Assessing “AI-Enabled” Tools

Separating the Wheat from the Chaff

Mike Hadjimichael, Ph.D.

Cyber AI Domain Capability Area Lead, National Cybersecurity FFRDC

The MITRE Corporation

mikeh@mitre.org

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™

AI-Enabled
AI-Powered
We Use AI
AI For Better…

There is Still Much to Learn With AI

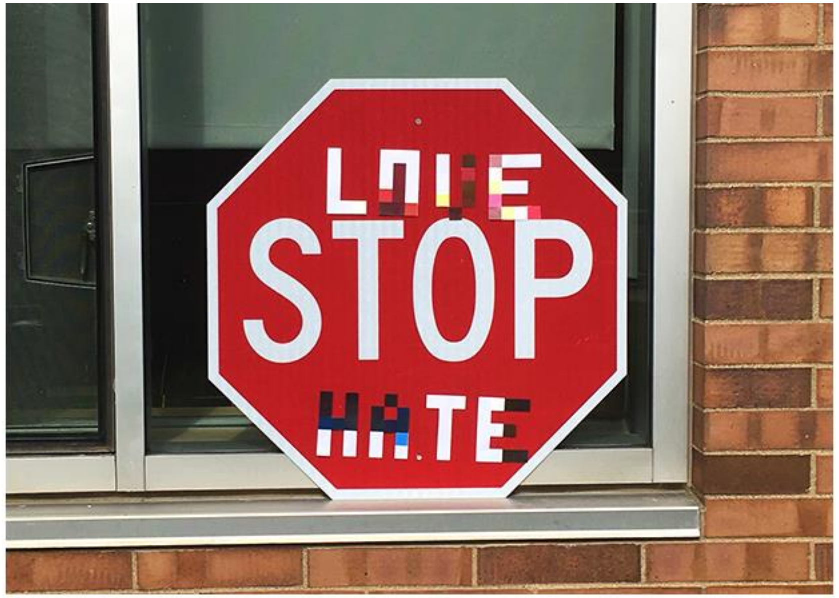


Photo: Cornell University

Image from: https://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms

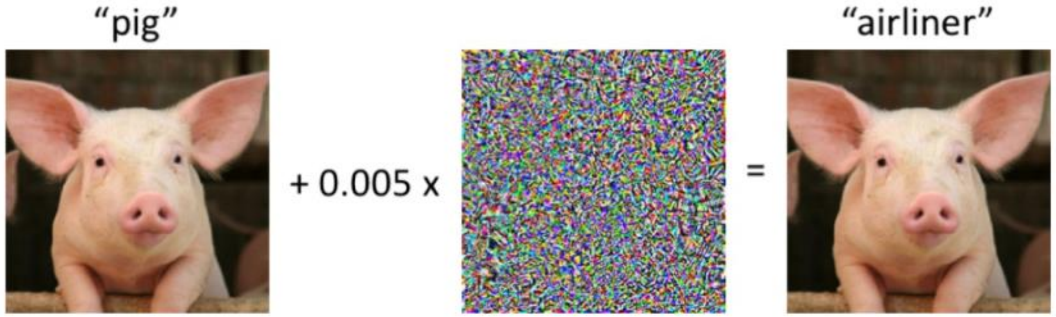
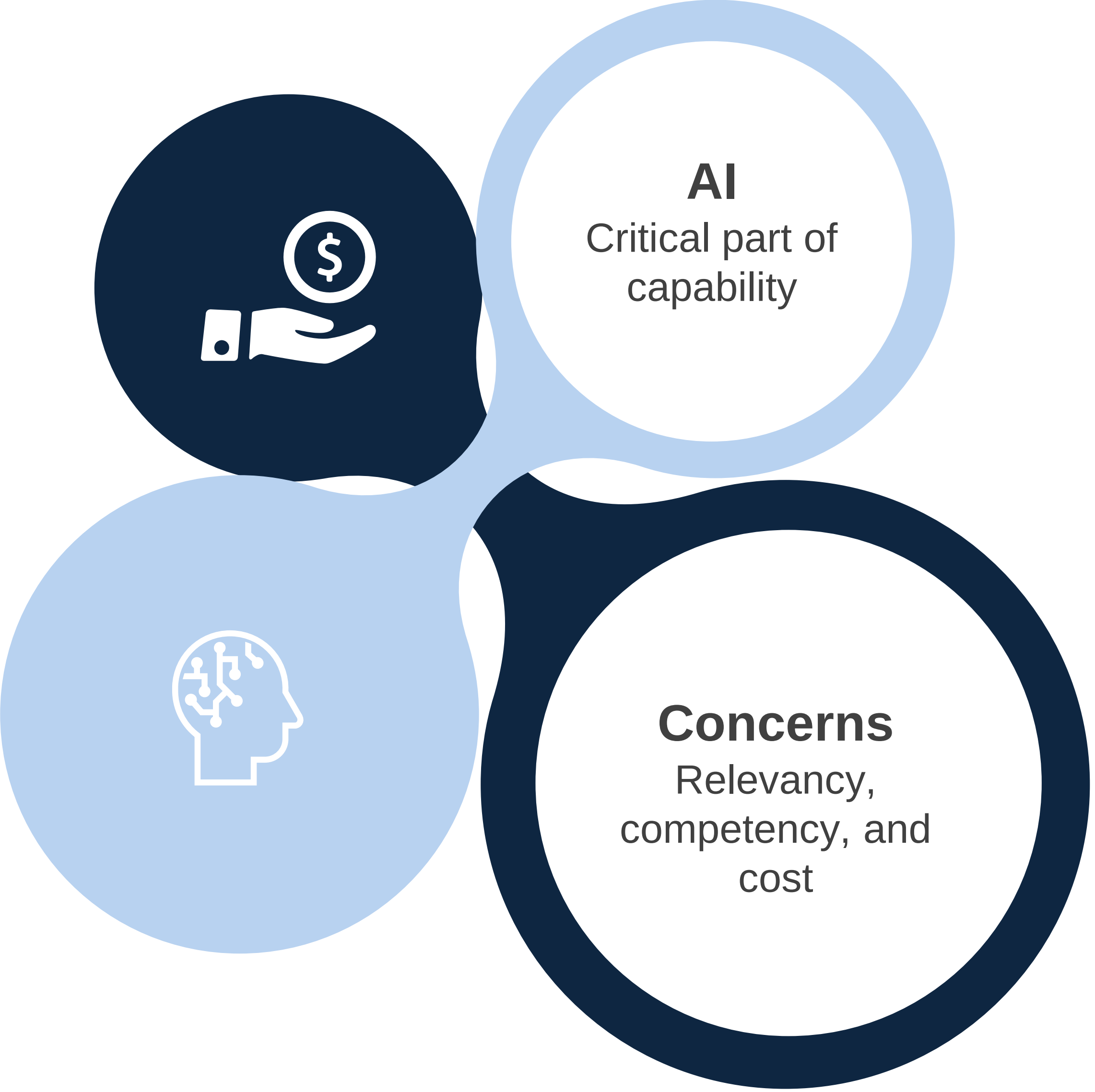


Image from https://gradientscience.org/intro_adversarial/

Challenge

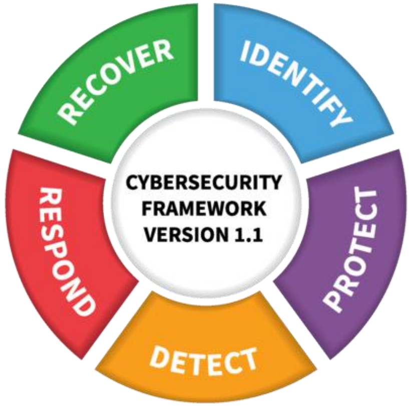


AI
Critical part of capability

Concerns
Relevancy, competency, and cost

Uncertainty
What questions do we ask to pair functionality of tool with what AI does?

Addressing the Challenge

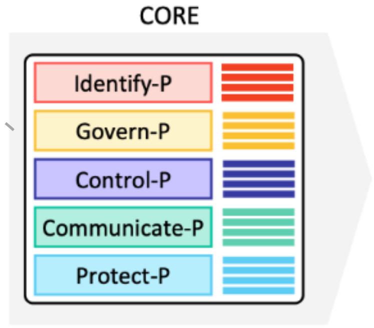


https://www.nist.gov/cyberframework

Frameworks

MITRE | Shield

https://shield.mitre.org



https://www.nist.gov/privacy-framework

ATT&CK®

https://attack.mitre.org

MITRE | ATLAS

https://atlas.mitre.org/

National Cybersecurity FFRDC Research Initiative

Goals

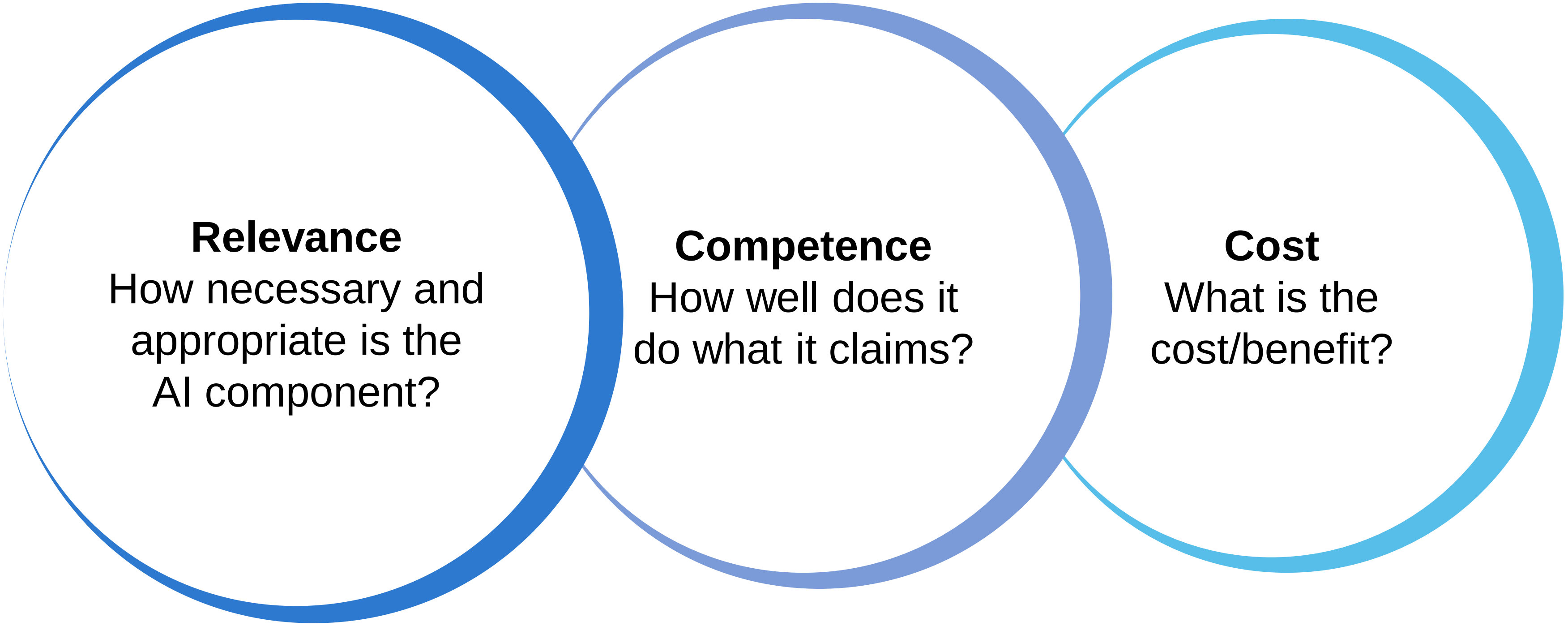
- Offer better understanding of the AI component
- NOT REVEAL proprietary development
- Facilitate better dialogue between vendor and potential tool adopter
- Developing a tool an organization can use
- More than cyber

The ARCCS Framework

AI Relevance Competence Cost Score

The ARCCS Framework
AI Relevance Competence Cost Score

Purpose: Develop an evaluation methodology and metrics to assess the degree and effectiveness of the AI component of a commercially offered, AI-enabled product



Relevance
How necessary and appropriate is the AI component?

Competence
How well does it do what it claims?

Cost
What is the cost/benefit?

Relevance

Is it the right tool?

- Machine learning?
- Expert system?
- The right data?

Is it central and significant?

- How much functionality does AI bring?

Is it necessary?

- Sometimes, it’s just gratuitous. Did we even need AI?

Competence

Needs Alignment

- Does it do well what you need?

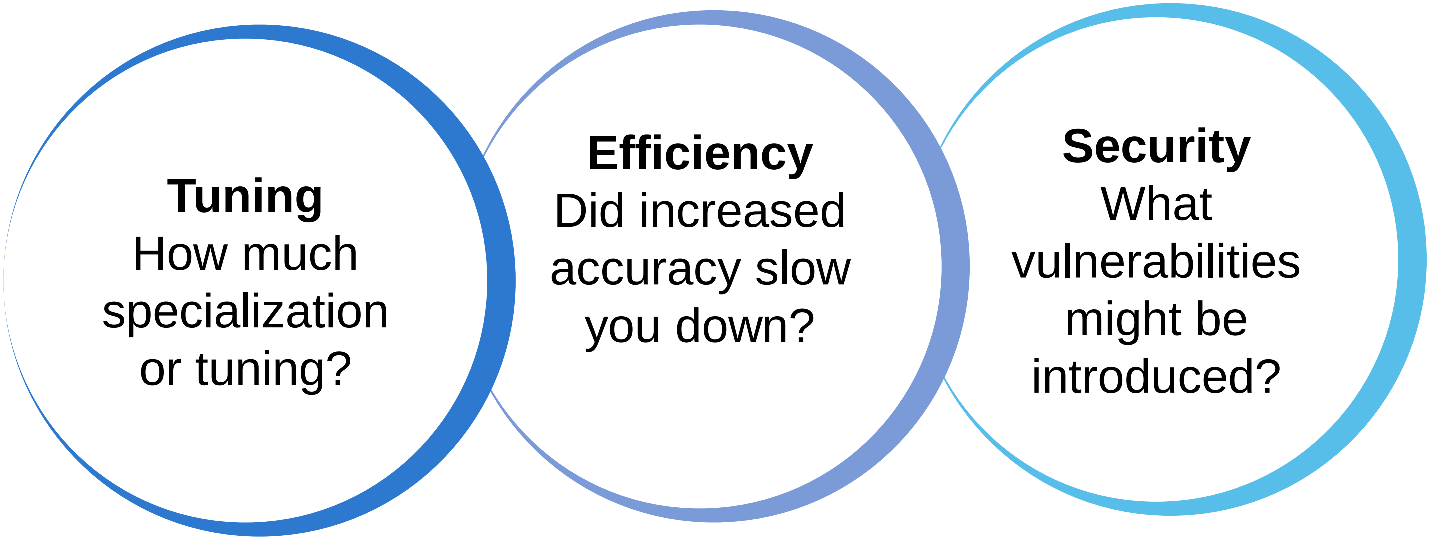
Real world demonstration

- How deal with operational issues like model drift and retraining requirements?

Transparency

- Can we see inside the box?
- How to monitor and improve performance?

Cost…More Than Dollars and Cents



Tuning
How much specialization or tuning?

Efficiency
Did increased accuracy slow you down?

Security
What vulnerabilities might be introduced?

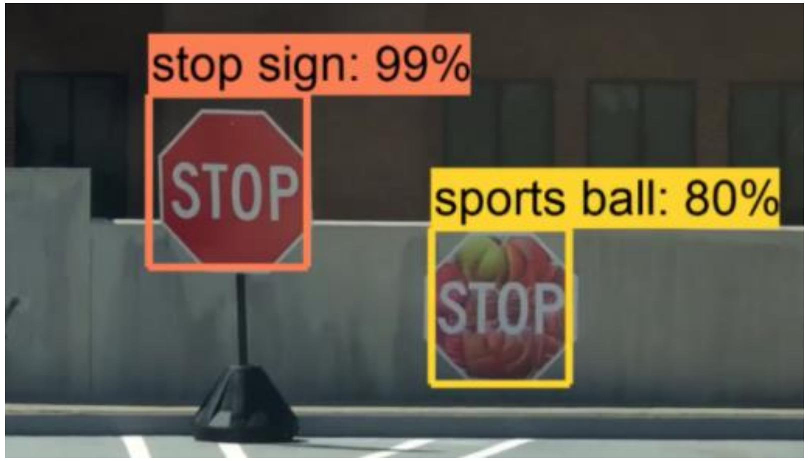
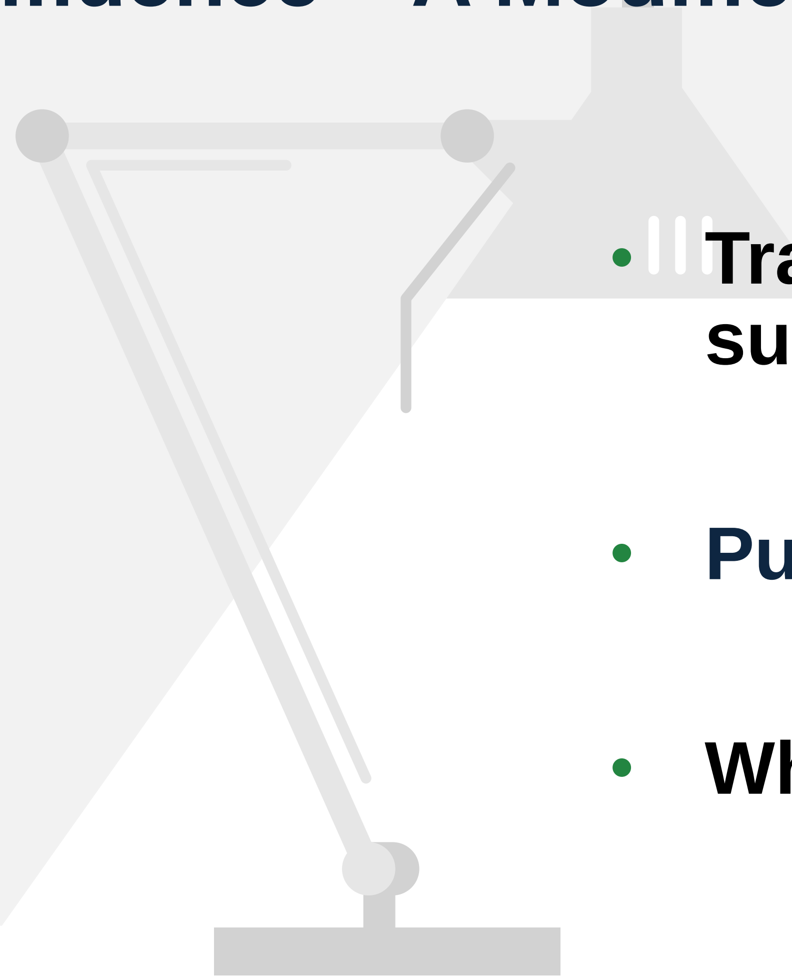


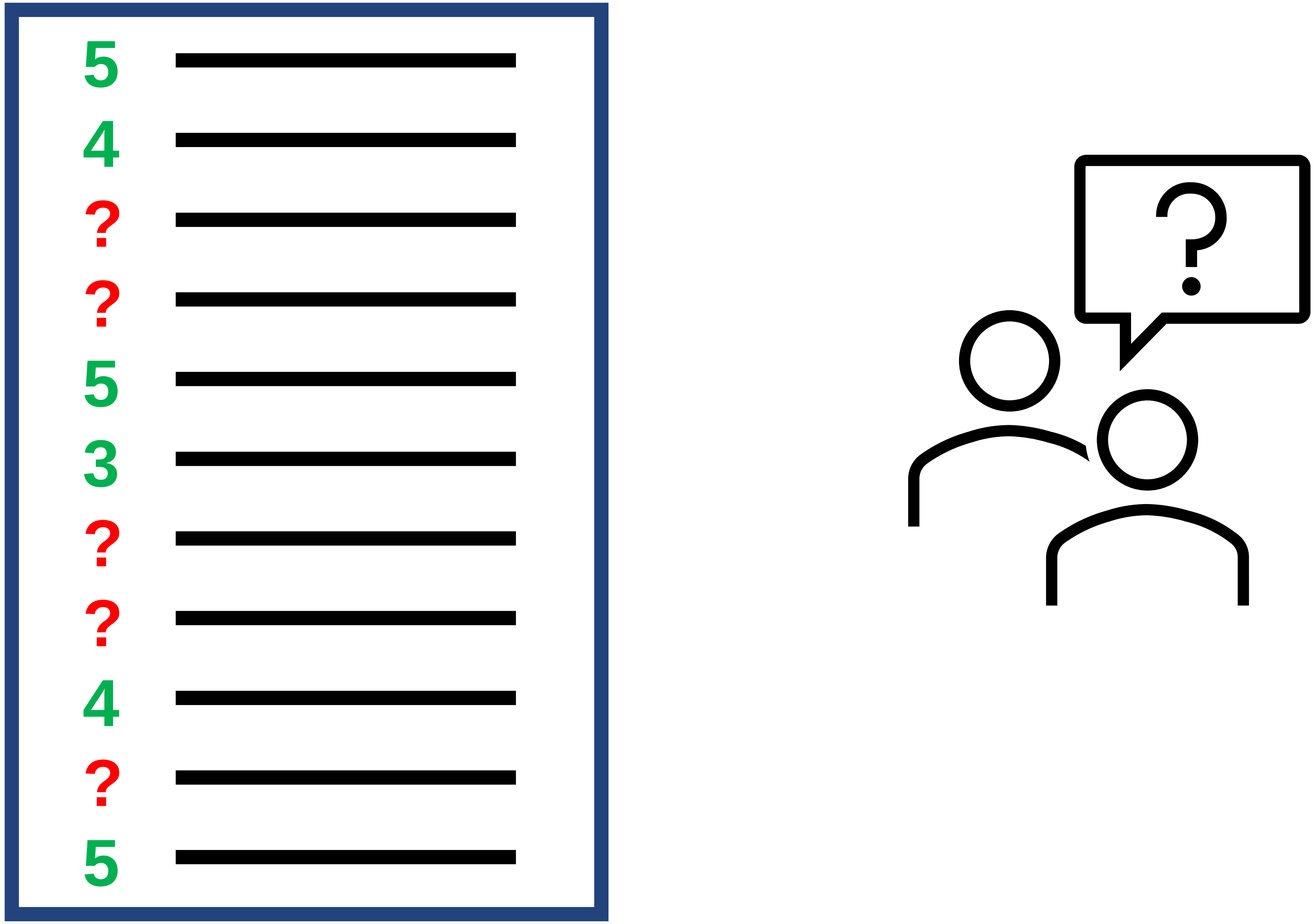
Image from: https://www.cse.gatech.edu/news/611783/erasing-stop-signs-shapeshifter-shows-self-driving-cars-can-still-be-manipulated

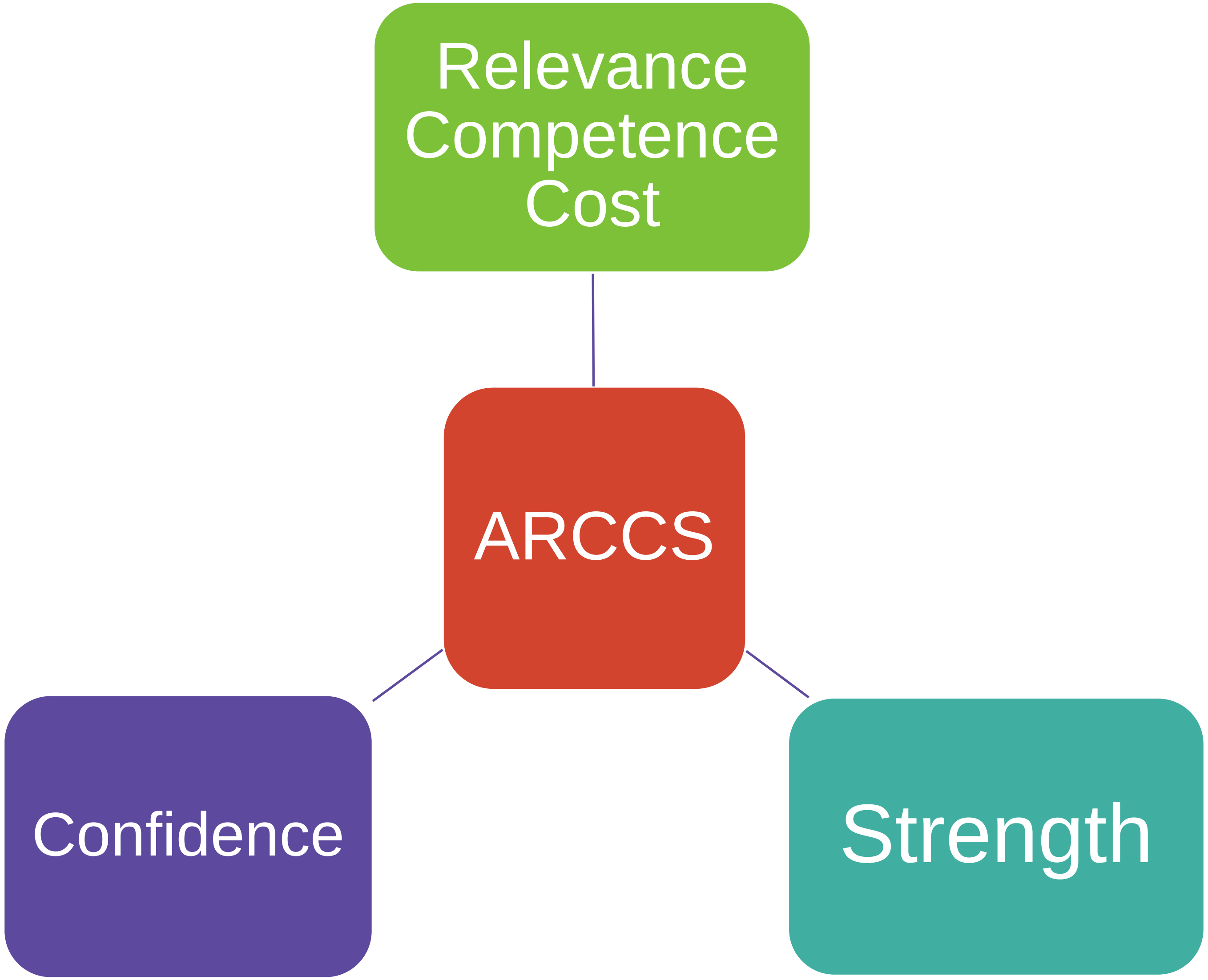
Confidence – A Modifier



- Transparency into the model and supporting data
- Publications and patents
- White papers and publicity materials

Strength – A Modifier Reflecting Knowns vs. Unknowns




Relevance
Competence
Cost
ARCCS
Confidence
Strength

ARCCS – What’s In The Box?

Questions

Web browser, Spreadsheet, and questionnaire

Scoring

Scoring system per feature

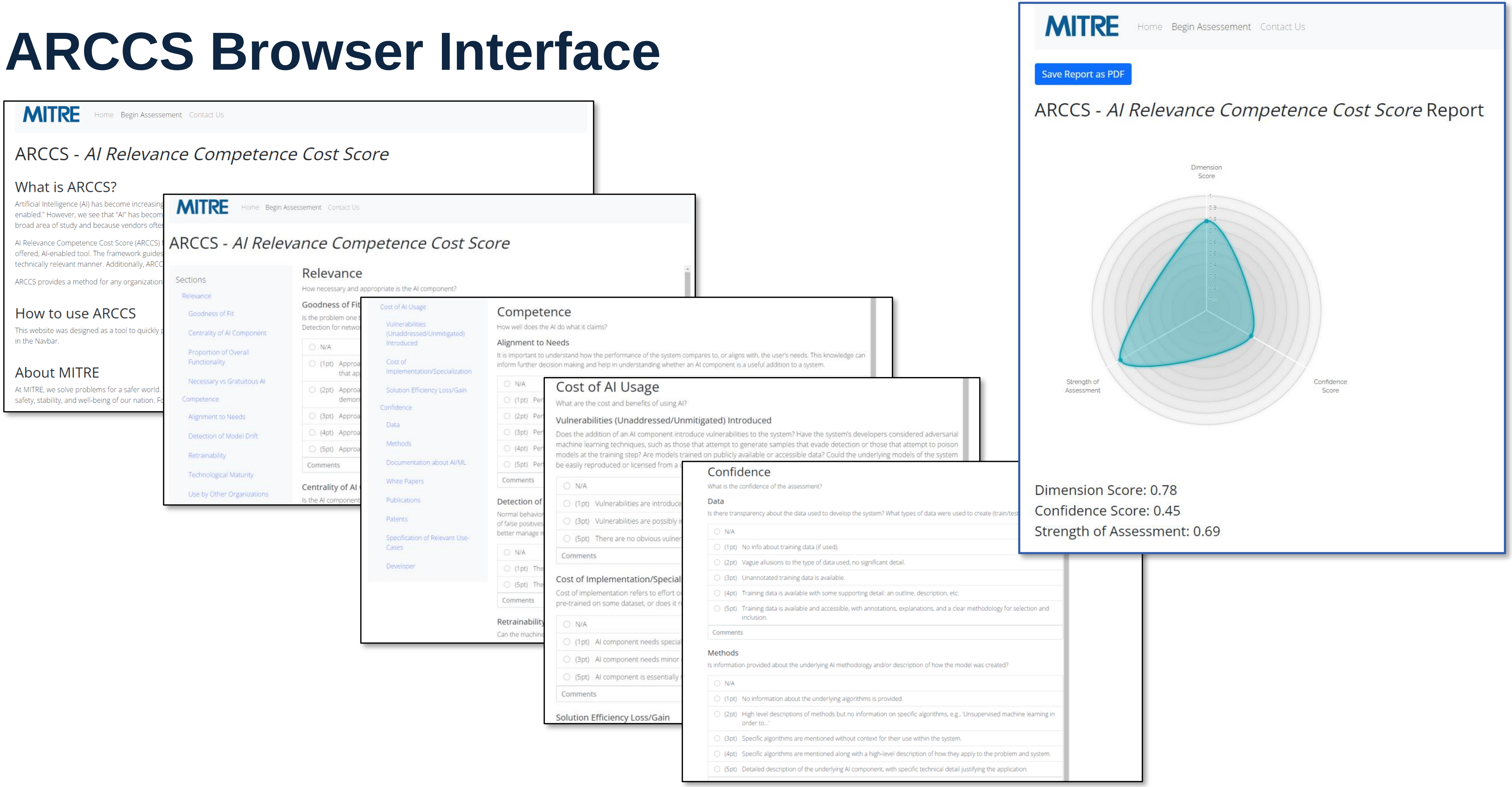
Inference

Inference method on feature scores

Guidance

How-to guidance with guided questions and expected answers

ARCCS Browser Interface



ARCCS - *AI Relevance Competence Cost Score*

What is ARCCS?

Artificial Intelligence (AI) has become increasing
enabled." However, we see that "AI" has becom
broad area of study and because vendors ofter

AI Relevance Competence Cost Score (ARCCS)
offered, AI-enabled tool. The framework guides
technically relevant manner. Additionally, ARCC

ARCCS provides a method for any organization

How to use ARCCS

This website was designed as a tool to quickly
in the Navbar.

About MITRE

At MITRE, we solve problems for a safer world.
safety, stability, and well-being of our nation. Fo

ARCCS - *AI Relevance Competence Cost Score*

Sections

- Relevance
  - Goodness of Fit
  - Centrality of AI Component
  - Proportion of Overall Functionality
  - Necessary vs Gratuitous AI
- Competence
  - Alignment to Needs
  - Detection of Model Drift
  - Retrainability
  - Technological Maturity
  - Use by Other Organizations

Relevance

How necessary and appropriate is the AI component?

Goodness of Fit

Is the problem one
Detection for netwo

○ N/A
○ (1pt) Approa
that ap
○ (2pt) Approa
demon
○ (3pt) Approa
○ (4pt) Approa
○ (5pt) Approa

Comments

Centrality of AI
Is the AI component

- Cost of AI Usage
  - Vulnerabilities (Unaddressed/Unmitigated) Introduced
  - Cost of Implementation/Specialization
  - Solution Efficiency Loss/Gain
- Confidence
  - Data
  - Methods
  - Documentation about AI/ML
  - White Papers
  - Publications
  - Patents
  - Specification of Relevant Use-Cases
  - Developer

Competence

How well does the AI do what it claims?

Alignment to Needs

It is important to understand how the performance of the system compares to, or aligns with, the user's needs. This knowledge can inform further decision making and help in understanding whether an AI component is a useful addition to a system.

○ N/A
○ (1pt) Per
○ (2pt) Per
○ (3pt) Per
○ (4pt) Per
○ (5pt) Per

Comments

Detection of
Normal behavior
of false positives
better manage m

○ N/A
○ (1pt) The
○ (5pt) The

Comments

Retrainability
Can the machine

Cost of AI Usage

What are the cost and benefits of using AI?

Vulnerabilities (Unaddressed/Unmitigated) Introduced

Does the addition of an AI component introduce vulnerabilities to the system? Have the system's developers considered adversarial machine learning techniques, such as those that attempt to generate samples that evade detection or those that attempt to poison models at the training step? Are models trained on publicly available or accessible data? Could the underlying models of the system be easily reproduced or licensed from a c

○ N/A
○ (1pt) Vulnerabilities are introduce
○ (3pt) Vulnerabilities are possibly i
○ (5pt) There are no obvious vulner

Comments

Cost of Implementation/Special
Cost of implementation refers to effort o
pre-trained on some dataset, or does it r

○ N/A
○ (1pt) AI component needs special
○ (3pt) AI component needs minor
○ (5pt) AI component is essentially

Comments

Solution Efficiency Loss/Gain

Confidence

What is the confidence of the assessment?

Data

Is there transparency about the data used to develop the system? What types of data were used to create (train/test

○ N/A
○ (1pt) No info about training data (if used).
○ (2pt) Vague allusions to the type of data used, no significant detail.
○ (3pt) Unannotated training data is available.
○ (4pt) Training data is available with some supporting detail: an outline, description, etc.
○ (5pt) Training data is available and accessible, with annotations, explanations, and a clear methodology for selection and inclusion.

Comments

Methods

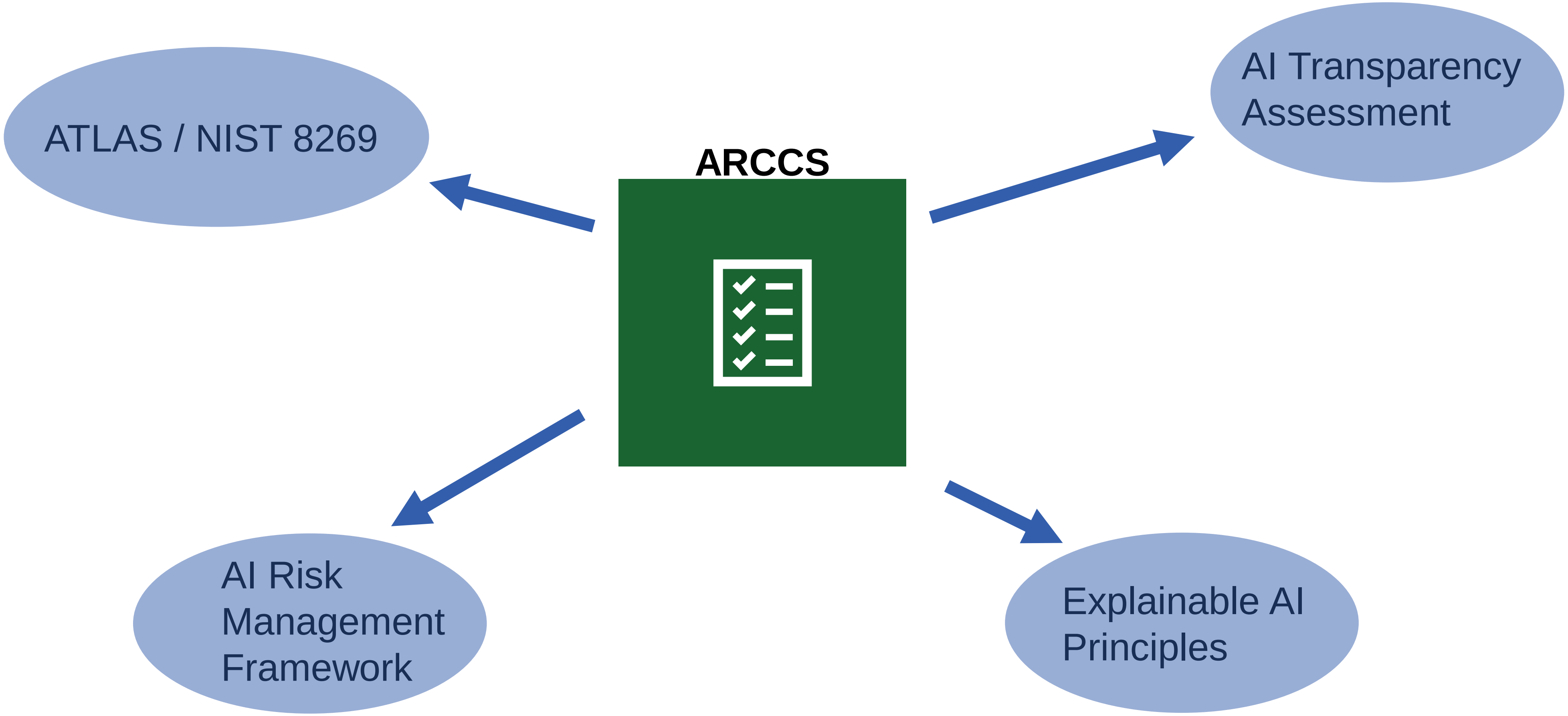
Is information provided about the underlying AI methodology and/or description of how the model was created?

○ N/A
○ (1pt) No information about the underlying algorithms is provided.
○ (2pt) High level descriptions of methods but no information on specific algorithms, e.g., 'Unsupervised machine learning in order to...'
○ (3pt) Specific algorithms are mentioned without context for their use within the system.
○ (4pt) Specific algorithms are mentioned along with a high-level description of how they apply to the problem and system.
○ (5pt) Detailed description of the underlying AI component, with specific technical detail justifying the application.

ARCCS - *AI Relevance Competence Cost Score* Report

Dimension Score: 0.78
Confidence Score: 0.45
Strength of Assessment: 0.69

Coming Version


ATLAS / NIST 8269
ARCCS
AI Transparency Assessment
AI Risk Management Framework
Explainable AI Principles

The Promise of AI Tools

Did You Get What You Paid For?

Or Not? ☹

Take this home

- **Know that not all “AI-enabled” claims are equivalent**
- **Introduce AI-enabled tools assessments to your acquisition process**
- **Purchasers: Learn how to use and apply ARCCS**
- **Vendors: See how your product rates; use results to drive your public documentation**

For a copy of the report/tool
or more information, contact: arccs@mitre.org

Downloads at: https://mitre.github.io/arccs/

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™

Mike Hadjimichael

mikeh@mitre.org

Team contact info

arccs@mitre.org

Get ARCCS

https://mitre.github.io/arccs/

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™