



DOI: 10.19181/smtp.2023.5.4.13

EDN: UGUHJX

## НАУКА О ДАННЫХ В СОВРЕМЕННОЙ СИСТЕМЕ НАУЧНЫХ ЗНАНИЙ

Рецензия на книгу Л. Цао «Образ мышления в науке о данных»<sup>1</sup>



**Гуреев  
Вадим Николаевич<sup>1</sup>**

<sup>1</sup> Институт нефтегазовой геологии и геофизики им. А. А. Трофимука  
Сибирского отделения РАН, Новосибирск, Россия



**Мазов  
Николай Алексеевич<sup>1</sup>**

<sup>1</sup> Институт нефтегазовой геологии и геофизики им. А. А. Трофимука  
Сибирского отделения РАН, Новосибирск, Россия

**Для цитирования:** Гуреев В. Н., Мазов Н. А. Наука о данных в современной системе научных знаний. Рецензия на книгу Л. Цао «Образ мышления в науке о данных» // Управление наукой: теория и практика. 2023. Т. 5, № 4. С. 209–214. DOI 10.19181/smtp.2023.5.4.13. EDN UGUHJX.

### АННОТАЦИЯ

В рецензии рассматривается монография «Образ мышления в науке о данных: Наступающая научно-техническая и экономическая революция» Лонбина Цао (ORCID: 0000-0003-1562-9429) – известного австралийского исследователя в области информатики и всего спектра ныне актуальных направлений, связанных с машинным обучением, искусственным интеллектом и др. В работе автор, несмотря на внушительный объём книги, компактно излагает становление новой дисциплины – науки о данных. В отличие от «традиционных» наук, это направление обладает целым рядом уникальных характеристик: более тесной связью теории с практикой, объединением методологических подходов общественно-гуманитарных и есте-

<sup>1</sup> Цао Л. Образ мышления в науке о данных: наступающая научно-техническая и экономическая революция / Пер. с англ. А. В. Климонтовича; науч. ред. В. И. Городецкий. СПб. : Издательство Европейского университета в Санкт-Петербурге, 2022. 552 с.

ственно-научных дисциплин, а также возможностью выхода в ранее недоступные для исследований пространства. Книга будет интересна как историкам развития науки, так и специалистам, непосредственно вовлечённым в работу с данными.

**КЛЮЧЕВЫЕ СЛОВА:**

наука о данных, научная революция, научное мышление, научная парадигма, цифровая экономика, междисциплинарные исследования

**В** прошлом, 2022-м, году вниманию отечественных читателей весьма оперативно был представлен перевод вышедшей из печати в 2018 г. фундаментальной монографии Лонбина Цао «Образ мышления в науке о данных». В последние годы представители практически всех научных дисциплин, а также множества других профессий так или иначе встречались в своей профессиональной деятельности с понятиями больших данных, искусственного интеллекта, машинного обучения, нейросетевых технологий или интеллектуальных языковых моделей, значительно изменяющих в настоящее время не только науку и технологии, но и повседневную жизнь.

В монографии предпринята амбициозная попытка охватить, охарактеризовать и систематизировать зарождение новой междисциплинарной области знаний уже на раннем этапе её формирования, с учётом тех глобальных изменений, которые приносит в нашу жизнь объект исследований дисциплины – окружающие человека данные. Несмотря на существование данных в том или ином виде со времён зарождения человечества, именно в последние десятилетия они обрели самостоятельную и особо высокую ценность в связи с появлением технологий их регистрации и последующей квантификации.

Поскольку объём генерируемых и фиксируемых данных растёт лавинообразно, а связи между ними усложняются, прежние подходы, прежде всего из области информатики, статистики, математики и науки об интеллекте, по отдельности не могут охватить всё разнообразие данных, что обуславливает необходимость в появлении новой дисциплины – науки о данных, а также соответствующего меж- и кроссдисциплинарного инструментария, связывающего физический мир и киберпространство.

Примечательно, что с началом нового тысячелетия данные стали сами активно воздействовать не только на проведение научных исследований, но и на человеческую деятельность в целом. Такая уникальная в истории человечества ситуация стала возможной из-за охвата данными – по сути числовыми характеристиками физического мира – различных сторон человеческой деятельности с учётом всей сложности их взаимодействия. Оказавшись зафиксированными, большие наборы данных в каждой из областей знаний открыли прежде недоступные, а порой никогда не планировавшиеся траектории для исследования универсума.

Основная характеристика науки о данных, как убедительно показывает автор, вытекает из самого объекта исследований – это так называемый data-driven подход, то есть исследования без заранее спланированного эксперимента или гипотезы, целиком основанные на данных, само изучение которых

в итоге формирует гипотезы и рабочие модели. В сумме такой «перевернутый» подход (от данных – к гипотезам, моделям, методам и даже целям) приводит к намного более значимым результатам в сравнении с теми, что можно получить традиционными методами. Во многом этому способствует большой объём данных в сравнении с традиционными выборками, а также их исследование во взаимосвязях друг с другом. Это позволяет выявлять не только наиболее частые закономерности (как в статистике), но и редкие, которые могут иметь гораздо большее значение.

Крайне важной, по замечанию автора, является впервые наметившаяся возможность выхода науки в неизвестную зону исследований, для которой верна ситуация «мы не знаем, чего мы не знаем». Обнаружение скрытых в данных зависимостей и закономерностей между ними, совместной встречаемости, взаимодействия, корреляции, причин тех или иных явлений, различных аномалий при использовании инструментария науки о данных позволяет получать принципиально новые и прежде скрытые знания, о существовании которых ранее никто не мог подозревать и появление которых прежде не представлялось возможным прогнозировать. При обнаружении этих скрытых закономерностей (в первую очередь, редких и неочевидных) данные сами раскрывают природу фактов, их проблематику и эволюцию.

Собственно, в этом методологическом повороте, уловить и усвоить который автор считает принципиальным в текущей парадигме развития науки, и заключается вынесенный в заглавие книги новый образ мышления: отталкиваясь от данных, исследователям необходимо выйти в невидимое пространство скрытых новых знаний, для чего следует разрабатывать соответствующий инструментарий на основе data-driven подхода. При этом понимать новый образ мышления рекомендуется расширительно и применять не только к науке о данных, но и ко всем остальным дисциплинам, где данные также начинают играть исключительно важную роль.

Автор скрупулёзно разбирает отношения и связи науки о данных со смежными и во многом давшими импульс её развитию дисциплинами. Кроме уже упомянутых информатики и статистики, это также широкий спектр прочих дисциплин, включая общественные и гуманитарные. Вбирая в себя их базовые основы, наука о данных в свою очередь обогащает каждую из традиционных областей новым методологическим аппаратом. В частности, в этом наметился важный поворот от традиционной аналитики, предоставляющей решения общего характера, к персонифицированным решениям, что особенно ярко проявилось в области медицины и различных рекомендательных систем.

Новые подходы могут оказаться принципиально важными в, казалось бы, далеко отстоящих от науки о данных общественных и гуманитарных дисциплинах. Традиционно в них применялись отличные от точных и естественных наук методология и исследовательский инструментарий. Нередкими были трудности с воспроизводимостью результатов, во многом из-за субъективной позиции исследователя и уникальных объектов анализа. Поскольку именно в общественных и гуманитарных областях в настоящее время происходит наиболее интенсивное накопление данных (преимущественно в социальных сетях), т. е. накопление фактов, поддающихся квантификации и обработке,

становится возможным и использование новой методологии, отталкивающейся от самих данных. Таким образом, и в точных, и в гуманитарных дисциплинах впервые за долгое время намечается выработка общих подходов, опирающихся на данные.

Одновременно с этим сама наука о данных в большой мере, по замечанию автора, пользуется инструментарием гуманитарных дисциплин, в первую очередь творческим мышлением (в противовес логике из точных и естественных наук). Это позволяет получать мультивариативное знание с учётом множества аспектов и точек зрения, проявлять гибкость и принимать альтернативные мнения об исследуемых объектах.

Очевидной практической пользой от результатов науки о данных является их непосредственная применимость в принятии решений и влияние на выполнение тех или иных действий. Сама аналитика переходит от традиционно дескриптивного подхода к прогнозному и предписывающему. В этом хорошо проявляется ещё одна черта науки о данных – изначальное отсутствие разрыва между теорией и практикой, поскольку в её инструментарии широко задействованы методы из бизнеса, теории управления, принятия решений и других, не относящихся к науке, областей деятельности.

В повседневной жизни высокая степень ценности результатов, полученных методами науки о данных, позволяет создавать новые сервисы и менять целые отрасли экономик. Автор убедительно демонстрирует это на примере различных рекомендательных услуг, рынка беспилотных или прокатных транспортных средств, бронирования отелей, образовательных онлайн-платформ, систем электронных платежей, «умных» домов и городов, вытеснения прямых продаж быстро растущей сферой аренды услуг, различных подписок на сервисы и перехода к динамичным моделям ценообразования в самых разных областях. К менее заметным, но не менее важным областям, претерпевающим критические изменения из-за накопления данных, относятся оптимизация различных отходов производств, распределение помощи, анализ чрезвычайных ситуаций или модели замедления климатических изменений.

Говоря о множестве преимуществ, которые несёт человечеству новая наука о данных, нельзя не сказать об оборотной стороне многих прорывных исследований последних лет, которой автор, возможно, уделяет меньше внимания, чем она заслуживает. Например, городские системы видеонаблюдения – часть «умного» города – могут использоваться не только для обеспечения безопасности граждан, но и в целях слежки за ними; эффективная обработка больших массивов данных часто применяется во время военных конфликтов, а широкие возможности интеллектуальных систем по распознаванию визуальных образов и голоса активно используются в том числе киберпреступниками.

В последние годы также можно наблюдать, что многие сервисы и данные могут использоваться как инструмент давления в условиях политических разногласий, при которых определённые регионы могут быть отключены от основанных на данных сервисов, что может приводить к снижению безопасности граждан, например, на транспорте (отключение обновлений программного обеспечения для автомобилей или систем обмена информацией при авиаперелётах). Во многих случаях нередко оказывается, что преимущество

имеют владельцы соответствующих данных и основанных на них сервисах. Не добавляет безопасности концентрация крупных информационных ресурсов среди узкого круга лиц, ответственных за принятие решений.

Отдельно стоит упомянуть проблему открытости данных, являющуюся одним из ключевых, по мнению автора, факторов успешного развития науки о данных. Во многих сферах человеческой деятельности можно наблюдать интенсивную борьбу за право доступа к данным, что выражается в теневой торговле информацией из различных организаций или проблемами с объединением данных от различных ведомств даже на государственном уровне. Очевидно, что многие спорные вопросы не могут быть решены в рамках самой науки о данных, и здесь может потребоваться опыт традиционных, прежде всего гуманитарных дисциплин.

В плане организации материала книга предполагает два варианта прочтения: быстрый – для общего ознакомления с предметом, и более вдумчивый. Каждую главу автор начинает с определённого утверждения (достаточно для получения общей информации), которое впоследствии подвергается уже подробному разбору со множеством примеров. Второй уровень чтения нельзя назвать лёгким, несмотря на безупречный перевод, выполненный сотрудниками Европейского университета в Санкт-Петербурге. Довольно часто автор преподносит материал чрезмерно детализировано – для повышения точности изложения, используя при этом исчерпывающее число примеров. Однако такой подход вмещает в монографию своеобразное справочное пособие по новой дисциплине, которое можно использовать для адресного обращения к нужным тематическим разделам. Книга сопровождается полезным списком рекомендуемой литературы и широким справочным аппаратом. Самостоятельную ценность имеет завершающий книгу 50-страничный Толковый англо-русский словарь терминов науки о данных, подготовленный проф. В. И. Городецким, – важный инструмент становления научной терминологии, необходимый в любой новой отрасли знаний.

*Статья поступила в редакцию 18.10.2023. Принята к публикации 15.11.2023.*

## **СВЕДЕНИЯ ОБ АВТОРАХ**

**Гуреев Вадим Николаевич** [GureyevVN@ipgg.sbras.ru](mailto:GureyevVN@ipgg.sbras.ru)

Кандидат педагогических наук, старший научный сотрудник, заведующий информационно-аналитическим центром, Институт нефтегазовой геологии и геофизики им. А. А. Трофимука Сибирского отделения РАН, Новосибирск, Россия

AuthorID РИНЦ: 663665

ORCID: 0000-0002-3460-0157

**Мазов Николай Алексеевич** [MazovNA@ipgg.sbras.ru](mailto:MazovNA@ipgg.sbras.ru)

Кандидат технических наук, ведущий научный сотрудник информационно-аналитического центра, Институт нефтегазовой геологии и геофизики им. А. А. Трофимука Сибирского отделения РАН, Новосибирск, Россия

AuthorID РИНЦ: 98887

ORCID: 0000-0003-4607-1122



DOI: 10.19181/smtp.2023.5.4.13

## DATA SCIENCE IN THE SYSTEM OF CONTEMPORARY SCIENTIFIC KNOWLEDGE

### Review of the Book “Data Science Thinking: The Next Scientific, Technological and Economic Revolution” by Longbing Cao

**Vadim N. Gureyev<sup>1</sup>, Nikolay A. Mazov<sup>1</sup>**

<sup>1</sup>Trofimuk Institute of Petroleum Geology and Geophysics, Siberian Branch of the RAS, Novosibirsk, Russia

**For citation:** Gureyev, V. N. and Mazov, N. A. (2023). Data Science in the system of contemporary scientific knowledge. Review of the book “Data Science Thinking: The Next Scientific, Technological and Economic Revolution” by Longbing Cao. *Science Management: Theory and Practice*. Vol. 5, no. 4. P. 209–214. (In Russ.). DOI 10.19181/smtp.2023.5.4.13.

**Abstract.** This paper reviews the book “Data Science Thinking: The Next Scientific, Technological and Economic Revolution” by Longbing Cao (ORCID: 0000-0003-1562-9429). He is a prominent Australian researcher specializing in informatics and wide range of current topical studies in machine learning, artificial intelligence, etc. Despite the large volume of the book, the author densely represent the formation of data science as a new discipline. This subject area demonstrates a number of unique features as compared to “traditional” sciences, i. e. a closer link between theory and practice, merging approaches from the humanities / social sciences and exact / natural sciences, as well as the possibility to study previously unknown and inaccessible phenomena. The book can generate interest among researchers of the history of science, as well as experts who work with data.

**Keywords:** data science, scientific revolution, scientific thinking, scientific paradigm, digital economics, interdisciplinary studies

*The article was submitted on 18.10.2023. Accepted for publication on 15.11.2023.*

#### INFORMATION ABOUT THE AUTHORS

**Vadim N. Gureyev** *GureyevVN@ipgg.sbras.ru*

Candidate of Pedagogics, Senior Researcher, Head of Information Analysis Center, Trofimuk Institute of Petroleum Geology and Geophysics, Siberian Branch of the RAS, Novosibirsk, Russia

AuthorID RSCI: 663665

ORCID: 0000-0002-3460-0157

**Nikolay A. Mazov** *MazovNA@ipgg.sbras.ru*

Candidate of Technical Sciences, Leading Researcher, Information Analysis Center, Trofimuk Institute of Petroleum Geology and Geophysics, Siberian Branch of the RAS, Novosibirsk, Russia

AuthorID RSCI: 98887

ORCID: 0000-0003-4607-1122