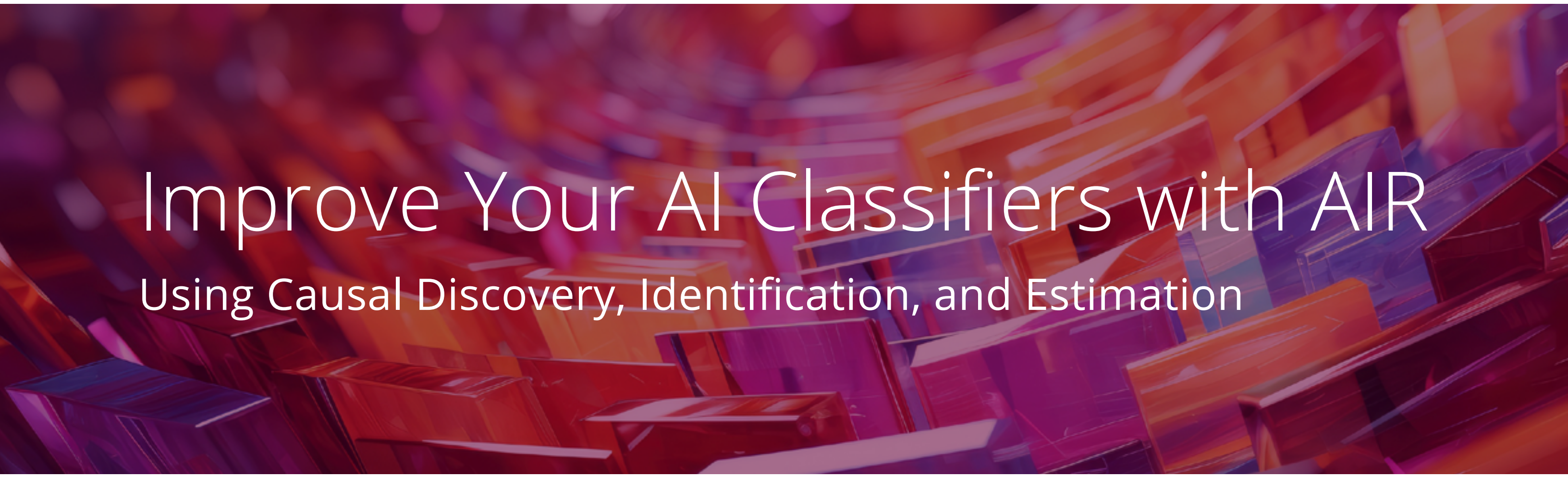
Carnegie Mellon University
Software Engineering Institute



Improve Your AI Classifiers with AIR

Using Causal Discovery, Identification, and Estimation

MODERN ANALYTIC METHODS, INCLUDING ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING (ML) CLASSIFIERS, DEPEND ON CORRELATIONS; however, such approaches fail to account for confounding in the data, which prevents accurate modeling of cause and effect and often leads to prediction bias. The SEI has developed a new AI Robustness (AIR) tool that allows users to gauge AI and ML classifier performance with unprecedented confidence.

How we can help

For the past several years, the SEI has been applying and adapting novel techniques from causal discovery (which produces cause–effect graphs) and causal inference (to evaluate treatment effects) to assess various classifier predictions with more nuance, resulting in

- AI and ML predictions that are less biased and more suitable for interventions/control
- better attribution of outliers and causes

This project is sponsored and funded by OUSD(R&E) to transition use of our AIR tool to AI users across the DoD. If you choose to participate in this project, you will receive custom setup of and training with our AIR tool. Your only cost is participation!

Why AIR?

DoD is increasing its use of AI classifiers and predictors; however, users may grow to distrust results because AI classifiers are subject to a lack of robustness (i.e., ability to perform accurately in unusual or changing contexts). Drift in data/concept, evolving edge cases, and emerging phenomena undermine the correlations relied upon by AI. New test and evaluation methods are therefore needed for ongoing evaluation.

How does it work?

The SEI AIR tool offers a precedent-setting capability to improve the correctness of AI classifications and predictions, increasing confidence in the use of AI in development, testing, and operations decision making (see Figure 1).

Improving classifier performance with AIR requires that we first build a causal graph (Step 1) that includes the treatment variable (X), the outcome variable (Y), any intermediate variables (M), and parents of either X (Z1) or M (Z2). Once we have a graph, we identify two conditioning sets (Step 2) that attempt to remove confounding effects associated with Z1 (top) or Z2 (bottom). Finally, we calculate the average risk difference and associated 95% confidence intervals for each conditioning set (Step 3) using causal inference and compare these to the AI Classifier’s predictions.

Collaborating for Success

We are looking for DoD collaborators to use and provide feedback on our technology. As a participant, your AI and subject-matter experts will work with our team to identify known causal relationships and build out an initial causal graph. Our process involves using a cutting-edge causal discovery tool, Tetrad, with custom causal identification algorithms and stacked super-learners using doubly-robust causal estimators to build an AI “health report.” Your report will include a confidence range of expected treatment effects from your data and interpretations of the causal graph to give you actionable insights into your AI classifier’s health.

For this project to be successful, your team must be willing to collaborate with the SEI and have an established AI classifier workflow, complete with data dictionaries and subject-matter experts to help us build out an optimized implementation.*

As part of our empirical research approach, we need timely access to subject-matter experts who know your training and test data well to help us develop valid causal models and identifications. This collaboration helps ensure that we apply the proper context and interpretations of the data.

Benefits to Partner

- Receive custom insights and recommendations on how to improve performance of AI and ML classifiers that support mission
- Obtain data-based confidence/trust in the robustness of the existing AI and ML classifiers
- Enhance staff capability and understanding of AI classifiers
- Become innovators in this domain and contribute to improving the state of practice across DoD
- Receive training and custom tool support assets

Contributions from Partner

- Deploy AIR for use on your AI/ML classifiers and rich data set
- Prepare infrastructure and ensure staff are ready to be mentored by SEI for AIR deployment
- Share results to validate effectiveness of AIR
- Provide feedback that leads to improvements and adaptation for broader DoD use

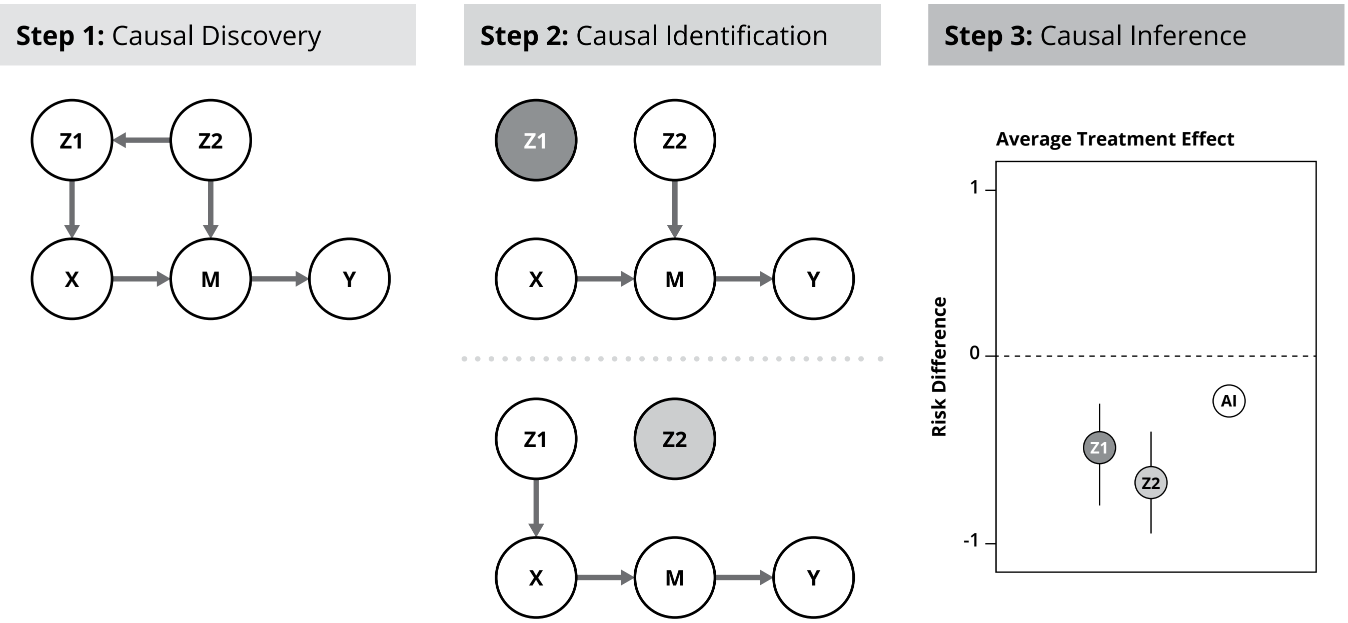


Figure 1: Steps in the AIR Tool Analysis Process. Results and interpretations given by the AIR tool are based on output from all three steps.

If you believe your organization could benefit from this research, please reach out to us.

**The SEI can accommodate classification levels up to Top Secret.*

About the SEI

Always focused on the future, the Software Engineering Institute (SEI) advances software as a strategic advantage for national security. We lead research and direct transition of software engineering, cybersecurity, and artificial intelligence technologies at the intersection of academia, industry, and government. We serve the nation as a federally funded research and development center (FFRDC) sponsored by the U.S. Department of Defense (DoD) and are based at Carnegie Mellon University, a global research university annually rated among the best for its programs in computer science and engineering.

Contact Us

CARNEGIE MELLON UNIVERSITY
SOFTWARE ENGINEERING INSTITUTE
4500 FIFTH AVENUE; PITTSBURGH, PA 15213-2612

sei.cmu.edu
412.268.5800 | 888.201.4479
info@sei.cmu.edu