Toward Trustworthy AI: An Analysis of Artificial Intelligence (AI) Bill of Materials (AI BOMs)

Technical Report · October 2023

DOI: 10.13140/RG.2.2.18893.61929

2 authors:

Omar Santos
Cisco Systems, Inc
47 PUBLICATIONS **701** CITATIONS
SEE PROFILE

Petar Radanliev
University of Oxford
136 PUBLICATIONS **1,934** CITATIONS
SEE PROFILE

Toward Trustworthy AI: An Analysis of Artificial Intelligence (AI) Bill of Materials (AI BOMs)

Omar Santos[*], Petar Radanliev[**]

[*] osantos@cisco.com, Cisco Systems, Inc.
[**] petar.radanliev@cs.ox.ac.uk, Department of Computer Sciences, University of Oxford

Abstract- As artificial intelligence (AI) and machine learning (ML) become more pervasive, the need for greater transparency and traceability is paramount. This paper aims to explore the concept of AI Bill of Materials (AI BOMs), a system designed to provide a comprehensive inventory of all components in an AI system, much like a traditional Bill of Materials does in manufacturing. By documenting every aspect, including model details, architecture, and usage, AI BOMs serve as a crucial tool for ensuring trust, security, and quality in AI systems.

Index Terms- Artificial Intelligence, Bill-of-Materials (BOMs), Cybersecurity, AI Security

I. Introduction

The rapid adoption of artificial intelligence (AI) and machine learning (ML) technologies has made them integral components of modern life. However, this has also raised important questions regarding their safety, ethics, and transparency. One critical measure that promises to address these concerns is the AI Bill of Materials (AI BOMs). This paper will delve into what AI BOMs are, why they are essential, their primary components, and how they differ from Software Bill of Materials (SBOMs).

II. What is an AI BOM?

Origin and Evolution

The concept of AI BOMs has been influenced by AI Model Cards, introduced by Ezi Ozoani, Marissa Gerchick, and Margaret Mitchell in 2022. Since then, AI BOMs continue to evolve. Manifest (a supply chain security company) also introduced an AI BOM concept that is being suggested to be included in OWASP's CycloneDX [1] and the Linux Foundation also created a project to standardize AI BOMs.

Comparisons with SBOMs

While SBOMs document the components of a software application, AI BOMs are designed to document every facet of an AI system. This includes model details, architecture, usage patterns, training data, and ethical and environmental considerations.

III. Proposed AI BOM Schema

Based on the work from Manifest, the following JSON schema that formalizes the structure of an AI BOM document, including required and optional fields, as well as data types.

```
{
    "$schema": "http://json-schema.org/draft-07/schema#",
```

```
      "type": "object",
      "properties": {
        "ModelDetails": {
          "type": "object",
          "properties": {
            "Name": { "type": "string" },
            "Version": { "type": "string" },
            "Type": { "type": "string" },
            "Author": { "type": "string" },
            "Licenses": { "type": "array", "items": { "type": "string" } },
            "Libraries": { "type": "array", "items": { "type": "string" }, "required": false },
            "Source": { "type": "string" },
            "BOMGeneration": {
              "type": "object",
              "properties": {
                "Timestamp": { "type": "string" },
                "Method": { "type": "string" },
                "GeneratedBy": { "type": "string" }
              },
              "required": false
            },
            "OtherReferences": { "type": "array", "items": { "type": "string" }, "required": false
},
            "Tags": { "type": "array", "items": { "type": "string" }, "required": false }
          },
          "required": ["Name", "Version", "Type", "Author", "Licenses", "Source"]
        },
        "ModelArchitecture": {
          "type": "object",
          "properties": {
            "Datasets": {
              "type": "array",
              "items": {
                "type": "object",
                "properties": {
                  "Name": { "type": "string" },
                  "Source": { "type": "string" },
                  "Usage": { "type": "string" }
                },
                "required": ["Name", "Source"]
              }
            },
            "Architecture": { "type": "string", "required": false },
            "ArchitectureFamily": { "type": "string", "required": false },
            "ParentModel": { "type": "object", "properties": {
                "Name": { "type": "string" },
                "Version": { "type": "string" },
                "Source": { "type": "string" }
              }, "required": false
            },
            "BaseModel": { "type": "object", "properties": {
                "Name": { "type": "string" },
                "Version": { "type": "string" },
                "Source": { "type": "string" }
              }, "required": false
            },
            "Input": { "type": "string" },
            "Output": { "type": "string" },
            "Hardware": { "type": "string", "required": false },
            "Software": { "type": "string", "required": false },
            "SoftwareRequiredForExecution": { "type": "boolean" }
          },
          "required": ["Datasets", "Input", "Output", "SoftwareRequiredForExecution"]
        },
        "Usage": {
          "type": "object",
          "properties": {
            "IntendedUse": { "type": "string" },
```

```
          "OutOfScopeUsage": { "type": "string" },
          "MisuseOrMaliciousUse": { "type": "string" }
        },
        "required": ["IntendedUse", "OutOfScopeUsage", "MisuseOrMaliciousUse"]
      },
      "Considerations": {
        "type": "object",
        "properties": {
          "EnvironmentalImpact": { "type": "string", "required": false },
          "EthicalConsiderations": { "type": "string", "required": false }
        },
        "required": []
      },
      "Authenticity": {
        "type": "object",
        "properties": {
          "Attestation": { "type": "string", "required": false }
        },
        "required": []
      }
    },
    "required": ["ModelDetails", "ModelArchitecture", "Usage"]
  }
```

Note: This schema was introduced in the following GitHub pull request: https://github.com/manifest-cyber/ai-bom/pull/31

Main Components of an AI BOM

The AI BOM is primarily comprised of the following sections: Model Details, Model Architecture, Model Usage, Model Considerations, and Model Authenticity or Attestations. Each serves to provide an exhaustive understanding of the AI system in question.

Schema Metadata

- $schema: Indicates the JSON schema draft version being used, which is Draft-07 in this case.
- type: Specifies the type of the root object, which is an "object".

Root Object

The root object contains the following required properties:
- ModelDetails
- ModelArchitecture
- Usage

ModelDetails Object

This object contains detailed information about the AI model and includes the following properties:

- *Name* (string): Name of the AI model.
- *Version* (string): Version of the model.
- *Type* (string): Type of the model.
- *Author* (string): Author or creator of the model.
- *Licenses* (array of strings): Licenses under which the model is released.
- *Libraries* (array of strings, optional): Libraries used in the model.
- *Source* (string): Source or repository where the model can be found.
- *BOMGeneration* (object, optional): Information about how the BOM was generated.
  - *Timestamp* (string): The time when the BOM was generated.
  - *Method* (string): The method used to generate the BOM.
  - *GeneratedBy* (string): The entity that generated the BOM.

- *OtherReferences* (array of strings, optional): Any other references related to the model.
- *Tags* (array of strings, optional): Tags for additional categorization or metadata.

ModelArchitecture Object

This object provides details about the architecture of the model:

- *Datasets* (array of objects): Information about datasets used in the model.
  - *Name* (string): Name of the dataset.
  - *Source* (string): Source of the dataset.
  - *Usage* (string): How the dataset is used.
- *Architecture* (string, optional): Description of the model architecture.
- *ArchitectureFamily* (string, optional): Family or class of the model architecture.
- *ParentModel* (object, optional): Information about the parent model, if any.
- *BaseModel* (object, optional): Information about the base model, if any.
- *Input* (string): Expected input type or format.
- *Output* (string): Expected output type or format.
- *Hardware* (string, optional): Required hardware for the model.
- *Software* (string, optional): Required software for the model.
- *SoftwareRequiredForExecution* (boolean): Flag indicating if software is needed for execution.

Usage Object

This object details the intended and potential misuses of the model:

- *IntendedUse* (string): Intended usage of the model.
- *OutOfScopeUsage* (string): What the model should not be used for.
- *MisuseOrMaliciousUse* (string): Potential misuses or malicious uses of the model.

Considerations Object

This object contains additional considerations regarding the model:

- *EnvironmentalImpact* (string, optional): Impact of the model on the environment.
- *EthicalConsiderations* (string, optional): Ethical considerations for using the model.

Attestations or Authenticity Object

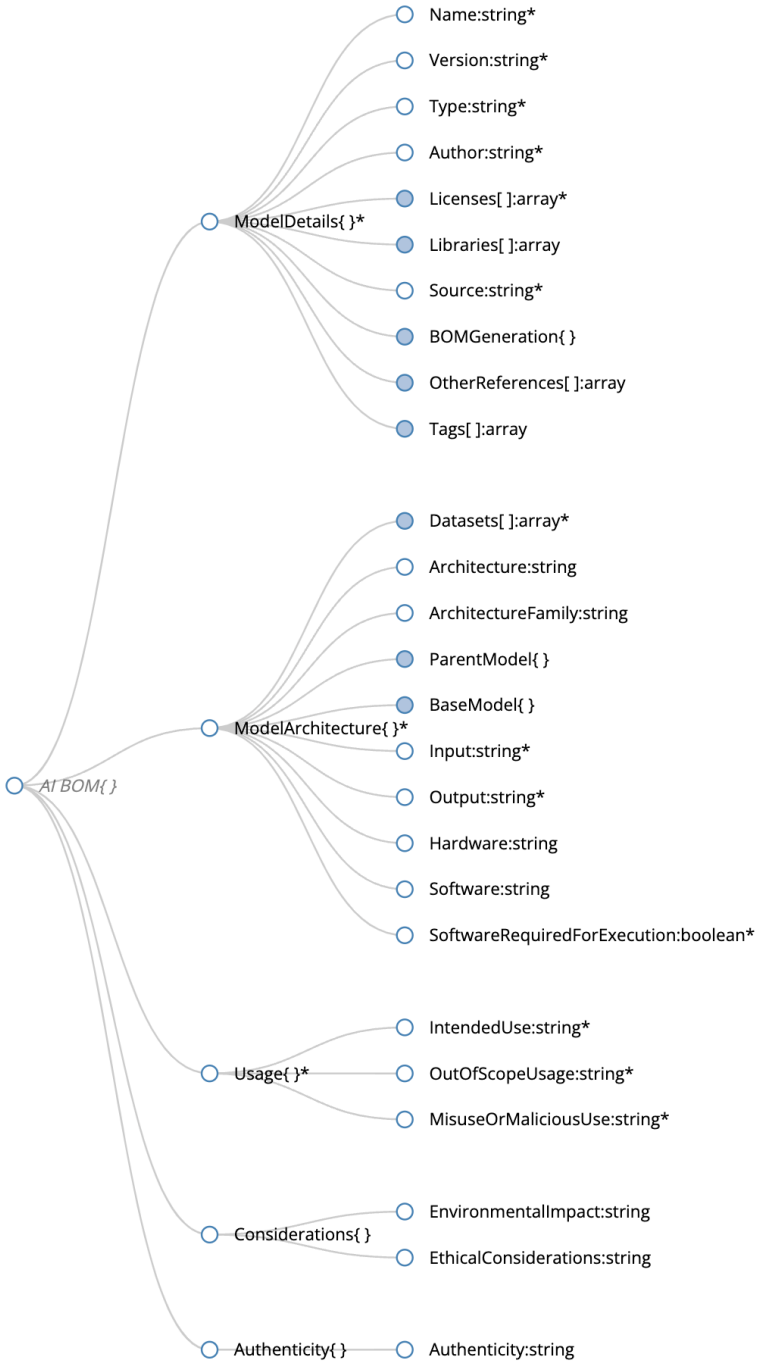
This object provides authenticity information regarding the model:

- *Authenticity* (string, optional): A digital signature, signed by the developer of the model, to ensure the authenticity and integrity of the given AI-BOM.

IV. AI BOM VISUALIZER

A visualizer tool to represent the AI BOM schema visually can be accessed at https://aibomviz.aisecurityresearch.org.

The following is a visual representation of the aforementioned AI BOM schema:



V. Why AI BOMs are Essential

Transparency and Trust

AI BOMs serve as a document of trust between all stakeholders, including users, developers, and auditors. They ensure that every element of an AI solution is transparently presented, fostering greater confidence in the system.

Supply Chain Security and Quality Assurance

With a detailed AI BOM, both developers and auditors can quickly assess the quality, reliability, and security of each component in an AI system.

Troubleshooting

AI BOMs can also facilitate rapid identification of problematic components in cases of system failure or bias, allowing for more effective troubleshooting.

The Future of AI BOMs

AI BOMs are set to become increasingly important as AI systems become more complex and as regulations around AI use loom on the horizon. They will serve as a cornerstone in ensuring responsible development and deployment of AI technologies.

VI. CONCLUSION

The emergence of AI BOMs as a tool for transparency, trust, and security marks a significant stride in responsible AI development. They not only contribute to building safer and more reliable systems but also promote ethical practices in AI development. Appendix

Appendixes, if needed, appear before the acknowledgment.

Acknowledgment

The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments.

References

[1] Santos, O. (2023). Artificial Intelligence Bill-of-Materials (AI BOMs): Ensuring AI Transparency and Traceability. Becoming a Hacker. Retrieved from https://becomingahacker.org/artificial-intelligence-bill-of-materials-ai-boms-ensuring-ai-transparency-and-traceability-82322643bd2a

[2] Manifest Cyber. (2023). AI BOM [Source code]. GitHub. https://github.com/manifest-cyber/ai-bom

[3] Software Package Data Exchange (SPDX). (2023). SPDX AI. https://spdx.dev/learn/areas-of-interest/ai

[4] OWASP. (2023) CycloneDX. https://cyclonedx.org

[5] Santos, O. (2023) AI BOM Visualizer. https://aibomviz.aisecurityresearch.org

Authors

Omar Santos – Distinguished Engineer, Cisco Systems, osantos@cisco.com
Petar Radanliev – Ph.D., Department of Computer Sciences, University of Oxford, petar.radanliev@cs.ox.ac.uk.