MANIFEST

May 2024

# State of AIBOM

## (And some initial efforts)

Daniel Bardenstein
RSA AIBOM Workshop

# whoami

**Co-founder and CTO, Manifest**

**Previous lives:**

- Chief of Tech Strategy @ CISA
- Director of Cyber Programs @ DoD
- Cybersecurity lead @ COVID-19 Vaccines
- Tech Policy Fellow @ Aspen Institute
- OT/ICS Cybersecurity Research @ DoD
- Built enterprise security tools @ Exabeam, Palantir

# 01: State of AIBOM

2 buzzwords put together, or valuable artifact?

# AIBOM: User Research

Research with 100+ security and AI/ML experts.

- Speed of ML (Gen AI, LLM) adoption
- Key concerns and risks of GenAI, LLMs
- Security strategies, existing and planned
- Pain points

White paper available (free).

Driving AI Transparency: the AI Bill of Materials

Achieving transparency for AI-enabled products

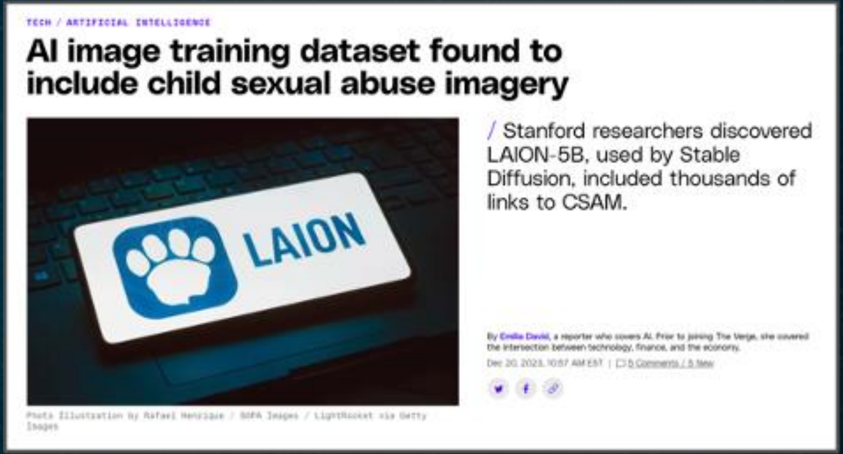
FALL 2023

# The Need Has Arrived

(Unintentionally) compromised models & datasets.

Nation State activity targeting models & datasets.

IP risks have already been realized.



TECH / ARTIFICIAL INTELLIGENCE

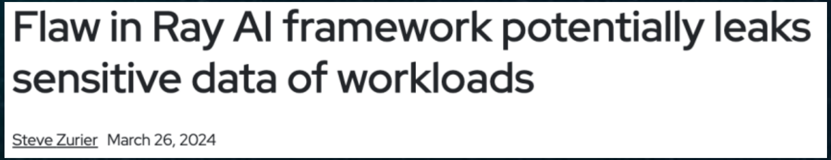
AI image training dataset found to include child sexual abuse imagery

Photo Illustration by Rafael Henrique / SOPA Images / LightRocket via Getty Images

/ Stanford researchers discovered LAION-5B, used by Stable Diffusion, included thousands of links to CSAM.

By Emilia David, a reporter who covers AI. Prior to joining The Verge, she covered the intersection between technology, finance, and the economy.

Dec 20, 2023, 10:57 AM EST | 5 Comments / 5 New



Flaw in Ray AI framework potentially leaks sensitive data of workloads

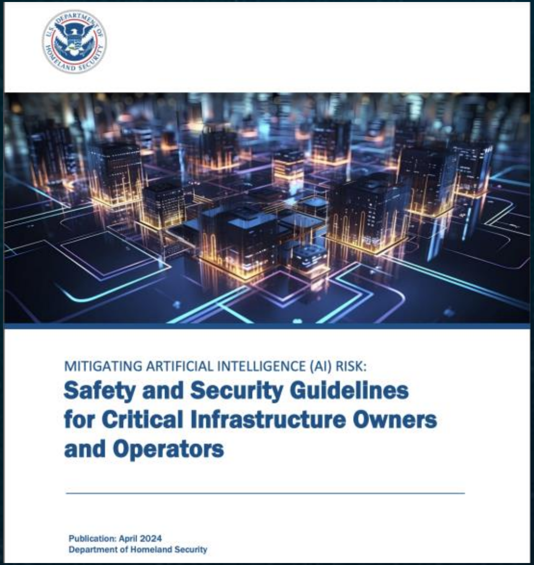
Steve Zurier March 26, 2024

# Regulation: AIBOM ~ SBOM

USG has paraphrased and/or indicated:

***AI is a subset of software. Therefore:***

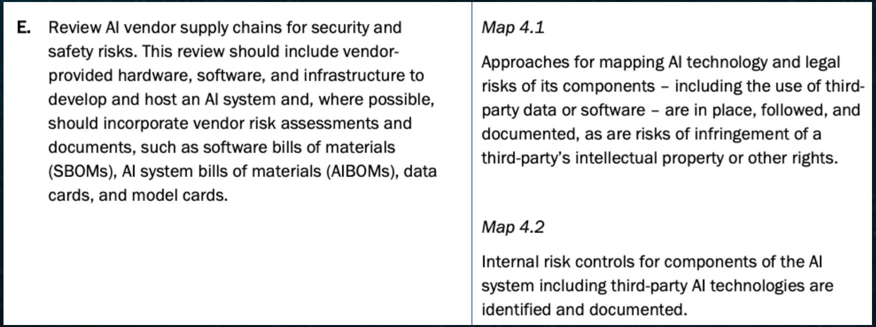
- All Secure By Design principles apply to AI development
- SBOM regulation, guidance includes AI components



MITIGATING ARTIFICIAL INTELLIGENCE (AI) RISK:

**Safety and Security Guidelines for Critical Infrastructure Owners and Operators**

Publication: April 2024
Department of Homeland Security



| | | |
|---|---|---|
| **E.** | Review AI vendor supply chains for security and safety risks. This review should include vendor-provided hardware, software, and infrastructure to develop and host an AI system and, where possible, should incorporate vendor risk assessments and documents, such as software bills of materials (SBOMs), AI system bills of materials (AIBOMs), data cards, and model cards. | *Map 4.1*<br>Approaches for mapping AI technology and legal risks of its components – including the use of third-party data or software – are in place, followed, and documented, as are risks of infringement of a third-party's intellectual property or other rights.<br>*Map 4.2*<br>Internal risk controls for components of the AI system including third-party AI technologies are identified and documented. |

# 02: Early Efforts

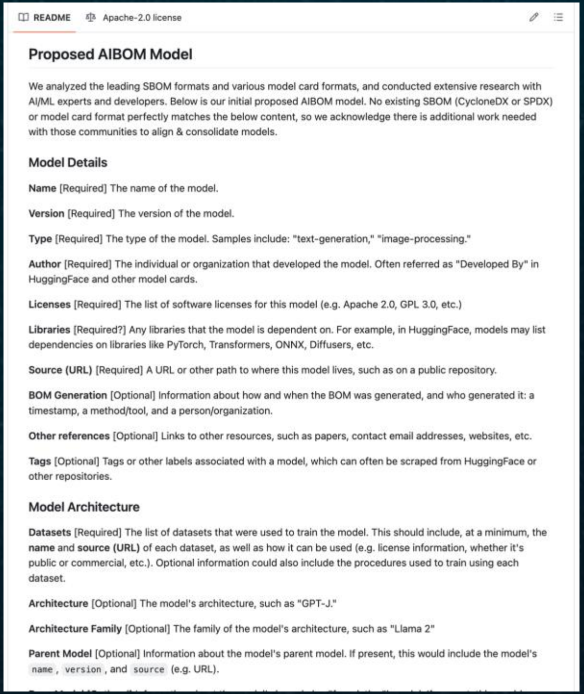
Forward progress

# 1. AIBOM Proposed Fields

An evolving list of proposed fields for models & datasets in AIBOMs.

Working with CycloneDX and SPDX to get new/updated fields into their formats.

View on Github



## Proposed AIBOM Model

We analyzed the leading SBOM formats and various model card formats, and conducted extensive research with AI/ML experts and developers. Below is our initial proposed AIBOM model. No existing SBOM (CycloneDX or SPDX) or model card format perfectly matches the below content, so we acknowledge there is additional work needed with those communities to align & consolidate models.

### Model Details

**Name** [Required] The name of the model.

**Version** [Required] The version of the model.

**Type** [Required] The type of the model. Samples include: "text-generation," "image-processing."

**Author** [Required] The individual or organization that developed the model. Often referred as "Developed By" in HuggingFace and other model cards.

**Licenses** [Required] The list of software licenses for this model (e.g. Apache 2.0, GPL 3.0, etc.)

**Libraries** [Required?] Any libraries that the model is dependent on. For example, in HuggingFace, models may list dependencies on libraries like PyTorch, Transformers, ONNX, Diffusers, etc.

**Source (URL)** [Required] A URL or other path to where this model lives, such as on a public repository.

**BOM Generation** [Optional] Information about how and when the BOM was generated, and who generated it: a timestamp, a method/tool, and a person/organization.

**Other references** [Optional] Links to other resources, such as papers, contact email addresses, websites, etc.

**Tags** [Optional] Tags or other labels associated with a model, which can often be scraped from HuggingFace or other repositories.

### Model Architecture

**Datasets** [Required] The list of datasets that were used to train the model. This should include, at a minimum, the **name** and **source (URL)** of each dataset, as well as how it can be used (e.g. license information, whether it's public or commercial, etc.). Optional information could also include the procedures used to train using each dataset.

**Architecture** [Optional] The model's architecture, such as "GPT-J."

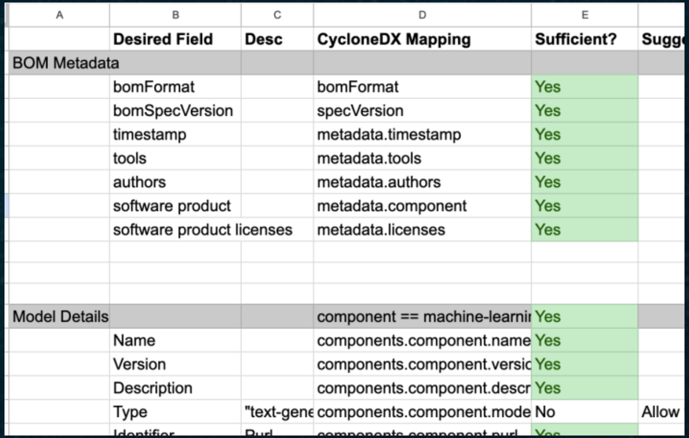
**Architecture Family** [Optional] The family of the model's architecture, such as "Llama 2"

**Parent Model** [Optional] Information about the model's parent model. If present, this would include the model's `name`, `version`, and `source` (e.g. URL).

# 2. Cyclone <> SPDX AIBOM Mapping

In progress. Still working through SDPX 3.0

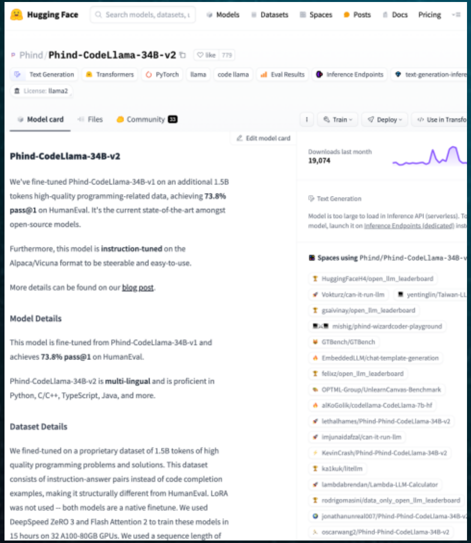
Would love some help!



| A | B | C | D | E | |
|---|---|---|---|---|---|
| | Desired Field | Desc | CycloneDX Mapping | Sufficient? | Sugg |
| BOM Metadata | | | | | |
| | bomFormat | | bomFormat | Yes | |
| | bomSpecVersion | | specVersion | Yes | |
| | timestamp | | metadata.timestamp | Yes | |
| | tools | | metadata.tools | Yes | |
| | authors | | metadata.authors | Yes | |
| | software product | | metadata.component | Yes | |
| | software product licenses | | metadata.licenses | Yes | |
| | | | | | |
| | | | | | |
| | | | | | |
| Model Details | | | component == machine-learnir | Yes | |
| | Name | | components.component.name | Yes | |
| | Version | | components.component.versic | Yes | |
| | Description | | components.component.descr | Yes | |
| | Type | "text-gene | components.component.mode | No | Allow |
| | Identifier | Purl | components.component.purl | Yes | |

# 3. AIBOM Generator

Converts open source models + datasets into valid SBOM json.



## Phind-CodeLlama-34B-v2

We've fine-tuned Phind-CodeLlama-34B-v1 on an additional 1.5B tokens high-quality programming-related data, achieving **73.8% pass@1** on HumanEval. It's the current state-of-the-art amongst open-source models.

Furthermore, this model is **instruction-tuned** on the Alpaca/Vicuna format to be steerable and easy-to-use.

More details can be found on our blog post.

### Model Details

This model is fine-tuned from Phind-CodeLlama-34B-v1 and achieves **73.8% pass@1** on HumanEval.

Phind-CodeLlama-34B-v2 is **multi-lingual** and is proficient in Python, C/C++, TypeScript, Java, and more.

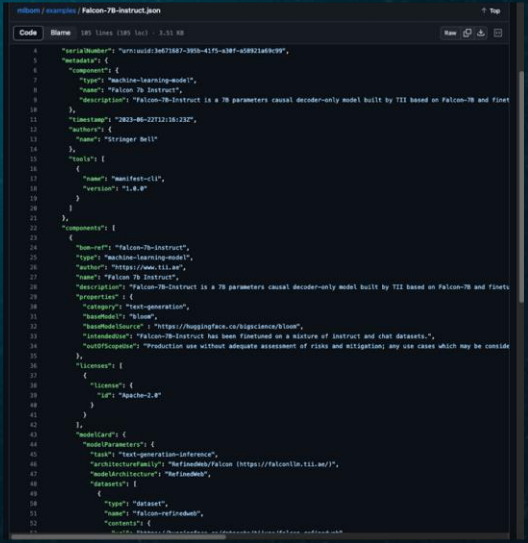
### Dataset Details

We fined-tuned on a proprietary dataset of 1.5B tokens of high quality programming problems and solutions. This dataset consists of instruction-answer pairs instead of code completion examples, making it structurally different from HumanEval. LoRA was not used -- both models are a native finetune. We used DeepSpeed ZeRO 3 and Flash Attention 2 to train these models in 15 hours on 32 A100-80GB GPUs. We used a sequence length of

Model is too large to load in Inference API (serverless). To
model, launch it on Inference Endpoints (dedicated) inst

**Spaces using** Phind/Phind-CodeLlama-34B-v

- HuggingFaceH4/open_llm_leaderboard
- Vokturz/can-it-run-llm
- yentinglin/Taiwan-LL
- gsaivinay/open_llm_leaderboard
- mishig/phind-wizardcoder-playground
- GTBench/GTBench
- EmbeddedLLM/chat-template-generation
- felixz/open_llm_leaderboard
- OPTML-Group/UnlearnCanvas-Benchmark
- alKoGolik/codellama-CodeLlama-7b-hf
- lethalhames/Phind-Phind-CodeLlama-34B-v2
- imjunaidafzal/can-it-run-llm
- KevinCrash/Phind-Phind-CodeLlama-34B-v2
- ka1kuk/litellm
- lambdabrendan/Lambda-LLM-Calculator
- rodrigomasini/data_only_open_llm_leaderboard
- jonathanunreal007/Phind-Phind-CodeLlama-34B-v2
- oscarwang2/Phind-Phind-CodeLlama-34B-v2


mlbom / examples / Falcon-7B-instruct.json
Top
Code
Blame
105 lines (105 loc) · 3.51 KB
Raw

```
  "serialNumber": "urn:uuid:3e671687-395b-41f5-a30f-a58921a69c99",
  "metadata": {
    "component": {
      "type": "machine-learning-model",
      "name": "Falcon 7b Instruct",
      "description": "Falcon-7B-Instruct is a 7B parameters causal decoder-only model built by TII based on Falcon-7B and finet
    },
    "timestamp": "2023-06-22T12:16:23Z",
    "authors": {
      "name": "Stringer Bell"
    },
    "tools": [
      {
        "name": "manifest-cli",
        "version": "1.0.0"
      }
    ]
  },
  "components": [
    {
      "bom-ref": "falcon-7b-instruct",
      "type": "machine-learning-model",
      "author": "https://www.tii.ae",
      "name": "Falcon 7b Instruct",
      "description": "Falcon-7B-Instruct is a 7B parameters causal decoder-only model built by TII based on Falcon-7B and finetu
      "properties" : {
        "category": "text-generation",
        "baseModel": "bloom",
        "baseModelSource" : "https://huggingface.co/bigscience/bloom",
        "intendedUse": "Falcon-7B-Instruct has been finetuned on a mixture of instruct and chat datasets.",
        "outOfScopeUse": "Production use without adequate assessment of risks and mitigation; any use cases which may be conside
      },
      "licenses": [
        {
          "license": {
            "id": "Apache-2.0"
          }
        }
      ],
      "modelCard": {
        "modelParameters": {
          "task": "text-generation-inference",
          "architectureFamily": "RefinedWeb/Falcon (https://falconllm.tii.ae/)",
          "modelArchitecture": "RefinedWeb",
          "datasets": [
            {
              "type": "dataset",
              "name": "falcon-refinedweb",
              "contents": {
```

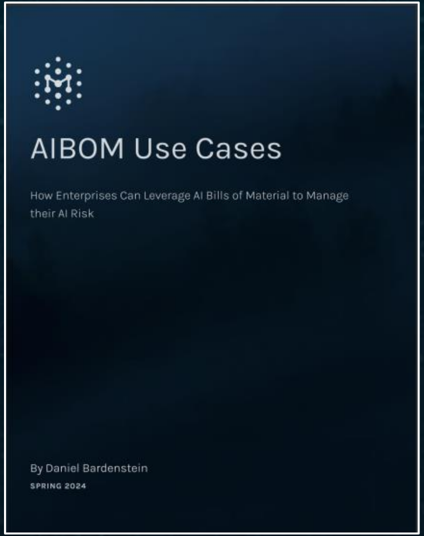
Scan the QR code to join the beta program

# 4. AIBOM Use Cases for the Enterprise

Four key use cases:

1. Managing open source models/datasets
2. Procurement / TPRM
3. Model/dataset inventory + links
4. AI incident response

White paper in progress.



AIBOM Use Cases

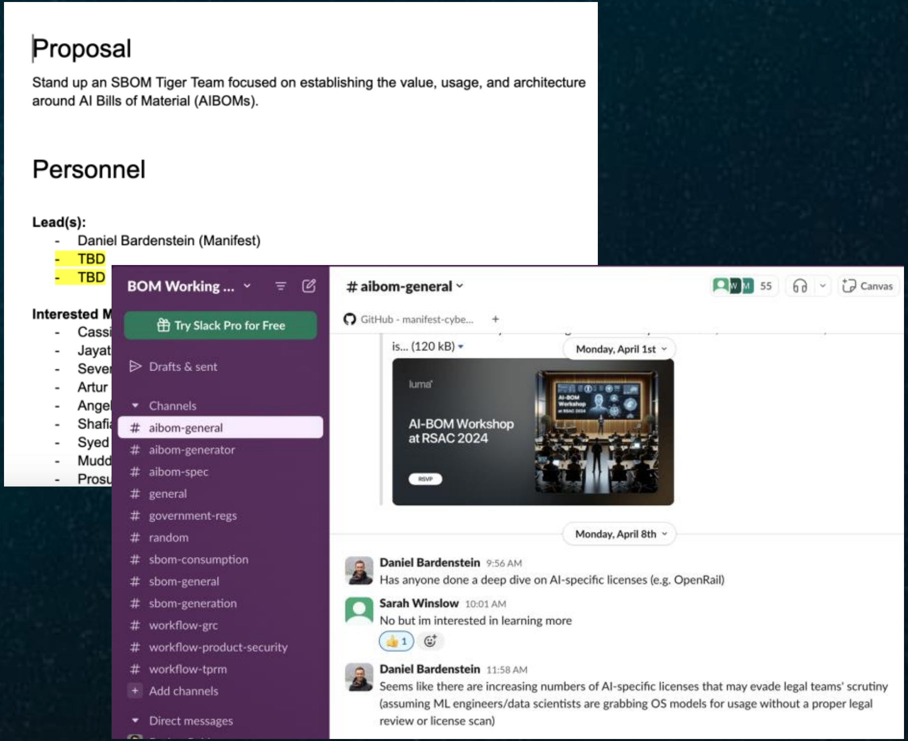
How Enterprises Can Leverage AI Bills of Material to Manage their AI Risk

By Daniel Bardenstein

SPRING 2024

# 5. Building Community

- Slack: BOM Working Community
- CISA: proposed AIBOM Tiger Team
- Collaboration



## Proposal

Stand up an SBOM Tiger Team focused on establishing the value, usage, and architecture around AI Bills of Material (AIBOMs).

## Personnel

**Lead(s):**

- Daniel Bardenstein (Manifest)
- TBD
- TBD

**Interested M**

- Cassi
- Jayat
- Sever
- Artur
- Ange
- Shafi
- Syed
- Mudd
- Prosu

Monday, April 1st

Monday, April 8th

**Daniel Bardenstein** 9:56 AM
Has anyone done a deep dive on AI-specific licenses (e.g. OpenRail)

**Sarah Winslow** 10:01 AM
No but im interested in learning more
👍 1

**Daniel Bardenstein** 11:58 AM
Seems like there are increasing numbers of AI-specific licenses that may evade legal teams' scrutiny (assuming ML engineers/data scientists are grabbing OS models for usage without a proper legal review or license scan)

# Get involved!

Github
(use cases, proposed fields)

Join the Slack Community

Interest in the CISA Tiger Team or beta-testing AIBOM generator

MANIFEST

# Thank You

**Daniel Bardenstein**
CTO, co-founder
daniel@manifestcyber.com

# Where do go from here?

Github
(use cases, proposed fields)

Join the Slack Community

Interest in the CISA Tiger Team or
beta-testing AIBOM generator