D5.7 TAIBOM Industry Recommendations

March 2025

COPPER HORSE

CONTENTS

Introduction

This document provides a set of recommendations based on the experience gained in the Innovate UK funded TAIBOM project. It is not exhaustive and is intended to provide an initial set of inputs to future roadmap planning by the project participants and to provide information for industry more broadly.

Recommendations

1. Understand and co-opt other emergent initiatives

The AI space is evolving at an incredible pace. As such there are other emergent activities and standards that should be considered that could be complementary to TAIBOM and that would enhance its effectiveness. A couple of examples are listed below:

IETF SCITT

An interesting project that could integrate well with TAIBOM is the IETF Supply Chain Transparency and Trust (SCITT) work[1]. The initiative may be able to complement TAIBOM by providing a method for transferring secured information about the state of an AI model, without having to transfer the entirety of the data for validation by third parties. This reduces workload and could potentially improve trust and accountability in AI systems.

Model Cards

The purpose of Model Cards is to aid informed use of an AI model. They can be used to aid SBOMs/AIBOMs with detailed information about the model. On top of the information on developer, model version, origin of training data, etc, generally they include specific fields for describing the area of scope, steps taken to mitigate unwanted bias, performance metrics, intended users and ethical considerations.

Many open source tools and platforms for managing model cards and their standards are available (for example Hugging face, Kaggle, Google Model Cards, AWS SageMaker, etc) at different stages of maturity.

They are a useful resource especially for the users' informed and trusted selection of pre-trained models. The tracking of original training dataset, model history and

[1] https://datatracker.ietf.org/group/scitt/about/

training performances, allows the insightful assessment of possible biases and gives a complete picture of the model status, which is essential for developing their contingent fine-tuning.

Reference examples:

- Hugging Face - https://huggingface.co/docs/hub/en/model-cards
- Kaggle - https://www.kaggle.com/code/var0101/model-cards/tutorial
- Google Model Cards - https://modelcards.withgoogle.com/about
- AWS SageMaker - https://docs.aws.amazon.com/sagemaker/latest/dg/model-cards.html

2. Alignment with existing initiatives in the (X)BOM space

More broadly it is recommended that efforts to harmonise and defragment standards are taken, including on operability.

It is recognised that other types of Bills of Materials could benefit from the developments in TAIBOM – this includes cryptographic bills of materials (CBOMs, HBOMs etc). Some study efforts are ongoing at the time of writing, for example in the mobile telecoms industry body GSMA, to understand the best way forward for those industries as the subject is nascent in some sectors.

The two most commonly adopted existing standards for SBOM are:

SPDX

https://spdx.dev/
Developed by the Linux Foundation, SPDX is an open standard for communicating SBOM information (components, licenses, copyrights, and security references). In September 2021 SPDX was standardised by ISO, recognising it as an international standard for security, license compliance, and other software supply chain artefacts as ISO/IEC 5962:2021. It has been adopted by many major companies including Intel, Microsoft, Siemens and Sony.

SPDX provides a common format to share important data between companies and communities, and to exchange metadata to monitor the software supply chain, with a data format being both machine and human readable.

CycloneDX

https://cyclonedx.org/

Developed by Open Web Application Security Project (OWASP), CycloneDX is considered a lightweight SBOM standard. Its main potential is its flexibility in being able to support myriads of use cases. The list of supported component types and classes, extends beyond software and applications and get into devices and even services. Adopting a nesting and hierarchical approach, CycloneDX is able to keep up with the growing complexity of modern software ecosystems.

The information on the origin of the software and on the validity of its digital signatures are tracked according to NIST's Cybersecurity Supply Chain Risk Management (C-SCRM) guidelines.

CycloneDX also supports the Vulnerability Exploitability Exchange (VEX), providing context on known vulnerabilities in software components and their patches from software producers.

Other Bills of Materials

A Bill of Materials (BOM) historically was, and is, used for tracking the components of a product during manufacture. There are other types of Bills of Materials that have evolved over time, primarily for supply chain provenance.

Hardware Bill of Materials (HBOM)

A Hardware Bill of Materials (HBOM) is used for tracking the make-up of product hardware, with its primary usage being supply chain traceability and provenance. This can be used, for example, to ensure compliance with export control requirements, to prove sustainable and environmental obligations and to assure the origin for security reasons within supply chains.

In September 2023, the Cybersecurity and Infrastructure Security Agency (CISA) issued a Hardware Bill of Materials Framework (HBOM) for Supply Chain Risk Management (SCRM). It was released by a public-private, cross-sector task force organized and co-chaired by CISA, through a National Risk Management Center (NRMC) representative, and representatives from the Information Technology (IT) and Communications Critical Infrastructure Sectors.
Reference examples:

- CISA HBOM - https://www.cisa.gov/resources-tools/resources/hardware-bill-materials-hbom-framework-supply-chain-risk-management
- Fortress Infosec HBOM - https://www.fortressinfosec.com/blog/sbom-vs-hbom-whats-the-difference

Extending the Concepts of Software Bills of Materials

Several initiatives have extended the concepts introduced through SBOMs to cover other elements of information and metadata surrounding software deployments.

These can collectively be termed ‘XBOMs' – that is different prefix letters which denote extensions to Software Bills of Materials. Some of these extensions should be aligned with as they will also aid the security of the TAIBOM project itself.

Cryptographic Bill of Materials (CBOM)

A CBOM records the detailed representation of cryptographic assets within a system, including algorithms, keys, certificates, and their relationships to specific software and hardware components. CBOM enables the management of cryptographic assets and policies and for organisations to assess the state of implemented cryptography. This can help in assessing future resilience, including when assessing the state of quantum computing readiness and implementation of post-quantum cryptographic algorithms.

IBM have developed an open-source CBOM Framework that extends the CycloneDX SBOM standard to include cryptographic assets. This framework allows cryptographic assets and their dependencies to be modelled, allowing for better management and assessment of the cryptographic implementations.

Reference examples:

- CycloneDX CBOM - https://cyclonedx.org/capabilities/cbom/
- IBM CBOM - https://www.zurich.ibm.com/cbom/

3. Alignment with existing initiatives for AI Bills of Materials for Artificial Intelligence components

The emerging domain of Artificial Intelligence (AI) has rapidly demonstrated the need to protect the information and data that supports and runs AI models so there are many organisations coming to similar conclusions as the TAIBOM project. This is an emerging domain extending SBOMs into the information and data that supports the software (such as datasets). It has been also termed ML-BOM (Machine Learning Bill of Materials) and also has been extended to GBOM (Generative AI Bill of Materials).

An AI BOM is tailored for AI systems, collecting the details of their typical components (training data, models, training algorithm, performance test), managing the external libraries and frameworks necessary for the AI system, and protecting the overall integrity of an AI model, including the authenticity of the authorship.

There is currently no harmonised standard for these types of BOMs, however there are some other initiatives to TAIBOM which the project members should seek to align with, such as SPDX AI BOM: https://spdx.dev/implementing-an-ai-bom/

Further reading:

- NIST - https://csrc.nist.gov/presentations/2024/securing-ai-ecosystems-the-critical-role-of-aibom
- Snyk -https://snyk.io/articles/ai-security/ai-bill-of-materials-aibom/

4. Track, communicate to and align with Government departments and initiatives

Many governments have recognised the very real issues presented by insecure AI models. The US Cybersecurity and Infrastructure Agency (CISA) has published a number of pieces on AI cyber security, as has the US National Institute of Standards and Technology. In 2025, the UK government Department for Science, Innovation and Technology (DSIT) published a Code of Practice for the Cyber Security of AI. Many of these principles align with the objectives of TAIBOM and indeed TAIBOM can provide the implementation solution to many of the listed requirements.

Ensuring that these departments from across the world are aware of the existence of TAIBOM, what it solves and how it can be used in the future is crucial in avoiding fragmentation and raising industry awareness to support adoption. It is therefore important that work is undertaken, not just in the UK, but further afield to explain and communicate the TAIBOM work and provide materials online for policy makers and engineers to read.

Further reading:

- CISA AIBOM Tiger Team: https://github.com/aibom-squad/AIBOM-Tiger-Team
- DSIT Code of Practice for the Cyber Security of AI: https://www.gov.uk/government/publications/ai-cyber-security-code-of-practice
- NIST Adversarial AI paper: https://csrc.nist.gov/pubs/ai/100/2/e2025/final

5. Tracking the state of the art in current and future AI attacks, in theory and practice

In the attacks on the use cases developed by Copper Horse in the Innovate UK TAIBOM project, it was not difficult to conceive of attacks that could be very damaging or manipulative, that an ordinary user or system would not understand and would easily fall victim to. These in turn were straightforward to implement because in a world where datasets, AI inference applications and models themselves are susceptible to manipulation, it is often a case that most attacks are very possible to implement, the only limit is the human imagination in coming up with the attack and its path of compromise.

Copper Horse has looked at many academic papers and proof-of-concept attacks in this and other work and it is very clear that the space needs to be tracked, as well as in the security research / hacking domain. Whether it is implementation of Stegonet[2] or backdoor attacks in Large Language Models (LLMs)[3], or simple manipulation of imagery as demonstrated in the TAIBOM abuse cases, they all represent significant threats to the trust in AI and an inherent future danger to human safety.

As new attacks are theorised, it is now possible to test those out with the AI models developed by Copper Horse and also with TAIBOMs applied in order to check their viability in a TAIBOM assured or protected environment. It should also be possible to conduct further security research as industry partners and with colleagues in the security research community and academia by making available the models and sample code for them to use.

The time to make future attacks a reality may be shorter than people realise, so being ahead of the curve on AI security research is absolutely fundamental to creating a robust TAIBOM protected AI ecosystem.

6. Broader TAIBOM Testing

While the Innovate UK TAIBOM project has been able to test early versions of TAIBOM, both functionally and from a security perspective, there will need to be a more approach to fully testing the system as it matures. Some domains that will require further testing are:

- Bailo implementation
- Web interface
- User enrolment platform

7. Increasing the visibility of TAIBOM and embedding within products

Integrating and embedding the TAIBOM specification into existing products will help to organically spread adoption and embed security principles throughout the

[2] https://dl.acm.org/doi/10.1145/3427228.3427268
[3] https://arxiv.org/html/2312.15867v1

development lifecycle of a project. It means that developers can ensure that trustable AI practices are not an afterthought but a foundational component.

Embedding the framework helps address security and trust downstream where product is deployed and ensures alignment with upstream processes such as the design, procurement, and supply chain verification. This also enables detection of vulnerabilities and more transparent communication between the stakeholders about the components and models being used in AI systems.

There are many possible outlets for TAIBOM integration and these should be fully explored. GCHQ's Bailo[4] is a prime example of the type of product that would benefit from an integration with TAIBOM. Bailo is a service that provides a centralised repository of ML models and datasets. If each component on the repository is signed with a TAIBOM, it provides transparency and gives accountability to the data that is shared.

[4] https://gchq.github.io/Bailo/docs

Some Industry Security Considerations and Recommendations

Following extensive testing of the TAIBOM implementation by Nquiringminds, the results have been provided back to their team so that they can build-in countermeasures to the theoretical and practical attacks highlighted in the security reports.

There are some broader industry-level security recommendations that should be considered beyond the implementation that was created for the Innovate UK project. These are not exhaustive but worthy of further consideration.

8. PQC implementation and preparedness

Companies and industries around the world are busy implementing post-quantum encryption algorithms and longer symmetric key lengths in order to prepare for the possibility of a cryptographically relevant quantum computer. While the estimated timeline for this moment continually gets pushed back another 15 years, it is true to say that it makes sense that vulnerable hard maths problems such as those used in public key cryptography and therefore digital signature schemes as used in TAIBOM should be insured. The NIST-recommended post quantum computing (PQC) algorithms should be adopted within cryptographic suites used within TAIBOM and implemented. This goes hand-in-hand with crypto agility.

9. Cryptoagility, algorithm replacement and cryptosystem implementation robustness

The cryptographic subsystem of TAIBOM should be refactored to support modular and configurable algorithm selection with the cryptographic configuration and selected algorithm used as part of TAIBOM clearly stated in the output attestations. It should be ensured that any cryptographically insecure algorithm cannot be used for signing of files in future, while still retaining backwards compatibility for validation of deprecated configurations. Backwards compatibility can be retained for validation but there should be clear warnings to ensure there is no ambiguity about the quality of the attestation and any trust that can be derived from it.

Cryptographic agility also enables the users of TAIBOM to determine their own cryptographic strength requirements.

Any algorithms that are selected must support the use of salt or be sufficiently salted beforehand. The use of keyed hash algorithms such as HMACs can be used in conjunction to salt to further increase security.

10. Solutions for security scalability

Larger datasets can create significant bottlenecks when generating cryptographic hashes. During TAIBOM testing it was found that a 1GB dataset would take around 17 seconds to create a hash, a 100 GB dataset would therefore take almost 8 minutes and in theory, a Large Language Model (LLM) dataset of 774.5 TBs would take around 122 days. The energy and time cost of applying security through TAIBOMs represents a significant risk to adoption and requires further study and consideration.

Future optimisation considerations include (but are not limited to):

- Hashing solutions should take advantage of processor optimisation capabilities such as threading and/or multi-processing.
- Batch processing inputs – hashing many inputs together rather than individual can also be used to reduce overhead.

11. Implementation security considerations

Some of the flaws discovered in the early TAIBOM tool implementation led to some observations of basic flaws. These are worth considering when using TAIBOM in practice and include:

- Perform validation checks. For example, in TAIBOM the functionality only checked if the directory provided was a directory, and did not check if it was a file instead.
- Perform basic input sanitisation on a user-provided directory string,
- Do not permit unresolved relative paths, even if such a path is provided by the user – a verified credential may state a relative path, but hash generation must only use absolute paths or resolved relative paths.

12. Hash creation scaling for large datasets

Future implementers of digital signing within AI models and as part of TAIBOM, should consider the use of Merkle trees to enable granular validation and detailed error messages. Using them may also allow validation of parts of the dataset against the whole dataset. This recommendation is for further exploration.

13. Employ Security by Design

As a software tool or function, TAIBOM should be business as usual – companies should not have to worry that a security flaw in TAIBOM could be the route to compromise. Of course, it will be a significant target for attack and is likely to be breached as it becomes a front-line defence in preventing the compromise of AI models.

Therefore, there is a significant requirement on all stakeholders to ensure that the ecosystem around TAIBOM adopts strong security practices such as 'Secure by Design' and 'Security by Default'. This includes ensuring that there are 'no known vulnerabilities' as far as possible. The security posture of the solution should be a conservative one; that is that the attack surface is minimised in terms of code and any associated data, and that configurations are not insecure. A range of good software quality and security practices should be employed including DRY (Don't Repeat Yourself) and the principle of least privilege.

Further Security Recommendations

For further reading on security recommendations from TAIBOM, the 5.6A and 5.6B documents of the TAIBOM evaluation provide detailed breakdowns of attacks against the use case AI models and penetration testing of TAIBOM itself.