

ISSN INTERNATIONAL STANDARD SERIAL NUMBER
ISSN 2782-4365
Проверить номер:

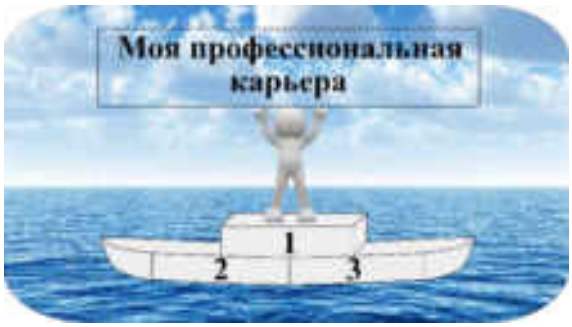
Научно-образовательный электронный журнал

ОБРАЗОВАНИЕ И НАУКА В XXI ВЕКЕ

Выпуск №79-2 (том 5) (октябрь), 2026)

Google Scholar

Периодичность выпуска: 1 раз в неделю
Сайт: mpcareer.ru/oinv21veke. Почта: obrmpcareer@mail.ru



Международный научно-образовательный электронный журнал «ОБРАЗОВАНИЕ И НАУКА В XXI ВЕКЕ»

ISSN 2782-4365
УДК 37
ББК 94

Международный научно-образовательный электронный журнал «ОБРАЗОВАНИЕ И НАУКА В XXI ВЕКЕ». Выпуск №79-2 (том 5) (октябрь, 2026). Дата выхода в свет: 12.10.2026.

Журнал объединяет авторов на территории стран СНГ и помогает обмениваться передовыми научно-образовательными исследованиями.

Содержит научные работы отечественных и зарубежных авторов по экономическим, техническим, философским, юридическим и другим наукам.

Миссия научно-образовательного электронного журнала «ОБРАЗОВАНИЕ И НАУКА В XXI ВЕКЕ» состоит в поддержке интереса читателей к оригинальным исследованиям и инновационным подходам в различных тематических направлениях, которые способствуют распространению лучшей отечественной и зарубежной практики в интернет пространстве.

Целевая аудитория журнала охватывает работников сферы науки и образования (педагоги, учителя, ученые, преподаватели, научные сотрудники, бакалавры, магистранты, аспиранты.

Материалы публикуются в авторской редакции. За соблюдение законов об интеллектуальной собственности и за содержание работ ответственность несут авторы работ. Мнение редакции может не совпадать с мнением авторов научных работ. При использовании и заимствовании материалов ссылка на издание обязательна.

© ООО «МОЯ ПРОФЕССИОНАЛЬНАЯ КАРЬЕРА»
© Коллектив авторов

СОДЕРЖАНИЕ

Название научной статьи, ФИО авторов	Номер страницы
Сарыев Нургулы, Байрамсяхедов Э. СИСТЕМНЫЕ МОДЕЛИ АЛГОРИТМОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОБРАБОТКЕ И ПОДГОТОВКЕ ДАННЫХ	13
Балканова Бахар, Сарыева Акбиби, Сапарова Махриджемал, Гельдимухаммедова Огулдженнет ВОСПИТАНИЕ ДЕТЕЙ РАННЕГО ВОЗРАСТА: ТЕОРЕТИЧЕСКИЕ ОСНОВЫ И ПРАКТИЧЕСКИЕ АСПЕКТЫ	24
Одаев Мерген, Аманмырадов Б., Сабуров И. СИСТЕМНЫЕ МОДЕЛИ АЛГОРИТМОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОБРАБОТКЕ И ПОДГОТОВКЕ ДАННЫХ	29
Гурбанова Мерджен, Атабаева Огулдурды, Аннагылыджов Перман, Мухамедниязова Айтач ТЕНДЕНЦИИ СОВРЕМЕННОГО ГЛОБАЛЬНОГО ЭКОНОМИЧЕСКОГО РАЗВИТИЯ И СТРУКТУРНЫЕ ВЫЗОВЫ: АНАЛИЗ НА ОСНОВЕ ПРОГНОЗОВ МЕЖДУНАРОДНЫХ ОРГАНИЗАЦИЙ	40
Мередова Марал, Атдиева Гульширин, Назарлыева Джумагозель, Нагмадова Гульджан СТРУКТУРНЫЙ РАСКОЛ В ГЛОБАЛЬНОМ ЭКОНОМИЧЕСКОМ ЛАНДШАФТЕ: СМЕНА ПАРАДИГМЫ — ОТ ЛОГИКИ ЭФФЕКТИВНОСТИ К ЛОГИКЕ БЕЗОПАСНОСТИ	45
Арсланова Алма, Дурдыева Эджеш, Мырадова Чынар, Беглиева Махрибан «ПАРАДОКС УСТОЙЧИВОСТИ» ГЛОБАЛЬНОЙ ЭКОНОМИКИ: РАСХОЖДЕНИЕ В АДАПТИВНОСТИ И ИНСТИТУЦИОНАЛЬНОЕ ОТСТАВАНИЕ	51
Алланазарова Шасенем, Аннагельдиев Дидар, Ахметджанова Йылдыз СЕЛЬСКОЕ ХОЗЯЙСТВО: ГЕНДЕРНЫЕ АСПЕКТЫ, РОЛЬ ЖЕНЩИН И СОЦИАЛЬНАЯ СТРУКТУРА АГРАРНОГО СЕКТОРА	56
Алланазарова Шасенем, Ашырова Язбиби СЕЛЬСКОЕ ХОЗЯЙСТВО: ОРГАНИЧЕСКОЕ ПРОИЗВОДСТВО, СЕРТИФИКАЦИЯ И ДОВЕРИЕ ПОТРЕБИТЕЛЕЙ	61
Тораева Огулджемал, Джумаева Айлар, Нурыев Непес СЕЛЬСКОЕ ХОЗЯЙСТВО: ГОСУДАРСТВЕННАЯ ПОДДЕРЖКА, СУБСИДИИ И ИСКАЖЕНИЯ РЫНКА	66

УДК: 004

ФИО автора(-ов): *Сарыев Нургулы*

преподаватель, Институт телекоммуникаций и информатики Туркменистана, Туркменистан, г. Ашхабад

Байрамсяхедов Э.

студент, Институт телекоммуникаций и информатики Туркменистана, Туркменистан, г. Ашхабад

Название публикации: «СИСТЕМНЫЕ МОДЕЛИ АЛГОРИТМОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОБРАБОТКЕ И ПОДГОТОВКЕ ДАННЫХ»

Аннотация

В статье рассматриваются системные модели алгоритмов искусственного интеллекта, применяемых на этапах обработки и подготовки данных. Показано, что качество исходных данных определяет точность, устойчивость и экономическую эффективность интеллектуальных систем не в меньшей степени, чем выбор самой модели машинного обучения. Представлена обобщённая модель конвейера подготовки данных, включающая сбор и интеграцию, очистку, преобразование и отбор признаков, разметку, аугментацию и балансировку, а также валидацию и мониторинг качества. Особое внимание уделяется алгоритмам восстановления пропущенных значений, обнаружения аномалий, масштабирования и кодирования признаков, активного обучения, слабой разметки и генерации синтетических данных. Выявлены основные проблемы: накопление ошибок на ранних этапах («каскады данных»), смещения в обучающих выборках, ошибки разметки, дрейф данных и высокая трудоёмкость ручной подготовки. Обоснована необходимость перехода к подходу, ориентированному на данные (data-centric AI), автоматизации типовых операций и внедрению систем непрерывного контроля качества. Сделан вывод о том, что

устойчивость интеллектуальных систем в бизнесе и управлении зависит от целостности всей цепочки подготовки данных, а не от отдельных алгоритмов.

Ключевые слова: искусственный интеллект, подготовка данных, очистка данных, конвейер обработки данных, машинное обучение, разметка данных, качество данных, data-centric AI.

Abstract

This article examines system-level models of artificial intelligence algorithms used for data processing and preparation. It shows that the quality of input data affects the accuracy, robustness, and economic efficiency of intelligent systems at least as much as the choice of the machine learning model itself. A generalized data preparation pipeline is presented, covering data collection and integration, cleaning, feature transformation and selection, labeling, augmentation and balancing, and quality validation and monitoring. Particular attention is paid to algorithms for missing value imputation, anomaly detection, feature scaling and encoding, active learning, weak supervision, and synthetic data generation. The study identifies major challenges, including error propagation from early pipeline stages ("data cascades"), bias in training samples, labeling errors, data drift, and the high labor intensity of manual preparation. It emphasizes the importance of the data-centric AI approach, automation of routine operations, and continuous data quality control. The article concludes that the reliability of intelligent systems in business and management depends on the integrity of the entire data preparation chain rather than on individual algorithms.

Keywords: artificial intelligence, data preparation, data cleaning, data processing pipeline, machine learning, data labeling, data quality, data-centric AI.

1. Введение

Данные стали одним из ключевых производственных ресурсов современной экономики. Интеллектуальные системы используются в прогнозировании спроса, оценке кредитных рисков, управлении запасами, образовательной аналитике и государственном управлении. Однако в большинстве практических проектов основная часть усилий приходится не на

построение модели, а на подготовку данных: их сбор, проверку, очистку, приведение к единому формату и разметку [1; 11].

Особенность этой области заключается в том, что ошибки, допущенные на ранних этапах, не исчезают в ходе дальнейшей обработки, а накапливаются и проявляются в итоговых решениях системы. Неполные, смещённые или неверно размеченные данные приводят к снижению точности прогнозов, росту затрат на исправление и потере доверия пользователей к результатам.

Актуальность исследования связана с быстрым расширением применения искусственного интеллекта в бизнесе и образовании, ростом объёмов разнородных данных и необходимостью обеспечить воспроизводимость и прозрачность их обработки.

Цель исследования — определить системную структуру алгоритмов искусственного интеллекта, применяемых при обработке и подготовке данных, проанализировать их возможности и ограничения и обосновать направления повышения эффективности подготовки данных.

2. Материалы и методы исследования

Теоретической основой исследования послужили фундаментальные монографии по искусственному интеллекту, интеллектуальному анализу данных и статистическому обучению, научные публикации по очистке данных, обнаружению аномалий, аугментации и качеству разметки, а также международный стандарт терминологии в области искусственного интеллекта [1–14]. Использованы методы системного, сравнительного и структурного анализа, а также обобщение научных подходов к построению конвейеров обработки данных.

В рамках исследования рассматриваются четыре взаимосвязанных направления: место подготовки данных в жизненном цикле интеллектуальных систем, обобщённая модель конвейера обработки данных, алгоритмические модели очистки, преобразования, разметки и аугментации, а также вопросы контроля качества и экономической эффективности. Такой подход позволяет

рассматривать подготовку данных не как набор разрозненных операций, а как целостную систему с внутренними связями и обратными зависимостями.

3. Результаты и обсуждение

3.1. Место подготовки данных в жизненном цикле систем искусственного интеллекта

Жизненный цикл интеллектуальной системы включает постановку задачи, сбор данных, их подготовку, обучение модели, оценку качества, внедрение и сопровождение. Подготовка данных занимает промежуточное положение: она преобразует «сырые» сведения из различных источников в форму, пригодную для алгоритмов обучения [2; 3].

Экономическое значение этого этапа определяется тем, что он влияет на все последующие. Модель, обученная на некачественных данных, не способна компенсировать их недостатки за счёт сложности архитектуры. Исследования практики разработки систем в ответственных областях показывают, что пренебрежение работой с данными приводит к «каскадам данных» — цепочкам негативных последствий, которые обнаруживаются лишь на стадии эксплуатации и требуют дорогостоящих доработок [11].

В связи с этим в научной литературе формируется подход, ориентированный на данные (data-centric AI). Его суть состоит в том, что приоритет отдаётся систематическому улучшению качества, полноты и согласованности данных, а не только совершенствованию алгоритмов обучения [12].

3.2. Системная модель конвейера обработки данных

С позиции системного подхода подготовка данных представляется как последовательность взаимосвязанных этапов, каждый из которых имеет входные данные, алгоритмы преобразования, критерии качества и выходной результат. Обобщённая модель включает пять блоков.

Первый блок — сбор и интеграция данных. На этом этапе информация из баз данных, документов, датчиков и внешних источников объединяется в единое

хранилище. Ключевыми задачами являются сопоставление записей, устранение противоречий между источниками и приведение данных к общей схеме [2].

Второй блок — очистка данных: обработка пропусков, выбросов, дубликатов и ошибочных значений. Третий блок — преобразование и отбор признаков: масштабирование, кодирование категориальных переменных, создание новых показателей и сокращение размерности [5]. Четвёртый блок — разметка, аугментация и балансировка выборки. Пятый блок — валидация и мониторинг качества, обеспечивающие контроль данных как на этапе обучения, так и в ходе эксплуатации системы.

Важной особенностью модели является наличие обратных связей. Результаты оценки качества модели и мониторинга в эксплуатации возвращаются на более ранние этапы, инициируя повторную очистку, дополнительную разметку или пересмотр источников данных. Следовательно, конвейер подготовки данных должен рассматриваться как непрерывный итерационный процесс, а не как однократная операция.

3.3. Алгоритмические модели очистки и преобразования данных

Очистка данных опирается на сочетание статистических методов и алгоритмов машинного обучения. Для обработки пропущенных значений применяются как простые подходы (замена средним, медианой или наиболее частым значением), так и более сложные: восстановление по ближайшим соседям, регрессионные модели и методы множественного заполнения. Выбор метода зависит от механизма возникновения пропусков: если они связаны с самими значениями показателя, простое удаление или усреднение способно исказить результаты анализа [6].

Для обнаружения аномалий и выбросов используются статистические критерии, методы на основе расстояний и плотности, ансамблевые алгоритмы (например, изолирующий лес) и нейросетевые автокодировщики, которые выявляют объекты, плохо восстанавливаемые по усвоенному «нормальному» образцу [7; 4]. Принципиально важно, что аномалия не всегда является ошибкой: в финансовых и производственных данных она может сигнализировать о

мошенничестве или сбое оборудования, поэтому автоматическое удаление таких записей недопустимо без анализа содержания.

Преобразование признаков включает масштабирование числовых переменных, кодирование категориальных признаков, обработку текстов и изображений, а также отбор наиболее информативных показателей. Эти операции обеспечивают сопоставимость данных и снижают риск переобучения. Для практической реализации широко используются библиотеки с унифицированными интерфейсами, позволяющие объединять этапы подготовки и обучения в единый воспроизводимый процесс [10].

3.4. Модели разметки, аугментации и балансировки данных

Разметка данных остаётся одной из наиболее трудоёмких операций. Для её удешевления применяются алгоритмы активного обучения, при которых модель сама выбирает наиболее неопределённые объекты и направляет их эксперту, а также слабая разметка, использующая правила, эвристики и сторонние модели для автоматического присвоения предварительных меток. Эти подходы позволяют сократить долю ручного труда, однако требуют последующего контроля точности меток.

Качество разметки имеет самостоятельное значение. Анализ широко используемых эталонных наборов данных показал, что в них присутствуют ошибки меток, способные изменять итоговые оценки и рейтинги моделей [13]. Для экономики это означает, что решения о внедрении систем могут основываться на искажённых показателях эффективности.

Аугментация расширяет обучающую выборку за счёт преобразования имеющихся объектов: поворотов, обрезки, добавления шума для изображений, перефразирования для текстов [9]. Для несбалансированных выборок, где редкие события (например, дефолты или отказы оборудования) представлены слабо, используются методы передискретизации, в частности синтетическое создание объектов миноритарного класса [8]. Дальнейшее развитие связано с генеративными моделями, создающими синтетические данные там, где реальные сведения недоступны или защищены требованиями конфиденциальности. При

этом синтетические данные наследуют смещения исходной выборки и не могут заменить реальные наблюдения без проверки.

3.5. Контроль качества, эффективность и перспективы развития

Качество данных не является постоянной характеристикой. Со временем распределение входных данных меняется (дрейф данных): изменяются потребительское поведение, экономические условия, форматы документов. Модель, показывавшая высокую точность при внедрении, постепенно теряет её. Поэтому в современных системах подготовка данных дополняется автоматическими проверками схемы, полноты и статистических свойств, а также мониторингом дрейфа [12].

С экономической точки зрения инвестиции в подготовку данных оправданы тем, что они снижают стоимость исправления ошибок на поздних стадиях, повышают воспроизводимость результатов и сокращают зависимость от отдельных специалистов. Значимым направлением становится автоматизация типовых операций, включая автоматизированный подбор приёмов преобразования данных. При этом полностью исключить участие человека не удаётся: постановка задачи, интерпретация аномалий и оценка этичности использования данных остаются зоной ответственности специалистов.

Дополнительным ограничением являются вопросы конфиденциальности, смещения и прозрачности. Терминологическая и нормативная база в этой области формируется, в частности, в международных стандартах по искусственному интеллекту [14]. Для организаций это означает необходимость документирования происхождения данных, правил их обработки и ответственности за качество на каждом этапе конвейера.

Таблица 1. Системная модель этапов подготовки данных и применяемых алгоритмов

Этап

Типовые алгоритмы

Экономический эффект

Основной риск

Сбор и интеграция

Сопоставление записей, унификация схем

Единая основа для анализа, сокращение дублирования

Противоречия между источниками

Очистка

Заполнение пропусков, обнаружение аномалий, удаление дубликатов

Снижение числа ошибок в решениях

Потеря значимых наблюдений

Преобразование и отбор признаков

Масштабирование, кодирование, сокращение размерности

Рост точности, снижение вычислительных затрат

Утечка информации, переобучение

Разметка, аугментация, балансировка

Активное обучение, слабая разметка, передискретизация

Снижение стоимости ручной разметки

Ошибки меток, перенос смещений

Валидация и мониторинг

Проверка схемы, контроль дрейфа, повторное обучение

Устойчивость системы в эксплуатации

Запоздалое обнаружение деградации

Источник: составлено автором на основе обобщения материалов [1–14].

Представленная модель показывает, что надёжность интеллектуальной системы определяется согласованностью всех этапов подготовки данных. Универсального набора алгоритмов не существует: выбор методов должен учитывать природу данных, цели использования результатов, допустимый уровень риска и располагаемые ресурсы организации.

4. Заключение

Проведённый анализ показывает, что подготовка данных является базовым элементом разработки и эксплуатации систем искусственного интеллекта и оказывает непосредственное влияние на точность моделей, стоимость проектов

и доверие к результатам. Системный подход позволяет рассматривать её как последовательность взаимосвязанных этапов с обратными связями: интеграция, очистка, преобразование, разметка, аугментация, валидация и мониторинг.

Вместе с тем практическое применение алгоритмов сопровождается рядом проблем. К ним относятся накопление ошибок на ранних стадиях, пропуски и выбросы, неточная разметка, несбалансированность выборок, смещения в данных, дрейф распределений и высокая трудоёмкость ручных операций. Эти факторы требуют пересмотра подходов к управлению данными в организациях.

Одним из приоритетных направлений является переход к подходу, ориентированному на данные: систематическое улучшение качества обучающих выборок, документирование их происхождения и непрерывный контроль на протяжении всего жизненного цикла системы. Одновременно необходимо расширять автоматизацию типовых операций, сохраняя экспертный контроль над ключевыми решениями.

Для бизнеса и образовательных организаций особое значение имеют развитие компетенций в области работы с данными, стандартизация процессов их обработки и оценка экономической эффективности инвестиций в качество данных. Для разработчиков приоритетными задачами остаются воспроизводимость конвейеров, прозрачность методов и учёт этических требований.

Таким образом, перспективы применения искусственного интеллекта определяются не только развитием алгоритмов обучения, но и качеством всей цепочки подготовки данных. Долгосрочная устойчивость интеллектуальных систем будет зависеть от зрелости процессов управления данными, глубины контроля качества и готовности организаций адаптироваться к изменению источников и характера информации.

Благодарности и финансирование

Автор выражает признательность научному сообществу за открытый доступ к публикациям и методическим материалам, используемым при изучении подготовки данных для систем искусственного интеллекта. Исследование

выполнено без привлечения внешнего финансирования на основе общедоступных информационных источников.

Список литературы

1. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Hoboken: Pearson, 2021.

2. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. 3rd ed. Waltham: Morgan Kaufmann, 2012.

3. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. 2nd ed. New York: Springer, 2009.

4. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge, MA: MIT Press, 2016.

5. Kuhn M., Johnson K. Feature Engineering and Selection: A Practical Approach for Predictive Models. Boca Raton: CRC Press, 2019.

6. Little R.J.A., Rubin D.B. Statistical Analysis with Missing Data. 3rd ed. Hoboken: Wiley, 2019.

7. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. 2009. Vol. 41, No. 3. Article 15.

8. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: Synthetic Minority Over-sampling Technique // Journal of Artificial Intelligence Research. 2002. Vol. 16. P. 321–357.

9. Shorten C., Khoshgoftaar T.M. A Survey on Image Data Augmentation for Deep Learning // Journal of Big Data. 2019. Vol. 6. Article 60.

10. Pedregosa F. et al. Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research. 2011. Vol. 12. P. 2825–2830.

11. Sambasivan N. et al. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI // Proceedings of the CHI Conference on Human Factors in Computing Systems. New York: ACM, 2021.

12. Whang S.E., Roh Y., Song H., Lee J.-G. Data Collection and Quality Challenges in Deep Learning: A Data-Centric AI Perspective // The VLDB Journal. 2023. Vol. 32. P. 791–813.

13. Northcutt C.G., Athalye A., Mueller J. Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks // NeurIPS Datasets and Benchmarks Track. 2021.
14. ISO/IEC 22989:2022. Information technology — Artificial intelligence — Artificial intelligence concepts and terminology. Geneva: ISO, 2022.