

ОЦЕНКА ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ПРЕДПРИЯТИЯХ

Аникин Максим Андреевич – аспирант, Институт проблем управления имени В.А. Трапезникова РАН, Москва, Россия, maxim.anikin1@yandex.ru

Аннотация. Искусственный интеллект (ИИ) быстро проникает в разные сферы жизни человека. В настоящее время данному феномену уделяется особо пристальное внимание ввиду его популярности, вызванной созданием чат-ботов, которые способны поддерживать разговор с человеком, генерируя правдоподобные ответы на сообщения. Популярность любой стремительно развивающейся технологии неизбежно приводит к постановке новых вопросов и проблем, что в свою очередь говорит о необходимости тщательного изучения влияния ИИ на человека и практики его применения.

Быстрый рост данной технологии вызывает заблуждение относительно «умственных способностей» ИИ, из-за которого возникает мнение о наличии «разума» у современных программ. Для рассмотрения данного вопроса было проведено сравнение естественного и искусственного интеллекта и сделан вывод о степени их сходства и возможностей.

В работе основное внимание уделено использованию ИИ на предприятиях. Рассмотрены экономические и этические последствия применения, а также психологические факторы, играющие роль при взаимодействии человека и ИИ в виде программного обеспечения. Помимо этого, дана оценка целесообразности применения ИИ, а также сделано предположение о необходимости понимания принципов его функционирования для эффективного взаимодействия.

Ключевые слова: искусственный интеллект (ИИ); эффекты применения; человеко-машинное взаимодействие; предприятия.

Для цитирования: Аникин М.А. Оценка применения искусственного интеллекта на предприятиях // Науковедческие исследования. – 2026. – № 1. – С. 17–33. – DOI: <http://www.doi.org/10.31249/naukoved/2026.01.02>

Статья поступила в редакцию 09.07.2025

Принята к публикации 19.02.2026

ASSESSMENT OF THE USE OF ARTIFICIAL INTELLIGENCE IN ENTERPRISES

Maxim A. Anikin – graduate student, Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia maxim.anikin1@yandex.ru

Abstract. Artificial intelligence (AI) is rapidly penetrating into different areas of human life. Currently, special attention is being paid to this phenomenon due to its popularity caused by the creation of chatbots that are able to maintain a conversation with a person, generating plausible responses to messages. The popularity of any rapidly developing technology inevitably leads to the raising of new questions and problems, which in turn suggests the need for a thorough study of the impact of AI on humans and its application practices. The rapid growth of this technology causes misconceptions about the “mental abilities” of AI, which leads to the opinion that modern programs have a “mind”. To consider this issue, a comparison of natural and artificial intelligence was carried out, and an appropriate conclusion was drawn about the degree of their similarity and capabilities. The paper focuses on the use of AI in enterprises. The economic and ethical consequences of the application, as well as the psychological factors that play a role in the interaction of humans and AI in the form of software, are considered. In addition, an assessment of the feasibility of using AI is given, and an assumption is made about the need to understand the principles of its functioning for effective interaction.

Keywords: artificial intelligence (AI); application effects; human-machine interaction; enterprises.

For citation: Anikin M.A. Assessment of the use of artificial intelligence in enterprises. *Naukovedcheskie issledovaniya = Science Studies*. 2026, N 1, P. 17–33. DOI: <http://www.doi.org/10.31249/naukoved/2026.01.02>

The paper was submitted 09.07.2025

Accepted for publication 19.02.2026

Введение

Системы ИИ все чаще используются в различных сферах человеческой деятельности, будь то медицина, логистика или производство. Особое внимание к этому феномену обращено в последнее десятилетие, после появления чат-ботов, способных правдоподобно имитировать человеческую коммуникацию. Широкая популярность данной технологии вызывает необходимость внимательно и объективно изучать, каким образом она влияет на жизнь людей, какие обнаруживаются возможности и риски ее использования.

В дискуссиях о создании и использовании ИИ, в том ракурсе, который нас интересует в данной работе, можно выделить два подхода. Сторонники первого считают, что обязательной предпосылкой интеллекта являются внутренние мыслительные процессы и рассуждения. Сторонники второго утверждают, что достаточно наблюдать внешнее поведение объекта для определения его интеллектуальности. Далее мы разберем оба

подхода и сделаем выводы. Также необходимо уточнить, что далее рассматриваются именно нейронные сети, хотя, строго говоря, ИИ не ограничивается только нейросетями.

В данной работе мы рассмотрим применение ИИ на предприятиях, тем не менее, некоторые из тезисов будут справедливы и для других сфер человеческой деятельности. Помимо экономических факторов в работе рассмотрены психологические аспекты, учет которых необходим при взаимодействии ИИ с человеком посредством программного обеспечения.

Таким образом, целью данной работы является оценка целесообразности применения ИИ на предприятиях.

Задачи данной работы: выявить основные характеристики ИИ в сравнении с естественным интеллектом; определить потенциальные положительные и отрицательные эффекты применения ИИ на предприятиях; привести примеры актуальных исследований в области взаимодействия человека и ИИ; дать оценку полезности и применимости ИИ на предприятиях.

ИИ и внутренние мыслительные процессы

Перейдем к первой задаче, поставленной в данной работе. Наличие внутренних мыслительных процессов может рассматриваться в качестве предпосылки наличия интеллекта. Попробуем описать внутреннее устройство нейросетей на уровне, достаточном для ответа на вопрос, как они «рассуждают», и какие процессы происходят внутри программы. Не погружаясь в детали технической реализации, работа нейросетей устроена следующим образом. Сначала подготавливается огромный объем данных, который классифицируется / маркируется / сортируется людьми. После этого подготовленный набор данных передается алгоритму, который мы и называем ИИ. Он обрабатывает полученные данные и создает статистические взаимосвязи между входными и выходными данными. Например, для того чтобы генерировать текстовый ответ на запрос, создается множество правил типа «Если до этого были слова А и Б, то нужно далее поставить слово В». Современные вычислительные способности позволяют нейросетям обрабатывать огромные объемы данных и, что не менее важно, сохранять в своей памяти большой объем контекста, что позволяет чат-ботам правдоподобно отвечать на вопросы людей.

Иными словами, мы имеем дело со сложным статистическим алгоритмом, который способен после обработки миллиардов последовательностей слов выявить закономерности их расположения и использовать эти закономерности для генерации новых последовательностей слов. Из этого легко сделать вывод, что ИИ не может создать новые знания, так как не может выйти за рамки тех статистических взаимосвязей, которые были получены за счет алгоритма. И даже если добавить вероятностную характеристику, которая позволит использовать не самое подходящее по стати-

стике слово, а какое-нибудь другое, то это не более, чем добавление разнообразия в генерируемые последовательности, а в некоторых случаях добавление бессмысленного информационного шума. Можно утверждать, что у ИИ нет функционала, аналогичного человеческому разуму. Алгоритм оперирует лишь словами, а не их смыслом, что в свою очередь подвергает сомнению гипотезу о наличии интеллекта у современных компьютерных программ.

ИИ и внешнее поведение

Перейдем к аргументам сторонников подхода, который за основу наличия интеллекта принимает внешнее поведение объекта. Когда дело касается вопроса о поведении машины, наиболее известным тестом является тест Тьюринга. Описание этого теста следующее. В процессе взаимодействия человека посредством текста с другим человеком и компьютером, он должен определить по их ответам на свои вопросы разговаривает он с компьютером или с человеком [Turing, 1950]. Если судья не может сказать, кто является компьютером, а кто человеком, то считается, что машина прошла тест и может мыслить.

В противовес этому стоит упомянуть об известном мысленном эксперименте – китайской комнате Джона Сёрла, целью которого является опровержение гипотезы сильного ИИ и теста Тьюринга, в частности. Для проведения этого мысленного эксперимента нужно представить изолированную комнату, в которой находится человек, не понимающий китайского языка (и в целом иероглифов), но имеющий подробную инструкцию о том, как оперировать иероглифами без объяснения их смысла. Человек снаружи комнаты, знающий китайские иероглифы, передает в комнату сообщение с иероглифами и ожидает получить осмысленный ответ. Следуя инструкции, человек внутри комнаты составляет ответ, который покажется человеку снаружи осмысленным, и таким образом он будет считать, что ведет диалог с человеком, знающим китайский язык. При этом у человека внутри комнаты нет возможности понять смысл составленных им или полученных извне сообщений. Этот эксперимент описывает ситуацию, аналогичную общению человека с чат-ботом на основе ИИ. У нейронных сетей есть только та самая «инструкция» по обращению со словами, но даже только благодаря ей она способна выдавать ответы, похожие на человеческие. Из-за такого результата некоторые люди начинают считать, что ведут разговор с объектом / программой, у которой действительно есть интеллект, хотя это не так.

Кроме указанных существуют и другие погрешности работы ИИ. Начнем с галлюцинаций. Этот термин часто применяется по отношению к чат-ботам в те моменты, когда они начинают выдавать совершенно бессмысленные ответы. Например, рецепт пасты из дерева и свинца. Ввиду

того, что нейросети действуют по принципу «черного ящика», трудно ответить на вопрос о конкретных причинах появления галлюцинаций, хотя соответствующие работы ведутся [Ferrando, Obeso, Rajamanoharan & Nando, 2025; Lindsey et al., 2025]. Основными предположениями являются следующие:

1) плохое качество данных, на которых обучалась модель. Если при обучении нейронная сеть получила противоречивые / недостоверные данные об одном и том же факте и если при разметке данных не определялись уровни доверия к источникам (например, пост в социальной сети по уровню доверия равен научной статье), то при ответе на вопрос вполне возможна ситуация, при которой будет дан неправильный или бессмысленный ответ;

2) генерация ответов на основе статистической правдоподобности. Это является фундаментальным ограничением нейросетевых моделей. Если мы спросим, каким спортом профессионально занимался известный музыкант, то ИИ может сгенерировать правдоподобный ответ типа «Музыкант занимался боксом», хотя такой информации в общем доступе не имеется. Причиной этому может быть наличие в вопросе слова «спорт», что при генерации ответа делает последовательность слов о том, что кто-то занимался боксом статистически правдоподобной;

3) неправильная / отсутствующая проверка фактов. У наиболее развитых моделей также присутствует функция проверки ответа по базе знаний или при помощи Интернета. Тем не менее такой функционал наблюдается не везде, а в некоторых случаях реализован некорректно, что также приводит к галлюцинациям.

Помимо текста рассмотрим интересный эффект ошибочного распознавания нейросетью предметов на изображении. Есть множество исследований, ориентированных на то, чтобы изменять изображения при помощи ИИ, и получать результаты, которые являются совершенно некорректными. Например, в одном из них [Bose, Aarabi, 2018] использовался метод, который позволяет скрыть от нейронной сети лицо человека на изображении, немного изменив оригинальную фотографию. При этом реальный человек даже не заметит разницы между оригиналом и подправленной версией и без труда определит лицо на обеих фотографиях. Аналогично существуют способы заставить классифицировать логотипы компаний как знаки «стоп» и наоборот [Sitawarin C. et al., 2018]. В этих работах используются различные приемы, позволяющие запутать нейросеть, и все они основаны на ее слабостях и особенностях функционирования. Часто при обучении ИИ он может фиксировать совершенно неожиданные признаки, по которым проводится классификация предметов. Если изображения, на которых обучалась нейросеть, подготовлены неправильно, то это всегда приводит к нелепым ошибкам. Для примера возьмем простую задачу классификации того, кто изображен на фотографии – собака или кошка, но с одним нюансом, который заключается в том, что все фотографии ко-

шек были случайным образом взяты с травой на фоне. В итоге высока вероятность того, что любое изображение, на фоне которого находится трава, будет классифицировано как содержащее кошку, даже если на самом деле там будет собака или что-то совершенно другое. Тут снова вступает в силу особенность обучения ИИ – это всего лишь алгоритм, высчитывающий статистические параметры.

Вышеперечисленные погрешности поведения ИИ берут свое начало в самой его структуре и принципах работы. Эти погрешности позволяют нам видеть, насколько далеки современные алгоритмы от понимания того, какой смысл у тех данных, на которых они обучаются и которые они выдают при запросе. Нужно помнить, что ИИ, как бы удачно он не имитировал отдельные аспекты человеческого поведения, является лишь инструментом, способным ускорить совершение конкретных действий, для которых он был создан.

Сравнение естественного интеллекта и ИИ не является корректным ввиду того, что последний представляет собой лишь компьютерную программу, удачно выделяющую закономерности из огромного количества данных без дальнейшего анализа этих закономерностей и тем более синтеза новых на их основе. Тем не менее это не означает, что ИИ является малоэффективной технологией. Сочетание наличия у человечества огромных вычислительных мощностей за счет эволюции ЭВМ, а также генерации огромного количества данных пользователями Интернета и датчиками на производствах, делает ИИ полезным инструментом на предприятиях, в повседневной жизни, обучении и сфере услуг.

Проблемы взаимодействия человека и ИИ

Ввиду широкой распространенности ИИ трудно встретить человека, который не взаимодействовал бы с ним. Даже если человек не пользовался чат-ботами, то наверняка знаком с другими цифровыми продуктами, которые основываются на ИИ. Это может быть рекомендательная система, подсказывающая, какая музыка или фильм понравится клиенту, или автоматическое улучшение качества фотографий с помощью нейросетей, используемое в некоторых моделях современных телефонов. Множество факторов влияет на то, как человек относится к ИИ. Если при оценке обычным пользователем мобильного приложения можно получить дополнительные данные, которые позволят в дальнейшем улучшить рекомендательные системы или другие составляющие цифрового продукта, то на производстве возникает также вопрос о безопасности других людей при взаимодействии ИИ и контролера технологических процессов. В одной из работ [Венгер, Дозорцев, 2022, с. 10–17] изучается проблема доверия оператора ИИ. Выделяются различные факторы, среди которых выявляются не только технические параметры автоматизированной системы (безопас-

ность, надежность, удобство использования), но и личностные характеристики оператора, такие как профессионализм, уровень самооценки, возраст. Помимо этого, на отношение к ИИ может также влиять и серьезность принимаемого решения, причем как в положительную, так и в отрицательную сторону. Если ситуация не является критической и цена ошибки низка, то склонность доверять решениям ИИ возрастает. Рассмотрим некоторые аспекты вопроса доверия к ИИ, и как их можно улучшить.

Одним из важнейших факторов для повышения доверия является увеличение надежности ИИ. Сбор релевантной для текущего производства информации, ее правильная подготовка и использование современных структур из области ИИ поможет с этой задачей. Тем не менее даже если ИИ будет отлично выполнять свою задачу, он все равно рано или поздно ошибется, так как идеальных систем не существует. И в случае ошибки необходимо не только устранить ее последствия, но также и выяснить причину ее появления. Для расследования отдельных случаев нужно понять, почему было выбрано ошибочное решение и как в дальнейшем его избежать. Таким образом, остро встает вопрос о прозрачности работы алгоритмов ИИ. Для решения данной проблемы существует отдельное направление, называемое объяснимым ИИ. Его целью является разработка методов объяснения результатов, полученных различными алгоритмами ИИ, а также разработка алгоритмов, которые поддерживают такую возможность. В работах [Linardatos, Papastefanopoulos, Kotsiantis, 2021, pp. 1–45; Long B. et al., 2025] представлены основные достижения в данной сфере. В некоторых алгоритмах удалось показать такие части предложений или фотографий (в зависимости от характера входных данных), которые имели наибольший вес при формировании нейросетью конечных результатов. Некоторые алгоритмы признаны необъяснимыми в силу природы их работы. Данная область исследований не теряет своей актуальности и поэтому будет и дальше решать проблемы прозрачности ИИ.

Вероятно, базовые параметры специалиста в ближайшие десятилетия начнут включать не только знания из профессиональной сферы работника, но и навыки по работе с ИИ, а также понимание принципов его работы. Неготовность к работе с ИИ может привести к излишнему доверию или абсолютному недоверию при работе с ним. Обе крайности не являются оптимальными. При безусловном доверии увеличивается вероятность пропуска сбоя алгоритма, а при абсолютном недоверии повышается вероятность ошибки, которая объясняется сложностью контроля процесса только за счет оператора процесса. Необходимо четко осознавать возможность возникновения аварийной ситуации даже под контролем ИИ и быть готовым взять инициативу на себя и сориентироваться в текущей ситуации без помощи этой технологии.

Также на доверие алгоритмам может влиять индивидуальная стратегия принятия решений, которая может описываться упрощенной моделью, включающей в себя субъективно допустимый уровень риска и степень

надежности информации, необходимой для принятия оператором решения. Немаловажной является и цена ошибки. Так, если человек понимает, что неверное решение не понесет за собой серьезное наказание, например, единичный пропуск бракованного материала, то склонность передать решение ИИ возрастает. Ошибка алгоритма практически не повлияет на положение работника, а возможность отдохнуть на рабочем месте может покрывать потенциальные отрицательные последствия. Тем не менее и здесь влияние оказывают личные качества работника, такие как ответственность, исполнительность, внимательность, недоверчивость. Наличие таких качеств подталкивает сотрудника перепроверять решения ИИ и вносить коррективы при необходимости. Поведение человека может кардинально измениться при наличии серьезных рисков (авария на производстве или смерть людей). Цена ошибки будет огромной, и каждое решение необходимо будет дополнительно взвешивать самому, даже если работнику будет дан совет от ИИ. Безусловно, в такой ситуации может оказаться и человек, который проявит халатность и не придаст должного значения аварийной ситуации. Для того, чтобы человек наиболее эффективно и гармонично взаимодействовал с «компьютерным советником», вероятно, нужно будет проводить дополнительное психологическое тестирование, которое позволит выявить наиболее ответственного кандидата, участвующего в процесс принятия решений и не слепо доверяющего решениям ИИ в критических случаях. Стоит заметить, что до этого рассматривались ситуации, где человеку приходит предупреждение или совет от ИИ, но возможна ситуация, при которой никакого сообщения по поводу возникновения аварии может и не быть, т.е. имеет место пропуск важного события. Специалист, работающий с алгоритмом, обязан следить за критическими параметрами системы и делать свои выводы параллельно с ИИ. Таким образом, оператор страхует алгоритм и наоборот. Программа способна обрабатывать огромные объемы информации быстрее, чем человек, и человек может отследить закономерности и ситуации, которые не были заложены в программу при ее обучении.

В бытовых ситуациях коммуницирование с ИИ происходит не только посредством чата, но и с помощью голоса. Плюсом данного технического достижения является естественность формирования запросов к различным умным поисковым системам. Вместо того, чтобы задумываться, как лучше задать вопрос или какую инструкцию использовать, чтобы программа тебя поняла, можно просто задать вопрос так, будто ты говоришь с человеком. Это ускоряет поиск необходимой информации в увеличивающихся до огромных масштабов базах данных. Также это позволяет уменьшить порог вхождения для использования технологии. Намного легче просто задать вопрос, чем изучать несколько десятков страниц документации, которая в свою очередь может предполагать наличие знаний из технических областей. Но сочетание естественных способов общения с похожими на человеческие ответами от ИИ склоняет людей наделять его

человеческими качествами. Это, безусловно, является одним из факторов, влияющих на отношение людей к ИИ. Похожесть на нас может влиять как положительно, так и отрицательно. При довольно сильной, но неидеальной похожести на человека возникает известный эффект «зловещей долины» [Masahiro, MacDorman, Kageki, 2012, pp. 98–100], который выражается в резком падении привлекательности объекта. Тем не менее из-за того, что ИИ представлен не в виде антропоморфного робота, а в виде программы на мобильном телефоне / компьютере, предмета техники (умная колонка, холодильник, чайник и т.д.) или игрушки, то такого явного отторжения не возникает. Вероятно, люди склонны представлять наиболее привлекательный образ ИИ и в какой-то степени привязываться к нему. Конечно, стоит отделять придуманный образ от реального объекта и не строить иллюзий по поводу интеллектуальности или даже эмпатии со стороны ИИ. В особенности важно доносить эту идею людям, которые плохо разбираются в технологиях или склонны доверять броским новостным заголовкам.

Тема взаимодействия ИИ и человека интересна в основном тем, что обращение с технологиями переходит из плоскости «субъект-объект» в плоскость «субъект-субъект», где ИИ наделяется равными человеку способностями, а в некоторых случаях речь заходит даже о юридических правах компьютерной программы. Предположительно, данный эффект возник из-за того, что множество людей оказались не готовы к возможности имитации человеческой деятельности (а, если быть более точным, имитацией результата этой деятельности) с такой точностью и ошибочно приняли это за признак интеллекта, сходного или альтернативного человеческому. Тем не менее, в том числе благодаря такой реакции, ведутся дополнительные работы по исследованию феномена ИИ и набирается база фактов, которые помогут в будущем более корректно относиться к такой технологии. Пока доподлинно неизвестно, возможно ли создать «сильный» ИИ в виде компьютерной программы, но можно утверждать, что тот ИИ, с которым мы сейчас взаимодействуем в наших домах и на рабочих местах, крайне далек от естественного интеллекта. Безусловно, данная технология меняет наш мир, и сейчас общество проходит этап адаптации к этим изменениям. Информационное поле и дальше будет пестрить новостями об очередных интеллектуальных прорывах компьютерных алгоритмов. По крайней мере до тех пор, пока это не станет обыденностью или не будет объяснен принцип устройства ИИ большинству людей.

Эффекты применения ИИ на предприятиях

Традиционный подход к автоматизации производств заключается в использовании автоматизированных систем управления (АСУ) различных видов. Появление ЭВМ позволило использовать вычислительные мощно-

сти для контроля технологических процессов (ТП) и расчета управляющего сигнала для достижения требуемых значений конкретных параметров процессов, что в последствии, в 60-х годах XX века, привело к созданию АСУ. Формирование новых направлений технической науки и разработка решений, позволяющих комбинировать информационные технологии с наработанным опытом производств, позволили автоматизировать управление различными параметрами ТП без участия человека (не считая контроля успешного выполнения заданной программы). АСУ до сих пор являются стандартным средством достижения целей на многих производствах и постоянно улучшаются.

Одним из актуальных направлений улучшения АСУ является внедрение ИИ. Использование данной технологии позволяет решать задачи [Эффективные отечественные ...], пока не поддающиеся аналитическим методам, такие как организация предиктивного технического обслуживания станков и его отдельных узлов [Turanoglu, Nyqvist, Skoogh, 2020, pp. 1–14], оценка состояния выходного изделия [Интеллектуальная система ... , 2022] или создание цифровых двойников [Huang Z. et al., 2021] производства для прогнозирования состояния и предложения вариантов управления процессами. Стоит обратить внимание, что даже указанные задачи могут решаться аналитическими методами, но только в ограниченном количестве случаев, когда существует соответствующая исследовательская база, а также возможность установки датчиков на конкретном станке, как физическая, так и экономическая. Использование ИИ подкупает тем, что он может выявить закономерности, потенциально заложенные в собираемых данных уже существующих датчиков или датчиков, интеграция в производство которых не вызовет трудностей. ИИ стоит рассматривать не как замену предыдущего набора инструментов в виде АСУ ТП, а как ее очередной виток развития. Распространение использования ИИ на предприятиях говорит о том, что данная технология имеет преимущества по сравнению с системой, которая его не использует. Тем не менее у каждой технологии есть свои минусы, а высокая популярность и скорость, с которой различные алгоритмы ИИ внедряются на предприятиях, диктуют необходимость исследования не только положительных эффектов его использования, но и отрицательных или даже опасных. В данной части мы рассмотрим различные последствия внедрения ИИ, сделав чуть больший упор на социальные и этические последствия.

С экономической точки зрения внедрение ИИ на производственных предприятиях значительно повышает его эффективность и объемы производства [Dejanovic, 2025]. Обслуживание производственных линий может занимать от 15 до 60 процентов от общих затрат на все производственные операции. Для уменьшения этих затрат были разработаны системы, позволяющие прогнозировать необходимость технического обслуживания и тем самым минимизировать время простоя. При правильной настройке ИИ позволяет более точно отслеживать брак на производстве, что, например,

позволило сэкономить компании Bosch 1,3 млн долл. за два года работы [Mikil, Kunal, 2022]. Положительный экономический эффект использования ИИ объясняет повсеместность его распространения и вложения в его развитие по всему миру.

А теперь рассмотрим некоторые негативные последствия применения ИИ. Одним из них является предвзятость алгоритмов [Харитонов, Савина, Паньини, 2021, сс. 488–515]. Предвзятость алгоритмов по отношению к конкретным группам лиц в большинстве случаев вызвана неправильной подготовкой данных, на которых алгоритм обучался. Предвзятость, например, может быть выражена в уменьшении шансов на получение работы лицами, которые имеют определенную расовую принадлежность или социальное положение в обществе.

Учитывая, что для работы ИИ необходимо собрать большое количество данных, возникает риск нарушения конфиденциальности не только в процессе сбора и хранения, но также и при их обработке [Bartneck C. et al., 2021, pp. 61–70]. Необходимо понимать, что при обучении ИИ важной является не только задача обучить алгоритм, выполняющий определенный круг задач, но и обеспечить безопасность используемых данных, т.е. предотвращение утечек и несанкционированного доступа лиц, не имеющих на это права.

Помимо этого, обострилась проблема ответственности за вред, причиняемый в результате использования ИИ. До этого редко вставал вопрос о том, кто ответственен за возникновение аварийной ситуации, так как задачи условно делились на полностью автоматизируемые и частично автоматизируемые / неавтоматизируемые. В случае полностью автоматизируемых ситуаций существовал четкий алгоритм действий системы при конкретных условиях, который обходился без участия человека. Частично автоматизируемые задачи изначально предполагали, что анализирующий алгоритм или инструмент является вспомогательным для человека, и именно человек несет ответственность за все предпринятые им действия. Появление ИИ привело к возникновению ситуаций, в которых алгоритм ИИ лучше справляется с задачей, чем среднестатистический сотрудник, и, тем не менее, может ошибиться. Иными словами, использование ИИ, по статистике, безопаснее и выгоднее, но в моменты, когда аварийную ситуацию все-таки не удастся предотвратить, возникает вопрос: кто же ответственен за аварию, если решения принимались лишь алгоритмом? Безусловно данный вопрос обостряется еще и потому, что большинство алгоритмов ИИ работают по принципу «черного ящика», что делает крайне затруднительным, а в некоторых случаях невозможным, получение объяснения принятия конкретного решения при данных условиях. Иными словами, получается, что безопасность и жизни людей зависят от алгоритма, который принимает решения необъяснимым нам образом, хотя при этом делает это надежнее, чем другие сотрудники. Несмотря на активную разработку этого вопроса [Апостолова, 2021, сс. 112–119], пока невозмож-

но создать алгоритм на основе технологий ИИ, который будет надежным на всем множестве возможных ситуаций. Поэтому, возможно, стоит относиться к алгоритмам ИИ как к инструментам, которые, как и любая другая система, могут дать осечку и потому требуют внешнего контроля человека.

Далее рассмотрим разные, но частично связанные проблемы: потерю управления и доверия в принятии решений автономных систем и подрыв способности принятия людьми самостоятельных решений. Ранее было отмечено, что алгоритмы могут быть более надежными в решении некоторых задач, чем человек. Тем не менее на производстве частой практикой является назначение оператора какого-то технологического процесса, даже если он управляется автономным алгоритмом. Большинство стандартных решений принимается алгоритмом, человек же принимает участие при возникновении критических ситуаций, а также при наличии какого-либо дефекта, который не был зафиксирован алгоритмом. Однако даже в критических ситуациях некоторые системы ИИ могут выдавать советы по устранению рисков. Учитывая, что система ИИ является лишь интегрированным модулем в конкретной АСУ ТП, совет может выглядеть как привычное всплывающее окно в SCADA, специальной программе, предназначенной для разработки или обеспечения работы в реальном времени систем сбора, обработки, отображения и архивирования информации об объекте мониторинга или управления, с рекомендацией к принятию действия или предупреждением о возникновении нежелательной ситуации [Таха, 2024]. В этом случае оператор должен решить, является ли текущая ситуация действительно критической, и правильный ли совет дала ему система. Во многом это вопрос доверия алгоритмам в целом. Может ли машина рассчитать вариант, который окажется более правильным, чем тот, который рассчитал оператор? Всегда ли она лучше человека? Адекватный ли вариант предложен системой в конкретной ситуации? Чтобы ответить на эти вопросы, оператор должен быть специалистом не только в технической сфере, непосредственно связанной с контролируемым им технологическим процессом, но также пройти дополнительное обучение по работе с автономными системами, при котором будут даны необходимые пояснения о внутренней работе алгоритма. Чрезвычайно важно, чтобы алгоритм воспринимался как инструмент, дающий информацию, а не как полноценная замена специалиста. Если этого не понимать, то может возникнуть опасная ситуация, при которой человек вынужден будет беспрекословно принимать все решения и советы, которые будут приходить от алгоритма [В гибели двух «боингов» ...].

Боязнь замены людей и потери рабочих мест [Okechukwu, Areji, Ugwu, 2023] вызывает сопротивление внедрению ИИ в некоторых сферах деятельности, а также инициирует создание новых законов, предъявляющих требования к безопасности данных и защите интеллектуальной собственности. Расширение сферы применения ИИ заставляет задуматься не только тех, на чью профессиональную сферу уже «посягнули», но и тех,

кто потенциально может быть затронут в ближайшие годы. Профсоюзы уже предпринимают действия [AI vs humans ...] по ограничению влияния набирающей силу технологии, чтобы защитить от потери источника доходов своих работников.

Таким образом, мы понимаем, что внедрение ИИ на предприятиях имеет не только очевидные экономические преимущества, но и вызывает множество дополнительных проблем в этической и социальной сферах, которые нужно исследовать и решать [Багдасарян, 2024, сс. 180–186]. Особое отношение к данной технологии во многом объясняется неправильным пониманием ее специфики. Слово «интеллект» чрезмерно для описания такого типа алгоритмов. Тем не менее современные достижения ИИ, а также реакции СМИ создают образ не полезного инструмента, а «программной сущности», обладающей способностями близкими или даже превосходящими человеческие. С уверенностью можно сказать, что данный феномен будет и дальше изучаться не только с точки зрения экономических и этических соображений, но также внутри- и внешнеполитических. Умение эффективно и правомерно использовать, а также правильно преподносить информацию о данном инструменте позволит достичь еще большей выгоды без тревожащих инфоповодов. Интерес составляет не только глобальное влияние ИИ на сферы деятельности человека или группы людей, но и локальное взаимодействие между человеком и ИИ, о чем мы рассуждали выше.

Заключение

ИИ является по своей сути развитием математического метода по обработке большого количества данных, которая стала возможна благодаря созданию высокопроизводительных вычислительных машин. При должном погружении в техническую реализацию ИИ, отношение к нему как к программе, имеющей способности человеческого разума, становится лишенным смысла и представляется несколько поспешным.

Положительный экономический эффект применения ИИ на предприятиях подкрепляется фактом его стремительного распространения и постепенного внедрения в АСУ производств. Эффективность принятия решений и использования имеющихся у компаний ресурсов становится выше. Тем не менее новые технологии порождают новые испытания для общества. Экономические выгоды, в частности, девальвируются, так как глобальные эффекты применения ИИ, такие как потенциальное лишение мест большого количества работников и вопрос использования чьей-то интеллектуальной собственности для обучения нейросетей вызвали волну сопротивления профсоюзов.

Мы считаем, что необходимо относиться к этой технологии как полезному инструменту. В таком случае мы избежим опасной ситуации,

в которой программа решает за человека, даже когда она для этого не предназначена.

Применение ИИ на предприятиях оправдано, но необходимо давать достоверную информацию о том, как он устроен и какие слабости имеет из-за своей специфики. Причем желательно не только предоставлять ее специалистам, непосредственно работающими с данной технологией на рабочем месте, но и повышать уровень информированности всего общества, чтобы не мистифицировать данный феномен и не спекулировать на слове «интеллект». Помимо этого, необходимо позаботиться о плавном переходе к использованию ИИ, который заключается в обеспечении заменяемым сотрудникам времени для поиска новой работы.

Список литературы

- Апостолова Н.Н. Ответственность за вред, причиненный искусственным интеллектом // Северо-Кавказский юридический вестник. – 2021. – № 1 – С. 112–119. – DOI: 10.22394/2074-7306-2021-1-1-112-119
- Багдасарян Д.А. Оценка и регулирование последствий внедрения искусственного интеллекта в деятельность предприятий и организаций // Телескоп: журнал социологических и маркетинговых исследований. – 2024. – № 3(15) – С. 180–186. – DOI: 10.24412/1994-3776-2024-3-180-186
- Дозорцев В.М., Венгер А.Л. О проблеме доверия человека-оператора искусственному интеллекту // Автоматизация в промышленности. – 2022. – № 2 – С. 10–17. – DOI: 10.25728/avtprom.2022.02.02
- В гибели двух «боингов» может быть виновен бортовой компьютер. – URL: <https://life.ru/p/1202034> (дата обращения: 09.06.2025).
- Интеллектуальная система поддержки принятия решений на основе распознавания видеобразов фурменных очагов доменной печи / Бахтадзе Н.Н. [и др.] // Заводская лаборатория. Диагностика материалов. – 2022. – Т. 88, № 1(1). – С. 98–110. – DOI: <https://doi.org/10.26896/1028-6861-2022-88-1-I-98-110>
- Харитонова Ю.С., Савина В.С., Паньини Ф. Предвзятость алгоритмов искусственного интеллекта: вопросы этики и права // Вестник Пермского университета. Юридические науки. – 2021. – Вып. 53. – С. 488–515. – DOI: 10.17072/1995-4190-2021-53-488-515
- Эффективные отечественные практики на базе технологий искусственного интеллекта в обрабатывающей промышленности. Аналитический отчет, ноябрь 2022 // АНО «Цифровая экономика». – 2022.
- Huang Z. et al. A Survey on AI-Driven Digital Twins in Industry 4.0: Smart Manufacturing and Advanced Robotics // Sensors. – 2021. – Vol. 21, N 19. – P. 1–35. – DOI: 10.3390/s21196340
- Bose A.J., Aarabi P. Adversarial Attacks on Face Detectors using Neural Net based Constrained Optimization // arXiv preprint arXiv:1805.12302. – 2018. – URL: <https://arxiv.org/abs/1805.12302> (дата обращения: 09.06.2025).
- Dejanović M. The role of artificial intelligence in enhancing economic efficiency and innovation // Forum for Economic and Financial Studies. – 2025. – Vol. 3, N 1. – P. 1–12. – DOI: <https://doi.org/10.59400/fefts1541>
- Ferrando J. et al. Do I know this entity? Knowledge awareness and hallucinations in language models // arXiv preprint arXiv:2411.14257. – 2025. – URL: <https://arxiv.org/abs/2411.14257> (дата обращения: 09.06.2025).

- Long B. et al.* Explainable AI – the latest advancements and new trends // arXiv preprint arXiv:2505.07005. – 2025. – URL: <https://arxiv.org/abs/2505.07005> (дата обращения: 09.06.2025).
- Linardatos P., Papastefanopoulos V., Kotsiantis S.* Explainable AI: a review of machine learning interpretability methods // *Entropy*. – 2021. – N 23(1) – P. 1–45. – DOI: <https://doi.org/10.3390/e23010018>
- Masahiro M., MacDorman K., Kageki N.* The Uncanny Valley [From the Field] // *IEEE Robotics & Automation Magazine*. – 2012. – Vol. 19, N 2. – P. 98–100. – DOI: 10.1109/MRA.2012.2192811
- Mikal P., Kunal J.* Artificial Intelligence in Manufacturing: Benefits, Challenges and Applications // *ResearchGate.net* – 2022. – DOI: 10.13140/RG.2.2.35456.92165 – URL: https://www.researchgate.net/publication/366320838_Artificial_Intelligence_in_Manufacturing_Benefits_Challenges_and_Applications (дата обращения: 09.06.2025).
- Okechukwu F., Areji A., Ugwu A.* Beyond the Fear of Artificial Intelligence and Loss of Job: a Case for Productivity and Efficiency // *ResearchGate.net* – 2023. – DOI: 10.32388/3BWNXG – URL: https://www.researchgate.net/publication/372617772_Beyond_the_Fear_of_Artificial_Intelligence_and_Loss_of_Job_a_Case_for_Productivity_and_Efficiency (дата обращения: 09.06.2025).
- Lindsey J. et al.* On the biology of a large language model. – URL: <https://transformer-circuits.pub/2025/attribution-graphs/biology.html>
- Bartneck C. et al.* Privacy Issues of AI // Bartneck C., Lütge C., Wagner A., Welsh S. *An Introduction to Ethics in Robotics and AI. SpringerBriefs in Ethics*. – Springer, Cham, 2021. – P. 61–70. – URL: https://doi.org/10.1007/978-3-030-51110-4_8
- Sitawarin C. et al.* Rogue signs: deceiving traffic sign recognition with malicious ads and logos // arXiv preprint arXiv:1801.02780. – 2018. – <https://arxiv.org/abs/1801.02780> (дата обращения: 09.06.2025).
- Taha M.* Integrating Artificial Intelligence with SCADA Systems: Enhancing Operational Efficiency, Predictive Maintenance, and Environmental Sustainability // *ResearchGate.net* – 2024. – DOI: 10.13140/RG.2.2.18624.75526 – URL: https://www.researchgate.net/publication/382537416_Title_Integrating_Artificial_Intelligence_with_SCADA_Systems_Enhancing_Operational_Efficiency_Predictive_Maintenance_and_Environmental_Sustainability (дата обращения: 09.06.2025).
- Turanoglu B.E., Nyqvist P., Skoogh A.* An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study // *Advances in Mechanical Engineering*. – 2020. – Vol. 12(5). – P. 1–14. – DOI: 10.1177/1687814020919207
- Turing A.* Computing Machinery and Intelligence // *Mind*. – Vol. LIX, N 236, October 1950. – P. 433–460.
- AI vs humans: Facing job losses, US unions demand protection; defend worker rights – URL: <https://timesofindia.indiatimes.com/business/international-business/ai-vs-humans-facing-job-losses-us-unions-demand-protection-defend-worker-rights/articleshow/121618261.cms> (дата обращения: 09.06.2025).

References

- AI vs humans: Facing job losses, US unions demand protection; defend worker rights. Retrieved from: <https://timesofindia.indiatimes.com/business/international-business/ai-vs-humans-facing-job-losses-us-unions-demand-protection-defend-worker-rights/articleshow/121618261.cms> (date of access: 09.06.2025)

- Apostolova, N.N. (2021). Liability for damage caused by artificial intelligence. *North Caucasian Legal Vestnik*. 1, 112–119. DOI: 10.22394/2074-7306-2021-1-1-112-119 (In Russ.)
- Bagdasaryan, D.A. (2024). Assessment of negative consequences of the introduction of artificial intelligence and managing of emerging social risks. *Telescope: journal of sociological and Marketing Research*. 3(15), 180–186. DOI: 10.24412/1994-3776-2024-3-180-186 (In Russ.)
- Bakhtadze, N.N., Beginyuk, V.A., Elpashev, D.V., Zakharov, E.A., Donchan, D.M., Salikhovm Z.G. & Pyateckij, V.E. (2022). Intelligent decision support system based on video recognition of the blast furnace tuyeres. *Industrial laboratory. Diagnostics of materials*. 88(1(I)), 98–110. DOI: <https://doi.org/10.26896/1028-6861-2022-88-1-I-98-110> (In Russ.)
- Bartneck, C., Lütge, C., Wagner, A., Welsh, S. (2021). Privacy Issues of AI. In *An Introduction to Ethics in Robotics and AI*. SpringerBriefs in Ethics. Springer, Cham. (pp. 61–70). https://doi.org/10.1007/978-3-030-51110-4_8
- Bose, A.J., & Aarabi, P. (2018). Adversarial attacks on face detectors using neural net based constrained optimization. *arXiv preprint arXiv:1805.12302*. Retrieved from: <https://arxiv.org/abs/1805.12302> (date of access: 09.06.2025).
- Dejanović, M. (2025). The role of artificial intelligence in enhancing economic efficiency and innovation. *Forum for Economic and Financial Studies*. 3(1), 1–12. DOI: <https://doi.org/10.59400/feffs1541>
- Dozortsev, V.M., & Venger, A.L. (2022). On the problem of human operator's trust in artificial intelligence. *Automation in Industry*. 2, 10–17. DOI: 10.25728/avtprom.2022.02.02 (In Russ.)
- Effective domestic practices based on artificial intelligence technologies in the manufacturing industry. Analytical report, November 2022. ANO «Digital Economy», 2022. (In Russ.)
- Ferrando, J., Obeso, O., Rajamanoharan, S. & Nando, N. (2025). Do I know this entity? Knowledge awareness and hallucinations in language models. *arXiv preprint arXiv:2411.14257*. Retrieved from: <https://arxiv.org/abs/2411.14257> (date of access: 09.06.2025).
- Huang, Z., Shen, Y., Li, J., Fey, M. & Brecher, C. (2021). A Survey on AI-Driven Digital Twins in Industry 4.0: Smart Manufacturing and Advanced Robotics. *Sensors*, 21(19), 1–35. DOI: 10.3390/s21196340
- Kharitonova, Yu.S., Savina, V.S. & Panini, F. (2021). Artificial intelligence's algorithmic bias: ethical and legal issues. *Perm University Herald. Juridical Sciences*. 53, 488–515. DOI: 10.17072/1995-4190-2021-53-488-515 (In Russ.)
- Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: a review of machine learning interpretability methods. *Entropy*. 23(1), 1–45. DOI: <https://doi.org/10.3390/e23010018>
- Lindsey, J., Gurnee, W., Ameisen, E., Chen, B., Pearce, A. ... & Batson, J. (2025). On the biology of a large language model. Retrieved from: <https://transformer-circuits.pub/2025/attribution-graphs/biology.html>
- Long, B., Liu, E., Qiu R. & Duan, Y. (2025). Explainable AI – the latest advancements and new trends. *arXiv preprint arXiv:2505.07005*. Retrieved from: <https://arxiv.org/abs/2505.07005> (date of access: 09.06.2025).
- Masahiro, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*. 19(2), 98–100. – DOI: 10.1109/MRA.2012.2192811
- Mikil, P., & Kunal, J. (2022). Artificial Intelligence in Manufacturing: Benefits, Challenges and Applications. DOI: 10.13140/RG.2.2.35456.92165 Retrieved from: https://www.researchgate.net/publication/366320838_Artificial_Intelligence_in_Manufacturing_Benefits_Challenges_and_Applications (date of access: 09.06.2025).
- Okechukwu, F., Areji, A., & Ugwu, A. (2023). Beyond the Fear of Artificial Intelligence and Loss of Job: a Case for Productivity and Efficiency. DOI: 10.32388/3BWNXG Retrieved from: https://www.researchgate.net/publication/372617772_Beyond_the_Fear_of_Artificial

- Intelligence_and_Loss_of_Job_a_Case_for_Productivity_and_Efficiency (date of access: 09.06.2025)
- Sitawarin, Ch., Bhagoji, A.N., Mosenia, A., Mittal, P. & Chiang, M. (2018). Rogue signs: deceiving traffic sign recognition with malicious ads and logos. *arXiv preprint arXiv:1801.02780*. Retrieved from: <https://arxiv.org/abs/1801.02780> (date of access: 09.06.2025)
- Taha, M. (2024). *Integrating Artificial Intelligence with SCADA Systems: Enhancing Operational Efficiency, Predictive Maintenance, and Environmental Sustainability*. DOI: 10.13140/RG.2.2.18624.75526 Retrieved from: https://www.researchgate.net/publication/382537416_Title_Integrating_Artificial_Intelligence_with_SCADA_Systems_Enhancing_Operational_Efficiency_Predictive_Maintenance_and_Environmental_Sustainability (date of access: 09.06.2025)
- The on-board computer may be responsible for the deaths of two Boeing aircraft. (In Russ.) Retrieved from: <https://life.ru/p/1202034> (In Russ.) (date of access: 09.06.2025).
- Turanoglu, B.E., Nyqvist, P., & Skoogh, A. (2020). An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. *Advances in Mechanical Engineering*. 12(5), 1–14. – DOI: 10.1177/1687814020919207
- Turing A. (1950, October) Computing Machinery and Intelligence. *Mind*. Vol. LIX, N 236. P. 433–460.