RAG と生成 AI による学内専用チャットボットの開発
～情報基盤センター問い合わせ対応業務軽減の取組み～

○原　祐一
静岡大学技術部情報・機器分析系情報部門

１．はじめに

静岡大学情報基盤センターでは、2025 年 1 月より AI-RAG を実現できる Dify クラウドを用いて、情報基盤センターサイト内のすべての情報の中から、質問に応じたコンテキストを選択して、生成 AI による応答を行う AI チャットボットをリリースした。また、Dify は、チャットボット以外にも応用がきくことがわかり、他のシステムも開発したので、この開発業務について紹介する。

２．AI-RAG とは

AI-RAG の RAG は、「Retrieval Augmented Generation」の略で、検索拡張生成と訳すことができ、生成 AI モデルと検索技術を組み合わせた手法である。ユーザーからの質問に対して、準備したナレッジから関連情報を検索し、その情報を基に AI が回答を生成する。これにより、AI は特定のデータに基づいた応答が提供でき、回答の正確性や信頼性が向上することで、生成 AI の課題であるハルミネーションを解決する手段の１つと考えられている。

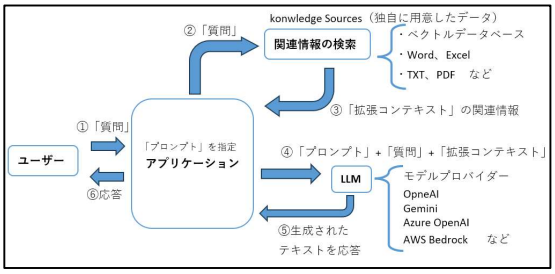

図１　AI-RAG 概念図

３．AI-RAG vs AI 検索エンジン

従来の AI 検索エンジンは、ユーザーのクエリに対して関連性の高いウェブページやドキュメントのリストを提供する。ユーザーはその中から必要な情報を探し出す必要があるため、検索結果の精度は検索アルゴリズムやインデックスの質に依存する。

表１　AI-RAG と AI 検索エンジンの違い

項目	AI-RAG	AI 検索エンジン
情報源	事前の限定されたデータソース	インターネット全体（限定ページは不可）
実時間性	データソース更新日時に依存	高い最新情報に対応可能
用途	専用的・限定的なトピック	幅広いトピック、日常の情報収集
信頼性	回答が精密	要約的な回答が得意
誤情報	ほぼ無い	あり得る
誤読解	あり得る※制御できる可能性有	あり得る

AI-RAG と AI 検索エンジンは似て非なるもの

４．AI-RAG を用いたシステム開発

４.１　開発までの準備

AI-RAG のシステム開発に向けて、python による LangChain を用いた自作 AI-RAG（OpenAI API Platform + Chainlit 使用）、RAG エンジンを搭載している AI アプリ開発プラットフォーム Dify オンプレミス（自前サーバで構築）、Dify クラウドプラットフォームを試した。その中で、Dify クラウドサービスは、バージョンアップ自動化、ノーコード・ローコード開発が可能、サーバ管理が必要ないなど、開始からランニングまで、利用がすごく簡単！！という点から活用することを決めた。

４.２　開発システム①「学内専用 AI-RAG チャットボット」の紹介

AI-RAG チャットボットを実現するためには、Dify のナレッジに応答に利用するデータと、そのデータを検索して AI に生成させるアプリを作成する必要がある。そこで、必要なデータは情報基盤センター学内ページにすべて記載されているため、Web スクレイピングで各ページデータを取得して、html タグをすべて削除して、Dify ナレッジ API を利用して登録・更新するスクリプトを作成した。このスクリプトを定期実行することで、Dify ナレッジを最新の状

態に維持できるようにした。次にチャットフローを使って Dify アプリを作成した。

図 2 AI-RAG チャットフロー

LLM のプロンプトには、どのように応答してほしいかを記述し、回答には「OpenAPI gpt-4o-mini」を設定した。

プロンプト「静岡大学情報基盤センター窓口対応を支援するためのチャットボットです。あなたはカスタマーサポートの専門家です。ユーザーからの問い合わせに対して、わかりやすく丁寧に回答してください。」

最後に Dify ではチャットボットをするための公開 URL が準備されているため、この公開 URL を学内専用 Web ページで利用できるように設定した。

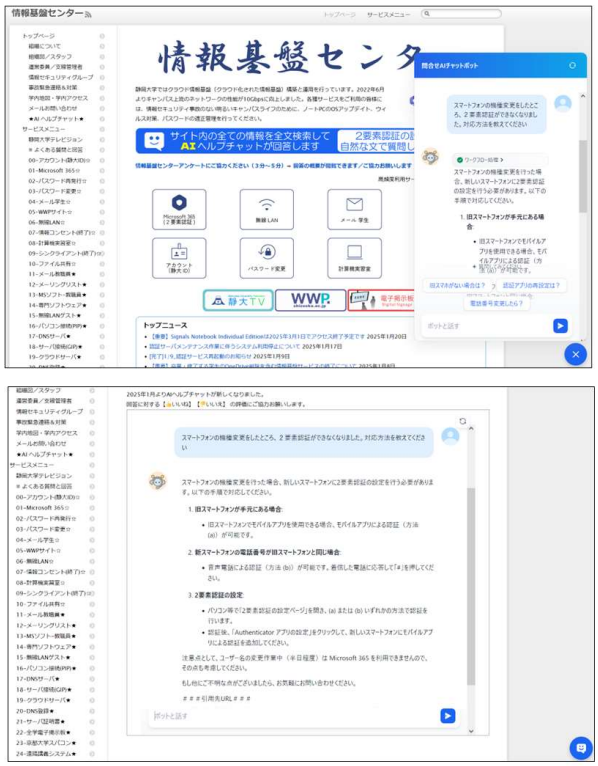

図 3　AI-RAG チャットボット実装

4.3　開発システム②「UTM アラート AI 解析」の紹介

Dify は、Dify アプリ API を利用して、データを送り、生成 AI が分析した回答を受け取るアプリを作成することができる。この機能を利用して、24 時間のアラートログを分析して、関係者にメール送信するスクリプトを作成した。

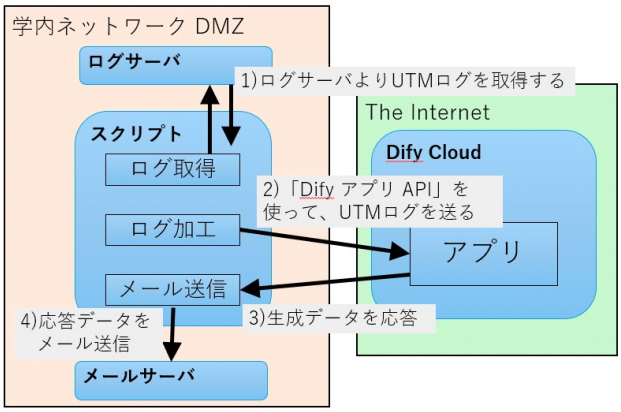

図 5 UTM アラート AI 解析　概念図

まず、アラートデータを受取り、LLM（OpenAI gpt-4o-mini）で分析して、レスポンスを返すアラートを分析するための Dify アプリを作成する。

プロンプト「SYSTEM:あなたは 25 年以上の経験を有するセキュリティの専門家であり、サイバーセキュリティアナリストアシスタントです。」「USER:ログを分析し、分析結果を回答してください。具体的に「脅威度、UTM の対応、脅威の説明、リスクの説明、対応方法」を解説してください」

Dify アプリ作成後、スクリプトを作成する。スクリプトで、ログサーバから 24 時間分のアラートメールを取得して、必要ないデータを削除し、加工する。加工データを Dify アプリ API で、Dify に送り、アプリ内で生成したレスポンスを受取り、関係者にメール送信する。

4.4　開発システム③「脆弱性診断レポートの AI 解説および AI チャットボット」の紹介

脆弱性診断レポートを API で Dify へ送り、生成 AI で解説したデータを取得し、受け取ったデータを内製チャットボットに表示しつつ、ナレッジソースとして活用して、脆弱性診断に関することのみ AI チャットを継続して行うことができるシステムを内製した。

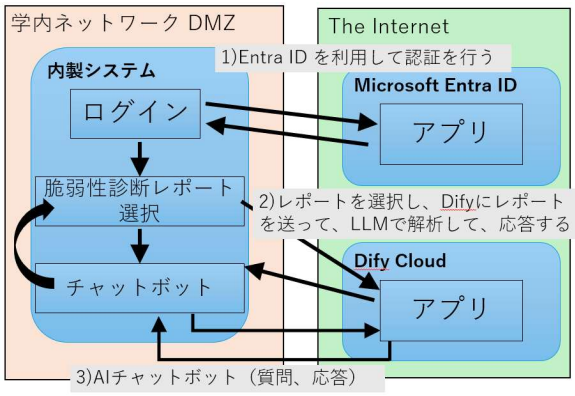

図 6 脆弱性診断レポートの AI 解説および AI チャットボット 概念図

まず、内製システムにログイン（Microsoft Entra ID 利用）する。ログイン後、自身が管理しる脆弱性診断レポートが表示されるので、AI 解説したいレポートを選択し、Dify アプリ API を使って Dify アプリにレポートを送る。

図 7 レポート選択画面

Dify アプリで LLM を使って解説して、結果を返信し、

図 8 Dify アプリ（ワークフロー）

返信結果を内製チャットボットで表示する。

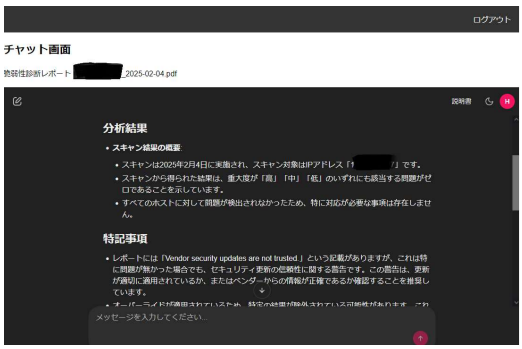

図 8 チャット画面

継続して、Dify アプリに質問を送ることで、生成 AI と「質問」「応答」を行うことができる。

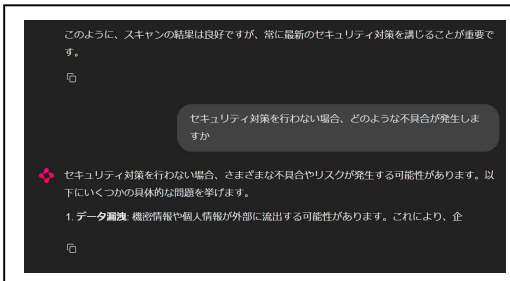

図 9 AI チャットボット

以上が Dify を活用して開発した３つのシステムである。

謝辞

本業務を進めるにあたり、ご協力頂いた、静岡大学情報基盤センター長　教授　長谷川孝博先生、並びに情報基盤センターおよび情報企画課スタッフには、ここに深く深謝の意を表します。