

РАЗДЕЛ I. МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ И ПАРАДИГМАЛЬНЫЕ СДВИГИ

1.1. Сдвиг парадигмы научных исследований: от гипотезо-центричной модели к data-driven исследованиям на платформах Big Data и ИИ

Введение

Классическая научная методология, сформировавшаяся в трудах Ф. Бэкона, Р. Декарта и, в особенности, К. Поппера с его концепцией фальсификационизма¹, традиционно базируется на гипотезо-центричной модели. В ее основе лежит гипотетико-дедуктивный метод, на протяжении столетий являвшийся несущим каркасом научного познания. Однако, вступление человечества в эпоху цифровых технологий, характеризующуюся экспоненциальным ростом объема, скорости и разнообразия генерируемых данных (феномен Big Data), а также появлением мощных инструментов искусственного интеллекта (ИИ) и машинного обучения (ML), инициировало глубинный пересмотр этих классических оснований. Tony Hey и его коллеги ввели понятие «Четвертой парадигмы» науки, основанной на данных (data-intensive science), которая приходит на смену экспериментальной, теоретической и вычислительной парадигмам². Современная наука переживает тектонический сдвиг – переход от проверки априорных гипотез к процессу индуктивного открытия знаний, скрытых в самих данных. Новая data-driven парадигма, опирающаяся на аналитические платформы, трансформирует не только инструментарий, но и саму онтологию и эпистемологию научного исследования³.

1. Кризис классической модели и генезис data-driven парадигмы

Традиционная гипотезо-центричная модель демонстрирует свою ограниченность в условиях, когда объектом исследования становятся сложные, многомерные и нелинейные системы. Крис Андерсон в своей провокационной статье «Конец теории: Data Deluge делает научный метод

¹ Popper, K. (2002). The logic of scientific discovery. Routledge. (Original work published 1959).

² Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). The fourth paradigm: Data-intensive scientific discovery. Microsoft Research.

³ Wong, D. T. W. Data-Driven Innovation: A Literature Review, Conceptual Framework, and Research Agenda / D. T. W. Wong, E. W. T. Ngai // IEEE Transactions on Engineering Management. – 2024. – Vol. 71. – P. 5815-5828. – DOI 10.1109/tem.2024.3371094. – EDN EYAMGF.

устаревшим» утверждал, что огромные объемы данных и корреляционный анализ делают ненужными традиционные модели, теории и гипотезы⁴. Основные вызовы классической модели включают: когнитивную предвзятость исследователя, проблему сложных систем и масштаб данных (табл.1). Ответом на эти вызовы стал переход к data-driven парадигме, который меняет последовательность исследовательского процесса.

Таблица 1

Основные вызовы традиционной гипотезо-центричной модели

Вызов	Пояснение
Когнитивная предвзятость	Гипотеза, формулируемая человеком, неизбежно ограничена рамками существующего знания, что сужает поле возможных открытий
Проблема сложных систем	В областях, где отсутствует развитый теоретический базис, выдвижение осмысленных гипотез зачастую не представляется возможным
Масштаб данных	Объемы данных, генерируемых современными научными установками, настолько велики, что их ручной анализ становится неэффективным, что требует новых подходов, как предсказывал Джим Грей в своем видении «Четвертой парадигмы» ⁵ [4]

Источник: составлено авторами

Исследование начинается не с гипотезы, а со сбора и комплексного анализа больших массивов данных, в процессе которого алгоритмы выявляют скрытые паттерны⁶. Таким образом, гипотеза становится не исходной точкой, а результатом начальной фазы исследования.

2. Эпистемологические основания data-driven науки: корреляция через каузальность

Фундаментальным эпистемологическим следствием новой парадигмы является переосмысление роли корреляции и каузальности. В классической науке открытие корреляции между переменными было лишь первым шагом на пути к установлению причинно-следственной связи.

В data-driven парадигме, как утверждал Андерсон⁷, «числа говорят сами за себя». Виктор Майер-Шёнбергер и Кеннет Кукир в своей книге «Big Data: A Revolution That Will Transform How We Live, Work, and Think»

⁴ Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, 16(7). <https://www.wired.com/2008/06/pb-theory/>

⁵ Gray, J. (2007). The fourth paradigm: Data-intensive scientific discovery [Transcript of talk]. In A. J. G. Hey, S. Tansley, & K. M. Tolle (Eds.), (2009). The fourth paradigm: Data-intensive scientific discovery. Microsoft Research.

⁶ University 4.0 concept: educational and scientific policies, innovative development of vocational education and training / I. P. Chernaya, N. N. Masyuk, V. S. Prosalova [et al.] // *Frontiers in Education*. – 2023. – Vol. 8. – DOI 10.3389/feduc.2023.1125361. – EDN YQZHWG.

⁷ Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, 16(7). (дата обращения – 01.11.2025)

подчеркивали, что ценность больших данных зачастую заключается не в точности, а в массовости, и что на смену каузальности может приходить простая корреляция⁸. Высокопрогнозные модели, построенные исключительно на корреляциях, зачастую оказываются достаточными для практического применения.

Это порождает серьезную эпистемологическую проблему: можно ли считать знанием результат работы «черного ящика» нейронной сети? Ответом становится развитие новых областей, таких как «объяснимый ИИ» (Explainable AI, XAI)⁹. Тем не менее, сама природа научного знания трансформируется, все больше смещаясь от поиска фундаментальных причин к построению высокоточных прогнозных моделей. Работа Джуды Перла и Дэны Маккензи «The Book of Why: The New Science of Cause and Effect» является попыткой вернуть каузальные рассуждения в область data science, предлагающая математический аппарат для анализа причинности¹⁰.

3. Триада технологических драйверов Big Data, ИИ/ML и аналитические платформы

Сдвиг парадигмы был бы невозможен без трех взаимосвязанных технологических драйверов.

3.1 Big Data как новая онтология научного объекта.

Понятие Big Data, всесторонне проанализированное Дугом Лэйни¹¹, характеризуется объемом (Volume), скоростью (Velocity) и разнообразием (Variety). Эти «3V» формируют новую онтологию – научный объект конституируется как целостный, многомерный поток данных.

3.2 Искусственный интеллект и машинное обучение как когнитивное ядро

Алгоритмы ML, подробно описанные в фундаментальном труде Тревора Хасти, Роберта Тибширани и Джерома Фридмана «The Elements of Statistical Learning»¹² [5], выступают в роли «усилителя» интеллекта исследователя. Методы обучения без учителя позволяют обнаруживать ранее неизвестные классы объектов. Методы обучения с учителем создают прогнозные

⁸ Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

⁹ Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138-52160.

¹⁰ Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect. Basic Books.

¹¹ Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note, 6(70), 1.

¹² Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

модели. Глубокое обучение (Deep Learning), революционизированное работами Яна Лекуна, Джошуа Бенджио и Джейфри Хинтона¹³, демонстрирует выдающиеся результаты в анализе неструктурированных данных.

3.3 Аналитические платформы как инфраструктурная основа

Современные облачные платформы интегрируют в себе вычислительные ресурсы, инструменты для работы с данными и предобученные модели. Дэвид Рю и его соавторы отмечали, как такие платформы способствуют развитию воспроизводимых и коллективных исследовательских практик, формируя то, что они называют «computational social science»¹⁴ [8]. Это способствует стандартизации, воспроизводимости и коллaborации в научной среде.

4. Новая организация исследовательского процесса и роль ученого

Data-driven парадигма трансформирует не только методологию, но и организацию научного труда. На смену индивидуальному исследователю приходят междисциплинарные коллективы, или, как их называет Генри Чесборо, «сообщества инноваций» с открытыми моделями collaboration¹⁵.

Исследовательский процесс приобретает итеративный и циклический характер, приближаясь к модели «конвейера данных» (data pipeline), описанной в работах по data science¹⁶.

Роль ученого-предметника эволюционирует: из «творца гипотез» он превращается в «интерпретатора» и «куратора». Его ключевые задачи – формулировка исследовательской задачи в терминах, понятных для ИИ, корректная интерпретация результатов, генерируемых алгоритмом, и обеспечение содержательного контекста для выявленных паттернов. Критическое мышление и глубина предметных знаний не теряют своей значимости, а, напротив, становятся важнейшими фильтрами для отсеивания статистических артефактов.

¹³LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

¹⁴ Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... & Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721-723.

¹⁵ Chesbrough, H. W. (2003). Open innovation: The new imperative for creating and profiting from technology. *Harvard Business School Press*.

¹⁶ Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking. *O'Reilly Media*.

5. Преимущества Data-Driven исследований

Переход к data-driven исследованиям предлагает множество значительных преимуществ по сравнению с традиционными подходами, фундаментально расширяя масштабы, скорость и глубину научного поиска (табл.2).

1. Повышение объективности и снижение человеческой предвзятости.

Традиционное исследование часто начинается с интуиции исследователя, на которую неизбежно влияют когнитивные искажения. Data-driven исследование начинается с самих данных, позволяя закономерностям проявляться без ограничений, накладываемых предшествующими предположениями. Это приводит к более объективным и беспристрастным открытиям. Алгоритмы могут выявлять корреляции и тренды, которые исследователь мог упустить, потому что они не вписывались в преобладающую теорию.

Таблица 2

Преимущества Data-Driven исследований

№	Преимущество	Ключевая выгода	Пример
1	Повышение объективности	Снижает предвзятость; обнаруживает неожиданные паттерны	ИИ находит неочевидную связь между микрофлорой кишечника и неврологическим заболеванием
2	Масштабируемость	Позволяет изучать целые сложные системы, а не их модели	Анализ данных со всех финансовых рынков для моделирования системного риска
3	Обнаружение паттернов	Генерирует новые гипотезы и выявляет скрытые инсайты	Использование кластеризации для выявления новых подтипов рака из генетических данных
4	Прогнозная сила	Точно предсказывает будущие события или классифицирует сложные случаи	Прогнозирование местных погодных условий с высокой точностью на основе данных сенсорных сетей
5	Повышение эффективности	Автоматизирует рутинные задачи, ускоряя цикл исследования	Автоматизированный скрининг тысяч научных статей для систематического обзора
6	Анализ неструктурированных данных	Открывает знания из текстов, изображений и аудио	Анализ исторических климатических паттернов из судовых журналов (текст)
7	Улучшенная воспроизведимость	Укрепляет научную строгость через прозрачный и повторяемый анализ	Публикация Jupyter notebook и набора данных для точного воспроизведения исследования
8	Демократизация исследований	Снижает барьеры для не-экспертов	Историк использует удобную платформу для анализа паттернов в архивных данных

Источник: составлено авторами

2. *Беспрецедентная масштабируемость и возможность анализа сложных систем.* Data-driven методы, особенно с использованием распределенных вычислений, предназначены для работы с большими объемами, высокой скоростью и разнообразием данных — ключевыми характеристиками Big Data. Это позволяет исследовать чрезвычайно сложные системы в их целостности, такие как глобальные климатические модели или социальные сети, вместо того чтобы изучать их упрощенные модели.

3. *Обнаружение неочевидных паттернов и скрытых корреляций.* Алгоритмы машинного обучения, особенно методы обучения без учителя, преуспевают в выявлении сложных, многомерных паттернов, невидимых для человеческого глаза. Эта способность способствует генерации гипотез и может привести к прорывным открытиям. Например, в фармакологии ML может выявить перспективную комбинацию молекул для нового лекарства среди миллионов возможных.

4. *Высокая прогнозная сила и точность.* Обучая прогнозные модели на больших исторических данных, можно с remarkable точностью предсказывать будущие события или классифицировать новые случаи. Это неоценимо в таких областях, как медицина (прогнозирование вспышек заболеваний), финансы (оценка кредитного риска) и маркетинг (прогнозирование поведения потребителей).

5. *Повышение эффективности и автоматизация исследовательских процессов.* Трудоемкие задачи, такие как предобработка данных и первичный скрининг, могут быть автоматизированы с помощью скриптов и AI-пайплайнов. Это освобождает исследователей для решения более сложных задач, таких как интерпретация результатов и теоретическое моделирование, значительно ускоряя жизненный цикл исследования.

6. *Возможность работы с неструктурированными данными.* Data-driven подходы, использующие NLP и компьютерное зрение, могут анализировать неструктурированные данные: текст, изображения и аудио. Это резко расширяет вселенную анализируемой информации, позволяя извлекать знания из ранее неиспользуемых источников, например, определяя общественные настроения по постам в соцсетях.

7. *Улучшенная воспроизводимость и проверяемость.* Data-driven исследование неразрывно связано с кодом и наборами данных. Когда эти компоненты открыто публикуются, любой исследователь может точно воспроизвести весь аналитический процесс. Это способствует прозрачности и укрепляет достоверность научных результатов, позволяя напрямую проверять выводы и строить более надежную совокупность знаний.

8. Демократизация исследований и стимулирование междисциплинарного сотрудничества. Облачные аналитические платформы предоставляют доступные инструменты, снижая порог входа для исследователей из неметрических дисциплин. Это стимулирует междисциплинарное сотрудничество. Социолог может работать с data scientist над анализом социальных трендов, объединяя предметные знания с передовой аналитической мощью.

Заключение

Сдвиг парадигмы научных исследований от гипотезо-центричной модели к data-driven подходу является закономерным откликом на вызовы цифровой эпохи. Этот переход, предсказанный Греем¹⁷ и развитый Хейем¹⁸, не отменяет ценности классического научного метода, но существенно расширяет его инструментарий и переопределяет логику открытия. Новая парадигма, опирающаяся на триаду Big Data, ИИ/ML и аналитических платформ, смещает фокус с верификации априорных гипотез на индуктивное открытие знаний непосредственно из данных. Это порождает новые эпистемологические вызовы, связанные с приматом корреляции над каузальностью, которые пытаются решить такие ученые, как Перл¹⁹. В результате формируется новая, более гибкая экосистема производства научного знания. Успешное развитие науки в XXI веке будет зависеть от способности научного сообщества адаптироваться к этой новой реальности, интегрируя мощь data-driven подходов с глубиной традиционного теоретического мышления, как подчеркивается в работах по взаимодействию человека и ИИ в научном творчестве. Парадигма data-driven не просто дополняет традиционные исследования — она фундаментально расширяет границы познаваемого, предлагая более масштабируемую, объективную и мощную основу для научных открытий в XXI веке.

¹⁷ Gray, J. (2007). The fourth paradigm: Data-intensive scientific discovery [Transcript of talk]. In A. J. G. Hey, S. Tansley, & K. M. Tolle (Eds.), (2009). The fourth paradigm: Data-intensive scientific discovery. Microsoft Research.

¹⁸ Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). The fourth paradigm: Data-intensive scientific discovery. Microsoft Research.

¹⁹ Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect. Basic Books.

1.2. Цифровой эволюционный скачок: от теоретических основ информатики к трансформации реальности

1. Научный фундамент: теоретические истоки и прорывные идеи

Любое значимое технологическое внедрение есть материализация фундаментальной научной идеи. Современные информационные технологии являются прямым следствием цепи интеллектуальных прорывов, совершененных гениями прошлого века.

Однако было бы ошибкой считать, что эти интеллектуальные прорывы возникали в отрыве от социального и исторического контекста. Напротив, они зачастую были прямым ответом на вызовы эпохи. Работа Алана Тьюринга по расшифровке «Энигмы» была мотивирована Второй мировой войной. Теория управления Норберта Винера и архитектура Джона фон Неймана формировались в условиях роста сложности управления большими системами — от глобальной телефонии до расчетов баллистических траекторий. Таким образом, эволюция информационных технологий — это не только внутренняя логика науки, но и диалог с насущными потребностями человечества, где теория, получившая практическую реализацию, сама становилась двигателем следующих исторических изменений.

1.1. От булевой алгебры к квантовым вычислениям

Основой всего цифрового мира является математическая логика. Работы Джорджа Буля (1815-1864) и его книга «Законы мышления» (1854) formalizовали бинарную логику, в которой любое утверждение может быть либо истинным (1), либо ложным (0). Однако ключевым внедрением этой теории в инженерию стала диссертация Клода Шеннона «Символический анализ релейных и переключательных схем» (1937). Шенон гениально показал, что булевы операторы (И, ИЛИ, НЕ) могут быть представлены электрическими схемами с реле и переключателями. Это открыло путь к проектированию цифровых электронных схем и, в конечном итоге, микропроцессоров.

Эволюция этой идеи подошла к новому рубежу — квантовым вычислениям. Если классический бит может быть 0 или 1, то квантовый бит (кубит) благодаря принципу суперпозиции может находиться в состоянии 0 и 1 одновременно. Теоретическая база была заложена Полом Бениоффом (1980), Ричардом Фейнманом (1982), который предложил использовать квантовые системы для моделирования квантовой же физики, и Дэвидом Дойчем (1985), разработавшим теорию квантового логического вентиля и квантовой машины Тьюринга. Современные эксперименты, такие как демонстрация