ИСПОЛЬЗОВАНИЕ ЯЗЫКОВЫХ МОДЕЛЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РЕШЕНИЯ ЗАДАЧ НАУЧНО-ТЕХНИЧЕСКОЙ ЭКСПЕРТИЗЫ: АКТУАЛЬНЫЕ ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ

М.В. Сергеев, гл. аналитик ФГБНУ НИИ РИНКЦЭ, канд. техн. наук, *sergeev@extech.ru*
И.М. Сергеев, ст. аналитик ФГБНУ НИИ РИНКЦЭ, *imsergeev@extech.ru*

Рецензент: Е.В. Ляпунцова, ФГАОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», д-р техн. наук, *lev7lev63@me.com*

В статье рассмотрены возможности применения больших языковых моделей (LLM[1]) искусственного интеллекта (ИИ) к решению задач в сфере научно-технической экспертизы. Отмечена актуальность использования LLM при анализе больших объемов текстовых документов, описаны методические подходы к развертыванию открытых LLM-моделей на локальной инфраструктуре организации с учетом требований информационной безопасности. Проанализированы актуальные возможности внедрения локальных LLM в работу научно-экспертных организаций, даются рекомендации по безопасному и эффективному использованию этих технологий.

Ключевые слова: научно-техническая экспертиза, технологии ИИ, текстовые документы, локальные LLM.

THE USE OF ARTIFICIAL INTELLIGENCE LANGUAGE MODELS FOR SCIENTIFIC AND TECHNICAL EXPERTISE: CURRENT CAPABILITIES AND PROSPECTS

M.V. Sergeev, Chief Analyst, SRI FRCEC, Doctor of Engineering, *sergeev@extech.ru*
I.M. Sergeev, Senior Analyst, SRI FRCEC, *imsergeev@extech.ru*

The article discusses the possibilities of applying large language models (LLM) of artificial intelligence (AI) to solving problems in the field of scientific and technical expertise. The relevance of using LLM in the analysis of large volumes of text documents is noted, methodological approaches to deploying open LLM models on the organization's local infrastructure, taking into account information security requirements, are described. The current possibilities of introducing local LLM into the work of scientific and expert organizations are analyzed, and recommendations are given on the safe and effective use of these technologies.

Keywords: scientific and technical expertise, AI technologies, textual documents, local LLMs.

Дисклеймер: упомянутые в тексте названия моделей ИИ не являются рекламой, авторы делятся личным опытом работы с конкретными моделями и литературными данными, которые находятся в открытом доступе на момент написания статьи. Материал предназначен для иллюстрации возможностей моделей и технологий и не является юридической консультацией или одобрением.

[1] Перечень использованных в статье терминов – см. в конце текста перед списком литературы.

Введение

В последние годы искусственный интеллект (ИИ) достиг качественно нового этапа своего развития, превратившись из экспериментальной технологии в повседневный инструмент, активно внедряемый в различные сферы деятельности. Еще два с половиной года назад взаимодействие с таким специализированным программным обеспечением, которое теперь принято называть ИИ, было уделом достаточно узкого круга специалистов, но все коренным образом изменилось в конце 2022 г., с появлением чата GPT фирмы OpenAI [1] в открытом доступе. Теперь пользователь[2], обладающий смартфоном или компьютером с выходом в Интернет, может вступить в диалог с ИИ и получить необходимую ему информацию. И уже в середине 2023 г. число ежемесячных обращений к этому чату превысило 300 млн.

С распространением коммерческих решений от технологических компаний LLM используются в разных областях деятельности. Список ниш весьма впечатляет: создание изображений, текстов, музыки, видео, а также успехи в таких сферах, как электронная коммерция, маркетинг и реклама, дизайн, образование и онлайн-курсы, здравоохранение и фитнес, недвижимость, фриланс и креативные услуги, логистика и доставка, туризм и финансовые услуги [2].

Далее в своем рассмотрении мы ограничимся только одним аспектом применения ИИ – достижениями в обработке текстовых документов (включая таблицы, графики и другие изображения), поскольку документы составляют преобладающую долю информации, которую обрабатывают в процессе научно-технической экспертизы. Здесь главенствующую роль играют большие языковые модели (LLM), построенные на архитектуре GPT (Generative Pre-trained Transformer), предварительно обученные на больших объемах текстовых данных.

О задачах, решаемых научно-технической экспертизой

Проблематика, требующая разрешения с привлечением научно-технической экспертизы, как отмечалось нами ранее [3], содержит задачи, различные как по масштабам и объектам, так и по целям экспертного анализа. Среди них, в частности:

– научно-технологическое развитие экономики;

– бюджетное финансирование научных исследований;

– конкурсное сопровождение грантовой поддержки науки.

Научно-техническая экспертиза по своей сути представляет собой высокоинтеллектуальную деятельность, сопряженную с анализом значительных объемов текстовой информации. К числу обрабатываемых документов относятся, в частности, заявки на получение грантов и участие в конкурсах, отчеты о выполнении научно-исследовательских работ, патентные описания, а также различные виды нормативно-технической документации.

При соблюдении общих методологических подходов к проведению экспертизы подобного рода задач в ФГБНУ НИИ РИНКЦЭ выработаны конкретные нормативы и правила, обеспечивающие системность организации экспертной работы [4]. Например, при проведении конкурсных процедур необходим объективный и всесторонний анализ заявок, что, в свою очередь, способствует целесообразному распределению конкурсного финансирования.

В табл. 1 представлены ключевые критерии, которые, как правило, входят в структуру экспертного заключения и служат основой для принятия решений о поддержке исследовательских инициатив.

[2] Односторонние санкции фирмы OpenAI против жителей России и Беларуси опытным пользователям не помеха.

Таблица 1

Критерии оценивания заявок на разработку научно-исследовательских проектов

Критерий	Предмет оценки
Анализ соответствия заявленного проекта целям конкурса	Оценка того, насколько проект соответствует заявленным целям и задачам конкурса, а также приоритетам в области науки и технологий
Оценка научной новизны и оригинальности	Анализ новизны предложенной идеи или технологии, определение, какой вклад проект вносит в развитие соответствующей области науки и технологий
Техническая осуществимость	Оценка возможности реализации проекта с технической точки зрения, включая доступность технологий, материалов и оборудования
Оценка научной и/или технической базы	Анализ научных и технических ресурсов, имеющихся у заявителя, включая квалификацию команды, наличие инфраструктуры и опыт решения аналогичных проектов
Экономическая целесообразность	Оценка бюджета проекта, включая его обоснованность, распределение средств и потенциальную экономическую эффективность
Оценка рисков	Выявление и анализ возможных рисков, связанных с реализацией проекта, а также наличие мер по их минимизации
Влияние на социально-экономическое развитие	Оценка потенциального вклада проекта в развитие региона, отрасли или общества в целом
Сравнительный анализ	Сравнение заявки с другими проектами, представленными на конкурс, для определения относительных преимуществ и недостатков
Рекомендации по улучшению	Формулирование рекомендаций для заявителя по улучшению проекта в случае выявления недостатков

Источник: составлено авторами на основе данных [4].

В ходе подобного анализа эксперту – специалисту высокой квалификации приходится проводить целый ряд рутинных операций типа валидации, не связанных напрямую с аналитической работой, но без которых подобное исследование не может обходиться.

Валидация заявки в контексте конкурсной документации включает несколько ключевых параметров проверки, которые помогают удостовериться, что поданная заявка соответствует установленным требованиям и критериям (табл. 2).

Изложенное позволяет утверждать, что проведение всесторонней и добросовестной экспертной оценки требует от специалиста значительных временных и интеллектуальных затрат. Наряду с аналитическими функциями эксперту приходится выполнять целый ряд вспомогательных операций, таких как проверка комплектности и формата документации, извлечение релевантных фактов и показателей из текстов, сопоставление содержимого заявок с установленными конкурсными критериями, а также формирование сводных таблиц и итоговых отчетов. Несмотря на важность этих действий для обеспечения надлежащего качества экспертизы, сами по себе они не предполагают глубоких предметных знаний и носят, по сути, рутинный характер. Именно в этой области возникают значительные резервы для автоматизации: современные инструменты ИИ, в частности большие языковые модели, обладают необходимым потенциалом для решения значительной части подобных задач.

Таблица 2

Основные параметры анализа заявки в процессе валидации

Параметр	Предмет проверки
Соответствие форме заявки	Проверяется, заполнены ли все необходимые поля, а также соблюден ли установленный формат подачи (например, наличие подписей, печатей и т.д.)
Соответствие требованиям документации	Анализируется, соответствует ли содержание заявки требованиям, указанным в конкурсной документации, таким как: – квалификационные требования к участникам (опыт, лицензии, сертификаты и т.д.); – технические требования к предлагаемым товарам или услугам
Полнота и достоверность информации	Оценивается, полно ли представлена информация и соответствует ли она действительности (например, наличие необходимых документов, финансовые показатели и т.д.)
Сроки подачи	Проверяется, была ли заявка подана в установленный срок
Критерии оценки	Анализируются соответствие заявки критериям оценки, указанным в документации, наличие всех необходимых материалов для их оценки
Условия участия	Проверка на выполнение условий, указанных в документации (например, отсутствие задолженности, соблюдение антикоррупционных норм и т.д.)
Финансовая состоятельность	Анализ финансовых документов для подтверждения способности участника выполнить условия контракта
Правовые аспекты	Проверка на наличие юридических ограничений для участия в конкурсе (например, наличие судебных разбирательств)

Источник: составлено авторами.

Специфика информации, которая содержится в документах, подлежащих экспертному анализу в научной и научно-технической сферах, состоит в том, что она носит закрытый характер (патенты, ноу-хау, персональные данные, коммерческие тайны, государственные секреты).

Утечка подобных «чувствительных» данных может привести к финансовым потерям, репутационным рискам или даже к правовым последствиям как для экспертной организации, так и для конкретных экспертов, обязанных соблюдать профессиональную тайну, закрепленную законодательством (например, Федеральный закон от 27.07.2006 № 152 «О персональных данных»). И даже при отсутствии прямых юридических ограничений эксперт несет моральную ответственность за защиту информации, доверенной ему заказчиком, поскольку, к примеру, утечка данных о перспективной разработке может позволить конкурентам скопировать технологии.

Таким образом, применение ИИ в научно-технической экспертизе представляется не только перспективным, но и практически целесообразным. Использование LLM может существенно ускорить обработку текстовых и графических материалов, повысить объективность и воспроизводимость оценочных процедур, а также снизить нагрузку на экспертов за счет автоматизации этапов предварительного анализа. Однако следует учитывать специфику экспертной деятельности, предъявляющую особые требования к конфиденциальности обрабатываемой информации и строгому соответствию формируемых выводов исходным данным.

Возможности и ограничения LLM

В течение последних двух лет LLM продемонстрировали значительный прогресс в области обработки текстовой информации и стали важнейшим инструментом анализа, извлечения и структурирования содержательных элементов из обширных массивов документов. Благодаря своей способности обобщать и систематизировать данные, LLM находят применение в самых различных сферах: от автоматизированной обработки юридических и технических текстов до анализа научных публикаций и подготовки кратких аннотированных обзоров. Эти функции особенно востребованы в ситуациях, когда специалисту необходимо в сжатые сроки ознакомиться с большим объемом информации и принять обоснованные решения на ее основе.

Определяющим аспектом работы с LLM является промпт-инжиниринг – навык формулирования запросов к модели для получения наиболее точных и релевантных ответов. Анализ результативности диалогов с моделями ИИ показал, что, несмотря на имеющийся прогресс в развитии ИТ-технологий, работает старый программистский принцип: «Что в компьютер заложишь – то и получишь!». Например, на общий и неопределенный запрос: «Проанализируй документ» – модель может сформировать не менее общий ответ: «Документ соответствует требованиям». Но формулировка запроса в более конкретной форме: «Составь перечень рисков и их причин в этом договоре» – позволяет получить развернутый, структурированный и содержательно насыщенный результат, отражающий как сами риски, так и их возможные источники.

Следует отметить, что за прошедшие два года промпт-инжиниринг превратился в самостоятельное направление, охватывающее разнообразные техники взаимодействия с LLM. На рис. 1 представлен перечень различных подходов, применяемых в промпт-инжиниринге, которые позволяют достичь высокой степени релевантности ответов LLM на запросы пользователя.

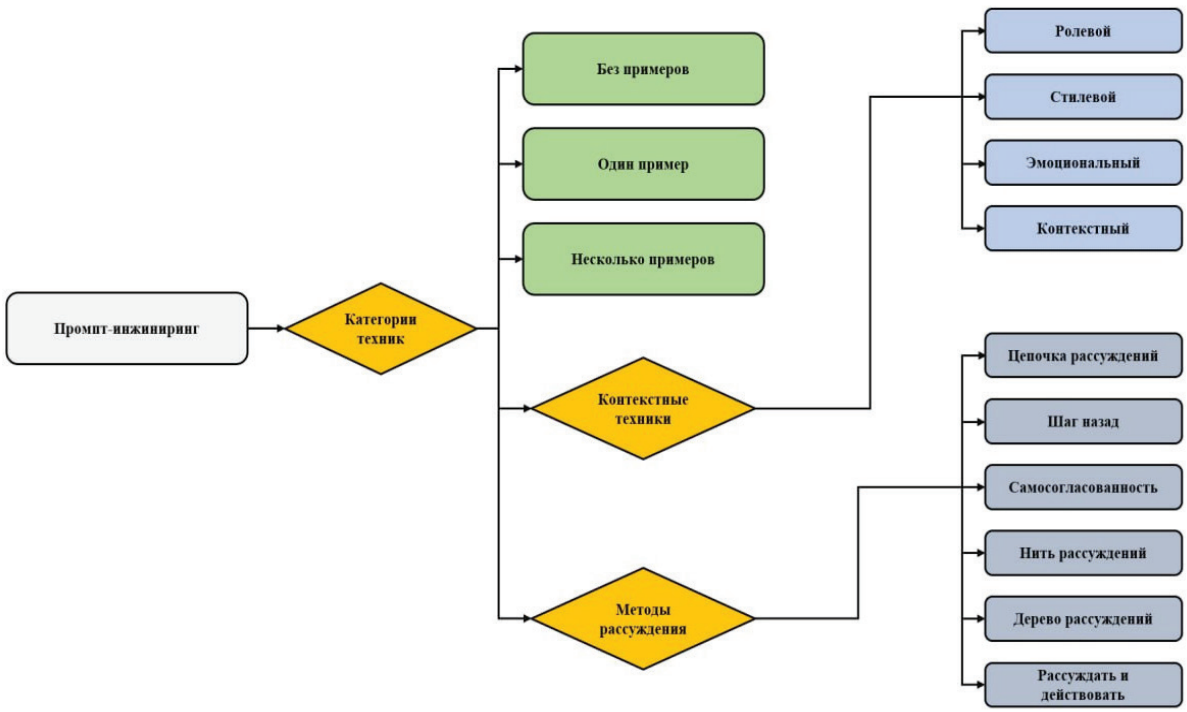

Рис. 1. Схема техник промпт-инжиниринга

Однако навык правильного формулирования запроса – не единственное, что определяет особенности общения с современными LLM [5–7]. Существенным ограничением выступает их закрытый (проприетарный) характер: исходные коды, архитектурные особенности и обучающие датасеты этих моделей недоступны для широкой исследовательской и профессиональной аудитории. Это означает, что пользователи полностью зависят от разработчиков в вопросах обновления, настройки, расширения функционала и обеспечения совместимости с внутренними системами.

Но для задач научно-технической экспертизы более принципиально то, что для получения достоверных и точных ответов LLM требуют «погружения в контекст», т.е. учета специфики предметной области, терминологии и логики рассуждения, характерной для конкретной задачи. Без такого контекстуального встраивания модель, как правило, демонстрирует поверхностные, обобщенные или даже искаженные результаты. Это означает, что проприетарная модель должна быть «настроена» на конкретную задачу, иначе говоря, требуется передача «чувствительных данных» (других в экспертизе, как показано выше, не бывает!) «в чужие руки», что неприемлемо для экспертной организации, поскольку одна из базовых предпосылок работы с проектной документацией – это строгий режим информационной безопасности.

Использование облачных сервисов, предлагаемых коммерческими провайдерами, также не позволяет гарантировать сохранность передаваемой информации: все данные – от пользовательских запросов до загружаемых документов – будут выходить за пределы внутреннего контура организации и обрабатываться на внешних серверах.

Альтернативой использованию проприетарных языковых моделей становится внедрение локальных LLM с открытым исходным кодом (open-source models), развертываемых непосредственно на вычислительной инфраструктуре самой организации. В последние годы открытое сообщество представило ряд моделей, сопоставимых по качеству с ведущими коммерческими системами. Среди них — семейство LLaMA [7] от компании Meta[3] (США) (включая версии LLaMA 2 и LLaMA 3), французские модели Mistral (с числом параметров 7 млрд и 16 млрд), китайская DeepSeek LLM (7 млрд и 67 млрд) [8] и др. Доступность архитектурных описаний и исходного кода обеспечивает высокую гибкость: организации получают возможность модифицировать модель под собственные задачи, а также интегрировать дополнительные компоненты и модули.

Главное преимущество open-source LLM заключается в возможности их локального развертывания, что гарантирует полный контроль над средой исполнения и сохранность обрабатываемых данных. Все операции осуществляются внутри защищенного информационного контура, исключая риски несанкционированного доступа или утечек информации. Более того, автономный характер таких решений устраняет зависимость от интернет-соединения и внешних связей, повышая надежность и отказоустойчивость систем. С экономической точки зрения, использование локальных моделей также может оказаться более целесообразным: отсутствуют затраты на подписки на облачные сервисы и на оплату запросов к удаленным ресурсам. Совокупность этих факторов делает open-source LLM особенно привлекательными для применения в научно-экспертной среде, где критически важны безопасность, адаптируемость и экономическая эффективность.

Текущие тренды в развитии искусственного интеллекта отмечают переход от громоздких, ресурсоемких языковых моделей к более компактным и эффективным архитектурам. Компактные языковые модели (SLM), обладая меньшими требованиями к вычислительным ресурсам и энергоемкости, обеспечивают возможность их внедрения непосредственно на пользовательских устройствах и в ограниченных инфраструктурах. SLM-модели становятся

[3] Meta – на территории России компания признана экстремистской организацией.

определяющим элементом нового этапа технологической революции, делая ИИ более устойчивым, этичным и управляемым. Некоторые SLM-модели, такие как Olmo 2 1B[4] с размерностью 1 млрд параметров, демонстрируют превосходство над LLM-моделями от крупных компаний в тестах на арифметические способности и фактическую точность. Модель Olmo 2 1B может эффективно работать на стандартных ноутбуках и даже на мобильных устройствах, что делает ее доступной для широкого круга пользователей [9]. Подобная децентрализация ИИ снижает барьеры доступа к технологиям, ускоряет адаптацию и персонализацию решений и способствует демократизации ИИ – аналогично тому, как распространение персональных компьютеров в прошлом радикально трансформировало цифровую среду.

Рис. 2. Этапы развертывания локальной LLM-модели в инфраструктуре организации

Развертывание локальной LLM-модели

Процесс установки малой модели LLM может включать 5 этапов, изображенных на рис. 2.

На начальном этапе осуществляется выбор модели, подходящей по необходимым для организации характеристикам (качество, требуемые вычислительные ресурсы, поддержка русского языка и т.д.). Среди популярных open-source моделей, доступных на данный момент, можно отметить LLaMA 2, LLaMA 3, Mistral 7B, XGen-8B, RuGPT-3 13B (Сбербанк) и другие – выбор зависит от конкретных задач и доступных технических ресурсов [10].

Далее модель развертывается в локальной среде. Существуют готовые инструменты, упрощающие этот процесс [11]: например, LM Studio (приложение для загрузки и запуска моделей), утилита командной строки Ollama, проекты типа LocalAI, Text-Generation-WebUI и др. Они позволяют запустить языковую модель в несколько шагов, зачастую используя технологии контейнеризации (Docker) для удобства развертывания. Аппаратные требования при

[4] Буквой B в обозначении типа модели принято указывать количество параметров модели в миллиардах.

этом зависят от размера модели: небольшие модели (до 7–13 млрд параметров) могут работать на современных персональных компьютерах (необходимо как минимум 16+ ГБ оперативной памяти, при наличии видеопамяти от 24 ГБ обработка данных ускоряется), более крупные (30–70 млрд параметров) требуют рабочих станций с высокопроизводительными графическими картами или развертывания на серверном оборудовании. В некоторых случаях применяют сжатие моделей (quantization), чтобы уменьшить объем занимаемой памяти ценой небольшого снижения точности.

После установки и запуска модель интегрируется в инфраструктуру организации – важно убедиться, что она функционирует в изолированном контуре, соответствующем политике информационной безопасности организации (без доступа к внешним сетям, с шифрованием хранимых данных и т. п.).

Следующий этап – адаптация модели к предметной области экспертизы. Базовые (pretrained) LLM, обученные на общих данных, не обладают специфическими знаниями, необходимыми для высокоточных ответов по узкой тематике. Важно отметить, что большинство текущих LLM-моделей не генерируют новых знаний, а оперируют только той информацией, которую «видели» при обучении. Если в исходном датасете не было сведений, скажем, о специфических показателях определенной технологии, то и ответ LLM на экспертный вопрос без дополнительной помощи будет неопределенным или ошибочным. Для решения этой проблемы применяются два подхода [12, 13] (которые можно использовать и совместно):

1) Retrieval-Augmented Generation (RAG) – механизм, при котором при каждом запросе к LLM выполняется поиск по внешней базе знаний (корпоративному хранилищу документов), и найденные релевантные тексты подставляются в подсказку модели;

2) Fine-tuning (дообучение) – дополнительное обучение модели на накопленных данных организации, включая архив экспертиз, отчеты и т. п.

В первом случае модель каждый раз получает актуальный контекст, в котором она сама не была изначально обучена, и за счет этого формирует более точные и фактически обоснованные ответы. Во втором случае происходит трансформация внутренних параметров модели под специфическую задачу за счет многократного прогона примеров вопросов-ответов из требуемой области (доучивание на отраслевом датасете). Fine-tuning позволяет «встроить» знания из специализированной области в модель, зафиксировав их на долгосрочной основе. Комбинация подходов (дообученная модель + RAG) считается наиболее мощным решением для достижения одновременно высокой точности и широкого охвата знаний.

Для реализации подхода RAG в условиях экспертной организации необходимо иметь систему управления базой знаний – электронный архив документов, по которому будет выполняться поиск. Это могут быть как простые утилиты (к примеру, скрипт, индексирующий папку с файлами PDF/DOCX и выполняющий поиск по ключевым словам, или семантический поиск), так и специализированные платформы (например, Haystack, Milvus, ChromaDB и др., позволяющие строить вопросы к базе знаний на естественном языке).

В свою очередь, Fine-tuning моделей требует подготовки размеченных данных (пар «вопрос – правильный ответ» по тематике экспертиз) и использования фреймворков глубокого обучения (по типу HuggingFace Transformers, PyTorch Lightning и т. п.) для обновления весов модели. В 2023 г. появились удобные средства для параметрического дообучения LLM с относительно малыми затратами ресурсов, например метод LoRA (Low-Rank Adaptation), позволяющие осуществлять fine-tuning больших моделей на одной-двух GPU за приемлемое время.

Перед вводом модели в эксплуатацию архитектуры организации крайне важно провести ее тестирование на контрольных примерах и убедиться, что ответы соответствуют ожидаемым. В методическом плане необходимо разработать регламент взаимодействия экспертов с LLM. Следует также учитывать, что даже после адаптации знаний LLM остается вероят-

ностной моделью, порой склонной к «галлюцинациям» – генерации правдоподобного, но не имеющего к действительности содержания. По данным исследований, современные LLM могут выдавать некорректные факты или ссылки в значительной доле случаев, особенно при выходе за пределы изученных данных [14]. Именно поэтому рекомендации ИИ должны проходить стадию валидации экспертом, прежде чем использоваться в финальных выводах экспертного заключения.

В ходе тестирования также стоит настроить систему запретов и фильтров (например, отключить или ограничить функции, ненужные экспертизе: генерацию программного кода, выполнение внешних запросов и т.д., если таковые имеются в модели). Для дополнительного повышения надежности можно использовать модульные решения: например, отдельный скрипт для проверки фактов (сопоставление ответа LLM с базой знаний) или несколько разных моделей для одних и тех же задач с последующим сравнением ответов модели. Также для удобства использования модели экспертами представляется важным разработать функциональный дизайн рабочей программы.

Предложенная методика включает полный цикл работ: от выбора и локального развертывания LLM-модели до ее донастройки и безопасной интеграции в процессы экспертизы.

Метод RAG позволяет модели встраивать в процесс генерации текстов элементы внешнего поиска и семантического сопоставления. В контексте экспертных задач это означает, что модель способна при обработке текста автоматически обращаться к нормативным документам, внутренним базам знаний и архивам, тем самым обеспечивая контекстуально точные и обоснованные ответы.

Технология Fine-tuning позволяет адаптировать поведение модели к требованиям конкретной предметной области: стилю, терминологическому полю и формату представления информации. Даже модели с относительно небольшим числом параметров демонстрируют высокую релевантность при условии корректной настройки на прикладные задачи отраслевой экспертизы. Таким образом, после дообучения на базе внутреннего массива документов и архивов организация получает инструмент, способный в сжатые сроки анализировать большие объемы текста, формировать структурированные таблицы соответствий и выдавать релевантные рекомендации. Это обеспечивает экономию времени эксперта, а также способствует стандартизации и воспроизводимости экспертных процедур.

Обсуждение

Несмотря на впечатляющие успехи ИИ, существующие LLM не лишены недостатков, которые особенно критичны в экспертной деятельности. Во-первых, как уже упоминалось, модель не генерирует новых знаний, а лишь комбинирует уже известную ей информацию. Это означает, что LLM не заменит полностью эксперта: она не способна сделать научное открытие или интуитивный вывод за пределами имеющихся данных. Роль модели – ускорить рутинные процессы, предоставить аналитику «черновой материал» для итоговых выводов, но не более. Во-вторых, остается проблема галлюцинаций и достоверности ответов. LLM может предъявить уверенно написанный текст, который выглядит правдоподобно, но не соответствует реальным фактам. В экспертной работе это недопустимо – ошибки ИИ могут привести к серьезным последствиям (неправильная оценка проекта, неверные рекомендации). Поэтому необходима многоуровневая система контроля качества ответов: автоматическое сопоставление с базой знаний, выявление потенциально выдуманных фактов, ручная проверка основных выводов. Только при условии такой проверки LLM может использоваться как вспомогательный инструмент.

Одним из существенных ограничений при внедрении LLM в практику экспертной деятельности являются высокие аппаратные и ресурсные требования. Крупные модели предъявляют значительные требования к вычислительным мощностям, особенно на этапах дообучения и адаптации к специфике задач. Не каждая организация располагает возможностями

инвестировать в специализированные серверы и соответствующую инфраструктуру. Частично данную проблему можно смягчить за счет применения методов оптимизации, таких как квантизация, сжатие модели или гибридные схемы хранения параметров, а также использования компактных моделей, демонстрирующих высокую эффективность при относительно невысоких затратах. Тем не менее с ростом сложности решаемых задач возрастает потребность в более мощных архитектурах, что вновь обостряет вопрос технического обеспечения. Помимо вычислительных ресурсов необходимы квалифицированные IT-специалисты, способные корректно настроить модель, осуществлять ее сопровождение, управлять обновлениями и обеспечивать соблюдение требований к информационной безопасности.

Принципиально важным аспектом является также то, что языковые модели обучаются на обобщенных корпусах данных, которые, как правило, не содержат специфической информации, подлежащей экспертной оценке. В связи с этим обязательным условием для практического применения LLM в экспертизе становится интеграция механизмов извлечения знаний (RAG) и дообучения (Fine-tuning) на локальных текстах, относящихся к предмету оценки. Такие процессы должны проводиться исключительно в рамках внутреннего контура безопасности организации, с соблюдением всех требований к защите данных, что делает их неотъемлемым этапом встраивания ИИ в экспертную инфраструктуру.

Важно также учитывать правовые и этические вопросы. При автоматизированной обработке заявок и проектов нужно гарантировать защиту персональных данных, отсутствие систематических смещений в оценках, равное отношение ко всем проверяемым объектам. LLM-модели обучены на больших массивах интернет-текстов, где могут присутствовать стереотипы и предвзятости; эти нежелательные паттерны могут проявляться в ответах. Необходимы меры по «очистке» и фильтрации выводов ИИ, особенно если речь идет об оценке людей или организаций.

Также требует дальнейшего анализа социальный аспект интеграции ИИ в экспертную деятельность. В частности, необходимо понять, как меняются роль, ответственность и квалификационные требования к экспертам в условиях гибридных систем, где функции анализа и суждения частично делегируются ИИ.

На данный момент локальные LLM уже доказали свою эффективность для ускорения экспертизы, обеспечивая подготовительную работу и позволяя эксперту сосредоточиться на высокоуровневом анализе. В дальнейшем можно ожидать появления все более специализированных моделей, обученных на узких тематических областях (химия, энергетика, биотехнологии и т.д.) – их интеграция повысит качество именно отраслевых экспертиз. Развитие методов RAG и связок LLM с внешними инструментами (например, с программами расчета или с онтологиями знаний) приведет к созданию своего рода «виртуальных экспертов», способных не только читать тексты, но и выполнять вычисления, строить логические цепочки обоснований. Однако полностью заменить человеческий фактор ИИ вряд ли сможет: эксперт-аналитик будет играть роль наставника и контролера ИИ, проверяя и направляя модель.

Заключение

Современные большие языковые модели (LLM) демонстрируют впечатляющие результаты, однако их использование в научной экспертизе и работе с чувствительными данными сопряжено с рядом ограничений. Проприетарные модели обучены на обширных, но закрытых датасетах, что делает их менее подходящими для задач, требующих прозрачности и воспроизводимости – ключевых требований в научной среде. Кроме того, их использование часто связано с необходимостью передачи данных на внешние серверы, что неприемлемо при работе с конфиденциальной информацией. Обучение собственных моделей с нуля требует значительных вычислительных ресурсов, доступных лишь крупным организациям.

В данной ситуации на первый план выходят open-source модели с возможностью локального развертывания. Принципы открытости и прозрачности кода позволяют обеспечить полный контроль над управлением данными и соответствие локальным требованиям безопасности, а использование методов адаптивного дообучения и извлечения знаний не только повышает точность и адаптирует модель к специфике анализируемых задач, но и позволяет интегрировать актуальные и специализированные знания без необходимости переобучения всей модели.

Таким образом, локальные open-source модели, дополненные методами RAG и Fine-tuning, представляют собой актуальное и своевременное решение для научной экспертизы и работы с чувствительными данными. Они обеспечивают необходимую точность, прозрачность и безопасность, позволяя организациям адаптировать модели под свои специфические потребности без риска компрометации данных в целях повышения эффективности процесса проведения научно-технической экспертизы.

Статья подготовлена к изданию при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках Государственного задания на 2025 г. № 075-00682-25-02.

Словарь актуальных терминов

Fine-tuning (файн-тюнинг) модели – процесс дополнительного обучения предварительно обученной модели на специализированном наборе данных для улучшения ее производительности в конкретной задаче, что достигается путем корректировки ее внутренних параметров (весов).

GPT (Generative Pre-trained Transformer) – семейство языковых моделей ИИ, основанных на архитектуре трансформеров, способных генерировать текст на основе заданного контекста.

LLM (Large Language Model) – большая языковая модель ИИ, обученная на обширных текстовых данных, способная генерировать и обрабатывать текст, отвечать на вопросы и выполнять другие задачи, связанные с языком.

Open-source model (опенсоурсная модель) ИИ – модель искусственного интеллекта, исходный код и данные которой доступны для общественности, что позволяет пользователям модифицировать и использовать ее свободно.

Prompt-engineering (промпт-инжиниринг) – процесс создания и оптимизации запросов (промптов) к языковым моделям для получения более точных и релевантных ответов.

Proprietary model (Проприетарная модель) ИИ – модель искусственного интеллекта, права на которую принадлежат частной компании или организации, и доступ к ней ограничен.

RAG (Retrieval-Augmented Generation) – модель ИИ, которая сочетает генерацию текста и извлечение информации из внешних источников для улучшения качества ответов и их актуальности.

Reasoning (ризонинг) – процесс объяснения и обоснования выводов, сделанных моделью, часто используемый для повышения прозрачности и доверия к решениям ИИ.

Sensitive (чувствительные) данные – данные, требующие особой защиты из-за их конфиденциальности или чувствительности, включая личную информацию, финансовые данные и медицинские записи.

SLM (Small Language Models) (компактные языковые модели) – это более легкие версии своих LLM-аналогов. Они поддерживают меньшее количество параметров, что позволяет оптимизировать их скорость и эффективность без ущерба для основных возможностей.

System model setting (системная настройка модели) – процесс адаптации и настройки модели под конкретные задачи или требования, включая изменение архитектуры или параметров обучения.

Облачный сервис – инфраструктура, которая предоставляет бизнесу доступ к вычислительным ресурсам, хранилищам и приложениям через Интернет.

Список литературы

1. Lock, Samantha. What is AI chatbot phenomenon ChatGPT and could it replace humans? // The Guardian (англ.) 05.12.2022. ISSN 0261-3077 (дата обращения: 15.03.2025).

2. Chat GPT: частное сообщение. URL: http://openai.com (дата обращения: 15.04.2025).

3. Сергеев М.В. Актуальные вопросы экспертизы в научно-технической сфере // Инноватика и экспертиза. 2023. Вып. 2 (36). С. 36–51.

4. Дивуева Н.А., Лукашева Н.А. Подход к формированию обобщенной методики организации и проведения научно-технической экспертизы // Инноватика и экспертиза. 2024. Вып. 2 (38). С. 32–42.

5. Секрет успеха Napkin AI: быстро, качественно и доступно. URL: https://habr.com/ru/companies/bothub/news/881526 (дата обращения: 15.04.2025).

6. Обзор видов лучших ИИ-моделей 2024: сравнение и рейтинг популярных LLM. URL: https://iaassaaspaas.ru/rating/ai/obzor-vidov-luchshih-ii-modeley-2024-sravnenie-i-reyting-populyarnyh-llm (дата обращения: 15.04.2025).

7. Быстрое введение в мир существующих больших языковых моделей (LLM) для начинающих. URL: https://habr.com/ru/articles/825032 (дата обращения: 15.04.2025).

8. Lambert G. DeepSeek R1: Is This the Open-Source Legal Tech Breakthrough We've Been Waiting For? // 3 Geeks and a Law Blog, Jan 2025. URL: https://geeklawblog.comgeeklawblog.com (дата обращения: 15.04.2025).

9. Новая небольшая модель искусственного интеллекта Ai2 превосходит аналогичные по размеру модели от Google и Meta*. URL: https://habr.com/ru/companies/bothub/news/906430 (дата обращения: 15.05.2025).

10. Сравнительный обзор LLM-моделей (LLM Model Comparison). URL: https://chatgpt.com/c/680a7562-044c-8007-ac74-a708db942dbc (дата обращения: 25.04.2025).

11. Развертывание кастомной LLM на собственной инфраструктуре (Custom build on-premise Large Language Model – Fine-tuning models on private business data). URL: https://unfoldai.com/build-custom-llm-business/?utm_source=chatgpt.com (дата обращения: 25.04.2025).

12. Как безопасно настроить локальные LLM и системы RAG (How to Set Up Local LLM and RAG Systems Securely). URL: https://www.chitika.com/local-llm-rag-security (дата обращения: 25.04.2025).

13. Дообучение LLM в 2025 году (Fine-tuning large language models (LLMs) in 2025). URL: https://www.superannotate.com/blog/llm-fine-tuning (дата обращения: 25.04.2025).

14. Причины галлюцинаций ИИ (и методы их уменьшения). URL: https://ru.shaip.com/blog/ai-hallucinations (дата обращения: 25.04.2025).

References

1. Lock, Samantha. What is AI chatbot phenomenon ChatGPT and could it replace humans? The Guardian (англ.) 05.12.2022. ISSN 0261-3077 (date of access: 15.03.2025).

2. Chat GPT: chastnoe soobshchenie [Chat GPT: private message]. Available at: http://openai.com (date of access: 15.04.2025).

3. Sergeev M.V. (2023) *Aktual'nye voprosy ekspertizy v nauchno-tekhnicheskoy sfere* [Current issues of expertise in the scientific and technical field] *Innovatika i ekspertiza* [Innovation and Expert Examination]. Issue. 2 (36). C. 36–51.

4. Divueva N.A., Lukasheva N.A. (2024) *Podkhod k formirovaniyu obobshchennoy metodiki organizatsii i provedeniya nauchno-tekhnicheskoy ekspertizy* [An approach to the formation of a generalized methodology for organizing and conducting scientific and technical expertise] *Innovatika i ekspertiza* [Innovation and Expert Examination]. Issue. 2 (38). C. 32–42.

5. *Sekret uspekha Napkin AI: bystro, kachestvenno i dostupno* [The secret of success Napkin AI: fast, high-quality and affordable]. Available at: https://habr.com/ru/companies/bothub/news/881526 (date of access: 15.04.2025).

6. *Obzor vidov luchshikh II-modeley 2024: sravnenie i reyting populyarnykh LLM* [Overview of the types of the best AI models in 2024: comparison and rating of popular ones LLM]. Available at: https://iaassaaspaas.ru/rating/ai/obzor-vidov-luchshih-ii-modeley-2024-sravnenie-i-reyting-populyar-nyhllm (date of access: 15.04.2025).

7. *Bystroe vvedenie v mir sushchestvuyushchikh bol'shikh yazykovykh modeley (LLM) dlya nachinayushchikh* [A quick introduction to the world of existing Large Language Models (LLM) for beginners]. Available at: https://habr.com/ru/articles/825032 (date of access: 15.04.2025).

8. Lambert G. DeepSeek R1: Is This the Open-Source Legal Tech Breakthrough We've Been Waiting For? 3 Geeks and a Law Blog, Jan 2025. Available at: https://geeklawblog.comgeeklawblog.com (date of access: 15.04.2025).

9. *Novaya nebol'shaya model' iskusstvennogo intellekta Ai2 prevoskhodit analogichnye po razmeru modeli ot Google i Meta** [The new small Ai2 artificial intelligence model is superior in size to similar models of Google и Meta*]. Available at: https://habr.com/ru/companies/bothub/news/906430 (date of access: 15.05.2025).

10. *Sravnitel'nyy obzor LLM-modeley* [LLM Model Comparison]. Available at: https://chatgpt.com/c/680a7562-044c-8007-ac74-a708db942dbc (date of access: 25.04.2025).

11. *Razvertyvanie kastomnoy LLM na sobstvennoy infrastrukture* [Custom build on-premise Large Language Model – Fine-tuning models on private business data]. Available at: https://unfoldai.com/build-custom-llm-business/?utm_source=chatgpt.com (date of access: 25.04.2025).

12. *Kak bezopasno nastroit' lokal'nye LLM i sistemy RAG* [How to Set Up Local LLM and RAG Systems Securely]. Available at: https://www.chitika.com/local-llm-rag-security (date of access: 25.04.2025).

13. *Doobuchenie LLM v 2025 godu* [Fine-tuning large language models (LLMs) in 2025]. Available at: https://www.superannotate.com/blog/llm-fine-tuning (date of access: 25.04.2025).

14. *Prichiny gallyutsinatsiy II (i metody ikh umen'sheniya)* [Causes of AI hallucinations (and methods to reduce them)]. Available at: https://ru.shaip.com/blog/ai-hallucinations (date of access: 25.04.2025).