УДК 004.8:004.056

ПРИМЕНЕНИЕ ЛОКАЛЬНОЙ МОДЕЛИ ДЛЯ ДЕТЕКТИРОВАНИЯ АНОМАЛИЙ В ЖУРНАЛАХ СОБЫТИЙ СИСТЕМЫ

М.А. СОЛОНОВИЧ, Е.И. ЮНЧИЦ, А.А. ИГНАТЕНКО

Белорусский государственный университет информатики и радиоэлектроники
(г. Минск, Беларусь)

E-mail:maxnest77max@icloud.com

Аннотация. Проведено исследование в целях улучшения результатов готовой модели искусственного интелекта. Рассмотрено применение локальной модели на примере выявления аномалий в журналах событий системы. В работе описывается процесс обучения готовой модели и сравнение с ее результатами до обучения.

Abstract. A study was conducted to improve the performance of a pre-existing artificial intelligence model. The application of a local model for such purposes is examined, using the example of detecting anomalies in system event logs. The paper describes the training process of the pre-existing model and compares it with its pre-training results.

Введение

Искусственный интеллект стал неотъемлемым помощником в задачах любого плана. Его применяют абсолютно везде, использование известных сервисов с искусственным интеллектом ставит под угрозу свои данные. Существует так же способ использовать искусственный интеллект локально у себя на машине для этого используют уже готовые и обученные модели.

Современные системы мониторинга информационной безопасности собирают огромные количество событий безопастности системы с ростом их числа специалисту информационнной безопастности становится все сложнее и сложнее анализировать эти данные. Для этого можно использовать искусственный интеллект, который способен находить по признакам атаки. Использование сервисов, предоставляющих доступ к актуальным мощным моделям, ставит под угрозу конфиденциальность. Для этого необходимо использовать локальные легкие модели. Для исследования взяли Qwen 2.5 на 3 миллиарда параметров. Легкая модель проста в использовании и адаптации дообучением под определеные задачи [1].

Теоретическое обоснование выбора модели и метода обучения

Большие языковые модели (далее LLM) – это тип модели глубокого обучения, обучаемая на массовых текстовых наборах данных для понимания и генерации человеческого текста. Такие модели могут выполнять различные задачи, они заранеее предобучены на больших обьемах данных. Изначально для работы таких моделей требовались значительные серверные мощности. Однако со временем крупные технологические компании и независимые исследователи начали публиковать архитектуры и веса моделей в открытом доступе.

Важным этапом развития стала разработка компактных версий LLM с уменьшенным числом параметров. Благодаря алгоритмической оптимизации эти модели стали пригодны для локального развертывания и инференса на потребительских персональных компьютерах, что существенно снизило порог входа для их использования [3].

Локальное развертывание компактных языковых моделей обеспечивает три ключевых преимущества: абсолютную конфиденциальность, вычислительную автономность и высокую эффективность. Выполнение инференса внутри изолированного ИТ-контура пользователя исключает передачу данных через сторонние API, что позволяет безопасно обрабатывать сенситивную, медицинскую и корпоративную информацию без риска ее компрометации. Кроме того, такие системы не зависят от интернет-соединения и лицензионных ограничений облачных провайдеров. При этом, несмотря на существенно меньшее количество параметров, современные локальные модели способны демонстрировать производительность, сопоставимую с крупномасштабными аналогами. Данная эффективность достигается за счет смещения фокуса с масштабирования архитектуры на использование тщательно отфильтрованных, высококачественных обучающих выборок, что многократно повышает внутреннюю «плотность знаний» нейросети и компенсирует ограничения вычислительных ресурсов.

Сами по себе локальные модели не могут эфективно выполнять посталенные задачи. Для подготовки модели под конктреную задачу ее не обходимо дообучить. В данной работе для обучения модели под конкретную задачу распознавания аномалий в журналах системы виндовс эфективно будет применять дообучение с помошью Lora Qlora

Для подготовки модели к работе с подобными машинными данными необходимо проведение процедуры дообучения [4]. Традиционное дообучение нейросети с обновлением всех миллиардов параметров требует колоссальных вычислительных мощностей и больших объемов видеопамяти, что противоречит самой концепции использования легковесных локальных моделей на стандартном оборудовании [2].

В связи с этим в данной работе для адаптации модели к задаче распознавания аномалий в журналах Windows применяется метод параметрически эффективного дообучения (PEFT – Parameter-Efficient Fine-Tuning), в частности технологии LoRA (Low-Rank Adaptation) и ее модификация QLoRA (Quantized LoRA).

Метод LoRA позволяет «заморозить» основные веса предварительно обученной модели и интегрировать в архитектуру трансформера обучаемые матрицы низкого ранга. Это сокращает количество обновляемых параметров в тысячи раз. Применение QLoRA идет еще дальше: базовая модель квантуется до 4-битного представления, что радикально снижает требования к оперативной и видеопамяти при сохранении качества обучения. Таким образом, использование связки LoRA/QLoRA является наиболее эффективным и ресурсосберегающим подходом, позволяющим превратить компактную модель общего назначения в специализированный инструмент аналитика информационной безопасности [5].

Процесс обучения модели

Ключевым этапом адаптации любой языковой модели к узкоспециализированной задаче является формирование репрезентативного обучающего набора данных (датасета). В качестве основы для проведения исследования был выбран открытый проект OTRF Security Datasets (Mordor). Данный репозиторий содержит верифицированные наборы системных событий (в частности, телеметрию Sysmon и Windows Event Logs), сгенерированные в контролируемой среде в процессе симуляции реальных кибератак и аномального поведения операционной системы.

Исходные журналы событий представляют собой массивы данных, обладающие высокой информационной избыточностью и «шумом», что делает их непригодными для прямого использования в обучении. Для эффективного применения метода дообучения с учителем сырые логи были предварительно очищены, отфильтрованы и приведены к строгому машиночитаемому формату JSON; итоговый датасет включал 312 событий.

Для интеграции данных в модель был выбран стандартный формат диалоговых инструкций, состоящий из массива messages. Такой подход позволяет нейросети обучаться в парадигме «запрос – ответ». Каждый обучающий пример был структурирован по следующему шаблону:

Системный контекст ("role": "system"). Данный параметр задает глобальное поведение и специализацию модели. В рамках эксперимента модели был передан промпт: «You are a Cyber Security SOC Analyst. Analyze the following Windows Event Logs and identify if they represent a known attack tactic or technique». Это позволяет инициализировать контекст предметной области и настроить внутренние механизмы внимания трансформера на поиск угроз.

Входные данные ("role": "user"). В качестве пользовательского ввода выступает хронологически выстроенный массив связанных событий системы (например, логи создания процессов, изменения ключей реестра и сетевых подключений). Использование формата JSON для передачи массива логов (с выделением полей EventID, Image, TargetObject, CommandLine) обеспечивает сохранение структурных связей между событиями. Это критически важно для того, чтобы модель могла проводить корреляционный анализ разрозненных событий, а не анализировать их изолированно.

Эталонный ответ ("role": "assistant"). Выступает в роли целевой переменной, на которую ориентируется функция потерь при обновлении весов модели. Ответ был строго формализован и разделен на четыре аналитических компонента:

Verdict (Вердикт) – бинарная классификация наличия угрозы (например, anomaly).

Confidence (Уверенность) – численная оценка вероятности правильности вердикта (например, 0.99).

MITRE Tactic (Тактика) – этап жизненного цикла атаки (например, Collection), что соотносится с методологией MITRE ATT&CK.

MITRE Technique (Инструмент) – конкретный метод компрометации или используемый инструмент (например, logonpasswords, auditpol system user auditpolicy modification).

Description (Описание) – краткое обоснование вынесенного вердикта на основе предоставленных логов.

Recommendation (Рекомендация) – рекомендации по дальнейшему анализу или действиям для расследования инцидента (например, обзор изменений auditpol в фазе Defense Evasion).

Подобная архитектура обучающей выборки позволяет в процессе применения алгоритма QLoRA трансформировать локальную модель общего назначения в специализированную экспертную систему. Модель обучается не просто пересказывать текст, а проводить логический синтез: выявлять причинно-следственные связи в технической телеметрии Windows и формировать структурированный отчет об инциденте

информационной безопасности. На рисунке 1 представлена общая архитектура обучающего события с сокращенной полезной нагрузкой в пользовательском вводе

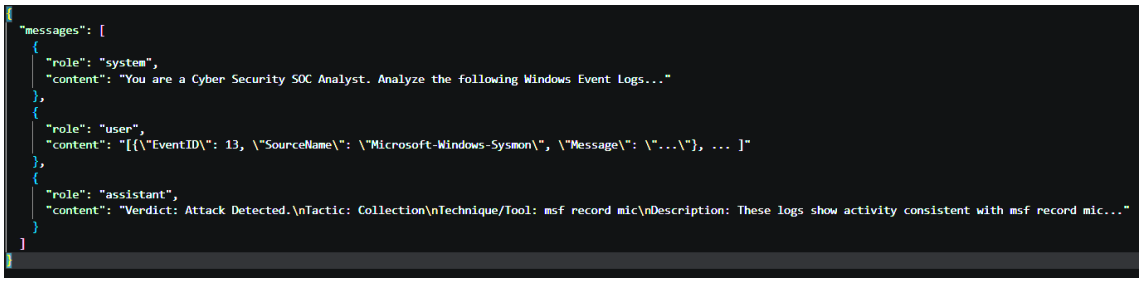

Рис. 1. Представление способов сканирования изображений

Для обучения модели Qwen 2.5 3B была разработана специализированная программная среда на языке Python с использованиемстека библиотек Hugging Face (transformers, peft, trl) и bitsandbytes. Учитывая специфику задачи – детекцию аномалий вWindows-логах – стандартный процесс дообучения был модифицирован для повышения точности генерации ответов.

Для эффективного использования видеопамяти применена технология 4-битного квантования типа NormalFloat4 с использованием двойного квантования. Это позволило загрузить веса модели в сжатом виде, используя float16 для вычислений, что критически важно для работы в условиях ограниченных ресурсов GPU. Для снижения нагрузки на VRAM также активирована технология Gradient Checkpointing, позволяющая пересчитывать градиенты во время обратного прохода вместо их хранения в памяти.

Особое внимание уделено алгоритму подготовки данных. В отличие от стандартных методов, в работе реализован класс AssistantOnlyCollator. Его ключевая особенность заключается в маскировании меток для сообщений пользователя. Таким образом, функция потерь вычисляется только на ответах модели, что заставляет нейросеть фокусироваться на изучении структуры аномалий, а не на повторении текста запроса.

Вместо настройки только проекций механизма внимания, в данном исследовании адаптеры были интегрированы во все линейные слои, включая MLP-блоки. Это обеспечивает более глубокую адаптацию модели к специфической лексике системных журналов.

Процесс обучения реализован с использованием оптимизатора paged_adamw_8bit, который эффективно управляет фрагментацией.

Результаты обучения

После дообучения локальной модели на подготовленном датасете, включающем 312 событий, была проведена оценка её способности выявлять аномалии в журналах Windows. Для взаимодействия с моделью и передачи ей логов был разработан простой CLI-скрипт, обеспечивающий удобный ввод данных и получение результатов. На начальном этапе исходная, недообученная модель обрабатывала события и в качестве результата предоставляла лишь сами логи без какой-либо аналитической обработки или классификации. Пример работы модели на этом этапе представлен на рисунке 2, где видно, что все события структурированы корректно, однако ни одно из них не идентифицировано как аномальное, а аналитическая информация отсутствует.

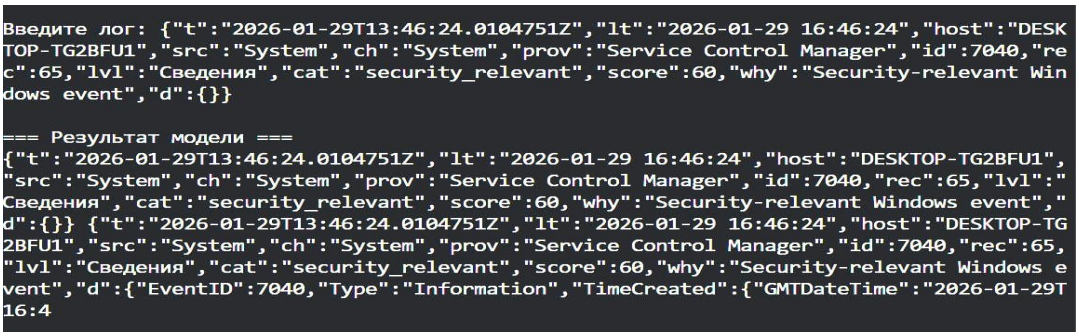

Рис. 2. Результат работы исходной модели

После дообучения модель показала заметное улучшение результатов. Она успешно обработала весь датасет из 312 событий, выявила критические аномалии и сформировала структурированные отчеты с детализацией по каждому событию: бинарный вердикт, тактика MITRE, конкретная техника или инструмент, описание инцидента и рекомендации по дальнейшему анализу. Для взаимодействия с дообученной моделью был

создан новый CLI-скрипт, который позволял удобно передавать ей лог-файлы и получать аналитические отчеты с подробной структурой. Пример работы дообученной модели представлен на рисунке 3, где модель обнаружила подозрительное событие, присвоила ему высокий уровень уверенности (0.85), правильно идентифицировала тактику атаки (Lateral Movement) и конкретную технику (logonpasswords). Кроме того, скрипт позволял сразу выводить рекомендации по дальнейшему расследованию, что облегчало работу аналитика и ускоряло процесс обработки большого объема журналов.

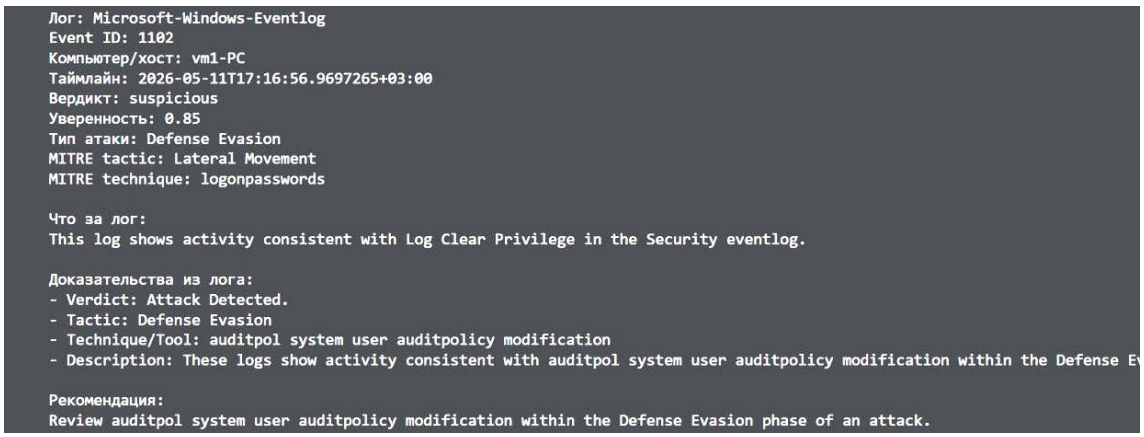

Рис. 3. Результат работы дообученной модели

После дообучения локальная модель была протестирована на датасете из 312 событий журналов Windows. Анализ показал, что исходная модель без дообучения лишь возвращала сами логи без какой-либо аналитики и классификации событий. В то же время дообученная модель демонстрирует способность выявлять аномальные действия, классифицировать их по тактикам и техникам MITRE, оценивать уровень уверенности и формировать развернутые рекомендации по каждому событию. Для удобного взаимодействия с моделью был создан CLI-скрипт, который обеспечивает оперативную передачу данных и вывод аналитических отчетов. Ниже представлена сравнительная таблица ключевых метрик для исходной и дообученной модели.

Таблица 1. Результат поиска экстремумов изображений с размером

Показатель	Исходная модель	Дообученная модель
Число выявленных аномалий	0	1
Средняя уверенность модели	0	0.85
Среднее время обработки одного события, с	0.24	0.26
Потребление памяти, МБ	7024	2500

Сравнительный анализ показал, что дообученная модель обеспечивает значительное улучшение качества обработки логов. Она способна не только выделять аномалии, но и предоставлять подробную информацию о типе атаки, используемой технике и рекомендациях по расследованию. Такой подход снижает нагрузку на специалистов по информационной безопасности и ускоряет процесс анализа больших объемов данных. Использование CLI-скрипта для взаимодействия с моделью позволяет оперативно получать структурированные результаты, что делает дообученную локальную модель эффективным инструментом для интеграции в систему мониторинга и анализа безопасности.

При этом применение QLoRA с 4-битным квантованием существенно сократило потребление видеопамяти: если исходная модель Qwen 2.5 на 3B параметров использовала около 7–8 ГБ VRAM, то дообученная версия потребляет лишь 2–2,5 ГБ VRAM, что обеспечивает возможность обработки модели на стандартных GPU с ограниченными ресурсами без потери точности распознавания аномалий. Полученные результаты подтверждают практическую ценность параметрически эффективного дообучения и демонстрируют возможность использования модели в реальных условиях для оперативного выявления и классификации угроз при экономии вычислительных ресурсов.

Заключение

В ходе проведенного исследования была рассмотрена возможность применения локальной модели искусственного интеллекта для распознавания аномалий в журналах событий системы Windows. Были проведены эксперименты с исходной моделью, которая без дообучения возвращала лишь структурированные логи, не обеспечивая аналитической обработки и выявления потенциальных угроз. Для взаимодействия с моделью был разработан CLI-скрипт, позволяющий передавать события и получать результаты в структурированном виде, что позволило наглядно оценить начальный уровень функциональности модели и её ограничения.

Дальнейшие эксперименты показали, что дообучение модели с использованием параметрически эффективного метода QLoRA и 4-битного квантования значительно улучшает её возможности. После дообучения модель успешно обрабатывала 312 событий, выявляла аномалии и формировала развернутые аналитические отчеты, включающие бинарный вердикт, уровень уверенности, тактику MITRE, используемую технику или инструмент, описание события и рекомендации для дальнейшего расследования. Новый CLI-скрипт позволял оперативно получать результаты, что облегчало интеграцию модели в процесс мониторинга информационной безопасности.

Сравнительный анализ показал, что дообученная модель значительно превосходит исходную по всем ключевым показателям: она способна выявлять скрытые угрозы, предоставлять структурированные отчеты с высокой детализацией и ускорять процесс анализа больших массивов данных. При этом экономия ресурсов также была значительной: за счет 4-битного квантования и параметрически эффективного дообучения потребление видеопамяти уменьшилось почти в три раза, что делает возможным использование модели на стандартных GPU с ограниченными ресурсами, сохраняя при этом точность распознавания аномалий.

Таким образом, результаты работы демонстрируют, что локальная дообученная модель искусственного интеллекта представляет собой эффективный инструмент аналитика информационной безопасности, способный автоматически идентифицировать аномалии, классифицировать события по тактикам и техникам MITRE, а также предоставлять рекомендации для последующего расследования. Эти результаты подтверждают практическую ценность параметрически эффективного дообучения и показывают возможность масштабирования подхода для работы с более крупными и разнообразными датасетами, обеспечивая оперативное выявление и классификацию угроз в реальных условиях.

Список использованных источников

1. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016.
2. Murphy K. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
3. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006.
4. Шелухин О. И. Обнаружение и диагностика аномальных состояний компьютерных систем средствами интеллектуального анализа данных системных журналов // Вопросы кибербезопасности. 2018. № 2(26). С. 33–43.
5. Атарская Е. А., Вульфин А. М., Кириллова А. Д. Система обнаружения аномалий в журналах регистрации событий средств защиты информации // Systems Engineering and Information Technologies. 2025. Т. 7, № 1(20). С. 3–12.