Семинар "Педагогика ИИ"

Дата и время: 17.09.2025, 10.00

Место проведения: НИУ «МЭИ», г. Москва, ул. Красноказарменная, д. 13с3, аудитория М-101.

Участники: Павел Юрьевич Анучин (НИУ «МЭИ»), Владимир Владимирович Чистяков (НИУ «МЭИ»), Шамиль Алиевич Оцоков (НИУ «МЭИ»), Эдуард Артурович Челышев (НИУ «МЭИ»), Павел Романович Варшавский (НИУ «МЭИ»), Евгений Александрович Волошин (НИУ «МЭИ»), Дмитрий Жоржевич Корзун (ПГУ), Данияр Альбекович Альжанов («Диасофт»), Алексей Георгиевич Марахтанов (ПГУ), Мария Александровна Дурова (НИУ «МЭИ»).

Повестка:

1) Тема основной дискуссии «Применение искусственного интеллекта для решения практических задач в образовании и в сфере обеспечения информационной безопасности».

В рамках семинара были представлены два доклада:

- о создании учебного ИИ-ассистента для помощи преподавателям и студентам;

- об автоматическом обнаружении экстремистского контента с помощью ИИ.

2) Выбор тематики следующего семинара.

Резюме по первому докладу:

С докладом выступал Шамиль Алиевич Оцоков (НИУ «МЭИ»)

Первый доклад был посвящён разработке ассистента на базе искусственного интеллекта (ИИ), предназначенного для оказания поддержки преподавателям и студентам в учебном процессе. Основная концепция заключается в использовании ИИ в качестве посредника между преподавателями и студентами для автоматизации выполнения рутинных задач и предоставления оперативных ответов на вопросы студентов в круглосуточном режиме.

Сценарии функционирования ИИ-ассистента:

1. Автоматический сценарий:

- Вопрос от студента поступает через Telegram-бот.
- Модель ИИ обрабатывает запрос и формирует ответ.
- Ответ передаётся обратно в бота и отображается студенту.

2. Полуавтоматический сценарий с участием эксперта:

- Задача поступает в систему.
- Если задача не требует сложных аналитических действий, она решается моделью ИИ, а ответ проверяется и утверждается экспертом.
- В случае сложных задач решение принимается экспертом.

Данный подход особенно актуален для областей с высокими рисками ошибок, таких как медицина.

Варианты реализации:

1. Дообучение универсальной модели (fine-tuning):

- Преимущества: возможность адаптации модели под специфические задачи.
- Недостатки: высокие требования к вычислительным ресурсам, необходимость регулярного обновления лекционных материалов, низкая скорость обработки данных.

2. Модели с длинным контекстом:

- Преимущества: способность обрабатывать большие объёмы данных (тысячи и десятки тысяч слов).
- Недостатки: сложности в обработке больших объёмов документов, возможность предоставления неточных ответов из-за избыточности информации.

Выбранный подход: использование «наивного RAG» (Retrieval-Augmented Generation):

- Документы разделяются на фрагменты фиксированной длины (чанки).
- Каждый чанк преобразуется в векторное представление с использованием специализированных моделей.
- Векторные представления сохраняются в векторной базе данных.
- При поступлении запроса от пользователя он также преобразуется в векторную форму.
- Система осуществляет поиск наиболее релевантных чанков в базе данных.
- На основе найденных чанков формируется промт для языковой модели.
- Языковая модель генерирует ответ.

Используемые технологии и модели:

- API DeepSeek и ChatGPT для быстрой генерации ответов и интеграции лекционных материалов.
- Локальная модель DeepSeek с количеством параметров 1,78 миллиарда.

Проблемы и ограничения:

- В некоторых случаях система не может найти подходящий чанк, что приводит к отсутствию ответа.
- Необходимость дальнейшей оценки качества работы системы с использованием специализированных метрик.
- Перспективы развития включают внедрение продвинутого RAG, который будет учитывать структуру текста (например, разделение на абзацы) и сможет обрабатывать более сложные запросы.

Также были обсуждены возможности улучшения системы, включая переработку лекционного материала с применением разметки Markdown для более удобного представления формул и диаграмм.

Резюме по второму докладу:

С докладом выступал Шамиль Алиевич Оцоков (НИУ «МЭИ»)

Второй доклад был посвящён разработке системы автоматического обнаружения экстремистского контента с использованием технологий искусственного интеллекта.

Актуальность темы:

- В сети Интернет генерируется огромный объём информации, среди которой встречаются тексты, содержащие экстремистские призывы.
- Наблюдается рост количества экстремистского контента в мировом информационном пространстве.
- Существует необходимость в создании системы, способной выявлять и анализировать экстремистские материалы, а также выделять ключевые слова и фразы в текстах.

Цель работы:

Разработка прототипа системы, способной идентифицировать экстремистский контент и выделять соответствующие слова и фразы в текстах.

Сложности и ограничения:

- Отсутствие доступных открытых датасетов с экстремистским контентом из-за их конфиденциального характера.
- Этические и правовые риски, связанные с поиском, анализом и обработкой экстремистских материалов.

Методы и подходы:

- Разработан скрипт для поиска в социальной сети «ВКонтакте» текстов с экстремистскими высказываниями по ключевым словам.
- Часть данных была сгенерирована вручную.
- Использована универсальная модель BERT, дообученная на собранном датасете, что позволило достичь высокого качества обработки текстов.

Архитектура системы:

- Клиентская часть разработана с использованием Django.
- В качестве базы данных применён SQLite.
- Для обработки текстов используется модель BERT+PyTorch.

Результаты:

- Собрано 1154 примера текстов, связанных с религиозным экстремизмом.
- Проведено четыре эпохи обучения модели.
- Достигнута точность 97% при определении наличия экстремистского контента.
- Разработана система, которая не только выявляет экстремистский контент, но и выделяет ключевые слова и фразы с использованием подсветки по весам (чем выше вес, тем более явно слово связано с экстремизмом).

Дополнительные возможности системы:

- Аутентификация пользователей.
- Сохранение истории анализа текстов в базе данных для дальнейшего улучшения модели.

Перспективы и вопросы для дальнейших исследований:

- Возможность повышения точности модели при дополнительном обучении.
- Адаптивность системы к различным типам контента (не только к постам в социальных сетях).
- Необходимость проведения дополнительных экспериментов для оценки адаптивности и эффективности системы.

Тема доклада на следующий семинар:

«Архитектура интеллектуальных решений на стыке ASR, TTS и LLM для задач унифицированных коммуникаций»

Транскрипт

Павел Юрьевич Анучин:
Коллеги, добрый день! Предлагаю начинать. Напомню, что семинар проводится на регулярной основе, как в МЭИ, так и на дружественных площадках. Имеет открытый формат и носит следующие цели, такие как локализация в России инструментов и систем ИИ, обобщение опыта применения инструментов искусственного интеллекта и популяризации лучших практик решения прикладных и отраслевых задач с применением искусственного интеллекта. К задачам семинара относятся как выявление актуальных проблем сферы и путей их решения, так и формирование инженерной культуры обмена знаниями в надвузовском сообществе инженеров. По традиции предлагаю коллегам представиться и рассказать о том, кто чем занимается и где применяет искусственный интеллект, какие уже есть достижения. Начать предлагаю по часовой стрелке с присутствующих в офлайне и продолжить с коллегами в онлайне. Шамиль Алиевич, давайте начнем с вас.

Владимир Владимирович Чистяков:
Здравствуйте, коллеги! Я Оцоков Шамиль Алиевич, я работаю по направлению машинное обучение, искусственный интеллект. то есть преподаю курсы, связанные с машинным обучением, а также блокчейн и ряд других курсов. Меня интересуют направления, связанные с машинным обучением, в том числе такие темы, как обезличивание и ряд других. Я сегодня буду докладывать по теме АИ-ассистент, на пути создания АИ-ассистента для учебных курсов. для организации учебного процесса. В общем-то, все.

Павел Юрьевич Анучин:
Я Павел Юрьевич Анучин, ассистент кафедры радиотехнических систем Московского энергетического института. В данный момент веду несколько работ по нашей теме, в которых мы с командой используем как методы обучения, так и дообучение нейросетей под разные задачи. Мой интерес здесь - посмотреть, как коллеги применяют ИИ у себя, как используют инструменты и где подобные практики находят свою нишу в отрасли. Сегодня я выступаю еще и в качестве модератора семинара.

Владимир Владимирович Чистяков:
Здравствуйте, коллеги. Меня зовут Чистяков Владимир Владимирович. Я сотрудник кафедры РТС Московского энергетического университета. В целом мы с Павлом работаем в одной команде, поэтому цели у нас с ним одинаковые. Мы занимаемся внедрением и развитием искусственного интеллекта на нашей кафедре. И хотим с

помощью этого семинара посмотреть на работу коллег и, может, что-нибудь для себя интересное подчеркнуть из этого.

Эдуард Артурович Челышев:
Челышев Эдуард Артурович, ассистент кафедры вычислительных машин, систем и сетей, также Московский энергетический институт. Интересуюсь искусственным интеллектом в различных сферах его возможного применения.

Павел Юрьевич Анучин:
Предлагаю перейти к коллегам в онлайне. Павел Романович, давайте начнем с вас.

Павел Романович Варшавский:
Добрый день, коллеги! Меня зовут Павел Романович Варшавский, я заведующий кафедрой прикладной математики и искусственного интеллекта Московского энергетического института. Ну, соответственно, опыт работы в этой области у подразделения и тематика моих научных интересов, это как раз вот то, что связано с проблематикой искусственного интеллекта, интеллектуальный анализ данных, моделирование рассуждений, в том числе достоверных и правдоподобных и все, что касается этого. Ну и, скажем так, определенный бэкграунд у меня и у кафедры уже за не один десяток лет имеется. Скажем так, с советских времен кафедра занималась этими вопросами, но вот сегодня это действительно в тренде, поэтому здесь, готовы и поделиться нашими, разработками, ну и послушать, что коллеги предлагают по, соответственно, новым моделям и методам.

Павел Юрьевич Анучин:
Спасибо, Павел Романович. Евгений Александрович, давайте тогда дальше вы.

Евгений Александрович Волошин:
Добрый день, коллеги! Меня зовут Волошин Евгений, я сотрудник кафедры релейной защиты и автоматики. Наша область научных интересов заключается в прикладном применении технологий искусственного интеллекта для решения задач управления в электроэнергетике. В частности, это и синтез алгоритмов, и мультиагентные распределенные системы интеллектуального управления оборудованием и системами электроснабжения в целом. И все, что с этим связано. Ну, а также технологии доверенного искусственного интеллекта. Мы сейчас постепенно в эту тему тоже погружаемся.

Павел Юрьевич Анучин:
Спасибо. Дмитрий Жоржевич?

Дмитрий Жоржевич Корзун:
Добрый день, коллеги! Корзун Дмитрий, Петрозаводский госуниверситет. Я доцент кафедры информатики и математического обеспечения, а также научный

руководитель Центра искусственного интеллекта ПГУ. Мы занимаемся проектами, связанными с промышленным искусственным интеллектом для различных предприятий, для медицины, для сельского хозяйства, для робототехнических задач. Такого рода применение искусственного интеллекта нас интересует. Также начали заниматься вопросами доверенного искусственного интеллекта, поскольку многие наши заказчики и партнеры - у них есть закрытый контур безопасности, за который нельзя информацию выдавать.

Павел Юрьевич Анучин:
Спасибо. Данияр Альбекович?

Данияр Альбекович Альжанов:
Добрый день, коллеги! Я Альжанов Данияр Альбекович, представляю компанию «Диасофт», старший Python-разработчик. Именно искусственным интеллектом начал заниматься недавно, начал изучать тему глубокого обучения. Поэтому пока что в качестве научного интереса присутствую, но и в дальнейшем планирую переквалификацию в сферу глубокого обучения.

Павел Юрьевич Анучин:
Спасибо. Алексей Георгиевич?

Алексей Георгиевич Марахтанов:
Марахтанов Алексей Георгиевич, директор Центра искусственного интеллекта Петрозаводского государственного университета. Вот мы коллеги с Дмитрием Жоржевичем Корзуном. В целом он рассказал про наши направления. Это использование искусственного интеллекта в промышленных задачах, активно сотрудничаем с заводами, сотрудничаем с предприятиями Росатома. Из таких, может быть, еще направлений, которые не озвучил Дмитрий Жоржевич, используем искусственный интеллект в аквакультуре. То есть задача промышленного выращивания рыбы. Реализуем методы подводной видеоаналитики, методы прогнозирования на основе искусственного интеллекта. Поэтому тема данная очень интересна и к семинару подключился, чтобы что-то новое узнать, обменяться опытом.

Павел Юрьевич Анучин:
Теперь предлагаю перейти к докладам. У нас на сегодня запланировано два доклада. Первый - на тему «На пути создания учебного ассистента». Ее доложит Шамиль Алиевич Оцоков. Вторая тема «Автоматическое обнаружение экстремистского контента на основе технологии искусственного интеллекта». К сожалению, докладчик второй темы не смогла прийти по причине болезни, доклад сделает Шамиль Алиевич.

Шамиль Алиевич Оцоков:
Тема моего первого доклада связана с актуальным направлением создания учебного ИИ-ассистента для помощи преподавателям и студентам в обучении. В чем суть этой

работы и как здесь используется искусственный интеллект? Искусственный интеллект будет являться посредником между преподавателем и студентом, потому что зачастую преподаватель не имеет возможности, чтобы объяснить каждому студенту, разложить все знания по полочкам, объяснить тему очень подробно. И здесь появляется вот такой ИИ посредник в качестве квалифицированного специалиста, который может ответить на вопросы в любое время. Идея была такая, и здесь я рассматривал разные варианты вместе со студентами.

Шамиль Алиевич Оцоков:
Ранее я отмечал, что рутинные операции, связанные с типовыми вопросами студентов, можно автоматизировать с помощью Telegram-бота и искусственного интеллекта. В этом случае система помогает преподавателю, отвечая на стандартные запросы и тем самым снижая его нагрузку.
На следующем слайде представлены сценарии работы AI-ассистента. Первый сценарий — автоматический: студент задаёт вопрос через Telegram-бот, далее запрос поступает в языковую модель искусственного интеллекта, которая формирует ответ и возвращает его в бот. Второй сценарий — полуавтоматический, когда в цепочку обработки включается человек-эксперт. AI-ассистент в этом случае решает задачу совместно с экспертом: если система определяет, что запрос несложный, он обрабатывается моделью; если же требуется более глубокое понимание или цена ошибки высока, решение принимает эксперт. Финальный ответ проходит проверку специалиста и затем направляется пользователю.

Шамиль Алиевич Оцоков:
Существуют области, где цена ошибки особенно велика — например, медицина. В таких случаях предпочтителен именно полуавтоматический режим работы, когда человек контролирует результат. Мы рассматривали подобные сценарии при разработке AI-ассистента и планировали подключать к процессу студентов, изучающих машинное обучение. Один из возможных вариантов развития — дообучение модели (fine-tuning). Мы используем универсальную языковую модель, содержащую знания по различным темам, включая блокчейн. Я разрабатываю ассистента, который консультирует студентов по курсу «Введение в блокчейн». Мы пытались адаптировать модель на основе собственных конспектов лекций — у меня их десять по данному курсу. Рассматривали возможность fine-tuning, но отказались: для этого нужны мощные вычислительные ресурсы, а учебные материалы регулярно обновляются, особенно в такой динамичной области, как блокчейн.

Шамиль Алиевич Оцоков:
В этой дисциплине постоянно появляются новые технологии, о которых важно рассказывать студентам, поэтому повторное обучение модели становится неэффективным. Кроме того, fine-tuning занимает слишком много времени: студент не может ждать несколько минут, пока система сформулирует ответ. Мы рассматривали также использование моделей с длинным контекстом, которые могут анализировать

документы объемом в десятки тысяч слов, но в нашем случае этот подход оказался непрактичным: слишком большой объем лекционных материалов. При использовании одной модели с длинным контекстом ответы становятся слишком общими, так как модель теряет фокус на конкретном запросе. Поэтому мы остановились на более простом, но эффективном решении — использовании системы типа RAG, а точнее, её базовой версии — «наивного RAG».

Шамиль Алиевич Оцоков:
Мы изучали публикации в этой области: научных работ немного, в основном встречаются технические описания реализации RAG-систем. Существует множество их разновидностей — от наивных до продвинутых и графовых. На текущем этапе мы применили простейший вариант. У нас есть множество документов, которые разбиваются на фрагменты фиксированной длины — чанки. В нашем случае получилось 35 чанков. Каждый из них векторизуется с помощью специализированных моделей, а полученные векторы сохраняются в векторной базе данных. Такая база хранит пары «ключ–значение»: ключом является документ, а значением — его векторное представление. Далее, когда студент задаёт вопрос через Telegram-бот, запрос также преобразуется в вектор с помощью embedding-модели.

Шамиль Алиевич Оцоков:
После этого система ищет наиболее релевантные фрагменты в векторной базе. Найденные чанки используются для формирования промпта к языковой модели. Модель при этом не дообучается, а лишь обрабатывает заданный контекст. Мы тестировали несколько языковых моделей. Первый вариант — использование API, совместимого с OpenAI, через DeepSeek. Такой клиент позволял без изменений кода переключаться между DeepSeek и ChatGPT. Эта модель показала хорошее качество ответов: они были не общими, а опирались на лекционный материал. Именно это отличает наш подход от обычного использования ChatGPT, где ответы формируются без учета конкретных учебных данных.

Шамиль Алиевич Оцоков:
Второй вариант — локальная модель DeepSeek с числом параметров 1,78 миллиарда. Она обладает базовыми знаниями, в том числе в области блокчейна, и умеет рассуждать. Мы запускали её на достаточно мощном компьютере, но время отклика составляло несколько минут, а качество ответов было ниже, чем у облачной версии. Поэтому локальный вариант мы сочли непрактичным. Однако если работать с конфиденциальными данными, использовать локальную модель всё же необходимо. Так как в нашем случае лекции не содержат закрытой информации, мы остановились на первом варианте — облачной модели: она быстрее и даёт более точные ответы, хотя и требует небольшой оплаты за запрос.

Шамиль Алиевич Оцоков:

Стоимость обращений к API невысока, поэтому такой подход экономически оправдан. Тем не менее мы продолжаем исследования и рассматриваем другие варианты — возможно, найдём локальные модели, которые обеспечат сопоставимое качество и скорость при меньших затратах. Планируем проводить дополнительные эксперименты с использованием более мощных графических ускорителей. Если у коллег есть вопросы, можно задавать прямо по ходу обсуждения.

Владимир Владимирович Чистяков:
Интересно, вы упомянули модель DeepSeek 1.5B, которая показала слабые результаты. Пробовали ли вы использовать более мощные модели, или пока нет?

Шамиль Алиевич Оцоков:
Пока нет. Даже при текущем размере модели время отклика составляет несколько минут, а при увеличении числа параметров оно только возрастёт. Я тестировал локальный вариант на достаточно производительном ноутбуке, но всё равно ответ занимал минуты. А при использовании облачной модели результат приходит за несколько секунд — это гораздо комфортнее для студентов.

Мария Александровна Дурова:
Подскажите, пожалуйста, какую векторную базу данных вы используете? И какой метод векторизации выбран? Проводили ли вы исследование, как способ векторизации влияет на качество поиска?

Шамиль Алиевич Оцоков:
Пока мы используем стандартный вариант из базовой реализации RAG-системы, без дополнительных модификаций. Специальных экспериментов по сравнению методов векторизации мы пока не проводили, но планируем это сделать со студентами — изучить влияние алгоритма векторизации и способов поиска на качество ответов.

Владимир Владимирович Чистяков:
На одном из слайдов вы показывали, что документы делятся на чанки и переводятся в вектора. Используете ли вы промежуточное хранение, например в формате JSON, или всё происходит сразу в векторном представлении?

Шамиль Алиевич Оцоков:
Механизм работает автоматически: документы делятся на чанки фиксированной длины (у нас их 35), и каждый фрагмент преобразуется в вектор. Явного промежуточного шага в виде JSON у нас нет — лекции изначально были в формате DOC, мы конвертировали их в текст, объединили и передали скрипту, который выполняет разделение и векторизацию.

Владимир Владимирович Чистяков:
И в целом система хорошо работает с RAG? Ответы корректные?

Шамиль Алиевич Оцоков:
В большинстве случаев да, но бывают запросы, на которые система не может ответить — подходящего фрагмента просто не находится. Мы думаем, как улучшить этот механизм: возможно, если релевантного материала нет, стоит позволить модели отвечать на основе собственных знаний. Мы пока не проводили количественную оценку качества — метрики RAG-систем ещё не применяли, экспериментов немного. В дальнейшем планируем перейти к более продвинутым версиям RAG.

Владимир Владимирович Чистяков:
Да, метрики требуют значительных трудозатрат, особенно при ручной подготовке тестов.

Шамиль Алиевич Оцоков:
Согласен.

Мария Александровна Дурова:
У меня ещё технический вопрос. В курсе по блокчейну формул немного, но если перенести ассистента на предметы, где формул много, как можно с ними работать? Есть ли идеи, как обрабатывать формулы — например, через LaTeX?

Шамиль Алиевич Оцоков:
Да, это интересный вопрос. Вероятно, для этого потребуется мультимодальная модель, способная анализировать и текст, и изображения, ведь студенту важно видеть формулу в привычном виде. Пока мы ограничились только текстовыми материалами, но в будущем этот вопрос нужно будет проработать.

Мария Александровна Дурова:
Да, сначала нужно отладить текстовую часть. Возможно, чтобы не нагружать систему, формулы стоит хранить в виде LaTeX-кода и просто красиво визуализировать их на стороне клиента.

Шамиль Алиевич Оцоков:
Согласен. Вопрос лишь в том, сможет ли языковая модель не только отобразить формулу, но и корректно рассуждать о ней — объяснять смысл, проводить доказательства. Мы используем облачную модель DeepSeek, и она уже обладает базовыми возможностями анализа формул, поэтому, думаю, мы сможем это реализовать.

Владимир Владимирович Чистяков:
А какую именно модель векторизации вы используете?

Шамиль Алиевич Оцоков:

Сейчас точно не скажу, нужно посмотреть в коде.

Владимир Владимирович Чистяков:
Хорошо, потом посмотрим, просто любопытно.

Шамиль Алиевич Оцоков:
Да, конечно, я поделюсь.

Евгений Александрович Волошин:
Хотел добавить по поводу формул и визуализации. Markdown поддерживает вставку кода, в том числе для построения диаграмм (например, Mermaid). Можно преобразовать часть лекционного материала в формат, где графические элементы описаны текстом, а на клиентской стороне они будут рендериться как изображения. Так, например, GigaChat корректно обрабатывает формулы в LaTeX и отображает их прямо в чате. Возможно, стоит рассмотреть подобный подход.

Шамиль Алиевич Оцоков:
Спасибо, это полезное замечание.

Павел Юрьевич Анучин:
Коллеги, продолжаем. Есть ли ещё вопросы по этой части? Если да, прошу задавать.

Шамиль Алиевич Оцоков:
На следующем слайде представлены примеры ответов ассистента на вопросы студентов. Видно, что система действительно использует загруженные лекционные материалы. Перспективы дальнейшей работы связаны с переходом на продвинутый RAG — где чанки имеют не фиксированную длину, а формируются по смысловым границам, например, по абзацам. Это позволит повышать релевантность извлекаемых фрагментов. Также планируется декомпозиция сложных вопросов на несколько простых и их последовательная обработка, а затем объединение ответов в единый результат. Такой подход описан в литературе как одно из направлений развития RAG-систем.

Шамиль Алиевич Оцоков:
Ещё одно направление — GraphRAG, где знания представляются в виде графа: вершины — сущности, а рёбра — связи между ними. Исследования в этой области активно ведутся. Один из моих студентов защитил магистерскую работу с использованием GraphRAG и сейчас работает в «Росатоме», где применяет этот подход на практике. Мы планируем пригласить его выступить на семинаре и поделиться опытом.

Павел Юрьевич Анучин:
По данной презентации есть ещё вопросы?

Евгений Александрович Волошин:
Да, хотел уточнить по поводу графов знаний. Мы в своих проектах используем графовое представление информации, но пока наполняем базы вручную, создавая онтологии предметных областей, которые затем заполняются фактами, извлечёнными из текста. В GraphRAG этот процесс происходит автоматически. Есть ли возможность увидеть, какие именно факты извлекла модель, чтобы убедиться, что она правильно поняла предметную область и не сделала ложных выводов?

Шамиль Алиевич Оцоков:
Пока мы не применяли GraphRAG на практике, поэтому ответить точно не могу. После проведения экспериментов смогу поделиться результатами.

Евгений Александрович Волошин:
Когда будете заниматься этим направлением, обратите внимание на визуализацию внутренних связей модели. Если будет возможность посмотреть, какие связи и факты она выделяет, это поможет не только валидации, но и интерпретации выводов. Мы сами будем рады использовать такие инструменты в прикладных задачах.

Павел Юрьевич Анучин:
Спасибо. Тогда предлагаю перейти к следующему докладу. Вам слово.

Шамиль Алиевич Оцоков:
Тема следующего выступления — «Распознавание экстремистского контента с помощью технологий искусственного интеллекта». Работа выполнена студенткой Пушининой Дианой Владимировной, однако она, к сожалению, приболела и не смогла присутствовать на семинаре, поэтому я представлю её исследование. Актуальность темы очевидна: в интернете ежедневно публикуется огромное количество информации, и среди этого потока периодически встречаются тексты, содержащие признаки экстремизма — призывы, лозунги и т.п. Согласно статистике Роскомнадзора, количество заблокированного экстремистского контента ежегодно растёт. Целью данной работы стало создание прототипа системы, способной автоматически распознавать экстремистский контент и выделять слова, указывающие на его наличие.

Шамиль Алиевич Оцоков:
Рост объёма подобного контента особенно заметен в веб-приложениях и социальных сетях. В рамках работы акцент был сделан на распознавании религиозного экстремизма. Мы не рассматривали политические, этнические и иные формы. Одной из основных трудностей стало отсутствие открытых датасетов, содержащих экстремистские тексты. Понятно, что такие данные не публикуются в свободном доступе, а их поиск связан с определёнными этическими и юридическими рисками. Поэтому было принято решение собрать собственный датасет.

Для этого был разработан специальный скрипт, который выполнял парсинг социальной сети «ВКонтакте» по ключевым словам, связанным с экстремистской тематикой. На основе результатов был сформирован корпус текстов, содержащих потенциально опасные фразы. Дополнительно часть текстов была сгенерирована вручную, чтобы сбалансировать датасет.

Шамиль Алиевич Оцоков:
Поскольку доступ к запрещённым источникам невозможен, все примеры создавались безопасным способом — искусственно. Далее была разработана информационная система с клиентской и серверной частями. Серверная часть написана на Python с использованием фреймворка Django, база данных — SQLite. В качестве модели применялась языковая модель BERT, которая продемонстрировала высокое качество обработки текстов. Модель была дообучена на нашем собранном датасете. Процесс выглядел следующим образом: выполнялся парсинг сообщений «ВКонтакте» по тематическим ключевым словам, связанным, например, с терроризмом. Датасет включал как экстремистские, так и нейтральные тексты (примерно 50 на 50). Всего собрано 1154 примера по религиозной тематике. Обучение проводилось в течение четырёх эпох, и по мере обучения наблюдалось снижение функции потерь.

Шамиль Алиевич Оцоков:
Результаты работы модели представлены в виде матрицы ошибок. Доля правильных классификаций составила 97%. Также были рассчитаны метрики точности, полноты и F1-скор. Архитектура системы включает несколько модулей: аутентификацию пользователя, ввод текста для анализа, обработку текста языковой моделью и визуализацию результата. Пользователь вводит текст, система анализирует его и сообщает, содержит ли он экстремистский контент.
Дополнительно реализована функция подсветки слов, которые наиболее вероятно относятся к экстремистской лексике. Слова с наибольшим весом выделяются красным цветом, нейтральные остаются серыми. Таким образом, можно визуально увидеть, какие фразы система определила как опасные.

Шамиль Алиевич Оцоков:
История анализов сохраняется в базе данных, что позволяет пополнять обучающую выборку и повторно обучать модель на новых данных. На данный момент это прототип, целью которого было проверить возможность автоматического выявления экстремистского контента. Эксперименты показали, что система действительно способна выполнять такую задачу, что делает направление перспективным. У меня всё, спасибо.

Павел Юрьевич Анучин:
Спасибо, Шамиль Алиевич. Коллеги, есть вопросы?

Владимир Владимирович Чистяков:

Да, есть вопрос. На графике видно, что после четырёх эпох обучение ещё не полностью вышло на плато. Возможно, стоило добавить ещё одну-две эпохи — может быть, точность поднялась бы до 98%. Пробовали ли вы увеличивать число эпох, или модель начинала переобучаться?

Шамиль Алиевич Оцоков:
Пробовали, но дальнейшее обучение не дало значительного улучшения. Видеокарта была сильно загружена, и прирост точности составлял доли процента, поэтому не имело смысла продолжать. К тому же обучение одной эпохи занимало около десяти часов, что слишком долго для локальной машины.

Владимир Владимирович Чистяков:
Понимаю. Просто даже один процент улучшения бывает важен, особенно в задачах сегментации или классификации.

Мария Александровна Дурова:
Подскажите, как в обучающей выборке различались нейтральные высказывания? Например, если текст на религиозную тему, но без признаков враждебности, относили ли вы его к нейтральным?

Шамиль Алиевич Оцоков:
Да, такие тексты присутствовали. В датасете были примеры, где доля экстремистского содержания составляла примерно 50%, а оставшаяся часть — нейтральные или просто религиозные высказывания без негативной окраски. То есть модель училась различать подобные случаи.

Евгений Александрович Волошин:
Спасибо.

Владимир Владимирович Чистяков:
Да, спасибо.

Евгений Александрович Волошин:
На одном из слайдов упоминался параметр «адаптивность». Рассматривался ли вопрос о способности системы адаптироваться к изменению языковой среды? Например, если появляются новые формы сленга или способы завуалированных призывов, сможет ли модель подстраиваться под эти изменения или потребуется полное переобучение?

Шамиль Алиевич Оцоков:
Специальных механизмов адаптации пока не реализовано. На слайде, о котором вы говорите, речь шла о сравнении с существующими решениями — там указаны оценки из литературы. В нашем случае адаптивность можно обеспечить только периодическим обновлением датасета и переобучением модели на новых данных.

Сейчас система анализирует посты в «ВКонтакте» и определяет их принадлежность к экстремистскому или нейтральному контенту, но автоматическая самообучающаяся адаптация пока не внедрена.

Евгений Александрович Волошин:
Понятно, спасибо.

Павел Юрьевич Анучин:
Шамиль Алиевич, вы упоминали, что готовятся и другие доклады. Уже есть темы для следующих выступлений?

Шамиль Алиевич Оцоков:
Да. Следующий доклад будет посвящён применению агентов искусственного интеллекта в жизненном цикле разработки программного обеспечения. Его представит коллега, который работает в компании, реализующей подобный проект. Он поделится практическим опытом.

Павел Юрьевич Анучин:
Отлично. Тогда, возможно, заслушаем этот доклад на следующем семинаре?

Шамиль Алиевич Оцоков:
Да, я уточню детали и согласую дату. Коллега хотел выступить сегодня, но не смог по рабочим причинам. Я свяжусь с ним и предложу перенести выступление на ближайшее заседание.

Павел Юрьевич Анучин:
Хорошо. Семинары у нас проходят каждые две недели. В сентябре был небольшой перерыв из-за начала учебного года, поэтому сегодняшний мы перенесли на 17-е. Обычно встречи проходят в гибридном формате, но, конечно, очное участие предпочтительно — дискуссии тогда проходят живее.
Коллеги, есть ли у кого-то из присутствующих или подключившихся предложения по темам следующих семинаров или вопросы по уже обсуждённым докладам?

Евгений Александрович Волошин:
Похоже, вопросов больше нет.

Павел Юрьевич Анучин:
Тогда подведём итог. На следующем семинаре предварительно планируется доклад на тему «Применение агентов ИИ в жизненном цикле разработки программного обеспечения». Также, возможно, коллеги из НИИСИ представят свою работу — детали уточним позже.
Напомню, что все материалы семинаров, включая презентации и краткие сводки обсуждений, размещаются на сайте. Текущие презентации также будут опубликованы

в разделе сегодняшнего семинара. Если у коллег нет дополнительных вопросов и предложений, предлагаю завершать заседание. Спасибо всем за участие.