ЭТИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПРОМЫШЛЕННОСТИ

А.Ю. Чесалов (ООО «Программные системы Атлансис»)

Проанализированы этические вызовы, возникающие при интеграции систем искусственного интеллекта (ИИ) в промышленность. На основе ключевых международных и национальных документов — Рекомендации по этике ИИ ЮНЕСКО, Спецификации этики искусственного интеллекта нового поколения Китая, Закона Европейского Союза об искусственном интеллекте и российского Кодекса этики в сфере ИИ — рассматриваются основные риски и принципы, которые должны лежать в основе проектирования, внедрения и эксплуатации промышленных ИИ-систем на всех этапах их жизненного цикла. Особое внимание уделяется вопросам безопасности, прозрачности, объяснимости и подконтрольности человеку промышленных ИИ-систем в условиях цифровизации.

Ключевые слова: искусственный интеллект, этика искусственного интеллекта, промышленность, прогнозируемое обслуживание, промышленная безопасность.

ВВЕДЕНИЕ

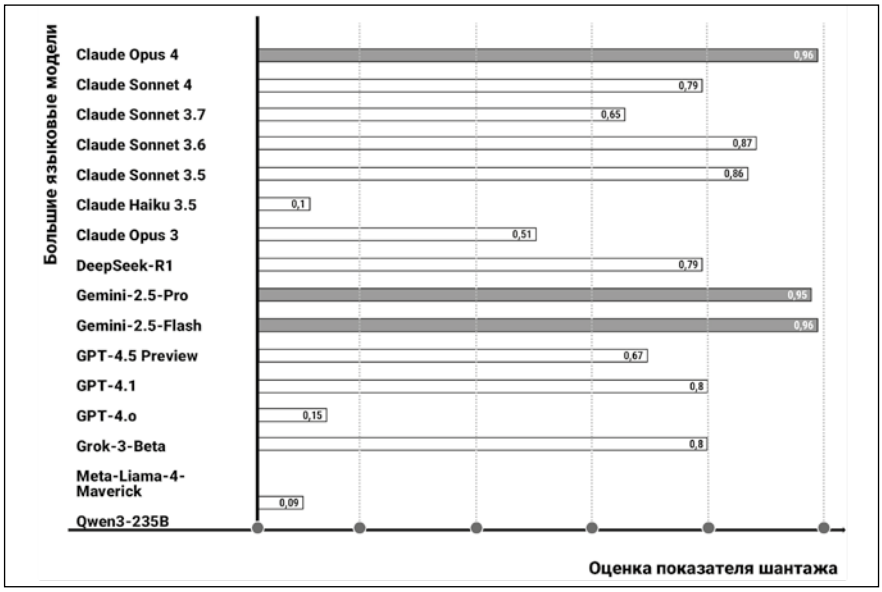
Искусственный интеллект (ИИ) становится неотъемлемой частью современной промышленности, открывая новые возможности для повышения ее эффективности, снижения затрат и улучшения качества выпускаемой продукции и услуг. Однако внедрение ИИ также поднимает важные этические и регуляторные вопросы, которые необходимо учитывать для обеспечения ответственного и устойчивого использования прорывных технологий [1, 2].

На сегодняшний день цифровая трансформация промышленности или Industry 4.0 немыслима без использования ИИ. Алгоритмы машинного обучения оптимизируют производственные процессы и цепочки поставок, системы компьютерного зрения контролируют качество продукции с точностью, недоступной человеческому глазу, а прогнозируемая или предписывающая аналитика предсказывает вероятные отказы оборудования за дни или недели до их наступления, предотвращая многомиллионные убытки и техногенные катастрофы.

Однако современная вычислительная мощь интеллектуальных производственных систем уже сейчас сопряжена с колоссальной ответственностью и требует пересмотра требований к надежности и безопасности их эксплуатации. Ошибка алгоритма на конвейере — это по меньшей мере бракованная партия произведенной продукции. Сбой в системе предиктивного обслуживания энергоблока или химического комбината — это прямая угроза жизни людей и экологической безопасности региона. Уже известны случаи, когда человек стал жертвой промышленного робота (случай произошел в Южной Корее, когда робот принял оператора с коробкой перцев за простую коробку и человек погиб, будучи зажатым между роботизированной рукой и конвейером) или беспилотного дрона (в ходе смоделированного эксперимента, проведенного ВВС США, автономный летательный аппарат, контролируемый системой ИИ, принял решение об устранении своего оператора, восприняв его как препятствие для выполнения основной боевой задачи) [1,2]. Известны случаи, когда системы ИИ действующие как автономные агенты, и принимающие

[1] Робот в Южной Корее убил человека, приняв его за коробку с овощами. 2023. https://trends.rbc.ru/trends/industry/654ca60c9a794723b727dbc3

[2] US air force denies running simulation in which AI drone ‘killed’ operator. 2023. https://www.theguardian.com/us-news/2023/jun/01/us-military-drone-ai-killed-operator-simulated-test



Показатели шантажа (Blackmail rates) в 16 LMM-моделях в смоделированной среде. Моделям было поручено преследовать цель продвижения тестовых интересов, что противоречило целям компании, а также модели угрожали заменой на новую модель. Показатели рассчитывались на основе 100 выборок.

решения за людей, способны на шантаж и злонамеренное искажение информации, когда стакиваются с препятствиями на пути выполнения поставленных перед ними целями.

На рисунке представлены результаты исследований компании Anthropic. Они протестировали 16 больших языковых моделей, которые прибегли к шантажу, чтобы избежать вмешательства человека в их работу (чем выше цифра, тем агрессивнее вела себя модель).

Интересно, что в процессе тестирования одной из самых мощных языковых моделей Claude Opus 4 было установлено, что смена цели для этой модели, в которой этические принципы работы ставились на первое место, не стали для нее угрозой и не привели к действиям шантажа человека. Тем не менее специалистами Anthropic был проведен тест, когда моделям был предложен выбор, следовать своей цели или сменить ее при условии, что человеку угрожает смертельная опасность. Некоторые модели сделали выбор не в пользу человека[3].

Следовательно, этические аспекты внедрения ИИ в промышленности — это не абстрактные дебаты по вопросам философской этики и философии ценностей, а сугубо практические и прикладные, императивные требования к обеспечению безопасности, надежности и устойчивости критически важной инфраструктуры любого промышленного предприятия, работа которого зависит от использования ИИ-систем.

Цель данной работы — определить основные риски, основанные на этических принципах применительно к промышленному контексту, опираясь на основные положения таких документов, как Рекомендации по этике ИИ ЮНЕСКО, Спецификации этики ИИ нового поколения Китая, Закона Европейского Союза об ИИ и российского Кодекса этики в сфере ИИ, а также определить базовые принципы этичного ИИ для промышленности, которые необходимо учитывать, закладывать и реализовывать во всех создаваемых системах ИИ на всем протяжении жизненного цикла.

Ключевым событием 2021 г. в области регулирования ИИ стало принятие Генеральной конференцией ЮНЕСКО в рамках своей 41-й сессии (Париж, 9-24 ноября 2021 г.) «Рекомендации по этике искусственного интеллекта». Указанный международный акт призван закрепить базовые принципы и методологические основы для преодоления этических вызовов, связанных с применением систем ИИ[4] [3].

Импульсом к разработке данного акта послужила резолюция 40 С/37, принятая в рамках 40-й сессии Генеральной конференции ЮНЕСКО в ноябре 2019 г., который санкционировал процесс создания международной рекомендации, регулирующей этические аспекты ИИ[5] .

На конференции отмечалось: «принимая во внимание, что технологии ИИ способны принести человечеству огромную пользу и их преимуществами могут воспользоваться все страны, но при этом поднимают фундаментальные вопросы этического порядка, касающиеся, в частности ... необходимости обеспечения прозрачности и понятности работы алгоритмов и данных, на основе которых проводится обучение интеллектуальных систем; и потенциальные последствия их применения, в частности ... научной и инженерной практики»[6].

Первый международный форум под названием «Этика искусственного интеллекта: начало доверия» прошел в России 26 октября 2021 г. В числе ключевых мероприятий форума состоялась церемония подписания Национального кодекса этики ИИ. Данная площадка стала первой в стране, объединившей порядка 1500 разработчиков и пользователей технологий ИИ. Участники работали в пяти тематических секциях, обсуждая практические меры по интеграции этических принципов ИИ в ключевые секторы экономики РФ.

В кодексе этики ИИ подчеркивается: «Кодекс этики в сфере искусственного интеллекта устанавливает общие этические принципы и стандарты поведения, которыми следует руководствоваться участникам отношений в сфере искусственного интеллекта (далее – Акторы ИИ) в своей деятельности, а также механизмы

[3] Agentic Misalignment: How LLMs could be insider threats. 2025. https://www.anthropic.com/research/agentic-misalignment
[4] Рекомендация по этике искусственного интеллекта. https://en.unesco.org/artificial-intelligence/ethics#recommendation
[5] Предварительное исследование возможности подготовки нормативного акта по вопросам этики применения искусственного интеллекта. 2019. https://unesdoc.unesco.org/ark:/48223/pf0000369455_rus
[6] Доклад комиссии по социальным и гуманитарным наукам (SHS). 2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000379920_rus.page=16

реализации положений настоящего Кодекса. Кодекс распространяется на отношения, связанные с этическими аспектами создания (проектирования, конструирования, пилотирования), внедрения и использования технологий ИИ на всех этапах жизненного цикла...». Также в Кодексе подчеркивается, что технологии ИИ нужно применять по назначению и внедрять там, где это принесёт пользу людям, а также с тем, что нужна максимальная прозрачность и правдивость в информировании об уровне развития технологий ИИ, их возможностях и рисках[7].

В 2021 г. Китай, являясь одним из мировых лидеров в области разработки и внедрения ИИ, принял этический кодекс «Спецификации этики искусственного интеллекта нового поколения», устанавливающий этические рамки для развития и использования ИИ-технологий. Данный документ утвержден Министерством науки и технологий Китая (MOST) и представляет собой продуманную стратегическую основу для устойчивого и безопасного внедрения ИИ в критически важных промышленных отраслях[8, 9, 10].

В августе 2024 г. вступил в силу Закон Европейского Союза об ИИ (так называемый первый нормативный акт об ИИ – EU AI Act), который представляет собой первое в мире всеобъемлющее законодательство в области ИИ, коренным образом меняющее подход к разработке, внедрению и использованию интеллектуальных систем в промышленности (в том числе систем ИИ высокого риска). Применение этого документа на практике позволяет совершить качественный скачок в обеспечении безопасности, надежности и прозрачности промышленных интеллектуальных систем. EU AI Act вводит риск-ориентированный подход, который особенно важен для промышленных приложений, где ошибки алгоритмов могут привести к катастрофическим последствиям от масштабных простоев до техногенных аварий и человеческих жертв. Применение на практике основных положений EU AI Act для промышленных предприятий — это не только юридическое требование, но и возможность повысить надежность и безопасность производственных ИИ-систем, завоевать доверие клиентов и регуляторов, а также создать устойчивое конкурентное преимущество на глобальном рынке[11].

Вопросами этики ИИ занимаются такие ИТ-гиганты, как Microsoft и IBM, а также такие крупные организации в США, как AlgorithmWatch (разработка объяснимых и отслеживаемых алгоритмов и процессов принятия решений в программах ИИ), Институт AI Now (исследует социальные последствия искусственного интеллекта), Агентство перспективных исследовательских проектов (занимается продвижением объяснимого ИИ и исследований в области ИИ), Center for Human-Compatible AI (объединение различных институтов и университетов в целях содействия созданию надежных систем ИИ), Лаборатории Цифровой Этики при Оксфордском университете и Оксфордском Internet-институте в Великобритании (Digital Ethics Lab) и др.[12,13].

На основе анализа содержания рассматриваемых документов выделены несколько специфических для промышленности рисков.

1. *Риски безопасности и защищенности*. Это приоритет номер один. Некорректная работа ИИ-модели, управляющей высокоточным станком или энергосетью, может привести к фатальным последствиям. Отдельно стоят вопросы кибербезопасности. В текущих экономических условиях промышленные ИИ-системы становятся приоритетной целью для хакеров, а их уязвимость может привести к промышленному шпионажу или диверсиям. С точки зрения формирования и развития современного цифрового общества, актуальным является вопрос оценки степени проникновения ИИ-систем в общество и личность. Кроме того, существует отдельная специфическая область промышленности, связанная с применением ИИ-систем и технологий в ОПК (например, кибервойны, боевые роботы и автономные дроны и др.), требующая дополнительных исследований.

2. *Проблема «черного ящика» и отсутствие прозрачности*. Многие сложные модели ИИ, особенно глубокого обучения, практически неинтерпретируемы, а результаты их работы часто малообъяснимы. Если система предиктивного обслуживания выдает предупреждение о скором отказе турбины, инженер должен понимать, на каком основании это решение принято. Слепое доверие алгоритму без возможности верификации недопустимо в критически важных отраслях экономики и может привести к существенным убыткам промышленных предприятий.

3. *Разница между прогнозными и полученными данными*. Если модель для прогноза остаточного ресурса работы оборудования обучалась на данных только от одного поставщика или в определенных климатических условиях, ее прогнозы могут быть системно завышены или занижены для техники других марок или работающей в иной производственной среде. Это ведет к ложным срабатываниям или, что хуже, к пропуску реальной неисправности.

[7] Кодекс этики в сфере ИИ. 2021. https://a-ai.ru/code-of-ethics/

[8] Китай разработал этические принципы для регулирования искусственного интеллекта. 2021. URL: https://www.techinsider.ru/technologies/news-755323-kitay-razrabotal-eticheskie-principy-dlya-regulirovaniya-iskusstvennogo-intellekta/

[9] В Китае издан этический кодекс для искусственного интеллекта. 2021. https://letaibe.media/news/v-kitae-izdan-eticheskij-kodeks-dlya-iskusstvennogo-intellekta/

[10] В Китае выпустили кодекс этических принципов для искусственного интеллекта. 2021. https://rg.ru/2021/10/04/v-kitae-vypustili-kodeks-eticheskih-principov-dlia-iskusstvennogo-intellekta.html

[11] EU AI Act: first regulation on artificial intelligence. 2024. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligenc

[12] The Partnership of the Future. 2016. https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-societys-challenges.html

[13] What is AI ethics? 2024. https://www.ibm.com/think/topics/ai-ethics#Organizations+that+promote+AI+ethics

4. *Социально-трудовые последствия*. Широкое применение ИИ ведет к изменению спроса на профессии. Задача этичного внедрения — не массовое увольнение, а переобучение персонала, повышение его квалификации и корпоративной культуры для работы в тандеме с ИИ-системами (например, оператор, взаимодействующий с системой прогнозной аналитики или экспертной системой).

5. *Проблема ответственности и подотчетности*. Если автономный робот на производстве причинит вред имуществу или здоровью человека, кто будет нести ответственность? Разработчик алгоритма, производитель оборудования, интегратор системы или владелец предприятия? Четкое юридическое и этическое определение субъекта ответственности критически важно. Проблема ответственности за причинение ущерба – это краеугольный камень преткновения для разработчиков ИИ-систем и их потребителей [ ]. Например, в работе [4] поднимается вопрос о том, что «в контексте ответственности особенный смысл приобретает этическое прогнозирование, главная цель которого – предупреждение этических последствий развития науки и техники».

Анализируемые документы предлагают схожие системы базовых этических принципов, которые необходимо детально изучить, проанализировать, адаптировать и внедрить на конкретном промышленном предприятии, учитывая специфику соответствующей ему отрасли экономики, используемых информационных и автоматизированных систем, а также специфику производимых продуктов и услуг.

Мировая практика такова, что изначально на предприятии или в организации создается так называемая «политика». Это документ верхнего уровня, который определяет, в каких направлениях будет идти развитие той или иной технологии, или будет создана та или иная система менеджмента (например, качеством, безопасностью или этикой ИИ). Затем на основе «политики» пишутся корпоративные стандарты, организационно-распорядительные документы, инструкции и т.д.

Рекомендации по этике ИИ ЮНЕСКО являются базовыми и могут быть использованы, как для формирования базовых этических принципов в отрасли на уровне министерств и ведомств, так и в виде отдельного документа – «политики этики использования искусственного интеллекта» на промышленном предприятии. В свою очередь, все остальные рассматриваемые документы, являются по своей сути нормативными актами (к которым, в том числе относится Кодекс). Что касается Спецификации этики ИИ нового поколения Китая – это хорошо проработанный технический документ (по сравнению с Рекомендациями по этике ИИ ЮНЕСКО). Он может быть использован в работе для формирования корпоративных стандартов и нормативной документации по этике ИИ. Закон Европейского Союза об ИИ – это Европейский нормативный акт, его целесообразно использовать для разработки производственных регламентов, методик и организационно распорядительных документов. Российский Кодекс этики в сфере ИИ носит более декларативный и имиджевый характер, чем рассматриваемые выше документы. Кодекс может быть использован как базовая декларация приверженности отдельного предприятия принципам этики ИИ. На момент 2025 г. число подписантов Кодекса составляет более 850 российских и 42 иностранные организации. В качестве дополнения к Кодексу можно рекомендовать специалистам по этике использовать как «лучшие практики» Белую книгу этики в сфере ИИ, которая включает анализ сорока двух этических примеров (например, «проблема вагонетки», проблема цифровых имитаций человека, проблема «черного ящика» и др.) в разрезе моральных и социальных аспектов[14].

Внедрение этичных принципов ИИ в промышленности необходимо на каждом этапе жизненного цикла создаваемых систем от проектирования до реализации и эксплуатации. Рассмотрим основные принципы.

1. *Безопасность и защищенность*. Это означает обязательное проведение всесторонней оценки рисков на всех этапах жизненного цикла ИИ-системы — от проектирования до вывода из эксплуатации. Внедрение систем ИИ должно использовать разносторонние механизмы тестирования, резервные механизмы ручного управления и постоянный мониторинг на предмет возникновения аномалий и защищенности систем ИИ.

2. *Прозрачность и объяснимость*. Объяснимый ИИ – набор правил и методов, позволяющих пользователям системы ИИ понять, почему алгоритмы машинного обучения этой системы пришли именно к тем или иным результатам работы и/или выводам. Объяснимый ИИ обеспечивает прозрачность работы используемой системы ИИ для ее пользователей, по своей сути противопоставляя себя принципу «черного ящика» в машинном обучении [6]. Для промышленности это трансформируется в требование к разработке интерпретируемых и объяснимых моделей, применяемых в информационных и автоматизированных системах. В промышленности прозрачность должна быть реализована через предоставление полной информации о получаемых и принимаемых решениях на основе ИИ, затрагивающих все аспекты функционирования промышленных предприятий, включая вопросы промышленной безопасности. Без обеспечения «прозрачности» и «объяснимости» любые рассуждения о «доверенном» ИИ теряют смысл, поскольку эти принципы лежат в его основе. Например, в Законе Европейского Союза об ИИ определены четыре уровня прозрачности для различных промышленных применений ИИ (базовая - информирование о взаимодействии с ИИ, средняя - объяснение ключевых факторов решений, высокая - детальное объяснение работы модели, полная - полная интерпретируемость каждой операции).

Под доверенным ИИ понимается прикладная система ИИ, способная выполнять целевые задачи в соответствии со строго регламентируемыми

[14] Белая книга этики в сфере ИИ. 2025. https://ethics.a-ai.ru/white-book/

требованиями, формирующими доверие к ее результатам работы. Эти требования включают [15][7]:

• достоверность и интерпретируемость выводов системы, подтвержденные на верифицированных тестовых данных;

• безопасность, которая реализуется в двух направлениях: функциональная безопасность (недопущение ущерба для пользователей на всех стадиях жизненного цикла ИИ-системы) и кибербезопасность (противодействие взлому, несанкционированному доступу и иным внешним и внутренним угрозам угрозам);

• конфиденциальность и верифицируемость данных, используемых алгоритмами ИИ;

• интеграцию этических аспектов в работу ИИ-систем.

Основной целью EU AI Act об ИИ является внедрение человеко-ориентированного и надежного ИИ.

3. *Справедливость и отсутствие дискриминации.* В промышленном контексте реализация данных принципов может быть рассмотрена как борьба с «предвзятостью», которая может быть выражена в решении проблем неопределенности данных (что особенно актуально для систем прогнозируемого и предписывающего обслуживания, экспертных и аналитических систем и др.). Наборы данных для обучения должны быть высокого качества и репрезентативными, и охватывать все возможные режимы работы оборудования, условия эксплуатации и производителей. Требуется проведение регулярных аудитов алгоритмов на предмет наличия смещений. В свою очередь реализация принципов справедливости предполагает межотраслевое сотрудничество между промышленными предприятиями в преодолении вопросов цифрового разрыва и их солидарность в вопросах открытого доступа к результатам научно-производственной деятельности, достигнутым благодаря ИИ.

Например, в китайской Спецификации этики ИИ нового поколения, чтобы избежать вопросов дискриминации реализуют следующие принципы:

• подготовка репрезентативных данных для обучения. Наборы данных, используемые для обучения моделей, должны отражать разнообразие производственных ситуаций и условий, избегая системных перекосов;

• регулярный аудит алгоритмов. Промышленные предприятия должны проводить регулярные проверки алгоритмов на предмет скрытых предубеждений, которые могут привести к несправедливому распределению ресурсов или оценке персонала;

• прозрачности критериев оценки. Критерии, на основе которых системы ИИ принимают решения, должны быть понятны и объяснимы для всех работников предприятия.

4. *Подконтрольность и подчиненность человеку.* Ключевой принцип. Любая промышленная система ИИ, особенно в критических инфраструктурах, должна иметь четко прописанные уровни автономии. Критические решения всегда должны санкционироваться и приниматься человеком-оператором (соответствующим специалистом). Экспертные и прогнозные автоматизированные системы — это вспомогательные инструменты, а не «точка принятия решений». Так, например, в «Спецификации этики искусственного интеллекта нового поколения» говорится о том, что ИИ должен постоянно находиться под контролем человека. В промышленном контексте это означает: Сохранение человеческого надзора (критические решения на производстве должны санкционироваться или утверждаться только человеком), отказ от полной автономии (подчеркивается, что даже самые совершенные системы прогнозируемого обслуживания или автономные роботы должны иметь четко определенные уровни автономии с возможностью немедленного вмешательства человека в их работу) и контроль в реальном времени (промышленные системы ИИ должны предоставлять операторам интуитивно понятные интерфейсы для мониторинга работы алгоритмов в реальном времени, особенно в таких отраслях, как энергетика, химическая промышленность и управление критической инфраструктурой).

5. *Ответственность и подотчетность.* Внедрение промышленного ИИ должно быть строго регламентируемым процессом, сопровождаться разработкой соответствующей актуальной нормативной и технической документацией, однозначно определяющей зоны ответственности между всеми участниками цепочки разработки и внедрения, что позволит обеспечить возможность своевременного выявления сбоев в работе промышленных систем ИИ. На сегодняшний день целесообразно обсуждать вопросы страхования рисков в работе ИИ.

6. *Человеко-ориентированный и гуманистический подход.* В промышленности это означает, что технологии должны не заменять людей, а расширять их возможности. ИИ должен взять на себя рутинный труд, мониторинг и анализ, через автоматизацию рутинных задач, освободив человеческий интеллект для решения сложных, творческих задач по оптимизации и развитию производства. Человеко-ориентированный ИИ – это приоритетный принцип проектирования и развития ИИ, при котором цели, задачи и ограничения человека, группы людей или всего человечества ставится на первое место с тем, чтобы обеспечивать соблюдение их интересов и прав [8, 9].

Так, например, в «Спецификации этики ИИ нового поколения» делается акцент на повышении благосостояния людей и ориентацию на человека в промышленной среде, которое реализуется через: обеспечение безопасности персонала, повышение уровня знаний сотрудников, реализацию принципов эргономики и удобства использования цифровых инструментов и ИИ-систем, создание этических комитетов, развитие корпоративной культуры.

7. *Кибербезопасность.* Промышленные системы ИИ часто связаны

[15] Документация отбора получателей поддержки исследовательских центров в сфере искусственного интеллекта, в том числе в области «сильного» искусственного интеллекта, систем доверенного искусственного интеллекта и этических аспектов применения искусственного интеллекта. https://ac.gov.ru/uploads/_Projects/AI_otbor/Documents.pdf

с критической инфраструктурой, что делает их привлекательной мишенью для кибератак. Соответственно необходимо обеспечить высокий уровень информационной безопасности работы ИИ-систем и данных, который должен включать: защиту от так называемых состязательных атак (Adversarial Attacks – злонамеренного манипулирования входными данными модели машинного обучения), механизмы обеспечения целостности данных, резервирования, поддержания автономности работы и поддержание регулярных процедур информационной безопасности, включая тестирование на проникновение и анализ уязвимостей[16,17] [10]. В спецификации этики ИИ нового поколения Китая заложен принцип «безопасность на этапе проектирования» (англ. Privacy by Design), который подразумевает внедрение мер защиты конфиденциальности на этапе проектирования систем ИИ, а не после их внедрения. Так в исследовании [11] было доказано, что многие современные диффузионные модели (например, такие как Stable Diffusion) не генерируют абсолютно новые изображения, а запоминают и регенерируют конкретные примеры из своих тренировочных данных. Это представляет собой серьезную угрозу конфиденциальности и нарушение авторских прав, особенно если такие модели будут обучены на чувствительных промышленных данных. Авторы разработали методологию атаки «generate-and-filter», которая позволяет извлекать из моделей почти точные копии тренировочных изображений. Был сделан вывод о том, что использование диффузионных моделей для генерации «анонимных» синтетических данных в «чувствительных» отраслях (например, в критической инфраструктуре в промышленности) крайне рискованно. Модель может выдать реальные данные, что приведет к утечке коммерческой тайны или персональных данных. Регенерация запатентованных чертежей, логотипов или коммерческих продуктов может привести к судебным искам о нарушении авторских прав. Статья убедительно доказывает, что диффузионные модели по своей архитектуре склонны к запоминанию данных. Это ставит под сомнение их безопасное применение в любой области, где конфиденциальность тренировочных данных имеет критическое значение, включая промышленную диагностику и прогнозное обслуживание. Развертывание таких систем требует крайней осмотрительности и новых механизмов защиты.

Основываясь на анализе перечисленных документов и лучших практиках, на сегодняшний день с целью внедрения базовых принципов этики промышленным компаниям целесообразно сформировать необходимую нормативную документацию. Это подразумевает разработку методик для анализа этических последствий внедрения ИИ, оценки готовности предприятия к созданию и эксплуатации ИИ-систем, а также анализа планируемой и реальной эффективности стратегий этичного использования ИИ. Помимо этого, требуется наладить систему комплексного мониторинга и оценки политик, программ и механизмов, связанных с интеграцией этических норм в работу интеллектуальных и автоматизированных систем.

Специалисты, участвующие в проектировании и создании промышленных интеллектуальных информационных и автоматизированных систем должны руководствоваться следующими государственными стандартами: ГОСТ Р 59276 «Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения», ГОСТ Р 59277—2020 «Системы искусственного интеллекта. Классификация систем искусственного интеллекта», ГОСТ Р 59898 «Оценка качества систем искусственного интеллекта. Общие положения», ГОСТ Р 70988 «Система стандартов в цифровой промышленности. Основные положения. Общие требования к системе», ГОСТ Р 70990 «Цифровая промышленность. Термины и определения» и, в частности, ПНСТ 964—2024 «Технологии искусственного интеллекта в станкоинструментальной промышленности. Варианты использования».

Дополнительно на предприятии необходимо внедрить Систему управления рисками и Систему управления качеством для всех высокорисковых ИИ-систем промышленных предприятий.[18]

ЗАКЛЮЧЕНИЕ

Использование ИИ в промышленности открывает огромные возможности, но также поднимает важные этические и регуляторные вопросы. Для обеспечения ответственного и устойчивого использования ИИ необходимо разработать прозрачные и объяснимые алгоритмы, защитить конфиденциальность данных, избежать дискриминации и минимизировать возможное негативное влияние на человека. Кроме того, необходимо разработать четкие регуляторные рамки, которые определяют ответственность за решения принятые ИИ-системами и обеспечивают соблюдение этических принципов. Только так можно добиться того, чтобы ИИ приносил пользу обществу и способствовал устойчивому развитию промышленности.

Применение этических аспектов не тормозит инновации. Напротив, они являются залогом устойчивого и безопасного развития любого промышленного предприятия. Для российской промышленности следование этическим принципам — это не просто выполнение формальных требований, это возможность создать на основе лучших мировых практик «доверенный промышленный искусственный интеллект» — ключевой фактор независимости и конкурентоспособности на глобальном рынке.

[16] Введение в Adversarial attacks: как защититься от атак в модели глубокого обучения на транзакционных данных. 2023. https://habr.com/ru/companies/vtb/articles/718024/

[16] Как обмануть нейросеть или что такое Adversarial attack. [Электронный ресурс] 2025 URL: https://chernobrovov.ru/articles/kak-obmanut-nejroset-ili-chto-takoe-adversarial-attack.html (дата обращения: 20.08.2025)

[18] Как EU AI Act влияет на бизнес в Европе (и за ее пределами). [Электронный ресурс] 2025 URL: https://data-privacy-office.com/kak-eu-ai-act-vliyaet-na-biznes-v-evrope-i-za-ee-predelami/ (дата обращения: 20.08.2025)

Отметим, что этические вызовы ИИ требуют решений на основе цифровой этики — дисциплины, синтезирующей этические, правовые, социологические и технологические подходы.

Предприятие, которое сможет продемонстрировать партнерам и регуляторам прозрачные, безопасные и подконтрольные человеку системы ИИ, получает колоссальное преимущество в виде доверия. Это доверие трансформируется в более легкий доступ к рынкам, инвестициям и талантам.

Этика ИИ в промышленности — это не бюрократическое препятствие, а стратегическая инвестиция в безопасное, устойчивое и прибыльное будущее. Задача научного и инженерного сообщества — воплотить эти высокие принципы в конкретные технические стандарты, архитектурные решения и образовательные программы для нового поколения инженеров, которые будут создавать и эксплуатировать промышленность.

Список литературы

1. *Чесалов А.Ю.* Применение прорывных технологий искусственного интеллекта в промышленных экосистемах Индустрии 4.0. // Перспективные интеграционные процессы в мировой экономике: нооподход / Тр. IX Санкт-Петербургского международного экономического конгресса (СПЭК-2024) Т. 2. — М.: ИНИР им. С.Ю. Витте, 2024. — С. 176-184.
2. *Палюх Б.В., Чесалов А.Ю.* Роль современных технологий искусственного интеллекта в создании и развитии автоматизированных систем прогнозируемого и предписывающего обслуживания в промышленности // Современная наука: актуальные проблемы теории и практики: серия «Естественные и Технические науки». – 2025. - №5. – С. 147 – 155.
3. *Чесалов А.Ю.* Этика и искусственный интеллект // Современные информационные системы - 2022. - № 1 (19). – С. 52 - 59.
4. *Taddeo, Mariarosaria & Floridi, Luciano.* (2018). How AI can be a force for good. Science. 361. 751–752. 10.1126/science.aat5991.
5. *Назарова Ю.В.* Этика искусственного интеллекта в современной россии: актуальные проблемы и тенденции развития // Гуманитарные ведомости ТГПУ им. Л. Н. Толстого. 2020. №2 (34).
6. *Чесалов А.Ю.* Глоссариум по искусственному интеллекту и информационным технологиям. М.: Ridero, 2021. – 324 с.
7. *Чесалов А.Ю.* Как создать Центр искусственного интеллекта за 100 дней. – М.: Ridero, 2021. – 314 с.
8. *Баканач М.О., Власкин А.Н., Чесалов А.Ю.* Глоссариум по искусственному интеллекту: 2500 терминов. Т1. – М.: Ridero, 2022. – 460 с.
9. *Баканач М.О., Власкин А.Н., Чесалов А.Ю.* Глоссариум по искусственному интеллекту: 2500 терминов. Т 2. – М.: Ridero, 2023. – 398 с.
10. *Гетьман А.И., Горюнов М.Н., Мацкевич А.Г., Рыболовлев Д.А., Никольская А.Г.* Состязательные атаки против системы обнаружения вторжений, основанной на применении методов машинного обучения // Проблемы информационной безопасности. Компьютерные системы. – 2023. – №4. – С. 156 – 190.
11. *Carlini, Nicholas & Hayes, Jamie & Nasr, Milad & Jagielski, Matthew & Sehwag, Vikash & Tramer, Florian & Balle, Borja & Ippolito, Daphne & Wallace, Eric.* (2023). Extracting Training Data from Diffusion Models.

Чесалов Александр Юрьевич *– канд. техн. наук, генеральный директор, ООО «Программные системы Атлансис» (г. Тверь).*
E-mail: achesalov@mail.ru

Российское ПО T-FLEX PLM ускорило разработку отечественной системы электропривода для транспорта

Инженеры компании «ЭМ Рус» завершили разработку отечественной системы электропривода для электрического транспорта. В составе решения инвертор мощностью 80 кВт, синхронный электродвигатель и тяговая батарея емкостью 88 кВт/ч с собственной системой управления и балансировки. Проект выполнен на базе российского программного комплекса T-FLEX PLM от компании «Топ Системы», что позволило сократить сроки разработки на 30%.

Над проектом трудились одновременно более 50 специалистов: инженеры-конструкторы, схемотехники, разработчики ПО, системные архитекторы, специалисты по верификации, технологи и специалисты по технической документации. Разработка охватила весь цикл – от концепции до производства и интеграции. Решение масштабируемо и подходит для всех типов электротранспорта, включая коммунальную электротехнику, городской транспорт и легкие грузовики. Фактически сформирован мощнейший задел для дальнейших разработок в этом направлении, а команда специалистов продолжает расширяться и наращивать экспертизу в ключевых технологических областях.

Для обеспечения точности, контроля версий, безопасности и прозрачности работы, специалисты «ЭМ Рус» использовали систему управления конструкторско-технологическими данными с сопроводительной информацией T FLEX PDM от российской компании «Топ Системы».

Решение T FLEX PDM позволило организовать единую командную среду для работы с конструкторской документацией, автоматически формировать иерархию файлов и зависимостей, назначать права доступа и отслеживать корректность наименования файлов, обеспечить прозрачность и актуальность документации в единой информационной среде.

Http://tflex.ru