

OpenSSF Tech Talk

Securing the AI Lifecycle: Trust, Transparency & Tooling in Open Source

WED, SEPT 24, 2025 | 1:00PM ET



OpenSSF

OPEN SOURCE SECURITY FOUNDATION

Welcome!

- Thank you for joining us today! We will begin at 1:02 pm ET.
- While we wait for everyone to join, please take a moment to do one (or more) of the following:
 - Please add questions using the Zoom Q&A feature
 - Follow us on X: [@openssf](#), Mastodon: [social.lfx.dev/@openssf](#), LinkedIn: [OpenSSF](#), Bluesky: [@openssf.org](#)
 - Visit our website: <https://openssf.org>
 - Sign up for training: <https://openssf.org/training/courses/>
- **This Tech Talk is being recorded and slides will be available!**



OpenSSF

OPEN SOURCE SECURITY FOUNDATION

Agenda

- Housekeeping
- Speaker Introductions
- White Paper Deep Dive
- Model Signing & Trust in AI Artifacts
- Opportunities and Emerging Efforts
- Panel Discussion & Audience Q&A
- Important announcements

Help us improve! Tech Talk Survey



OpenSSF

OPEN SOURCE SECURITY FOUNDATION

Antitrust Policy Notice

Linux Foundation meetings **involve participation by industry competitors**, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.

Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at <http://www.linuxfoundation.org/antitrust-policy>. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrave of the firm of Gesmer Updegrave LLP, which provides legal counsel to the Linux Foundation.

Code of Conduct

- The Linux Foundation and its project communities are **dedicated to providing a harassment-free experience** for participants at all of our events, whether they are held in person or virtually.
- All event participants, whether they are attending an in-person event or a virtual event, **are expected to behave in accordance with professional standards**, with both this Code of Conduct as well as their respective employer's policies governing appropriate workplace behavior and applicable laws.
- <https://openssf.org/community/code-of-conduct/>

Q&A

Please submit your questions during the meeting by using the Q&A feature on Zoom.



Thank you!

Introductions

Christopher “CRob” Robinson





Christopher “CRob” Robinson - Security Lorax, Chief Architect of OpenSSF

Christopher Robinson (aka CRob) is the Chief Security Architect for the Open Source Software Foundation (OpenSSF). With over 25 years of experience in engineering and leadership, he has worked with Fortune 500 companies in industries like finance, healthcare, and manufacturing, and spent six years as Program Architect for Red Hat’s Product Security team.

CRob has spoken at major events such as RSA, BlackHat, and DefCon, and was recognized as a top presenter at Red Hat Summits in 2017 and 2018. He holds certifications like CISSP and CSSLP. He leads several OpenSSF working groups, chairs its Technical Advisory Committee, and contributes to the FIRST PSIRT SIG.

CRob enjoys hats, herding cats, and moonlit beach walks.



Sarah Evans - Distinguished Engineer, Dell

Sarah Evans delivers technical innovation for secure business outcomes through her role as the security research program lead and Distinguished Engineer in the Office of the CTO at Dell Technologies. She is an industry leader and advocate for extending secure operations and supply chain development principles to AI. Sarah also ensures the security research program explores the overlapping security impacts of emerging technologies in other research programs, such as quantum computing. Sarah leverages her extensive practical experience in security and IT, spanning small businesses, large enterprises (including the highly regulated financial services industry and a 21-year military career), and academia (computer information systems). She earned an MBA, an AIML professional certificate from MIT, and is a certified information security manager (CISM). Sarah is also a strategic and technical leader representing Dell in OpenSSF, a foundation for securing open-source software.



Mihai Maruseac - Staff Software Engineer, Google

Mihai Maruseac is a member of Google Open Source Security team (GOSST), working at the intersection of AI and security. He co-leads the OpenSSF AI/ML working group and co-maintains two supply-chain security projects: model-signing and GUAC.

Before joining GOSST, Mihai worked on OSS TensorFlow, creating the TensorFlow Security Team. Prior to Google, Mihai worked on adding Differential Privacy to ML algorithms, following a PhD on Differential Privacy.

Mihai blogs at mihai.page, likes Haskell, hikes, books, and some games (Zelda, Hollow Knight).

Marcela Melara - Research Scientist & OpenSSF TAC Member



Dr. Marcela Melara is a research scientist at Intel leading internal, academic and open-source projects that advance software and AI supply chain security. Besides serving as member of the OpenSSF TAC, she's co-author of the OpenSSF SLSA attested build environments track, and maintainer of the in-toto Attestation Framework.

Marcela's work appears in various publications, OSS conferences and patents. Prior to joining Intel, she received her PhD in Computer Science from Princeton University. Marcela is a Siebel Scholar and OpenSSF Golden Egg Award recipient. She co-chairs the OpenSSF BEAR working group and has served as mentor for Científico Latino. Outside of work, Marcela is an avid hiker, creative writer and crocheter.

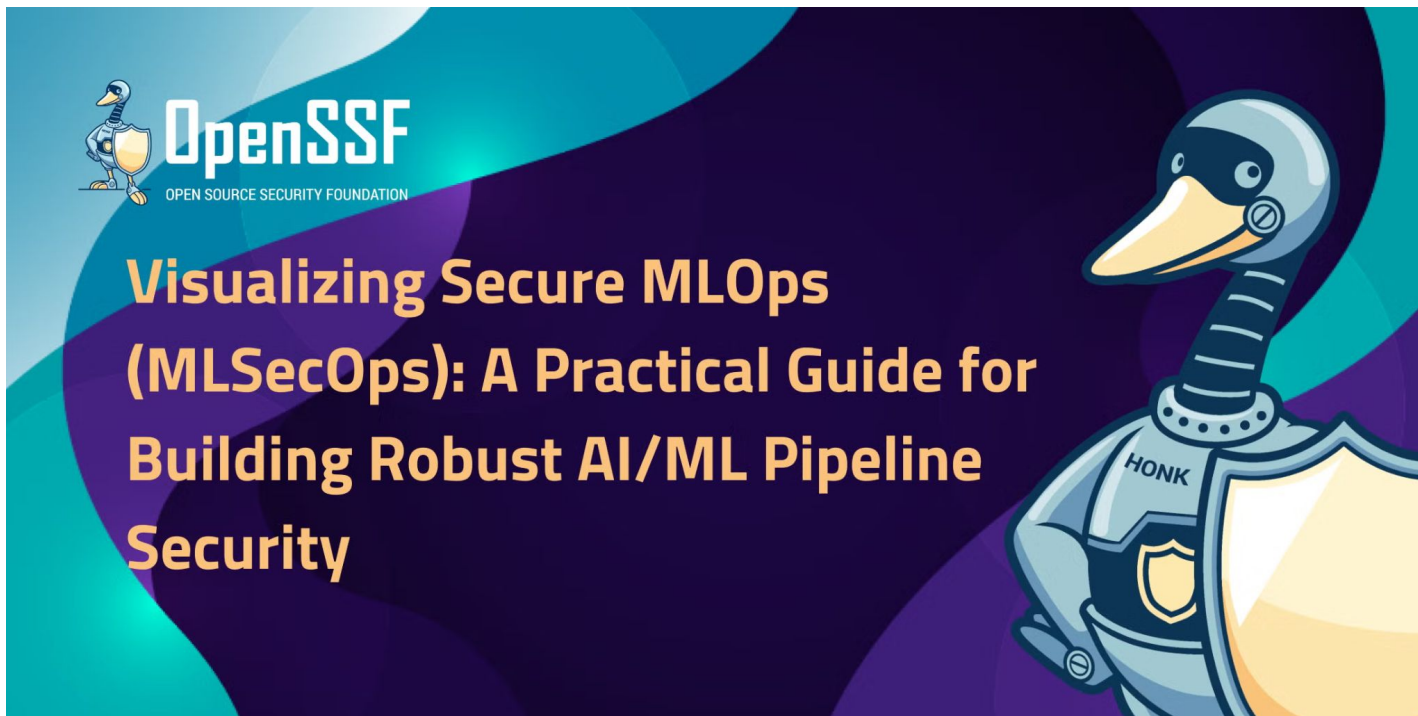
Whitepaper Deep Dive

Sarah Evans

Distinguished Engineer, Dell Technologies



White Paper



<https://openssf.org/resources/visualizing-secure-mlops-mlsecops-a-practical-guide-for-building-robust-ai-ml-pipeline-security/>

MLSecOps white paper

- Who should read this
- What's Inside
- When should I use this
- Where does this apply
- Why now

References:

- About the white paper:
<https://openssf.org/resources/visualizing-secure-mlops-mlsecops-a-practical-guide-for-building-robust-ai-ml-pipeline-security/>
- Blog:
<https://openssf.org/blog/2025/08/05/visualizing-secure-mlops-mlsecops-a-practical-guide-for-building-robust-ai-ml-pipeline-security/>

The first layer and the origin story

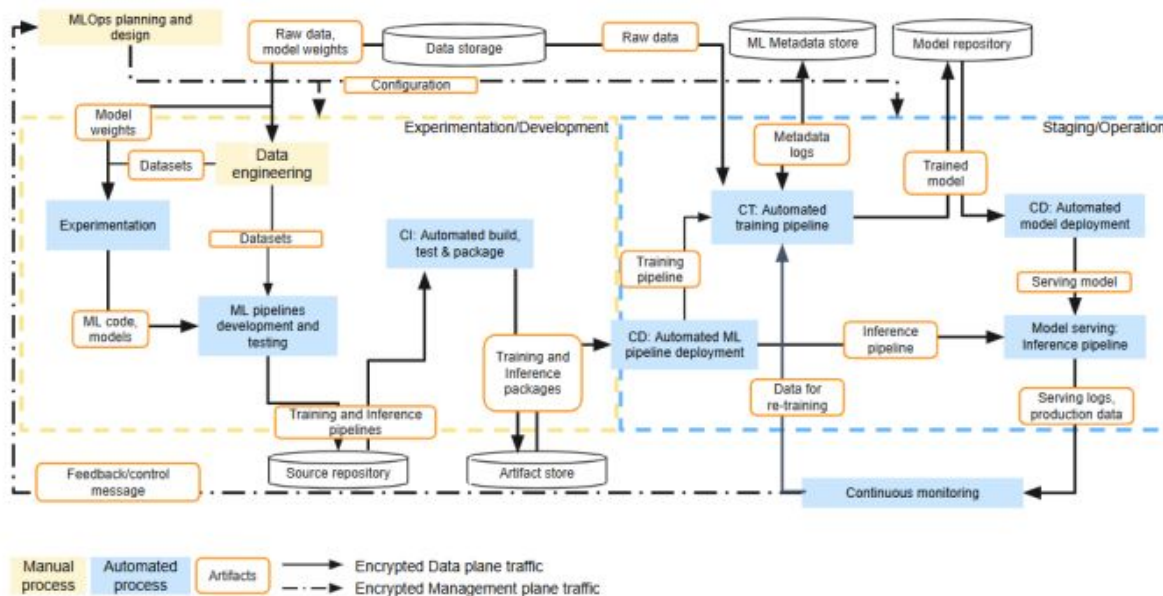


Figure 4: A generalized MLOps reference architecture

The final layer

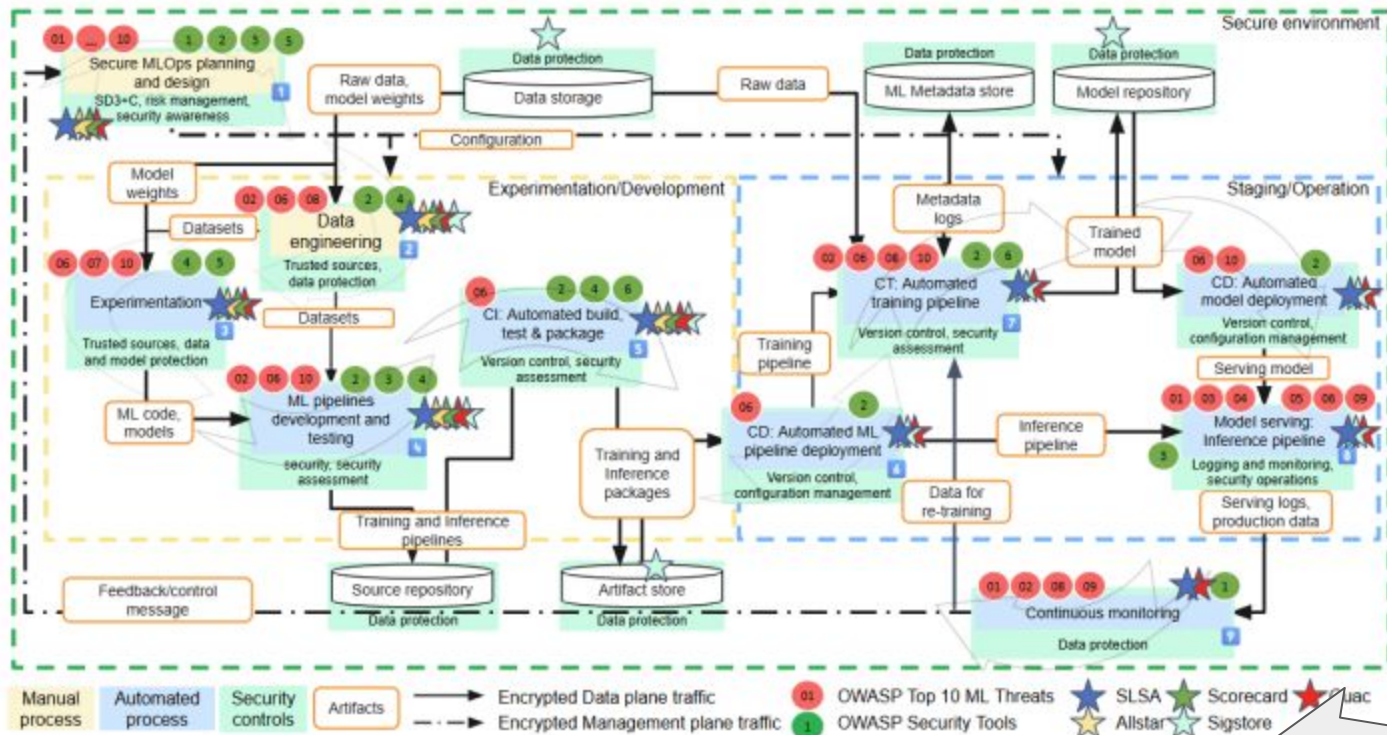


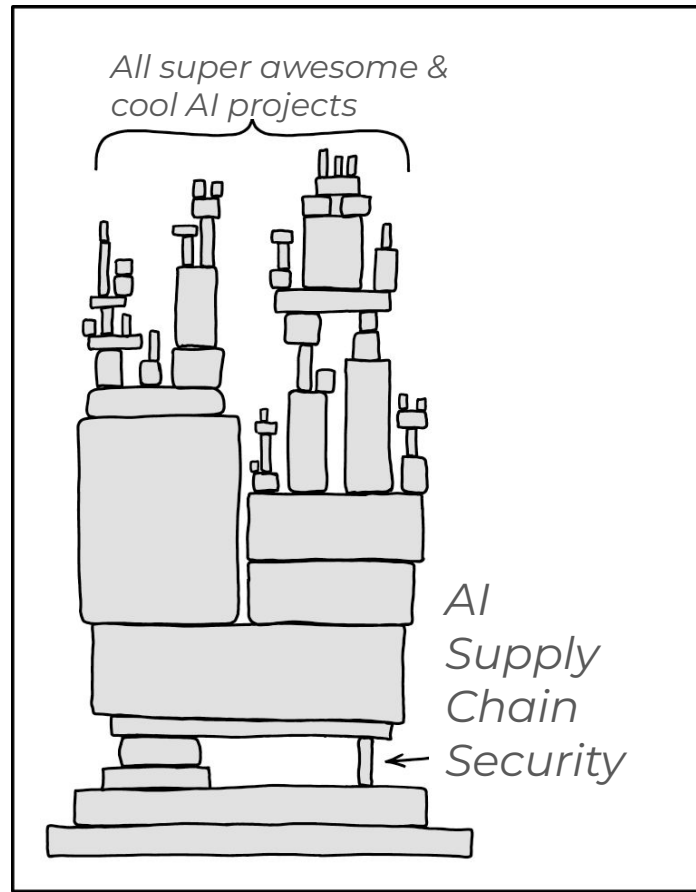
Figure 9: Mapping of security measures and tools to MLSecOps stages

Model Signing & Trust in AI Artifacts

Mihai Maruseac

Staff Software Engineer, Google





XKCD 2347 (Kinda)

Initial training

Fine-tuning

Deploy



ML Developer 1



ML Developer 2



SWE

Train
Foundation
Model



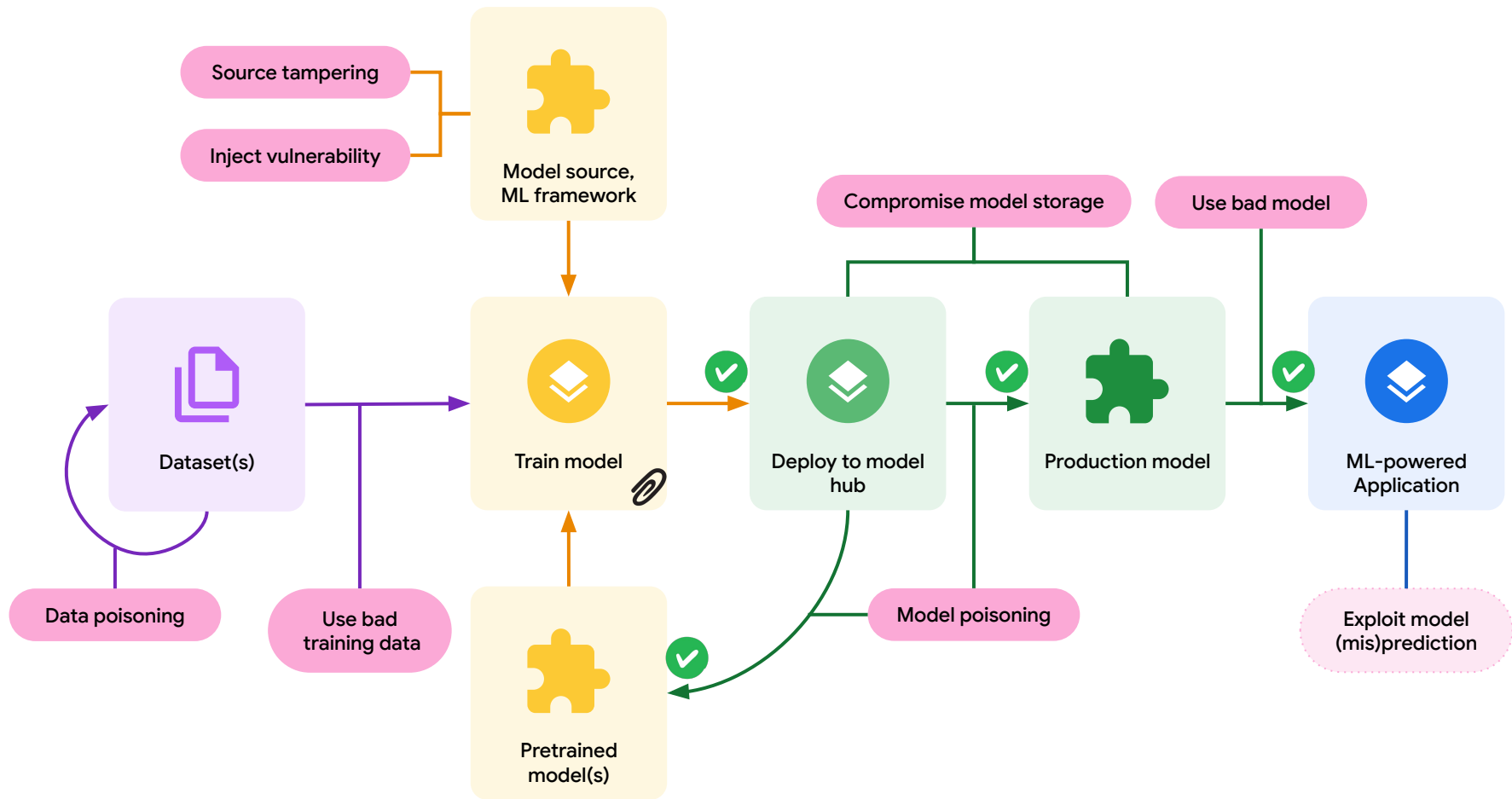
Public/Private
Model Hub

Train
Fine-tuned
Model



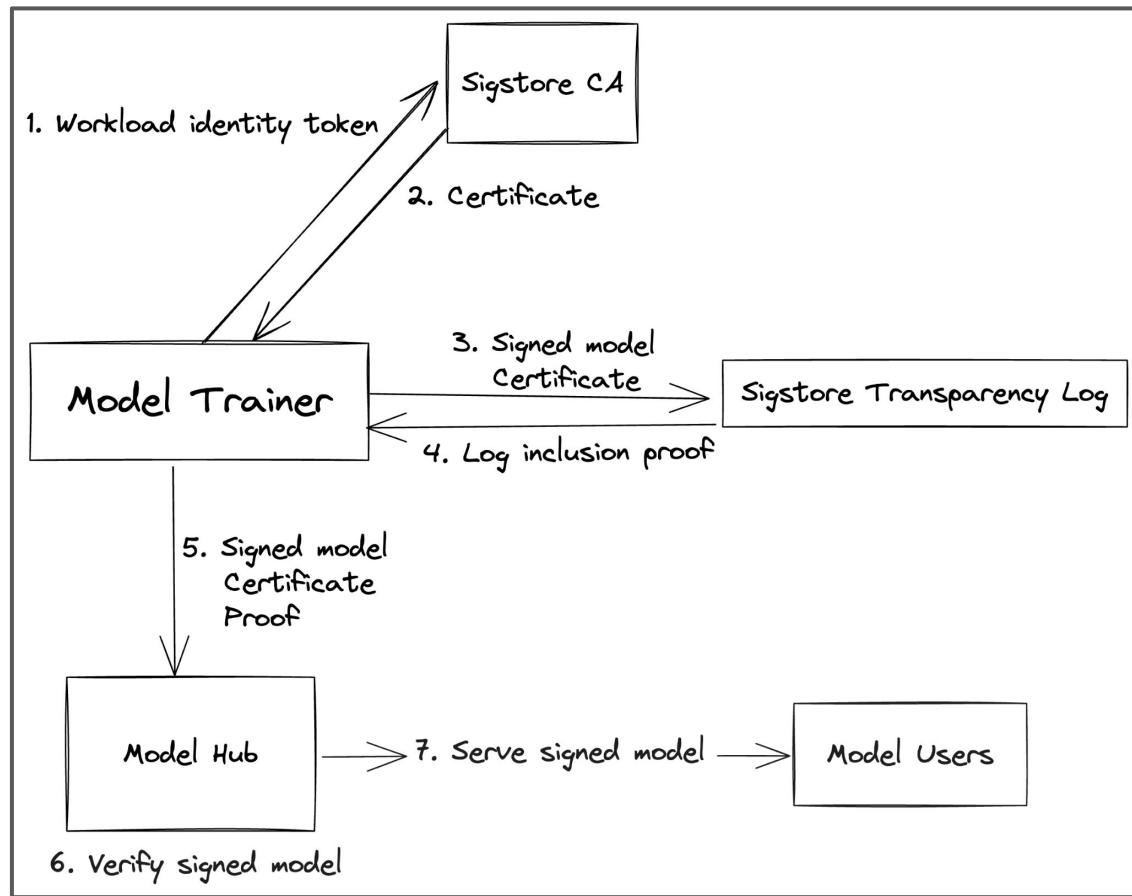
Internal Model
Hub

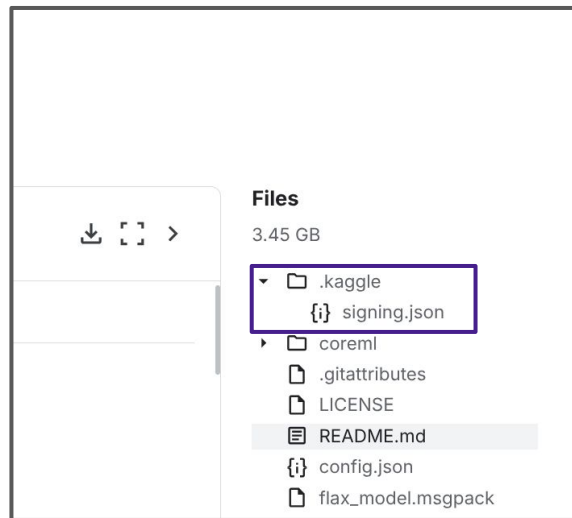
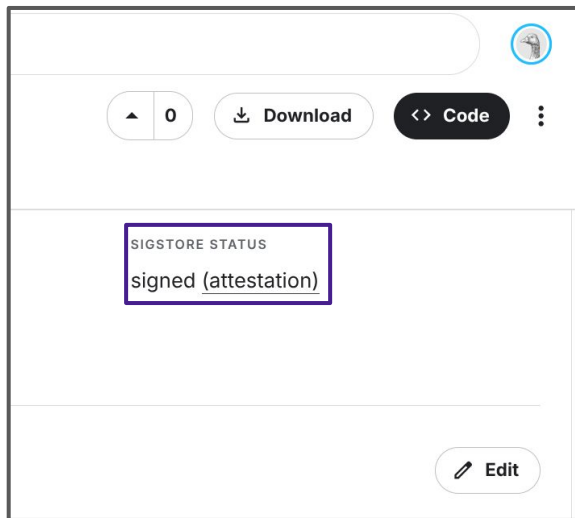
Deploy Model
into Application



Model signing

- Sign a collection of arbitrary files, verify a subset
 - Meet ML developers where they are: model hub with multiple formats, users need only one
 - Support distributed training and inference
- Detached signature
 - Model format agnostic, easier to integrate with model hubs and trusted publishing
- Private Key Infrastructure agnostic
 - Sign with Sigstore or keys/certificates
- Integrate with model hub
 - Trainer: Train model, publish to hub from pipeline, get supply chain assurances by default
 - Model user: Download model from hub with verification incorporated
 - Also: Check model signature from model hub and/or verify locally





Entry UUID: [108e9186e8c5677a009d9d38452e9bb768de9e11dc83631e09de16f913211d19c4c120f58cd3677f](#)

TYPE	LOG INDEX	INTEGRATED TIME
dsse	275041887	2 months ago (2025-07-15T09:47:16-07:00)

Hash


```
sha256:d1691eda734544bbaf64130dbdf3e8e6a594677ee99f866d3d70dc0ac2e379b3
```


Signature


```
MEYCIQCpINFAWwqu0ic+hL35ycJZKEJUOk14IXm/jfzHN7Da0gIhALqRqPJL2U7Tu50ZXAFAP87k2R4a0xwWqj9L0h8INQ86
```

Public Key Certificate



```
data:
  Serial Number: '0x5f7984a25dab11a75fdd7dafb018d69e3ef43493'
Signature:
  Issuer: 0=sigstore.dev, CN=sigstore-intermediate
  Validity:
    Not Before: 2 months ago (2025-07-15T09:47:16-07:00)
    Not After: 2 months ago (2025-07-15T09:57:16-07:00)
  Algorithm:
    name: ECDSA
    namedCurve: P-256
  Subject:
    extraNames:
      items: {}
    asn: []
X509v3 extensions:
  Key Usage (critical):
    - Digital Signature
  Extended Key Usage:
    - Code Signing
  Subject Key Identifier:
    - E7:B9:2B:F4:0B:30:76:9C:DC:79:3A:6F:A6:13:63:7E:B5:7F:72:C4
  Authority Key Identifier:
    kevid: DF:D3:E9:CF:56:24:11:96:F9:A8:D8:E9:28:55:A2:C6:2E:18:64:3F
  Subject Alternative Name (critical):
    email:
      - 21935169@privatekaggle.com
  OIDC Issuer: https://www.kaggle.com/api/v1/models/signing
  OIDC Issuer (v2): https://www.kaggle.com/api/v1/models/signing
1.3.6.1.4.1.11129.2.4.2: 04:79:00:77:00:75:00:dd:3d:30:6a:c6:c7:11:32:63:19:1e:1c:99:67:37:02:a2:4a:5e:b8:de:3c:ad:ff:87:8a:72:80:2f:29:ee:8e:00
```



 **NVIDIA** NGC Catalog Explore Search


 Model


 **NVOF**

NVOF is a deep learning based optical flow estimation



 1.0  Signed

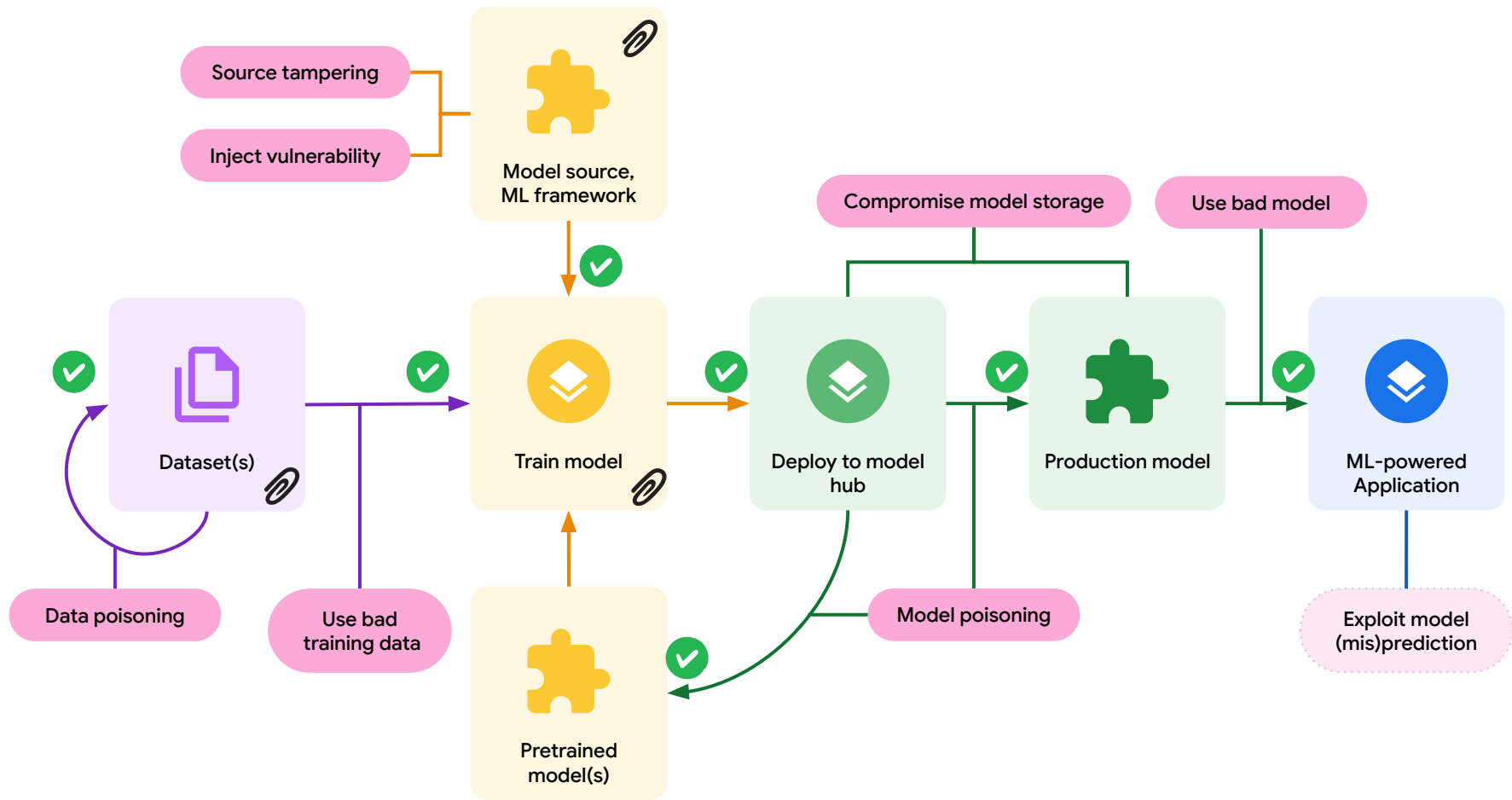
 **NVIDIA** NGC Catalog Explore Search Support

 Model

 **Llama 3.2 3B Instruct**

The Meta Llama 3.2 collection of multilingual large language models (in/text out).

 2.0  Signed



Opportunities and Emerging Efforts

Marcela Melara

Research Scientist, Intel Corporation



What does it really mean to “trust a model”?

Actually asking:

- Where did this model originate?
 - What components make up this model?
 - Was my model produced from the expected components and pipeline?
 - Can I demonstrate that I ran my customer’s expected pipeline?
 - What platform was the pipeline run on?
 - How was the pipeline configured? What parameters, weights, etc. were used?
 - What dataset was used to train the model?
 - Was a legitimate version of the foundation model, PyTorch, etc. used?
 - Does the pipeline contain any known vulnerabilities?
 - etc.
- } Model Signing

Emerging Effort: ML Model Lifecycle Provenance

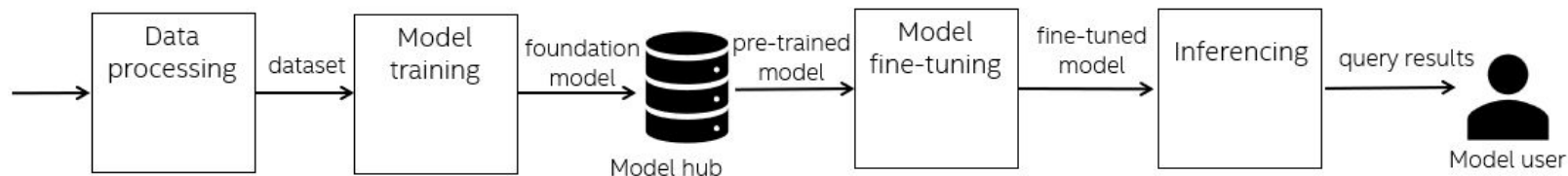
Goals:

- Collect ML artifact & pipeline metadata
- Establish a cryptographically verifiable graph of model lineage
- Leverage trusted execution hardware as root of trust for integrity

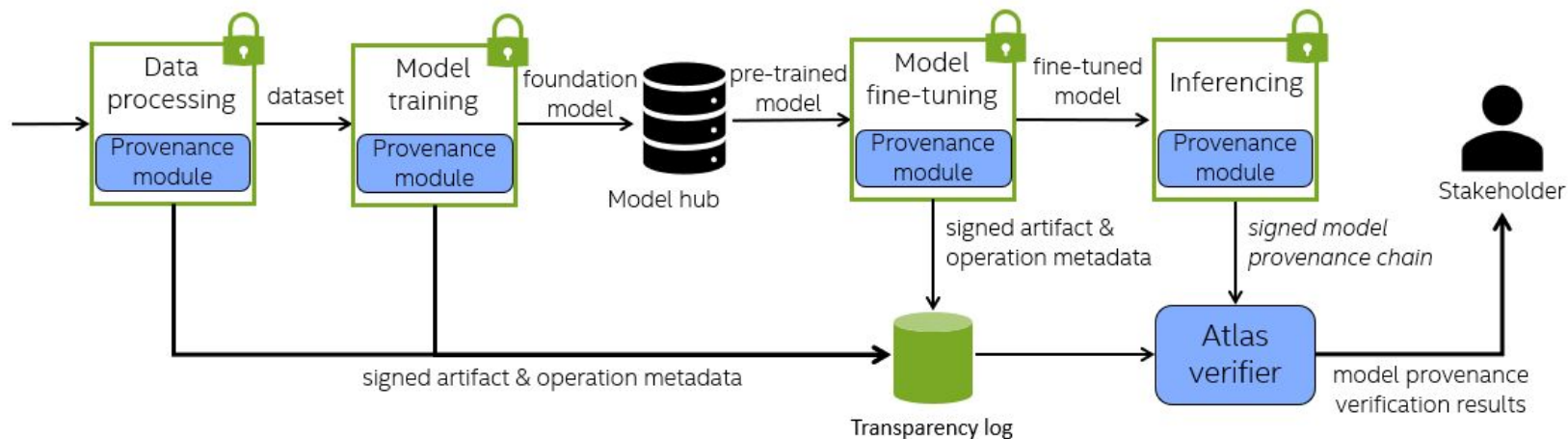


Learn more in our OSS NA '25 talk!

Atlas: Mapping Model Lifecycle Provenance



Atlas: Mapping Model Lifecycle Provenance



Atlas: Framework for ML Model Lifecycle Provenance

- Focus on Integrity: Attestable foundation for model privacy, confidentiality, quality, safety, etc.
- Supports Model Signing!

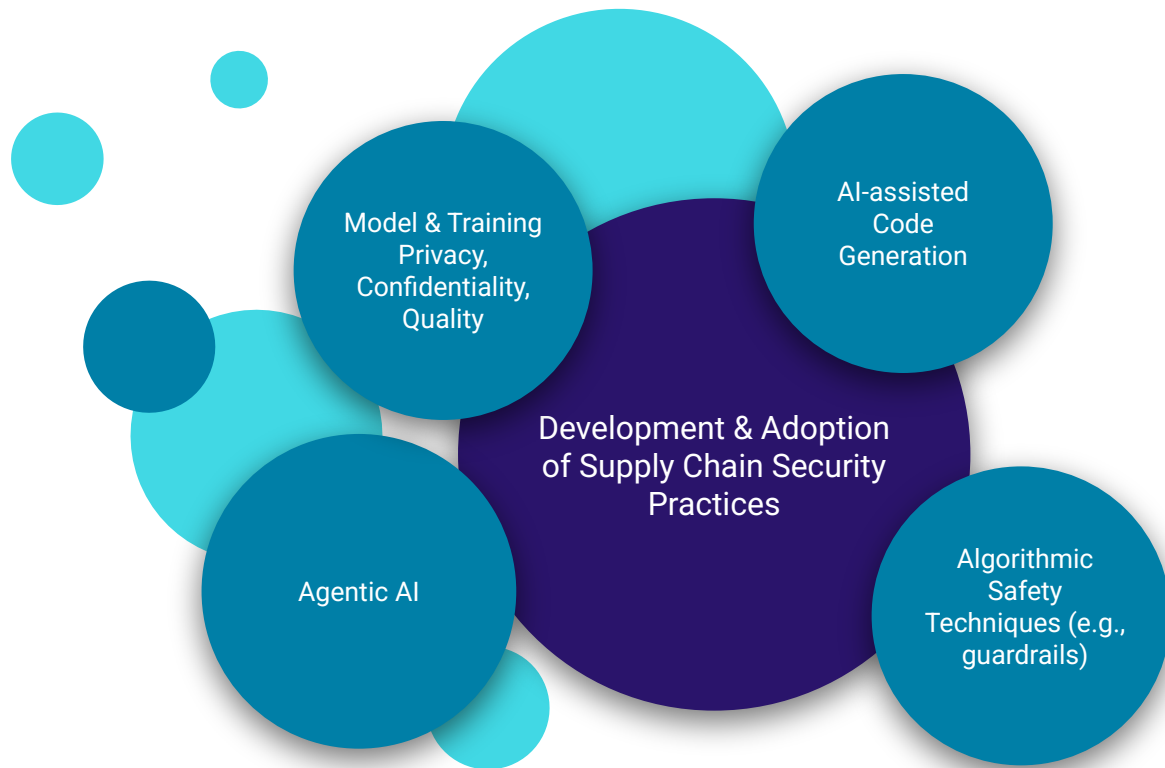


github.com/IntelLabs/atlas-cli

Emerging Effort: GPU-based Integrity for LLMs

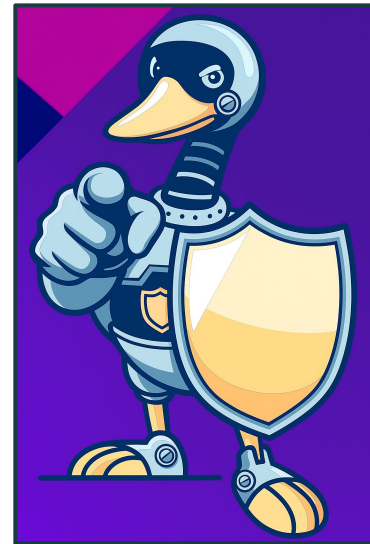
- Goal: Offload integrity verification from CPU to GPU for improved performance
 - Efficient cryptographic hashing and signing of LLMs
- Combine new and existing trusted hardware features and standards:
 - Confidential virtual machines for CPU-based integrity and attestation
 - TEE Device Interface Security Protocol (TDISP) for secure CPU-to-GPU communications
- Ongoing development of proof-of-concept
 - Establish a unified framework for different hardware architectures
 - Integration options with Model Signing!

Opportunities for AI Lifecycle Security Abound



Get involved!

- Join the AI/ML Security Working Group
 - Next meeting: Sept 29, 10 AM Pacific
- ... or any of its Special Interest Groups
 - Cyber Reasoning Systems: from AIxCC to security at scale
 - Safe MCP: securing the agentic world
- Cross-pollinate between different communities
 - CoSAI (next supply chain workstream meeting: Oct 1, 10 AM Pacific)
 - LF AI & Data (next meeting: Sept 30, 7 AM Pacific)
- Try out / contribute to existing and emerging AI lifecycle security projects



Panel Discussion & Audience Q&A



Get the Launch Notification for LFEL1012!

Our new course is almost here! Sign up to receive a notification the moment it goes live, along with early access to resources and updates.



https://docs.google.com/forms/d/e/1FAIpOLsFwW8M6PwOM62VHgc-YyogzT-eK_scJVk21BtezFUnJmMx6DQ/viewform

Get Involved & Helpful Resources

Join the [AI/ML Security Working Group](#)!

Resources:

- [Visualizing Secure MLOps \(MLSecOps\): A Practical Guide for Building Robust AI/ML Pipeline Security](#)
- [OpenSSF Model Signing \(OMS\)](#)
- [OpenSSF Model Signing Specification](#)
- [Model Signing tools](#)



Upcoming Events

[The Linux Foundation Europe Roadshow](#)

October 29, 2025 | Ghent, Belgium

[Register European Open Source Security Forum](#)

- October 30, 2025 | Brussels, Belgium
- [Express your interest to join](#)

[OpenSSF Community Day Korea](#)

November 4, 2025 | Seoul, South Korea

- [Schedule Live](#)
- [Register](#), Select "Attendee"

[Open Source SecurityCon](#)

November 10, 2025 | Atlanta, Georgia (Co-located with KubeCon)

- [Schedule Live](#)
- [Register and select your colo event](#)

Ways to Participate



Join a [Working Group/Project](#)



Come to a Meeting (see [Public Calendar](#))



Collaborate on [Slack](#)



Contribute on [GitHub](#)



Become an [Organizational Member](#)



Keep up to date by subscribing to the [OpenSSF Mailing List](#)

Engage with us on social media



X
[@openssf](https://twitter.com/openssf)



LinkedIn
[OpenSSF](https://www.linkedin.com/company/openssf)



Mastodon
social.lfx.dev/@openssf



YouTube
[OpenSSF](https://www.youtube.com/OpenSSF)



Facebook
[OpenSSF](https://www.facebook.com/OpenSSF)



Bluesky
[OpenSSF.org](https://bluesky.com/OpenSSF.org)

Is your organization a member?

Questions? Contact membership@openssf.org

openssf.org/join



Take our quick Tech Talk Survey

Help us improve!



Thank You



OpenSSF
OPEN SOURCE SECURITY FOUNDATION

Legal Notice

Copyright © [Open Source Security Foundation](#)®, [The Linux Foundation](#)®, & their contributors. The Linux Foundation has registered trademarks and uses trademarks. All other trademarks are those of their respective owners.

Per the [OpenSSF Charter](#), this presentation is released under the Creative Commons Attribution 4.0 International License (CC-BY-4.0), available at <<https://creativecommons.org/licenses/by/4.0/>>. You are free to:

- Share — copy and redistribute the material in any medium or format for any purpose, even commercially.
- Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms:

- Attribution — You must give appropriate credit , provide a link to the license, and indicate if changes were made . You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.