

Предисловие

Задумывались ли вы на рубеже XX и XXI века о том, как будет в будущем выглядеть мир вокруг? И что будет вдохновлять и определять развитие и изменение почти всех сфер нашей повседневной жизни, учебы, работы и развлечений, а также государственного управления, бизнеса и общества в целом? Наиболее уместным ответом может быть: данные, а точнее, так называемые *большие данные*, экономика данных, наука о данных и специалисты по данным. Наступает, без сомнения, эпоха больших данных — науки о данных, экономики данных и новых специальностей, связанных с данными.

В последние несколько лет мы наблюдаем огромный ажиотаж по отношению к облачным вычислениям, большим данным, науке о данных, а в наше время еще и по отношению к искусственному интеллекту. Несомненно, что объем, разнообразие, скорость и значение данных продолжают увеличиваться с каждой миллисекундой. Именно данные и связанные с ними интеллектуальные возможности участвуют в преобразованиях и интеграции прошлого, настоящего и будущего. Данные рассматриваются как новый бренд наподобие Intel Inside, или как новое «топливо», или как стратегический актив. Данные являются движущей силой или даже определяют будущее науки, технологии, экономики и, возможно, всего в завтрашнем мире.

Быстроразвивающийся и безграничный мир данных вызывает споры о *научном поиске, связанном с интенсивным использованием данных*, — науке о данных как новой парадигме, то есть о так называемой четвертой научной парадигме, которая объединяет эксперимент, теорию и вычисления, то есть экспериментальные, теоретические и вычислительные науки. В то же время возникает несколько фундаментальных вопросов: Что такое наука о данных? Как наука о данных связана с другими научными дисциплинами? Как наука о данных проникает в различные профессии, образование и экономику? Как наука о данных изменяет существующие науки, технологии, промышленность, экономику и образование? Как наука о данных конкурирует с наукой, технологиями, экономикой, сферами профессиональной деятельности и образованием будущего? Возникают также более конкретные вопросы: например, что формирует склад ума и набор навыков специалистов по данным?

Научные исследования и инновации занимаются этими и смежными вопросами и продвигают четвертую революцию в истории науки, технологии и экономики, а именно *науку о данных, технологии данных*

и экономику данных. Все вышеизложенное стало первопричиной написания этой книги.

Существует не так уж много изданий, которые позиционируются на книжном рынке как относящиеся к науке о данных. Настоящая книга не содержит описания технических деталей или каких-либо аспектов математики, статистики, машинного обучения, поиска данных в разных источниках, облачных вычислений, языков программирования и других тем, относящихся к науке о данных. Техническим аспектам науки о данных и ее приложений будет посвящена другая книга автора: «Наука о данных: Методы и приложения» (“Data Science: Techniques and Applications”).

Эта книга навеяна желанием найти и изучить ответы на поставленные выше фундаментальные вопросы в эпоху науки о данных и экономики данных. Мы намерены нарисовать полную картину науки о данных: как новой научной парадигмы с точки зрения эволюции науки, как даталогического мышления с точки зрения мышления, как трансдисциплинарной науки с точки зрения входящих в нее дисциплин и как новую область профессиональной деятельности и экономики с точки зрения бизнеса.

Таким образом, книга охватывает широкий спектр существенных аспектов науки о данных — от ее оснований и до ее роли в индустриализации, в появлении новых профессий, в образовании и в широком круге предлагаемых ей возможностей. Книга состоит из трех частей.

В первой части описаны эволюция, основные понятия, неверные представления и мышление в науке о данных. В главе 1 рассматриваются эволюция, характеристики, свойства, направления развития и повестка дня в эпоху данных. Глава 2 посвящена ответу на вопрос «Что такое наука о данных?» с высокоуровневой, мультидисциплинарной и процессной точек зрения. Ажиотаж, связанный с большими данными и наукой о данных, сопровождается большим количеством мифов и ошибочных представлений, которые также обсуждаются в этой главе. Даталогическое мышление играет важную роль в исследованиях, инновациях и приложениях науки о данных, что обсуждается в главе 3.

Во второй части рассматриваются вызовы науки о данных и основания работы в ней. Эта часть состоит из трех глав. Различные вызовы рассматриваются в главе 4. Глава 5 посвящена методологиям, дисциплинарным рамкам и направлениям исследований с точки зрения разных научных дисциплин. В главе 6 исследуется роль различных дисциплин и знаний в формировании оснований науки о данных. Наконец, глава 7 резюмирует основные направления исследований, теории, методы и приложения аналитики и обучения в различных предметных областях.

Последняя, третья часть книги посвящена индустриализации, движущей силой которой является наука о данных. Она состоит из четырех глав. Наука о данных и ее разнообразные приложения двигают вперед экономику данных, индустрию данных и сервисы данных, что обсуждается в главе 8. Наука о данных, экономика данных и приложения, связанные с данными, способствуют развитию профессий, связанных с данными,

а также стимулируют формирование должностей в науке о данных и моделях зрелости — этому посвящена глава 10. Эпоха науки о данных должна строиться специалистами по данным и инженерами, которые должны получить соответствующую квалификацию, образование в науке о данных, технологиях и экономике. Многозадачность делает написание такой книги, как эта, чрезвычайно сложным и рискованным предприятием. В книге не обсуждаются теории и алгоритмы машинного обучения и поиска данных или связанные с ними понятия и методы — их можно найти в других специальных книгах, многочисленные ссылки на которые есть в списке литературы.

Книга предназначена для менеджеров данных (например, менеджеров аналитических портфелей (*англ.* analytics portfolio manager), менеджеров бизнес-аналитики, ведущих аналитиков данных и ведущих специалистов по данным), руководителей и лиц, принимающих стратегические решения, руководителей исследований и преподавателей, которые отвечают за продвижение новой научной, инновационной и промышленной повестки, для тех, кто отвечает за стратегическое планирование на предприятиях и за разработку учебных курсов следующего поколения, а также для всех, кто занимается наукой о данных, технологиями и экономикой в перспективе высокого уровня.

Студенты по различным специальностям, связанным с наукой о данных, также найдут ее полезной для начала своего научного пути, планирования своей карьеры и подготовки к конкуренции в науке, технологиях и экономике следующего поколения.

Можете ли вы представить себе, как будет далее развиваться мир данных и как данные будут влиять на будущую науку, технологии, экономику и общество во второй половине XXI века? Для того чтобы претендовать на звание специалиста по данным и на «занятия данными», необходимо найти ответы на многие вопросы, извлечь пользу из текущего состояния науки о данных и осознать возможности, которые возникнут в будущем. Таким образом, я надеюсь, что эта книга внесет свой вклад в обсуждение науки о данных.

*Лонбин Цао
Сидней, Австралия
июль 2018 года*