РОЛЬ БОЛЬШИХ ДАННЫХ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СОВРЕМЕННОЙ НАУКЕ

к.ф.-м.н., доцент ***Аминжан Мавлянов***
филиал Астраханского государственного технического университета в Ташкентской области
старший преподаватель ***Ихтиёр Ёкубович Эшматов***
"Perfect-university" г.Ташкент
старший преподаватель ***Сарвар Самиевич Садиков***
"Perfect-university" г.Ташкент
Нарманова Барно Ёкубовна
Ташкентский химико-технологический институт
г.Ташкент

В статье рассматривается влияние больших данных (Big Data) и технологий искусственного интеллекта (ИИ) на парадигму современных научных исследований. Анализируются методы обработки массивов информации в различных дисциплинах, приводятся примеры интеграции нейросетевых алгоритмов и машинного обучения в естественнонаучный и точный сектора науки. Особое внимание уделяется переходу к «интенсивной на основе данных» науке и вызовам, возникающим при этом.

Ключевые слова: большие данные, искусственный интеллект, машинное обучение, современная наука, цифровая трансформация, анализ данных.

Введение

В последние годы мировая научная общественность наблюдает качественный скачок, сопоставимый по масштабам с научно-технической революцией. Стремительное развитие сенсорных систем, интернет-вещей (IoT), секвенирования геномов, спутникового мониторинга и высокопроизводительных вычислительных комплексов привело к экспоненциальному росту объемов генерируемой информации. Возникла концепция Big Data (больших данных), которая охватывает не только огромные объемы, но и высокую скорость поступления, а также разнообразие форматов.

В этих условиях традиционные ручные или полуавтоматические методы анализа данных перестали справляться с поставленными задачами. На стыке математики, статистики и информационных технологий ключевым инструментом познания стал искусственный интеллект (ИИ) и его подразделы - машинное обучение и глубокое обучение.

Переход к науке, управляемой данными

Исторически наука развивалась по индуктивно-дедуктивному пути, опираясь на эксперимент и теоретические умозаключения. Сегодня формируется четвертая парадигма научных открытий - data-intensive scientific discovery.

Автоматизация поиска закономерностей: ИИ способен выявлять скрытые корреляции и паттерны в терабайтах многомерных данных, которые человеческий ум или стандартные статистические пакеты не способны зафиксировать из-за ограничений восприятия.

Снижение эмпирических затрат: вместо проведения сотен дорогостоящих натурных экспериментов ученые могут обучать модели на исторических и синтетических данных, сокращая время поиска новых материалов, лекарств или астрономических объектов.

Ключевые направления применения Big Data и ИИ в современных дисциплинах

Синергия больших данных и интеллектуальных систем охватывает практически все сферы фундаментальной и прикладной науки:

Точные науки и астрономия

Телескопы нового поколения (например, обзоры неба LSST) еженощно фиксируют петрабайты изображений космоса. Алгоритмы компьютерного зрения на базе сверточных нейросетей автоматически классифицируют галактики, экзопланеты и аномальные космические объекты в режиме реального времени.

Науки о жизни и биоинформатика

Прорыв в области предсказания структуры белков (например, системы типа AlphaFold) продемонстрировал, как искусственный интеллект решает фундаментальные биологические задачи, на решение которых уходили десятилетия лабораторных экспериментов. Анализ экспрессии генов на основе Big Data открывает эру персонализированной медицины.

Науки о Земле и экология

Мониторинг климатических изменений требует обработки глобальных массивов метеорологических данных. ИИ-модели прогнозируют траектории стихийных бедствий (ураганов, засух, лесных пожаров) гораздо точнее традиционных численных прогнозов погоды за счет мгновенной обработки исторических архивов.

Архитектура и технологический стек обработки больших данных

Для эффективного функционирования научных систем на базе ИИ используется современный технологический стек:

распределенные вычисления: инструменты вроде Apache Hadoop и Apache Spark для параллельной обработки петабайтных датасетов;

высокопроизводительные аппаратные платформы: использование графических процессоров (GPU) и тензорных процессоров (TPU), способных ускорить обучение глубоких нейросетей в сотни раз;

интеллектуальные конвейеры: автоматизированные системы сбора, очистки, нормализации и разметки данных перед подачей их в нейросетевые модели.

Вызовы и ограничения

Несмотря на колоссальный потенциал, внедрение Big Data и ИИ в науку сопряжено с рядом проблем:

проблема «черного ящика»: многие модели глубокого обучения выдают точный результат, но логика принятия решения остается непрозрачной для исследователя, что затрудняет научное обоснование теории;

качество и чистота данных: зашумленные, неполные или смещенные данные могут приводить к ложным научным выводам («мусор на входе — мусор на выходе»);

этические и правовые вопросы: вопросы конфиденциальности медицинских данных, авторства открытий, сделанных автономными алгоритмами, и безопасности алгоритмических систем.

Решение проблем и преодоление ограничений, возникающих при внедрении математического моделирования, больших данных и искусственного интеллекта в научные исследования, требует комплексного междисциплинарного подхода. Ниже приведены основные пути и методы решения ключевых вызовов.

Решение проблемы «черного ящика» (интерпретируемость ИИ)

Развитие направления XAI (Explainable AI): внедрение методов и алгоритмов объяснимого искусственного интеллекта, которые позволяют визуализировать и обосновывать логику принятия решений нейросетями.

Физически информированные нейросети (PINNs): использование гибридных моделей, в которые на этапе обучения жестко встраиваются фундаментальные законы физики, химии или биологии. Это не позволяет модели выдавать результаты, противоречащие законам природы.

Повышение качества и очистка данных

Автоматизация пайплайнов данных: создание строгих алгоритмов предварительной фильтрации, нормализации и удаления шумов из сырых данных с использованием статистических методов контроля качества.

Синтетические данные и аугментация: генерация искусственных датасетов на основе точных математических моделей для обучения нейросетей в условиях дефицита реальных натурных измерений.

Верификация и калибровка моделей

Постоянный обратный контроль: создание «цифровых двойников» с механизмами непрерывной обратной связи, где модель регулярно сверяется с реальными датчиками и лабораторными экспериментами.

Мультимодельные ансамбли: использование не одной модели, а ансамбля различных независимых алгоритмов с последующим усреднением прогнозов для снижения вероятности системной ошибки.

Преодоление вычислительных ограничений

Распределенные и облачные вычисления: переход на масштабируемую инфраструктуру (кластерные системы, облачные научные платформы), позволяющую параллелить ресурсоемкие расчеты.

Подготовка к квантовым вычислениям: инвестиции в разработку алгоритмов, адаптированных под будущие квантовые процессоры, способные экспоненциально ускорить решение нелинейных уравнений.

Ликвидация кадрового дефицита (междисциплинарное образование)

Модернизация учебных программ: интеграция дисциплин по программированию, анализу данных и машинному обучению в образовательные программы классических естественных факультетов (физиков, биологов, экологов).

Подготовка «цифровых ученых»: развитие совместных образовательных треков, где готовят специалистов на стыке точных наук и ИКТ.

Гармонизация правовых и этических норм

Разработка научного кодекса этики ИИ: установление четких стандартов прозрачности исследований, где зафиксирована ответственность ученого за корректность применения алгоритмических инструментов.

Принципы «Открытой науки» (Open Science): поощрение обмена открытыми датасетами и исходным кодом моделей для независимого аудита и воспроизводимости научных результатов мировым сообществом.

Заключение

Большие данные и искусственный интеллект перестали быть вспомогательными инструментами и превратились в фундамент современной исследовательской деятельности. Они кардинально меняют методологию науки, ускоряя темпы открытий и расширяя границы познания от микромирных структур до мегамасштабов Вселенной. Дальнейшее успешное развитие этого

направления в академической и университетской среде требует глубокой междисциплинарной интеграции специалистов по аналитике, математиков и профильных экспертов естественных наук.

Список литературы

1. Kitchin, R. The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. -SAGE Publications, 2014.
2. Russell, S., Norvig, P. Artificial Intelligence: A Modern Approach. -4th ed. -Pearson, 2020.
3. Hey, T., Tansley, S., Tolle, K. M. The Fourth Paradigm: Data-Intensive Scientific Discovery. -Microsoft Research, 2009.
4. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. -MIT Press, 2016.
5. Каримов, Б. Х., Рахимов, Д. Ш. Использование технологий Big Data в научных исследованиях. -Ташкент: Издательство Университета «Ренессанс», 2024.