CNNP/EPRI Collaboration: Generating Synthetic Ultrasonic Testing (UT) Data with Deep Learning

Caleb Watson, Data Scientist/Engineer II
NDE Technology

What is Deep Learning?

- Deep learning is a category of AI that uses neural networks to identify complex patterns in data (particularly useful for high-dimensional data).

- Can be used for a variety of applications (such as classification, decision making, and data generation).

Importance of UT Data

- UT data availability is key for Non-Destructive Evaluation (NDE) training, qualification, and maintaining examiner proficiency.

- Multiple Petabytes of UT data can be needed to validate advanced NDE technologies such as Artificial Intelligence models, Augmented Reality (AR) platforms, and digital twin platforms.

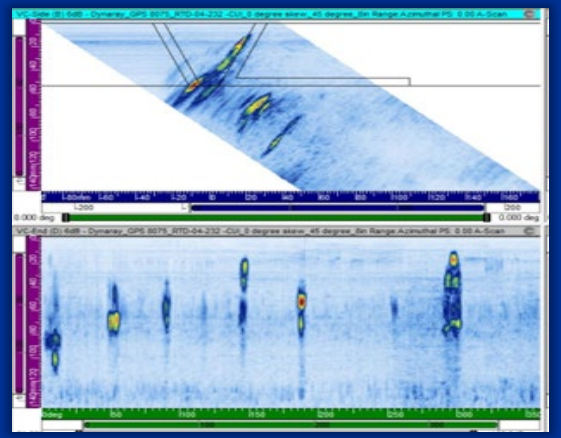


Figure 1: Example of UT Data [1]

Figure 2: Virtual NDE Simulator

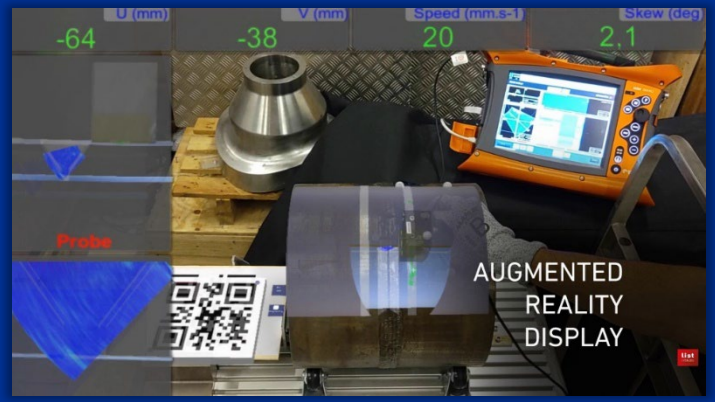


Figure 3: NDE AR Platform

Current Methods of UT Data Collection

- Currently in industry, there are only two methods to acquire UT data:
  - Collect data on a physical sample with an encoding setup.
  - Create simulated data via simulation and modeling tools (i.e. CIVA [1]).

Figure 4: Physical Sample and Encoding Setup [1]

- Collecting data on a physical sample:
  - Can be costly and time consuming.

- Creating simulated data via simulation and modeling tools:
  - May be costly and time consuming due to limiting factors in computing performance.
  - In some cases, not fully realistic.

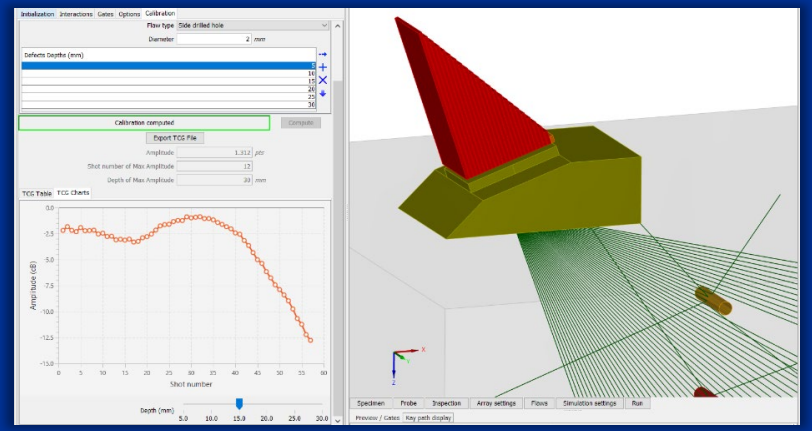


Figure 5: UT Simulation Software – CIVA [2]

Project Objective and Value Statement

Objectives: Create a Deep Learning (subset of AI) algorithm capable of producing and extrapolating synthetic UT data, targeting flaw types or geometries such as corrosion or EDM notches.

- Based on prior EPRI work, previously created algorithms may be leveraged. (publicly available product: 3002028007)

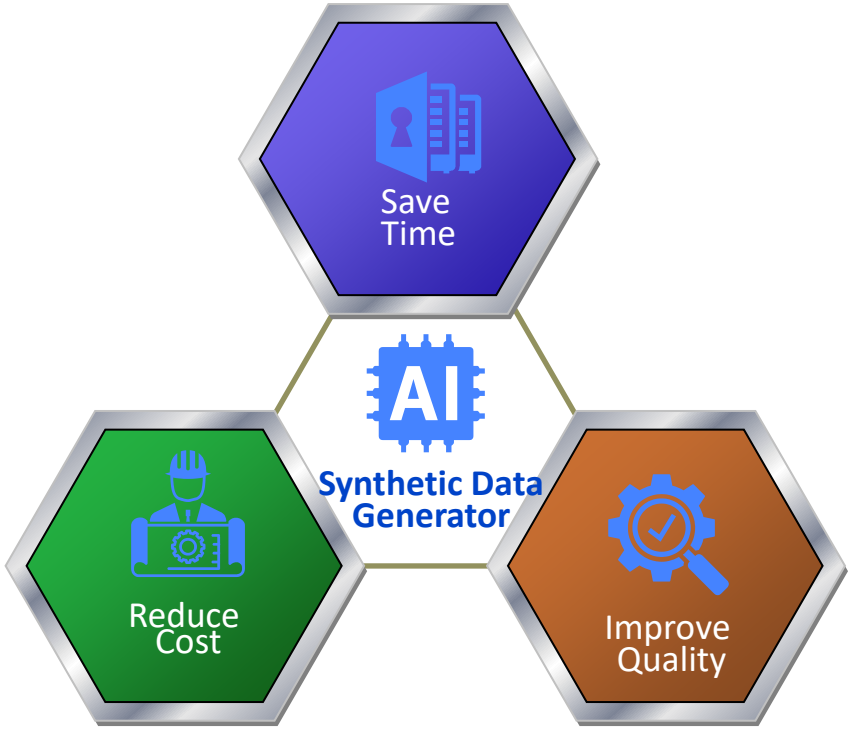


Fig 1: High Level Project Statement

Value Statement: UT Data Synthesizers are computationally lightweight, produce realistic data, and are less cost-intensive than traditional methods.

Previous & Future EPRI/CNNP Work

- Results from initially developed algorithm made public (corrosion flaw types).
- Training data made public (corrosion flaw types).
- Results from initially developed algorithm compared with FE simulations (EDM notch geometries) and made public.
- Create new model (predictive/generative AI approach) or revise current model to improve data realism and performance.

- Generating Synthetic Data with Deep Learning (https://www.epri.com/research/products/000000003002028007)
- Program on Technology Innovation: Corrosion Ultrasonic Data Set of Aluminum Plate (https://www.epri.com/research/products/000000003002031138)
- Artificial Intelligence in Nondestructive Testing: Comparing Simulated and Deep Learning Based Ultrasonic Flaw Responses (https://www.epri.com/research/products/000000003002032320)
- **New research with CNNP**

Statement of Work Overview

1. Exploratory Data Analysis
2. Model Development
3. Performing Experiments with Developed Models
4. EPRI Feedback and Model Revision
5. User Interface Development
6. Reporting

Exploratory Data Analysis

- CNNP will conduct an exploratory data analysis (EDA) on the EPRI supplied ultrasonic datasets.
  - The goal of the EDA is to understand the data, attributes, and features that will be useful in the modeling effort.
- The EDA phase will allow the CNNP to obtain domain knowledge and establish modeling expectations.
- Data integrity, sparsity, and subsequent cleaning will also be evaluated and performed in this phase of the analysis.
- The contractor will examine the current state-of-the-art implementations of generative/predictive artificial intelligence algorithms, specifically determining which model frameworks are best fit for the problem given.

Model Development

- CNNP will:
  - Create a predictive, or generative/predictive hybrid model to synthesize UT data.
  - Explore the feasibility of data extrapolation.
  - Explore latent space features.
  - Documenting other technical parameters and metrics that can be made for assessing the realistic behavior/appearance of the synthesized UT data such as signal to noise ratio, response accuracy in time and amplitude, and/or mean square error.

Performing Experiments with Developed Models

- Experiments shall be performed by the CNNP with EPRI provided ultrasonic data (with or without labels). These provided datasets are to be used with newly developed CNNP model(s) from task 2 to generate synthetic ultrasonic data.

- Additionally, the CNNP shall assist EPRI personnel in creating evaluation metrics, and visualizing the results of the experiments conducted.

EPRI Feedback and Model Revision

- The CNNP will use previously created evaluation metrics and qualitative examination of results to revise the initially developed model framework.

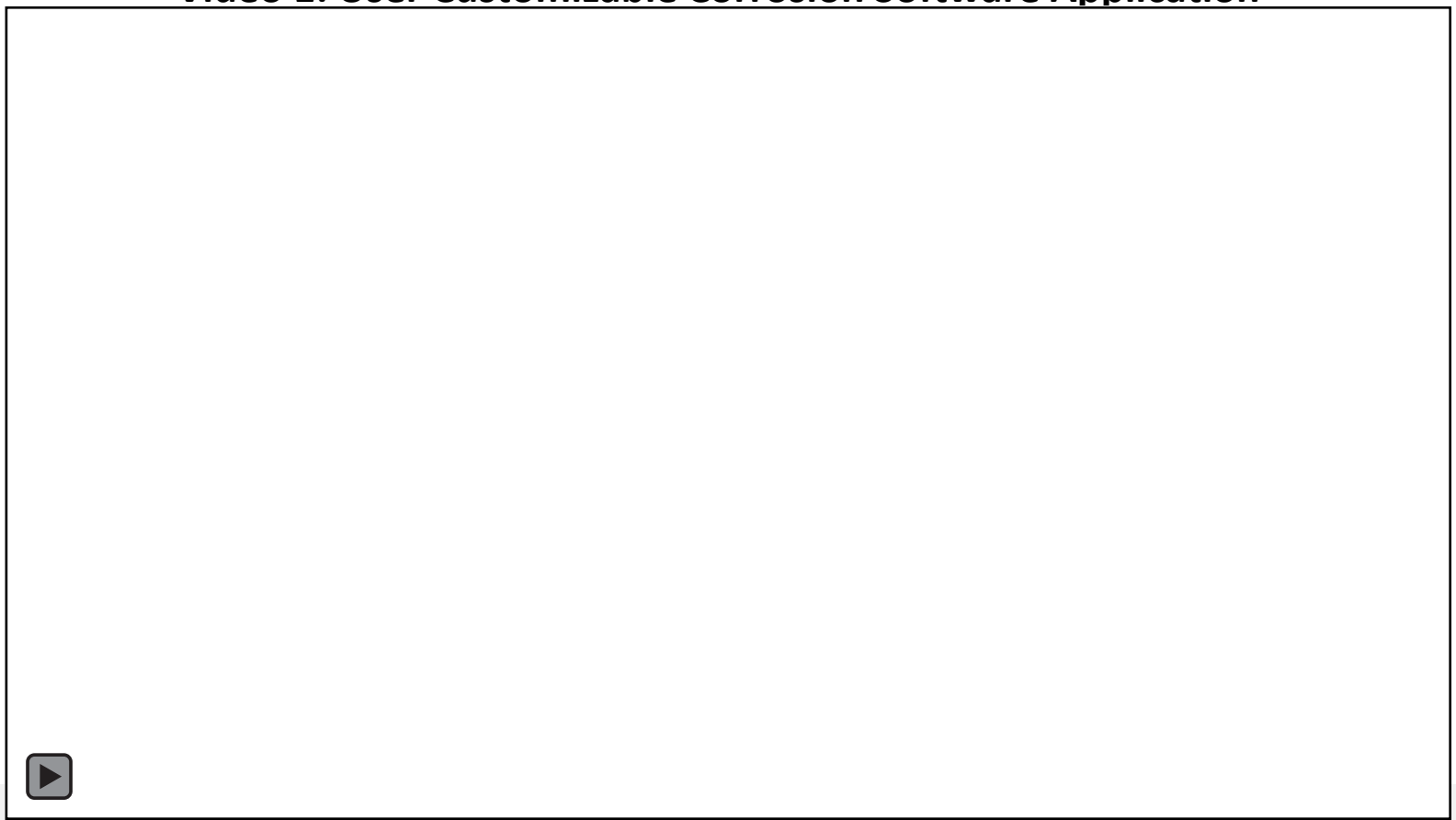
- EPRI will provide feedback regarding the level of realism in the synthetic data by which the CNNP will use to revise the model framework.

- EPRI will also assist with creation of evaluation metrics.

User Interface Development

- After satisfactory model performance has been achieved in the time given and model development has concluded, a user interface will be created to interact with the developed model framework and showcase the utility of research

Video 1: User Customizable Corrosion Software Application



Reporting

- During the life of the project, EPRI and the CNNP will hold bi-weekly meetings to discuss progress and obtain feedback and relevant guidance.

- At the conclusion of the project, the CNNP will help prepare a written technical brief of mutually agreed upon length with a targeted 10 page minimum detailing the approach, results (including model assessment), and conclusions.

- In addition, the CNNP will develop a brief document with code documentation and install/run/model training instructions for all code developed. The contractor will also make a presentation to EPRI staff summarizing the results and showing how to run and use all developed code.

Technology Transfer Plan and Value Statement

- The majority of work and project cost will be dedicated to algorithm development.

- A technical brief will be published to outline performance of the new or revised synthetic data generator.

- Implementation Category: Reference (Technical Basis)

Expected Project Logistics	
Estimated Cost	TBD
Duration	12 Months
Deliverable Type	**Technical Brief**

Table 1: Project Logistics

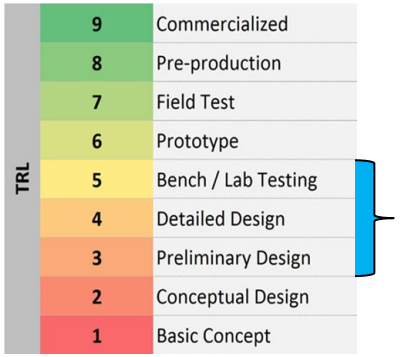


Fig 2: TRL Readiness of this Project

TOGETHER…SHAPING THE FUTURE OF ENERGY®

www.epri.com