

Полякова Елена Георгиевна
Студентка СПб ГБПОУ «СПбТОТФиП»
Руководитель – Быстрова Елена Адольфовна
преподаватель СПб ГБПОУ «СПбТОТФиП»

Секция: Цифровая трансформация производства

Тема: Экономический потенциал и риски при внедрении
искусственного интеллекта

Содержание

Введение.....	3
Искусственный интеллект – драйвер изменений экономики.....	4
Литература.....	8

В первой половине XVII Томас Гоббс заявил, что мышление – это вычисление, а Блез Паскаль создал Паскалину, одну из первых счетных машин. Таким образом, появление искусственного интеллекта стало вопросом времени.

Искусственный интеллект - это процесс познания и постоянного обучения.

В последние годы развитие искусственного интеллекта получило большую динамику. ИИ уже внес большой вклад в прогресс общества, начал преобразовывать методы производства и изменять способы мышления в человеческом обществе.

Искусственный интеллект представляет собой особым образом настроенные технологические решения, которые позволяют имитировать когнитивные функции человека и получать результаты, сопоставимые с реализацией его интеллектуальной деятельности.

Ключевое отличие систем ИИ от обычных программных средств заключается в способности ИИ к обучению. Системы на основе ИИ обучаются, извлекая закономерности из данных и настраивая внутренние скрытые параметры, необходимые для получения решения.

ИИ постепенно становится технологией общего назначения и способен трансформировать бизнес-процессы и рыночные практики организаций. Внедрение ИИ может способствовать росту эффективности бизнеса участников рынка, в том числе за счет снижения издержек, ускорения процессов, ресурсной оптимизации, обработки больших массивов данных, а также повышения лояльности клиентов, благодаря более качественным и таргетированным предложениям продуктов и сервисов.

Перечень сфер применения ИИ постоянно расширяется. С учетом экспансии ИИ, роста глубины интеграции данной технологии в бизнес-процессы организаций он может оказать существенное влияние на развитие российской экономики.

Дальнейшая реализация потенциала технологии ИИ на российском рынке зависит от повышения доступности и качества данных, развития и повышения доступности программного обеспечения и вычислительной инфраструктуры, развития научной базы в области ИИ, развития профессиональных компетенций в области ИИ, обеспечения доверия граждан к технологиям ИИ, развития конкуренции, а также от создания регуляторных условий, способствующих развитию технологии ИИ.

Искусственный интеллект (ИИ) – это системы и сервисы, в основе которых лежат модели машинного обучения; в свою очередь, машинное обучение (ML) – системы, которые самостоятельно формулируют правила принятия решений в процессе обучения на данных на основе заданной человеком архитектуры такой системы. Построение моделей машинного обучения возможно тогда, когда возможна формулировка задачи для модели – предсказание числового значения, категоризация данных и т. д.

К областям технологий искусственного интеллекта относят:

- 1). компьютерное зрение (CV) - область технологий, объединяющая решения для распознавания объектов реального мира. Решения на основе компьютерного зрения встречаются повсеместно – к ним относится распознавание лиц камерой смартфона, сканирование штрихкодов в магазинах, распознавание номеров машин камерами на дорогах;
- 2). обработка естественного языка (NLP) - область технологий, позволяющая компьютерным системам анализировать и интерпретировать человеческий язык. Используется, например, в чат-ботах в клиентской поддержке, в поисковых системах при поиске текстов, для задач машинного перевода, в голосовых или текстовых ассистентах;
- 3). речевые технологии – это технологии, которые позволяют работать с человеческой речью, состоят из двух блоков – распознавание речи и синтез речи., используются в голосовых ассистентах, таких как Siri или Алиса;
- 4). рекомендации и персонализация — это технология, позволяющая сделать пользователю наилучшее предложение исходя из его характеристик, используется в маркетинге, рекламе и продажах, преимущественно в цифровых отраслях;
- 5). генеративные модели — это модели, которые при помощи данных, использовавшихся для обучения, создают новые данные, используются для разного типа задач в бизнесе, преимущественно связанных с генерацией текста или изображений, примером таких моделей являются GPT-4 (OpenAI), YandexGPT (Яндекс), GigaChat (СБЕР).

Искусственный интеллект занимает все более значимое место в мировой экономике. По оценкам международных экспертов, потенциал влияния искусственного интеллекта на экономику составляет 17–26 трлн долл. США в год:

- около 70% приходится на внедрение традиционного ИИ (машинного обучения, глубокого обучения и продвинутой аналитики),

- около 30% на генеративный ИИ, который создает как прямой эффект для отраслей, выражающийся во внедрении новых продуктов и сервисов, так и дополнительный эффект за счет роста продуктивности сотрудников, использующих генеративный ИИ в работе для решения базовых задач.

С точки зрения отраслей внедрения решений около 75% рынка сосредоточено в шести ключевых индустриях с наибольшим потенциалом внедрения моделей традиционного ИИ – банки и финансовые институты, здравоохранение, медиа и реклама, транспорт и логистика, производство, ритейл. В России в топ входят также добывающая промышленность и ИТ-отрасль, которые вытеснили из топ-6 здравоохранение, а также медиа и рекламу. Полный экономический потенциал искусственного интеллекта в России составляет 22–36 трлн руб., к 2028 г. реализованный эффект на рост выручки и сокращение затрат компаний может составить 4,2–6,9 трлн руб. (среднее внедрение ИИ среди компаний России – около 20%), что эквивалентно влиянию на ВВП до 4%. Эффект от генеративного ИИ составит от этого значения 0,8–1,3 трлн руб. (около 20%).

Важно отметить, что внедрение ИИ с высокой долей вероятности будет иметь более широкие последствия для страны, чем непосредственно экономический эффект, – например, повышение качества и продолжительности жизни (за счет развития превентивной медицины), снижение числа ДТП и травматизма на производствах (автономный транспорт, автономные склады, фабрики и т. д.), повышение качества образования (высокоадаптивные образовательные системы на базе ИИ), создание новых рабочих мест (за счет появления новых профессий, например AI-тренеров и промпт-инженеров) и повышение производительности труда. Глава Минцифры РФ Максуд Шадиев заявил, что искусственный интеллект будет основным трендом цифровизации в России до 2030 года.

Нельзя не отметить то, что при внедрении решений на базе ИИ компании сталкиваются со сложностями, среди которых ключевыми являются:

1). дефицит профильных специалистов (дата-сайентистов, дата-инженеров, дата-аналитиков). Согласно заявлению Максуда Шадиева, главы Минцифры, общий дефицит кадров в российской ИТ-отрасли – около 500–700 тыс. человек. А по заявлениям Германа Грефа, дефицит составляет более 1 млн специалистов;

2). высокий уровень затрат на разработку ИИ-инструментов: стоимость разработки ПО и сервисов на основе ИИ в России с июня 2022 г. по июнь 2023 г. увеличилась на 30–40%, по оценкам разработчиков таких систем (июль 2023 г.). Основная причина – рост стоимости и сложности с закупками иностранного оборудования, например графических процессоров NVIDIA.

Рост технологий ИИ во всех сферах экономики ставит и такой важный вопрос как этические риски внедрения искусственного интеллекта. К важнейшим рискам относятся:

1). Несовершенство искусственного интеллекта. Искусственный интеллект может принимать предвзятые решения, в основном вследствие предвзятости данных, на которых был обучен. Обучающая выборка может быть неполной, нерепрезентативной или уже содержать элементы предвзятости. Например, еще в 2018 г. Amazon приостановил использование инструмента для автоматического скрининга резюме соискателей – экспериментальный алгоритм использовал ИИ, чтобы выставлять соискателям оценки от одной до пяти звезд. Ключевая причина – алгоритм не был гендерно-нейтральным. Модель Amazon была обучена отбирать кандидатов на базе воронки найма за 10-летний период, в который большинство резюме поступало от мужчин. Бороться с предвзятостью можно путем улучшения качества исходных данных и полноты выборки.

2). Галлюцинации генеративных моделей. Генеративные модели, даже уровня GPT-4, могут «галлюцинировать», то есть предоставлять некорректные с фактологической точки зрения ответы. Галлюцинации, например, активно обсуждали после презентации чат-бота Bing от Microsoft. Bing сравнивал финансовые отчеты двух компаний. Часть цифр оказалась «сгаллюцинированной», то есть подобных значений не было в финансовых отчетах, которые модель использовала для ответа на пользовательский запрос. Проблема «галлюцинаций» – следствие не только качества данных для обучения, но и вероятностной природы модели, и сейчас отсутствуют методы, гарантирующие ее стопроцентное устранение. Однако вероятность галлюцинаций можно существенно уменьшить за счет правильных промптов (текстовых запросов) и встраивания автоматической проверки на адекватность. Ожидается, что проблему галлюцинаций генеративных моделей удастся решить за 1,5–2 года. Специалисты по машинному обучению предлагают разные решения проблемы. Среди них – создание специализированных наборов данных, использование обучения с подкреплением на основе человеческих предпочтений, использование поисковой нейросети, которая будет разбирать и анализировать внешнюю информацию для генеративной нейросети, и др. Пользователи также могут повлиять на снижение вероятности галлюцинаций, даже при работе с моделями напрямую в режиме чата: давать контекст и четкие инструкции при запросе, формулировать точные, недвусмысленные промпты, разбивать сложные задачи на более легкие, давать обратную связь на ответы модели, уточняя исходный запрос.

3). Осознанное стремление пользователей добиться от моделей ложных или неэтичных ответов. Есть люди, которые готовы взломать технологии, чтобы сгенерировать вредоносный контент. Так, актуальна проблема перехвата работы модели – осознанного

стремления пользователя подавить механизм безопасности модели за счет специально сконструированных запросов. Идея заключается в том, чтобы изменить или добавить промпты, которые применяются для обучения модели, таким образом, чтобы вводные данные были искажены и выходные данные не соответствовали им. Это может привести к некорректной обработке данных моделью и, следовательно, к ошибочным результатам. Последствия – распространение дезинформации, фейковых новостей, нежелательных и вредоносных текстов.

Чем более популярной становится технологии ИИ, тем более возрастает значимость вопросов этики. Этические риски, связанные с природой ИИ, скорее всего, никуда не исчезнут. Вероятность галлюцинаций в ближайшие 5–10 лет будет неуклонно снижаться, но не опустится до 0%, как и возможности перехвата работы модели, хотя взломать модели станет сложнее.

Несмотря на все проблемы, для того чтобы извлечь максимум пользы из развития технологий, развитие искусственного интеллекта стоит воспринимать как стандартный технический прогресс.

Литература

- 1). Бритвина П.В. «Этические вопросы в применении искусственного интеллекта и машинного обучения» // Международный научный журнал «ВЕСТНИК НАУКИ» № 1 (70) Том 4. ЯНВАРЬ 2024. с. 445-447.
- 2). Лешкевич Т.Г. «Парадокс доверия к искусственному интеллекту и его обоснование» // Философия науки и техники. – Москва, 2023. – Т. 28, № 1. – с. 34-39
- 3). Любимов А.П., Майстренко Г.А. «Перспективы искусственного интеллекта в России и за рубежом» // Философия науки и техники. – Москва, 2023. – Т. 28, № 1. – с. 121–132
- 4). Миндигулова А.А. «Этика и искусственный интеллект: проблемы и противоречия» // МЕДИЦИНА. СОЦИОЛОГИЯ. ФИЛОСОФИЯ. Прикладные исследования. №3 2022. с.146-149
- 5). Ю. В. Назарова Ю.В. «Этика искусственного интеллекта в современной России: актуальные проблемы и тенденции развития» // Гуманитарные ведомости ТГПУ им. Л. Н. Толстого № 2 (34), октябрь 2022г. с. 14-23
- 6). «Искусственный интеллект в России – 2023: тренды и перспективы // Отчет консалтинговой компании «Яков и Партнёры». URL: <https://www.yakov.partners/publications/ai-future/>