Описание передовых направлений развития сферы искусственного интеллекта, в том числе «сильного» искусственного интеллекта, систем доверенного искусственного интеллекта и этических аспектов применения искусственного интеллекта, для определения тематик Центров

1. Направление «Искусственный интеллект для промышленности»

Сложные технические системы, к которым, в частности, относятся промышленные предприятия, энергетические системы, транспортные системы и т.д., как правило, оснащены средствами измерения технических параметров, которые позволяют получать большие объемы данных по технологическим процессам и техническому состоянию системы. Эти данные в большинстве случаев используются лишь частично, обеспечивая базовую диагностику состояния или визуализацию параметров.

Более глубокое использование больших объемов данных, включает в себя прогнозирование развития технологических процессов, раннюю диагностику проблем оборудования, построение рекомендаций по техническим параметрам.

Основные особенности подходов Центра:

- комбинирование традиционного моделирования, методов оптимизации и искусственного интеллекта, что позволит получить наилучшие результаты, опирающиеся как на теорию предметной области, так и на результаты обработки данных;
- повышенное внимание к предварительной аналитической обработке данных, что позволит более эффективно использовать квалифицированный труд специалистов в области искусственного интеллекта и машинного обучения.

Методы искусственного интеллекта могут быть применены в первую очередь для анализа временных рядов технических параметров и изображений; комбинация этих аналитических методов с более традиционными подходами прикладной математики приводит к созданию многофункциональных моделей.

В ходе работы Центра могут быть разработаны методы и инструменты предварительной аналитической обработки данных с использованием подходов искусственного интеллекта, а также примеры построения или прототипы прогнозных и рекомендательных систем, комбинирующих традиционное математическое моделирование с машинным обучением.

Результатом работы Центра могут быть прототипы в виде программного инструментария/программно-аппаратного комплекса на основе искусственного интеллекта, решающие прикладные задачи индустриального партнера.

Сферами применения являются: тяжелая промышленность; добывающая промышленность; транспорт; энергетический комплекс; любые другие отрасли промышленности.

2. Направление «Искусственный интеллект для медицины»

Получатель поддержки осуществляет исследования в области методов анализа медицинских изображений, текстов, видео совокупности разнородной клинической информации, а также последовательностей подобных типов данных.

Получатель поддержки развивает принципы и подходы к созданию интерпретируемых методов для прозрачного взаимодействия врача и ИИ и общего повышения эффективности принятия врачебных решений на основании ИИ. Важный фокус целесообразно сделать на исследованиях методов для работы с ограниченным числом данных с высокоточной разметкой и/или данных со слабой разметкой.

В ходе исследований могут быть разработаны методы обучения ИИ для анализа разнородных данных, данных со слабой разметкой, методы самообучения, с особым фокусом на обобщаемость этих методов на разные источники данных. Важные результаты деятельности Центра – создание эффективных метрик и инженерных практик верификации, валидации и мониторинга работы систем ИИ в медицинских организациях.

Создаваемые Центром технологии ИИ, должны быть реализованы в том числе в виде программного инструментария для создания систем поддержки принятия медицинских решений, обеспечивающего их достоверность (общая надежность выводов, полученных с помощью ИИ и проверенных на тестовых данных), безопасность (не причинят вреда пациенту, защита от взлома, несанкционированного доступа и др. негативных внешних воздействий), приватность (включая анонимизацию этих данных и разграничение доступа к ним).

Примерами являются технологии обработки мультимодальных медицинских данных из разных источников, построение моделей по неполным, несбалансированным, неточно аннотированным данным и др.

Разработанные системы могут быть ориентированы на широких круг задач первичной, вторичной и третичной медицинской помощи, включая первичную диагностику и маршрутизацию пациентов, проведение

контрольных диагностических исследований, выбор тактики лечения, ассистирование в планировании и проведении терапевтических и хирургических вмешательств.

3. Направление «Биометрические технологии искусственного интеллекта»

Биометрические технологии (БТ) основаны на измерении уникальных характеристик человека и включают в себя распознавание лица, речи, поведенческих особенностей. Задача БТ состоит в построении наиболее точных и надежных алгоритмов идентификации человека по его биометрическим данным. БТ уже сейчас повсеместно применяются для осуществления доступа к персональным (в том числе, банковским) данным, мобильным устройствам, а также для обнаружения и предотвращения преступлений. В основе современных технологий биометрии лежат алгоритмы искусственного интеллекта, которые уже достигли высокой точности и скорости работы, однако при масштабном внедрении возникает все больше сложностей, требующих решения. В частности, такие системы могут быть подвержены злонамеренным атакам. Такие атаки основаны на специально созданных модификациях цифровых данных, которые приводят к неправильной работе биометрических систем, например, к идентификации человека как другого лица для получения несанкционированного доступа. Такие ситуации могут возникать и из-за того, что в обучающей выборке отсутствовали определенные классы данных, и необходимо создавать подходы, которые могут быстро и надежно это детектировать.

Для нахождения ответов на фундаментальные и прикладные вопросы, связанные с развитием искусственного интеллекта для биометрии, предлагается создать Центр. Центр осуществляет исследования в том числе по следующим базовым направлениям: компьютерное зрение, распознавание образов, обработка трехмерных данных, анализ и обработка речи, робастность и интерпретируемость алгоритмов искусственного интеллекта. Все эти разрабатываемые технологии так или иначе связаны с обработкой больших объемов видеоданных, изображений и данных сенсоров, их эффективным хранением, в том числе в условиях существенных ограничений памяти, а также разработке эффективных моделей их обработки.

Центр должен развивать технологии искусственного интеллекта в своих предметных областях, наращивать экспертизу и организовывать площадку для доступа к ней. Ключевой особенностью работы Центра является в том числе создание отечественного программного обеспечения, которое может быть использовано для обеспечения национальной безопасности и решения

большого набора государственных задач в предметных областях работы Центра.

Центром могут быть разработаны новые и модифицированы существующие технологии компьютерного зрения и распознавания образов. Результаты могут публиковаться на ведущих конференциях в области ИИ, а также в ведущих международных журналах.

К важным результатам деятельности Центра могут относиться в том числе:

- повышение качества биометрических систем, включая разработку новых технологий распознавания изображений, захвата движения, распознавания жестов, создания 3Д моделей и обработки облаков точек и их программная реализация;
- разработка методов детектирования аномалий для биометрических систем и их программная реализация;
- создание подходов противодействия злонамеренным атакам на системы распознавания образов, детекторов и классификаторов и их программная реализация;
- создание технологий борьбы со спуфингом, детектирования фейковых цифровых данных (фотографии, видео, звук) и их программная реализация;
- создание алгоритмов биометрии, эффективно работающих на мобильных устройствах с минимально возможной передачей персональных данных и их программная реализация;
- разработка подходов к биометрии на основе мультимодальных (визуальных, звуковых и др.) данных.

Центр совместно с индустриальными партнерами должен разработать отечественное ПО, в котором воплощены технологии искусственного интеллекта, в том числе создаваемые в ходе реализации программы Центра.

Компьютерное зрение, доверенное распознавание образов востребованы в широком спектре индустриальных приложений: банковская сфера, безопасность, ритейл, ИТ-индустрия и индустрия развлечений. Кроме того, такие технологии, как захват движения, распознавания жестов и поз могут быть интегрированы в концепцию умного дома, умного производства и умного города.

4. Направление «Искусственный интеллект для оптимизации управленческих решений в целях снижения углеродного следа»

Вопросы экологии и снижения углеродных выбросов являются ключевым вызовом российской экономики. Вводимые углеродные пошлины на ввозимое в ЕС, США и Китай сырье и материалы ставят на грань рентабельности, не только формирующие 20% ВВП страны (угольную нефтегазовую, химическую и иные отрасли), но и отражаются на смежных секторах экономики: финансах, ИТ, социальной сфере. При отсутствии своевременных действий наша страна может потерять промышленный потенциал, накопленный за десятилетие развития экономики.

Энергетическая стратегия развития Российской Федерации до 2035 г. (распоряжение Правительства Российской Федерации от 9 июня 2020 г. № 1523-р) уделяет особое внимание вопросам устойчивого развития (ESG: Environment, Social & Governance), поэтому необходима разработка методов мониторинга и снижения прямых выбросов парниковых газов (промышленность, транспорт, добыча) и косвенных выбросов за счет потребления электроэнергии. Решение задач подобного типа предполагает обработку данных разной модальности (показаний датчиков, спутниковых данных, данных с камер видеонаблюдения и т.п.) и требует применения и развития технологий искусственного интеллекта (ИИ).

В рамках Центра целесообразно предусмотреть разработку системы определения ESG рисков (выбросы метана, CO2, разливы нефти и т.п.), их мониторинга и контроля в масштабах предприятий, регионов и страны. Такие системы используют ИИ для сбора и консолидации разнородных данных и создания гибкой иерархии предиктивных моделей на их основе. Это также предполагает развитие фундаментальных методов ИИ, ориентированных на конкретные приложения (physics-informed AI), и энергоэффективных методов ИИ (быстрые алгоритмы обучения, сжатия и т.п.) для обработки мультимодальных данных.

Внедрение разрабатываемых технологий в индустрии позволит не только снизить прямые и косвенные выбросы парниковых газов, но и увеличить выручку компаний, снизит затраты на применение ИИ и повысит привлекательность у акционеров. Трансфер разработанных технологий в индустрию, продвижение и развитие соответствующих продуктов могут осуществляться за счет компаний при экспертной поддержке Центра.

Направления исследований Центра могут включать в том числе:

Направление I. Разработка необходимых прикладных инструментов ИИ для мониторинга, прогнозирования и оптимизации ESG рисков

Направление II. Решение конкретных прикладных задач в области декарбонизации (снижения углеродного следа) и экологии, создание сервисов на основе разрабатываемых прикладных инструментов ИИ в интересах промышленности РФ.

Направление III. Развитие фундаментальных технологий ИИ и разработка соответствующего вычислительно эффективного программного инструментария для консолидации (data fusion) мультимодальных данных (результатов математического моделирования на основе Первых принципов, сенсорных данных и данных дистанционного зондирования) в целях предиктивного моделирования процессов, происходящих в окружающей среде.

Разработки Центра могут быть использованы в том числе для:

1. Создания систем экологического мониторинга, в частности, для построения на основе ИИ комплексных гибридных систем, использующих данные дистанционного зондирования Земли и иные современные сенсоры, и датчики, передающие информацию в режиме реального времени, для контроля и управления качеством окружающей среды, оценки динамики изменения углеродного следа и разработки оптимального комплекса мероприятий для его снижения;

2. Моделирования и оптимизации подходов к улавливанию и хранению углерода с текущими источниками энергии (газ, нефть, уголь) и компенсационных (лесоклиматических) проектов. Создания комплексных моделей оценки стоимости улавливания, транспортировки и переработки/утилизации/захоронения CO2 на производстве для повышения качества капиталоемких решений с целью снижения углеродного следа производства;

3. Повышения энергоэффективности и экологического нефтесервиса, в частности, для моделирования и подбора оптимальных значений параметров функционирования промышленных предприятий и добычи полезных ископаемых, что позволит минимизировать антропогенное воздействие на окружающую среду и снизить риски возникновения экологических проблем;

4. Оценки инфраструктурных инвестиций и финансового обеспечения ESG перехода, в частности, для создания комплексных систем оценки степени соответствия юридических лиц принципам «зеленой экономики» на основе технологий ИИ и консолидации большого количества разнородных данных.

5. Направление «Анализ естественного языка методами искусственного интеллекта»

Центр осуществляет исследования в том числе в области моделирования всех уровней естественного языка (фонетика, морфология, синтаксис, семантика, прагматика) и его модальностей (текст, речь) с помощью вычислительных систем и методов ИИ, таких как глубокое обучение. Особое внимание целесообразно уделить исследованию методов и моделей вычислительной семантики по двум направлениям:

- исследованию методов и моделей анализа, поддерживающих большое количество языков и учитывающих особенности обработки русского языка;
- исследованию методов интеграции графов знаний, таких как Wikidata, в модели обработки естественного языка.

Центр развивает принципы и подходы к созданию систем анализа естественного языка с помощью современных методов глубокого обучения и других подходов. Соответствующие исследования могут включать в себя создание языковых моделей, обеспечивающих моделирование естественного языка на уровне здравого смысла человека, например, за счет интегрирования базовой информации о мире из графов знаний.

В ходе исследований могут быть разработаны методы обучения, которые позволят получить повышение уровня анализа документов и ведения диалога. Важными результатами деятельности Центра могут быть создание новых языковых моделей, объединяющих графы знаний с нейросетевыми языковыми моделями, создание мультиязычных языковых моделей, а также, применение разработанных моделей в различных приложениях, таких как создание вопросно-ответных систем, систем информационного поиска и машинного перевода.

Результатом работы Центра могут быть в том числе:

- набор из нескольких предобученных мультиязычных языковых моделей (разного размера), интегрирующих информацию из текстов и графов знаний, которую можно использовать для решения в области автоматической обработки текстов.
- вопросно-ответная система по базе знаний (KBQA), созданная на базе разработанной модели;
- библиотека с открытым исходным кодом для обучения языковых моделей с использованием графов знаний и для решения различных прикладных задач в области автоматической обработки текста с использованием данной языковой модели.

Сферы применения создаваемых систем могут быть ориентированы на здравоохранение, образование, торговлю и электронную коммерцию и любые другие сферы, в которых возможно использование электронного оборота текстовых документов или возможно решение задач с использованием диалогового агента (голосового либо в виде чат-бота).

6. Направление «Искусственный интеллект для решения задач развития ТЭК и энергетики»

Основное направление деятельности Центра – вопросы внедрения инструментов ИИ для решения задач развития технологий топливно-энергетического комплекса и энергетики, в том числе задач увеличения нефтеотдачи. Решение задач ТЭК подобного типа предполагает обработку больших объемов разнородных данных (параметры нефтеносных пластов, параметры скважин, данные по добыче, и т.п.). С точки зрения энергетики также необходимо применение машинного обучения для обработки больших объемов данных (параметрические данные состояния и режимов работы энергетического оборудования, потребителей и источников генерации энергии и т.д.) и других технологий ИИ для оптимизации работы энергосистем объектов и зданий. Все эти задачи требуют применения и развития технологий ИИ.

В рамках Центра возможна разработка комплекса систем на основе методов искусственного интеллекта для оптимизации процессов добычи полезных ископаемых и повышения эффективности энергосистем, в частности: предсказательные модели по добыче скважин (предсказание временных рядов), машинное обучение на полевых данных для минимизации нежелательных явлений во время бурения, метамодели на синтетических данных для моделирования притока к скважине, для построения моделей ФЕС пласта и соотношений проницаемости-пористости гранулярных сред в приложении к технологии цифровой керн, методов машинного обучения для оптимизации подбора геолого-технических мероприятий (ГТМ), построение метамоделей процессов при реализации нефтесервисных технологий, а также модели для оптимизации и управления энергосистемами просьюмеров, «смарт гридов», производственных предприятий, зданий, использующих накопители энергии и распределенную генерацию, применение технологий ИИ для осуществления программно-аппаратного моделирования энергообъектов в реальном времени. Внедрение разрабатываемых технологий в индустрии позволит не только снизить затраты, но увеличить добычу, значительно повысить энергоэффективность производственных процессов и энергообъектов в целом.

Направления исследований Центра могут включать в том числе:

Направление I. Разработка необходимых систем сбора и хранения промысловых данных и параметричеких данных оборудования энергообъектов.

Направление II. Решение конкретных прикладных задач в области повышения нефтеотдачи, обеспечения надежности и эффективности энергосетей:

- создание систем предсказания выбытия добывающих скважин на основе алгоритмов интеллектуального анализа данных в нефтегазовой отрасли;
- создание прототипов расчетного ядра симуляторов на основе метамоделей и аппроксиматоров с использованием машинного обучения на синтетических, лабораторных и/или полевых данных (дизайн ГРП, моделирование добычи и т.д.);
- развитие ИТ решений для управления рисками при бурении нефтяных и газовых скважин на основе инструментов машинного обучения на полевых данных;
- создание цифровой лаборатории анализа керна (материала горных пород) на основе анализа изображений сканирования образцов и применения метамоделей для восстановления фильтрационно-емкостных свойств (ФЕС);
- решения для прогноза добычи на основе интеллектуального анализа исторических данных;
- рекомендательные системы по дизайну нефтесервисных технологий и управлению добычей на основе методов машинного обучения на исторических полевых данных;
- технологии предсказательного техобслуживания (например, для электроцентробежных насосов, ЭЦН).

Направление III. Развитие фундаментальных технологий ИИ для моделирования процессов, сопутствующих процессам добычи полезных ископаемых, и для программно-аппаратного моделирования и управления энергообъектов (в том числе зданий) в реальном времени с целью максимизации оптимизационного потенциала.

Разработки Центра могут быть использованы для:

- создания систем для центров управления добычей нефтяных компаний;
- повышения энергоэффективности нефтесервиса, в частности, для моделирования и подбора оптимальных значений параметров функционирования промышленных предприятий и добычи полезных

ископаемых, что позволит увеличить рентабельность добычи и отдачу на вложенный капитал;

– повышения энергоэффективности любых производственных и других энергообъектов (включая объекты распределенной генерации, смарт гриды, здания и т.д.) с целью снижения затрат и выбросов парниковых газов.

7. Направление «Искусственный интеллект для «Умного города» и транспорта»

Умный город – сложная система, состоящая из подключенных устройств, датчиков и интеллектуальных технологий. Но с появлением устройств возникает острая необходимость анализа, обработки и хранения огромных массивов данных. В свою очередь, грамотная работа с данными приносит пользу как жителям городов, так и правительству, и бизнесу. Искусственный интеллект становится одной из основных движущих сил цифровой трансформации экономики и социальной сферы. Изменяются общественный уклад, организация производства, предоставления услуг, логистические решения. Рутинные операции будут выполнять роботизированная техника, решения будут приниматься на основе технологий искусственного интеллекта. Искусственный интеллект и широкое применение интернета вещей позволят избежать управленческих ошибок и принять оптимальное решение во всех отраслях экономики и городского управления: определение необходимых объемов финансирования и распределения ресурсов, управление ЖКХ (в значительной степени будет осуществляться на основе больших данных с применением предиктивной аналитики, а отдельные схемы энерго-, тепло-, газо- и водоснабжения сформируют общую «Систему систем»).

Для нахождения ответов на фундаментальные и прикладные вопросы, связанные с развитием искусственного интеллекта для решения задач «Умного города», целесообразно создать Центр, осуществляющий исследования в том числе по следующим базовым направлениям: робототехника, распознавание речи, изображений и видео, автопилотирование в контролируемом окружении, предиктивная аналитика, блокчейн, смарт-вещи с машинным обучением, диалоговые платформы, чат-боты, персональные ассистенты.

В рамках Центра целесообразно предусмотреть разработку комплекса систем на основе методов искусственного интеллекта, которые позволят кардинально улучшить городскую среду и условия жизни, труда и отдыха горожан; уменьшить непроизводительные затраты времени на перемещение

по городу, административные процедуры, оформление документов, получение услуг.

Результатом работы Центра может быть общегородская платформа данных в целях поддержки быстрого и обоснованного принятия городских решений на основе больших данных и искусственного интеллекта и формирования условий для создания новых услуг на основе городских данных, в том числе в целях учёта/экономии потребления ресурсов, своевременного предупреждения аварий и сокращения времени их устранения, контроля степени изношенности инженерных коммуникаций, повышения прозрачности управления и решения других задач жилищно-коммунального хозяйства.

Правительство и бизнес получат качественные профессиональные инструменты поддержки принятия решений с использованием последних достижений в области искусственного интеллекта и аналитики больших данных для применения в следующих сферах: безопасность, розничная торговля, логистика, доставка, транспорт, финансы, медицина, образование, ЖКХ и др.

8. Направление «Искусственный интеллект для робототехники и управления беспилотными системами»

В робототехнике существует множество задач, связанных с высокоуровневыми системами принятия решений. Современные роботы работают в неструктурированной динамической среде. Например, коллаборативные роботы, взаимодействуя с человеком могут оперативно и безопасно реагировать на поведение человека, которое является непредсказуемым. Прогнозы поведения и анализ логики принятия решений не всегда может быть описан формальными математическими правилами, поэтому зачастую используются методы искусственного интеллекта для эффективной обработки большого количества данных. Автономное вождение – одна из ключевых областей применения искусственного интеллекта, что позволяет значительно расширить количество сценариев, которые способен обработать робот без вмешательства человека.

Основными сферами применения в предметной области Центра могут быть в том числе:

- системы распознавания автономных роботов, позволяющие реагировать на объекты в окружающем пространстве. Беспилотные роботы оснащенные, как правило, несколькими датчиками, такими как камеры, радары и лидары, которые помогают им лучше понимать окружающую обстановку и планировать путь. Эти датчики генерируют огромное

количество данных, обработка которых может быть ускорена комбинированием традиционных алгоритмов и искусственного интеллекта, что позволит получить наилучшие результаты по точности определения объектов;

- система прогнозирования поведения участников, взаимодействующих с роботом, это могут быть участники дорожной сцены, такие как пешеходы или другие транспортные средства, а также объекты инфраструктуры.
- система предиктивной аналитики для оценки и анализа состояния робота, за счет получаемых данных с датчиков и интегрированного в нейронные сети опыта по техническому обслуживанию;
- планирование и оптимизация транспортных потоков, для повышения общей скорости и безопасности передвижения в динамических средах, таких как города или отдельно взятые территории.

Методы искусственного интеллекта могут быть применены в первую очередь для анализа данных с сенсоров беспилотного робота. Комбинация этих аналитических методов с более традиционными подходами приведет к созданию многофункциональных моделей.

В ходе работы Центра могут быть разработаны методы и инструменты предварительной аналитической обработки данных с использованием подходов искусственного интеллекта, в том числе проведены:

- анализ и исследование методов формирования структуры гибридного интеллектуального управления роботами с применением новых подходов;
- разработка гибридных алгоритмов на основе классических и новых подходов;
- разработка программной архитектуры и модулей, реализующих новые методы.

Результатом работы Центра могут быть прототипы в виде программного инструментария/программно-аппаратного комплекса на основе искусственного интеллекта, отработка разработанных алгоритмов и решение прикладных робототехнических задач. Применение методов искусственного интеллекта позволит повысит уровень автономности роботов за счет расширения сценариев работы.

Сферами применения являются беспилотные наземные и воздушные транспортные средства; системы мониторинга состояния техники; интеллектуальные транспортные системы; складские робототехнические системы; коллаборативные роботы; другие области робототехники.

9. Направление «Искусственный интеллект в сельском хозяйстве и производстве продуктов питания»

Центр развития ИИ в сельском хозяйстве и производстве продуктов питания осуществляет исследования в том числе в области методов предиктивной аналитики, систем принятий решений, компьютерного зрения, систем ИИ для робототехники и дронов, и других направлениях ИИ. Результаты исследований могут быть направлены на повышение эффективности работы сельского хозяйства, рост урожайности, эффективность разведения скота, производство продуктов питания, снижение нагрузки на персонал.

Сельское хозяйство охватывает огромные территории по всей стране. Исследования в области искусственного интеллекта могут учитывать влияние большого количества внешних факторов (почва, экология, климат и пр.) в условиях постоянного изменения этих условий.

Результаты Центра могут обладать высокой степенью автономности работы всех систем, чтобы снижать нагрузку на персонал и повышать эффективность всех процессов.

Результатом работы Центра могут быть прототипы в виде программного инструментария и (или) программно-аппаратного комплекса на основе технологий искусственного интеллекта, решающие прикладные задачи индустриального партнера.

Разработанные системы могут быть ориентированы на снижение затрат и повышение эффективности операций в результате внедрения перспективных решений с ИИ по всей цепочке создания стоимости, например:

1. Растениеводство:
    a. Посев, выращивание, осмотр полей:
        - прогнозирование урожайности;
        - инструменты точного земледелия, дифференцированное внесение удобрений, планирование севооборота.
        - использования автономных самоходных роботов для ухода за растениями;
    b. Сбор урожая, сортировка и первичная переработка:
        - роботизированный сбор урожая автономным транспортом;
        - автоматизированная сортировка с помощью компьютерного зрения;
        - автоматизированная переработка;
2. Животноводство:

    Разведение и набор массы скота:

- мониторинг и анализ поведения и здоровья скота с использованием технологий ИИ;
- инструменты ухода за скотом (в зависимости от изменений климата, экологии и пр.).

Сферами применения являются сельское хозяйство (растениеводство, животноводство), производство продуктов питания. Другие сферы, где могут применятся разработанные технологии ИИ: экология, промышленность, роботизация и пр.

10. Направление «Искусственный интеллект в биотехнологиях и генной инженерии»

Центр осуществляет исследования в области методов анализа данных, предиктивной аналитики, систем принятий решений и по другим направлениям ИИ. Результаты исследований могут быть направлены на решение актуальных задач фармацевтики, селекции, генной инженерии, а также на повышение эффективности работы высококвалицированных специалистов в этих областях.

Исследования в биотехнология и генной инженерии не могут существовать в отрыве от других областей. Это кросс-отраслевое направление, результаты работы которого являются фундаментом для практического применения в медицине, сельском хозяйстве, промышленности и пр.

Результатом работы Центра могут быть прототипы в виде программного инструментария и (или) программно-аппаратного комплекса на основе технологий ИИ, решающие прикладные задачи в одной или нескольких областях:

Ключевые области в том числе:

a) Фармацевтика:
   - поиск новых лекарственных препаратов для лечения;
   - подбор эффективных лекарственных препаратов для лечения болезней (в том числе персонализация);

b) Селекция:
   - прогнозирование и подбор оптимальных пар для скрещивания;

c) Геномные технологии и структурная биология:
   - прогноз структуры белка по последовательности аминокислот;
   - моделирование структуры клетки и отдельных органелл;
   - прогноз фенотипа по генотипу.

Сферами применения являются сельское хозяйство (растениеводство, животноводство), фармацевтика и другие области.

11. Направление «Доверенный искусственный интеллект»

Приоритетное направление Центра – разработка методов и технологий мирового уровня, необходимых для создания доверенной конкурентоспособной продукции с применением технологий ИИ (ТИИ) в различных прикладных областях.

Деятельность центра затрагивает весь спектр прикладных задач и систем с применением ТИИ в различных отраслях экономики и социальной сферы, включая системы обработки изображений, аудио, текстовой и другой информации (системы распознавания лиц, системы обнаружения вторжений, системы автономного вождения и др.). Ключевыми в деятельности Центра являются анализ и верификация указанных систем на всех этапах их жизненного цикла, включая в том числе управление процессами:

- разработки доверенных систем с ТИИ;
- тестирования систем с ТИИ на этапе их ввода в эксплуатацию с целью подтверждения соответствия требованиям в области эффективности, продуктивности, безопасности и надёжности;
- диагностики и исправления ошибок систем с ТИИ в ходе эксплуатации;
- дообучения систем с ТИИ и регулярного тестирования систем с ТИИ на стадии эксплуатации, и др.

Перечисленные процессы должны учитывать требования к **безопасности**, **надежности**, **эффективности** и **продуктивности** систем с ТИИ в каждой из прикладных областей, а также основанные на этих требованиях **понятие доверенной системы** с ТИИ и **критерии** доверия в этих прикладных областях.

Управление перечисленными процессами подразумевает разработку методик и прикладных инструментов для оценки и обеспечения безопасности и эффективности платформ машинного обучения, моделей машинного обучения и наборов данных, а также создание верифицированных в соответствии с разработанными методиками платформ машинного обучения (в том числе на основе свободного программного обеспечения).

Внедрение ТИИ требует учитывать ряд возникающих рисков и новых угроз. ТИИ включают целый стек технологий, состоящий из методов и алгоритмов ИИ, платформ машинного обучения, а также инфраструктурных решений для их поддержки (облачные системы, распределенные системы, специализированные аппаратные системы и др.). Эти технологии являются источниками новых типов ошибок и уязвимостей, которые отличаются от классических уязвимостей программного обеспечения и предоставляют новые

возможности для проведения атак злоумышленниками. К числу таких атак можно отнести атаки:

- с использованием состязательных примеров;
- с порчей наборов данных;
- с внедрением программно-аппаратных закладок в нейросетевые модели;
- с кражей моделей и пользовательских данных из облачных сред.

Главным результатом может стать **многоуровневый стек системного ПО для анализа и верификации систем, основанных на применении технологий ИИ,** на всех этапах их жизненного цикла. Такой стек включает в себя в том числе:

- платформы машинного обучения (не менее одной на основе свободного программного обеспечения), прошедшие проверку в соответствии с разработанными методиками;
- технологии для выявления классических уязвимостей и атак, а также состязательных атак на этапе сбора и обработки данных, на этапе (до-)обучения и функционирования модели, которые специфичны для систем с ТИИ;
- методы и программные средства оценки наборов данных, используемых для создания доверенных систем с ТИИ, на соответствие предъявляемым требованиям;
- методы синтеза доверенных систем для решения типовых задач ИИ, включая методы очистки данных и предварительно обученных моделей, методы федеративного и состязательного обучения, методы определения сдвига данных и понятий и иные методы;
- эталонные наборы данных с примерами уязвимостей, состязательными примерами, моделями с программными закладками для оценки безопасности и эффективности существующих и новых платформ машинного обучения;
- программные инструменты для оценки соответствия систем с ТИИ предъявляемым требованиям;
- технологии для улучшения интерпретируемости нейросетевых моделей, для объяснения результатов в системах поддержки принятия решений;
- прототипы доверенных систем с применением ТИИ, обеспечивающих решение конкретных прикладных задач в различных областях (транспорт, медицина, образование и др.).

Помимо самой платформы, в рамках деятельности Центра целесообразно формализовать понятия доверенной, безопасной и интерпретируемой системы с ТИИ и соответствующие критерии в различных

прикладных областях, разработать методики и рекомендации по разработке систем с использованием ТИИ, соответствующих заданным требованиям, в том числе предложения по модификации процессов жизненного цикла для анализа качества ТИИ и оценки их уязвимостей, научно-методические основы формального описания предусмотренных (допустимых) условий эксплуатации систем с применением ТИИ, рассчитанных на решение типовых задач.

Перечисленные результаты затрагивают все информационные системы, а также все стадии разработки и внедрения информационных систем, в которых применяются ТИИ. В первую очередь это системы, предъявляющие самые жесткие требования к безопасности, и системы, в которых обрабатывается конфиденциальная информация, включая персональные данные.

12. Направление «Межотраслевые технологии искусственного интеллекта и искусственный интеллект для иных приоритетных отраслей экономики и социальной сферы»

К приоритетам Центра относится деятельность в одном или нескольких направлениях:

1) Создание отраслевых платформенных решений на базе межотраслевых технологий искусственного интеллекта;

Отраслевые платформенные решения включают набор разнотипных программных инструментов и решений, согласованных по входным и выходным данным, подходам и требованиям, в том числе связанным со спецификой ИИ, позволяющей решать широкий спектр отраслевых задач, как правило выходящих за пределы обычного прикладного программного обеспечения. Такие решения могут быть направлены на применение в отдельных отраслях, но при этом базироваться на общих компонентах, инструментах, методах и алгоритмах ИИ, применимых для использования в различных отраслях в рамках решения технологически схожих задач (межотраслевые технологии ИИ), в том числе в сложных комплексах корпоративного, отраслевого, муниципального и федерального управления.

В ходе создания платформенных решений центральную роль играет решение базовых задач машинного обучения, актуальных не только внутри разнотипных задач одной отрасли, но и для различных отраслей, развитие новых кросс-отраслевых и междисциплинарных технологических направлений (новые материалы, хемоинформатика, нейротехнологии, робототехника, геоинформационные технологии и др.) с применением технологий ИИ.

2) Создание отраслевых прикладных решений для иных отраслей, в том числе для приоритетных отраслей экономики, не вошедших в направления центров (образование, финансовые услуги, госуправление и госуслуги).

Фокус отраслевых прикладных решений – повышение эффективности процессов планирования, мониторинга, прогнозирования и принятия управленческих решений, повышения качества обслуживания, в том числе за счет индивидуализации предоставляемых услуг.

Научная деятельность предлагаемого Центра связана с решением проблем, актуальных для создания отраслевых и межотраслевых технологий ИИ, платформенных и прикладных решений.

Решение проблем может включать:

- создание моделей ИИ в условиях неопределенностей с данными при создании отраслевых платформенных решений;
- обобщение моделей ИИ в условиях обучающих выборок малого объема, для применения в отраслях, где создание наборов данных большого объема затруднено;
- дизайн экспериментов методами машинного обучения с целью оптимального выбора подмножества данных для обучения, в том числе для применения в междисциплинарных технологических направлениях;
- создание интерпретируемых методов машинного обучения, в том числе для отраслей с высокой ценой ошибки при принятии решений;
- комбинирование традиционного моделирования, методов оптимизации и искусственного интеллекта, что позволит получить наилучшие результаты, опирающиеся как на теорию предметной области, так и на результаты обработки данных;
- предварительная аналитическая обработка данных, что позволит более эффективно использовать квалифицированный труд специалистов в области искусственного интеллекта и машинного обучения.

При создании отраслевых прикладных решений могут применяться технологии распределенных интеллектуальных систем анализа данных, моделирования взаимодействия мультиобъектных систем и поддержки принятия управленческих решений.

Разрабатываемые методы ИИ могут быть применимы к различным видам данных (изображения, текстовая информация, аудио, временные последовательности сигналов и др.), в том числе разнородным.

В ходе работы Центра целесообразно предусмотреть разработку алгоритмов, методов и инструментов для решения наиболее актуальных проблем машинного обучения, препятствующих созданию эффективных прикладных и платформенных решений в различных отраслях экономики. С

использованием полученных научных результатов целесообразно развитие отдельных технологических направлений (в том числе тех, где технологии ИИ являются принципиально новой парадигмой решения задач).

С учетом масштабов и сложности отношений в указанных сферах, объемов данных и требований к работе с ними и их защите, особую роль в развитии и внедрении систем искусственного интеллекта начинают играть распределенные интеллектуальные системы анализа данных, моделирования взаимодействия мультиобъектных систем и поддержки принятия управленческих решений.

В ходе работы Центра могут быть разработаны методы и инструменты предварительной аналитической обработки данных с использованием подходов искусственного интеллекта, а также примеры построения или прототипы прогнозных и рекомендательных систем, комбинирующих традиционное математическое моделирование с машинным обучением.

В ходе создания отраслевых прикладных решений могут быть разработаны методы и инструменты предварительной аналитической обработки данных с использованием подходов ИИ, подходы к интеграции информационных систем как внутри одной отрасли, так и между отраслями, а также примеры построения или прототипы прогнозных и рекомендательных систем, комбинирующих традиционное математическое моделирование с машинным обучением.

Результатом работы Центра могут быть прототипы в виде программного инструментария / программно-аппаратного комплекса на основе ИИ, решающие прикладные задачи индустриального партнера.

Перечисленные результаты затрагивают все информационные системы, а также все стадии разработки и внедрения информационных систем, в которых применяются технологии ИИ. В первую очередь это системы, предъявляющие самые жесткие требования к безопасности, и системы, в которых обрабатывается конфиденциальная информация, включая персональные данные.

Проектирование результатов работы Центра целесообразно осуществлять с учетом приоритетных отраслей экономики и социальной сферы согласно перечню, указанному в абзаце 8 подпункта «б» пункта 11 Указа Президента Российской Федерации от 7 мая 2018 года № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года».

13. Направление «Этические аспекты применения искусственного интеллекта»

Технологии искусственного интеллекта (ТИИ) обладают мощным трансформирующим потенциалом. Они быстро и масштабно воздействуют на человека и общество, приводя к системным изменениям в поведении людей, общественном сознании, социокультурных и иных моделях развития общества.

На современном этапе развития ТИИ существует уникальная возможность своевременно сформулировать общие ценностные основания и ориентиры, которые будут стимулировать технический прогресс и в то же время определять границы допустимого влияния и вмешательства ТИИ в ключевые сферы жизни человека и общества (когнитивные функции, интеллект, личностные качества, социальные навыки, поведение).

Разработка научно выверенных подходов в этой сфере позволит создать среду доверия к ТИИ и сообществу их разработчиков со стороны российских и зарубежных потребителей, а практическое применение этических стандартов при проектировании и разработке отечественных ТИИ повысит их конкурентоспособность на мировых рынках.

Особенности работы Центра:

- междисциплинарный подход и работа на стыке технических и гуманитарных дисциплин (нейробиология, нейроинформатика, медицина, право, психология, философия, антропология, филология и др.);
- универсальный характер разработок Центра и возможность их межсекторального применения на уровне регуляторов, разработчиков, эксплуатантов, пользователей и других акторов ИИ;
- новизна подходов, основанная на научной обоснованности понимания предмета этики ИИ, методик классификации рисков и систем прогнозирования последствий разработки, внедрения и использования ТИИ.

Деятельность Центра будет направлена на:

- фундаментальное исследование влияния ТИИ на когнитивные функции и естественный интеллект человека, состояние общества и эволюционные процессы в различных сферах человеческой деятельности;
- научное исследование рисков, связанных с разработкой, внедрением и использованием ТИИ, а также моделей прогнозирования последствий и сценариев развития ТИИ;

- определение базовых ценностей и этических ориентиров для различных акторов ИИ.

В ходе исследований Центра разрабатываются:

- научные методы прогнозирования последствий и сценариев развития ТИИ и оценки влияния ИИ на различные сферы жизни человека и общества, когнитивные функции и личностные характеристики человека; определены механизмы профилактики и предотвращения деструктивного влияния ТИИ на человека и общество;
- научно выверенная классификация рисков, принципы и методики риск-ориентированного подхода к развитию «слабого» и (в перспективе) «сильного» ИИ в рамках морально-нравственных и этических вопросов, недискриминации, обеспечения прав человека и т.д.;
- методики и критерии оценки этичного поведения акторов ИИ на всех этапах жизненного цикла ТИИ;
- подготовка образовательных модулей по вопросам этики в сфере ИИ.

Продукты деятельности Центра, содержащие научно обоснованные этические подходы и ориентиры, будут востребованы и доступны для практического применения представителями государства, бизнеса, экспертного и научного сообщества, а также широкой общественности, в том числе международной.

Они могут служить основой при разработке государственными органами документов стратегического планирования, нормативно-правовых актов и иных инструментов регулирования и саморегулирования в сфере ИИ.

14. Направление «Искусственный интеллект для обеспечения кибербезопасности»

Одним из приоритетов научно-технологического развития Российской Федерации является противодействие техногенным, биогенным, социокультурным угрозам, терроризму и идеологическому экстремизму, а также киберугрозам и иным источникам опасности для общества, экономики и государства. Этим обуславливается предметная область работы Центра, заключающаяся в разработке методов и технологий мирового уровня, необходимых для мониторинга и противодействия атакам как в компьютерных и киберфизических системах и сетях, включая Интернет вещей (IoT), так и в системах социального взаимодействия, таких как мессенджеры, социальные сети и т.д.

Деятельность Центра затрагивает весь спектр прикладных задач и систем с применением ИИ, включая системы обработки сетевого трафика,

событий безопасности, связей между сетевыми и социальными объектами, текстовой визуальной, звуковой и другой информации, а также системы контроля разграничения доступа, анализа и прогнозирования поведенческой активности и защищенности информации. Ключевыми в деятельности Центра также могут являться процессы интеллектуального управления распределенной обработкой больших объемов данных, в том числе на основе свободного программного обеспечения.

Работа Центра в области анализа данных и управления процессом анализа может, в том числе, быть направлена на разработку методик и прикладных инструментов для обеспечения кибербезопасности самой системы анализа и управления.

Постоянный рост объема, разнородности и изменчивости информации, генерируемой компьютерными и социальными информационными системами, а также требования к оперативности обработки этой информации в условиях лавинообразно возрастающего количества меняющихся по алгоритмам своей реализации атак (программно-информационных воздействий) обуславливают необходимость применения новейших методов ИИ для повышения эффективности мониторинга и противодействия угрозам кибербезопасности. Кроме того, отличительной особенностью предметной области Центра является то, что необходимо не просто найти закономерности в исходных данных, а требуется обнаружение с высокой достоверностью нарушителей, которые специально скрывают свою активность. Кроме того, внедрение технологий искусственного интеллекта определяет необходимость учета ряда возникающих рисков и новых угроз. Так, нарушители могут учитывать особенности применяемых методов ИИ для целенаправленного снижения их эффективности.

Главным результатом может стать комплекс теоретических и практических результатов, основанных на технологии ИИ для мониторинга и противодействия киберугрозам. Данный комплекс должен включать в себя:

- концепцию предлагаемого комплекса для мониторинга и противодействия киберугрозам;
- комплекс моделей, методик и алгоритмов на базе методов ИИ для мониторинга и противодействия киберугрозам в компьютерных сетях;
- комплекс моделей, методик и алгоритмов на базе методов ИИ для мониторинга и противодействия киберугрозам в киберфизических сетях, включая IoT;
- комплекс моделей, методик и алгоритмов на базе методов ИИ для мониторинга и противодействия киберугрозам в социальных информационных сетях (мессенджерах, социальных сетях и т.д.);

- архитектура программного комплекса на базе разработанных теоретических результатов;
- программный комплекс, базирующийся на разработанных теоретических результатах, предназначенный для мониторинга и противодействия киберугрозам в компьютерных, киберфизических и социальных информационных сетях, работающий как в интерактивном, так и в автоматическом режиме;
- пользовательский интерфейс с системой разграничения доступа для обеспечения возможности управления разработанным программным комплексом;
- программа и методики экспериментальной оценки разработанного программного комплекса;
- программные инструменты и наборы данных для оценки соответствия разработанного программного комплекса предъявляемым требованиям, а также результаты экспериментальных исследований;
- научно-технические предложения по внедрению результатов Центра на практике, а также описание путей масштабирования использования комплекса в интересах новых потенциальных заказчиков;
- описание этических аспектов внедрения результатов Центра с указанием возможных рисков применения систем искусственного интеллекта и мер по недопущению их возникновения;
- научно-технические предложения по внедрению результатов Центра в рамках образовательных программ, учебных курсов, методических материалов и т.д.;
- публикации в высокорейтинговых рецензируемых изданиях, выступления на конференциях в Российской Федерации и за её пределами и распространение информации в региональных и федеральных СМИ;
- охраноспособные результаты, зарегистрированные в ФИПС.

Результаты работы Центра могут применяться для построения ситуационных центров и иных органов управления кибербезопасностью, позволяющих сократить ущерб и затраты государственных структур и коммерческих организаций за счет эффективного мониторинга и противодействия киберугрозам. Защищаемой средой при этом могут быть как компьютерные и киберфизические системы и сети, так и социальные информационные системы, такие как мессенджеры, социальные сети и т.д.