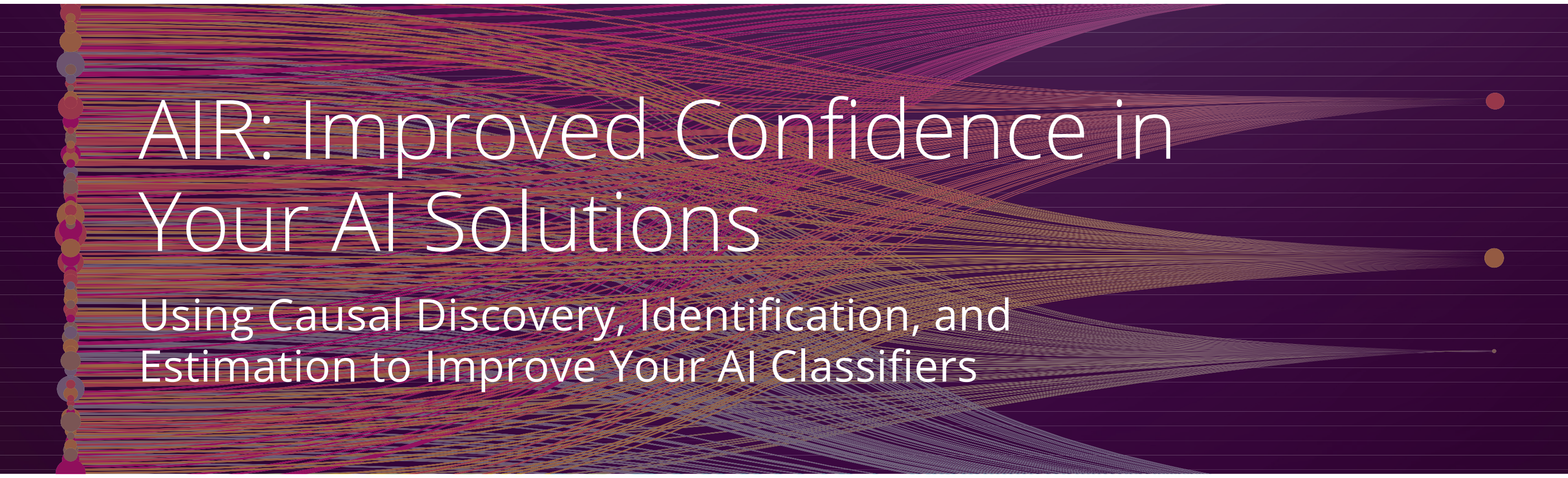
Carnegie Mellon University
Software Engineering Institute



AIR: Improved Confidence in Your AI Solutions

Using Causal Discovery, Identification, and Estimation to Improve Your AI Classifiers

MODERN ANALYTIC METHODS—including artificial intelligence (AI) and machine learning (ML) classifiers—provide powerful tools that have revolutionized prediction capabilities and automation through their capacity to analyze and classify data. To produce their results, most AI and ML methods depend on correlations. However, an overreliance on correlations can lead to prediction bias and reduced confidence in AI outputs.

Drift in data and concept, evolving edge cases, and emerging phenomena can undermine the correlations that AI classifiers rely on, resulting in a lack of robustness (i.e., the ability to perform accurately in unusual or changing contexts). As the Department of Defense (DoD) is increasing its use of AI classifiers and predictors, users may be growing to distrust results because of these issues. To regain user trust in these tools, we need new methods for ongoing testing and evaluation of AI and ML accuracy.

We Can Help Improve Your Classifiers

The SEI has developed a new AI Robustness (AIR) Tool that allows users to gauge AI and ML classifier performance with unprecedented confidence.

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

- AI and ML predictions that are less biased and more suitable for guiding intervention and control of a system's performance
- better attribution of outliers and causes

How Does AIR Work?

Improving classifier performance with AIR begins with building a causal graph (see Step 1 in Figure 1) from variables in the dataset including treatment (X), outcome (Y), and intermediate variables (M). Then, in the resulting graph, the AIR Tool specifies two adjustment sets (Step 2) consisting of the parents of X (Z1, top) and M (Z2, bottom), which allows it to remove any potential bias introduced between X and Y. Finally, the tool calculates the average risk difference and associated 95% confidence intervals for each adjustment set (Step 3) using causal estimation and compares these to the AI classifier's predictions. In this way, the AIR Tool shows when the AI classifier can't be trusted and suggests where it could be improved.

Collaborate with Us to Improve AI Robustness

Do you want to improve the confidence you have in your AI classifiers? Please reach out to work with us on AIR! The tool is free. Your only cost is participation.

We are looking for collaborators to use our technology and provide feedback. As a participant, your AI and subject matter experts will work with our team to identify known causal relationships and build an initial causal graph. In addition, we will provide you a summary of our findings, including 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. If you choose to participate in this project, you will receive custom setup of and training with our AIR Tool.

Benefits of Collaborating with Us

There are many benefits your organization can gain by collaborating on this project, including the following:

- receive custom insights and recommendations on how to improve performance of AI and ML classifiers that support your mission
- obtain data-based confidence and trust in the robustness of your 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

Collaboration with the SEI will involve the following work:

- deploying AIR for use on your AI and ML classifiers and rich data set[*]
- preparing infrastructure and ensure staff are ready to be mentored by SEI for AIR deployment
- sharing results to validate effectiveness of AIR
- providing feedback that leads to improvements and adaptation for broader DoD use

Learn More about the SEI's Research in Causal Learning

You can learn more details about the SEI's research in causal learning and the work that led to AIR in our blog post, "Measuring AI Accuracy with the AI Robustness (AIR) Tool." To read the post, use the QR code to the right.

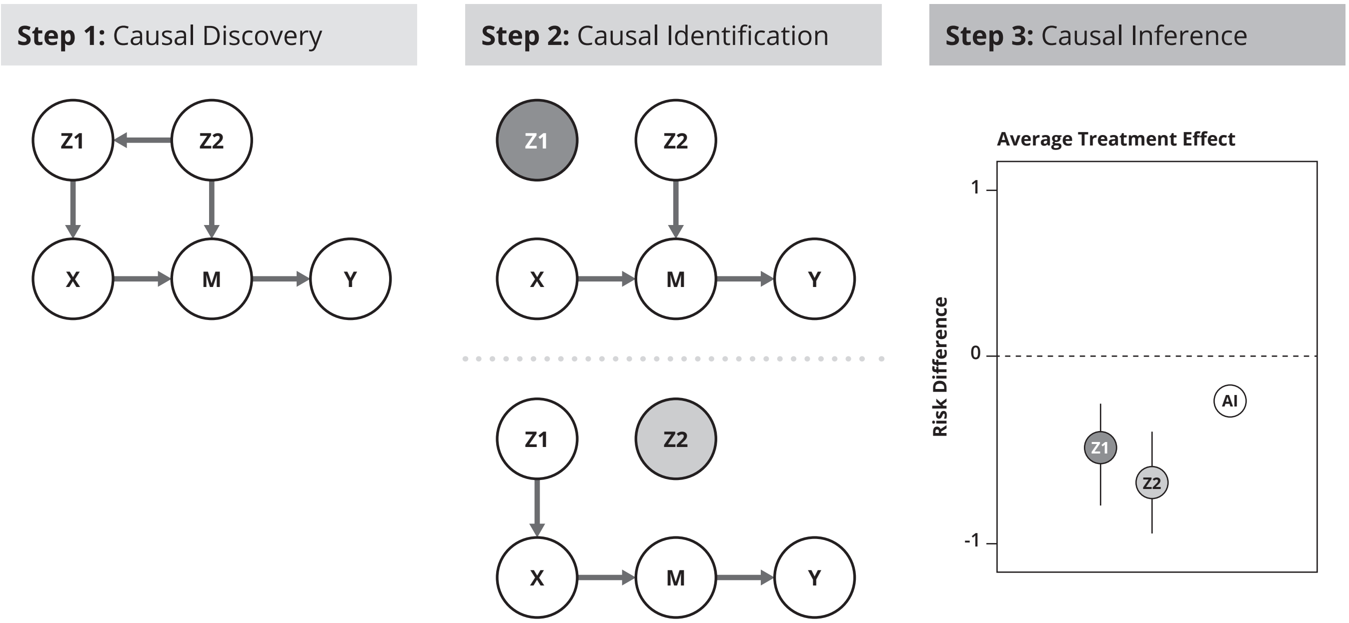


Figure 1: Steps in the AIR Tool Analysis Process. Results and interpretations given by the AIR Tool are based on output from all three steps.

If you believe your organization could benefit from this research, please reach out to us at info@sei.cmu.edu.

**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