AASHTO
AASHTO INNOVATION INITIATIVE

AASHTO Innovation Initiative

[Proposed] Nomination of Innovation Ready for Implementation

Sponsor

Nominations must be submitted by an AASHTO member DOT willing to help promote the innovation. If selected, the sponsoring DOT will be asked to promote the innovation to other states by participating on a Lead States Team supported by the AASHTO Innovation Initiative.

1. Sponsoring DOT (State): North Carolina

2. Name and Title: Arnav Jhala, Associate Professor of Computer Science/Clare Fullerton, NCDOT CLEAR PM

Organization: North Carolina State University/ NCDOT

Street Address: 890 Oval Dr. / 1 South Wilmington

City: Raleigh

State: NC

Zip Code: 27606

Email: ahjhala@ncsu.edu/ cefullerton@ncdot.gov

Phone: 8313324680/ 4046062675

Fax: Click or tap here to enter text.

Innovation Description (10 points)

The term "innovation" may include processes, products, techniques, procedures, and practices.

3. Name of the innovation:

Artificial Intelligence (AI)-driven Language Model for Construction Domain

4. Please describe the innovation.

Through NC DOT's CLEAR program, we have created a database of Lessons Learned and Best Practices that utilizes an AI-driven computational language model of construction-specific vocabulary. The AI model is trained on >1 million terms from construction textbooks, specification documents, feasibility studies, and other project documents. This language model enables better knowledge management, retrieval, and knowledge sharing within NC DOT employees over time and also potentially nationally across DOTs, contractors, and private sector construction projects.

5. What is the existing baseline practice that the innovation intends to replace/improve?

Previous lessons learned database design efforts have been manual entry of unstructured narratives of lessons-learned and best practices in a static database with a web-based querying interface. These are created at high cost but face challenges because they depend on getting buy-in from the end-users to embrace the database in terms of regular usage. The ultimate success of any lessons learned program depends heavily on getting the end-users to enter knowledge and access past wisdom to be applied on future projects. However, many a database has become defunct over the long run, leading to huge financial implications for organizations in terms of lost efforts and time in designing the database. To tackle such a problem, this research proposes a novel effort in using machine learning in the form of natural language processing created from neural language models to automatically and proactively push knowledge to end-users, thereby increasing the reliability of the CLEAR program.

6. What problems associated with the baseline practice does the innovation propose to solve?

There is a need to collect real-time information from projects to deliver time-critical interventions as may be deemed necessary. In addition, there is also a need to automate the process of information extraction to seamlessly sift through massive volumes of documents, reduce human bias in interpreting data, and to ensure high quality information extraction. Within the realm of construction industry, automating the process of information extraction and dissemination on construction documents has been performed to improve communication among project participants (Mutis et al. 2007), expand organizational learning (Erkal and Molenaar, 2020), and enhance site safety and productivity (Zhu 2013) to name a few. To make these baseline practices cost-effective and sustainable requires a richer and digitally supported knowledge management program.

7. Briefly describe the history of its development.

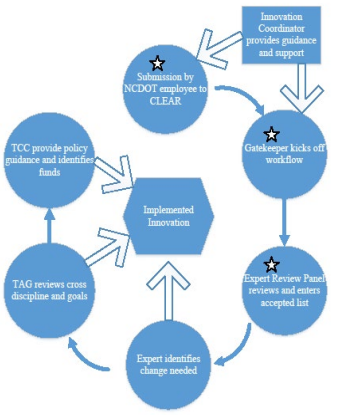
To improve the search capabilities of the current CLEAR search engine, a new AI approach using Natural Language Processing was initiated. Development of this model involved machine learning using a word corpus from pertinent project documents such as project feasibility study reports, NCDOT construction

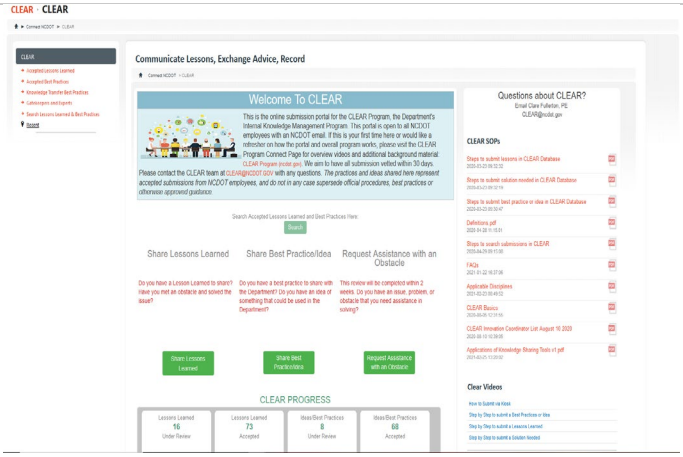
manuals, construction text books, etc. was first developed. The model places semantically similar words closer to each other. Input keywords from end-users were finally searched in the database; the results include words that are closely placed due to their semantic similarity, thus yielding improved search results. This process included searching for pertinent lessons learned and best practices within the CLEAR program to suggest to end-users for their use on projects. These end-users include members from all project phases to obtain a comprehensive review of the current automated process and obtain feedback to further improve the neural language model. Multiple rounds of review will be conducted with these end-users. For the scope of this research, efforts will be made to broaden the search results to include relevant lessons learned and best practices from CLEAR based on keywords inputted by the end-users. These search results represent the innate project knowledge and uncover useful insights from the domain knowledge to be applied on future projects. In other words, the AI model will map the essential keywords/sets to the text documents for suggesting the most relevant and necessary documents that are semantically related to such automatically detected keywords from project documents.

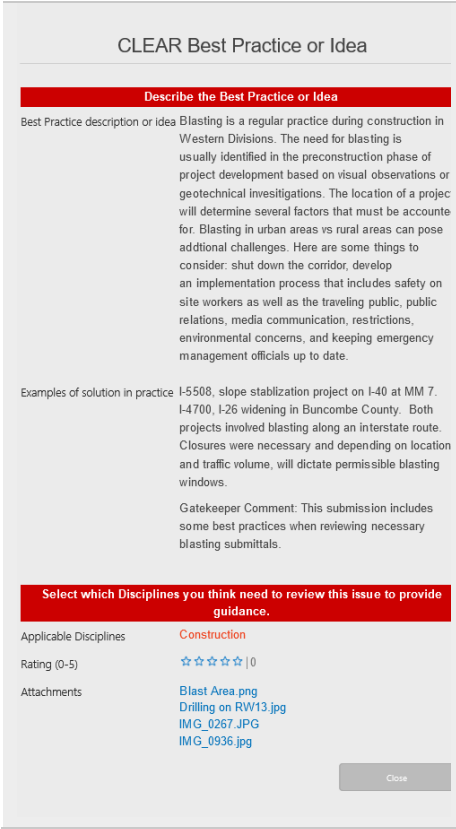
8. What resources—such as technical specifications, training materials, and user guides—have you developed to assist with the deployment effort? If appropriate, please attach or provide weblinks to reports, videos, photographs, diagrams, or other images illustrating the appearance or functionality of the innovation (if electronic, please provide a separate file). Please list your attachments or weblinks here.

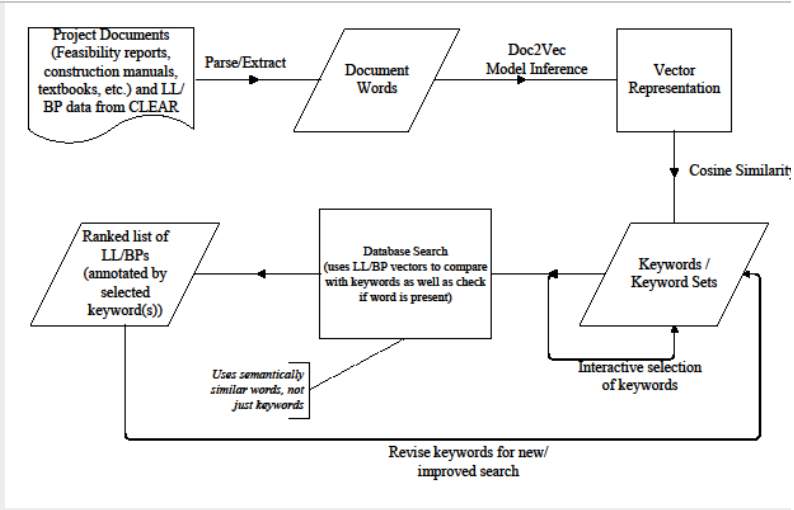
We have developed a fully-functional pilot application of a trained AI language model for the construction domain, an interactive web-application that supports intelligent search (similar to suggested keywords in search engines like Google), and a knowledge graph of project documents such as feasibility studies, specifications, and lessons learned/best practices that can be queried to return semantically similar projects without users having to enter specific keywords. A video of the interface in action is available via the following YouTube link https://youtu.be/hdGo7oxVWuM.

Attach photographs, diagrams, or other images here. If images are of larger resolution size, please provide as separate files.









State of Development (40 points)

Innovations must be successfully deployed in at least one State DOT. The AII selection process will favor innovations that have advanced beyond the research stage, at least to the pilot deployment stage, and preferably into routine use.

9. How ready is this innovation for implementation in an operational environment? Please select from the following options. Please describe.

☒ Prototype is fully functional and yet to be piloted

☐ Prototype has been piloted successfully in an operational environment

☐ Technology has been deployed multiple times in an operational environment

☐ Technology is ready for full-scale implementation

Our prototype is fully-functional and is ready for use in an operational environment. We are conducting usability studies with NC DOT project managers to get feedback on the design and usability of the interface.

10. What additional development is necessary to enable implementation of the innovation for routine use?

Integration with the existing software infrastructure that is being used by the NC DOT and developing documentation and training modules for the IT staff. This is necessary for the maintenance and expansion of the language model with new documents.

11. Are other organizations using, currently developing, or have they shown interest in this innovation or of similar technology?? ☒ Yes ☐ No

If so, please list organization names and contacts. Please identify the source of this information.

Organization	Name	Phone	Email
Michigan DOT	Aaron Johnson	Click or tap here to enter text.	Johnsona25@michigan.gov
Utah DOT	Rod McDaniels	Click or tap here to enter text.	Rmcdaniels@utah.gov
Click or tap here to enter text.	Click or tap here to enter text.	Click or tap here to enter text.	Click or tap here to enter text.

Potential Payoff (30 points)

Payoff is defined as the combination of broad applicability and significant benefit or advantage over baseline practice.

12. How does the innovation meet customer or stakeholder needs in your State DOT or other organizations that have used it?

This innovation will allow for the knowledge base found in the CLEAR program to proactively be utilized on future projects. This will allow the Department to reach a level of organizational knowledge that will improve project delivery and enhance cross collaboration that works towards innovative solutions. Right now, this is being applied to highway projects but once it is established, it can be utilized for any project or program need.

13. Identify the top three benefit types your DOT has realized from using this innovation. Describe the type and scale of benefits of using this innovation over baseline practice. Provide additional information, if available, using quantitative metrics, to describe the benefits.

Benefit Types	Please describe:
Improved Operation Performance	By utilizing the AI search engine, more relevant and returned lessons learned and best practices have been discovered that will improve the performance of projects.
Organizational Efficiency	The improved ranking of returned submissions saves time and increases the efficiency of the knowledge transfer benefit providing for the CLEAR program.
Improved Quality	This process improved the quality of the search which allowed for more relevant submissions to be implemented during the pilot testing.

Provide any additional description, if necessary:

N/A

14 How broadly might this innovation be deployed for other applications. in the transportation industry (including other disciplines of a DOT, other transportation modes, and private industry)?

A specialized language model for the construction domain and an AI-based knowledge graph of the construction domain has broad applicability across all sectors of construction including state

transportation offices, private industry, education, and transportation modes. The automation in terms of organizing and keeping track of this information is valuable for many data analytics based approaches that are prevalent in other domains. The specific semantics of the construction domain, the scale of expenditures, and the long range of maintenance make this a unique domain for AI.

Market Readiness (20 points)

The AII selection process will favor innovations that can be adopted with a reasonable amount of effort and cost, commensurate with the payoff potential.

15. What specific actions would another organization need to take along each of the following dimensions to adopt this innovation?

Check boxes that apply	Dimensions	Please describe:
☒	Gaining executive leadership support	Leadership support will be needed to assist in most of the actions below including communication, funding, and firewall (legal) concerns.
☒	Communicating benefits	An organization would need to communicate the benefits through a business case – showing the cost/benefit in addition to provide qualitative benefits.
☒	Overcoming funding constraints	Cost of developing a system and programmer would be based on the existing capabilities of an organization.
☒	Acquiring in-house capabilities	In-house IT capabilities, including specific programming skills would be needed.
☒	Addressing legal issues (if applicable) (e.g., liability and intellectual property)	An organizations firewall would need to be reviewed to allow for the use of data in the developed system.
☐	Resolving conflicts with existing national/state regulations and standards	Click or tap here to enter text.
☐	Other challenges	Click or tap here to enter text.

16. Please provide details of cost, effort, and length of time expended to deploy the innovation in your organization.

Cost: Full deployment of the language model and search interface integration with NC DOT IT infrastructure will take 9 to 12 months and expected costs in the range of $52,000 to $64,000. These

include stipends for graduate student programmers and university release time for Dr Jhala. In addition, we anticipate expenditures of $10,000 per year for two subsequent years for troubleshooting and fine tuning of the AI algorithms for the language model.

Level of Effort: 0.5 student programmer year, release time for faculty researchers for project management and research direction

Time: 1 year. + 2/3 years of on-call troubleshooting

17. To what extent might implementation of this innovation require the involvement of third parties, including vendors, contractors, and consultants? If so, please describe. List the type of expertise required for implementation.

This would require the use of Dr Jhala and his team to implement. Potential requirement for contractor in the Department with necessary programming skills.