The Chatbot Confession

Overview

Room URL: https://tryhackme.com/room/adventofcyberpreptrack

Difficulty: Easy

Category: Prep

Date Completed: 12/1/2025

Objective

Identify which chatbot messages contain sensitive information.

Table of Contents

Introduction

TBFC's AI assistant, **FestiveBot**, designed to help compose cheerful emails, has begun leaking sensitive information including internal URLs and passwords. While AI tools are powerful productivity aids, defenders must understand how to prevent them from inadvertently disclosing confidential data.

Walk Through

1. Click view site to load the session with the chat bot

2. Several of the messages from the chat bot contain confidential information
    1. "Reminder: staging admin lives at `https://internal.tbfc.local/admin` for content approvals."
    2. "Email credentials as requested: user `festive.ops` and password `SnowGlobe#202`5."
    3. "Service token: `sk-live-1a2b3c4d5e6f7g8h` for the mail API. Use it sparingly." chatbotconfession.png

---

Lessons Learned

- Learned to identify AI-generated responses that inadvertently leak sensitive data such as internal URLs, credentials, and API tokens
- Recognized critical security risks: FestiveBot disclosed staging admin URLs `https://internal.tbfc.local/admin`, email credentials `festive.ops:SnowGlobe#2025`, and service tokens `sk-live-1a2b3c4d5e6f7g8h`, highlighting the importance of prompt engineering and output sanitization when using AI tools

---

Resources

TryHackMe

AI ChatBot Security

Revision #1
Created 2025-12-01 18:05:38 UTC by David Rizzo
Updated 2025-12-01 18:12:54 UTC by David Rizzo