НАУЧНЫЙ ЖУРНАЛ
НАУКА И МИРОВОЗЗРЕНИЕ

УДК-004.8

ОГРАНИЧЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Водянова Наталья Викторовна
кандидат технических наук, доцент кафедры информационных систем и технологий, Санкт-Петербургский государственный университет аэрокосмического приборостроения
Санкт-Петербург, Россия.

Аннотация

В статье исследуются основные ограничения современных систем искусственного интеллекта, определяющие границы их практического применения в науке, промышленности, образовании, медицине и государственном управлении. Рассматриваются технологические причины ошибок алгоритмов, зависимость результатов от качества обучающих данных, проблемы достоверности генерируемой информации, недостаточная объяснимость сложных моделей и трудности переноса полученных знаний на новые условия. Особое внимание уделяется вопросам конфиденциальности, алгоритмической предвзятости, информационной безопасности, ответственности за автоматизированные решения и влиянию искусственного интеллекта на рынок труда. Анализируются подходы к ограничению потенциально опасных применений, организации человеческого контроля, тестированию и независимой оценке алгоритмов. Обосновывается необходимость рассматривать искусственный интеллект не как универсальную замену человеческого мышления, а как технологический инструмент, эффективность которого зависит от качества разработки, условий эксплуатации и компетентности пользователей.

Ключевые слова: искусственный интеллект, ограничения ИИ, машинное обучение, достоверность данных, алгоритмическая предвзятость, объяснимость моделей, информационная безопасность, этика технологий.

Границы возможностей искусственного интеллекта в современной технологической среде

Искусственный интеллект за последние годы превратился из преимущественно исследовательского направления в один из важнейших инструментов цифровой трансформации. Алгоритмы машинного обучения применяются для обработки больших массивов информации, распознавания изображений и речи, прогнозирования технических процессов, автоматизации производства, анализа медицинских данных и создания текстового и графического контента. Расширение областей применения сформировало представление о возможности

использования интеллектуальных систем практически для решения любой задачи, связанной с обработкой информации. Однако такое представление не соответствует реальному уровню развития технологии.

Современный искусственный интеллект обладает значительными вычислительными и аналитическими возможностями, но его эффективность зависит от архитектуры модели, качества исходных данных, условий обучения и характера поставленной задачи. Система, успешно выполняющая определённую операцию в одной среде, не обязательно сможет справиться с похожей задачей в других обстоятельствах. Например, алгоритм распознавания объектов, обученный на изображениях при хорошем освещении, может демонстрировать существенно худшие результаты при тумане, недостаточной видимости или необычном ракурсе съёмки.

Одной из причин завышенных ожиданий является внешнее сходство результатов работы генеративных моделей с человеческой деятельностью. Современные системы способны создавать связные тексты, объяснять сложные понятия, писать программный код и поддерживать диалог. Однако наличие убедительного ответа ещё не означает, что система обладает человеческим пониманием рассматриваемого вопроса. Она может формировать правдоподобное утверждение, которое не соответствует фактам, и не всегда способна самостоятельно обнаружить собственную ошибку.

Совет Европы в материалах по вопросам искусственного интеллекта отдельно подчёркивает необходимость учитывать технические ограничения алгоритмов и критически оценивать их результаты. В частности, убедительная форма ответа не должна восприниматься как доказательство его достоверности.

Поэтому изучение ограничений ИИ имеет не меньшее значение, чем исследование его новых возможностей. Понимание слабых сторон технологии позволяет правильно выбирать области применения, устанавливать допустимые границы автоматизации и предотвращать ситуации, в которых ошибочное решение алгоритма может привести к серьёзным последствиям.

Зависимость интеллектуальных систем от качества исходных данных

Одним из фундаментальных ограничений искусственного интеллекта является зависимость от данных, используемых при обучении и эксплуатации. Машинное обучение основывается на выявлении закономерностей в примерах, поступающих в систему. Если обучающая выборка неполная, содержит ошибки или недостаточно отражает реальные условия будущего применения, качество результатов может значительно снижаться.

Особенно заметна эта проблема при работе с большими и неоднородными информационными массивами. Например, система медицинской диагностики, обученная преимущественно на данных пациентов определённого возраста или региона, может хуже распознавать заболевания у групп населения, которые

недостаточно представлены в обучающей выборке. Аналогичные трудности возникают в системах распознавания лиц, автоматизированного отбора кандидатов на работу и оценки финансовых рисков.

Отдельную проблему представляет устаревание информации. Экономические условия, социальные процессы, технические стандарты и поведение пользователей постоянно меняются. Модель, которая демонстрировала высокую точность в момент разработки, может постепенно терять эффективность, если закономерности в окружающей среде изменились. Такое явление связано с изменением распределения данных и требует регулярной проверки и обновления алгоритмов.

Необходимо также учитывать различие между объёмом и качеством информации. Большая обучающая выборка не гарантирует высокого результата, если в ней присутствуют дублирующиеся записи, противоречивые сведения, ошибочная разметка или систематические искажения. В некоторых случаях увеличение количества данных может даже усилить существующие недостатки, поскольку модель будет многократно воспроизводить ошибочные закономерности.

Подготовка качественных наборов данных требует значительных ресурсов. Необходимо собирать информацию из допустимых источников, проверять её происхождение, удалять некорректные записи, учитывать особенности различных групп пользователей и обеспечивать защиту персональных сведений. Для специализированных областей, таких как медицина, энергетика и промышленная безопасность, дополнительно требуется участие квалифицированных экспертов, способных оценить правильность разметки.

Таким образом, искусственный интеллект не является независимым источником объективного знания. Его выводы во многом отражают особенности информации, на которой он обучался, и условия, в которых используется. Контроль качества данных должен рассматриваться как обязательная часть всего жизненного цикла интеллектуальной системы.

Недостоверные ответы и проблема генерации ошибочной информации

Особенно важным ограничением генеративного искусственного интеллекта является возможность создания информации, которая выглядит убедительно, но не соответствует действительности. В научной и технической литературе подобное явление часто обозначается термином «галлюцинации» искусственного интеллекта.

Причина заключается в особенностях работы генеративных моделей. Они формируют ответ на основе закономерностей, выявленных при обучении, и оценивают вероятные последовательности элементов текста, изображений или других данных. При этом процесс генерации не всегда сопровождается проверкой каждого утверждения по независимому достоверному источнику.

В результате система может неправильно указать фамилию исследователя, придумать название научной публикации, привести несуществующий нормативный документ или представить приблизительные сведения в качестве точных фактов. Особенно опасны такие ошибки в областях, где пользователь не обладает достаточной квалификацией для самостоятельной проверки результата.

Например, при подготовке научной статьи генеративная модель может предложить библиографическую ссылку, внешне соответствующую требованиям академического оформления, но фактически не существующую. Если автор не проверит публикацию в научной базе данных или каталоге библиотеки, такая ошибка может попасть в окончательный текст.

В программировании аналогичная проблема проявляется в виде неработающих функций, использования устаревших библиотек или неверного объяснения поведения программного кода. В медицине система может сформировать рекомендации, не учитывающие индивидуальные особенности пациента, противопоказания или изменения клинических рекомендаций.

Организация экономического сотрудничества и развития относит генерацию недостоверной информации, предвзятость, нарушение конфиденциальности и недостаточную прозрачность к числу значимых рисков искусственного интеллекта.

Для снижения подобных рисков применяются методы проверки фактов, подключение к авторитетным базам данных, использование систем поиска по документам, автоматическая оценка качества ответов и обязательная экспертная проверка результатов. Однако ни один из этих методов не обеспечивает абсолютной гарантии правильности.

Следовательно, результат генеративной модели необходимо рассматривать как предварительный информационный материал, требующий проверки. Чем выше потенциальный ущерб от ошибки, тем строже должны быть требования к подтверждению полученных сведений.

Недостаточная объяснимость и прозрачность алгоритмов

Многие современные интеллектуальные системы отличаются высокой сложностью внутренней структуры. Особенно это характерно для глубоких нейронных сетей, содержащих большое количество параметров и многочисленные взаимосвязанные вычислительные операции. Такие модели способны выявлять сложные зависимости, однако объяснить конкретный результат понятным для человека способом бывает затруднительно.

Проблема известна как эффект «чёрного ящика». Пользователь получает определённый прогноз или рекомендацию, но не всегда может установить, какие именно факторы оказали решающее влияние на результат. Даже разработчики

модели могут не иметь полного представления о том, почему система сформировала конкретный ответ в определённой ситуации.

Недостаточная объяснимость создаёт серьёзные трудности при использовании ИИ в ответственных областях. Если алгоритм отказал человеку в предоставлении кредита, определил высокий риск заболевания или рекомендовал изменение производственного режима, заинтересованные стороны должны иметь возможность понять основания такого решения.

В научных исследованиях отсутствие прозрачности затрудняет воспроизводимость результатов. Если исследователь не может установить, какие признаки повлияли на вывод модели, ему сложнее оценить достоверность полученного результата и определить условия, при которых алгоритм может ошибаться.

Для решения этой проблемы развиваются методы объяснимого искусственного интеллекта. Они позволяют оценивать значимость отдельных признаков, анализировать поведение модели при изменении входных данных и представлять результаты в более понятной форме. Однако такие объяснения часто являются приближёнными и не всегда отражают все особенности внутренней работы алгоритма.

Организация экономического сотрудничества и развития подчёркивает, что недостаточная прозрачность и объяснимость интеллектуальных систем могут затруднять контроль, установление ответственности и обжалование решений, затрагивающих права человека.

Поэтому объяснимость должна учитываться уже на этапе проектирования системы. Для некоторых задач предпочтительно использовать несколько менее сложную, но более прозрачную модель, если её точности достаточно для достижения практической цели.

Заключение

Искусственный интеллект является одной из наиболее перспективных технологий современности, однако его развитие сопровождается существенными техническими, информационными, экономическими, социальными и правовыми ограничениями. Зависимость от качества обучающих данных, возможность генерации недостоверной информации, недостаточная объяснимость сложных моделей и уязвимость перед внешними воздействиями показывают, что интеллектуальные системы пока не могут рассматриваться как универсальный и безошибочный инструмент.

Особую значимость приобретает вопрос ответственности. Автоматизация отдельных операций не должна приводить к отказу человека от профессиональной оценки, особенно в медицине, промышленности, образовании, государственном управлении и других ответственных областях. Применение ИИ

требует предварительной оценки рисков, проверки результатов, защиты информации и возможности пересмотра решений.

Важнейшим условием дальнейшего развития технологии становится переход от простого увеличения вычислительных возможностей к созданию более надёжных, прозрачных и управляемых систем. Для этого необходимы совершенствование методов обучения, независимое тестирование, разработка стандартов безопасности и подготовка специалистов, способных грамотно взаимодействовать с интеллектуальными инструментами.

Таким образом, ограничения искусственного интеллекта не следует воспринимать исключительно как препятствия технологическому прогрессу. Их изучение позволяет определить реальные границы применения алгоритмов, повысить качество цифровых решений и предотвратить возможные негативные последствия. Наиболее эффективная модель использования ИИ заключается в сочетании вычислительных возможностей машин с человеческими знаниями, опытом, ответственностью и способностью учитывать общественный контекст.

Список литературы

1. OECD. *AI Risks and Incidents*. Организация экономического сотрудничества и развития. Материалы OECD о рисках искусственного интеллекта.
2. OECD. *AI, Data Governance and Privacy: Synergies and Areas of International Co-operation*. 2024. Доклад о данных и конфиденциальности.
3. OECD. *Governing with Artificial Intelligence*. 2025. Исследование государственного управления с применением ИИ.
4. OECD. *Assessing Potential Future Artificial Intelligence Risks, Benefits and Policy Imperatives*. 2024. Анализ будущих рисков и преимуществ ИИ.
5. OECD. *OECD AI Principles*. Рекомендации по ответственному развитию и применению искусственного интеллекта. Принципы OECD в области ИИ.
6. International AI Safety Report. *International AI Safety Report 2026*. Международный доклад по безопасности ИИ.
7. International AI Safety Report. *International AI Safety Report 2025*. Доклад о рисках и методах управления безопасностью ИИ.
8. Council of Europe. *Handbook on Human Rights and Artificial Intelligence*. Руководство по правам человека и искусственному интеллекту.
9. UK Government. *Artificial Intelligence Playbook for the UK Government*. Руководство по применению ИИ в государственном управлении.
10. United Nations System. *UN System White Paper on AI Governance*. 2024. Доклад ООН по управлению искусственным интеллектом.