APPENDIX A: U.S. DOT SBIR Proposed FY24.2 Phase I Research Topic

Complete Streets Artificial Intelligence Initiative
Generating, Integrating, and Activating Data for Mobility

The Complete Streets Artificial Intelligence (CSAI) Initiative is a multi-phase effort to develop a suite of powerful new decision-support software tools for state, local, and tribal transportation agencies that assists in the siting, design, and deployment of Complete Streets. The initiative is nested within the U.S. Department of Transportation's (DOT) Small Business Innovation Research (SBIR) program. DOT anticipates funding small businesses for up to three phases of activity. These phases, in their entirety, will propel early-stage concepts out of research and development and into nationwide products for public and private use. Small businesses are expected to utilize teams with expertise in transportation planning, data science, and software development. The anticipated funding for each phase is described below.

Phase	Period of Performance	Estimated Funding per Award	Estimated Number of Awards	Estimated Total Funding
I	6 months	$200k	10	$2 million
II	12 to 18 months	$1.8 million	5	$9 million
Follow-on funding	TBD	$2 million	2	$4 million
				$15 million

Background:

Complete Streets are "streets and networks that prioritize safety, comfort, and connectivity to destinations for all people who use the street network."[1] This includes people who walk, bike, use micromobility, or use mobility assistance devices (collectively, "active transportation"); people who use public transportation and on-demand paratransit services; children, older individuals, and individuals with disabilities; people who operate freight and delivery vehicles; and people who drive and ride in personal vehicles.

The CSAI Initiative groups "data" into three categories: Infrastructure, Traveler Behavior/Safety, and Context.

- Infrastructure: Complete Streets treatments are reflected in Infrastructure, the primary category that tells us *how complete* a street or network is. Complete Streets treatments provide localized benefits such as safe pedestrian crossings, reliable transit service, convenient bicycle facilities, and efficient freight routes. The benefits of Complete Streets treatments accrue at scale as *complete networks* that provide continuous and connected routes, especially for people using active and public transportation. Complete networks increase opportunities for people to travel by active and public transportation.

- Traveler Behavior/Safety: Traveler Behavior/Safety data help us know *what kind* of Complete Streets treatments to implement. Not every road can serve each transportation mode equally.

[1] https://highways.dot.gov/complete-streets

Prioritizing different modes on different parts of the network requires an understanding of travelers' current and future needs and behaviors, including their travel modes, trip purposes, and the behavior and safety of other travelers.

- Context: Context data inform both *where* and *what kind* of Complete Streets treatments to implement. Understanding context allows practitioners to allocate limited rights-of-way to prioritize the modes that best serve the needs of those traveling to and within that context. This requires an understanding of land use, key destinations, urban design, local and regional population trends, and other contextual characteristics that are the settings for travel.

To achieve the safety, equity, economic, and climate benefits of a truly multimodal transportation system, the people who shape its ongoing evolution need rigorous analytical tools and methods. The solutions resulting from the CSAI Initiative aim to enable practitioners to better plan, design, build, manage, and evaluate infrastructure and operational strategies that increase people's opportunity to use active and public transportation, while facilitating safe and efficient freight and vehicular travel as well.

Decision Support Tool:

This SBIR topic, the CSAI Initiative, will leverage novel approaches in data science, machine learning, and artificial intelligence to accelerate implementation of Complete Streets and the complete networks they support. Awardees will use improved data collection, data fusion, and analysis methods to build tools that help identify network gaps, test and/or propose solutions, and display the impact of network and operational changes on mobility, access and accessibility, equity, safety, and other characteristics. Awardees are expected to utilize big data, internet-of-things, artificial intelligence, machine learning, and computer vision, among other methods. The resultant tools will support state, local, and tribal practitioners and other users by providing quantitative, segment-scale and network-scale insights to help shape decision making for planning, designing, and implementing Complete Streets projects.

At the completion of the multi-phased effort, the DOT expects functional software tools that can, at a minimum, identify, represent, and analyze physical transportation infrastructure. The tool(s) will represent this infrastructure as a high-fidelity graph network or a similar format that reflects the physical and operational details that impact the safety and utility of active transportation, transit, freight and delivery, and personal vehicles.

The following examples illustrate just two of many different potential capabilities of a final tool(s):

Example #1:

- Prompt: Identify existing routes for pedestrians (people walking and using mobility devices) to get from their homes to a grocery store; suggest infrastructure improvements to increase by 15 percent the number of people who can reach a grocery store by such modes within 20 minutes.

- Output: Tool identifies absent sidewalk segments, missing curb ramps, and indirect routing on otherwise suitable infrastructure that prevents a 20-minute trip by pedestrians. Tool proposes changes to the network to increase by 15 percent the number of pedestrians who can complete the trip in 20 minutes. Changes include new sidewalks, retrofitted curb ramps, and new crosswalks. Tool offers user choices to fine tune automated solutions, and outputs the resultant impact on the population.

Example #2:

- Prompt: Propose infrastructure and operational changes to provide low-traffic-stress ("all-ages-and-abilities")1F[2] bicycle facilities that afford access between the residence of 50 percent of the population and 50 percent of civic, institutional, and employment destinations.

- Output: Tool identifies roadway segments and other alignments with proposed changes, including new and/or upgraded infrastructure such as bike markings, lanes, and physical separation to reduce conflict points; bike signals, intersection treatments, and traffic calming; and operational changes such as reduced speed limits and signal timing changes that achieve low-traffic-stress access between 50 percent of the population and 50 percent of noted destinations.

CSAI Initiative Phasing:

This multi-phased effort is designed to scale the volume and completeness of data elements, the geographic coverage of those data, and the analytical capabilities of the final products. The anticipated progression of this initiative is presented in Table 1. The requirements beyond Phase I are subject to change in response to Phase I results. The purpose of outlining an intended progression is to provide insight into anticipated expectations and outcomes beyond Phase I. These are presented to help interested parties more fully contextualize the Initiative.

Table 1: Anticipated Phase Progression

Phase	Goals	Expected Functionality & Outcomes
Phase I	Demonstrate the potential of novel data generation and processing methods to produce useful content for proposed Complete Streets analysis capabilities.	1) Select one (1) contiguous geographic area representing both urban and suburban development patterns and different roadway contexts (residential, commercial, industrial, civic, recreational/natural, etc.) and comprising at least 250,000 people. 2) Generate data elements for the fullest extent of the geographic area practicable from among each of the three primary data categories: Infrastructure, Traveler Behavior/Safety, Context, (see Table 2). 3) Describe source input data, data processing pipeline, data validation, and planned Phase II analytical capabilities.2F[3]
Phase II	Scale promising novel data generation and processing methods to greater geographic extents and completeness; implement Complete	1) Expand coverage to include five new, non-contiguous geographic areas representing both urban and suburban development patterns and different roadway contexts (residential, commercial, industrial, civic, recreational/natural,

[2] https://highways.dot.gov/safety/proven-safety-countermeasures/bicycle-lanes

[3] While proprietary information should be protected, some explanation of sources and methods is expected.

	Streets analysis capabilities in a functioning software solution.	etc.) and comprising at least 250,000 people each (1.5 million total, including Phase 1 geography). 2) Generate additional data elements and higher fidelity data, building on Phase I, for the full extent of the expanded geographic areas, from among each of the three primary data categories: Infrastructure, Traveler Behavior/Safety, Context (see Table 2). 3) Update the description of the source input data, data processing pipeline, data validation, and anticipated future analytical capabilities.[3] 4) Build a working beta software with functioning automated and query-based analytical capabilities using the generated data.
Follow-on funding	Scale software to national scale.	1) Expand coverage to greater than 50 percent of the U.S. population. 2) Generate additional data elements and higher fidelity data, building on Phase II, for the full extent of the geographic area, from among each of the three primary data categories: Infrastructure, Traveler Behavior/Safety, Context (see Table 2). 3) Update the description of the source input data, data processing pipeline, data validation, and analytical capabilities.[3] 4) Build a fully functioning software application of the automated and query-based analytical tool using the generated data.

Data Gaps and Analysis:

The 2022 FHWA Complete Streets Report to Congress identifies data collection and analysis as the first of five critical opportunity areas to advance Complete Streets efforts. Basic data about parts of the transportation network like roadway assets (such as bike lanes and sidewalks) and traveler volumes (such as counts of pedestrians and bicyclists) are often incomplete or unavailable. The available data can also be biased towards certain populations, income levels, or geographies. National crash data sources represent an incomplete picture of safety and underrepresent injury crash events, especially for pedestrians and bicyclists.

The CSAI Initiative categorizes data elements into three categories: Infrastructure, Traveler Behavior/Safety, and Context as described above. Table 2 provides a list of priority data elements that are most relevant to Complete Streets. Some elements may need to be generated or enhanced because nationwide data is unavailable or incomplete. These are the data for which novel identification and processing methods are most needed and are expected to be a significant focus of the effort throughout the

Phases. Other elements within the table are identified because there are nationally available data of sufficient quality to support Complete Streets planning and analysis objectives.

The inclusion of as many data elements as practicable is desired, given the parameters of the performance period and funding for each phase. Offerors are not required to include all listed data elements and may add other elements at their discretion. A balance of data volume, fidelity, accuracy, and extent is needed to ensure the development of tools that can perform the highest impact Complete Streets analyses possible. Additional data sets and other resources are available to help offerors consider their choices in approaching this effort and understand how practitioners would ultimately use the final tools created. The additional datasets and other resources include links to nationally available datasets and Federal resources on relevant Complete Streets data, analysis, and safety topics.

These resources can be found at https://its.dot.gov/csai/ in the *Datasets & Other Resources* section at the bottom of the page.

Table 2: Example Data Elements

Infrastructure

Data elements to generate or enhance:

- Roadway features
  - Centerline roadway network
  - Travel direction, number of lanes, lane markings and widths, shoulder width, edge condition barriers, guardrails
  - Mid-block, intersection-approach, and through-intersection conditions (turn lanes, intersection markings, intersection format e.g., T, X, multi-leg, roundabout)
- Signals, markings, signs (including posted speed limits)
- Signal timing and phasing
- Sidewalks, crosswalks, driveways, curb ramps, medians, refuges, curb extensions
- Transit stop accessibility (ADA) features (ramp deployment clear space, tactile warning strips)
- Bicycle facilities and facility type (unmarked, unseparated, and separated facilities; intersection-approach and through-intersection treatments)
- Speed management countermeasures (speed humps, tables, and raised intersections)
- Sidewalk surface condition (roughness, cracking, heaving)
- Fixed obstructions within pedestrian and bicycle facilities (utility poles, signposts)
- On-street curb management (parking, loading)
- Multi-use paths and traversable public open spaces

Generally available data:

- Transit stops and routes
- Rail routes and crossings

Traveler Behavior/Safety

Data elements to generate or enhance:

- Vehicle volumes and types (cars, trucks, freight, other)
- Vehicle operating speeds and speed distributions
- Injuries, injury severity, and near misses
- Bicycle and pedestrian user and/or trip counts
- Transit ridership

Generally available data:

- Crash data (fatalities, location, crash rate, roadway classification and contexts, and crash severity)

Context

Data elements generate or enhance:

- Land use (residential, commercial, industrial, civic, academic, mixed-use)
- Tree cover / landscaping (including shade)
- Urban design / frontage (including adjacent site design and building setback)

Generally available data:

- Key destinations (healthcare, academic, civic, employment, retail-commercial, recreation)
- Housing and employment density
- Environment (surface heat index, green space, open space)
- Social determinants of health
- Socio-economic and demographic data
- Environmental Justice and transportation disadvantaged communities
- Topography

Expectations and Outcomes:

The CSAI Initiative seeks outcomes that use novel methodologies to fill data gaps and provide validation of processes to increase data coverage, fidelity, and accuracy; reduce biases; and fuse data sources in novel ways to generate more complete information. These methodologies may include applying computer vision and other artificial intelligence / machine learning approaches to a range of data sources and types. Data may include (but not be limited to) satellite, aerial, or street-level video and still photography; satellite, aerial, or street-level lidar data; sensor-based and crowdsourced data; vehicle probe data and telemetry, including shared micromobility and transit modes; multimodal volume/count data; and other sources and types.

The initiative is sequenced in phases to support collective progress in addressing different parts of this effort; see the narrative description of expected outcomes for each phase below, as a complement to Table 1. As noted above, future phases may change to reflect the results of prior phases.

In all phases:

- The data used should not pre-date 2020, unless being used to inform or validate broader data sources, or unless a more recent source is unavailable. Offerors should keep in mind that many travel behaviors and other patterns were atypical during the height of the COVID-19 pandemic and strive to choose data that is representative of the most current and/or typical context.

- The data generation process should anticipate and be able to accommodate regular updates to reflect evolving real-world conditions. Planned update frequency should be annual, at least, and preferably seasonal (e.g., spring, summer, fall, and winter).

- The data and methodologies generated from this initiative that reflect existing bicycle, pedestrian, and accessibility infrastructure shall be hosted publicly in an open and accessible format, where possible, to facilitate immediate use to improve access and safety. Such data should use prevailing data standards, where appropriate.

- Outputs from the tool should be built upon and/or easily interface with existing, widely used planning and mapping platforms and tools; this expectation is intended to facilitate integration with existing data analysis processes, network quality assessments or scoring rubrics for specific modes, and support broad adoption and use of the tool and generated data among practitioners.

- Data should be maintained in a secure environment, with proper administrative and technical safeguards and measures in place to prevent unauthorized access, disclosure, acquisition, destruction, use, or modification. Any Personally Identifiable Information (PII) and other sensitive data should be kept confidential. Offerors should be mindful of reidentification and other risks of data sharing, and have an approach to minimize these risks in the final product.

Expected Phase I Outcomes:

Phase I will focus on generating data layers that geographically represent elements from among those listed in Table 2. To help offerors make choices about which elements, processes, and formats to employ, Phase I also requires awardees to outline a description of the intended analysis capabilities those data will support in future phases, if awarded.

The outcome of Phase I shall be a proof-of-concept report, as well as the digital data layers that have been generated for the chosen geographic region in a standard geospatial format (e.g., GeoJSON). The proof-of-

concept report should include a list of the data that awardees have generated, a list of data they plan to generate if selected for Phase II, and the characteristics or features of those data; a description of methods used to generate the data, an evaluation of the scalability of those methods, and assessment of data coverage limitations and gaps; and, a transparent validation of data accuracy. Validation can be carried out remotely or digitally. The proof-of-concept report should include the required description of intended analytics capabilities the data will support.

Following the completion of Phase I, DOT anticipates hosting a virtual technology presentation to provide Phase I awardees the opportunity to share the data generated for areas within their selected geography. This event is expected to be open to the public and awardee participation in the event, if held, is voluntary and is not to be proposed in a Phase I proposal submission.

Expected Phase II Outcomes:

Small businesses selected to continue in Phase II will build out and add new data layers, achieve higher data fidelity and accuracy, and scale their methodologies for data generation to cover additional geographies. During Phase II, awardees will build a functioning beta software tool.

The tools shall enable practitioners and other users to explore each data layer individually to ensure transparency, explainability, and trust. Additionally, the tools are expected to enable users to add other data layers and types. The tools will use generated data, along with other available data sources, to perform analysis in response to queries from users. These queries may be for automated actions, such as detecting network gaps and proposing solutions, and may also allow users to fine tune queries with filters or parameters. The software is also expected to allow users to run analyses on specific, proposed changes to the transportation network. The resultant software tools are expected to facilitate network analysis for people who use active and public transportation and may feed data into other external tools or analyses. The tools are expected to allow users to visualize existing and proposed multimodal conditions and traveler outcomes at the segment- and network-scale.

DOT intends to host a second technology presentation towards the conclusion of Phase II. The purpose of the event is to provide Phase II awardees the opportunity to present capabilities of the software tool and its use of the data generated for areas within their selected geographies. Additional detail will be provided with instructions for submitting a Phase II proposal.

Expected effort outcomes from an additional follow-on funding agreement:

Additional follow-on funding agreement(s) may be awarded to support expanding Phase II work to a national scale covering at least 50 percent of the U.S. population. At the conclusion of this phase, the tool should be poised for commercialization to serve the needs of a broad cross-section of public, private, academic, and non-profit users.