Secure Software Development Framework for Generative AI and for Dual Use Foundation Models

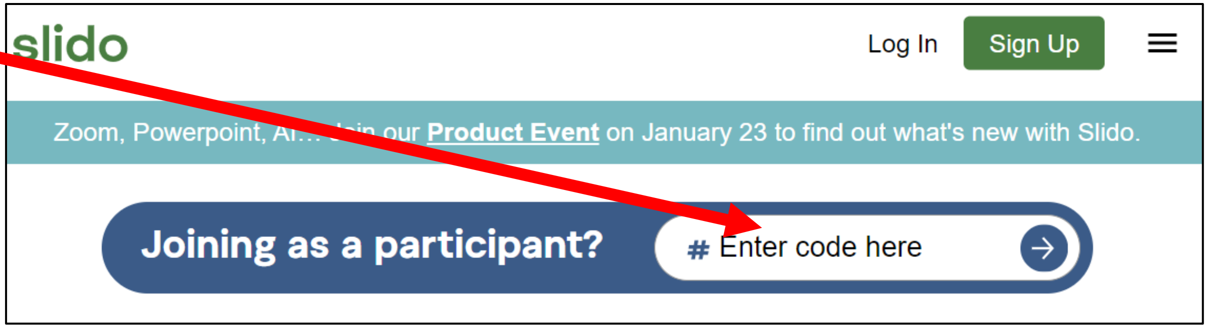
Virtual Workshop

NIST

NATIONAL
CYBERSECURITY
CENTER OF EXCELLENCE

HOUSEKEEPING

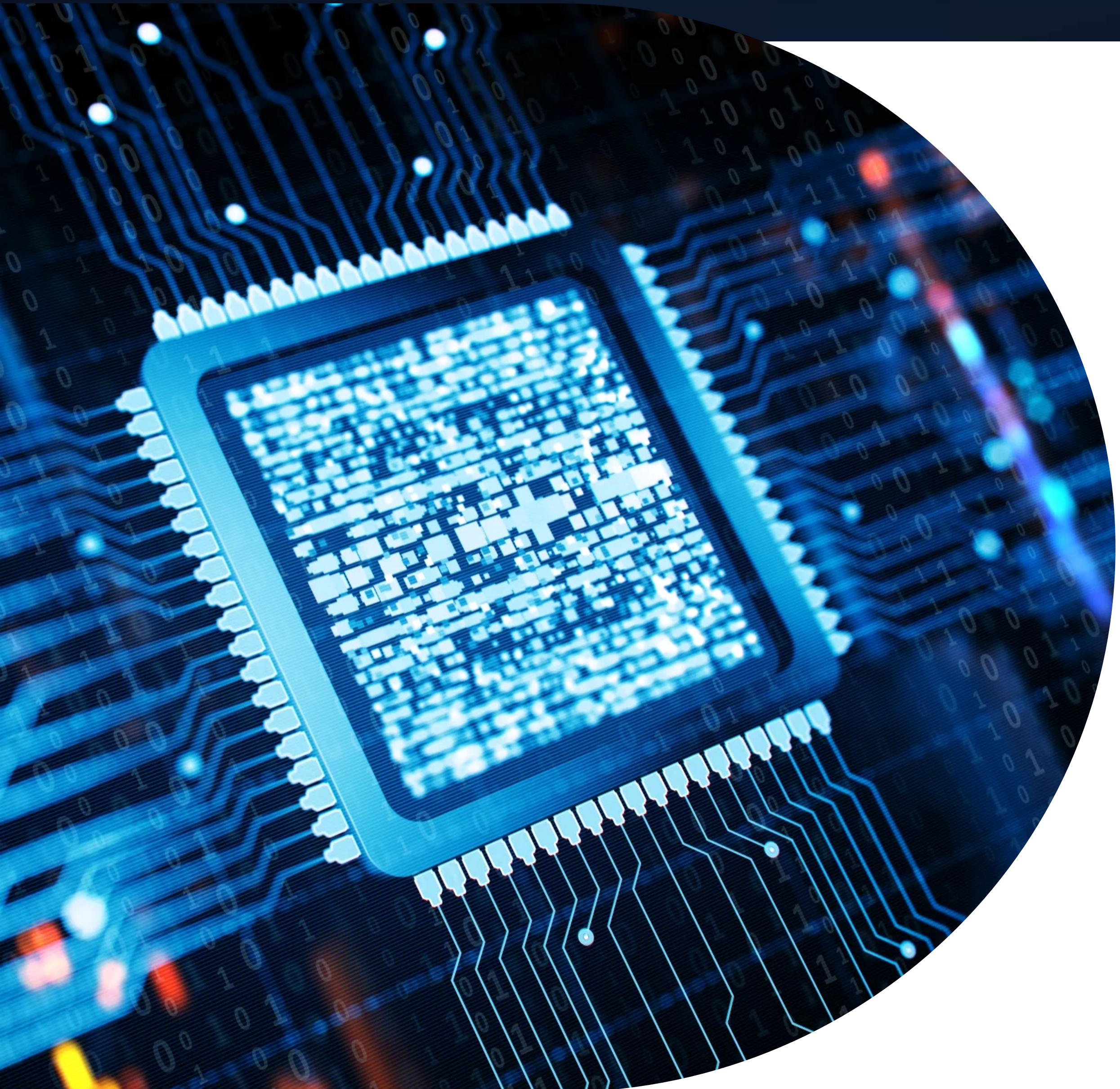
- Submit questions using Slido: https://app.sli.do/event/pdUcG7qTxsHEZa2nNjvAaD
  - Event Code: #SSDF



Zoom, Powerpoint, AI... Join our **Product Event** on January 23 to find out what's new with Slido.

- Today's session is being recorded
- Any feedback can be submitted to: ssdf@nist.gov

WORKSHOP AGENDA



9:00 AM – 9:15 AM	**Introduction and Overview**
9:15 AM – 10:15 AM	**Session 1 - Secure Software Development Challenges with Large Language Models (LLMs) and Generative AI Systems**
10:15 AM – 11:15 AM	**Session 2 - Secure Development of LLMs and Generative AI Systems**
11:15 AM – 11:30 AM	**Break**
11:30 AM – 12:30 PM	**Session 3 - Secure Use of LLMs and Generative AI Systems**
12:30 PM – 12:45 PM	**Next Steps & Wrap Up**

Introduction and Overview

NIST

NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

NIST Trustworthy & Responsible AI Program

- Cultivate trust in the design, development, use, and governance of artificial intelligence (AI) technologies and systems
  - Conduct fundamental and applied research
  - Establish benchmarks, data and metrics for evaluation
  - Lead and participate in development of technical AI standards
  - Contribute technical expertise to inform AI policies

https://www.nist.gov/trustworthy-and-responsible-ai

Secure Software Development Framework (SSDF)

- *Version 1.1 issued in February 2022*
  - Provide a set of fundamental, sound, and secure software development practices
  - Provide a common language for describing secure software development practices for software producers and acquirers
  - Help reduce the number of vulnerabilities in released software and the potential impact of the exploitation of undetected or unaddressed vulnerabilities, and address the root causes of vulnerabilities to prevent recurrences
  - Focuses on outcomes, not tools and techniques, and applicable to various types of software development, including AI models

https://csrc.nist.gov/pubs/sp/800/218/final

Objectives of Today's Workshop

- Inform development of "*a companion resource to the SSDF to incorporate secure development practices for generative AI and for dual-use foundation models* " directed in EO 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.

- Bring together AI developers and researchers and secure software practitioners from industry and government to discuss the secure software practices and considerations in developing generative AI models and dual-use foundation models