Модель для фильтрации коммерческой тайны

Н.А. Сурков, Д.В. Садова, А.М. Дмитриев

Филиал Самарского государственного технического университета, Сызрань, Россия

Обоснование. На сегодняшний день все большую актуальность набирает использование больших языковых моделей (преимущественно это модели от OpenAI и Google, в виде chatGpt и Gemini). В результате чего появляются такие риски, как отсутствие нормативных требований, этнические и социальные риски, а самое главное — это риск утечки конфиденциальной информации.

В условиях нарастающей важности информационной безопасности и соблюдения конфиденциальности данные проекты приобретают критическую значимость. Защита коммерческой тайны необходима для поддержания конкурентоспособности предприятий и предотвращения утечек информации, которые могут привести к значительным финансовым потерям и репутационному ущербу. Быстрое и точное распознавание таких данных является важной задачей для обеспечения качественной работы с информацией [1].

Актуальный на сегодняшний день вопрос почему бы не использовать локальные или отечественные языковые модели? Ответ прост, они не в полной мере проработаны, чтобы давать более точные ответы, а зачастую они галлюцинируют.

Актуальность выбранной темы заключается в том, что защита коммерческой тайны критически важна для предотвращения утечек, финансовых потерь, а также для исключения урона имиджу компании, а для повышения конкурентоспособности необходимо оперативное распознавание конфиденциальных данных.

Цель — создание эффективного инструмента для автоматического распознавания и фильтрации текстов, содержащих коммерческую тайну, что обеспечит безопасное использование систем искусственного интеллекта и защиту конфиденциальной информации в электронных коммуникациях.

Методы. Создание модели для фильтрации коммерческой тайны» подразумевает разработку высокопроизводительной и быстрой искусственной нейронной сети, предназначенной для определения наличия коммерческой тайны в текстах. Для создания модели были использованы передовые методы машинного обучения для анализа текстовых данных, с целью классификации и количественной оценки содержания конфиденциальной информации.

Результаты. Предлагаемая технология построена на принципе машинного обучения и классификации текста путем преобразования этого текста в набор векторных числовых представлений при помощи модели BERT.

Для тестирования работоспособности архитектуры были обучены три модели, отличающиеся размерами своих датасетов. Отличием также является внедрение в датасет большой модели строк, похожих на запросы к языковым моделям, которые могут содержать коммерческую тайну.

В результате выполнения научно-исследовательской работы были поставлены и выполнены следующие задачи:

1. Исследование рынка:
– проведение анализа текущего состояния рынка;
– изучение существующих решений и технологий;
– идентификация ключевых конкурентов.
2. Определение потребностей и болевых точек клиента:
– интервьюирование функциональных менеджеров.
3. Обоснование экономической целесообразности:
– оценка влияния существующих проблем качества на финансовые показатели компании (EBITDA);
– разработка экономической модели.
4. Произведен выбор инструментов разработки.
5. Выполнена разработка структуры системы обучения.
6. Проведено научное обоснование разработанной архитектуры.

7. Написаны инструменты создания данных.

8. Проведен этап первоначального тестирования на больших данных.

Выводы. Предлагаемое решение актуально и обладает высоким потенциалом развития, применения инновационных интеллектуальных процессов.

В целом, рынок защиты коммерческой тайны в облачных моделях имеет значительный потенциал для роста и привлекает внимание крупных компаний, что создает благоприятные условия для развития проекта.

Таким образом, результаты исследования подтверждают возможность эффективного применения предлагаемого решения.

Ключевые слова: контекстная фильтрация; информационная безопасность; конфиденциальная информация; коммерческая тайна; языковые модели; искусственный интеллект.

Список литературы

1. ec-rs.ru [Электронный ресурс] Информационная безопасность: Полное руководство. Режим доступа: https://www.ec-rs.ru/blog/informacionnaja-bezopasnost/informatsionnaya-bezopasnost-polnoe-rukovodstvo/ Дата обращения: 01.07.2024.

Сведения об авторах:

Никита Александрович Сурков — студент, группа ЭИЗ-21(с); Филиал Самарского государственного технического университета, Сызрань, Россия. E-mail: acinit2@yandex.ru

Дарья Владимировна Садова — студентка, группа ЭИ-22; Филиал Самарского государственного технического университета, Сызрань, Россия. E-mail: sadova_daria@mail.ru

Александр Максимович Дмитриев — студент, группа ЭИ-21; Филиал Самарского государственного технического университета, Сызрань, Россия. E-mail: shipuchka.ad@yandex.ru

Сведения о научном руководителе:

Кристина Владимировна Садова — старший преподаватель кафедры «Информатика и системы управления»; Филиал Самарского государственного технического университета, Сызрань, Россия. E-mail: crazyojj@mail.ru