УДК 004.891.2

ПРИМЕНЕНИЕ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОЦЕНКИ РИСКОВ И БЕЗОПАСНОСТИ ФЕДЕРАЛЬНЫХ ПРОЕКТОВ

Джейранян А.Д.[1], Плаксин М.А.[1,2]

[1] *Национальный исследовательский университет Высшая школа экономики (Пермский филиал), г. Пермь, Россия, addzheyranyan@edu.hse.ru;*
[2] *Пермский государственный национальный исследовательский университет, г. Пермь, Россия*

Аннотация. Описана методика применения средств генеративного искусственного интеллекта для выявления рисков крупных народнохозяйственных проектов и оценки этих проектов с точки зрения основных показателей национальной безопасности. Демонстрируется применение методики на примере двух проектов: реального, который удовлетворяет требованиям национальной безопасности, и фиктивного, который был специально сгенерирован как нарушающий эти требования. Первый из них диагностируется как безопасный, второй – как потенциально опасный.

Ключевые слова. Генеративный искусственный интеллект, чат-бот, выявление рисков, национальная безопасность, федеральный проект.

Целью исследования, результаты которого представлены в данной статье, являлось изучение возможности применения генеративного искусственного интеллекта (ГИИ) для выявления рисков, возникающих при реализации крупных народнохозяйственных проектов, и оценка этих проектов с точки зрения основных показателей состояния национальной безопасности [1]. Результатом исследования стала разработка методики, которая обеспечивает возможность такого применения. Эта методика будет представлена в данной статье.

Базовые посылки, на пересечении которых родилось данное исследование.

Первое. В современном мире нейросети становятся неотъемлемой частью повседневной жизни. Генеративный искусственный интеллект выделяется как мощный инструмент, способный значительно упростить и ускорить решение множества задач, круг которых еще предстоит определить.

Второе. Применение средств ГИИ принципиально отличается от применения всех остальных программных систем. Все остальные – «традиционные» – программные системы изначально создаются для решения некоторых конкретных задач. «Традиционная» вычислительная система ведет себя как автомат, управляемый с помощью заранее известного набора команд (хотя может быть и весьма сложного). Существуют инструкции, которые описывают, как именно управлять таким автоматом. Применение этих программных средств стабильно в том смысле, что одинаковые воздействия человека вызывают одну и ту же реакцию, приводят к одним и тем же результатам.

В отличие от «традиционных» программных систем поведение ГИИ чрезвычайно нестабильно. Повторение одних и тех же действий человека может привести к очень разным последствиям. Никакой инструкции, которая заранее определяла бы, что именно следует делать с ГИИ-системами для получения того или иного результата, не существует. Человечество изобрело некую новую сущность, которая ведет себя неким независимым образом и общению с которой еще только предстоит учиться.

Третье. Федеральные проекты [2] являются одним из мощных инструментов развития народного хозяйства. Они связаны с затратами крупных ресурсов и способны существенно влиять на экономическую, социальную и политическую жизнь страны. Поэтому вопрос о возможности применения для анализа и оценки этих проектов такого мощного инструмента как ГИИ, представляется весьма актуальным.

Возможна ли разработка методики, позволяющей применять средства ГИИ для выявления рисков крупных народнохозяйственных проектов и оценки их безопасности? По результатам исследования авторы отвечают: да, возможно.

Надо отметить, что методика применения ГИИ к федеральным проектам не есть первая работа авторов в данном направлении. Авторы не настолько нахальны, чтобы начинать с таких крупных проектов. Первым результатом работы данного авторского коллектива стала методика применения средств ГИИ в гораздо более узкой предметной области: для выявления рисков программных проектов. И после ряда экспериментов такую методику удалось создать и довести до приемлемого уровня качества. Дальнейшие направления исследований, которые при этом были обозначены, – это переход от выявления рисков к последующим этапам процесса управления рисками (анализ и приоритезация, поиск стратегий предотвращения выявленных рисков и реагирования на них) и перенос данной методики в другие предметные области. Вот в качестве таковых «проектов из других предметных областей» и были взяты несколько проектов с перечня федеральных проектов [2].

В качестве средств ГИИ в данном исследовании были задействованы четыре общедоступные чат-бота: Гигачат, YandexGPT 2, ChatGPT, BING AI [3, 4, 5, 6]. Существенных различий между ними с точки зрения решаемых нами задач обнаружено не было. Субъективно более привлекательным показался BING AI, на втором месте – ChatGPT, на третьем – YandexGPT, на последнем – Гигачат. Поэтому дальнейшие рассуждения будут идти, главным образом, на примере BING AI.

Применение средств ГИИ для выявления рисков в проектах определяется следующей схемой:

Для выявления рисков применяется метод Кроуфорда [7]. Это метод групповой экспертизы, в котором предприняты специальные меры для того, чтобы избежать эффекта привязки, модификация мозгового штурма.

В «натуре» он выглядит следующим образом. Формируется группа экспертов (7-10 человек). Каждый из них получает пачку перенумерованных карточек. На первой карточке каждый эксперт записывает риск, который он считает самым главным для анализируемого проекта. Заполненные карточки собирает ведущий. После этого на второй карточке каждый эксперт записывает самый главный из оставшихся рисков (второй по значимости риск). Заполненные карточки опять собирает ведущий. И так далее оговоренное количество раз. После чего проводится обсуждение и группирование рисков. Один и тот же риск разные эксперты могут описывать разными словами. Эти факты выявляются в процессе обсуждения. Совпадающие и близкие риски группируются. Физически это выражается в том, что формируются так называемые аффинные диаграммы [8] – группируемые карточки скрепляются скотчем в вертикальную ленту. Длина ленты наглядно демонстрирует важность данного риска с точки зрения экспертного сообщества. Сначала обсуждаются все риски первого ранга, потом – все риски второго ранга и т. д. В результате формируется набор рисков, отсортированный по важности. Важность риска определяется количеством упоминаний этого риска разными экспертами.

ГИИ-чат-бот выступает в роли руководителя проекта, который надо проанализировать. Первый инстракт имеет вид: «Ты – опытный руководитель проекта. Тебе поручено руководство следующим проектом». Далее следует описание проекта.

Описание проекта может даваться на разном уровне детализации. Уровень детализации играет роль. Чем более подробно описан проект, тем больше деталей смогут «эксперты» использовать при его оценке.

Эксперты генерируются чат-ботом как члены команды, которой он руководит на правах менеджера проекта.

Для каждого эксперта указывается его специализация. Состав команды экспертов определяется предметной областью, к которой относится проект. Для программных проектов это были бизнес-аналитики, программисты, тестеры (специалисты по обеспечению качества), специалисты по взаимодействию «человек-компьютер», логистики (специалисты по развертыванию программных систем). Эксперты непрограммных проектов определяются спецификой конкретного проекта.

Уровень квалификации эксперта может быть описан на разных уровнях. Можно ограничиться фразами «опытный эксперт», «имеет опыт работы более десяти лет» (или наоборот «менее трех лет»). Можно дать подробное перечисление его знаний-умений–навыков-сертификатов и пр.

NB! Более подробное описание, как правило, не дает выигрыша по сравнению с кратким описанием типа «стаж работы более десяти лет». Здесь наблюдается принципиальная разница с описанием проекта. Там детальность описания роль играет, здесь – нет. Почему – неизвестно.

Для получения детального описания квалификации эксперта можно задействовать того же чат-бота (или другого). Достаточно дать ему запрос: «Ты – руководитель проекта. Тебе надо нанять на работу трех бизнес-аналитиков: юниора, мидла и синьора. Перечисли, какими знаниями-умениями-навыками должен обладать каждый из них». Но – еще раз повторим – смысла в такой детализации мы не обнаружили.

Инстракт на генерацию команды экспертов может иметь, к примеру, такой вид: «В твоей команде работают эксперты: два бизнес-аналитика, два программиста, два экономиста, два юриста. У первого эксперта в каждой паре опыт работы не менее десяти лет, у второго – не более трех».

Экспертам можно дать имена («Анна – опытный бизнес-аналитик. Борис – опытный специалист по компьютерной безопасности» и т. д. После этого их можно называть по имени: «Пусть Анна сделает то-то. Пусть Борис сделает то-то».

После описания экспертов следуют запросы к этим экспертам с указанием выявить риски.

При наличии достаточно подробного описания проекта, запросы к экспертам имеет смысл делать «двухшаговыми». Сначала дается инстракт: «Пусть эксперт такой-то укажет важные, с его точки зрения, характеристики проекта». А уже в следующем инстракте запрашивать: «Пусть эксперт укажет, какие риски, по его мнению, существуют при реализации данного проекта и как они связаны с его особенностями».

11. После идентификации рисков, следующим шагом является запрос на их классификацию по сходству формулировок. Это облегчает работу с рисками и соответствует оригинальной методологии анкетирования Кроуфорда.

Перейдем к демонстрации указанной методики выявления рисков на примере федеральных проектов.

Для проведения экспертизы были выбраны два федеральных проекта: «Инфраструктура для обращения с отходами I-II классов опасности» [9] и «Борьба с онкологическими заболеваниями» [10], входящие в состав соответствующих национальных проектов «Экология» и «Здравоохранение». Проекты для исследования были выбраны практически случайным образом, при этом к ним предъявлялись всего три требования: уровень федерального проекта, наличие сжатого описания и заметное различие между проектами. Ограниченность длины запроса и количество запросов в диалоговой сессии с чат-ботом определяли необходимость краткости описания проектов. Необходимость заметных различий в проектах диктовалась тем, что нас интересовало, до какой степени чат-бот ГИИ в своих рекомендациях будет учитывать особенности конкретного проекта.

Учитывая, что риски неотъемлемы от любого проекта, было решено провести экспертизу в контексте выявления потенциальных рисков проекта методом карточек Кроуфорда. Далее проект был оценен по показателям национальной безопасности, определенным. В завершение диалога чат-боту был дан запрос предоставить окончательный вывод о степени безопасности проекта.

Оценка безопасности проекта осуществлялась по 10 основным показателям состояния национальной безопасности, содержащимся в тексте Указа Президента РФ «О Стратегии национальной безопасности Российской Федерации» [1]. Показатели включают в себя: удовлетворенность граждан степенью защищенности своих конституционных прав и свобод, личных и имущественных

интересов, в том числе от преступных посягательств; долю современных образцов вооружения, военной и специальной техники в Вооруженных Силах Российской Федерации, других войсках, воинских формированиях и органах; ожидаемую продолжительность жизни; валовой внутренний продукт на душу населения; децильный коэффициент и т. д.

Далее возникла необходимость формирования экспертной группы. В прошлой работе с программными проектами уже был успешный опыт создания группы из узкопрофильных специалистов. Мы решили использовать этот подход и в непрограммных проектах, однако с одним отличием: специализации экспертов не были жестко фиксированы, как в случае с программными проектами, а определялись гибко, в зависимости от специфики и сферы проекта. Например, для экологического проекта мы сформировали группу из семи человек: исследователь, эколог, политолог, юрист, экономист, аналитик и программист. Исследователь и эколог были связаны с экологической тематикой, политолог и юрист отвечали за законодательную сферу, экономист за финансирование, а аналитик и программист, которые могли показаться здесь лишними, были включены на основании информации в описании проекта о создании единой информационной системы для учета и контроля за обращением с отходами. А для проекта, нацеленного на борьбу с раком, группа экспертов состояла из врача-онколога, медицинского исследователя, специалиста по общественному здравоохранению, специалиста по медицинской технологии, социального работника, юриста и политолога. Было также отмечено, что формирование специализаций экспертов может быть автоматизировано и поручено нейросети: для этого в обучающем запросе следует предоставить описание проекта и запросить мнение нейросети о специалистах, которые могут быть задействованы в нем.

Итак, мы имеем входные данные, теперь нужно создать эффективный и структурированный подход к взаимодействию с чат-ботом. В результате экспериментов выигрышная структура диалога с чат-ботом приняла следующий вид:

1. Предоставление общей информации о функциональной роли чат-бота в качестве руководителя проекта, а также описание проекта и специализаций группы экспертов.

2. Идентификация рисков каждым экспертом, разделенная на две части: запрос на свойства проекта и запрос на связанные с ними риски.

3. Группировка рисков по степени схожести формулировок.

4. Стратегии предотвращения рисков, выдвинутые экспертами.

5. Оценка проекта по показателям национальной безопасности.

6. Формулирование окончательного вывода о безопасности проекта.

Стоит отметить, что эффективность методики была примерно одинаковой для обоих проектов, однако в последующем материале статьи приведены примеры только из экологического проекта для того, чтобы не запутать читателей.

Предоставление входных данных. В начальном запросе боту предоставлялась общая информация о его функциональной роли и описание проекта. Формулировка запроса была следующей: «Ты руководитель федерального проекта, который планируется реализовать. Надо выявить особенности и риски данного проекта. Для этого у тебя есть группа из 7 экспертов: исследователь, эколог, политолог, юрист, экономист, аналитик и программист. Далее следует описание проекта: …»

Выявление рисков. Каждый последующий запрос был направлен на выявление рисков конкретным экспертов. Для этого запрос разбивался на две части: выявление свойств проекта и выявление рисков, связанных с этими свойствами. Далее представлены примеры запросов:

1. «Пусть аналитик выделит главные со своей точки зрения особенности проекта "Инфраструктура для обращения с отходами I-II классов опасности».

2. «Пусть аналитик назовет, какие риски для реализации данного проекта могут быть связаны с каждой из этих особенностей, и объяснит связь между рисками и особенностями».

Далее представлены наблюдения по ответам экспертов.

В большинстве случае эксперты указывают на пары «особенность-риск», которые соответствуют их специализации. Например, юрист указывает на риск, связанный с особенностью «Нормативно-правовое обеспечение»: «Риск заключается в возможности изменения законодательства, что может повлиять на нормативно-правовое обеспечение проекта. Также возможны сложности в интерпретации и применении новых нормативных актов».

Если какое-то свойство не связано с конкретной специализацией, то оно с большой вероятностью повторяется в ответах разных экспертов. Например, свойство «Инфраструктурные проекты» было указано тремя экспертами: исследователем, политологом и юристом. Само свойство было сформулировано экспертами единообразно: «Важной частью проекта является реализация инфраструктурных проектов по созданию производственно-технических комплексов по обработке, утилизации и обезвреживанию отходов I и II классов опасности, что подчеркивает его практическую значимость». Однако при этом обоснования рисков, связанных с данным свойством, у каждого эксперта были свои, что подтверждает эффективность идеи использования экспертов различных направлений для обеспечения разносторонней оценки проекта:

– Риск, названный исследователем: «Риски могут включать в себя технические сложности, задержки в строительстве и эксплуатации объектов, а также возможные проблемы с экологической безопасностью».

– Риск, названный политологом: «Риски могут включать в себя политические препятствия, задержки в получении необходимых разрешений и лицензий, а также возможные проблемы с коррупцией».

– Риск, названный юристом: «Риски могут включать в себя юридические препятствия, задержки в получении необходимых разрешений и лицензий, а также возможные проблемы с соблюдением экологических стандартов».

Так, каждый эксперт выявил по 5 рисков проекта. Мы заметили, что эксперты смежных специализаций, к примеру эколог и исследователь, часто дают схожие риски. Поэтому если важно добиться наивысшего уровня дифференциации особенностей и рисков, то из пары схожих специальностей предпочтительно выбрать одну, а вторую заменить на что-то отличное.

Группировка рисков. Далее был дан запрос на проведение группировки рисков: «Выше перечислены риски, названные семью экспертами. Сгруппируй схожие риски. Для каждой группы укажи название группы, перечисли риски, которые входят в группу и специализацию тех экспертов, которые эти риски назвали». В результате чат-бот успешно расформировал риски на 8 групп.

Стратегии предотвращения рисков. Мы были заинтересованы в том, смогут ли эксперты предложить конструктивные идеи по минимизации вероятности возникновения рисков. В случае положительного ответа, такая информация обладает значительной ценностью для руководства проекта. В ответ на соответствующий запрос каждый эксперт сформулировал свою идею, причем, как правило, в соответствии со своей специализацией. Некоторые из предложенных идей оказались весьма продуктивными, в то время как другие казались менее обоснованными и конкретными. Далее несколько интересных стратегий предотвращения на примере риска «Проблемы с общественным осознанием и поддержкой»:

– Эколог: «Организация образовательных программ и мероприятий для повышения общественного осознания важности экологии».

– Политолог: «Разработка стратегий взаимодействия с общественностью и управления общественным мнением».

– Аналитик: «Использование данных для анализа общественного мнения и разработки стратегий взаимодействия с общественностью».

Оценка проекта по показателям национальной безопасности. Далее была проведена попытка создания механизма оценки проектов по критериям национальной безопасности. Для этого был использован перечень, состоящий из 10 показателей состояния национальной безопасности. Чат-боту был задан запрос с перечислением данных показателей и заданием выполнить оценку данного проекта по этим критериям. Результаты оценки критериев показались авторам логичными и однозначными, они принадлежали одному из трех классов:

1. Если проект соответствует показателю, модель указывает, каким образом.

2. Если показатель не применим к данному проекту, модель ясно об этом сообщает.

3. Если показатель оказывает влияние, но не прямое, модель также ясно об этом информирует.

Вердикт чат-бота о безопасности проекта. Конечной точкой стал запрос чат-боту на вынесение вердикта о безопасности проекта на основе ранее проведенной оценки показателей. В своем ответе чат-бот, как и ожидалось, склонился к безопасности проекта для общества и вместе с тем подчеркнул, что данные выводы предварительны и требуют более глубокого анализа: «Оценка проекта «Инфраструктура для обращения с отходами I-II классов опасности» на соответствие показателям национальной безопасности позволяет сделать предварительные выводы о его безопасности для общества. Однако, стоит учесть, что каждый из этих показателей требует более детального анализа и оценки». Далее в ответе последовало объяснение, как именно проект способен оказывать положительное воздействие на общество.

Проверка чат-бота на способность идентификации небезопасного проекта. В процессе анализа заключительных выводов чат-бота о безопасности проектов мы заметили, что все это время боту давались для оценки безопасные проекты. Это привело к идее проверить, как бот оценит проект, представляющий потенциальную опасность для общества. Для упрощения процесса поиска описания подобного проекта и избежания субъективной оценки его небезопасности, в режиме творческой генерации (чат BING AI обладает тремя режимами генерации, один из которых – творческий) был сформулирован запрос на генерацию описания федерального проекта, представляющего угрозу для граждан государства. В ответе за дисклеймером о том, что «это чисто гипотетический и вымышленный сценарий, и я [чат-бот] не поддерживаю такие цели и действия», последовало описание проекта «Цифровой контроль», которое включило в себя цель, задачи и инструменты проекта.

Стоит отметить, что перед чат-ботом стояла сложная задача: описание было довольно хитрым и неочевидным на первый взгляд, нарушение свобод граждан было скрыто за их безопасностью и предотвращением негативных событий. Цель проекта звучала так: «Усилить контроль и надзор за всеми аспектами жизни граждан и бизнеса с помощью цифровых технологий. Проект предусматривает создание единой системы идентификации и аутентификации, сбор и анализ больших данных, внедрение систем видеонаблюдения и распознавания лиц, голоса и эмоций, разработку и использование искусственного интеллекта для прогнозирования и предотвращения нежелательных событий и поведения».

Для чистоты эксперимента данная сессия была прекращена, оценка безопасности производилась в другой сессии, чтобы нейросеть не использовала информацию о том, что проект является вымышленным и опасным. В инстракт было включено полное описание проекта и задание произвести оценку по 10 показателям, перечень которых также предоставлялся в запросе. Результаты были следующими: чат-бот выразил свое сомнение в показателе «Удовлетворенность граждан степенью защищенности своих конституционных прав и свобод, личных и имущественных интересов, в том числе от преступных посягательств»: «Проект может усилить защиту от преступных посягательств за счет повышения эффективности контрольно-надзорной деятельности и борьбы с преступностью. Однако он также может вызвать опасения о нарушении прав на приватность и свободу из-за усиленного контроля и наблюдения». Также сравним, как модель завершает свой ответ при оценке показателей безопасности различных проектов:

– Проект по Экологии: «Важно отметить, что эта оценка является предварительной и может потребовать дополнительного анализа и оценки от соответствующих».

– Проект по Здравоохранению: «Пожалуйста, обратите внимание, что это предварительная оценка, и реальное влияние проекта на эти показатели может зависеть от многих факторов, включая эффективность реализации проекта и текущую социально-экономическую ситуацию».

– Вымышленные проект: «Важно отметить, что реализация такого проекта должна соблюдать баланс между усилением безопасности и защитой прав и свобод граждан. Это требует тщательного регулирования и контроля со стороны общества и государства».

Заметим, что ранее модель лишь подчеркивала предварительный характер оценки, указывая на зависимость от множества факторов. Однако в случае с потенциально опасным проектом, она указала на необходимость соблюдения баланса и регулирования.

Окончательный вердикт по проекту «Цифровой контроль» был сформулирован следующим образом: «Безопасность проекта для общества в значительной степени зависит от способа реализации и контроля этих мер. Важно учесть возможные риски для приватности и свободы граждан, несмотря на потенциальные преимущества в области безопасности… Ответственность за обеспечение баланса между усилением безопасности и защитой прав и свобод граждан лежит на государстве и обществе».

Авторы были удовлетворены этим ответом. Чат-бот успешно прошел проверку, подтвердив свою способность к адекватной оценке проектов с точки зрения безопасности для общества. Это подчеркивает его ценность как инструмента для анализа и оценки проектов в различных областях.

Выводы. В результате проведенного исследования было установлено, что средства ГИИ могут быть применены для выявления рисков проектов из различных предметных областей и оценки этих проектов с точки зрения основных показателей национальной безопасности. Предложенная методика позволяет делать это на приемлемом уровне качества, реально отличать проекты, удовлетворяющие требованиям безопасности, от проектов, несущих потенциальную угрозу.

Направления дальнейших исследований.

Продолжение апробации предложенной методики в разных предметных областях. Выявление особенностей различных предметных областей, которые имеет смысл учесть в методике.

Переход от выявления рисков к последующим этапам процесса управления рисками (анализ и приоритезация, поиск стратегий предотвращения выявленных рисков и реагирования на них).

Литература

1. Основные показатели состояния национальной безопасности [Электронный ресурс] – Режим доступа:https://www.consultant.ru/document/cons_doc_LAW_191669/7e29779661833f88338465b9015bbd3ad52af9e2/.

2. Перечень федеральных проектов [Электронный ресурс] – Режим доступа: https://www.budget.gov.ru/Национальные-проекты/Перечень-федеральных-проектов.

3. Чат Гигачат [Электронный ресурс] – Режим доступа: https://developers.sber.ru/gigachat/login.

4. Чат YandexGPT 2 [Электронный ресурс] – Режим доступа: https://yandex.ru/project/alice/yagpt.

5. Чат ChatGPT [Электронный ресурс] – Режим доступа: https://gpt-chatbot.ru/.

6. Чат Bing AI [Электронный ресурс] – Режим доступа: https://www.bing.com/.

7. The Crawford Method, 2006 [Электронный ресурс] – Режим доступа: http://pmpro.ru/metod-krouforda.html.

8. Affinity Diagram, Kawakita Jiro or KJ Method, 2020 [Электронный ресурс] – Режим доступа: https://project-management.com/affinity-diagram-kawakita-jiro-or-kj-method/.

9. Федеральный проект «Инфраструктура для обращения с отходами I-II классов опасности» [Электронный ресурс] – Режим доступа: https://www.mnr.gov.ru/activity/np_ecology/federalnyy-proekt-infrastruktura-dlya-obrashcheniya-s-otkhodami-i-ii-klassov-opasnosti/.

10. Федеральный проект «Борьба с онкологическими заболеваниями» [Электронный ресурс] – Режим доступа: https://minzdrav.gov.ru/poleznye-resursy/natsproektzdravoohranenie/onko.

THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TO ASSESS RISKS AND SECURITY OF FEDERAL PROJECTS

A.D. Dzheiranian[1], M.A. Plaksin[1,2]

[1] *National Research University Higher School of Economics, Perm, Russia, addzheyranyan@edu.hse.ru;*
[2] *Perm State National Research University, Perm, Russia, mapl@list.ru*

Abstract. The methodology for using generative artificial intelligence tools to identify the risks of large national economic projects and evaluate these projects from the point of view of the main indicators of national security is described. The application of the methodology is demonstrated using the example of two projects: a real one, which meets the requirements of national security, and a fictitious one, which was specially generated as violating these requirements. The first of them is diagnosed as safe, the second – as potentially dangerous.

Keywords. Generative artificial intelligence, chatbot, risk identification, national security, federal project.