CoSAI TSC Meeting

Inaugural Meeting
September 27, 2024

Agenda

- Welcome and Introductions (Akila & JR)
- TSC Governance (OASIS staff)
- Workstreams (Akila & JR)
- Meeting Cadence (Akila & JR)
- Next Steps
- AOB

Introductions

- Co-Chairs: Akila Srinivasan (Anthropic) and J.R. Rao (IBM)
- Introductions of all members on the call
  - Company
  - Current Role
  - Why are you interested in participating in the TSC
  - Which Workstreams do you want to contribute to?

TSC and Workstream Governance

- TSC and Workstream Governance document can be found on GitHub
- TSC is responsible for technical health and direction of the project and advises the PGB
- TSC is also responsible for releases and overseeing work of Workstreams, WS Chairs, Contributors, and Maintainers. The TSC reports to the PGB and is subject to the CoSAI Governance.
- TSC members: founding and Premier sponsors and non-sponsor seats
- Co-Chairs are elected or re-confirmed every 2 years

TSC and Workstream Governance

- TSC Member Responsibilities:
  - TSC members are expected to attend committee meetings on a regular basis and contribute to the objectives and outcomes of those meetings.
  - Every TSC member is required to be involved in at least one workstream.
  - The TSC in coordination with the PGB, is responsible for setting a roadmap for each Workstream. PGB provides a workstream scope and vision. The TSC then oversees the implementation of the roadmap by the individual workstream leads, ensuring its successful execution.
- TSC Attendance policy
  - A TSC member will lose their voting right at the end of the second consecutive TSC meeting missed where neither the member nor their designated alternate is present, without prior notice to the TSC Chair
  - A TSC member who has lost their voting rights may regain voting rights by (a) declaring to the Chair their desire to regain voting rights and then (b) attending (or having their alternate attend) two consecutive meetings of the TSC.

Contributions and Maintainer policy

Contributing to CoSAI

The TSC is responsible for creating a contribution policy for the workstreams which can be approved by lazy consensus.

Maintainer Policy

The TSC will establish a maintainer policy which covers the number of maintainers per repository. The Workstream Leads should also serve as the maintainer for the relevant repositories.

Action:

The TSC shall determine the number of maintainers required to merge a contribution into the master branch of the WS repo. This shall be done during the first TSC meeting. Changes to this number require a simple majority of TSC members.

TSC Tools and Communication

- Mailing list is the official channel
  - TSC Mailing List: cosai-tsc@lists.oasis-open-projects.org
    - Public Archive: https://lists.oasis-open-projects.org/g/cosai-tsc
- GDrive (eCLA required for access): TSC Folder
- GitHub: https://github.com/cosai-oasis
- Slack established by OASIS can be found slack cosai

CoSAI: Three Project Workstreams

Addressing the AI Developer, Security Analyst & Compliance Officer Persona

Software Supply Chain Security for AI Systems

Enhance AI security through provenance and risk evaluation. By building on SSDF and SLSA security principles, CoSAI aims to expand and refine these efforts.

Preparing Defenders for a Changing Security Landscape

Develop frameworks to guide defenders in prioritizing investments to counter offensive cybersecurity capabilities of current and potential AI models.

AI Risk Governance

Create AI security assessment tools: risk taxonomy, controls checklist, and readiness scorecard for practitioners.

W1: Software Supply Chain Security for AI systems

Description

- Expand provenance controls in the AI domain and lower the barriers to AI provenance adoption and risk management.
- Extend SSDF and SLSA principles to the security of AI development.

Targeted Persona	Classes of Use Cases	Possible Outcomes
• AI application developers • Data scientists • Model creators	• Model selection and usage • Technical controls for model governance • Model usage/prompt security • Secure deployment and containment practices • Standardization of model weight serialization/deserialization	• AI-specific SSDF and SLSA frameworks • Security controls for deployment, containment, use of AI • Guide standards for secure model weight serialization/deserialization. • Industry standards for secure model weight handling • **Not in scope:** Aspects of Provenance such as copyright information, licensing information and terms and conditions

W2: Preparing Defenders for a Changing Cybersecurity Landscape

Description
• Address the security impacts of use of AI by business applications including attackers, defenders, mitigation techniques and best practices

Targeted Persona	Classes of Use Cases	Possible Outcomes

• CISO
• Security practitioners
• AI app developers
	• Threat and vulnerability management
	• AI Attack surface management
	• Human-AI collaboration interfaces
		• Helpful prompt libraries and templates to promote effective use of AI in security
		• Helpful prompt libraries and templates to prevent abuse of AI
		• Best practices for:
		  ○ Preventing data leaks in fine-tuned models
		  ○ Sharing model threat/vulnerability information
		• **Not in scope:** Security of Agentic Frameworks

W3: AI Risk Governance

Description
• Develop a risk and controls taxonomy, checklist, and scorecard to guide practitioners in readiness assessments, management, monitoring, and reporting of their AI products, services, and components

Targeted Persona	Classes of Use Cases	Possible Outcomes
• Chief Risk Officers • Compliance Officers • CISOs, CIOs, and CEOs • Policy and regulation framers	• Governance when **introducing** AI to your company • Governance when **using** AI • Demonstrating compliance to users and external authorities	• Transparent, consistent model cards to help enterprises assess AI adoption risks • Define technical controls to help enterprises achieve regulatory compliance • Discover and govern shadow AI usage and secure AI artifacts throughout the lifecycle. • Manage security risks of AI artifacts, including training data and tuning processes • **Not in scope:** Security of Agentic Frameworks

Launch Timelines

Date	Action
Friday, Sep 27, 2024	Launch Day of the CoSAI Technical Steering Committee (TSC)
Tuesday, Oct 1, 2024	Each TSC member to volunteer for a workstream: CoSAI Workstream sign up sheet Each organization has volunteered non-TSC members to participate in a workstream
Wednesday, Oct 2, 2024	(If necessary) TSC co-chair(s) will send a form to TSC members to vote for the WS leads (2 leads per workstream)
Friday, Oct 4, 2024	Second CoSAI TSC Meeting: Workstream leads will be announced
Week of Oct 7, 2024	TSC co-chair(s) and OASIS rep(s) to meet with workstream leads to plan workstream launch
By Oct 16, 2024	Workstream leads have set the first meeting with their workstream to kick off the effort
Friday, Oct 18, 2024	Schedule periodic comebacks by WS leads to report on their scope and deliverables

TSC Members: Action Items

- Review workstream scope and content (this slide deck)
- Review GitHub policies (CoSAI github)
- By Oct 1, 2024:
  - Volunteer for ≥1 workstream Sept 2024: CoSAI Workstream sign up sheet
  - Nominate non-TSC members for workstreams
- By Oct 3, 2024:
  - Vote for workstream leads (2 per stream)

Any Other Business?