Utilizing SBOM for Transparent AI Risk Communication

Lennard Helmer[1], **Lisa Fink**[1], **Maximilian Poretschkin**[1,2,3]

[1]Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Sankt Augustin, Germany
[2]Lamarr Institute for Machine Learning and Artificial Intelligence, Sankt Augustin, Germany
[3]University of Bonn, Bonn, Germany
{firstname.lastname}@iais.fraunhofer.de

Abstract

Value chains for AI systems are becoming increasingly complex and can consists of multiple actors that contribute services, tools, data, models and code. An efficient risk management along this value chain requires all actors to communicate potential risk sources and recommendations for mitigation. The Software Bill of Materials (SBOM) is a method from cybersecurity, that enables organizations to communicate information like licences, security vulnerabilities and dependencies of software components. SBOM raises increasing interest in the AI community to share information about AI components, like data and models. In this paper we discuss the suitability of SBOM for AI risk management along a value chain and show the potential but also gaps in current approaches.

Introduction

The development of modern AI systems relies on distributed value chains involving multiple internal and external actors contributing software, data, models, services and hardware (Engler and Renda 2022; Widder and Nafus 2023; Thomas et al. 2024). While this enables rapid integration of specialized components, it also introduces a high degree of opacity: actors often do not know about all upstream and downstream actors and their contributions, which creates unique challenges when it comes to identifying sources of harm, assigning accountability, or implementing effective mitigation—posing challenges for users, providers, and deployers alike (AI Assurance Club 2024).

Although the need for risk management is well known for software projects and in modern project management, recent efforts have highlighted the breadth and complexity of AI-related risks. The MIT AI Risk Repository (2024) cataloged over 1,000 AI-specific risks and proposed a taxonomy spanning seven domains and 24 subdomains, including discrimination, toxicity, and socioeconomic harms (Slattery et al. 2024). These risks can arise at any stage of the AI life cycle and manifest as technical malfunctions, undesirable system properties, or negative impacts (Schmitz et al. 2025), necessitating coordinated risk communication and governance across all contributors in the value chain (AI Assurance Club 2024).

Regulatory frameworks such as the EU AI Act and AI-related standards such as ISO 23984 recognize this issue. They require providers to obtain sufficient information about system components in order to assess and mitigate potential risks. The lately publicized Code of Practice for General-Purpose AI Models[1] from the EU required signatories to provide additional information to downstream providers to enable them to have a good understanding of the capabilities and limitations of the general-purpose AI model relevant for its integration into the downstream providers' AI system. However, its not specified how to do so efficiently. Furthermore, in practice, providers often lack knowledge about upstream contributors, as well as all relevant downstream providers, and whether appropriate risk mitigation measures have been implemented by third parties. This creates a pressing need for solutions that enabled efficient sharing of of risk-relevant information across organizational boundaries.

In this paper we describe an approach to leverage the Software Bill of Materials (SBOM) to support the communication of AI-specific risk causes. Originally developed for software engineering in cybersecurity and license management, SBOMs are used to share information about licenses, vulnerabilities, and the relationships between system components. Specifications, like SPDX and CycloneDX, already enhance their frameworks with AI specific data field and support the communication about datasets and models. Furthermore, the G7 Cybersecurity Working Group has advocated for the introduction of a concept for an SBOM for AI (G7 Cybersecurity Working Group 2025), envisioned as a structured record of details and supply chain relationships for the various components used in building an AI system with the goal of enhancing transparency, traceability, and ultimately the security of AI systems through standardized component tracking (G7 Cybersecurity Working Group 2025).

We argue that SBOMs are especially suitable for communicating risk causes in AI systems. By embedding risk-relevant metadata into SBOMs, stakeholders can gain visibility into potential sources of harm and make informed decisions about risk mitigation. The remainder of this paper is structured as follows: In chapter 1, we will discuss related work and how SBOM is already used in AI; in chapter

[1]https://digital-strategy.ec.europa.eu/de/policies/contents-code-gpai Accessed 30.07.25

2 we will describe the SBOM methodology in more detail and the two main specifications. In chapter 3, we will show how SBOM can be leveraged for risk cause communication along a value chain and discuss the approach, its limitations and future work in chapter 4.

Related Work

SBOMs are attracting increasing interest from researchers in various fields, such as international affairs (Radanliev 2025), defense (Beninger et al. 2024) and industry (Stalnaker et al. 2024). Studies show that practitioners are aware of SBOM (Stalnaker et al. 2024; Xia et al. 2023) but still face challenges. Participants expressed interest in using SBOMs for reporting risks, e.g. the limitations of software with regard to its suitability for certain tasks due to security risks (Stalnaker et al. 2024). One of the concerns about SBOMs is that the current standards are not meeting requirements in regard to specific use cases (Xia et al. 2023). Maybe in response, more specialized SBOMs are developed, e.g for data, to capture dependencies and stakeholder of datasets (Liu et al. 2024). Also AI specific SBOMs are being suggested. In (Xia et al. 2024), the authors describe how AI-specific SBOMs can encompass the potential risks of AI systems that could arise from trade-offs in development or context-specific assessments. They argue that AI-specific SBOMs could facilitate informed decision-making by being used to communicate about risks.

The two dominant specifications take steps into integrating AI specific information into their models. For example, CycloneDX integrates machine learning-related information into its SBOM specification, e.g reference to a model card[2] and SPDX developed AI profiles to enhance the transparency of AI components [3].

While it is clear that SBOMs are gaining interest within the AI community, and that they could support the robustness and security of AI systems by communicating known risks, to the best of our knowledge, no clear recommendations on how to achieve this using SBOMs have yet been made.

Software Bill of Materials

In response to Executive Order 14028 on improving the Nation‘s cybersecurity, the National Telecommunications and Information Administration (NTIA) defined SBOM as formal record containing the details and supply chain relationships of components used for building complex software systems (NTIA 2021). By documenting the origin, versioning, dependencies, and licensing of each component, SBOMs provide downstream actors with essential insights into the composition and provenance of the software they rely on. This transparency is particularly valuable in distributed development environments, where components are sourced from external suppliers. SBOMs allow stakeholders to track updates, assess compatibility, and identify potential vulnerabilities without requiring direct communication between producers and consumers. While one could argue that a well done technical documentation could serve equal means, the benefit of SBOM is the standardized format that is used. By leveraging key-value pairs, following either the XML or JSON format, SBOMs are machine readable and can be consumed and automatically processes (and created) by a variety of tools. In its report, the NTIA specifies the minimum elements for a SBOM, like *Component Name*, *Version of the Component*, and *Dependencies*. These elements form the basis for traceability and risk assessment in software supply chains.

Two dominant SBOM specifications have emerged: SPDX and CycloneDX, each offering distinct strengths and extensibility features relevant to AI systems. CycloneDX focuses on information sharing about vulnerabilities and potential mitigation measures and SPDX incorporated the AI specific AI-BOM (Bennet et al. 2025) to enable producers of AI components to share information about a variety of information related to the AI functionalities.

SPDX

Developed by the Linux foundation, the System Package Data Exchange (SPDX) specification is an open standard for communicating bill of materials information. It aims to enhance the transparency of software systems and to ease the management licenses. SPDX follows a graph-oriented structure, meaning that it structures information as nodes and relationships between those objects as edges. This allows it to visualize the relationships between different objects and components of the system. The SPDX specification was codified in an international open standard (ISO 2021)

CycloneDX

Developed by the Open Web Application Security Project (OWASP), CycloneDX was developed to support users in the identification of risks in the value chain of software systems. Vulnerability description remains one of its primary use cases and allows the incorporation of many different information related to software vulnerabilities.

In the context of AI, CycloneDX introduced the Machine Learning BOM (ML-BOM), which aims to standardize model cards and provide an inventory of models and datasets. CycloneDX is designed to be highly extensible, offering extension points within the object model that support domain-specific adaptations, including those required for AI governance and compliance.

SBOM for Risk Communication

AI Risk Management

An AI system is a software system that may be integrated into a larger system, which potentially consists of other components, such as hardware or sensors, and it can consist of multiple AI models that are connected to achieve a certain objective (Schnitzer, Hapfelmeier, and Zillner 2025). The components that define the inner workings of an AI system can be either AI model components, that receive input data in order to infer output data. And non-AI components, which

[2]https://cyclonedx.org/guides/OWASP_CycloneDX-Authoritative-Guide-to-SBOM-en.pdf Accessed: 26.07.25

[3]https://spdx.dev/learn/areas-of-interest/ai/ Accessed: 26.07.25

Domain/ Subdomain
1. Discrimination & Toxicity
1.1 Unfair discrimination and misrepresentation
1.2 Exposure to toxic content
1.3 Unequal performance across groups

Table 1: Example for one domain of the MIT risk taxonomy

Domain/ Subdomain
1. discrimination-toxicity
1.1 discrimination-misrepresentation
1.2 toxic-content
1.3 unequal-performance

Table 2: Example on how to use the MIT risk taxonomy in SBOM

are all components that were developed without AI techniques (Schnitzer, Hapfelmeier, and Zillner 2025). These different components can be developed and contributed by different actors, and it is necessary to gain a holistic perspective on all components and their respective strengths and limitations, especially because AI systems introduce unique risks due to their probabilistic nature.

The MIT risk repository (Slattery et al. 2024) identified over 1,000 AI-specific risks through a comprehensive literature review and the management of AI-related risks requires a deep understanding of the inner working of an AI system.

The goal of risk management is to prevent harmful risks from occurring, and planning for an appropriate response if that fails (Gillanders 2003). Effective risk analysis requires structured information about existing risk sources, which should as a minimum include (Gillanders 2003):

- Cause: Definite events or sets of circumstances that exists within a system
- Risk description: Uncertain events that produce (harmful) impacts
- Effect: Anticipated impact if the risk is occurring

While risk mitigation is of high importance, it might not always be feasible or appropriate for every actor in the value chain to spend resources on it, even if a providers becomes aware of a potential risk in its system. Many risks are context-dependent. For example, the representativeness of training data becomes critical only when the AI system affects underrepresented user groups. While a provider could be aware of missing evaluations of the system for certain subgroups, he might not take any measures to mitigate these risks. However, by documenting known limitations and risk causes, and sharing this information with downstream users, the provider can still contribute meaningfully to the robustness of the system and overall risk management process. This collaborative approach enables developers to assess whether these known risks are relevant to their specific use case and to implement appropriate mitigation strategies, enhancing the risk documentation with their own input.

AI and SBOM

In complex systems, consisting of many components, it is essential to have a structured, machine-readable format for communicating such risks. This enables automated analysis, monitoring, and updating of risk information, particularly when components are updated or newly integrated. SBOM is highly suitable for this purpose. Although they have not yet having appended their standard model with AI-specific risk fields, CycloneDX provides extension points within its specification to support specialized and future use cases, thereby benefiting a wider community (OWASP 2024). The project explicitly encourages the development of extensions targeting specialized or industry-specific use cases. Using CycloneDX properties allows users to represent complex data in the BOM that is not provided by the standard model. At their core, CycloneDX properties are name-value stores. Unlike key-value stores, they support duplicate names, with each name potentially having a different values.

SPDX 3.0.1 allows developers to specify the safety risk level (*SafetyAssessmentType*), in accordance with the EU's general risk assessment methodology, inform about the usage of sensitive personal data (*UseSensitivePersonalInformation*) and limitations (*limitations*). A free-text field that can be used to capture the limitations of the AI package, e.g. that a dataset is not representative for certain subgroup[4]. To capture the result of a general risk assessment, *safetyRiskAssessment* can be used. Although it is promising that the risk transparency is already included in the standard model and could be used to capture the information about risks that we suggested, we believe that free-text fields make it difficult to analyze SBOMs for AI-related risks using tools and are prone to processing errors.

Our Approach

In the following, we describe how SBOM can (already) be used to communicate about risks, and recommend enhancing both standard models. Two types of information must be shared along a value chain. Firstly, risk sources must be identifiable by being assigned an identifier. This allows potential mitigation measures to be monitored, relevant risks to be analyzed automatically, and communication between actors to be supported. In our approach, we leverage the MIT risk taxonomy, resulting from the MIT AI risk repository (Slattery et al. 2024). Based on an enhanced analysis of risks that were defined in the literature, the taxonomy is organized into seven domains and 24 subdomains, covering areas such as discrimination, toxicity, and socioeconomic or environmental harm. Table 1 illustrates an example domain. Using a common taxonomy enables each risk source to be tagged with a well-defined identifier. For instance, a risk which is related to the representativeness of the data would be categorized under the domain *Discrimination & Toxicity* and the subdomain *Unfair discrimination and misinterpretation*. In table 2 we show an example, on how such an identifier could be defined to comply with Cyclone DX-specific naming requirements.

[4]https://spdx.dev/wp-content/uploads/sites/31/2024/12/SPDX-3.0.1-1.pdf, Accessed 28.07.2025

Algorithm 1: Example on how to use CycloneDX properties

```
{
  "properties": [
    {
      "name": "ai-risk:<domain>:<
          subdomain>",
      "value": "Cause: ...; Risk: ...;
          Effect: ...; Recommendation:
          ...;"
    }
  ]
}
```

Secondly, in order to enable an appropriate response it is necessary to provide additional information about the risk, namely the already mentioned *cause*, *risk description*, and *effect*. In addition to the information required for assessing the risk, we recommend adding *recommendation* to the information set. As we intend to communicate with other actors in the value chain, it is beneficial to share additional information that might be useful, like potential mitigation strategies or references to additional information sources.

In Algorithm 1, we present the recommended structure tailored to the CycloneDX properties field in the standard model. The identifier based on the MIT risk taxonomy is added to the key field 'name'. In a complex system, consisting of multiple SBOMs for different components, a provider can easily identify all risk sources related to a certain AI risk that is relevant for the use case. The 'value' field can be used to store the meta information about the risk.

We evaluate our approach against five requirements that a structured AI system description needs to fulfill in order to support an assessor during a risk assessment (Schnitzer, Hapfelmeier, and Zillner 2025) as these requirements are equally relevant for a structured description of AI risks.

Flexibility Across Diverse AI Systems: The method is flexible across diverse AI systems, because the structure and format are standardized yet still adaptable, and it is already in use for a wide range of traditional software systems. Furthermore, its format allows to be used for legacy systems and components.

Ensuring Reproducibility: The method ensures reproducibility, as SBOM is intended to be reproduced for each new version of a component and the automated creation is supported by a large variety of tools.

Integration with Risk Assessment: The method integrates well with risk assessment efforts, as it enhances the transparency in complex systems, enabling developers to mitigate relevant risks and assessors to track mitigation measures back to the risk source.

Grounded in best Practices: Our approach is firmly grounded in best practices, as we have tried to leverage and combine existing methods and taxonomies as much as possible.

Reflecting AI Specific Characteristics: The method reflects AI-specific characteristics, as the strategy for creating identifiers is based in an AI-specific taxonomy, enabling developers to assess their systems based on AI-specific risk sources and risks.

Discussion and Conclusion

This paper has argued that Software Bills of Materials (SBOMs), originally developed to enhance transparency and security in software supply chains, offer a promising foundation for communicating AI-specific risk sources across complex value chains. In their current form, SBOM specifications such as SPDX and CycloneDX provide extensibility mechanisms that can be leveraged to encode structured risk-related metadata.

To fully realize this potential, we propose that SBOM specification models be extended to support the communication of risk sources. At a minimum, this should include standardized fields for:

- Cause
- Risk description
- Effect
- Recommendation

Such structured information would enable actors along the AI value chain to identify known risk sources, assess their relevance in specific deployment contexts, and implement appropriate mitigation measures. Moreover, the development of centralized, machine-readable repositories-similar to the Common Vulnerabilities and Exposures (CVE) database in cybersecurity could further enhance the utility of SBOMs by enabling the reference to publicly documented risks and vulnerabilities in AI systems.

In summary, extending SBOMs to include AI risk metadata represents a scalable and interoperable approach to fostering transparency, accountability, and robustness in AI systems. It empowers all actors along the AI value chain to contribute to and benefit from shared infrastructure and information for AI risk governance.

Acknowledgments

This research has been funded by the Fraunhofer Cluster of Excellence Cognitive Internet Technologies CCIT and the Ministry of Economic Affairs, Innovation, Digitalization and Energy of the State of North Rhine-Westphalia as part of the flagship project ZERTIFIZIERTE KI

References

AI Assurance Club. 2024. MAGF – A Call and Proposal for Assurance Information Sharing Standards. https://intaigovassoc.org/ai-value-chain. Accessed: 2025-07-14.

Beninger, M.; Charland, P.; Ding, S. H.; and Fung, B. C. 2024. ERS0: Enhancing Military Cybersecurity with AI-Driven SBOM for Firmware Vulnerability Detection and Asset Management. In *2024 16th International Conference on Cyber Conflict: Over the Horizon (CyCon)*, 141–160. IEEE.

Bennet, K.; Rajbahadur, G. K.; Suriyawongkul, A.; and Stewart, K. 2025. Implementing AI Bill of Materials (AI BOM) with SPDX 3.0: A Comprehensive Guide to Creating AI and Dataset Bill of Materials. *arXiv preprint arXiv:2504.16743*.

Engler, A. C.; and Renda, A. 2022. *Reconciling the AI Value Chain with the EU's Artificial Intelligence Act*. CEPS.

G7 Cybersecurity Working Group. 2025. A Shared G7 Vision On Software Bill of Materials for AI. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/SBOM-for-AI_Food-for-thoughts.pdf?__blob=publicationFile&v=6. Accessed: 2025-07-14.

Gillanders, C. 2003. When Risk Management turns into Crisis Management. In *AIPM National Conference, Sydney, Australia*. Citeseer.

ISO. 2021. Information technology — SPDX® Specification V2.2.1. Standard, International Organization for Standardization, Geneva, CH.

Liu, Y.; Zhang, D.; Xia, B.; Anticev, J.; Adebayo, T.; Xing, Z.; and Machao, M. 2024. Blockchain-Enabled Accountability in Data Supply Chain: A Data Bill of Materials Approach. In *2024 IEEE International Conference on Blockchain (Blockchain)*, 557–562. IEEE.

NTIA. 2021. The Minimum Elements For a Software Bill of Materials (SBOM). https://www.ntia.gov/files/ntia/publications/sbom_minimum_elements_report.pdf. Accessed: 2025-07-17.

OWASP. 2024. Authortative Guide to SBOM: Implement and optimize use of Software Bill of Materials. https://cyclonedx.org/guides/OWASP_CycloneDX-Authoritative-Guide-to-SBOM-en.pdf. Accessed: 2025-07-23.

Radanliev, P. 2025. Cyber diplomacy: defining the opportunities for cybersecurity and risks from Artificial Intelligence, IoT, Blockchains, and Quantum Computing. *Journal of Cyber Security Technology*, 9(1): 28–78.

Schmitz, A.; Mock, M.; Görge, R.; Cremers, A. B.; and Poretschkin, M. 2025. A global scale comparison of risk aggregation in AI assessment frameworks. *AI and Ethics*, 5(2): 1407–1432.

Schnitzer, R.; Hapfelmeier, A.; and Zillner, S. 2025. EAM Diagrams-A Framework to Systematically Describe AI Systems for Effective AI Risk Assessment (Academic Track). In *Symposium on Scaling AI Assessments (SAIA 2024)*, 3–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Slattery, P.; Saeri, A. K.; Grundy, E. A.; Graham, J.; Noetel, M.; Uuk, R.; Dao, J.; Pour, S.; Casper, S.; and Thompson, N. 2024. The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence. *arXiv preprint arXiv:2408.12622*.

Stalnaker, T.; Wintersgill, N.; Chaparro, O.; Di Penta, M.; German, D. M.; and Poshyvanyk, D. 2024. BOMs Away! Inside the Minds of Stakeholders: A Comprehensive Study of Bills of Materials for Software Systems. In *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*, 1–13.

Thomas, C.; Roberts, H.; Mökander, J.; Tsamados, A.; Taddeo, M.; and Floridi, L. 2024. The case for a broader approach to AI assurance: addressing "hidden" harms in the development of artificial intelligence. *AI & SOCIETY*, 1–16.

Widder, D. G.; and Nafus, D. 2023. Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data & Society*, 10(1): 20539517231177620.

Xia, B.; Bi, T.; Xing, Z.; Lu, Q.; and Zhu, L. 2023. An Empirical Study on Software Bill of Materials: Where We Stand and the Road Ahead. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, 2630–2642. IEEE.

Xia, B.; Zhang, D.; Liu, Y.; Lu, Q.; Xing, Z.; and Zhu, L. 2024. Trust in Software Supply Chains: Blockchain-Enabled SBOM and the AIBOM Future. In *Proceedings of the 2024 ACM/IEEE 4th International Workshop on Engineering and Cybersecurity of Critical Systems (EnCyCriS) and 2024 IEEE/ACM Second International Workshop on Software Vulnerability*, 12–19.