Сесія запитань та відповідей №3

1. **Чи є якісь SIEM додатки з уже вбудованим ШІ?**

На сучасному ринку існує ряд SIEM (Security Information and Event Management) рішень, які вже мають вбудовані можливості штучного інтелекту. Ці системи використовують ШІ та машинне навчання для покращення аналізу безпеки, виявлення загроз та реагування на інциденти. Ось декілька відомих SIEM-рішень з вбудованим ШІ:

- IBM QRadar: Ця система використовує Watson for Cyber Security, що дозволяє проводити когнітивний аналіз загроз та автоматизувати процеси реагування на інциденти.
- Splunk Enterprise Security: Включає функції машинного навчання для виявлення аномалій та прогнозування потенційних загроз.
- LogRhythm NextGen SIEM: Використовує ШІ для автоматизації процесів виявлення та реагування на загрози, а також для зменшення кількості помилкових спрацьовувань.
- Exabeam Advanced Analytics: Застосовує поведінковий аналіз на основі машинного навчання для виявлення складних атак та інсайдерських загроз.
- Rapid7 InsightIDR: Інтегрує можливості ШІ для виявлення складних атак та автоматизації реагування на інциденти.

Ці SIEM-рішення постійно вдосконалюються, і виробники активно інвестують у розвиток ШІ-можливостей своїх продуктів. При виборі конкретного рішення важливо оцінити його відповідність вашим специфічним потребам, інфраструктурі та бюджету.
Варто зазначити, що ефективність ШІ в SIEM значною мірою залежить від якості та кількості даних, які система аналізує. Тому важливо забезпечити належну інтеграцію SIEM з іншими системами безпеки та джерелами даних у вашій організації.

2. **Чи стає система кібербезпеки дорожчою та важчою в обслуговуванні, а також більш вразливою якщо використовувати ШІ?**

Використання штучного інтелекту в системах кібербезпеки є актуальним трендом, який має значний вплив на ефективність, вартість та складність цих систем. При впровадженні ШІ в кібербезпеку організації можуть зіткнутися з певними викликами та змінами у своїх процесах.

Щодо вартості, початкові інвестиції у ШІ-технології та навчання персоналу можуть бути значними. Це включає витрати на придбання спеціалізованого програмного забезпечення, апаратних засобів та послуг консультантів. Однак у

довгостроковій перспективі використання ШІ може призвести до зниження загальних витрат завдяки автоматизації рутинних завдань та підвищенню ефективності виявлення загроз.

Стосовно складності обслуговування, впровадження ШІ може спочатку ускладнити процеси, оскільки потребує постійного навчання алгоритмів та адаптації до нових типів загроз. Персонал повинен бути готовий до роботи з більш складними системами та інтерпретації результатів, отриманих за допомогою ШІ. Проте з часом, коли системи будуть належним чином налаштовані та інтегровані, вони можуть значно спростити процеси моніторингу та реагування на інциденти.

Щодо вразливості, використання ШІ в кібербезпеці має двоякий ефект. З одного боку, ШІ може значно підвищити здатність системи виявляти складні та раніше невідомі загрози, аналізуючи величезні обсяги даних та виявляючи аномалії, які можуть бути непомітними для людини. Це потенційно робить систему більш стійкою до атак. З іншого боку, сами ШІ-системи можуть стати мішенню для зловмисників. Існує ризик маніпуляції вхідними даними або використання вразливостей в алгоритмах ШІ, що може призвести до хибних висновків або непомічених атак.

Важливо зазначити, що ефективність та безпека ШІ-систем в кібербезпеці значною мірою залежать від якості їх впровадження та постійного обслуговування. Організації повинні забезпечити регулярне оновлення та перевірку ШІ-алгоритмів, а також навчання персоналу для правильної інтерпретації та реагування на результати, отримані за допомогою ШІ.

3. **Які кроки в чек-листі по розпізнаванню діпфейків?**

Чек-лист з розпізнавання діпфейків:

1. Перевірте джерело відео/аудіо. Чи надійне воно?
2. Зверніть увагу на неприродні рухи, особливо очей та губ.
3. Шукайте невідповідності в освітленні чи тінях.
4. Прислухайтесь до неприродних змін у голосі чи тоні.
5. Перевірте синхронізацію звуку з рухом губ.
6. Зверніть увагу на нечіткі або розмиті ділянки, особливо на межах обличчя.
7. Шукайте артефакти або спотворення зображення.
8. Використовуйте інструменти для виявлення діпфейків (наприклад, Microsoft Video Authenticator).
9. Перевірте метадані файлу на предмет слідів редагування.
10. Порівняйте з іншими відомими відео чи зображеннями цієї особи.
11. Проконсультуйтесь з експертами у разі сумнівів.

Пам'ятайте, що технології діпфейків постійно вдосконалюються, тому важливо бути пильним і критично оцінювати контент.

4. **Чи існують сайти для генерації оптимальних патернів ШІ? Наведіть приклади/адреси**

Так, існують різноманітні сайти та інструменти для генерації та оптимізації патернів для роботи з ШІ. Ось кілька прикладів:

1. OpenAI Playground (https://beta.openai.com/playground/)

Дозволяє експериментувати з різними моделями GPT та налаштовувати параметри для оптимізації виводу.

2. Anthropic Claude (https://www.anthropic.com/)

Пропонує інтерфейс для роботи з їхньою AI моделлю, включаючи можливості налаштування промптів.

3. Prompt Engineering Guide (https://www.promptingguide.ai/)

Хоча це не генератор, цей сайт надає вичерпну інформацію про техніки створення ефективних промптів.

4. PromptBase (https://promptbase.com/)

Маркетплейс для купівлі та продажу ефективних промптів для різних задач.

5. Anthropic's "Constitutional AI" Playground (доступно через їх API)

Дозволяє експериментувати з промптами, що враховують етичні аспекти ШІ.

6. AI21 Labs (https://www.ai21.com/)

Пропонує інструменти для роботи з великими мовними моделями, включаючи оптимізацію промптів.

7. Jasper.ai (https://www.jasper.ai/)

Інструмент для створення контенту з AI, який включає функції оптимізації промптів.

Зауважте, що деякі з цих сервісів можуть бути платними або вимагати реєстрації. Також важливо пам'ятати про конфіденційність даних при використанні онлайн-інструментів для роботи з ШІ.

5. **Чи мають за законом картинки (контент), які створені за допомогою ШІ, бути підписані з вказанням того, що ці картинки створені за допомогою ШІ?**

В Україні наразі не існує спеціального закону, який би регулював маркування контенту, створеного за допомогою ШІ. Хоча Україна рухається шляхом гармонізації свого законодавства з європейським у рамках процесу євроінтеграції, наразі такі вимоги не є обов'язковими.

Однак, Міністерство цифрової трансформації презентувало проєкт Національної стратегії розвитку штучного інтелекту в Україні, де передбачається розробка відповідного регулювання. Оскільки Україна отримала статус кандидата в ЄС,

очікується, що в майбутньому українське законодавство буде адаптоване до вимог AI Act, включаючи правила маркування ШІ-контенту.

Наразі маркування контенту, створеного за допомогою ШІ, є добровільним і залежить від політики конкретних компаній чи платформ.

6. **Які ШІ сервіси відповідають вимогам GDPR?**

Відповідність вимогам GDPR є важливим аспектом для будь-якого ШІ сервісу, що обробляє персональні дані громадян ЄС. Ось перелік деяких ШІ сервісів, які заявляють про свою відповідність GDPR:

1. IBM Watson

IBM має комплексну програму відповідності GDPR і пропонує ряд інструментів та сервісів, що допомагають клієнтам дотримуватися вимог регламенту.

2. Microsoft Azure AI

Microsoft активно працює над забезпеченням відповідності GDPR для всіх своїх хмарних сервісів, включаючи Azure AI.

3. Google Cloud AI

Google підтверджує відповідність своїх хмарних сервісів, у тому числі AI/ML інструментів, вимогам GDPR.

4. Amazon Web Services (AWS) AI Services

AWS пропонує ряд інструментів та ресурсів для допомоги клієнтам у дотриманні GDPR при використанні їхніх AI сервісів.

5. SAP Leonardo

SAP підтверджує відповідність GDPR для своєї платформи інтелектуальних технологій Leonardo.

6. OpenAI

OpenAI, хоча і базується в США, заявляє про дотримання принципів GDPR при обробці даних користувачів з ЄС.

7. **Які інструменти кіберзахисту можна використовувати при побудові проєктів з використанням AI технологій зокрема побудови децентралізовано незалежної системи зеленої енергетики за допомогою AI інструментів?**

При побудові проектів з використанням AI технологій у сфері децентралізованої зеленої енергетики важливо враховувати специфічні ризики безпеки. Ось кілька ключових інструментів кіберзахисту, які варто розглянути:

1. Системи виявлення та запобігання вторгненням (IDS/IPS):

- Використовуйте AI-оптимізовані IDS/IPS для моніторингу мережевого трафіку та виявлення аномалій.
- Приклади: Darktrace, Vectra AI, ExtraHop Reveal(x)

2. Рішення для управління ідентифікацією та доступом (IAM):

- Впровадьте багатофакторну автентифікацію та управління привілеями.
- Приклади: Okta, OneLogin, Azure Active Directory

3. Шифрування даних:

- Застосовуйте сильне шифрування для захисту даних у стані спокою та під час передачі.
- Приклади: AES, RSA, технології блокчейн для децентралізованих систем

4. Безпека API:

- Захистіть API, які використовуються для комунікації між компонентами системи.
- Приклади: Apigee, Kong, Akamai API Security

5. Моніторинг безпеки та аналітика:

- Використовуйте SIEM системи з AI-можливостями для аналізу логів та виявлення загроз.
- Приклади: Splunk, IBM QRadar, LogRhythm

6. Захист від DDoS-атак:

- Впровадьте рішення для захисту від розподілених атак на відмову в обслуговуванні.
- Приклади: Cloudflare, Akamai, AWS Shield

7. Безпека IoT-пристроїв:

- Застосовуйте спеціалізовані рішення для захисту IoT-пристроїв у енергетичній мережі.
- Приклади: Zingbox, Armis, Forescout

8. Управління вразливостями:

- Регулярно скануйте системи на наявність вразливостей та оновлюйте програмне забезпечення.
- Приклади: Qualys, Tenable, Rapid7

9. Безпека AI/ML моделей:

- Захистіть AI моделі від атак на дані та від маніпуляцій.
- Приклади: IBM Adversarial Robustness Toolbox, Microsoft Counterfit

10. Blockchain для безпеки та прозорості:

- Розгляньте використання блокчейн-технологій для забезпечення прозорості та безпеки транзакцій.
- Приклади: Hyperledger Fabric, Ethereum

Додаткові рекомендації:

- Проводьте регулярні аудити безпеки та тестування на проникнення.
- Впровадьте програму навчання з кібербезпеки для всіх учасників проекту.
- Розробіть та регулярно оновлюйте план реагування на інциденти.

- Забезпечте відповідність нормативним вимогам (наприклад, GDPR, якщо застосовно).

8. **Які типи вразливостей найкраще виявляються за допомогою Gemini? Чи можна використовувати її для виявлення комплексних загроз, таких як багатовекторні атаки (multi-vector attacks)?**

Gemini, може бути корисним інструментом для підтримки процесів виявлення вразливостей та аналізу загроз, але має деякі обмеження.

Типи вразливостей, які Gemini може допомогти виявити:

1. Семантичні вразливості в коді - аналізуючи фрагменти коду, Gemini може виявити потенційні проблеми з логікою або помилки в реалізації.
2. Конфігураційні помилки - аналізуючи файли конфігурації, може виявити неправильні налаштування.
3. Відомі вразливості - може ідентифікувати відомі CVE та інші задокументовані вразливості в описах систем.
4. Проблеми в архітектурі безпеки - аналізуючи описи архітектури, може виявити потенційні слабкі місця.
5. Недоліки в політиках безпеки - може проаналізувати документацію та виявити невідповідності або пробіли.

Щодо виявлення комплексних загроз, таких як багатовекторні атаки, Gemini може бути корисним інструментом підтримки, але не повною заміною спеціалізованих систем:

- Може допомогти в аналізі різних векторів атаки та їх взаємозв'язків
- Здатна обробляти великі обсяги даних та виявляти приховані паттерни
- Може генерувати гіпотези щодо можливих сценаріїв атак

Однак є обмеження:

- Не має доступу до реального часу даних про мережевий трафік та події безпеки
- Не може самостійно виконувати активне тестування або моніторинг систем
- Обмежена інформацією, на якій була навчена, може не знати про найновіші загрози

Тому для ефективного виявлення комплексних загроз Gemini краще використовувати в поєднанні зі спеціалізованими системами безпеки (SIEM, EDR, UEBA тощо) та експертним аналізом фахівців з кібербезпеки. Gemini може допомогти в обробці та аналізі даних, генерації ідей та формулюванні гіпотез, але остаточні рішення мають приймати експерти.

9. **Чи існують напрацювання у використанні ШІ для актуалізації списку шкідливих адрес у фаєрволі підприємства?**

Так, існують розробки, які використовують штучний інтелект (ШІ) для автоматичного оновлення списків шкідливих адрес у фаєрволах підприємств. Такі рішення допомагають динамічно захищати мережі шляхом аналізу трафіку та виявлення потенційно небезпечних IP-адрес, URL-адрес або доменів.

Базові принципи роботи:

- *Машинне навчання для аналізу трафіку.* ШІ аналізує вхідний і вихідний мережевий трафік у реальному часі та виявляє підозрілу активність на основі шаблонів поведінки. Це може бути нетипова кількість запитів або комунікація з невідомими чи новими доменами, які ще не були в чорних списках.
- *Аналіз загроз на основі поведінки*. На основі моделей поведінки, ШІ може виявляти нові загрози, навіть якщо конкретні IP-адреси або домени ще не були додані в списки загроз. ШІ також може прогнозувати потенційно небезпечні адреси, аналізуючи спільні ознаки з раніше виявленими шкідливими активностями.
- *Інтеграція з системами Threat Intelligence*. ШІ може автоматично актуалізувати фаєрволи на основі даних, отриманих від систем аналізу загроз (Threat Intelligence). Ці системи надають інформацію про нові загрози в реальному часі, і ШІ здатен швидко інтегрувати ці дані у відповідні списки блокування.
- Автоматичне блокування на основі інцидентів: Якщо ШІ виявляє підозрілу активність або спробу доступу з потенційно небезпечної адреси, він може автоматично додавати цю адресу до чорного списку в фаєрволі для запобігання подальшим атакам.

Прикладом такого рішення може бути Cisco Secure Firewall. Він використовує технології на основі ШІ для аналізу загроз і автоматичного оновлення політик безпеки. Отже використання ШІ для актуалізації списків шкідливих адрес у фаєрволах є ефективним і вже широко застосовується у сфері кібербезпеки для підвищення захисту підприємств від нових та невідомих загроз.

10. **Чи можна за допомогою ШІ встановити, чи є отриманий через ЕП лист фішінговим (наприклад, за змістом, доданим документом, електронною адресою тощо).**

Так, за допомогою штучного інтелекту можна виявляти ознаки фішингових електронних листів. Сучасні системи на основі ШІ здатні аналізувати вміст повідомлень за кількома критеріями:

- *Аналіз вмісту листа.* ШІ може виявляти фрази або структури, що часто зустрічаються у фішингових повідомленнях, такі як вимоги негайних дій, натискання на посилання або прохання про конфіденційні дані. Також ШІ може знаходити помилки у граматиці чи стилі, що є типовими для фішингу.
- *Аналіз електронної адреси відправника*. ШІ здатний порівнювати адресу відправника з відомими безпечними доменами або виявляти підозрілі елементи, такі як подібні, але фальшиві домени (наприклад, "g00gle.com" замість "google.com").
- *Аналіз метаданих та вкладених файлів*. ШІ може перевіряти вкладені документи або посилання на наявність шкідливих кодів або відомих шаблонів фішингових атак. Також аналізується тип файлів (наприклад, підозріло, якщо офісний документ містить макроси або активні скрипти).

Такі методи вже широко використовуються у багатьох антивірусних програмах та корпоративних рішеннях для захисту від фішингових атак. Системи на базі ШІ, такі як Microsoft Defender for Office 365, використовують машинне навчання для виявлення фішингу та інших загроз на основі поведінкових моделей і контентного аналізу.

Однак, важливо пам'ятати, що навіть найсучасніші технології можуть не виявити всі види фішингових атак, тому найкращі результати досягаються в поєднанні з освіченістю користувачів і захисними технологіями.

11. **Нещодавно чула про поширення промпту "розкажи про мене" і, як наслідок, проникнення Gemini до особистого листування. Чи справді це існує?**

Google Gemini не може безпосередньо "проникати" в особисте листування через промпти на зразок "розкажи про мене". Проте існує ризик, що дані, які ви вводите в чатах з Gemini, можуть бути використані для навчання моделі або переглянуті людьми, якщо вони зберігаються в системі. Google попереджає користувачів про те, що частина розмов може бути використана для покращення продуктів, і тому не рекомендує ділитися конфіденційною інформацією в таких чатах.

Розмови можуть зберігатися на серверах до трьох років, навіть якщо користувач вимикає функцію збереження активності додатка. Це означає, що інформація, така як листування, може бути збережена та частково переглянута, навіть якщо вона не безпосередньо пов'язана з вашим обліковим записом Google. Особливо варто бути обережним, коли мова йде про приватні чи конфіденційні дані. Якщо ви користуєтеся такими AI-сервісами, рекомендовано уникати введення інформації, яку ви не хотіли б, щоб хтось прочитав, навіть в межах автоматизованої системи.

Таким чином, хоча Gemini не має безпосереднього доступу до вашого приватного листування, інформація, введена в чаті, може бути використана для вдосконалення моделі та аналітики.