



**ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР** МГТУ им. Н. Э. Баумана

Data Science

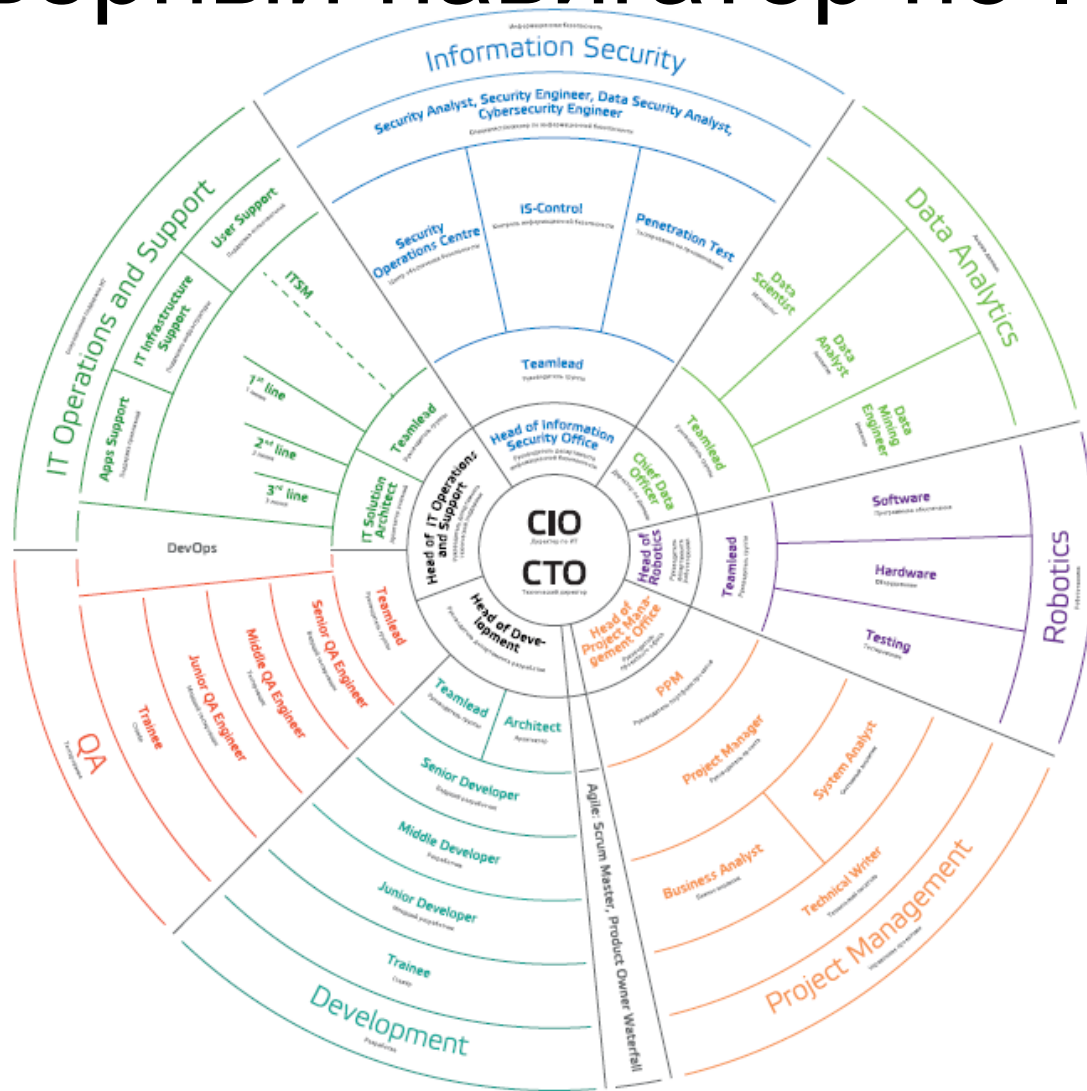
Знакомство с основными понятиями

Корпоративное обучение на базе
Образовательного центра МГТУ им. Н. Э. Баумана
под управлением МИЦ «Композиты России»
Докладчик: Панфилов И.А. — канд. техн. наук, доцент

Почему вы решили стать аналитиком данных?



Карьерный навигатор по IT



Анализ данных

Data Analytics

Data Scientist
Методолог

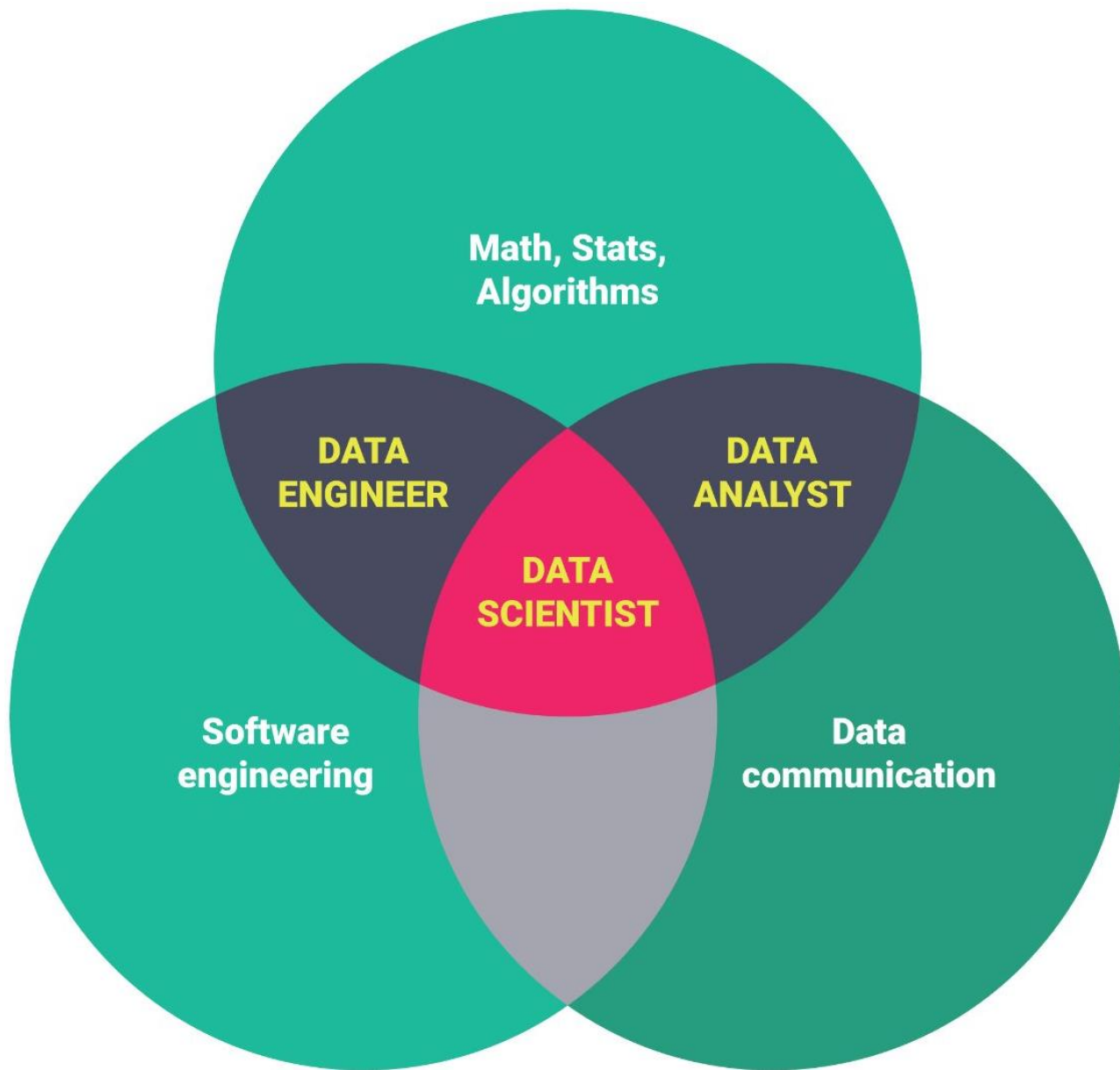
Data Analyst
Аналитик

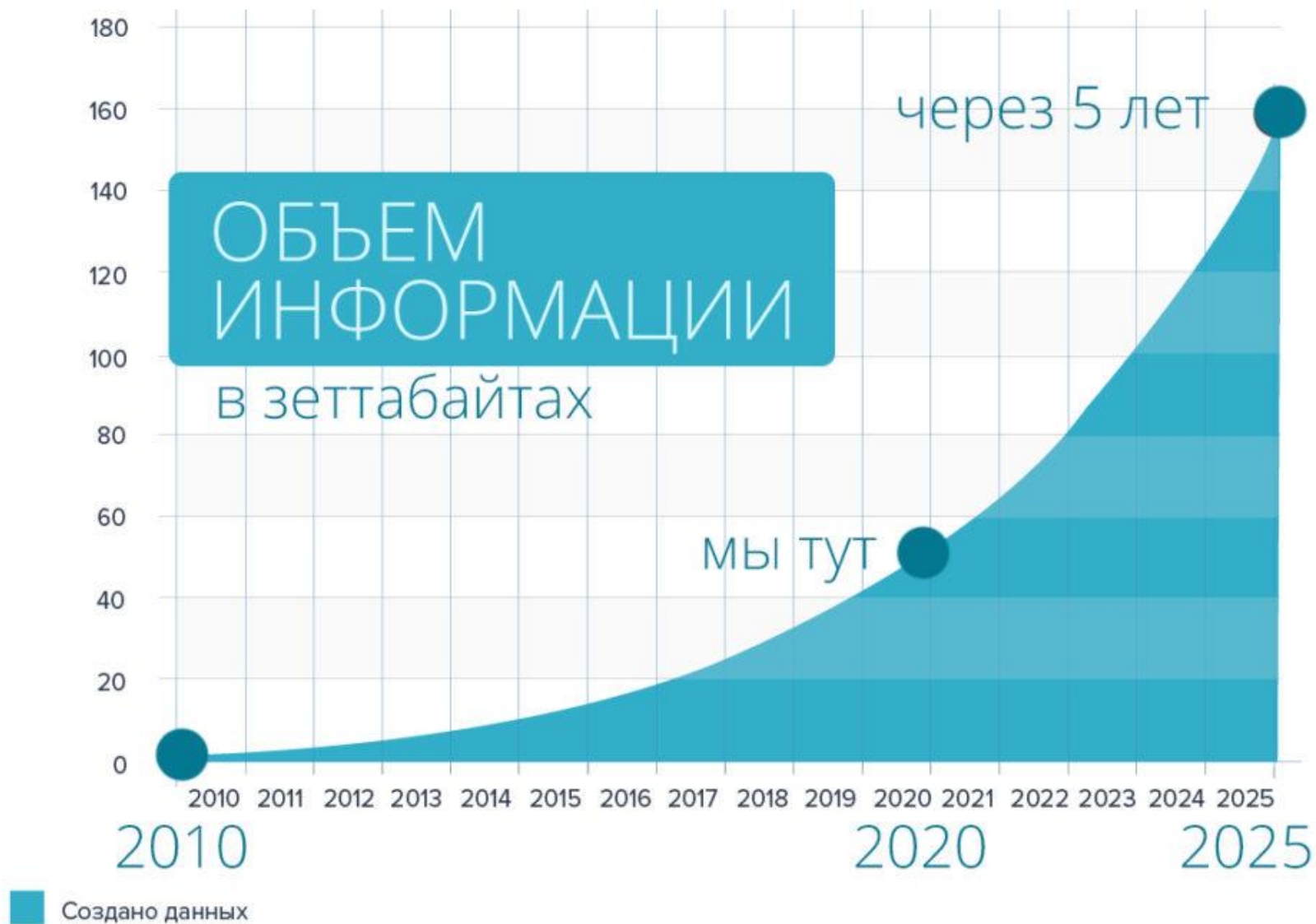
Data Mining Engineer
Инженер

Teamlead
руководитель группы

Chief Data Officer
Директор по данным







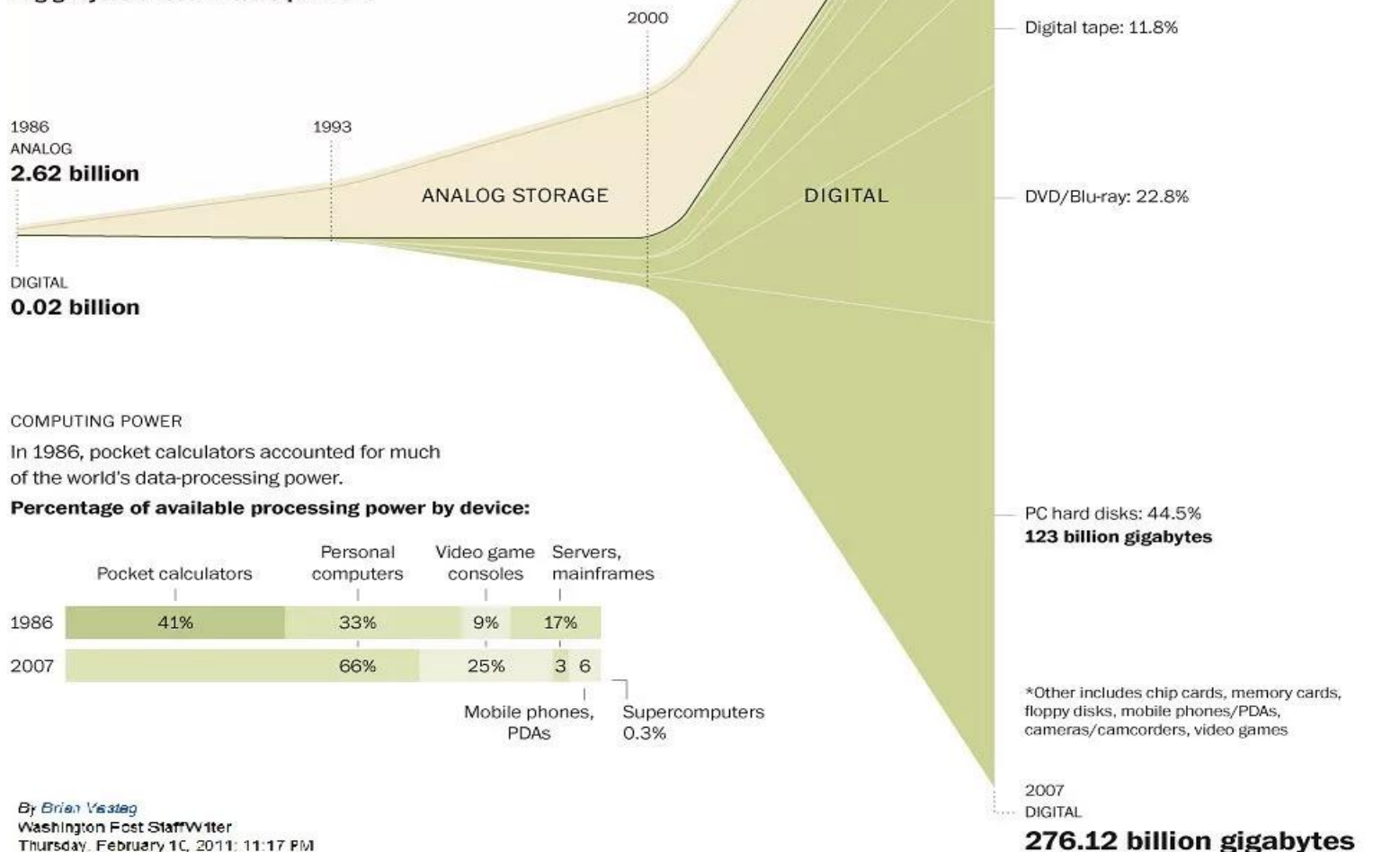
The Washington Post

Exabytes: Documenting the 'digital age' and huge growth in computing capacity

THE WORLD'S CAPACITY TO STORE INFORMATION

This chart shows the world's growth in storage capacity for both analog data (books, newspapers, videotapes, etc.) and digital (CDs, DVDs, computer hard drives, smartphone drives, etc.)

In gigabytes or estimated equivalent



By Brian Vesteg
 Washington Post Staff Writer
 Thursday, February 10, 2011: 11:17 PM

Big Data

огромные наборы разнообразных данных. Огромные, потому что их объемы такие, что простой компьютер не справится с их обработкой, а разнообразные — потому что эти данные разного формата, неструктурированные и содержат ошибки. Большие данные быстро накапливаются и используются для разных целей.

Под **Big Data** будем понимать не какой-то конкретный объём данных и даже не сами данные, а методы их обработки, которые позволяют эффективно обрабатывать информацию.

Источники данных



Источники медицинских данных

МЕДИЦИНСКИЕ ДАННЫЕ

- Изображения (КТ, МРТ, видео-операции и др.)
- Текст (ЭМК, клинические протоколы и рекомендации и др.)
- Звук (записи голоса пациентов, хрипов, кашля и др.)
- Сигналы (ЭЭГ, ЭКГ, данных с прикроватных мониторов и носимых устройств и др.)
- Генетические (NGS, ДНК-чип и др.)

ФАРМАКОЛОГИЧЕСКИЕ ДАННЫЕ

- Данные клинических исследований
- Данные о препаратах
- Данные о продажах
- Данные фармнадзора

ФИНАНСОВЫЕ ДАННЫЕ

- Данные о движении денежных средств ФОМС, МО
- Данные о движении денежных средств страховых компаний



АДМИНИСТРАТИВНЫЕ ДАННЫЕ

- Данные о работе МО (загрузка кабинетов, оборудования, расписание приемов и др.)
- Реестры ФРМО, ФРМР, НСИ
- Данные о жалобах и отзывах на МО
- Данные о распространении заболеваний, эпидемиях и др.

СТРАХОВЫЕ ДАННЫЕ

- Страховые отчеты
- Отчеты по скорингу клиентов
- Данные о претензиях

ВНЕШНИЕ ДАННЫЕ

- Демографические данные
- Биомедицинская литература
- Данные с форумов о здоровье
- База медицинских онкологий

Data Driven Decision Making

ОПТИМИЗАЦИЯ

Предиктивное моделирование

Прогнозирование/экстраполяция

Статистический анализ

Уведомления

Запрос данных / Углубленный анализ

Отчеты ad hoc

Стандартные отчеты

Прогноз

Понимание
инсайтов

Анализ
ретроспективы

24 апреля 2013. Ложное сообщение в Twitter агентства Associated Press о двух взрывах в Белом доме вызвало распродажу на фондовом рынке.

Трех минут до опровержения хватило, чтобы отправить индекс Dow Jones вниз на 145 пунктов (почти 1%) и уменьшить капитализацию входящих в него компаний на \$200 млрд.

Доступ к данным

Мобильные системы и поисковики

1. Мобильные системы

- Записи телефонных разговоров
- SMS и интернет-трафик

2. Яндекс

- Местоположение
- Уникальный идентификатор мобильного устройства
- Уникальный идентификатор пользователя
- Имя, фамилия, электронный адрес и фотография пользователей (если есть), авторизованных с помощью учетной записи на Яндексе
- Записи микрофона
- Изображения из галереи устройства
- Данные об установках приложения пользователей, авторизованных с помощью учетной записи на Facebook
- Данные о движении по маршруту пользователей, с помощью и на Facebook

Банки

- | | |
|----------------------------------|---------------------------------|
| – имя | – Размер дохода |
| – Пол | – Счет за коммунальные услуги |
| – Возраст | – выписка из банка |
| – Адрес, включая почтовый индекс | – Банковские справки |
| – Национальность | – Источник средств |
| – Эл. адрес | – номер банковского счета |
| – Номер телефона | – Дата рождения |
| – Мобильный номер | – место рождения |
| – Временный адрес проживания | – биометрические данные |
| – Адрес занятости | – личный номер |
| – Имя работодателя | – Расходы |
| – Личный статус | – Информация о родственниках |
| – Номер удостоверения личности | – Возраст детей |
| – Внутренний номер паспор | – Дети |
| – Номер загранпаспорта | – Данные о здоровье |
| – Семейный статус | – образование |
| – профессия | – Квалификация |
| – Страна налогообложения | – Членство в любых организациях |
| – Налоговый номер | – Судимости |
| – Сумма дохода | – статус банкротства |
| – Источник дохода | – религиозные верования. |

Этический контракт или положение О защите данных (GDPR) (ЕС) созданное для гармонизации законов о конфиденциальности данных по всей Европе, обеспечивая защиту личных данных граждан ЕС и изменяя подход организаций, как европейских, так и международных, к конфиденциальности данных. Постановление было одобрено парламентом ЕС в апреле 2016 года и вступило в силу в мае 2018 года во всех странах ЕС.



- Более жёсткие требования к предоставлению информации. Однозначные и прозрачные формы разрешений на сбор данных. Человеку однозначно и ясно что он даёт согласие на обработку данных. Исключить согласие на обработку данных «по умолчанию»
- Определить возрастной порог, после которого человек может давать согласие на обработку данных
- Расширить понятие «персональные данные» до генетических и биометрических данных
- Псевдонимизация — техника обработки персональных в таком виде, чтобы её не могли соотнести с конкретным лицом без дополнительной информации
- Расширенные права граждан на управление собственной информацией, в том числе право на «перенос данных, забвение, право не подвергаться существенному автоматизированному принятию решений.
- Обязательства по уведомлению об утере данных.

Для всех участников цифрового пространства наиболее существенными становятся риски управления большими данными и поведения в социальных сетях. Особенно высокими эти риски становятся для определенных профессий, к которым предъявляются конкретные требования, связанные с репутацией и служебными обязанностями.

Эти профессии можно назвать «сенсорными».

В первую очередь это касается:

государственных служащих



преподавателей и
учителей



работников
здравоохранения



То есть всех тех, кто по роду своей деятельности максимально включен в социальное взаимодействие.



edu.bmstu.ru

+7 (495) 120-30-75

E-mail: edu@bmstu.ru

**Москва, ул. 2-я Бауманская,
дом 5, стр. 1**