-powered GPT across evaluation metrics such as AlignScore, ParaScore, BERTScore, and ROUGEL. RAG-powered GPT outperforms all others, demonstrating the highest scores in contextual alignment and factual grounding, particularly with an AlignScore of 0.73201 and a ROUGEL F1-Score of 0.37340. ChatGPT-4o ranks second in most categories, while ChatGPT-3.5 shows the lowest performance overall. These results confirm the superiority of retrieval-augmented models in producing accurate and coherent legal and regulatory responses. Table 3 assesses the response times and effectiveness of two inference methods, Whole Document Inference (WDI) and Segmented Window Inference (SWI), across three query types. SWI performs faster on single-document queries, while WDI is faster on multi-document queries. However, for complex queries, SWI is more accurate despite taking longer, correctly identifying relevant jurisdictions where WDI does not. Collectively, these tables highlight the trade-offs between speed and accuracy and underscore the value of retrieval-augmented strategies in improving both precision and factual reliability in domain-specific applications.

Table 3: Comparison of Response Time Between Two Different Variations of the RAG-driven LLM.

Query Type	Response Time (Seconds)	
	Method 'WDI'	Method 'SWI'
Single Document Query: "What is the definition of identification documents according to Alabama acts?"	16.21	**14.66**
Multi-Document Query: "Give me an extensive comparison between Colorado Transportation act and Connecticut transportation act."	**10.72**	13.25
Complex Query: "Give me a comparison between the Digital Crime Acts of Florida and its neighboring states."	18.03 (States are not correctly identified)	**23.12** (States are correctly identified)

These findings suggest that even state-of-the-art LLMs face challenges when responding to specialized queries, particularly those involving recently enacted legislation. This limitation is significant, as inaccurate output can misidentify legislative gaps and complicate the work of legal and policy stakeholders. Nonetheless, the results point to a promising direction LLMs, when paired with recent legal developments, show clear potential in identifying gaps within the transportation cybersecurity landscape. The takeaway is straightforward: RAG-powered LLMs offer a strong foundation for detecting legislative deficiencies, with the added benefit of grounding responses in factual accuracy, an essential condition for meaningful legal analysis.

The scalability of our solution is a crucial factor in its utility. The computational and operational cost of maintaining a vector database in an RAG system largely depends on the need for reindexing. If the embedding model used to generate vector representations remains stable, the cost of adding new documents is generally linear and does not require full reindexing. However, periodic full reindexing may be necessary, particularly in domains where the source content is subject to regular updates, such as annual revisions to state legislation or corrections to existing laws. Although such reindexing incurs significant computational overhead, it is not required frequently and should be planned as part of scheduled maintenance cycles. In a nutshell, incremental document additions are computationally inexpensive as long as the embedding model remains stable. However, periodic full-scale reindexing may be warranted if previous reindexing efforts have demonstrated significant improvements in retrieval quality or model performance. Incremental updates involving thousands of documents can typically be completed within seconds, especially when using GPU acceleration. In contrast, full-scale reindexing may take minutes to hours, depending on the size of the dataset and the embedding model used, as it requires re-generating embeddings for all documents and rebuilding the entire vector index from scratch.

4.2 Policy Analysis

The team was able to construct the above-described LLM, ("TraCR AI") which then proved capable of analyzing cybersecurity legislation in the U.S. To train this LLM, "TraCR AI," researchers constructed a database of currently enacted cybersecurity laws at the federal and state levels. Through a months-long process of question-and-answer verification, the team brought TraCR AI to a point of competent research assistance that has enabled us to begin an in-depth analysis of American cybersecurity policy.

Researchers were able to gain a clearer understanding of the national schematic of state-level laws on issues related to cybersecurity and data management. A few examples of the legislative analysis the team performed is shown in the figures below.

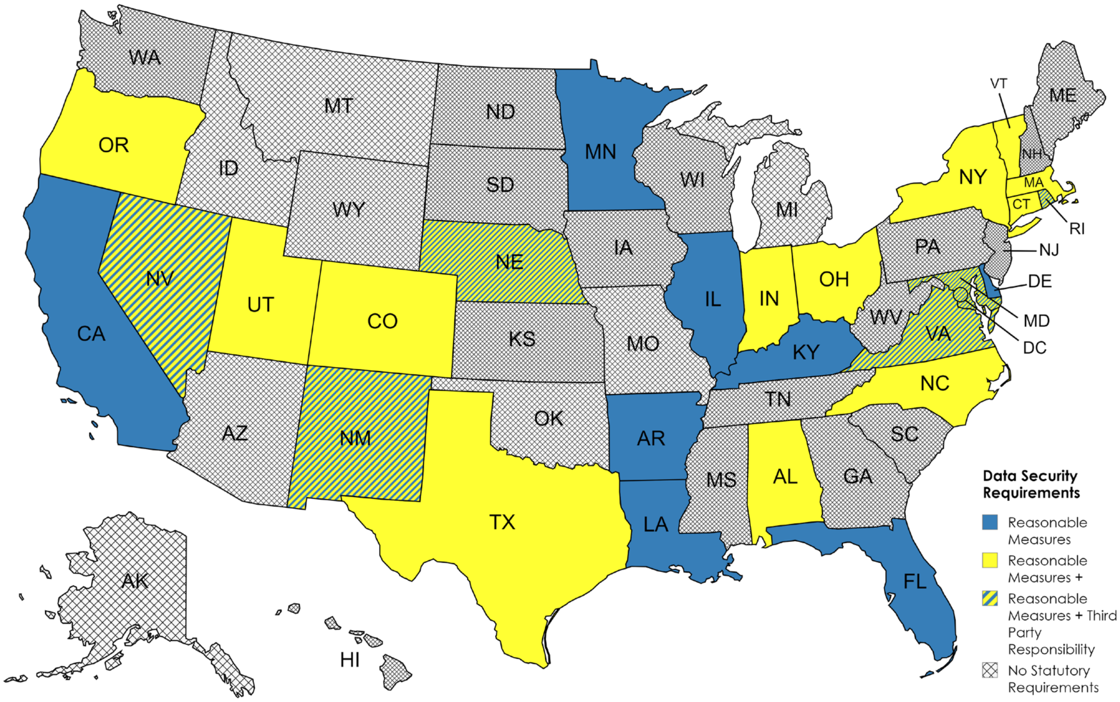

Figure 9. Data Security Requirements.

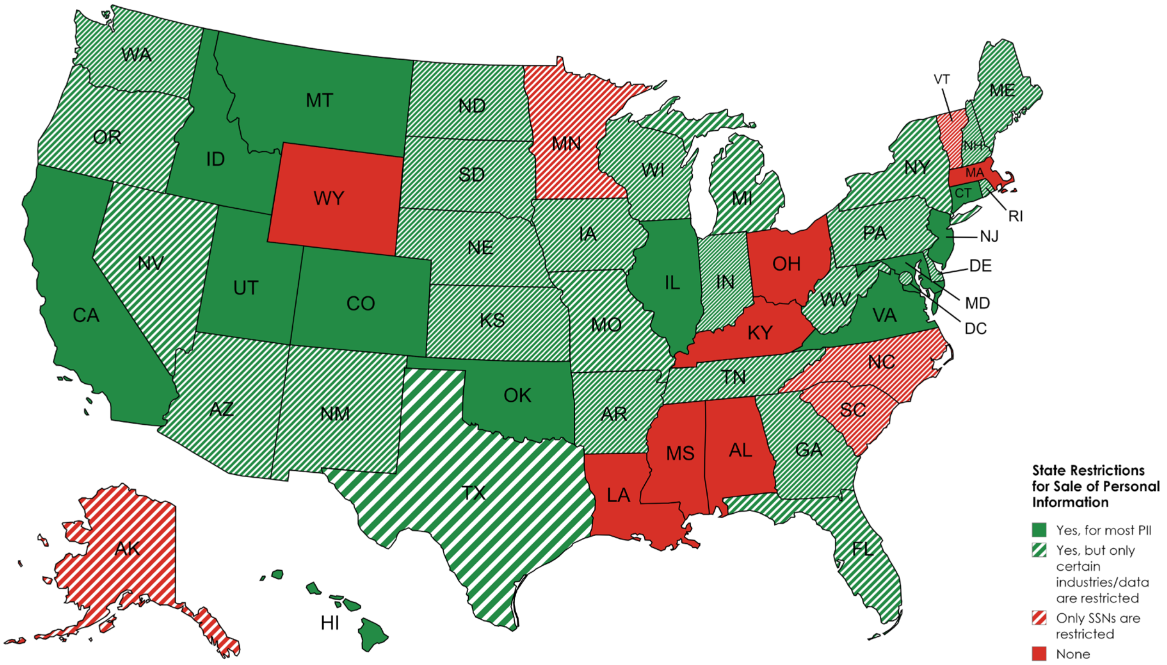
WA
MT
ND
MN
VT
ME
OR
ID
SD
WI
NY
NH
MA
WY
MI
CT
RI
PA
NJ
IA
NE
NV
OH
DE
UT
IL
IN
CO
WV
MD
CA
KS
MO
VA
KY
DC
NC
TN
OK
AZ
NM
AR
SC
MS
AL
GA
TX
LA
AK
FL
HI
State Restrictions for Sale of Personal Information
🟩 Yes, for most PII
▨ Yes, but only certain industries/data are restricted
▨ Only SSNs are restricted
🟥 None

Figure 10. Restrictions for Sale of Personal Information Data.

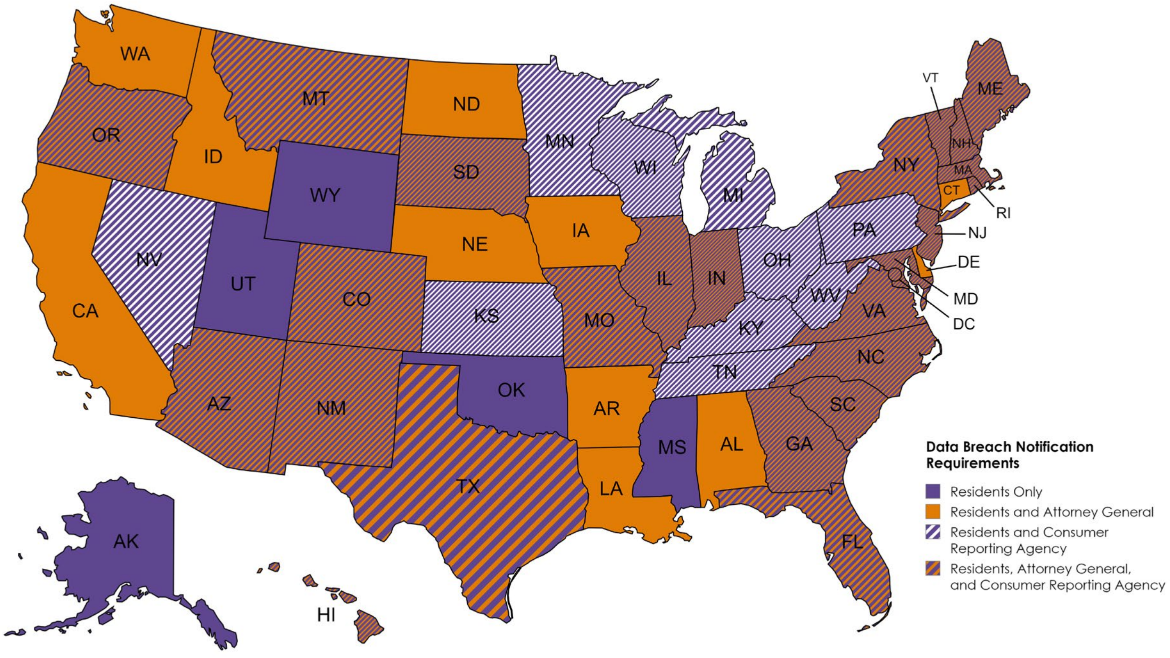
WA
MT
ND
VT
ME
OR
MN
SD
ID
WI
NH
NY
WY
MA
MI
CT
RI
IA
PA
NJ
NE
NV
OH
IL
IN
DE
UT
CO
WV
MD
CA
KS
MO
VA
KY
DC
NC
TN
OK
AR
SC
AZ
NM
MS
AL
GA
LA
TX
AK
FL
HI
Data Breach Notification Requirements
■ Residents Only
■ Residents and Attorney General
▨ Residents and Consumer Reporting Agency
▨ Residents, Attorney General, and Consumer Reporting Agency

Figure 11. Data Breach Notification Requirements.

While data security regulation is gaining traction among legislative bodies, existing laws remain fragmented in their approach to cybersecurity. Some industries, such as finance, healthcare, and insurance, have had sector-specific regulations for decades, while data protection laws outside these areas are relatively new and advancing slowly. At the same time, data collection, storage, and transfer have surged with the rise of connected and smart technologies, introducing unprecedented cybersecurity risks worldwide. Despite the growing volume of personal data collected daily, federal data security measures for critical infrastructure sectors have been slow to develop. Recognizing this gap, some states have begun implementing broad data privacy regulations.

Certain jurisdictions have enacted comprehensive data security laws applying to all entities that collect or manage personal data. While many states only mandate "reasonable" security measures, others specify actions data collectors must take to avoid liability for breaches. A few states have adopted liability frameworks requiring data collectors to ensure third-party recipients also safeguard shared data, preventing careless transfers that could expose sensitive information.

States also vary significantly in their restrictions on selling personally identifiable information (PII). Some impose no limits, while others regulate how PII is exchanged within the data marketplace. Many states with consumer data privacy laws grant individuals the right to request data deletion, which may extend to instances where PII has been sold to third-party vendors. However, the impact of these provisions on interstate data trading remains uncertain, raising legal questions about data ownership and whether personal data should be treated as "property" rather than an extension of identity.

Data breach notification laws are similarly inconsistent nationwide. States differ on whether data managers must notify affected individuals, state authorities, or consumer reporting agencies, and no clear regional patterns exist for analyzing these variations. Many states impose threshold requirements for breach notifications, but these thresholds vary widely. In states without consumer data protection laws, residents lack guaranteed rights to access, control, delete, or opt out of data collection altogether. However, data privacy should not be dictated by geography; an individual's personal information remains the same whether stored in Maine, Arizona, or elsewhere. Yet, differing security measures, collection practices, and breach notification laws create disparities in consumer protections and complicate compliance for data managers in critical infrastructure sectors. Given these challenges, further research is needed to assess the compatibility of emerging data security regulations and to establish a more unified and coordinated national approach to cybersecurity governance.

4.3 Publications, Presentations, and Demonstrations

The foregoing research has been produced publishing the above discussed information. More details on our work above can be found in the following:

Journal Papers – Peer-reviewed:

- Khandakar, A., M. Uddin, T. Hockstad, L. Khan, M. Rahman, M. Chowdhury, M. Salek, B. Thuraisingham, S. Jones. "Retrieval Augmented Generation-Based Large Language Models for Bridging Transportation Cybersecurity Legal Knowledge Gaps,"(under review by the Transportation Research Record Editorial Board);
- Hockstad, T., M. Rahman, M. Chowdhury, K. Akbar, M. Uddin, L. Khan. "Data Security & Privacy Regulation in the U.S.: A 50-State Legislative Survey," (under review by the Transportation Research Record Editorial Board);
- Hockstad, T., M. Rahman, S. Jones, M. Chowdhury. 2024. "A Regulatory Gap Analysis in Transportation Cybersecurity and Data Privacy." *Transportation Journal* 64(1): e12036. https://doi.org/10.1002/tjo3.12036.

Conference Presentations:

- Khandakar, A., M. Uddin, L. Kahn. "Bridging Legal Knowledge Gaps in Cybersecurity for Connected and Automated Transportation Systems with Large Language Models," TraCR Annual Conference, Greenville, SC, May 5-6, 2024;
- Khandakar, A., M. Uddin, L. Kahn, T. Hockstad, M. Rahman, M. Chowdhury. "Mitigating Hallucinations in Transportation Cybersecurity Legislation Analyses," Transportation Research Board 104th Annual Meeting, Washington, D.C., January 5-9, 2025;
- Hockstad, T., A. Khandakar, M. Uddin, "A Regulatory Gap Analysis in Transportation Cybersecurity and Data Privacy," TraCR Annual Conference, Greenville, SC, May 5-6, 2024;
- Hockstad, T., M. Rahman, S. Jones, M. Chowdhury, L. Khan. "Resolving Legislative Gaps in Transportation Cybersecurity Policy," Future of Transportation Summit, Washington, D.C., August 13-15;
- Hockstad, T. and J. Fisher, "Grid Modernization and Cybersecurity: Policy Implications for Electric Vehicle Infrastructure," Transportation Research Board 104th Annual Meeting, Washington, D.C., January 5-9, 2025;
- Hockstad, T., M. Rahman, M. Chowdhury, K. Akbar, M. Uddin, L. Khan. "Data Security & Privacy Regulation in the U.S.: A 50-State Legislative Survey," Transportation Research Board 104th Annual Meeting, Washington, D.C., January 5-9, 2025;
- Thomas, O., M. Salek, J. Tine, M. Rahman, T. Hockstad, M. Chowdhury. "Cybersecurity in Transportation Systems: Policies and Technology Directions," Transportation Research Board 104th Annual Meeting, Washington, D.C., January 5-9, 2025 (Accepted for presentation).

Technology Demonstration:

Latifur Khan, Ashrafi Akbar, Md Nahiyan Uddin, and Trayce Hockstad conducted a live demonstration of TraCR AI – a large language model trained on current cybersecurity legislation across the U.S. at the 2024 TraCR Annual conference in Greenville, SC on May 6, 2024.

CHAPTER 5

Conclusions

This project evaluates the current state of U.S. transportation cybersecurity policy by comparing domestic regulations and industry concerns with international models to identify gaps and propose future directions. Using a trained LLM and the TraCR AI platform, researchers analyzed legislative inconsistencies, developed new data visualizations, and incorporated industry survey feedback. The study highlights a persistent disconnect between practitioner needs and existing U.S. regulatory frameworks, emphasizing that U.S. policies lack the coherence and scope seen in international counterparts. Despite some state-level progress, the absence of unified federal guidance results in fragmented cybersecurity protections across the transportation sector. The research underscores the value of retrieval-augmented LLMs for pinpointing legislative gaps and improving response accuracy, although additional tools are needed for cross-jurisdictional comparison. Moving forward, the team aims to enhance legal data integration, ensure reliable AI-generated insights, and support policymakers with decision tools and updated policy guidance that address evolving cybersecurity threats.

5.1 Technology Transfer

This project compared what researchers have learned about domestic transportation policy and industry cybersecurity concerns with international models to assess the current status and identify paths forward. Specifically, this research considered: 1) how do international cybersecurity regulations compare to those in place in the U.S. with regard to scope and efficacy, and 2) how does industry insight inform domestic transportation security policy development? To answer these questions, the team trained this LLM as an expert on an expanded database of legislation and then utilized TraCR AI to identify inconsistencies and gaps between national and international cybersecurity policies. The team then created new ways of visualizing this data, the product of which will be disseminated through the TraCR Technology Transfer program. Researchers will include the industry expertise and feedback and consider those survey results alongside the policy analysis. Finally, the results of this study will be disseminated in an updated policy guidance report as well as a decision-making support tool.

5.2 Future Directions

This research seeks to build on these first-year efforts to enhance transportation cybersecurity policy in the U.S. by providing detailed analyses of existing regulations, anticipating changes in forthcoming legal updates, and drawing on international examples to supplement gaps in the domestic schematic. The team also intends to compare the results of the legislative analysis with responses from two surveys administered to different stakeholders in the transportation sector. These surveys are aimed at identifying the cybersecurity needs and concerns of practitioners who face on-the-ground cyberthreats every day. Prior survey work indicates that a gap exists between what experts identify as pressing concerns in the industry and what exists as regulatory guidance. This research to date also suggests that the U.S. model is insufficient to address existing threats and can be enhanced by drawing from international examples.

Regarding TraCR AI, researchers will ensure factual accuracy in the dataset's generated responses by using comprehensive validation methods. These methods assess both questions and outputs for adherence to established standards. These observations indicate that identical queries can yield varying responses, even when referencing the same set of retrieved nodes in the RAG pipeline. One effective approach to address this issue is leveraging LLMs for answer distillation, selecting the most accurate response from multiple generated versions. Subsequently, valuable metrics for assessing open-ended question-answer generation can be explored by analyzing natural language nuances, especially in policy and legal contexts.

5.3 Closing Remarks

Cybersecurity policy in the U.S. remains fragmented, particularly in the transportation sector, with no overarching federal guidance to unify efforts across states. This lack of cohesive policy has led to inconsistent approaches, as states vary widely in how they regulate and protect transportation infrastructure from cyber threats. Some states have begun addressing cybersecurity for critical transportation systems like highways, ports, and EV charging networks, but there is little agreement on best practices, leaving gaps in national security. Without clear federal standards, both state-level inconsistencies and private sector challenges in securing transportation systems continue to pose significant risks. This policy vacuum highlights the need for a coordinated federal response to ensure a uniform approach across jurisdictions.

One way this might be accomplished is by aiding policymakers in the analysis of existing legislation to understand existing legal regulations as well as to draft future laws. These findings show that state-of-the-art LLMs struggle with specialized queries on recent legislation, leading to potential misidentification of legislative gaps. However, RAG-powered LLMs offer promise in identifying gaps in transportation cybersecurity by integrating recent legal advancements for factually accurate responses. The framework supports concept-based and state-specific analysis to uncover missing legislative elements, but requires an additional module to compare responses across legislative bodies. Future work will enhance user guidance and seamless legal data integration. By mitigating hallucinations, this approach strengthens legislative analysis and supports evolving regulatory landscapes.

REFERENCES

Belcher, S. 2020. In the Transit Industry Prepared for the Cyber Revolution? Policy Recommendations to Enhance Surface Transit Cyber Preparedness. Mineta Transportation Institute. p. 15.

Bowcut, S. 2023. Cybersecurity in the transportation industry. Cybersecurity Guide. (https://cybersecurityguide.org/industries/transportation/#:~:text=According%20to%20Cybertalk.org%2C%20between,the%20brunt%20of%20this%20trend).

Chowdhury, A., G. Karmakar, J. Kamruzzaman, A. Jolfaei, and R. Das. 2020. Attacks on self-driving cars and their countermeasures: A survey. *IEEE Access*, 8, pp.207308-207342.

Taeihagh, A. and H.S.M. Lim. (2019) Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport reviews*, *39* (1), pp.103-128.

Cybersecurity and Infrastructure Security Agency. 2020. Transportation Systems Sector Cybersecurity Framework Implementation Guide. (https://www.cisa.gov/resources-tools/resources/transportation-systems-sector-cybersecurity-framework-implementation-guide.)

Dempsey, J. 2021. Cybersecurity Law Fundamentals. p. 276. IAPP, Portsmouth, NH.

Devlin, J., M.W. Chang, K. Lee, K. Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* pp. 4171–4186. Association for Computational Linguistics.

DiFurio, D. 2023. 4 States Passed Nearly Half of All New Cybersecurity Laws Enacted Across the US in 2022. DRATA. (https://drata.com/blog/4-states-passed-nearly-half-of-new-cybersecurity-laws).

Dutta, N., N. Jadav, S. Tanwar, H. Kumar, D. Sarma, E. Pricop. 2022. *Cyber Security: Issues and Current Trends*. ch. 9. Springer, New York.

Dziri, N., S. Milton, M. Yu, O. Zaiane, S. Reddy. 2022. On the Origin of Hallucinations in Conversational Models: Is it the Datasets or the Models? *ArXiv.*/abs/2204.07931.

Grzadkowska, A. 2018. Transportation is now the third most vulnerable sector exposed to cyber attacks. *Insurance Business*. (https://www.insurancebusinessmag.com/us/news/cyber/transportation-is-now-the-third-most-vulnerable-sector-exposed-to-cyberattacks-106900.aspx).

Hockstad, T., M. Rahman, M. Chowdhury, K. Akbar, M. Uddin, L. Khan. 2025. Data Security & Privacy Regulation in the U.S.: A 50-State Legislative Survey. *Transportation Research Record* (under review).

Hu, Y., M. Hosseini, E. Skorupa Parolin, J. Osorio, L. Khan, P. Brandt, V. D′Orazio. 2022. ConfliBERT: A Pre-trained Language Model for Political Conflict and Violence. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* pp. 5469– 5482. Association for Computational Linguistics.

Huang, L., W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, T. Liu. 2023. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ArXiv*. /abs/2311.05232.

Khan, M.A., H. El Sayed, S. Malik, M.T. Zia, N. Alkaabi, and J. Khan. (2022) A journey towards fully autonomous driving-fueled by a smart communication system. *Vehicular Communications*, *36*, p.100476.

Khan, Z., S.M. Khan, M. Rahman, M. Islam, and M. Chowdhury. (2022) Deep Learning in Transportation Cyber-Physical Systems. In *Leveraging Artificial Intelligence in Engineering, Management, and Safety of Infrastructure* (pp. 331-347). CRC Press, Boca Raton, FL.

Khandakar, A., M. Uddin, T. Hockstad, L. Khan, M. Rahman, M. Chowdhury, M. Salek, B. Thuraisingham, S. Jones. 2024. Bridging Legal Knowledge Gaps in Cybersecurity for Connected and Automated Transportation Systems with Large Language Models. TraCR Annual Conference. Greenville, SC.

Khandakar, A., M. Uddin, T. Hockstad, L. Khan, M. Rahman, M. Chowdhury, M. Salek, B. Thuraisingham, S. Jones. 2025. 'Retrieval Augmented Generation-Based Large Language Models for Bridging Transportation Cybersecurity Legal Knowledge Gaps. *Transportation Research Record* (under review).

Khayyam, H., B. Javadi, M. Jalili, and R.N. Jazar. (2020) Artificial intelligence and internet of things for autonomous vehicles. Nonlinear Approaches in Engineering Applications: Automotive Applications of Engineering Problems, pp. 39-68.

Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, D. Kiela. 2021. Retrieval-Augmented Generation for Knowledge Intensive NLP Tasks. *NeurIPS Proceedings*. University College, London.

Lukings, M., A. H. Lashkari. 2022. *Understanding Cybersecurity Law and Digital Privacy: A Common Law Perspective*. p. 59. Springer, New York.

National Automobile Dealers Association. 2021. Personal Data in Your Car. (https://fpf.org/wp-content/uploads/2017/01/consumerguide.pdf).

PUBLIC LAW 117–103—MAR. 15, 2022.
https://www.congress.gov/117/plaws/publ103/PLAW-117publ103.pdf.

Sun, H., G. Xu, J. Deng, J. Cheng, C. Zheng, H. Zhou, N. Peng, X. Zhu, M. Huang. 2021. On the Safety of Conversational Models: Taxonomy, Dataset, and Benchmark. *ArXiv*. /abs/2110.08466.

Zhang, T., S.G. Patil, N. Jain, S. Shen, M. Zaharia, I. Stoica, J.E. Gonzalez. 2024. RAFT: Adapting Language Model to Domain Specific RAG. *ArXiv*./abs/2403.1013.

Zhang, Y., Y. Li, L. Cui, D. Cai, L. Liu, T. Fu, X. Huang, E. Zhao, Y. Zhang, Y. Chen, L. Wang, A.T. Luu, W. Bi, F. Shi, S. Shi. 2023. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *ArXiv*. /abs/2309.01219.

APPENDIX

TraCR 2024 State Survey

1. How well-informed do you personally feel about transportation cybersecurity issues and challenges?
   - ○ Not at all informed
   - ○ A little informed
   - ○ Somewhat informed
   - ○ Well-informed
   - ○ Very well-informed

2. How well-informed do you think your agency is about transportation cybersecurity issues and challenges?
   - ○ Not at all informed
   - ○ A little informed
   - ○ Somewhat informed
   - ○ Well-informed
   - ○ Very well-informed
   - ○ Not Sure

3. Do you feel that cybersecurity issues are a major concern for your agency?
   - ○ Strongly agree.
   - ○ Agree.
   - ○ Neither agree nor disagree
   - ○ Disagree.
   - ○ Strongly disagree.

4. How often does your organization consider cybersecurity when implementing, integrating, or upgrading transportation-related systems?
   - ○ All the time.
   - ○ Most of the time.
   - ○ Sometimes.
   - ○ Barely.
   - ○ Never.

5. Does your agency have specialized people or a department that focuses primarily on cybersecurity?
   - ○ Yes.
   - ○ No.

6. Which cybersecurity measures are currently implemented in your organization? (Select all that apply)
   ○ Firewalls (controls and monitors network traffic to block unauthorized access.)
   ○ Anti-virus Software (detects, prevents, and removes malware from systems.)
   ○ Intrusion detection systems (monitors network activities for malicious actions and gives alerts on potential intrusions.)
   ○ Multi-Factor Authentication (requires multiple forms of verification to access systems.)
   ○ Regular Security Audits by an internal or external entity(evaluates and improves an organization's security practices and compliance.)
   ○ Incident response plan (outlines procedures to handle and recover from security incidents.)
   ○ Employee cybersecurity training (educates staff on cybersecurity best practices and threat prevention.)

7. Has your agency ever experienced the following?

   i. A phishing attack (An attack where attackers impersonate trusted entities to deceive individuals into sharing sensitive information)
   [ ] Yes [ ] No [ ] Not sure

   ii. A ransomware attack (Malicious software that prevents access to computer systems and demands a ransom for restoration)
   [ ] Yes [ ] No [ ] Not sure

   iii. A denial-of-service attack (An attack designed to overwhelm systems with excessive traffic, making them inaccessible to legitimate users)
   [ ] Yes [ ] No [ ] Not sure

   iv. A data breach (An incident where unauthorized individuals access, steal, or disclose sensitive data without permission)
   [ ] Yes [ ] No [ ] Not sure

   v. A malware attack (Use of malicious software to damage, disrupt, or gain unauthorized access to computers, networks, or data)
   [ ] Yes [ ] No [ ] Not sure

   vi. A computer virus (A type of malware that attaches to files or programs in a computer system, which spreads across the system, causing damage or disruption)
   [ ] Yes [ ] No [ ] Not sure

   vii. Any other attack not mentioned above
   Please Specify: ____

8. What are your organization's major challenges while ensuring security for transportation-related hardware or software? (Check all that apply.)

   ○ Lack of understanding of how vendors from outside your agency ensure security for

their products.
[ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- The tools to address our cybersecurity problems do not exist
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- Integrating software and/or hardware from different vendors and a lack of collaboration on security.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- Lack of training for the transportation workforce in cybersecurity issues
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- Lack of financial resources to retain in-house cybersecurity experts.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- Lack of financial resources to purchase cybersecurity tools and consulting from outside your agency.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- Too many other pressing priorities for my agency, such as road and transit building, operations and maintenance.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- The public is not concerned enough about cybersecurity for it to be an issue we focus on.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- Elected officials in my state are not concerned enough about cybersecurity for it to be an issue we focus on.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- Cyberthreats change so frequently it is hard for our agency to keep up with new developments.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- The legal basis and best practices for implementing cybersecurity solutions are lacking.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- Cybersecurity tools to accomplish our goals have not been developed yet.
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- A lack of standards and policies for agencies to follow about transportation cybersecurity
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

- A lack of urgency about addressing cybersecurity issues
  [ ] Major problem [ ] Minor problem [ ] Not a problem [ ] Not sure

○ Others (Please tell us what they are) ____

9. How likely is it for you or your organization to collaborate with our national University Transportation Center (UTC), which focuses on transportation cybersecurity, to help you and your agency to better understand your security-related needs?

○ Very likely.
○ Likely.
○ Somewhat likely.
○ A little likely.
○ Not likely at all.

10. Which initiatives are your agency or the agencies in your region currently undertaking or planning? (Please check all that apply)

i. Installing adaptive traffic signals
[ ] Currently undertaking [ ] Planning to undertake [ ] Not planning to undertake

ii. Adding infrastructure to accommodate connected vehicles, such as roadside communication infrastructure and computing resources
[ ] Currently undertaking [ ] Planning to undertake [ ] Not planning to undertake

iii. Adding cameras that can identify vehicles, for example, red light cameras or speed cameras
[ ] Currently undertaking [ ] Planning to undertake [ ] Not planning to undertake

iv. Implementing or further developing electronic tolling systems that collect information from vehicles
[ ] Currently undertaking [ ] Planning to undertake [ ] Not planning to undertake

v. Implementing a vehicle-miles-traveled road user fee that collects information from vehicles
[ ] Currently undertaking [ ] Planning to undertake [ ] Not planning to undertake

vi. Exchanging sensitive data with other agencies (for example, your state department of transportation) or private firms
[ ] Currently undertaking [ ] Planning to undertake [ ] Not planning to undertake

vii. Conducting travel demand surveys that include personal information
[ ] Currently undertaking [ ] Planning to undertake [ ] Not planning to undertake

viii. Are you developing and implementing any other intelligent transportation systems that involve confidential data (Please tell us what these may be)

____

11. Are cybersecurity issues making it more difficult to pursue any of the initiatives just listed? If so, tell us whether cybersecurity is a major problem, a minor problem, or not a problem,

or whether your agency does not plan to pursue the initiative at all.

i. Installing adaptive traffic signals
[ ] Cybersecurity is a major problem [ ] Cybersecurity is a minor problem
[ ] Cybersecurity is not a problem. [ ] We will not be pursuing this

ii. Adding infrastructure to accommodate connected vehicles, such as roadside communications infrastructure and computing resources
[ ] Cybersecurity is a major problem [ ] Cybersecurity is a minor problem
[ ] Cybersecurity is not a problem. [ ] We will not be pursuing this

iii. Adding cameras that can identify vehicles, for example, red light cameras or speed cameras
[ ] Currently undertaking [ ] Planning to undertake [ ] Not planning to undertake

iv. Implementing or further developing electronic tolling systems that collect personal information from vehicles
[ ] Cybersecurity is a major problem [ ] Cybersecurity is a minor problem
[ ] Cybersecurity is not a problem. [ ] We will not be pursuing this

v. Implementing a vehicle-miles-traveled road user fee that collects personal information from drivers
[ ] Cybersecurity is a major problem [ ] Cybersecurity is a minor problem
[ ] Cybersecurity is not a problem. [ ] We will not be pursuing this

vi. Exchanging sensitive data with other agencies (for example, your state department of transportation) or private firms
[ ] Cybersecurity is a major problem [ ] Cybersecurity is a minor problem
[ ] Cybersecurity is not a problem [ ] We will not be pursuing this

vii. Conducting travel demand surveys that include personal information
[ ] Cybersecurity is a major problem [ ] Cybersecurity is a minor problem
[ ] Cybersecurity is not a problem. [ ] We will not be pursuing this

viii. Developing and implementing other intelligent transportation systems that involve confidential data
[ ] Cybersecurity is a major problem [ ] Cybersecurity is a minor problem
[ ] Cybersecurity is not a problem [ ] We will not be pursuing this

12. What are the ways our national UTC can help your agency? (Check all that apply.)
For each, tell us whether it would be very helpful, somewhat helpful, or not very helpful.

i. Publishing a newsletter about transportation cybersecurity issues
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

ii. Creating new transportation cybersecurity products and practices
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

iii. Creating new transportation cybersecurity products and practices
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

iv. Developing new online transportation cybersecurity training programs for working professionals.
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

v. Creating new transportation cybersecurity educational materials for colleges and universities
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

vi. Creating new transportation cybersecurity educational materials for K-12 schools
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

vi Developing a shared platform for incident reporting and threat intelligence sharing among transportation agencies
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

vii. Convening a regional cybersecurity task force that includes representatives from various sectors. For instance, (government, private sector, academia).
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

viii. Disseminating information on OT/ITS cybersecurity best practices and standards.
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

ix. Creating a validation testbed to test new cybersecurity products and practices.
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

x. Performing feasibility studies to gauge the benefits and costs of cybersecurity.
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

xi. Developing standards, guidelines and best practices for implementing cybersecurity.
[ ] Very helpful [ ] Somewhat helpful [ ] Not very helpful

xii. Others. (Please tell us that they are) ____

13. Do you feel the organizations listed below are addressing transportation cybersecurity issues well?

i. Your MPO

○ Not at all well
○ A little well
○ Somewhat well
○ Well
○ Very well
○ Not sure

ii. The local governments (such as municipalities and counties) in your region?

- ○ Not at all well
- ○ A little well
- ○ Somewhat well
- ○ Well
- ○ Very well
- ○ Not sure

iii. Your state government and state department of transportation?

- ○ Not at all well
- ○ A little well
- ○ Somewhat well
- ○ Well
- ○ Very well
- ○ Not sure

iv. The federal government and the US Department of Transportation (USDOT)?

- ○ Not at all well
- ○ A little well
- ○ Somewhat well
- ○ Well
- ○ Very well
- ○ Not sure

14. Which of the following resources does your DOT dedicate to cybersecurity? (Check all that apply)

- ○ At least one specialist internal staff person (such as a full-time IT expert)
- ○ Consultation with an outside government agency (like your state DOT or the USDOT)
- ○ Consultation with a private firm, such as an IT consulting firm
- ○ Cybersecurity software or hardware purchased from a third-party vendor.
- ○ Any other: ____

15. Have you experienced any cybersecurity issues in any of the following applications you may have? If so, tell us which ones. (Check all that apply.)

- ○ Electronic tolling.
- ○ Connected and/or self-driving vehicles.
- ○ “Smart” roadside infrastructure.
- ○ Red light cameras.
- ○ A vehicle-miles traveled fee.
- ○ Others. (Please tell us what they are). ____________________________________

16. Do you specialize in cybersecurity at your agency, or is most of your time spent on other issues?

- ○ Specialize in cybersecurity.
- ○ Work mostly on other issues.

17. What is your best source of information about cybersecurity issues? (Check one.)
    ○ Popular media.
    ○ Word of mouth.
    ○ The ITE.
    ○ AASHTO.
    ○ TRB.
    ○ University Transportation Centers.
    ○ Other. (Please tell us what it is.)
    ○ I am not well-informed about transportation cybersecurity issues from any source.

18. How often does your agency conduct cybersecurity training and awareness programs for staff?
    [ ] Very often [ ] Often [ ] Not too often [ ] Rarely [ ] Never

19. Are you aware of any specific regulatory requirements or standards that your agency must comply with regarding cybersecurity? If yes, could you please provide some examples that do not fall under any NDA?

    ____

20. What is your job title?

    ____

21. How long have you served at your agency?
    ○ Under 2 years
    ○ 2-5 years
    ○ 5-10 years
    ○ 10 years to 20 years
    ○ Over 20 years

22. What would you say your primary responsibility at your agency is?
    ○ IT
    ○ Cybersecurity
    ○ Planning
    ○ Engineering
    ○ Other (tell us what it is) _____________________

23. Is there any question missing that you think we should ask?
    ____