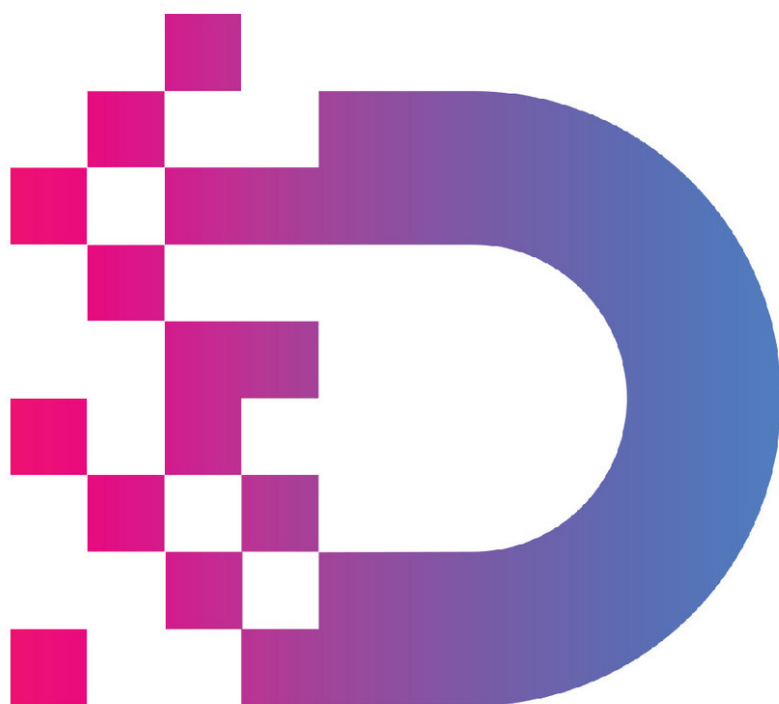


АЛЕКСЕЙ БЛАГИРЕВ



BIG DATA
ПРОСТЫМ ЯЗЫКОМ

Бизнес-бук

Алексей Благирев

Big data простым языком

«Издательство АСТ»

2019

УДК 004.67
ББК 32.973.233-018

Благирев А.

Big data простым языком / А. Благирев — «Издательство АСТ»,
2019 — (Бизнес-бук)

ISBN 978-5-17-111829-7

Наш телефон знает о нас больше, чем мы думаем. Он умеет собирать и анализировать информацию о том, как мы передвигаемся по городу, какие посты лайкаем и какими приложениями пользуемся. Он сообщит о пробках и поторопит на работу, чтобы мы не опоздали; подберет музыку под наше настроение и составит список персональных рекомендаций, чем можно занять себя в течение дня. Телефон – больше не устройство, по которому звонят, это уже средство управления окружающим нас миром. Незаметно мы окружили себя такими интерфейсами, которые создают невидимый барьер между человеком и окружающей средой. Планирование, управление, коммуникация, все теперь строится через эти программы и девайсы. Даже человеческие отношения. Но насколько глубока кроличья нора? Каждому предстоит разобраться в этом самом. Эта книга поможет донести основные принципы проектирования и создания таких интерфейсов управления бизнесом, обществом и окружающим нас миром посредством Больших данных. Читайте, наслаждайтесь и помните: сожжение книг противозаконно.

УДК 004.67
ББК 32.973.233-018

ISBN 978-5-17-111829-7

© Благирев А., 2019
© Издательство АСТ, 2019

Содержание

Предисловие	7
Глава 1	9
Марсианские диалекты	9
Что же это все-таки такое и откуда взялось?	13
Постинформационное общество[3]	16
Конец ознакомительного фрагмента.	21

Алексей Благирев

Big data простым языком

© Благирев А., текст, иллюстрации

© ООО «Издательство АСТ»

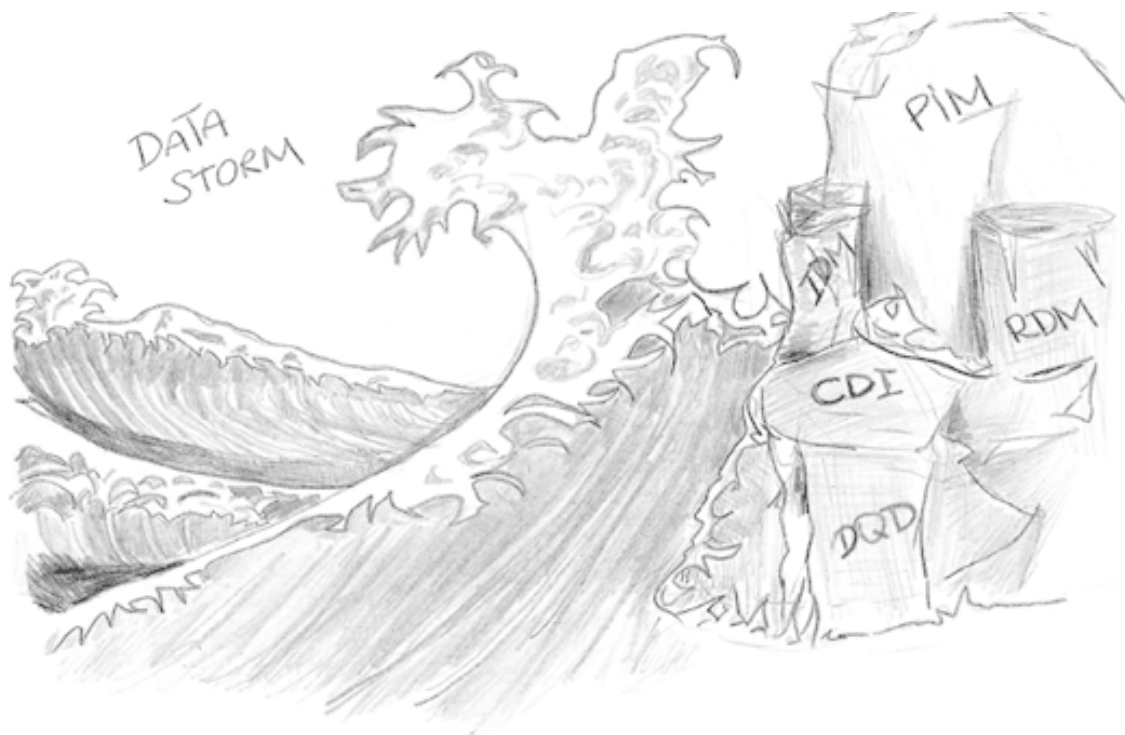
* * *

Предисловие

Люблю людей.

Именно такие мысли остаются в голове, когда тебе предлагают полностью переписать книгу. А если вы читаете это, значит, мне удалось, и я все-таки ее переписал.

Началось все с того, что один мой друг спросил, не знаю ли я людей, которые могли бы простым языком написать про Большие данные. Тогда я сразу представил бесконечное количество писем от издательства, разговоров, уточнений, переписываний, – всей этой суеты, и первое, что мне хотелось ответить: «Нет, таких разумных существ я не знаю».



Да и смысл писать про Большие данные, если про них уже столько всего написано и рассказано? Вероятность написать что-то умное – минимальна.

И вот я начал писать... Я даже уже представлял себе, как героически заканчиваю эту книгу, становлюсь миллионером и на все деньги с продаж бесконечных тиражей иду погашать ипотеку.

План был гениален, оставалось только его воплотить.

Но, когда я показал плод своих трудов редактору, он сказал, что книга сложна для восприятия, иными словами, подходит только для ботанов. Я честно писал ее с использованием книжной лексики, сложных эвфемизмов, деепричастных оборотов и кропотливо вставлял в текст ссылки на источники, если вдруг упоминал материалы других авторов.

Один раз даже пришлось взять DMBOK, такую специальную «поваренную книгу» с инструкциями и стандартами организации работы с данными. Я перевел из нее целую главу на русский, но мне сказали, что это точно «слишком мощно» для читателя, как и попытка проанализировать существующее регулирование данных.

Итак, в поисках правды, баланса и закрытия личных гештальтов – как сейчас популярно говорить, мне дали книгу «Хулиномика» как пример образцовой книги жанра нон-фикшн.

Когда я взял в руки «Хулиномику», помимо ненормативной лексики в мыслях у меня появились смелые очертания нового эксперимента, поэтому вы держите в руках книгу про Большие данные, изданную под влиянием уникальной простоты и творческой логики изложения.

Мир данных – это компот, из которого трудно отделить то, что нужно знать, а что нет. И вроде бы все интересно, про все можно рассказать, но как понять, что из этого важно, например, учителю физкультуры, который на досуге решил погрузиться в данные?

Задача оказалась сложнее, чем я думал.

Если вы пишете, скажем, про физику, то план изложения поправит научный редактор. А тут – технологии, англицизмы, и людей, знающих ключевые понятия, широту и многогранность Больших данных в издательстве просто-напросто нет.

И я взялся за дело. Сам.

Для начала я решил, что в каждой главе будет два уровня сложности. Первый – для тех, кто собрался почитать про данные, сидя на белом друге в тихой комнате, второй – для тех, чья сфера деятельности связана с данными.

Я написал большую главу про стратегию данных для тех, кто вынужден проектировать стратегию с нуля; попытался разобраться, как данные влияют на корпоративное управление компаниями; показал на ошибках людей, рисующих сложные, малопонятные графики, что формат изложения информации не менее важен, чем сам процесс получения знания.

Конечно, то, что вы держите в руках, – сильно переработанный вариант, но не менее достойный. Наверное.

Сегодня этот компот под названием «мир данных» – уже целая экономика, которая сильно повлияла на все вокруг, включая людей. Теперь нашими данными располагают голосовые помощники, а банки и компании, с которыми мы когда-либо имели дело, все чаще напоминают о себе и требуют внимания. Наш телефон знает, когда мы собираемся на работу, и заранее подгоняет нас к выходу, чтобы мы не опоздали из-за пробок, а когда мы выбираем песню, которую хотим послушать в машине, он выдает нам подходящий плейлист.

Важно знать, что за данные, а точнее за искусственный интеллект, начали активно «топить» в обществе и бизнесе, поднимая проблемы этики их использования.

Просто задумайтесь, вся цифровая среда уже оперирует такими понятиями как «лайки», «репосты», «конверсии». Люди уже обсуждают, как и где подешевле купить трафик себе на сайт, а накруткой подписчиков в Инстаграме не пользуется только ленивый.

Мы оставили позади (в первой версии книги) весь романтизм и большие надежды, поместив в новую версию экспертное мнение по основным блокам работы с данными.

Читайте, наслаждайтесь и помните: сожжение книг противозаконно.

Алексей Благирев

Глава 1

Что такое Big Data?

Марсианские диалекты

О Больших данных, или Big Data сегодня знают все.

Или еще нет?

Регулярно данные обсуждаются на сложных конференциях, где популярные компании собирают под своими тентами от дождя пару тысяч молодых людей, размещают роботов и плюшевые пуфики, предлагают даже сыграть в игру с ботом, чтобы посетители могли поучаствовать в машинном обучении. Происходит это примерно так: за ограниченное количество ходов игроку необходимо как можно быстрее спойть девушку-робота.

В общем, кто чем пытается покорить свою аудиторию, рассказывая о работе сервисов с данными. Вот только ни у кого нет единой картины.

Одни компании говорят про конфиденциальность, другие – про машинное обучение, перечислять можно бесконечно. Есть даже гипотеза о том, что общая картина больше никому не нужна.

«Как это не нужна?» – спросите вы и поспешите на ее поиски.

Выйдете вы из зоны комфорта, пройдетесь по ключевым конференциям, связанным с данными, прочтете статейки известных умных авторов, но все равно толком ничего не соберется вместе.

Чтобы погрузиться в эту тему, надо взять лопату и копать, копать, копать: по кусочкам собирать смыслы, общаться с разными людьми. Администраторы баз данных могут рассказать вам о том, как настраивать кластеры, а ребята, которые копаются в аналитике, помогут разобрать общую логику процесса.

Только вот почему-то каждый эксперт понимает один и тот же термин по-своему. Будто люди строили Вавилонскую башню из данных, чтобы достучаться до небес, а в конце концов все равно заговорили на разных языках, как написано в Ветхом завете. И эти эксперты вкладывают в, казалось бы, обычные слова, какое-то свое понимание, близкое только им.

Конечно, всех бы мог спасти робот-переводчик, который знает тридцать три наречия межпланетных иезуитов. Но, боюсь, пока его функционал не вырос до такого уровня, придется прикидываться оленеводами, которые впервые услышали о Больших данных. Надо признать, что в некоторых историях мне пришлось разбираться прям с самого что ни на есть нуля, так что расслабьтесь и получайте удовольствие. Будет весело!

А начнем с того, что познакомимся с народом.

#1.

Есть такие важные и бессмертные инженеры по машинному обучению. Задача их проста – проектировать логику и обучать алгоритмы, известные как нейронные сети, заводя в них все новые и новые данные. Если спросить этих инженеров о чем-нибудь другом из области данных, то в большинстве случаев они понятия не будут иметь, о чем их спрашивают – например, кто такие дата-стюарды?

#2.

Дата-стюарды и инженеры качества данных – это такие человечки, которые все правят, чинят и спасают, как Мастер Феликс-младший из игры Fix-It Felix Jr, по ней еще несколько лет назад сняли мультфильм «Ральф». Миссия стюардов и инженеров велика и необъятна. В данных всегда происходит переполох, и нужны те самые brave ребята, которые прибегут со словами «я почию!». Они измеряют искажения в данных и исправляют те самые ошибки, которые допускают пользователи, работая с информацией.

Если спросить у них, в чем роль инженеров по машинному обучению и почему они вообще так называются, то, очень вероятно, что ответа мы не получим. И это нормально.

Разные бригады экспертов занимаются разной работой.

#3.

Архитекторы и аналитики данных – это олицетворение разума. Они опираются на различные правила и методологию, чтобы структурировать данные внутри организации. Например, вместо обозначения таблички «N45» они напишут какое-нибудь гордое «Контрагент» и определят, что в этой табличке должна содержаться информация, касающаяся только контрагента, – например «ИМЯ» / «НАЗВАНИЕ», «ПАСПОРТ» / номер регистрации компании и так далее.

Суть архитекторов и аналитиков – стандартизировать взаимоотношения пользователей с данными и сделать самое главное: навести в этих данных порядок.

Результаты работы этих незаурядных личностей влияют через данные на управление организациями. По-умному их называют data-driven организациями. Они бывают разных типов и устроены все по-разному, но описать data-driven организации или отличить их друг от друга сможет далеко не каждый из описанных специалистов. И это еще один большой вызов.

Разные профессии работы с данными разговаривают на разных языках и формируют собой организации нового типа, где люди не имеют единого представления о том, как ими управлять. Вопрос «чем отличается data-driven организация от data-informed организации?» введет в дичайший ступор не только читателя, но и экспертов, которые работают с данными каждый день.

Перспектива восприятия нового во многом касается наличия практических навыков. Конечно, сегодня мало кто из экспертов имеет руководящий опыт и был тем самым директором по данным, который пытался изменить мир, запуская трансформационные процессы в своей организации для того, чтобы повысить значение использования данных. Это прерогатива людей, которые стоят у руля, а они обычно не разбираются в технике, считая, что она не влияет на принимаемые с точки зрения развития бизнеса решения.

А это все не так. Свойства информационной среды, которые заложены в ней при ее проектировании, оказывают непосредственное влияние на объем и качество принимаемых решений в этой среде.

Когда люди учатся писать на таком языке программирования как Python, им не рассказывают, какие фреймворки проектирования хранилища данных существуют, и что работает, а что уже устарело. Не важно, откуда специалист, интересуется его бизнес или IT, картина везде одна.

Получается, что знание сегментировано, утрировано и преподносится как тайное сокровище, хотя это не так.

Даже разработка на Python проста и похожа на обыкновенную разработку макросов в Excel.

Разбирая управленческие вопросы в организации, в части управления данными, стоит отметить самое важное и, наверное, самое главное. Гештальт, где должно определиться место функции управления данными или так называемого «директора по данным», до сих пор не закрыт и полон споров и противоречий.

ИТ-сфера активно определяет себя как поставщика данных и, соответственно, хочет играть в них ключевую роль, хотя большинство директоров в ИТ-сфере понятия не имеют, как правильно проектировать хранилища данных или функцию управления ими. Все ждут постановки от бизнес-подразделений.

Но сейчас ситуация, конечно, намного лучше, чем несколько лет назад, когда бюджеты заливались в бессмысленные проекты, обреченные на смерть еще в пубертатном периоде использования технологии. Тогда пожилые дядечки в возрасте, которые рулили ИТ-департаментами, с большой долей вероятности были поклонниками Билла Инмона (автора первой книги по созданию хранилища данных) или Ральфа Кимбалла (антагониста Билла). Конечно, согласия между этими концептами мало, и все споры всегда превращаются в дедовские войны на лазерных мечах. Причем, у них разное мнение даже на счет того, как и какими инструментами правильно обрабатывать данные в этих хранилищах.



Например, основной подход – это обрабатывать данные по расписанию, используя специальные инструменты – программы (ETL или ELT) для этой задачи.

Современные эксперты запустили уже свою собственную религию о том, как правильно использовать данные и собирать их в специальную штуку под названием Data Lake. Некоторые из этих экспертов пошли так далеко, что даже отказались от привычных инструментов обработки данных (ETL или ELT), заменив их малопонятной парадигмой, – разбивая все алгоритмы обработки на одинаковые шаги и превращая эти шаги в отдельные программы (сервисы) для создания сложных алгоритмов обработки данных.

Я вам скажу так: все, что можно было когда-либо сделать в Больших данных и машинном обучении – уже сделано. Теперь нужно просто брать существующие методы и сервисы и показывать им новые данные, обучая тем самым алгоритмы адаптироваться.

Перевожу на отечественный. Все, что осталось большинству специалистов – это участвовать в решении только одной задачи, загружать все больше данных для обучения уже существующих алгоритмов. Так ли это? Еще разберемся. Но такие мировые компании как Gartner, уже признают, что роль человека в кооперации с искусственным интеллектом отходит на задний план: необходимо предоставить искусственному интеллекту возможность учиться решать ежедневные задачи. Называется этот подход Augmented Intelligence.

В этой книге вместе представлены различные подходы и методы, которые в совокупности с заумной точкой зрения ведут читателя по новым путям работы с данными. Разобщенность терминологии и понятий, собственно, и подтолкнула меня к идее описать практический опыт тех решений, которые можно использовать для получения практического результата. Это должно помочь определить и выявить новые перспективы в работе с данными, чтобы освоить те дальние рубежи экономики, куда еще не проникла цифровизация.

Что же это все-таки такое и откуда взялось?

Начну со сложного. Понятие Big Data – это такое облако тегов, которое имеет несколько измерений, то есть зависит от ракурса, с которого смотрят.

Пространство координат, благодаря которым можно легко разобраться в том, что такое Big Data, постоянно меняется, создавая отдельные группы понятий, практически не связанных друг с другом. Трудно представить, да?

В Интернете есть известный мем о том, что в одном сперматозоиде содержится 37,5 мегабайт информации ДНК¹. А в результате генерального «салюта» выдается порядка 1500 терабайт.

К слову, в 2013 году мне удалось стать участником крупнейшего внедрения в банковском секторе размером в 51 терабайт. Я внедрял хранилище данных Vertica от Hewlett-Packard. Когда моя команда поместила все транзакции одного крупного банка в это хранилище, у нас получилось немногим больше десяти терабайт. А тут почти в 30 раз больше. В 30!



<https://www.quora.com/Is-it-true-that-a-single-sperm-carries-37-5-MB-of-DNA-info>

Так что самые «большие» данные еще впереди.

А теперь просто. Понятие Big Data можно сравнить с термином «инди-рок», который появился в 80-х годах. Так называли стиль, напоминающий гаражный рок или брит-поп, который играли группы в колледжах или университетах. Благодаря журналистам этот термин обрел множество значений, трактовок и представлений, поэтому инди-роком все стали называть любой стиль музыки, который хотя бы издавна напоминал Oasis, Blur и другие подобные группы.

К чему это? Любую активность, которую я считаю хоть как-то связанной с жизненным циклом данных, я называю Big Data.

Когда понятие попадает в мейнстрим, оно становится #хэштегом, который позволяет привлекать общественное внимание. Да всем плевать на смысл этого хэштега, главное – чтобы было прикольно.

Это происходит, например, потому, что большинство журналистов и копирайтеров не понимают, с чем они столкнулись, что это за технология, и как она будет вести себя дальше. Никого особо не парит, как ее назвать.

Прямо как в издательствах. Это ведь прикольно. Ты пишешь книгу, а ее вычитывает редактор, который не понимает, что это такое.

¹ По некоторым оценкам используется цифра 760,6 мегабайт для XX-хромосом и 735,9 мегабайт для XY-хромосом, или используется оценка в 400 мегабайт на один сперматозоид, что, в принципе, еще больше.

Однажды мне рассказали историю о том, как один высокопоставленный чиновник участвовал в реализации законопроекта в области платежей, но сам при этом ни разу в жизни не сделал ни одного банковского перевода. С Big Data так же.

Лет десять назад термин Big Data воспринимался исключительно как инфраструктурный – под ним понимался специальный класс баз данных, которые позволяли быстро обрабатывать большие объемы информации. То есть, Big Data называлась просто категория железок (серверов), которые умели выполнять определенные вычисления.

Зачем они были нужны? Затем, что обычные железки не умели работать с большим количеством записей. Им было сложно. Памяти не хватало, процессоры грелись, пыхтели бедняги, а скорость расчетов оставляла желать лучшего. Железяки или сервера категории Big Data позволяли решить эту проблему. Потом придумали, что дело вовсе не в железяках, и что можно создавать программное обеспечение («софт»), которое будет работать на самых обычных настольных компьютерах, объединенных в единые узлы. Такие конструкции могли работать параллельно над конкретной задачей из области обработки данных. По-научному их называли «программными комплексами» и «кластерами».

Аудиофайлы, изображения, сложные и слабоструктурированные файлы в то время мало обрабатывались. Существовало сильное ограничение по их исследованию. Для них также требовалось специальное программное обеспечение, а у обычных баз данных не было возможности быстро провести анализ.

Технологии очень быстро эволюционировали. В какой-то момент на смену традиционному понятию Big Data пришел еще один новый термин – Smart Data. Он означал, что «Умные данные» – это сигнал, а «Большие данные» – шум. Таким образом появилась парадигма, разделяющая методы анализа: исследования «шумов» и выявления «сигналов».

За какие-то двадцать лет мир потрянуло так, что он перешел от рынка, где нельзя было купить данные интернет-трафика со «следами», оставленными пользователями, к рынку, где любые данные можно достать в любой момент.

И все бы ничего, но мир перевернулся с ног на голову. Данных стало так много, что их внезапно начали регулировать. Беспощадно и беспристрастно.

Одним из первых пострадал банковский сектор. Все процессы и продукты пришлось пересматривать, потому что теперь даже для кредитного решения банк не может купить данные у кредитного бюро, чтобы проверить потенциального заемщика без его согласия.

В 2018 году появилось регулирование GDPR в Европе. Оно стало настолько жестким, что банки вынуждены были остановить привычные процессы привлечения клиентов в Интернете.

Конечно, если смотреть на все со стороны, то трансформация, которая произошла, – колоссальна. Только представьте, раньше данными занимались где-то внутри IT, в специальных операционных хранилищах (еще они называются ODS), дешево и сердито эти данные сваливались в одну кучу из разных источников. Но теперь мир перешел на новую стадию, данные – это новая нефть, из данных начали строить большой бизнес.

Новые технологии неизбежно приведут человечество к изменению мышления. Об этом уже писали эксперты², анализирующие влияние изучения другого языка на мышление человека. Новые технологии – это еще и переход к новой терминологии, который повлечет за собой новую форму организации взаимодействия потребителей и компаний. А она еще не выработана. Это значит, что данные как актив еще не имеют своей утвержденной и принятой формы по ведению бизнеса.

Поэтому теперь термин Big Data, скорее, отражает новую модель зрелости бизнеса, общества и государства, он больше не является просто олицетворением технологий хранения дан-

² Горелов И.Н., Седов К.Ф. Основы психолингвистики. М., 2001. С. 105–106. Тер-Минасова С.Г. Язык и межкультурная коммуникация. М., 2000. С. 29–30.

ных. Сегодня Big Data подразумевает, что пользователь понимает, как быстро и легально обработать информацию, и как ее структурировать таким образом, чтобы результаты этой работы были понятны окружающим.

Постинформационное общество³

Взрывной рост технологий использования данных приблизил человечество к новой модели своей работы – постинформационному обществу.

Звучит слишком заумно? Вообще префикс «пост» уже много где используется: постистория, постмодернизм, постиндустриальное общество и так далее.

Смысл постинформационного общества в том, что полезные знания среди разнообразной информации теперь могут находить алгоритмы, а не люди, которые их спроектировали.

Ну, то есть, учась в школе, ребенок может решать домашнюю работу вместе с алгоритмами, а не с родителями.

А еще с алгоритмами можно анализировать диагнозы множества пациентов или симптомов одновременно, не полагаясь на человеческую экспертизу.

Это реально?

Ага. Google со своим умным «движком» TensorFlow или Яндекс с CatBoost сделали возможным создание уникальных сервисов с использованием данных в домашних условиях (без каких-либо специальных лабораторий).

И чем больше мы используем алгоритмы, тем больше они учатся. Это можно гордо назвать демократизацией – когда всем понемногу достается кусочек счастья.

Демократизация технологий запустила новые процессы по унификации роли человека в процессах обработки, управления данными и развития искусственного интеллекта. Ручной труд стал больше не нужен. Всякие сверки и контроли – работа, которую теперь можно поручать алгоритмам, и они, в отличие от человека, умеют справляться с ней без ошибок.

Даже последний рубеж, которые машины взять никак не могли – тоже покорился. За несколько лет алгоритмы смогли освоить решение ранее сложных творческих и коллаборативных задач. Причем, этот рывок невозможно было спрогнозировать еще пять лет назад.

Такие системы как Alexa, Siri, Алиса и другие, ускоренными темпами захватывают рынок персональных ассистентов.


В 2015 году эксперты даже в своих самых смелых ожиданиях не могли сойтись в том, что алгоритмы смогут пройти этот рубеж всего лишь через год.

Сегодня есть ощущение, что близится еще один большой рывок, и он может произойти в ближайшие несколько лет.

По одной из гипотез им станет трансформация работы с данными для производств. Тогда собираемая информация будет использоваться с целью анализа и выявления аномалий операционного цикла производства, упрощая управление конвейером, будь это надой молока с установленными датчиками на коровах или завод по производству металлической продукции. Я говорю о едином управлении жизненным циклом продукта или услуги, например – локомотива. Компании взаправду разрабатывают единую концепцию жизненного цикла локомотивов и цифровизации депо. Это уже происходит в России.

³ Горелов И.Н., Седов К.Ф. Основы психолингвистики. М., 2001. С. 105–106. Тер-Минасова С.Г. Язык и межкультурная коммуникация. М., 2000. С. 29–30.

ПОЧЕМУ ЭТО ВАЖНО

ТРАДИЦИОННЫЙ БАНКИНГ	ИНДУСТРИАЛЬНЫЙ БАНКИНГ
<p>ДЛЯ БАНКА</p>  <ul style="list-style-type: none"> • СЛОЖНЫЕ И ДОРОГИЕ ПРОЦЕССЫ • УПРАВЛЕНИЕ СДЕЛКОЙ КАК ПРОЕКТОМ • БОЛЬШИЕ РИСКИ • СЛОЖНОСТЬ МОНИТОРИНГА ОБЪЕКТОВ ЗАЛОГА 	<ul style="list-style-type: none"> • ЦИФРОВЫЕ И ОПТИМИЗИРОВАННЫЕ ПРОЦЕССЫ • АНАЛИЗ РИСКОВ И ПОТРЕБНОСТЕЙ НА ОСНОВАНИИ ДАННЫХ ЧЕРЕЗ ЦИФРОВЫЕ КАНАЛЫ • ДОСТУПНОСТЬ МОНИТОРИНГА СЛОЖНЫХ СДЕЛОК
<p>ДЛЯ КЛИЕНТА</p>  <ul style="list-style-type: none"> • МНОГО ДОКУМЕНТОВ ДЛЯ ЗАКЛЮЧЕНИЯ СДЕЛКИ • МЕДЛЕННАЯ РЕАКЦИЯ БАНКА • ВЫСОКАЯ СТОИМОСТЬ 	<ul style="list-style-type: none"> • ПОЛУЧЕНИЕ ФИНАНСИРОВАНИЯ И КРЕДИТНЫХ ПРОДУКТОВ (ГАРАНТИЙ) БЕЗ ДОПОЛНИТЕЛЬНЫХ БУМАГ И СПРАВОК И БЕЗ ВЫЕЗДА СПЕЦИАЛИСТОВ
<p>ДЛЯ РЫНКА</p>  <ul style="list-style-type: none"> • НЕТ СИНЕРГИИ МЕЖДУ РАЗНЫМИ ФИНАНСОВЫМИ ПРОДУКТАМИ • ВЫСОКИЕ БАРЬЕРЫ ДЛЯ ПОЛУЧЕНИЯ ФИНАНСОВЫХ ПРОДУКТОВ 	<ul style="list-style-type: none"> • ЕДИНАЯ ЭКОСИСТЕМА И МАРКЕТПЛЕЙС ДЛЯ УЧАСТНИКОВ ФИНАНСОВОГО РЫНКА

ПОЭТАПНОЕ ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ БАНКИНГА

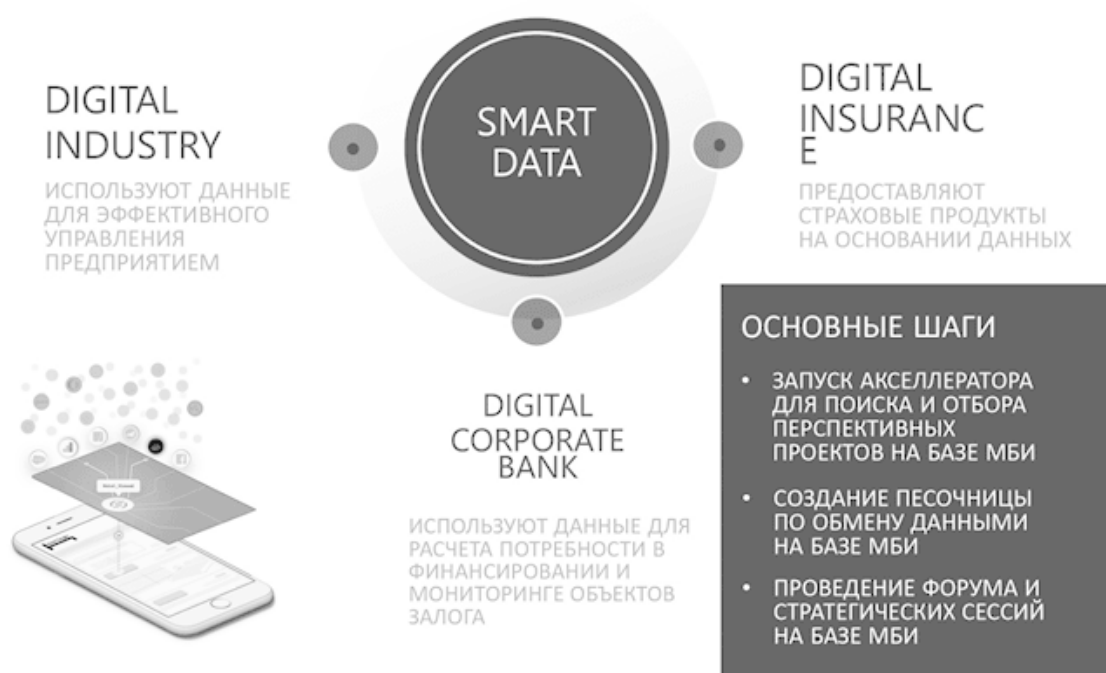


Создание подобных центров управления предприятиями сегодня не имеет технологических барьеров, проблема исключительно в кооперации участников. Решив ее, мир откроет невообразимую возможность создания адаптивной экономики, когда плановые значения заменяются на стандартные нормы производства, которые высчитывают алгоритмы в зависимости от множества факторов.

Но большинство людей все еще мыслит устаревшими категориями.

Для людей, проработавших много лет на производствах, все кажется достаточно понятным и простым. Сначала рисуешь и проектируешь с инженерами деталь, потом готовишь документацию, где прописываешь, как эту деталь обслуживать, потом производишь и, наконец, обслуживаешь.

ДАННЫЕ – ОСНОВА ИНДУСТРИАЛЬНОГО ФИНТЕХА



Казалось бы, все цели ясны, все пути определены – вперед, товарищи!

А на деле все сложнее. Упомянутый выше локомотив может быть старой развалиной без документации. И вот тут людям приходится креативить. Иными словами, инженеры пытаются решить проблему на месте, прямо в депо. Таких примеров много. Что это означает? Только то, что привычного конвейера, который придумал в свое время Генри Форд, больше не существует. Признать это сложно.

Людям хочется верить, что всем можно управлять, а если запустить какой-нибудь сверх-масштабный проект, то вообще получится все вокруг цифровизировать и изменить. Поток данных будут передаваться от производства к другим участникам рынка, например к страховой компании, которая будет выписывать страховой продукт, используя данные без выезда специалистов.

Страховая премия в этом случае может быть ниже рынка как минимум на десять процентов, при этом сам продукт будет более маржинален для страхового бизнеса, так как начнет использовать более точную оценку наступления риска, построенную на данных. Аналогичные продукты с использованием данных может предоставлять банковский сектор. Все соединится в единую экосистему обработки информации и извлечения из нее ценности.

Захватывает, не так ли?

Но на практике никакие масштабные программы не работают, потому что они медленные и не говорят на одном языке с технологией.

Государство, как и бизнес, тоже движется в сторону повышения роли данных. Но как сравнить, где находится государство с точки зрения роли Больших данных?

В 1965 году ООН ввел разделение на развивающиеся рынки и развитые страны.

Не важно, что это разделение уже не работает – его опроверг Ханс Рослинг⁴. Важно, что была попытка предложить систему оценки для сравнения экономического развития стран.

Сейчас, конечно, в национальном плане, единых критериев оценок до сих пор не выработано, хотя каждый малозначимый институт развития пытается предложить свою модель оценки для Больших данных. Короче, не понятно, кто где находится и куда идет.

Например, модель зрелости цифрового государства исследовательской компании Gartner, предполагает пять ступеней зрелости, где data-centric государство – это третья средняя ступень в развитии, этап, когда власть понимает, какие данные есть, когда она отладила процессы их получения и управления качеством.

Россия сегодня успешно завершает переход с первой ступени (E-Gov⁵) на вторую – когда для создания новых информационных сервисов федеральные и муниципальные органы власти предоставляют возможность получения открытых данных, хранящихся в государственных учреждениях. Но сами данные еще разрознены, некачественны, и, по сути, пользоваться ими пока что нельзя.

В Министерстве цифрового развития один из важнейших проектов – создание платформы управления классификаторами (для статистики), когда бизнес и общество могут стать основными источниками данных друг для друга. В идеале разработка платформы может устранить разобщенность классификации, например, номенклатуры товаров. Представьте себе, что больше не надо заполнять никакие накладные, таможенные декларации и прочие бумаги, весь товар регистрируется при производстве и отслеживается. Можно забыть про бумагу.

Единые классификаторы товарных позиций позволяют существенно упростить взаимодействие между несколькими торговыми рынками. В какой-то момент классификаторы позволят создать между ними уникальные зоны свободной торговли. Допустим, что вы приехали в аэропорт и идете через «зеленый коридор», вас никто не трогает, а рядом, в «красном коридоре», происходит принудительный досмотр вещей. Мысленно вы улыбаетесь, радуетесь тому, что вас там нет. Представьте, что таким может быть производство, банкинг, страхование и торговля.

Помимо бизнеса или государства, конечно же, данные сами по себе точно так же оказывают непосредственное влияние на рядового пользователя, например, упрощая процедуру идентификации и получения тех или иных сервисов, в том числе и финансовых. Так, можно удаленно открыть банковский счет, используя только биометрические данные и информацию из учетной записи в государственных системах. Вот вам и опять какие-то новые интерфейсы, которые уже вроде как работают. Пора в них разобраться.

В мире давно существует множество платформ, таких как id.me, tupas, bank.id и других, позволяющих использовать единую учетную запись без необходимости хранить десятки паролей.

Эти платформы формируют будущее цифровой идентичности.

С другой стороны, новое общество, которое уже десятилетием пользуется социальными сетями, электронной почтой и мессенджерами, обзавелось уникальными артефактами и привычками, которых нет как в настоящей жизни, так и в юридической практике.

Такие понятия как «лайк», «репост», «шер», «трансляция», оказывают большее влияние на пользователей, чем пощечина. Из-за лайков люди ссорятся, расходятся или строят целые бизнес модели.

⁴ Ханс Геста Рослинг – шведский врач, академик, профессор Каролинского института по вопросам международного здравоохранения, специалист по статистике и всемирно известный лектор.

⁵ E-Gov – технологично-центрированная, реактивная среда предоставления государственных сервисов в электронном формате. Начальный этап развития цифрового государства, который измеряется процентом покрываемых существующих сервисов в электронном виде.

Появилось такое явление как «цифровое рабство» которое стало следствием того, что данные пользователей не принадлежат им самим. Во многом это помогло цифровым платформам проектироваться без учета общественного диалога. Но парадокс в том, что такой диалог был невозможен на момент зарождения таких платформ ввиду отсутствия пользовательского опыта по использованию данных у самого общества. Соответственно, нас будут ждать еще и этические дилеммы в отношении тех или иных данных.

Сегодня общество уже переживает рефлексии о том, что такое «хорошо» и что такое «плохо» в отношении своих данных. Что делать можно, а что делать не нужно, даже если это явно не запрещено законом, еще предстоит определить. Определение этой границы в использовании данных откроется в самое ближайшее время.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.