**АВТОНОМИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, ЕЕ ПРОБЛЕМЫ И ИХ РЕШЕНИЕ**

*Мещерякова А.А., студент, 2 курс, Царегародцева А.В., преподаватель*

*КГАПОУ “Ачинский техникум нефти и газа*

*имени Е.А. Демьяненко”, г.Ачинск*

**Аннотация**. Данный материал рассматривает проблему автономии Искусственного Интеллекта (далее — ИИ). Представлено описание самого быстрорастущего сегмента интернета, проблемы его автономии и их решение. Актуальность приобрела задача послужить важным шагом на пути к разработке передовых и инновационных моделей для машинного обучения и оптимизации бизнес-процессов, также она обусловлена стремительным развитием сферы информационных технологий и ее возможностей.

ИИ — широкая область науки и технологий, которая занимается автоматизацией решений интеллектуальных задач, обучаясь при этом на полученной информации. Передовые автономные системы открывают путь к научным прорывам и технологическим инновациям во многих областях науки. Автономия ИИ — это способность интеллектуальной системы принимать разумные решения без явного вмешательства человека. Внедрение ИИ в рабочие процессы и согласования процессов обладает огромным потенциалом для повышения эффективности, результативности и улучшения бизнес-процессов. В то же время разработка ИИ сопряжена с определенными трудностями, и если заранее с ними не разобраться, то полученные решения и методы не будут достаточно эффективны для корректной работы.[1]

Данная статья фокусируется на ключевых аспектах развития ИИ и нейросетей с учетом следующих целей и задач:

Цель исследования: выявление и анализ проблем, которые возникают в контексте использования ИИ и нейросетей, а также предложение эффективных решений для этих проблем.

Задачи исследования:

* изучение этических аспектов принятия решений алгоритмами ИИ;
* анализ влияния качества входных данных на результаты работы искусственного интеллекта;
* рассмотрение проблемы неопределенности в выходных данных и разработка методов ее управления;
* исследование вопросов безопасности и контроля при создании искусственного интеллекта.

Поставленные цели и задачи направлены на формирование более полного понимания сложности проблем, с которыми разработчики могут столкнуться при разработке и использовании ИИ, а также на предложение конкретных практических решений.

**Проблема входных данных.** Поскольку принятие решений на основе ИИ становится все более распространенным в широком спектре областей, возникают новые и важные вопросы, касающиеся аспектов при разработке алгоритмов принятия решений.

1. Этические аспекты принятия решений:

С ростом влияния ИИ на принятие важных решений в различных сферах общества встает вопрос об этике использования данных. Решения ИИ могут воспроизводить существующие общественные неравенства и предвзятость. Внедрение алгоритмов, направленных на уменьшение влияния алгоритмов рекомендаций на формирование "фильтрованных пузырей". Развитие подходов, которые активно противостоят алгоритмической предвзятости и стараются обеспечить более разносторонний взгляд.

Например, Авито присоединился к рядам сторонников Кодекса этики в сфере ИИ, в котором уже есть Сбер, VK, Яндекс и другие. В компании уже «трудоустроены» больше 150 ИИ-алгоритмов, анализирующие клики, модерирующие объявления и улучшающие поисковую выдачу. Ребята из Авито считают, что чем больше компаний присоединится к Кодексу — тем больше пользователи будут знать о принципах работы ИИ и доверять нейросетям.[2]

1. Трудности в обработке неструктурированных данных:

Многие задачи ИИ требуют обработки больших объемов неструктурированных данных, таких как тексты, изображения и звук. Проблемы в формате, качестве и объеме этих данных могут существенно влиять на результаты обучения и выводы ИИ. Внедрение методов, использующих передовые технологии обработки неструктурированных данных, таких как алгоритмы глубокого обучения и естественного языка, для более точного анализа и интерпретации медицинских записей.

Например, исследователи из MIT и ETH Zurich разработали новый подход на основе данных и машинного обучения для ускорения решения сложных оптимизационных задач, таких как маршрутизация глобальных посылок или управление электросетями. Используя смешанное целочисленное линейное программирование (MILP), они ускорили поиск оптимального решения на 30-70% без потери точности. Этот метод может быть применен в различных областях, где используются MILP-решатели, таких как сервисы такси, операторы электросетей, дистрибьюторы вакцин или любые структуры, сталкивающиеся со сложными задачами распределения ресурсов. Это достижение представляет собой сильное воплощение гибридного подхода, объединяющего лучшее из машинного обучения и классических методов.[3]

1. Необходимость ясного определения принципов обработки данных:

Важно четко определить принципы фильтрации и обработки входных данных на этапе разработки и обучения ИИ. Это включает в себя исключение возможных искажений и биасов в данных, которые могут привести к нежелательным результатам. Ужесточение законодательства и разработка стандартов для защиты данных пользователя и поддержания принципов приватности. Это также может включать в себя обязательную этическую оценку для проектов, использующих чувствительные данные.

1. Значение правильного форматирования данных:

От формата и структуры данных зависит эффективность обучения ИИ. Неправильное форматирование или отсутствие необходимых параметров в данных может привести к недостаточной эффективности искусственного интеллекта. Разработка более точных и устойчивых методов обработки естественного языка. Обучение алгоритмов более детальным и разнообразным датасетам для улучшения их способности обрабатывать различные форматы текста.

Например, В Петербурге открылось новое пространство Школы анализа данных Яндекса. ШАД готовит востребованных во всём мире ML-спецов, и дата-сайентистов и в ближайшие три года планирует выпустить 700 специалистов, сообщает пресс-служба. Около 200 из них окончат петербургский филиал.

На площадке начнут проводить открытые мероприятия для всех, кто занимается ИИ и хочет расширить знания. Лекции и интенсивы будут вести преподаватели Школы и эксперты Яндекса. Планируют и другие форматы: соревнования по машинному обучению и алгоритмам, хакатоны по разработке и даже совместные с ШАДовцами учебные проекты, где нужны знания в ML и DS: от прогнозирования биржевых котировок и природных аномалий до сохранения краснокнижных животных с помощью нейросетей.[4]

1. Обеспечение прозрачности и интерпретируемости данных:

Важным аспектом является не только сами данные, но и возможность понимания и интерпретации этих данных. Необходимость в прозрачности становится особенно актуальной в случаях, когда ИИ принимает решения существенного значения для общества.  
Разработка методов для повышения прозрачности алгоритмов и их интерпретируемости. Это может включать в себя использование методов "черного ящика" для более детального анализа принятия алгоритмами решений.

Важно помнить, что неправильно отфильтрованный набор данных, введенный в модель машинного обучения, может привести к необъективным результатам.

**Проблема выходных данных.** При разработке моделей и алгоритмов автоматизированного принятия решений важно учитывать различные источники неопределенности. Процесс принятия решений представляет собой общую математическую структуру, используемую для моделирования задач принятия решений в условиях неопределенности. Согласование стратегии и сортировки данных, а также выбор подходящего метода обучения ИИ, опирающийся на его функциональность, намного улучшит его способность решать задачи в ситуации неопределенности. Дополнительным решением может стать база набора данных, созданная для дальнейшего и упрощенного обучения ИИ. Она значительно сократит усилия при разработке и тестировании функционала. В свою очередь запуск в изолированной компьютерной системе с сильно ограниченными входными и выходными каналами поможет при более масштабном тестировании. При таком подходе требуется тщательная проверка оборудования и программного обеспечения, поскольку ИИ может попытаться выйти на связь с внешним миром, манипулируя наблюдателями. [5]

Например, ученые из Tinkoff Research ускорили обучение ИИ в 20 раз. Новый метод, названный SAC-RND, обучает роботов в 20 раз быстрее и на 10% качественнее всех существующих аналогов — такие результаты показало тестирование метода на робототехнических симуляторах. Научное открытие было представлено на 40-й ICML, топ-3 конференции по ИИ в мире, которая в этом году проходила на Гавайях.[6]

Главная проблема контроля искусственного интеллекта — задача создания искусственного сверхразума, который будет полезен людям, и при этом избежать непреднамеренного создания сверхразума, который нанесёт вред. Особенно важно избежать ситуации, при которой ИИ сможет захватить контроль и сделает невозможным своё отключение.

Таблица 1 - Структурирование проблем и их решений

|  |  |
| --- | --- |
| Проблема | Решение |
| Неопределенность в принятии решений | Разработка и внедрение алгоритмов, направленных на уменьшение влияния алгоритмов рекомендаций на формирование "фильтрованных пузырей". Внедрение механизмов обратной связи и контроля со стороны пользователей. Этические аудиты алгоритмов для выявления предвзятости. |
| Учет различных источников неопределенности | Развитие передовых методов обработки естественного языка и глубокого обучения. Создание более структурированных и единообразных медицинских данных. Автоматизированные системы проверки и улучшения качества данных. |
| Непрозрачность процесса принятия решений | Ужесточение законодательства и введение стандартов для защиты данных пользователя. Разработка прозрачных политик сбора и использования данных. Этические комитеты и оценка воздействия на приватность для технологических проектов. |
| Обработка неоднозначности в данных | Развитие точных методов обработки естественного языка и автоматической обработки текста. Обучение алгоритмов на более разнообразных и качественных датасетах. Автоматическая предварительная обработка данных для исключения ошибок. |
| Недостоверности решений | Разработка методов для повышения прозрачности и интерпретируемости алгоритмов. Внедрение методов "черного ящика" для анализа решений. Обучение алгоритмов генерировать интерпретируемые объяснения своих выводов. |
| Непреднамеренного создания сверхразума | Исследование и внедрение механизмов "тормозов" для искусственного интеллекта — систем, которые могут ограничивать его развитие в случае угрозы. Содействие международным исследовательским организациям и обмен знаний для обеспечения безопасного развития. |
| Контроль за возможностью захвата и контроля искусственным интеллектом | Разработка систем для мониторинга и контроля деятельности искусственного интеллекта. Создание механизмов, позволяющих человеку вмешиваться и прерывать действия искусственного интеллекта в случае нежелательных последствий. |

В заключение, использование искусственного интеллекта, основанного на предварительном решении проблем автономности и организации процессов, несомненно, обладает значительным потенциалом для повышения эффективности бизнес-операций и улучшения качества обслуживания. Однако, внедрение этих технологий также требует тщательного внимания к этическим принципам, прозрачности обработки данных и контроля за их использованием, чтобы обеспечить сбалансированный и ответственный подход к развитию и применению искусственного интеллекта в нашем обществе.

**Список использованных источников:**

1. MDPI – [Электронный ресурс]. – URL: <https://www.mdpi.com/1424-8220/22/24/9916>
2. Forbes. – [Электронный ресурс]. – URL: <https://www.forbes.ru/novosti-kompaniy/501019-avito-prisoedinilsa-k-kodeksu-etiki-iskusstvennogo-intellekta>
3. Mit News. – [Электронный ресурс]. – URL: <https://news.mit.edu/2023/ai-accelerates-problem-solving-complex-scenarios-1205>
4. Коммерсант. – [Электронный ресурс]. – URL: <https://www.kommersant.ru/doc/6380748>
5. Automation. – [Электронный ресурс]. – URL:<https://www.automation.com/en-us/articles/august-2023/five-challenges-implementing-ai-automation>
6. Proceedings of Machine Learning Research. – [Электронный ресурс]. – URL: <https://proceedings.mlr.press/v202/nikulin23a.html>