Общество с ограниченной ответственностью «Маинд Крафт»

ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК JAY GUARD

г. Санкт-Петербург

2024 г.

СОДЕРЖАНИЕ

1. Цели ПО 3
2. Задачи ПО 4
3. Сценарии использования ПО 5

Платформа Jay Guard (далее ПО) представляет собой специализированный программный комплекс, предназначенный для интеллектуального анализа и обработки информации, при взаимодействии с большими языковыми моделями (далее LLM, от англ. Large Language Model) посредством выявления и замены различных данных в передаваемой информации. Платформа предоставляет публичный API для интеграции ее функциональности в сторонние приложения, а также бесшовно интегрирована с продуктом Jay CoPilot и предоставляет пользовательский интерфейс для администрирования.

1. Цели ПО

- Обеспечение фильтрации передаваемых данных.

  Контроль содержимого информации при передаче в сервисы, использующие облачные LLM.
- Сохранение семантической целостности данных.

  Обеспечение сохранения семантической связанности и контекста диалога при обработке запросов к LLM, что позволяет гарантировать правильное понимание информации и исключить искажение смысла при обработке данных.
- Гибкость настройки фильтров данных.

  Предоставление пользователям возможности гибкой настройки правил фильтрации данных в соответствии с их специфическими требованиями.
- Прозрачность и контроль над обработкой данных.

  Обеспечение прозрачности процесса обработки данных путем ведения журнала сработавших правил фильтрации.
- Интеграция с продуктом Jay CoPilot.

  Обеспечение возможности интеграции с Jay CoPilot — ассистентом, который помогает использовать генеративный искусственный интеллект в работе и повседневных задачах. Обеспечение возможности фильтрации данных, передаваемых LLM в интерфейсе Jay CoPilot или по API, с заданным уровнем выявления.

- Интеграция с внешними системами.

  Обеспечение возможности интеграции с SIEM-системами и другими системами информационной безопасности, а также с любыми сторонними приложениями, для централизованного мониторинга и анализа событий, связанных с обработкой данных.

2. Задачи ПО

- Управление ключами доступа.

  Предоставление пользователям возможности создавать и удалять ключи доступа, а также привязывать к каждому ключу необходимый набор правил фильтрации данных и использовать этот ключ для фильтрации данных передаваемых через API или в интерфейсе Jay CoPilot.

- Настройка правил фильтрации данных.

  Осуществление проверки содержимого входящих запросов к LLM на согласно правилам и фильтрам, задаваемым администратором ПО. Выявление данных, попадающих под фильтры — в том числе с использованием ML-сервисов, функционирующих на базе технологий искусственного интеллекта (обработка естественного языка), например, slovnet, pymorphy-inflect, deeppavlov и др., — и последующая обработка (передача, модификация, блокировка) запросов, содержащих выявленные данные.

- Замена данных.

  Замена первоначальных данных в запросах к LLM на специальные переменные в соответствии с установленными пользователями правилами фильтрации, что позволяет обеспечивать безопасность данных без потери смысла запроса.

- Обратная замена данных.

  Восстановление первоначальных данных из ответов LLM, замена переменных на первоначальные значения, сохраняя семантический смысл ответа.

- Настройка правил фильтрации.

  Предоставление пользователям возможности установки и редактирования правил фильтрации данных в соответствии с их требованиями.
- Интерфейс добавления пользовательских сущностей.

  Предоставление пользователям возможности создавать свои собственные сущности, помимо предустановленных, для более точной настройки правил фильтрации данных.
- Интерфейс тестирования работы инструмента.

  Предоставление инструмента для тестирования работы фильтров, настроенных в ПО. Инструмент позволяет пользователям проверить, как Jay Guard обнаруживает и заменяет данные в тексте.
- Ведение журнала аудита.

  Фиксирование всех сработавших правил фильтрации и информации о запросах, которые были отфильтрованы или изменены.
- Предоставление данных аудита для интеграции с внешними системами.

  Предоставление данных аудита в формате json, для централизованного сбора и анализа событий.
- Предоставление публичного API.

  Обеспечение доступа к API для интеграции со сторонними приложениями, а также проксирования запросов к различным LLM, таким как GigaChat или YandexGPT.

1 Возможные сценарии использования ПО

- Выявление и фильтрация пользовательских данных таких как имена, телефоны, адреса, при использовании LLM в сфере обслуживания клиентов — недопущение передачи данных в облачные сервисы с заданным уровнем выявления.

- Фильтрация данных сотрудников, таких как email-адреса, IP-адреса, при использовании LLM во внутренних системах компании — недопущение передачи данных в облачные сервисы с заданным уровнем выявления.
- Анализ произвольных текстов для выявления и замены данных для последующего использования текстов в целях обучения ML-сервисов.

Программное обеспечение Jay Guard предлагает широкий спектр возможностей для обеспечения анализа и фильтрации данных при взаимодействии с LLM в различных сферах.

Функционал платформы Jay Guard может быть адаптирован под конкретные требования и задачи пользователей.