

УДК [304.2]

## БОЛЬШИЕ ДАННЫЕ: КАК ТЕХНОЛОГИИ BIG DATA МЕНЯЮТ НАШУ ЖИЗНЬ



**В.В. Шаталова**  
заместитель декана ФКП,  
канд. техн. наук, доцент



**Д.В. Лихачевский**  
декан факультета  
компьютерного  
проектирования, кандидат  
технических наук, доцент



**Т.В. Казак**  
заведующий кафедрой ИПиЭ, член-  
корреспондент Международной  
академии психологических наук,  
доктор психологических наук,  
профессор

Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь.  
E-mail: shatalova@bsuir.by, kazak@bsuir.by.

### **В. В. Шаталова**

Окончила Белорусский государственный университет информатики и радиоэлектроники, факультет компьютерного проектирования. Работает в должности заместителя декана факультета компьютерного проектирования.

### **Д.В. Лихачевский**

Декан факультета компьютерного проектирования.

### **Т. В. Казак**

Заведующий кафедрой инженерной психологии и эргономики Белорусского государственного университета информатики и радиоэлектроники, доктор психологических наук, профессор, член-корреспондент Международной академии психологических наук.

**Аннотация.** В статье представлен теоретический анализ Big Data как совокупность технологий, которые позволяют обрабатывать огромный массив данных, генерируемых в таких источниках, как соцсети, информационные сайты, файлообменники и многие другие ресурсы. Рассматриваются положительные и отрицательные аспекты использования больших данных в повседневной жизни.

**Ключевые слова:** Big Data, большие данные, личные данные, открытые данные, Интернет вещей, информация, перспективы, риски.

Последнее время Big Data определяет наш ежедневный образ жизни и влияют на наши действия и решения различных задач. Сегодня главным потребителем и генератором данных является каждый из нас: ежедневные селфи; модные приложения; социальные посты; любой поисковый запрос; шаг, посчитанный фитнес-трекером; видео, просмотренное на YouTube – огромные и разнообразные массивы данных оставляют цифровой след. Соответственно, чем больше у нас гаджетов, тем больше данных о нас собирают [6].

До недавних пор любые компании управляли традиционными активами – имуществом, деньгами, интеллектуальной собственностью. Цифровая эпоха принесла новый тип активов – данные. Это сырье, из которого производятся прогнозы, инсайты и большие деньги.

Совсем недавно документы на персональном компьютере, рабочем сервере и в личном сотовом телефоне были просто структурированной информацией, сейчас – начиная от личных фотографий в смартфоне, операций по линии мобильной связи, кредитных карт, а также данные различных социальных сетей до сохраненной информации в облаке, а также адресов сайтов в истории, все это пассивно собранные данные, полученные в результате повседневных цифровых операций. Объем таких данных, которые называют большими данными, важно быстро обработать

и структурировать для того, чтобы в будущем получить что-то полезное из них. Частью больших данных становится информация о наших действиях практически в любой области - от покупки финансовых продуктов до записи к врачу, а затем она обрабатывается, анализируется и используется для различных целей [1].

Большие данные собираются из нескольких источников:

- Open data («Открытые данные») - социальная, экономическая и прочая публичная информация о городах, странах, данные о законодательствах, статистика спортивных мероприятий, т. е. любая открытая справочная информация;

- Social networking service («Социальные сети») - то, что мы с вами добровольно рассказываем Facebook, ВКонтакте, Instagram и другое. Даже если пост не публичный, он становится частью Big Data. В среднем пользователи Facebook пишут около 31,25 миллионов сообщений и просматривают 2,77 миллионов видео каждую минуту [1].

- Internet of things, IoT («Интернет вещей») - любые сенсоры в телефоне, смарт-часах, фитнес-трекерах передают самую различную информацию о нас и наших занятиях, а с каждым годом появляется всё больше вещей с интернет-доступом;

- Personal data («Личные данные») - тоже становятся частью больших данных, но данные обезличены: данные вашей медицинской карты, списки дел и другое;

- Коммерческие транзакции - это не только банковские транзакции, но и также любые платежи в интернете.

Свойства Big Data определяются четырьмя «V», каждое из которых, играет определенную роль в анализе и подтверждает их бизнес-ценность [2]:

- VOLUME - масштаб;
- VARIETY - разнообразие;
- VELOCITY - скорость передачи;
- VERACITY - достоверность.

В настоящее время упоминается еще «пятый V», который играет весьма существенную роль. Это VALUE - стоимость.

Однако, не всегда много данных - это хорошо, иногда имеющиеся данные скорее вред, чем польза.

1. «Грязные данные» - неверные, неточные, вводящие в заблуждения, дубликаты, занесенные с нарушением бизнес-процессов, а также с банальными нарушениями пунктуации или правописания. Причиной возникновения таких данных является человеческий фактор, из-за ежедневных рутинных операций, а также критический срок для сбора, когда люди просто делают исправления в ошибках, которые очевидны, на поверхности, не вникая в причину или первоисточник их возникновения.

2. Скрытые или «темные данные» - данные, которые большинство компаний собирают, обрабатывают и хранят как часть повседневной деловой активности, но не используют для каких-либо значимых исследований и решений.

3. Неструктурированные данные - иногда данные доступны, но не готовы к использованию. Их необходимо сделать совместимыми с системой, которая должна использовать данные. Для подготовки данных для алгоритмов машинного обучения, используется информация автономных систем и все они должны быть структурированы. Только после этого неструктурированные данные преобразуются в структурированные и используются для создания и обучения моделей машинного обучения [2].

Свободное движение данных, рост краудсорсинга, новые средства для сбора данных, возникновение больших данных, а также появление искусственного интеллекта и Интернета Вещей - приводят не только к изменениям в жизни общества, но и задают направление развития на ближайшее будущее. Прогресс в области вычислительной техники и информатики теперь позволяет обрабатывать и анализировать большие данные в реальном режиме времени. Новые элементы, полученные в результате обработки таких данных, могут служить важным дополнением к данным официальных опросов и статистической информации, помогая тем самым

анализировать поведение людей и их опыт. Использование этих новых данных совместно с данными, полученными традиционным путем, позволяет производить своевременный анализ информации на качественно более высоком уровне [3]. Возможности, которые открываются перед человечеством при реализации больших данных достаточно велико [4].

В медицине и фармакологии мониторинг текущего состояния пациентов и накопление этих данных поможет проверить, насколько эффективен тот или иной вид терапии или лекарственный препарат. В других ситуациях большие данные позволяют предсказывать распространение вирусных заболеваний и принимать меры для предотвращения эпидемий, что подтверждает глобальная пандемия коронавирусной инфекции, вызванная коронавирусом SARS-CoV-2, которая вызвала активный интерес к использованию больших данных со стороны властей различных государств, которые таким образом пытаются отслеживать распространение инфекции и планировать профилактические меры.

Бизнес-процессы любой компании можно совершенствовать путем анализа больших данных, за счет оптимизации цепочек поставок, сети продаж, инструментов маркетинга, оценивают платежеспособность клиентов, управляют персоналом, прогнозируют загруженность касс, отделений и терминалов. Анализируя полученную информацию о поведении, вкусах и образе жизни клиентов, предприятия и фирмы будут точнее таргетировать рекламу и получать от нее максимальную отдачу.

Большие данные используются для прогнозирования мест и времени возможных преступлений. Алгоритмы учитывают историю совершенных правонарушений и модели криминального поведения, соответственно можно построить карту самых опасных районов и заранее выставить в них полицейские патрули.

От информации о предпочтениях зависит, как выглядят для пользователя новостные ленты в соцсетях. Персонализированная реклама на сайтах, которая тоже подбирается с помощью анализа больших данных, помогает сэкономить время на поиске нужных товаров и услуг в интернет-магазинах или не пропустить выхода интересного фильма.

Учреждения образования благодаря сбору больших данных находят интересный и полезный обучающий контент, а учащиеся получают информацию в том виде (текстовом, видео, аудио), в котором лучше ее усваивают. Так растет интерес к образованию и уровень вовлеченности в обучение.

Новые источники данных, новые технологии и новые аналитические подходы, если они будут использованы ответственно, могут позволить лучше отслеживать прогресс в достижении целей устойчивого развития (рисунок 1) [5].

При этом Big Data несут не только перспективы, но и большие риски (рис. 1.).

1. Необходимо соблюдать принципы конфиденциальности и нормы человеческой морали и этики при работе с информацией, применять необходимые меры защиты данных, чтобы предотвратить неправомерное использование или неправильное обращение с данными. Большая часть новых данных собирается пассивно на основе «цифровых следов», оставляемых людьми, и анализируется на основе определенных алгоритмов, что приводит к включению некоторых лиц или групп лиц в определенную категорию и нанести им потенциальный вред.

2. Велика вероятность дискриминации и предвзятого отношения к определенным лицам или группам лиц. Уже заметен разрыв между теми, кто имеет доступ к данным и теми, кто не имеет, что в дальнейшем приведет к новым границам неравенства: тем кто обладает информацией и кто ею не обладает. Многие люди лишены возможности быть частью нового мира данных и информации по самым различным причинам: языковые барьеры, бедность, отсутствие образования, отсутствие технологической инфраструктуры, удаленность, предрассудки и дискриминация [5].

3. Многие компании и различные службы будут стремиться контролировать наше поведение для уменьшения собственных рисков. Например, страховщикам пригодятся данные о физической активности и привычках каждого клиента: если вероятность выплаты возрастает, это станет поводом повысить индивидуальный тариф; банкам будет полезна информация о том, какие статьи расходов у человека преобладают, чтобы принять решение о целесообразности выдачи ему

кредита, максимальной сумме и процентной ставке и другое.



Рисунок 1. Большие данные для целей развития и гуманитарных действий [5]

4. Большую опасность представляет дискриминация определенных групп населения. Например, неоднократно замечено, что люди с темным цветом кожи априори идентифицируются различными системами распознавания лиц как потенциальные преступники, а на сайтах с вакансиями, где пользователи адресно получают предложения по трудоустройству, мужчинам, как правило, демонстрируются более интересные объявления.

5. Одна из актуальнейших проблем – утечка информации. Повсеместное внедрение цифровых технологий делает управление бизнесом и государством на всех уровнях прозрачным, что повышает риски. С продвижением Хайтек – стратегии «Индустрия 4.0» возникают особые риски, связанные не только с мошенничеством, но и с настоящими диверсиями (перехват управлением промышленными объектами, транспортом и энергетикой, сферой здравоохранения и многое другое) [6].

6. Доступность самой разной информации о каждом человеке поможет нечистоплотным политтехнологам манипулировать общественным сознанием за счет использования ботов и таргетированного продвижения поддельных новостей. Такие манипуляции девальвируют политические процессы и угрожают социальной стабильностью [4].

Современные технологии сами по себе нейтральны, а вот влияние технологий на общество – вопрос их правомерного применения. Спрогнозировать вектор развития, дать оценку их использования, нанесут ли они больше вреда или пользы мы не можем сказать однозначно. Однако в наших силах – установить четкие цели, задачи и ограничения, связанные с обработкой больших данных. И тогда стремление улучшить жизнь совершенно точно не приведет к наступлению технократической антиутопии.

### Список литературы

- [1] Как Big Data влияет на нашу жизнь [Электронный ресурс]. – Режим доступа: <https://zen.yandex.ru/media/id/592e5e13d7d0a6f53d9a2c24>
- [2] Какие данные скорее вред, чем польза? [Электронный ресурс]. – Режим доступа: <https://www.datanomics.ru/articles/kakie-dannye-skoree-vred-chem-polza/>.
- [3] Инновации в сфере данных для целей развития [Электронный ресурс]. – Режим доступа: <https://www.un.org/ru/sections/issues-depth/big-data-sustainable-development/index.html>
- [4] Осторожно, данные собираются: как технологии big data меняют мир [Электронный ресурс]. – Режим доступа: <https://finparty.ru/opinions/166219/>.
- [5] Инновации в сфере данных для целей развития [Электронный ресурс]. – Режим доступа: <https://www.un.org/ru/sections/issues-depth/big-data-sustainable-development/index.html>
- [6] Кушнир, Е. А. Противостояние XXI века / Е. А. Кушнир, Л. А. Телегина. в 2-х томах. Т. 1. – Саров: ФГУП «РФЯЦ-ВНИИЭФ», 2019. – 354 с.

## BIG DATA : HOW BIG DATA TECHNOLOGIES ARE CHANGING OUR LIVES

**V.V. SHATALOVA**

*Deputy Dean of the Faculty of  
Computer-Aided Design, PhD,  
Associate Professor*

**D.M. LIKHACHEVSKY**

*Dean of the Faculty of  
Computer Design, candidate of  
Technical Sciences, Associate  
Professor*

**T.V. KAZAK**

*Head of the Department of  
Engineering Psychology and  
Ergonomics, Corresponding  
Member of the International  
Academy of Psychological Sciences,  
Doctor of Psychological Sciences,  
Professor*

*Belarusian State University Informatics and Radioelectronics, Republic of Belarus*

*E-mail: shatalova@bsuir.by, kazak@bsuir.by*

**Abstract.** The article presents a theoretical analysis of Big Data as a set of technologies that allow you to process a huge array of data generated in such sources as social networks, information sites, file sharing sites, and many other resources. The positive and negative aspects of using big data in everyday life are considered.

**Keywords:** Big Data, big data, personal data, open data, Internet of Things, information, prospects, risks.