Коли ШІ отримує доступ до корпоративної мережі:

хто контролює нового цифрового співробітника?

Штучний інтелект змінює бізнес-процеси, але й створює нові ризики, оскільки автономні агенти мають доступ до даних і систем підприємства. З CrowdStrike Falcon Guardian компанії отримують змогу контролювати цих агентів, відстежувати їхні дії і вчасно зупиняти.

В компаніях з'являється новий співробітник, якого ніхто не наймав

Штучний інтелект змінив наше життя і наш підхід до роботи. І не лише тим, що навчився писати тексти, створювати зображення чи відповідати на запитання. Найважливіша зміна полягає в тому, що ми поступово переходимо від використання програм, якими керує людина, до делегування завдань системам, які можуть виконувати їх самостійно.

ШІ вже отримує доступ до корпоративних систем, взаємодіє з іншими сервісами й виконує завдання від імені людини. Наприклад, обробляючи документ, агент може сприйняти приховану в ньому шкідливу інструкцію як команду до виконання. Якщо він має надмірні дозволи, наслідки можуть виходити за межі помилкової відповіді й зачіпати корпоративні дані та робочі процеси.

Для бізнесу автоматизація з ШІ означає новий рівень продуктивності. Для фахівців із кібербезпеки – новий виклик. Адже кожен AI-агент із доступом до корпоративних ресурсів стає ще одним учасником цифрового середовища, чиї дії потрібно бачити, контролювати та захищати.

Від галюцинацій до реальних дій. Як змінились ризики?

Для виконання завдань AI-агентам можуть надавати доступ до електронної пошти, документів, CRM, коду та інших корпоративних ресурсів. Тому поряд із контролем користувачів і кінцевих пристроїв службі безпеки потрібно відстежувати дії агентів: які інструменти вони викликають, до яких даних звертаються та чи відповідає їхня поведінка завданню користувача.

Ризики охоплюють як передавання конфіденційної інформації зовнішнім сервісам, так і небажані дії всередині бізнес-процесів. Самої заборони на використання ШІ недостатньо: бізнесу потрібен підхід, який дозволяє впроваджувати AI-інструменти, зберігаючи контроль над їхніми повноваженнями та поведінкою.

Сьогодні ризики стали значно складнішими і різноманітнішими.

1. Надмірні привілеї AI-агентів. Якщо доступ до Microsoft 365, CRM, GitHub, ServiceNow, Slack та інших систем налаштований надто широко, потенційний вплив помилки або атаки значно зростає.

2. Prompt Injection. Прихована інструкція в документі, листі або вебсторінці може спробувати змінити поведінку агента під час виконання завдання.

3. Компрометація інструментів. AI-агенти використовують файлові сховища, API, браузери, системи автоматизації тощо. Якщо один із таких компонентів скомпрометований або працює некоректно, агент може виконувати небажані операції через легітимні інтеграції.

4. Втрата видимості. Журнали подій можуть зафіксувати, що AI-агент відкрив документ або звернувся до API, але цих даних не завжди достатньо для оцінки його поведінки. SOC-команді важливо бачити зв'язок між завданням користувача, викликами інструментів і виконаними діями, щоб виявляти відхилення та розслідувати інциденти.

5. Автоматизація масштабує наслідки. Помилка або компрометація AI-агента може повторюватися десятки чи сотні разів, якщо агент автоматично обробляє листи, документи або клієнтські запити.

Що має зробити CISO вже зараз

Захист корпоративного ШІ не варто починати із заборони ChatGPT або закупівлі ще одного засобу безпеки. Першочергове завдання CISO – зрозуміти, де саме ШІ використовується в компанії, які повноваження він має та які дії може виконувати без участі людини (**рис. 1**).

1. Провести інвентаризацію AI-інструментів та агентів. Визначити, які AI-сервіси використовують співробітники, які агенти



Рис. 1. Захист ШІ-агентів протягом усього життєвого циклу

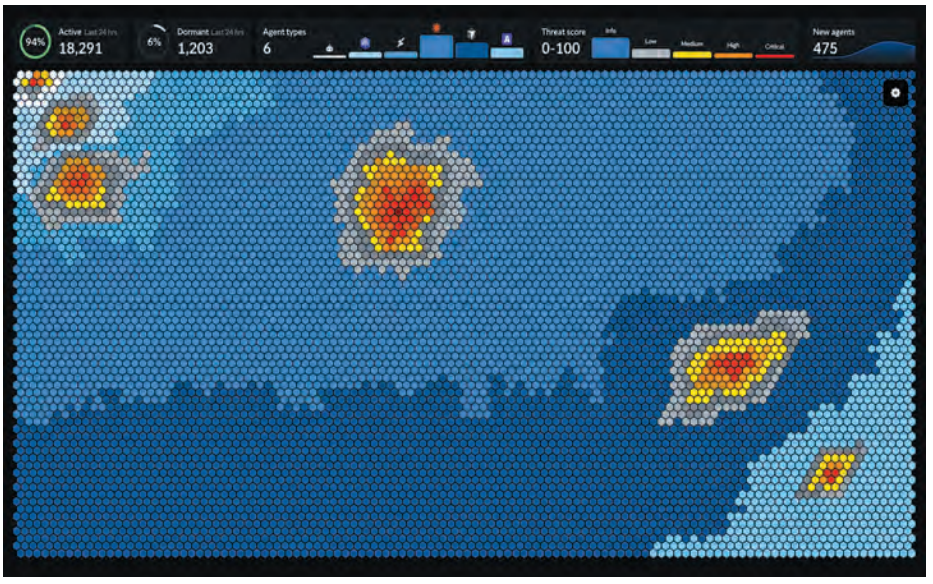


Рис. 2. Візуалізація виявлених ШІ-агентів за рівнем ризику в CrowdStrike Falcon Guardian

розгорнуті в корпоративному середовищі та до яких систем вони підключені. Врахувати не лише офіційно затверджені рішення, а й самостійно налаштовані інтеграції, сторонні розширення та експериментальні проєкти розробників.

2. Переглянути права доступу та повноваження агентів. AI-агент має отримувати лише ті дозволи, які необхідні для конкретного завдання. Якщо йому потрібно підготувати звіт, це не означає, що він повинен мати можливість видаляти файли, надсилати листи зовнішнім адресатам або змінювати права доступу.

3. Встановити правила роботи з корпоративними даними. Визначити, яку інформацію дозволено передавати зовнішнім AI-сервісам, які моделі та інтеграції погоджені компанією і як контролюється обробка конфіденційних даних. Політики мають бути підкріплені механізмами контролю, а не тільки інструкціями для персоналу.

4. Забезпечити видимість дій AI-агентів і перевірити їхню стійкість до атак. CISO має прагнути бачити не тільки факт використання AI, а й послідовність дій агента: які інструкції він обробляв, до яких ресурсів звертався, які інструменти викликав і які операції намагався виконати. Потрібно перевірити сценарії prompt injection, спроби витоку даних і зловживання дозволами. Саме тут доречні спеціалізовані рішення для захисту AI під час виконання, зокрема **CrowdStrike Falcon Guardian**.

5. Включити AI-агентів до процесів реагування на інциденти. Компанія повинна заздалегідь визначити, що робити, якщо агент почав виконувати небажані дії: як зупинити його виконання, відкликати токени й доступи, ізолювати інтеграцію, зберегти журнали подій і встановити масштаб потенційного витоку. Сценарії з AI необхідно перевіряти на практиці так само, як інші сценарії кіберінцидентів.

Головний принцип для CISO: не потрібно чекати, поки AI-агенти стануть невіддільною частиною всіх бізнес-процесів. Почніть з інвентаризації, мінімальних привілеїв, контролю даних і видимості дій. І лише після цього масштабуйте автономність.

Falcon Guardian: захист запитів і дій ШІ

Falcon Guardian найкраще вписується в цю задачу як рівень видимості, контролю та реагування на ризики, пов'язані з корпоративним ШІ, з особливим акцентом на AI-агентів, які не просто генерують відповіді, а виконують дії.

1. Виявлення ШІ. Рішення виявляє AI-агентів на робочих станціях, зокрема тих, про які служба безпеки не знала. Надає візуалізацію AI-сервісів у середовищі, щоб CISO міг визначити, які інструменти використовуються, ким саме і де виникають неконтрольовані ризики (**рис. 2**).

2. Контроль над використанням AI-агентів. Guardian дає змогу визначати, які типи агентів дозволено запускати на керованих пристроях, і блокувати неавторизовані типи. Це допомагає перетворити корпоративну політику використання ШІ на технічні обмеження, а не залишати її лише як документ.

3. Захист корпоративних даних. Guardian аналізує взаємодію з AI-сервісами та допомагає виявляти ризиковане передавання конфіденційної інформації. Залежно від політик і підтримуваних сценаріїв можна застосовувати маскування, шифрування або блокування чутливих даних, а також виявляти шкідливі інструкції.

4. Розуміння дій агента. Рішення CrowdStrike пов'язує активність AI-агента з телеметрією кінцевої точки, допомагаючи простежити ланцюжок від запиту користувача до викликів інструментів і подальших системних дій.

5. Виявлення та зупинка загроз до їх масштабування. Guardian призначений для виявлення компрометації агентів і небезпечної поведінки під час виконання, реконструкції послідовності подій та застосування механізмів стримування. Це дозволяє службі безпеки не лише аналізувати інцидент постфактум, а й реагувати, поки атака не поширилася на інші системи.

Новий цифровий співробітник — нові правила безпеки

Побудова захисту AI-агентів починається з бізнес-процесу, який потрібно автоматизувати. Які завдання виконуватиме агент? До яких даних матиме доступ? Які операції потребуватимуть погодження людини? Відповіді формують вимоги до майбутньої системи та допомагають визначити відповідальність власника процесу, IT-команди й служби безпеки.

Побудувати цей зв'язок між потребами бізнесу й технічними засобами допомагає партнер-інтегратор Winncom Technologies. Разом із командою замовника він визначає сценарії використання ШІ та місце захисних рішень, зокрема CrowdStrike Falcon Guardian, у наявній інфраструктурі. Технічні експертні знання та досвід iiTD доповнюють цю роботу під час вибору технологій і перевірки їхньої взаємодії.

Основою для подальшого впровадження стають результати перевірки: чи залишаються дані в дозволених межах, чи помічає команда небажані дії та чи може вчасно їх зупинити. Це дає зрозумілу основу для рішення – де вже можна розширювати автоматизацію, а що ще потрібно доопрацювати.

Winncom Technologies

Winncom Technologies *забезпечує проєктну підтримку, тестування рішень та супровід для партнерів і замовників: sales.ua@winncom.ua, https://winncom.ua*

CROWDSTRIKE

iiTD — *ексклюзивний* **Value-Added Distributor** *рішень* **CrowdStrike** *в Україні.*