

NICT

サイバーセキュリティ研究所の取り組み

- CYNEX, CYXROSS, AI x Cybersecurity -

国立研究開発法人 情報通信研究機構
サイバーセキュリティ研究所長
井上 大介

NICT サイバーセキュリティ分野の体制図 (2024.6.1)

【第5期中長期計画 サイバーセキュリティ分野】

(1) サイバーセキュリティ技術

- (ア) データ駆動型サイバーセキュリティ技術
- (イ) エマージングセキュリティ技術

(2) 暗号技術

- (ア) 安全なデータ利活用技術
- (イ) 量子コンピュータ時代に向けた暗号技術の安全性評価

(3) サイバーセキュリティに関する演習

(4) サイバーセキュリティ産学官連携拠点形成

(5) IoT機器のサイバーセキュリティ対策の促進

サイバーセキュリティ研究所



井上 大介
研究所長



中尾 康二
主管研究員



大西 祥浩
主管エキスパート

セキュリティ 基盤研究室



篠原 直行 室長

サイバーセキュリティ 研究室



笠間 貴弘 室長



総合企画室
馬淵 奈緒子 室長

ナショナルサイバー トレーニングセンター



園田 道夫
センター長



西浦 智幸
副センター長



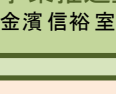
杉本 香純
マネージャー

サイバートレーニング 研究室



花田 智洋 室長

サイバートレーニング 事業推進室



金濱 信裕 室長

ナショナルサイバー オブザベーションセンター



衛藤 将史
センター長

サイバーオブザベーション 運用室



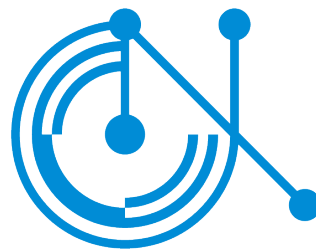
衛藤 将史 室長

サイバーオブザベーション 事業推進室



大西 祥浩 室長

サイバーセキュリティ統合知的・人材育成基盤



CYNEX
CYBERSECURITY NEXUS

サイバーセキュリティ産学官連携拠点の形成に向けて

●サイバーセキュリティ自給率の低迷

✓ サイバーセキュリティ戦略本部 研究開発戦略専門調査会（2019年5月17日）

●データ負けのスパイラル

✓ データが集まらない → 研究開発/人材育成できない → 国産技術を作れない
→ 国産技術が普及しない → データが集まらない → …

●今、日本に必要なこと

- ✓ 実データを大規模に収集・蓄積する仕組み
- ✓ 実データを定常的・組織的に分析する仕組み
- ✓ 実データで国産製品を運用・検証する仕組み
- ✓ 実データから脅威情報を生成・共有する仕組み
- ✓ 実データによる人材育成をオープン化する仕組み



これらの仕組みの実現を目指す
産学官の結節点を構築

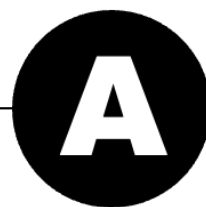
CYNEX：サイバーセキュリティ統合知的・人材育成基盤

- サイバーセキュリティ情報を国内で**収集・蓄積・分析・提供**するとともに、社会全体でサイバーセキュリティ人材を育成するための**共通基盤を構築**し、産学官の**結節点**として開放



71組織
参画中
(2024/6/1)

4つの“Co-Nexus”によるプロジェクト推進



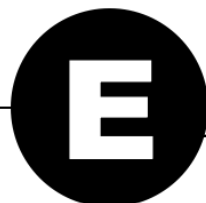
Co-Nexus A (Accumulation & Analysis)

- ✓ 各種観測機構によるデータ収集・蓄積
- ✓ 解析者コミュニティ醸成と共同分析の実現



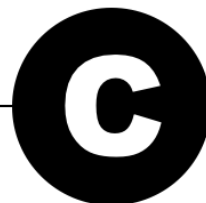
Co-Nexus S (Security Operation & Sharing)

- ✓ 高度SOC人材育成 (Online自主学習 & OJT)
- ✓ 国産脅威情報の生成・提供・情報発信



Co-Nexus E (Evaluation)

- ✓ 国産セキュリティ製品の長期運用・検証
- ✓ 国産セキュリティ製品へのフィードバック



Co-Nexus C (CYROP)

- ✓ サイバーセキュリティ演習基盤のオープン化
- ✓ 演習環境の運用と演習教材の継続的開発

*CYROP: Cyber Range Open Platform

Co-Nexus Chairs



安田 真悟
NICT



毛利 公一
立命館大学



佐藤 隆行
日立製作所



久保 正樹
NICT



piyokango
セキュリティインコ



安部 小百合
NICT

Walküre
CYNEX Red Team



佐藤 公信
NICT



島 成佳
長崎県立大学



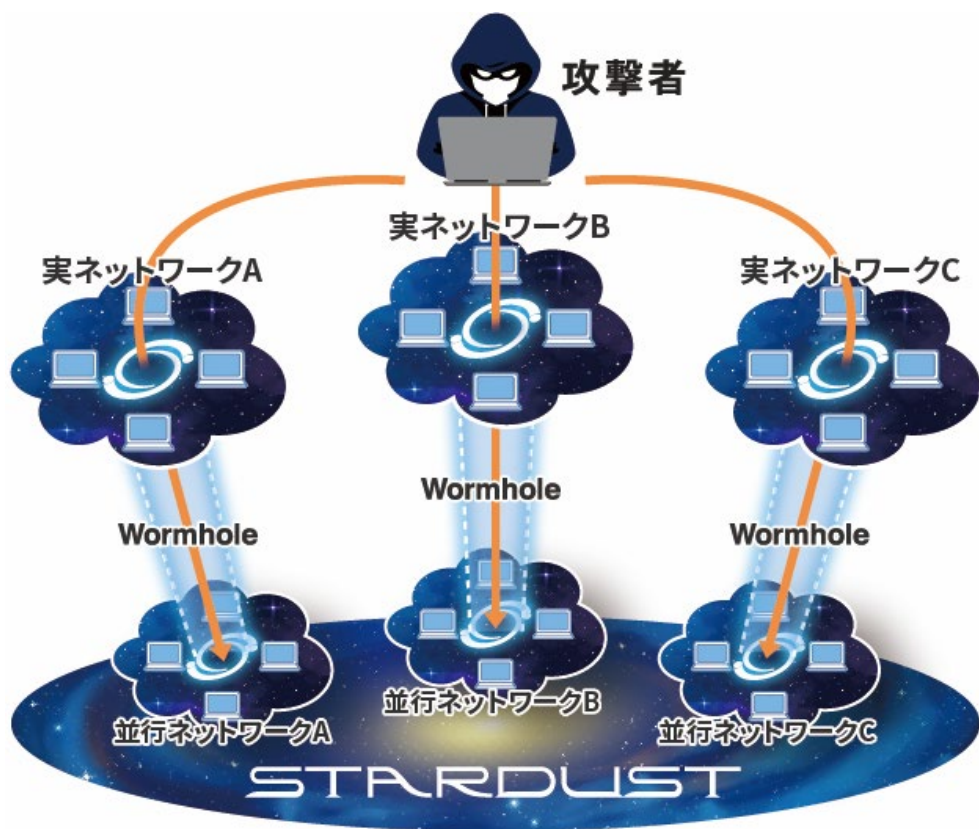
井田 潤
トレノケート

ユニーク参画組織数：71 (A: 36, S: 17, E:7, C:47) (2024年6月時点)

Co-Nexus A : STARDUST & 解析者コミュニティ形成

● 目的：STARDUSTを核とした共同解析と解析者コミュニティ形成

- ✓ STARDUST：人間の攻撃者を誘い込むサイバー攻撃誘引基盤
- ✓ 定常的な攻撃誘引の試行と解析結果を共有する **解析者コミュニティの形成**



サイバー攻撃誘引基盤STARDUST NextGen

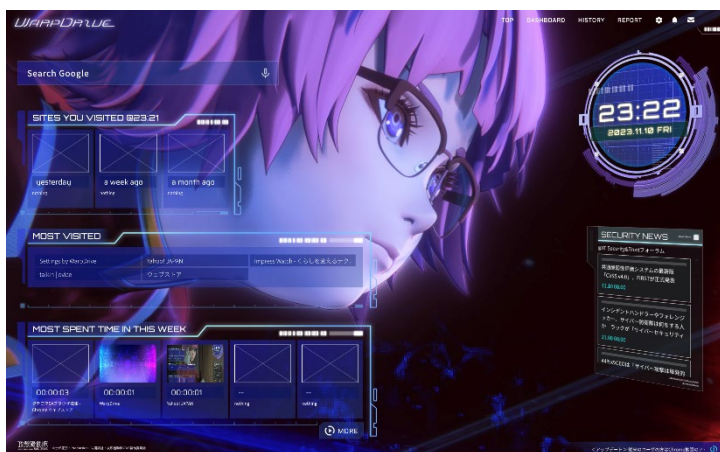
● R5年度活動状況

- ✓ **STARDUST NxtGen** 貸与開始
 - 17組織 (39模擬環境) 提供・利用開始
- ✓ **CURE** データ開放 (R5年度末予定)
- ✓ 年間**300日以上**の攻撃誘引実験
- ✓ **解析者コミュニティ会合** 定期開催
 - 毎回**90名**規模の参加者
- ✓ **脅威情報分析チーム“LETTICE”** 創設
 - piyokango氏を核にしたCYNEXアライアンス選抜チーム

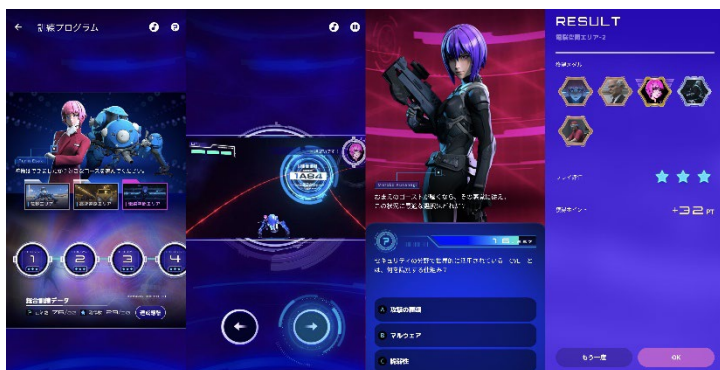
Co-Nexus A：WarpDriveプロジェクトの継承と進化

●目的：Web媒介型攻撃の対策確立のためのデータ収集・分析

- ✓ WarpDrive：ユーザ参加型のWeb媒介型攻撃大規模観測プロジェクト
- ✓ タチコマ・セキュリティ・エージェント（PC/Android版）を無償配布



新PC版
タチコマSA



新Android版
タチコマモバイル



タチコマ

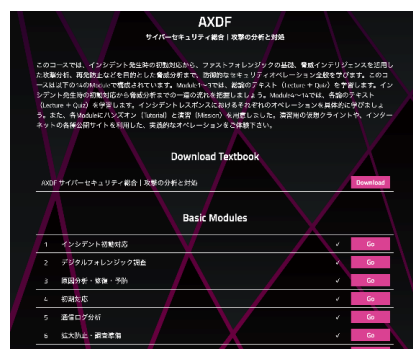
●R5年度活動状況

- ✓ 第2弾アップデート（2023/10/16 リリース）
 - Android向けタチコマモバイルにゲーム機能を開発
 - クイズゲームでセキュリティやITの知識を反復学習
 - 1日最大約3,600ダウンロードを記録
- ✓ 第3弾アップデート（2024/2 リリース予定）
 - PC向けタチコマSAにゲーム機能を開発
 - Android向けにローカルスキャン機能を開発
 ➡ 家庭内のネットワークデバイスを発見可能に
- ✓ WarpDriveコミュニティ発足
 - 産学の研究機関が複数参画
 - データ解析や対策展開を加速

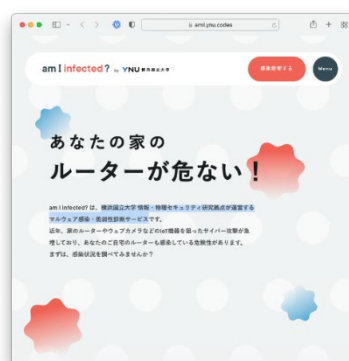
Co-Nexus S：高度SOC人材育成と国産脅威情報発信

●目的：高度な解析者の育成とCYNEX独自の脅威情報の生成・発信

- ✓ 自主学習型 **オンラインコース** と CYNEX 解析チームでの **OJTコース**
- ✓ サイバーセキュリティ関連情報の発信機能のCYNEXへの集約と強化



自主学習型オンライン
SOC研修システム



am I infected



フロー情報分析による
C&Cサーバ検知

● R5年度活動状況

✓ オンラインコース&OJTコース

- オンラインコース：3期生6名 修了、**4期生16名 研修中**
- OJTコース：CYNEX解析チームで1名修了、**1名育成中**

✓ NICTERレポート, Blog, Twitter

- NICTER観測レポートQ1, Q2, Q3, Q4
- NICTER観測レポート2023 (2024/2 公開)
- NICTER Blog : 6件、Twitter 随時更新

✓ am I infected? への情報提供

* 横浜国立大学 情報・物理セキュリティ研究拠点が運営するマルウェア感染・脆弱性診断サービス

✓ C&Cサーバ検知に関する調査 への情報提供

* 電気通信事業者におけるフロー情報を用いたC&Cサーバ検知に関する手法や有効性の調査

Co-Nexus E：国産セキュリティ製品の運用・検証

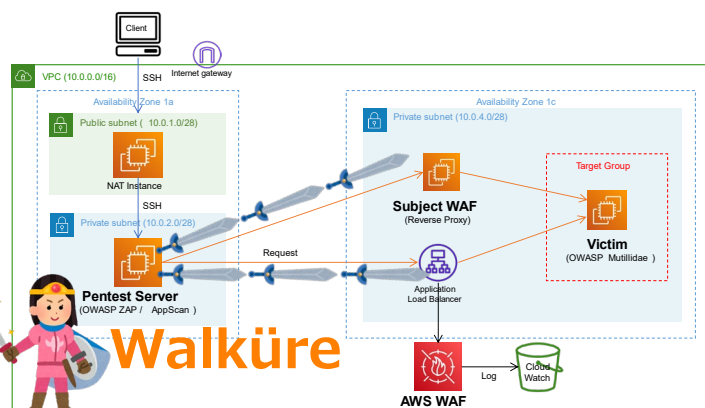
- **目的：国産セキュリティ製品のテスト環境提供による実用化支援**
 - ✓ NICT内部ネットワークにおける **国産セキュリティ製品の長期運用・検証**
 - ✓ **Walküre (CYNEX Red Team)** の模擬攻撃によるセキュリティ機能検証

国産セキュリティ製品検証リスト

製品種別	製品フェーズ	運用・検証の概要
ペネトレーションツール	商用化前技術	ツールの高度化及び長期運用
ファジングツール	商用化前技術	アルゴリズムの精度検証
IPレピュテーションサービス	商用化済技術	運用されている製品の精度検証 新たな分析軸の検証
ランサムウェア対策ソフト	商用化済技術	新たなマルウェアへの適応検証



IoT機器検証環境



WAF製品検証環境

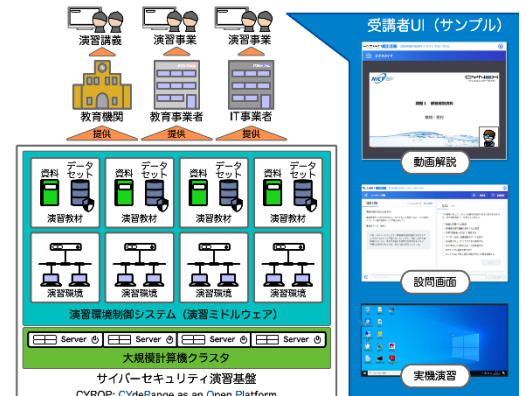
● R5年度活動状況

- ✓ **延べ6製品の長期運用・検証**を実施
 - 2製品は検証を終了し、商用化フェーズに移行
 - 4製品（左記）は引き続き検証継続中
- ✓ 各製品ごとに**カスタム検証環境構築**
 - IoT機器検証環境、WAF製品検証環境、etc...
- ✓ **Walküre**による模擬攻撃の実施
- ✓ 海外有力製品群との**比較検証**
- ✓ 民間企業への**フィードバック**

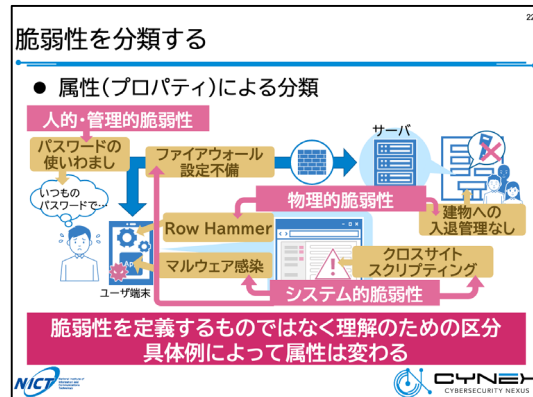
Co-Nexus C：人材育成オープンプラットフォーム

●目的：演習基盤開放による国内セキュリティ人材育成事業の活性化

- ✓ サイバーセキュリティ演習に必要な演習環境と演習教材をオープン化
- ✓ 産学官のニーズに基づき、NIST NICE Frameworkに沿って演習教材整備



サイバーセキュリティ演習基盤
CYROP (Cyber Range Open Platform)



CYNEXオリジナル演習教材

●R5年度活動状況

✓ CYROP本格稼働開始

- 大学・民間企業等21組織が参画

✓ 3組織が商用演習サービスを開始

- CYDER Aコース由来 演習教材 (順次受け入れ)
- CYDER Bコース由来 演習教材 (順次受け入れ)
- CYNEX オリジナル 演習教材 (拡充中)

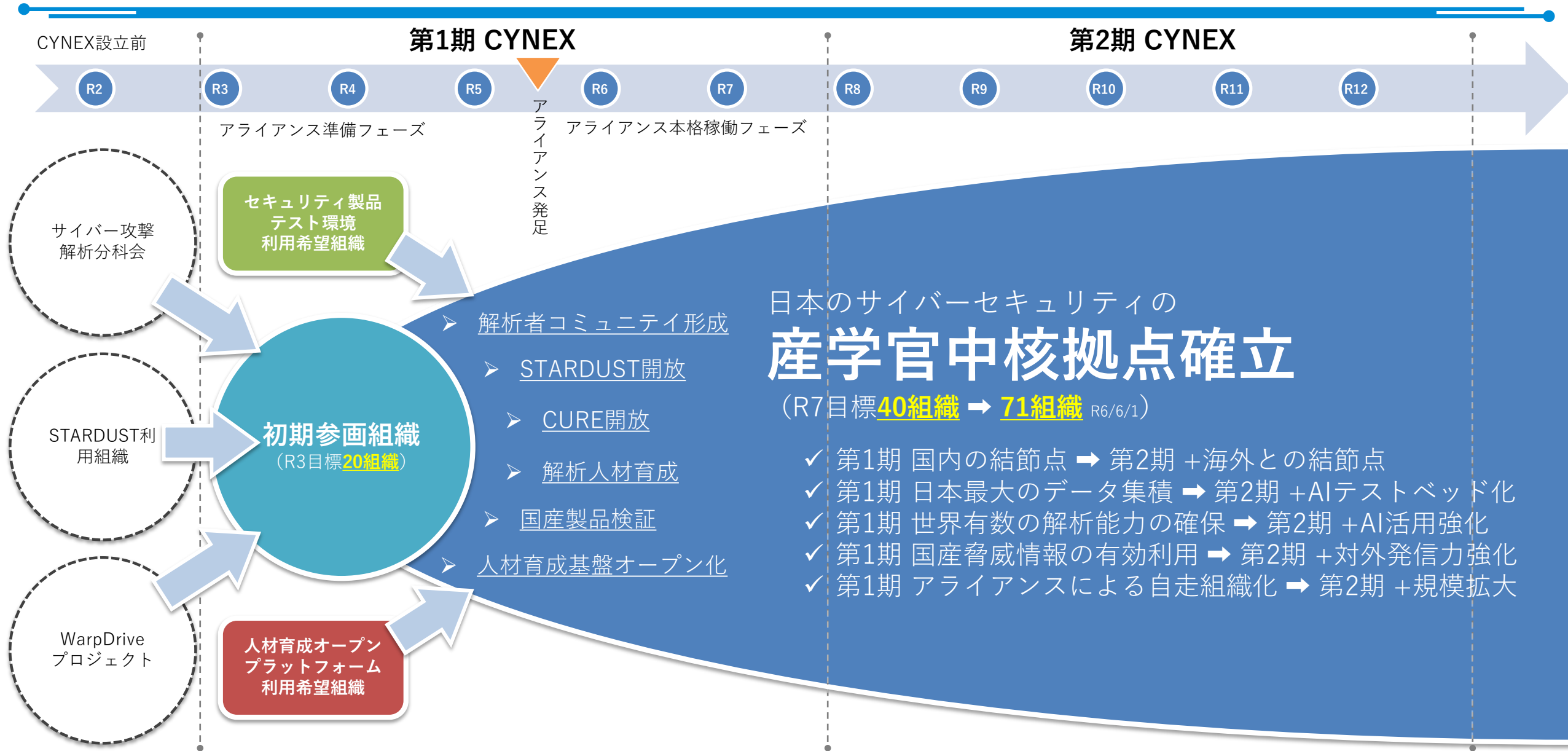
✓ 新規演習教材の共同開発を実施

- 2021年：CYDERコンテンツ、パケット解析等 (18種)
 - 2022年：セキュリティ管理、ペネテスト等 (18種)
 - 2023年：フォレンジック、ログ収集・分析等 (29種)
- ➡ 2022年度までに作成した教材のうち71%が実利用される

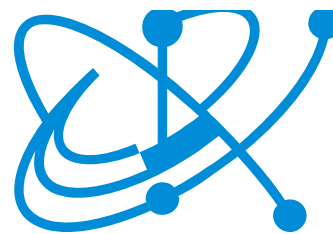


大学での演習教材利用事例

CYNEXの事業展開のタイムライン



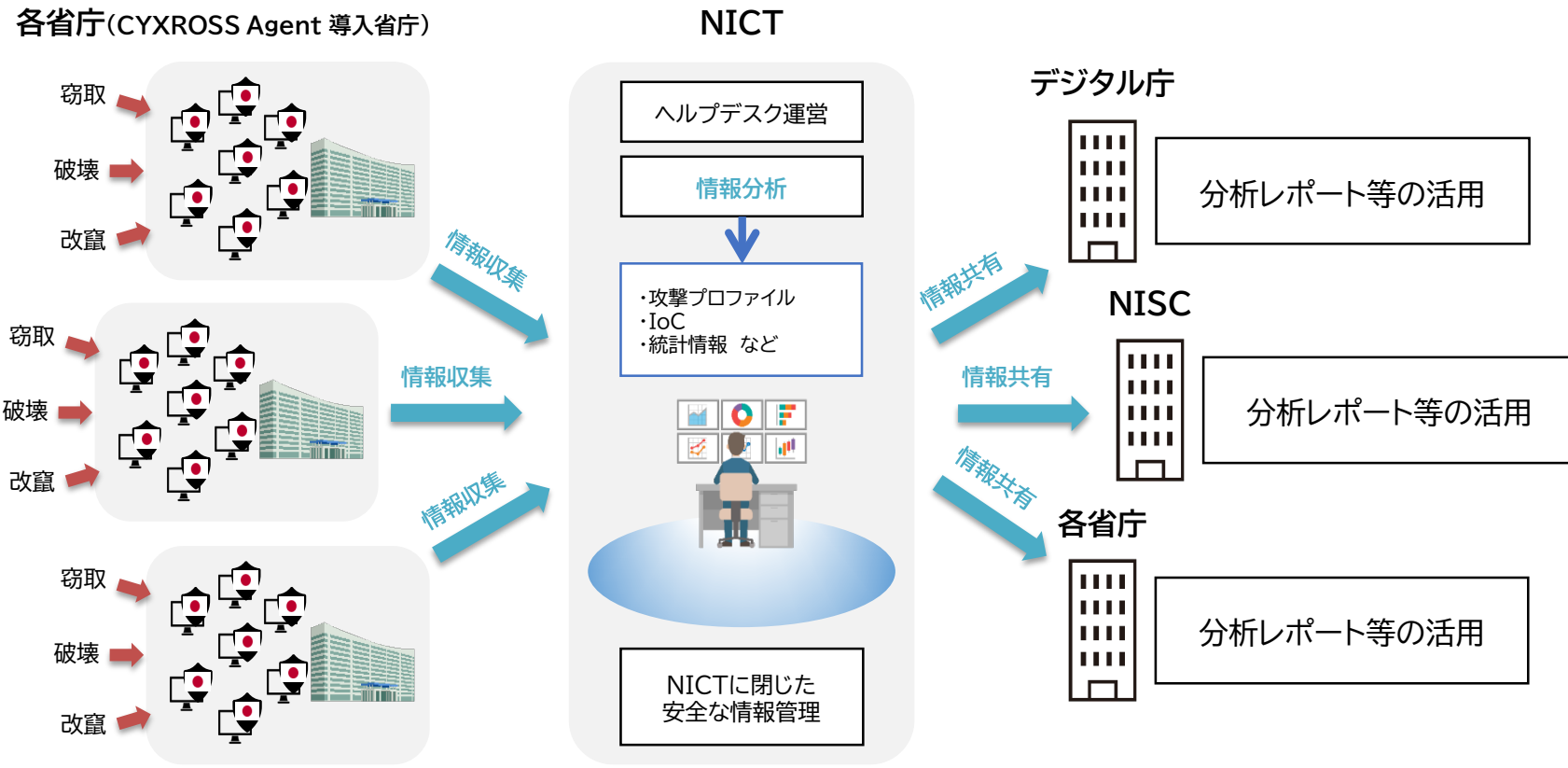
政府端末情報を活用した サイバーセキュリティ情報の収集・分析に係る実証事業



CYXROSS



- 安全性や透明性の検証が可能なNICT開発のセンサーを政府端末に導入し、得られたマルウェア情報等をNICTにおいて集約・分析する実証事業を実施



- ✓ NICT開発のセンサー『[CYXROSS Agent](#)』を政府端末に導入し情報収集
- ✓ NICT CYNEXで[組織横断的情報分析](#)を行い、デジタル庁やNISC等に情報共有、今後GSOCとも連携予定

CYXROSSの意義と活動状況

● 自国の健康診断は自国の技術で

- ✓ 国産技術『CYXROSS Agent』による日本独自の政府端末情報収集
- ✓ CYNEXを中心に複数省庁が連携する「ネクサス型」の横断分析環境の確立
- ✓ CYXROSS Agent x CYNEXデータ x CYNEXアナリスト = 日本の情勢分析能力強化

● 各省庁におけるセキュリティ対策を強化

- ✓ NICTによる分析結果をデジタル庁やNISC、CYXROSS Agent導入省庁に情報共有
- ✓ CYXROSS Agentは各省庁の既存のエンドポイント対策製品と共存（セカンドオピニオン）

➡ 我が国のサイバー安全保障の確保に貢献

● R5～R6活動状況

- ✓ CYXROSS Agent Ver 1.0 実装完了、CYXROSS CORE（オペレーションルーム）運用開始
- ✓ 総務省端末への導入開始、今後政府全体の端末に導入拡大へ

AI x Cybersecurity

USENIX Security 2023



USENIX Security 2023

- 2023年8月9～11日@Anaheim, US
- 参加者 > 1,000人
- 論文投稿数：1,444
- 採択数：422（採択率29.2%）
 - ✓ Major Revisionからのシェパードニングプロセス導入により採択率が昨年比11.1%上昇
- 6パラレルセッション
- 合計セッション：90
- 機械学習（ML）関連セッション：11
 - ✓ 他のセッションの論文も大多数は評価でMLを利用
- ドイツCISPA無双（採択論文30本以上）
- 中国系研究大增（特にML関連研究）
- 韓国系研究躍進（採択論文10本）
- 日本系研究希少（採択論文4本）

<https://www.usenix.org/conference/usenixsecurity23>

USENIX Security 2023 Program（黄色が機械学習関連セッション）

AIセキュリティに関する米国の動向

● バイデン大統領令（2023年10月）

- ✓ 人工知能の安全・安心・信頼できる開発と利用に関する大統領令
- ✓ 国際的な同盟国やパートナーと協力してAIのリスク管理をリード

NSA

AI Security Center 設立 (2023年9月)

- ✓ 米国国家安全保障および防衛産業において、AI機能の安全な開発、統合、導入を促進
- ✓ Cybersecurity Information Sheet “Deploying AI Systems Securely: Best Practices for Deploying Secure and Resilient AI Systems” 発行（2024年4月）

NIST

AI Safety Institute 設立 (2023年11月)

- ✓ AIの安全性と信頼性、最先端のAIモデル評価の実施組織として設立
- ✓ AISI Consortium (AISIC) を構築し、AI設計者、ユーザ、学会・政府・産業界の研究者を集結
- ✓ 既にAIを用いたシステムの安全性、セキュリティ、信頼性を向上するためのドラフトを4本公開

MITRE

AI Assurance and Discovery Lab 設立 (2024年3月)

- ✓ 国家安全保障、医療、運輸などの重要アプリケーションで使用するAI対応システムを評価
- ✓ シミュレーション環境でのリスク発見、AIレッドチームング、LLM評価、人間参加型実験
- ✓ MITRE ATLAS (AIベースシステムへの攻撃手法を体系化)

MITRE ATLAS

● ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems)

- ✓ AIベースのシステムに対する攻撃手法を体系化したナレッジベース（2021年6月公開）
- ✓ MITRE ATT&CKに準拠し、AIへの攻撃のTTPを掲載
- ✓ AI/MLに対する攻撃の脅威認識を高め、対応・準備させることが目的

● ATLASの発展状況

- ✓ 14 Tactics, 82 Techniques, 20 Mitigations, 22ケーススタディ（2024年5月時点）
- ✓ コンテンツは拡充していく必要があり、そのための研究開発や連携が求められている

Reconnaissance ^{&}	Resource Development ^{&}	Initial Access ^{&}	ML Model Access	Execution ^{&}	Persistence ^{&}	Privilege Escalation ^{&}	Defense Evasion ^{&}	Credential Access ^{&}	Discovery ^{&}	Collection ^{&}	ML Attack Staging	Exfiltration ^{&}	Impact ^{&}
5 techniques	7 techniques	6 techniques	4 techniques	3 techniques	3 techniques	3 techniques	3 techniques	1 technique	4 techniques	3 techniques	4 techniques	4 techniques	6 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	ML Model Inference API Access	User Execution ^{&}	Poison Training Data	LLM Prompt Injection	Evade ML Model	Unsecured Credentials ^{&}	Discover ML Model Ontology	ML Artifact Collection	Create Proxy ML Model	Exfiltration via ML Inference API	Evade ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities ^{&}	Valid Accounts ^{&}	ML-Enabled Product or Service	Command and Scripting Interpreter ^{&}	Backdoor ML Model	LLM Plugin Compromise	LLM Prompt Injection		Discover ML Model Family	Data from Information Repositories ^{&}	Backdoor ML Model	Exfiltration via Cyber Means	Denial of ML Service
Search Victim-Owned Websites	Develop Capabilities ^{&}	Evade ML Model	Physical Environment Access	LLM Plugin Compromise	LLM Prompt Injection	LLM Jailbreak	LLM Jailbreak		Discover ML Artifacts	Data from Local System ^{&}	Verify Attack	LLM Meta Prompt Extraction	Spamming ML System with Chaff Data
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application ^{&}	Full ML Model Access						LLM Meta Prompt Extraction		Craft Adversarial Data	LLM Data Leakage	Erode ML Model Integrity
Active Scanning ^{&}	Publish Poisoned Datasets	LLM Prompt Injection											Cost Harvesting
	Poison Training Data	Phishing ^{&}											External Harms
	Establish Accounts ^{&}												

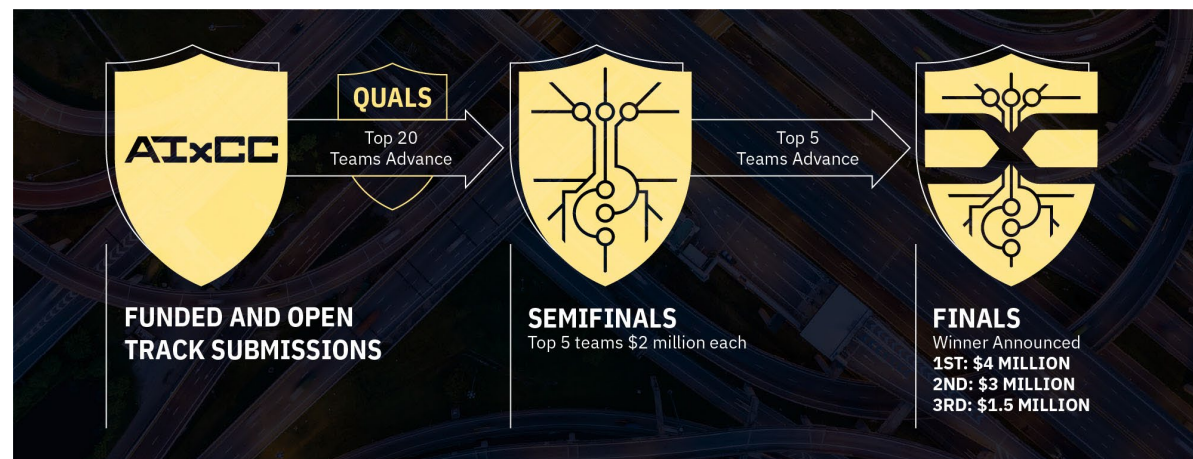
AIによるサイバーチャレンジ：DARPA: AI x CC

● Cyber Grand Challenge (2016)

- ✓ 優勝賞金200万ドル（約3億円）
- ✓ 全ての攻防戦がコンピュータにより自動で実施され、ヒトは見ているのみ
- ✓ 自動で脆弱性を発見、パッチ作成、対処
- ✓ CMUのシステム『Mayhem』が優勝

● AI x CC: AI Cyber Challenge (2024-2025)

- ✓ ソフトウェアの脆弱性を自動的に発見・修正し、サイバー攻撃から重要インフラを守る競技
- ✓ AIとサイバーセキュリティの統合にイノベーションの推進が目的
- ✓ 準決勝：Black Hat 2024
- ✓ 決勝：Black Hat 2025



NICTにおける AI x Cybersecurity 研究（これまで）

AI x Cybersecurity領域の研究開発

ダークネット分析

- マルウェア活動の早期検知
- 広域スキャナのフィンガープリント

侵入検知

- セキュリティアラートの分析・削減
- 説明可能な侵入検知

Webセキュリティ

- Webのアクセスログ分析
- 説明可能なフィッシングサイト検知

オペレーション 自動化

AI for Security

マルウェア分析

- IoTマルウェア分析
- Androidアプリ分析

脅威情報分析

- ダークウェブからの脅威情報抽出
- 脆弱性情報の分析・アノテーション

研究成果の活用・社会展開

- 研究用データセット公開
- CYNEX解析チーム/NICT-CSIRTでの実オペレーションへの活用
- 国際連携：データセット共有、インターン受入、リクルーティング、共同分析、論文共著

NICTにおける AI x Cybersecurity 研究（これから）

● NICTの強みを活かした AI x Cybersecurity 研究へ

✓ AI for Security

- 例：セキュリティ分野のコーパスを強化した自動翻訳技術で日/英以外の情報源を拡大

✓ Security for AI

- 例：大規模言語モデルの開発・利用に関するリスク評価、脅威分析、対策確立

● 日本の AI x Cybersecurity 研究の国際競争力強化

- ✓ サイバーセキュリティ分野は政策と研究開発が密接不可分

- ✓ 政策面で緊密に連携している日米

→ 研究開発面でも連携強化が重要（研究開発の成果を政策にフィードバック）

- ✓ MITREやNISTとの連携や人材交流

→ NICT内に米国研究機関との結節点の形成を模索