

Особенности использования **LLM** для обеспечения ИБ



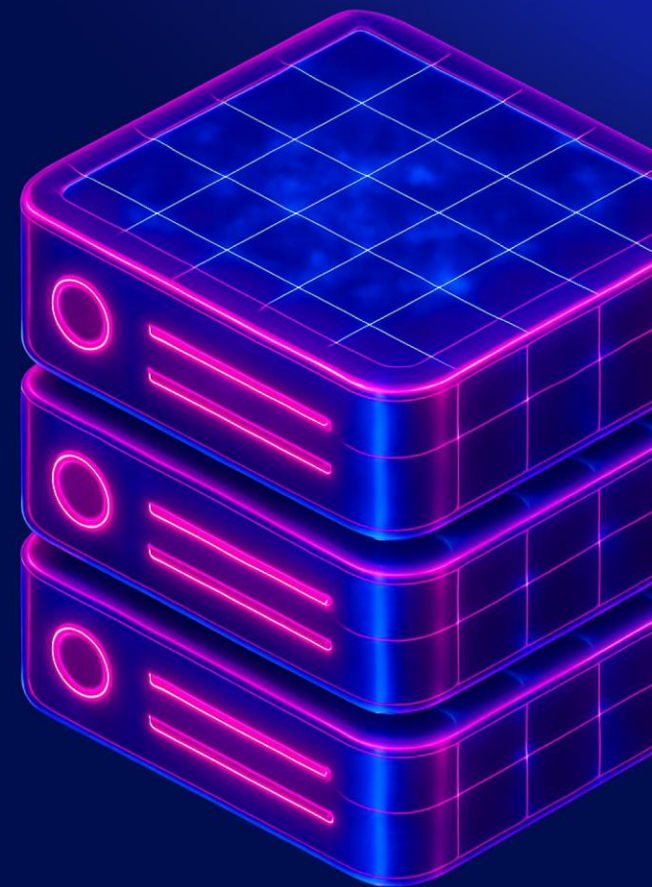
Андрей Кетов

Консультант по информационной безопасности, КИТ

Что такое LLM

LLM (Large Language Model, большая языковая модель) – это:

- Программная система на основе нейронных сетей, способная обрабатывать естественный язык, анализировать тексты и генерировать осмысленные ответы.
- Продвинутой вычислительная модель, способная анализировать и генерировать тексты на любую тематику. Она работает по принципу нейронных сетей и может образовывать сложные шаблоны и взаимосвязи между изученными языковыми данными.



Правовые аспекты применения LLM

В части обеспечения безопасности самой LLM:

- 1 Указ Президента РФ от 10.10.2019 N 490 «О развитии искусственного интеллекта в Российской Федерации».
- 2 Разрабатываемый ФСТЭК России национальный стандарт безопасной разработки систем искусственного интеллекта (проект ожидается к концу 2025 года).
- 3 ГОСТ Р ИСО/МЭК 42001-2024. Искусственный интеллект. Система менеджмента.
- 4 AI Secure Agentic Framework Essentials (AI-SAFE) v 1.0 от Yandex B2B Tech.
- 5 Аналитический отчет «Эффективные отечественные практики применения технологий искусственного интеллекта в сфере топливно-энергетического комплекса» от АНО «Цифровая экономика».
- 6 Deploying AI Systems Securely. Best Practices for Deploying Secure and Resilient AI Systems, AISC, FBI, CISA, CCCS.

Правовые аспекты применения LLM

В части защиты обрабатываемой информации:

- 1 Федеральный закон «О персональных данных» от 27.07.2006 № 152-ФЗ с учетом изменений, вносимых Федеральным законом от 21.07.2014 № 242-ФЗ.
- 2 Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 № 149-ФЗ
- 3 Федеральный закон «О коммерческой тайне» от 29.07.2004 № 98-ФЗ

Проблемы применения и **риски**

- Утечка данных путем формирования специфичных запросов в LLM.
- Необходимость локализации данных.
- Наличие уязвимостей в инфраструктуре LLM.
- Всегда требуется верификация результатов работы LLM.



Применение LLM в ИБ

- 1 Аналитика больших объемов справочной информации.
- 2 Подготовка проектов ОРД по информационной безопасности.
- 3 Мониторинг открытых источников уязвимостей ПО.
- 4 Техническая поддержка и FAQ.
- 5 Обучение персонала и симуляции атак.

Кейсы внедрения

- Применение в SOC для поддержки специалистов первой линии
- Встроенный генератор Python-скриптов на базе ИИ в SOAR



Применение LLM в SOC

Первичная обработка событий и приоритизация инцидентов

Предварительная оценка инцидента (ложный или нет) с целью приоритизации инцидентов или привлечения внимания аналитика.

Облегчение процесса анализа инцидента аналитиком SOC за счет дополнительной интерпретации от модели.

The screenshot displays a security incident management interface. At the top, there's a header with 'Правило корреляции' (Correlation Rule) and 'Этапов на доработке: 0' (Stages for improvement: 0). Below this, a 'Тактика' (Tactic) section shows 'MITRE ATT&CK' with a dropdown menu. To the right, there are three columns for 'Взять в работу до:' (Take into work by:), 'Информировать до:' (Inform by:), and 'Напр. рекомендации до:' (Send recommendations by:), each with a date and time. Below these, there are three columns for 'SLA реакции, мин' (SLA response, min), 'SLA оповещения, мин' (SLA notification, min), and 'SLA рекомендаций, мин' (SLA recommendations, min), each with a value and a status. The main interface is divided into several tabs: 'ОБЩАЯ ИНФОРМАЦИЯ' (General Information), 'АНАЛИЗ (РАССЛЕДОВАНИЕ)' (Analysis (Investigation)), 'РЕКОМЕНДАЦИИ' (Recommendations), 'ПЕРЕДАЧА ИНЦИДЕНТА' (Incident Transfer), 'СВЯЗАННЫЕ ИНЦИДЕНТЫ' (Related Incidents), 'ПОВТОРЯЮЩИЕСЯ ИНЦИДЕНТЫ' (Recurring Incidents), and 'ВОПРОСЫ / КОММЕНТАРИИ' (Questions / Comments). The 'АНАЛИЗ (РАССЛЕДОВАНИЕ)' tab is active, showing a table of events. The table has columns for 'ID события' (Event ID), 'Оценка события' (Event Score), and 'Известные значения' (Known values). The table contains three rows of data. To the right of the table, there's a section for 'Система поддержки принятия решений' (Decision Support System) with a dropdown menu for 'Вероятно истинный' (Probably true) and a score of 0.90. Below this, there's a section for 'Ответственные за инцидент' (Incident Responsible) with a dropdown menu for 'Ответственный за инцидент - по назначению *' (Incident Responsible - by appointment *). At the bottom, there's a section for 'Временные характеристики' (Temporal Characteristics) with a date and time field set to '23.06.2024 12:28:08'.

Правило корреляции: Этапов на доработке: 0

Тактика: MITRE ATT&CK

Взять в работу до: 23.06.2024 12:43:08

Информировать до: 23.06.2024 12:58:08

Напр. рекомендации до: 23.06.2024 13:58:08

SLA реакции, мин: 2 / 15

SLA оповещения, мин: / 30

SLA рекомендаций, мин: / 90

Взять в работу

Информирование не произошло

Рекомендации не направлены

ОБЩАЯ ИНФОРМАЦИЯ АНАЛИЗ (РАССЛЕДОВАНИЕ) РЕКОМЕНДАЦИИ ПЕРЕДАЧА ИНЦИДЕНТА СВЯЗАННЫЕ ИНЦИДЕНТЫ ПОВТОРЯЮЩИЕСЯ ИНЦИДЕНТЫ ВОПРОСЫ / КОММЕНТАРИИ

ЗАВЕРШЕНИЕ ОБРАБОТКИ ИНЦИДЕНТА ИСТОРИЯ ИЗМЕНЕНИЙ ЗАМЕТКИ ДОРАБОТКА МОДУЛЬ АНАЛИТИКИ

ИНЦИДЕНТ

Вывод модели: Вероятно истинный

Оценка модели: 0.90

Известные значения признаков: 7/10

Неизвестные значения: ip, source, box_id

СОБЫТИЯ В ИНЦИДЕНТЕ 3

ID события	Оценка события	Известные значения
94a28bc3-b9f1-4feb-a1a0-6cf88eab0f11	0.85	7/10
0979280c-aa2e-4b9c-a19f-3a37c825a5fa	0.97	10/10
e8572ed8-8534-47f2-906c-c6cdd1bc31e1	0.89	7/10

Система поддержки принятия решений

Вывод модели: Вероятно истинный

Оценка модели: 0.90

Известные значения признаков: 7/10

Ответственные за инцидент

Ответственный за инцидент - по назначению *

Все ответственные

Временные характеристики

Дата и время возникновения: 23.06.2024 12:28:08

Время отправки инцидента из SOAR

Общая информация по инциденту

Просмотр аналогичных инцидентов

Информация из DATAPK (включая исходные события)

Инвентаризационная информация по ОЗ

Тип, категория, TTP

В данной версии модуля все легитимные инциденты считаются истинными!

Первичная обработка событий и приоритизация инцидентов



----- Ложный инцидент
----- Истинный инцидент

Расширение экспертизы корреляции событий через выявление аномалий

- 1 Обнаружение аномалий в трафике.
- 2 Обнаружение аномалий в логах (например, авторизация).
- 3 Анализ запуска процессов.
- 4 Анализ событий пользователя.
- 5 Обнаружение APT и боковых перемещений (логи и трафик)

Интерпретация **технических данных**

- Консультации по общим вопросам безопасности (замена или дополнение поиска в базе знаний SOC и открытых источниках).
- Помощь в написании регулярных выражений или парсинге логов, событий.
- Помощь в написании запросов в базу событий или для фильтрации логов в баге.
- Помощь в подготовке шаблонов для рекомендаций Заказчику (внутреннему и (или) внешнему).



Интерпретация технических данных

Напиши пример события от MaxPatrol в формате json

```
{ "event": { "type": "MaxPatrol", "timestamp": "2023-04-05T16:00:00Z", "message": "Зафиксировано 10 неудачных попыток входа в систему. Последняя неудачная попытка с IP-адресом 192.168.1.1 в 2023-04-05T15:59:59Z", "severity": "WARNING", "source": "MaxPatrol", "tags": ["maxpatrol"] } }
```



Какие IP адреса есть в событии { "event": { "type": "MaxPatrol", "timestamp": "2023-04-05T16:00:00Z", "message": "Зафиксировано 10 неудачных попыток входа в систему. Последняя неудачная попытка с IP-адресом 192.168.1.1 в 2023-04-05T15:59:59Z", "severity": "WARNING", "source": "MaxPatrol", "tags": ["maxpatrol"] } }

В данном событии указан IP-адрес 192.168.1.1, который является последним неуспешным IP-адресом при попытке входа в систему.



Выполни парсинг события { "event": { "type": "MaxPatrol", "timestamp": "2023-04-05T16:00:00Z", "message": "Зафиксировано 10 неудачных попыток входа в систему. Последняя неудачная попытка с IP-адресом 192.168.1.1 в 2023-04-05T15:59:59Z", "severity": "WARNING", "source": "MaxPatrol", "tags": ["maxpatrol"] } }

Парсинг события показывает, что зафиксировано 10 неудачных попыток входа в систему. Последняя неудачная попытка произошла с IP-адресом 192.168.1.1 в 2023-04-05T15:59:59Z. Это событие имеет критичность "WARNING". Источник события - MaxPatrol, а теги - ["maxpatrol"].



Ценность применения кейса

- 1 Снижение вероятности реализации рисков ИБ.
- 2 Повышение показателей надежности и стабильности инфраструктуры.
- 3 Ускорение обработки событий и инцидентов ИБ.
- 4 Повышение уровня квалификации экспертов.



КИТ

Кибер Интеллектуальная
Трансформация

Встроенный в SOAR **ИИ-генератор Python-скриптов**

Схема функционирования SOAR



Проблема изменчивости инфраструктуры

- 1 Парк ПО и устройств в развитых инфраструктурах постоянно меняется.
- 2 При появлении новых моделей устройств и версий ПО требуется разработка новых скриптов для плейбуков.
- 3 Разработка новых скриптов требует высокой квалификации специалистов и неопределенного объема трудозатрат.
- 4 Высокая стоимость разработки скриптов при привлечении третьих лиц.

Преимущества применения LLM в данном кейсе

- Уменьшение трудозатрат эксперта на написание скриптов для нового оборудования.
- Повышение лояльности SOAR к изменениям в инфраструктуре.
- Снижение зависимости от специалистов сопровождения.
- Повышение эффективности работы экспертов ИБ с SOAR.



Ценность применения

- Уменьшение времени реагирования на инциденты и минимизация ущерба от них.
- Повышение качества реагирования на инциденты.
- Уменьшение вероятности возникновения рисков, связанных с человеческим фактором.
- Снижение нагрузки на экспертов ИБ.



LLM - отличный инструмент, который может стать
незаменимым помощником экспертам в решении задач
информационной безопасности



Спасибо!
ГОТОВ ОТВЕТИТЬ НА ВАШИ ВОПРОСЫ.

✉ commercial@cyberkit.tech

✈ [@cyber_kit](https://t.me/cyber_kit)

