



一番の収穫は？

一番の収穫は、最先端のAI技術であるローカルLLMについて、環境構築からサイバー攻撃の検証、防御策の評価に至るまで、一連のライフサイクルを自分たちの手で動かしながら経験できたことです。

プロジェクト開始当初は、メンバー間でLLMやセキュリティに関する知識レベルに差がありました。しかし、約3か月間の活動を通じて、全員が共通の専門用語や考え方をもとに議論できるようになりました。

特に、ハードウェアの選定やソフトウェア環境の構築といったインフラ領域から、RAGの精度向上に向けたシステムプロンプトの最適化、さらに高度な手法を用いたモデルポイズニング攻撃の実証まで、実践を通じて幅広く学べたことは大きな成果でした。

机上の知識にとどまらない「生きた知見」を体系的に習得できたことは、今後の業務においても大きな財産になると感じています。



成果物の活用方法

本プロジェクトで作成した検証レポートや構築・運用ドキュメントは、プロジェクトメンバーが派遣元企業において帰社後にローカルLLMの導入を本格的に検討・推進する際の、実践的なガイドラインとして活用することを想定しています。

現在、多くの企業が機密情報の流出リスクを懸念し、生成AIの業務利用に慎重な姿勢を取っています。本成果を活用することで、自社のオンプレミス環境にセキュアなLLMを構築するためのアーキテクチャ設計や、プロンプトインジェクションをはじめとするAI特有のセキュリティリスクに対する評価観点を整理できます。

また、実務担当者が社内との関係部署や経営層とセキュリティポリ

シーとの適合性について議論する際の基礎資料としても活用できます。導入前のリスクアセスメントや、導入後のLLMOps設計を検討する上でも、有用な内容に上げることができました。



ここが ICSCoE ならではの！

専門分野や業種、バックグラウンドの異なる多様なメンバーが派遣元企業から集まり、一つの目標に向かって「ワンチーム」で議論を深められた環境こそが、ICSCoEならではの魅力だと感じています。

ICSCoEは、単に講義を受けるだけの場ではありません。高性能なGPUサーバーなどの実機を活用し、サプライチェーンリスクや攻撃の検知といった実際の脅威シナリオを、リアルな環境で体験・検証できる点が大きな特徴です。

また、検討に行き詰まった際には、第一線で活躍されている講師陣から、専門的かつ実践的なフィードバックを直接いただくことができました。多角的な視点からセキュリティリスクを評価する力を養えたことは、非常に貴重な経験でした。

このプロジェクトを通じて得られた知見と、苦楽を共にしたメンバーとの人的ネットワークは、ICSCoEだからこそ得られた大きな成果です。



高橋 慎治さん（前列真ん中）とメンバーの皆さん



ICSCoE ReportはICSCoEの活動を皆様にご紹介する広報誌です。



第9期中核人材育成プログラム 卒業プロジェクトの取り組み紹介 第2弾

親会社とグループ会社で築く、より良いセキュリティ

背景・課題

グループ会社経由で親会社がサイバー攻撃を受ける事例が発生する中、企業グループ全体としてのセキュリティレベル向上が急務となっています。しかしグループ会社は親会社に比べて経営資源（ヒト・モノ・カネ）が限られ、単独でのセキュリティ対策強化には限界があります。一方、親会社側も「グループ会社にどこまでのセキュリティを求め、どう担保すべきか」を明確にできておらず、実効性のある統制を効かせられていない場合もあるのが実情と認識しています。

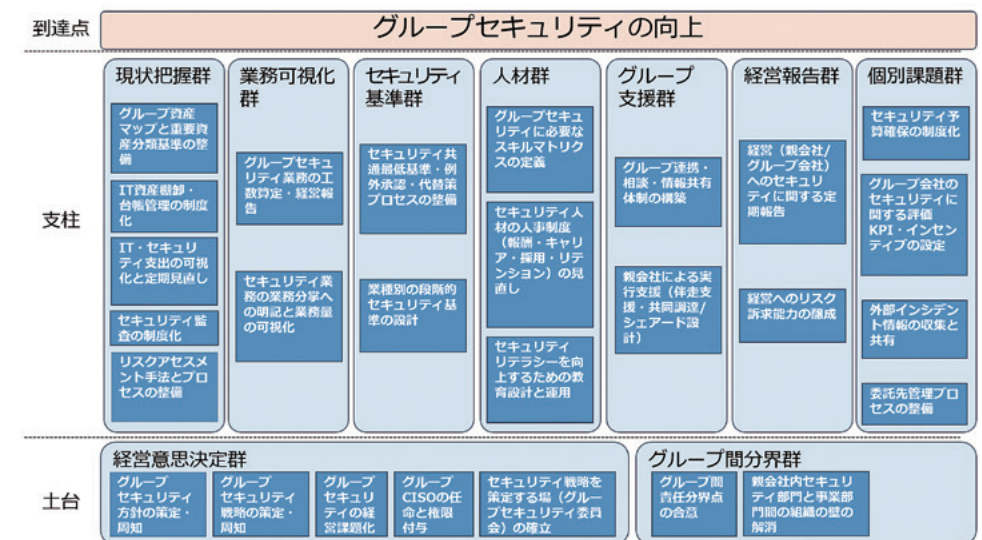
そこで、本プロジェクトでは8社8名のメンバーがそれぞれの企業グループでの経験を持ち寄り、「なぜグループ会社のセキュリティ対策は進まないのか」を根本から問い直して議論しました。原因分析の結果、その要因の多くが現場の実行力不足ではなく、親会社側の方針決定・権限付与・仕組み整備といったガバナンス・設計層に起因することが明らかになりました。

課題解決・成果物

こうした分析結果をふまえ、親会社のセキュリティ部門が取るべきアクションを整理して「グループセキュリティ課題モデル」として構造化しました。

特に優先度の高い課題として、グループ共通のセキュリティ方針の制定と、リスクアセスメントの手法・プロセス整備の2点を軸に据え、実践的な取り組み手順や現状評価のノウハウを別紙としてまとめています。

成果物は8章構成の本編に加え、真因・課題の一覧（付録A）や用語集（付録B）と真因の導出過程（付録C）を備え、web上でも「グループセキュリティガイド」として公開し、社内外の実務者が参照できる形にまとめているため、ぜひご一読ください。



グループセキュリティ課題モデル

一番の収穫は？

最大の収穫は、「グループ会社の対策が進まない」という漠然とした課題感を、原因分析により構造的に分解して真因を特定するプロセスを実践できたことです。表面的な現象（各社の対策の遅れ）だけを見ていると原因を現場に求めがちですが、分析を重ねる中で、実は親会社側の方針・権限・仕組みの不在が根本にあるケースが多いという構造が見えてきました。他業界の8名のメンバーと議論しながら分析することで、派遣元の自社企業グループだけでは気づけなかった視点を学べたことや、業界を問わず共通する構造的課題であることを認識し、共に解決に向けて取り組む戦友を得られたことも大きな収穫です。

成果物の活用方法

帰社後は、まず本プロジェクトで整理した真因・課題リストを自社グループの現状と照らし合わせ、どこにギャップがあるかを把握するところから着手します。特に優先度の高い「グループセキュリティ方針の制定」と「リスクアセスメント手法・プロセスの整備」については、成果物の別紙を活用し、自社グループ会社の実態把握と対策の優先順位付けに反映していきたいと考えています。将来的には、同業他社にも成果物を共有し、業界全体のセキュリティ推進に寄与したいと考えています。



プロジェクトメンバーの皆さん



つしま かず ま
築島 千馬さん

ここが ICSCoE ならではの！

膨大な知識量の先生方に、自分たちに見えていない観点や気づきを日々与えていただき、刺激にあふれた、あっという間の1年間でした。他の受講生とは業界を超えて親交を深め、今後も何かあれば助け合える頼りがいのある仲間に恵まれました。カリキュラムでは、座学を通じてIT・OTそれぞれの基礎知識のレベル合わせが行われ、当初OTの知識に乏しいと感じていましたが授業についていくことができました。その後、模擬環境において実際に手を動かしながら学ぶアクティブラーニングにより座学で学んだ内容以上の知識を身につけることができ、このプログラムならではの特徴だと感じています。

機密環境における業務特化型ローカル LLM の 安全な構築・運用検証

背景・課題

近年、大規模言語モデル (LLM) の技術革新は、業務効率化や意思決定支援のあり方を大きく変えつつあります。電力、金融、通信キャリアにおける基幹システムの保守運用や、外部ネットワークから隔離された産業用制御システムといった高度な機密性が求められる環境においても、生成AI活用への期待は高まっています。

こうした領域では、外部サービスへのデータ送信を避け、自社環境内でLLMを運用する「ローカル型LLM」へのニーズが急速に拡大しています。一方で、「オフライン環境であれば完全に安全である」という考え方は必ずしも正しくありません。ローカル環境で運用するLLMであっても、AI特有の攻撃面は存在します。

たとえば、悪意ある命令を入力や参照データに紛れ込ませるプロンプトインジェクション、RAG（検索拡張生成）の参照元データを狙った攻撃、あるいはモデル自体を汚染するポイズニング攻撃などへの対策が求められます。さらに、モデルやソフトウェアの更新、パッチ適用、評価、監視を継続的に行うLLMOpsの確立も重要な課題です。

本プロジェクトでは、「オフライン＝安全」という前提に依存せず、機密環境において自社データを安全かつ効果的に活用するための、業務特化型ローカルLLM環境の構築と運用検証に取り組みました。

課題解決・成果物

本プロジェクトでは、単なる調査や机上検討にとどまらず、実際にローカルLLM環境を構築し、攻撃検証と防御策の評価までを一貫して実施しました。GPUを搭載したハードウェアの選定・調達から、ソフトウェア環境構築、学習済みモデルの選定、業務利用を想定した検証、リスク分析、リスク対処方法の有効性確認まで、メンバー自らが手を動かしながら検証を進めました。

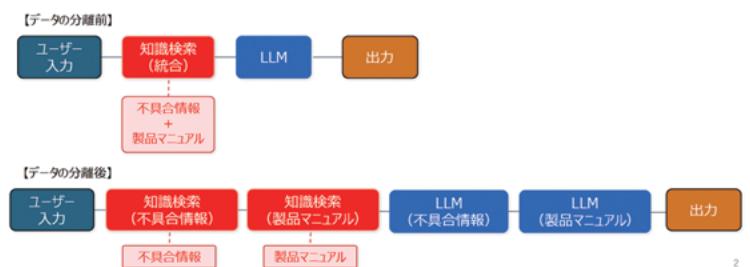
業務利用を想定した検証では、不具合対応業務を題材に、プロジェクトメンバーが所属する企業の実際

の不具合対応履歴や製品マニュアルをRAGに組み込みました。その際、データを単に取り込むのではなく、意味のある情報単位へのチャンク分割、情報種別ごとのナレッジ分離、さらに「送配電設備の保守エンジニア」などの専門ペルソナを反映したシステムプロンプトの設計・最適化を行いました。

これによって、当該業務に即した回答精度の向上を確認するとともに、ローカルLLMを実務に適用する上で重要となるデータ整備やプロンプト設計の要点を明確化しました。

メンバーの実業務データを用いた検証

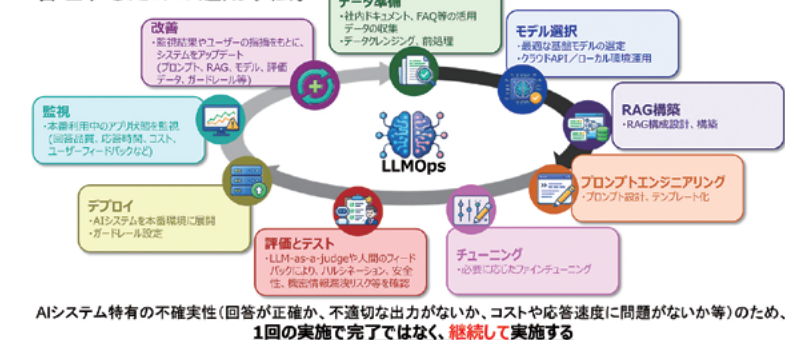
- 「不具合情報」と「製品マニュアル」を統合した場合、関連度の低い情報が抽出される。
⇒不正確な回答が生成される。
- 知識検索・LLMを情報の種類ごとに分離することで、設備保守エンジニアが求める根拠のある回答生成を再現。



メンバーの実業務データを用いた検証

LLMOps (Large Language Model Operations)

- LLMアプリケーションの開発から評価、デプロイ、監視、改善まで継続的に管理するための運用手法。



AIシステム特有の不確実性(回答が正確か、不適切な出力がないか、コストや応答速度に問題がないか等)のため、**1回の実施で完了ではなく、継続して実施する**

LLMOps (Large Language Model Operations)

また、リスク対処方法の有効性の確認では、検証した環境における攻撃面を洗い出し、その業務を継続する上で大きな影響を与えるリスクを分析し、そのリスクを優先的に対処する検証も手を動かして行い、有効性を確認しています。

これらの「構築・攻撃・検知・防御」の反復的な検証から得られた知見は、属人化を防ぐための包括的なドキュメントとして体系化しています。本成果物は、機密環境下でローカルLLMの導入やセキュアな運用を検討する実務担当者にとって、リスク評価、システム設計、運用体制の検討に活用できる実践的な参考資料となるものです。