AI Robustness (AIR)

AIR Knowledge File Guidance

For AIR Step 1 – Causal Discovery – the user is asked to upload a Knowledge File that defines rough hierarchies of levels of causation between variables as determined logically by subject matter experts. This input supports the accuracy and efficiency of the AIR causal learning algorithms.

Levels are defined as follows:

* **Level 0 – Exogenous variables**: These variables are not influenced by any other variables. Often used as starting points for causal graphs.
* **Level 1 – Endogenous variables:** Variables in this level are potentially influenced by those in level 0 and possibly other level 1 variables.
* **Level 2 and up – Higher-level variables:** These variables may be influenced by any preceding levels or its own. There can be any number of levels, though as few as three are strictly necessary.

Although it’s not necessary to assign every variable to a level, subject matter experts are encouraged to assign as many variables to as many hierarchy levels as logically correct, which will improve the performance of the tool. Two levels are sufficient, but three or more are generally recommended.

The Knowledge File should be a CSV with two columns. Here is an example about growing crops with reminders for appropriate formatting and content of the Knowledge File. (Next page)

v. 0.9.1

11. Knowledge file is saved as a CSV (text file) for upload.

1. The first row contains headers “Level” in column A and “Variable” in column B.

10. All other cells remain blank.

Table, Excel

AI-generated content may be incorrect.

9. All variables are assigned a level in column A.

8. The Experimental (X) variable should be at a level that precedes the Outcome (Y) variable to be analyzed.

7. All variables appear in column B exactly as they appear in uploaded dataset with no spaces. If unsure about causal influence of a variable, do not include it in the Knowledge File.

4. Variables are only influenced by variables in subsequent level or the same level.

3. There is only one variable per row.

2. Variables in first tier “0” are not influenced by any variables in other levels.

6. Rows are sorted by level.

5. Knowledge file contains at least 2 levels.