AIR Partner Prospective Case Studies

THIS PROSPECTIVE USE CASE DESCRIBES A POTENTIAL EXPERIENCE of a partner who is an early adopter of the Artificial Intelligence Robustness (AIR) Tool and is working with the Software Engineering Institute (SEI) to deploy and mature the prototype AIR Tool.

CONTEXT	A commercial drone company is responsible for gathering pictures after a disaster (e.g., a hurricane, tornado, or a widespread fire). The company needs to understand the variables that can impact the drone's ability to produce useful images efficiently so that responders get the information they need.
DETERMINE IF THE AIR TOOL IS A GOOD FIT	The chief engineer is leveraging an artificial intelligence (AI) classifier that is trained on historical data (including drone, environmental, and imaging performance variables) in order to predict mission success, as indicated by the emergency responders' ability to use the images. While this AI classifier has proven to be invaluable, the introduction of new weather patterns or regional politics may impact its accuracy and thus its trustworthiness. The chief engineer and their team determine that leveraging the AIR Tool will provide insight into the robustness of the current AI classifier and provide insight into what type of actions can be taken to extend the classifier's accuracy in these new situations.
PREPARING INPUTS	The team and the SEI discuss how to leverage the AIR Tool for the insight the team needs about their AI classifier. Subject matter experts' (SMEs') understanding of the data and context are critical to the selection and feature engineering of the appropriate data. The SEI team provides guidance on expectations and use of the AIR Tool. The preparation of inputs includes the following. **Scenarios:** The team identifies more precisely what types of mission scenarios are emerging and required. For example, scenarios could include the need to collect images of hurricane damage while the winds in the area are still strong or the need to capture images of tornado damage while there are still rainstorms in the area. From the existing variables, the team derives an indicator feature for each such scenario, which acts like a treatment label corresponding to that scenario. The indicator feature will be used for each scenario to partition the data into two subsets: one depicting the scenario being pursued (i.e., on-scenario) and the other depicting the scenario not being pursued (i.e., off-scenario). **Select Representative Data:** The team considers two possible datasets: • training dataset from the historical image collection from the past two years (which reflects the impact of different seasons), which is used to train the current AI classifier model • new dataset that is gathered from the image collection in the past month and includes new mission sets that may have been impacted by storms of greater strength than the storms that impacted training data *Continued*

PREPARING INPUTS, *continued*

All the sensor data in the datasets are continuous valued, except for the mission scenario indicator features, which are binary. Approximately 200 variables in the dataset are related to the drone flight and the ability to take quality images. There are 50–5,000 cases with at least 8–15 cases per mission type (e.g., hurricane, tornado, and fire). None of the datasets are classified.

The team elects to apply the AIR Tool to the new dataset (detailed in the second bullet point above). This decision should make the labeling exercise easier (because it's easier to create new mission scenario indicator variables) and should provide insight into the impact of the changing context on the AI classifier's performance.

The team determines what is needed to prepare the selected dataset for application of the AIR Tool. Missing data is identified; therefore, the SMEs on the company's team and the SEI discuss an appropriate remedy for the missing data, considering simple and sophisticated ways to handle missing data based on the circumstances. Some of the missing data is due to random error in processing the data, and the team knows which cases are affected. Therefore, the team chooses to delete those cases from the dataset. Other data is missing due to systemic reasons; therefore, the team decides to replace that missing data by the means of the associated variable and a new binary indicator feature that is constructed to indicate just those cases.

Knowledge File:
The chief engineer quickly identifies variables that aren't systematically impacted by other variables (e.g., the day of the year). These variables are also called "exogenous variables." Through team discussions, other systematic impacts between variables are identified and used to organize the remaining variables into tiers that adhere to the rule that earlier tiers (those of lower values) must be able to feasibly impact and receive no potential causal information from later tiers (those of higher values).

Tier 1 contains clearly exogenous variables;

Tier 3 includes clearly resultant (or outcome) variables;

Tier 2 consists of intermediate and mediating variables that none of the variables in Tier 3 can directly cause but can be influenced by what values they take.

Classifier Model/ATE:
The chief engineer exercises the to-be-evaluated classifier on the data for each scenario. These predictions are separated into two groups: on-scenario and off-scenario. From these two groups, the average is calculated for each group, and the difference of the on-scenario average versus the off-scenario average is calculated. In this way, the average treatment effect (ATE, which is also known as the average causal effect or ACE) is calculated for each new mission scenario.

RUNNING THE AIR TOOL FOR RESULTS

Installation:
With the SEI's training and support, the company team is easily able to access and install the necessary AIR Tool software. The team confirms that their current system supports the technical environment required by the AIR Tool, which requires Java and R. The SEI-delivered AIR Tool files are scanned and installed on the developer desktop, and the SEI-provided tests are run to ensure successful deployment of the tool software.

Use of Application:
The team uses the prepared materials to run the AIR Tool for the scenarios of interest. The tool processes all the inputs for each scenario and generates images and results.

Capture of Results:
The AIR Tool generates a set of causal models (known formally as directed acyclic graphs or DAGs), reflecting the relationships discovered between variables in the data.

Continued

RUNNING THE AIR TOOL FOR RESULTS, *continued*

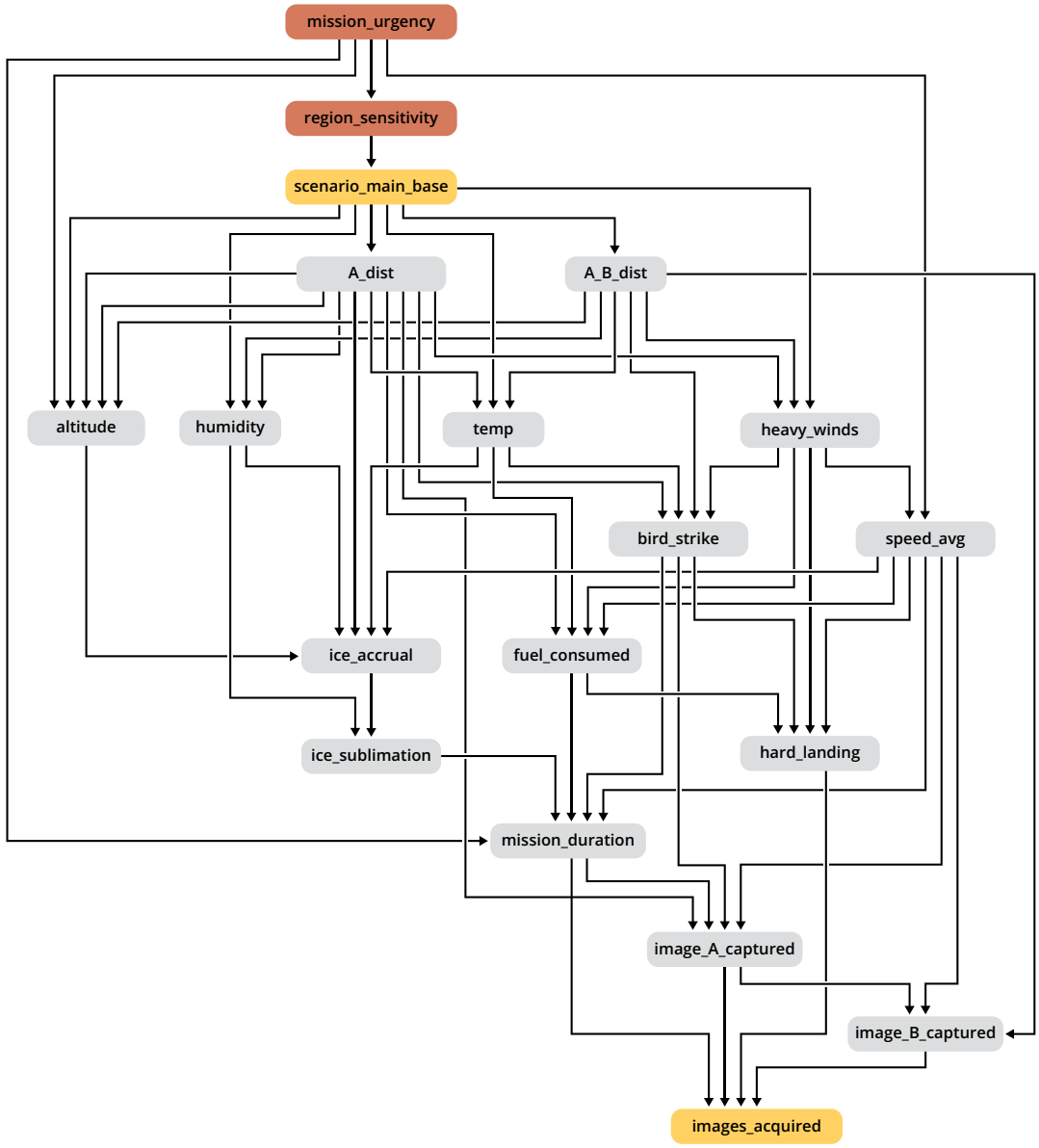


Figure 1. Causal DAG: This is an example of a labeled DAG that the AIR Tool might display as output. Both the scenario and outcome variables are highlighted in yellow. Additional nodes may be highlighted in either red or orange depending on how strongly the AIR Tool's results differ from the AI or ML classifier.

The AIR Tool also produces adjustment sets for each scenario, which offer insight into which variables are particularly important for the function of the classifier. In the DAG, the corresponding nodes are highlighted when the adjustment sets' confidence intervals do not overlap with the AI classifier (orange highlighting indicates that one set doesn't overlap while red highlighting indicates that both sets don't overlap).

The AIR Tool also produces a plot of results for each scenario.

Results Plot:

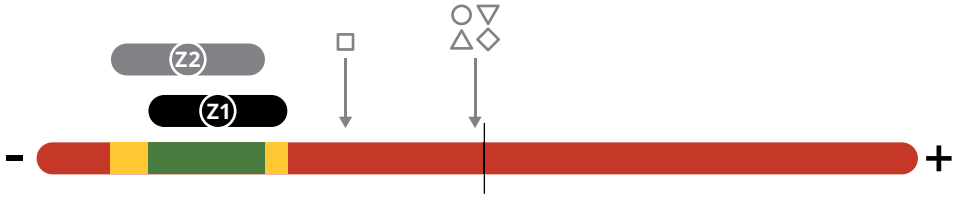


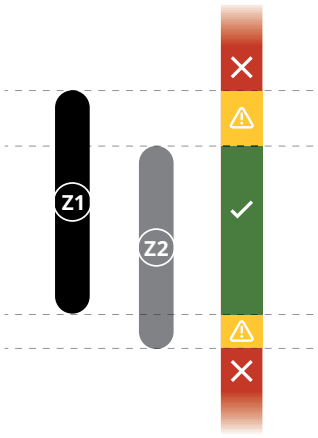
Figure 2. This chart represents the difference in outcomes resulting from a change in the scenario variable. The x-axis ranges from negative to positive effect, where the treatment either decreases the likelihood of the outcome variable or increases it, respectively. The midpoint corresponds to "no significant effect."

ACTIONS AND RESULTS

The company team discusses the interpretation of these results. As the team expected, the measured correlation between on-scenario and mission success was confounded by common ancestor variables to both treatment and outcome (i.e., scenario indicator and weather conditions), leading the classifier to produce biased predictions regarding the likelihood of collecting quality images.

The risk difference plot shows the team that the current classifier's indication of the likelihood of mission success for the first new mission scenario did not fall into the high confidence bands generated by the AIR Tool.

Risk difference plot:



With these insights, the team agrees that these results confirm that the AI classifier requires maintenance. The team considers two options:

- retraining the model on new performance data
- identifying variables that are not included in the current dataset and that would better inform the AI classifier's reliability

After discussion with model SMEs and drone SMEs, the chief engineer decides to assign a small group to a project (1) to consider whether additional variables should be collected and (2) to retrain the model and iteratively employ the AIR Tool for these scenarios until the classifier evaluation is more aligned with the high confidence bands.

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