Responsible AI: CyberSecurity and Global Cooperation

张雅琪
Alphatu

OpenSSF Beijing Meetup
2023 June 15th

About Me

Grace Zhang(Alphatu)

- Researcher
- Newbie of CyberSecurity
- Lover of Open Source

What Is The Situation?

- RSA 2023 : Focus more on AI and ML （Everyone’s Buzzing About AI）
- BAAI June 2023： Sam Altman & Hinton （AI Safety & Alignment）
- United Nations June 2023: Will appoint a scientific advisory （AI）
- British Prime Minister Rishi Sunak: The summit on the risks and regulation of AI in UK

What Is The Situation?

" With the emergence of the increasingly powerful AI systems, the stakes for global cooperation have never been higher

Sam Altman

Keynote Speech in the conference hosted by the Beijing Academy of Artificial Intelligence

OpenSSF

OpenSSF Head Delivers AI Warning for Application Security

by Michael Vizard on February 2, 2023

The overall state of application security is likely to worsen if organizations fail to take note of advances in artificial intelligence (AI).

Brian Behlendorf, general manager for the Open Source Security Foundation (OpenSSF) this week warned attendees of the CloudNative Security North America conference that organizations need to assume it is only going to get easier to launch, for example, automated sphear phishing attacks against development teams.

OpenSSF

How do we deal with an uncertain future? How do we collaborate to address the security threats posed by the future AGI era through the power of open source?

WORKING GROUP

Identifying Security Threats in Open Source Projects

This group enables informed confidence in the security of OSS by collecting, curating, and communicating relevant metrics and metadata.

Working Group Git Repo	Working Group Leads	Working Group Membership
github.com/ossf/wg-identifying-security-threats	Michael Scovetta, Microsoft	10-15 members

How can we Identify Security Threats in Open Source LMM Projects?

This is just a Beginning...

A journey of a thousand miles begins with a single step

In Globalization We Trust
In Code We Trust
In OpenSource We Trust