Extending S2C2F to AI Use Cases

This is a community-driven workstream to identify how S2C2F can be extended to provide security guidance for consuming AI use cases.

Work stream leads:

- Please add your name

Work stream contributors:

- Manish Shah, projectvail.org
- Jose Miguel Parrella, Microsoft
- Adrian Diglio, Microsoft

<this is a suggested doc outline. Feel free to revise it as needed>

Purpose

This document is intended to serve as a way for the community to collaboratively work on scoping the problem, and defining how S2C2F can be extended to address AI use cases (i.e. consuming open source models and/or data from public registries such as HuggingFace).

Defining the Scenario and User Workflow

Describe the typical scenario and workflow. Insert text here

Personas:

1. Data scientist. Typically in charge of training or fine tuning a model. Would like to verify integrity, provenance of the inputs to this process, which can be other models and datasets. Also would like to create provenance of their own, binding models, datasets, and code together.

2. Enterprise model catalog owner. Would like to gate admission of models into the enterprise catalog by compliance with internal policies. Would like to verify integrity and provenance, and probably also AIBOM.

3. Inference application operators. Would like to verify the integrity of code and models at deployment time.

4. End users. As they interact with the application, they would like to verify the provenance.

Scenarios:

Here are the list of scenarios that S2C2F could be extended to provide coverage for:

- Data Scientist ingests an open source model from HuggingFace (public model repository???)
-

Define the threats

OWASP LLM Top 10:

- LLM03 - Training Data Poisoning
- LLM05 - Supply Chain Vulnerabilities

Scope

Insert text here

S2C2F Practices

Do the existing set of practices also apply to the AI Use Case?

Recommendations to Mitigate against Threats

This will likely be where we describe the requirements. Requirements fit within a Practice. Insert text here

Organizing the Requirements into Maturity Levels

Insert text here

Implementation Guide

Describe any existing tools (paid or free) that help meet the requirements

Definitions

This section defines the terms and ontology used in this document. These terms align with industry standard terminology and references are cited for traceability.

Data Scientist

Public Model Repository

References

- Guidelines for secure AI system development (ncsc.gov.uk)
- OWASP-Top-10-for-LLMs-2023-v1_1.pdf