



Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников

НАУКА О ДАННЫХ

БОЛЬШИЕ ДАННЫЕ

НАУКА О ДАННЫХ



Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников



НАУКА О ДАННЫХ

DATA SCIENCE

— это раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме.



Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников



ПЕТЕР НАУР

ДАТСКИЙ УЧЕНЫЙ В ОБЛАСТИ
ИНФОРМАТИКИ

— один из пионеров компьютерной науки, который в своей книге 1974 года ввел определение **«науки о данных»** — дисциплины, которая изучает жизненный цикл цифровых данных.

НАУКА © ДАННЫХ



Вклад
в будущее
СБЕР

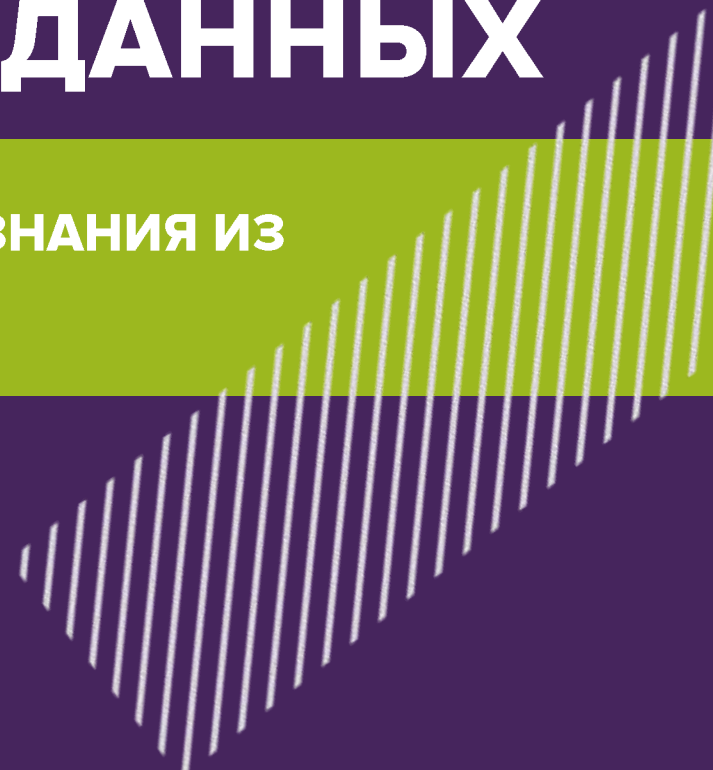


АКАДЕМИЯ
искусственного интеллекта
для школьников



ГЛАВНАЯ ЗАДАЧА НАУКИ О ДАННЫХ

ОБНАРУЖИТЬ ЗАКОНОМЕРНОСТЬ В ДАННЫХ И ИЗВЛЕЧЬ ЗНАНИЯ ИЗ
ДАННЫХ В ОБОБЩЕННОЙ ФОРМЕ





Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников



БОЛЬШИЕ ДАННЫЕ

BIG DATA

— это совокупность непрерывно увеличивающихся объемов информации одного контекста, но разных форматов, а также методов и средств для эффективной и быстрой обработки.



Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников

Термин **«большие данные»** ввел в широкое употребление Клиффорд Линч, редактор журнала Nature, подготовивший к 3 сентября 2008 года специальный выпуск с темой «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных?».



КЛИФФОРД ЛИНЧ

РЕДАКТОР ЖУРНАЛА NATURE

БОЛЬШИЕ ДАННЫЕ



Вклад
в будущее
СБЕР



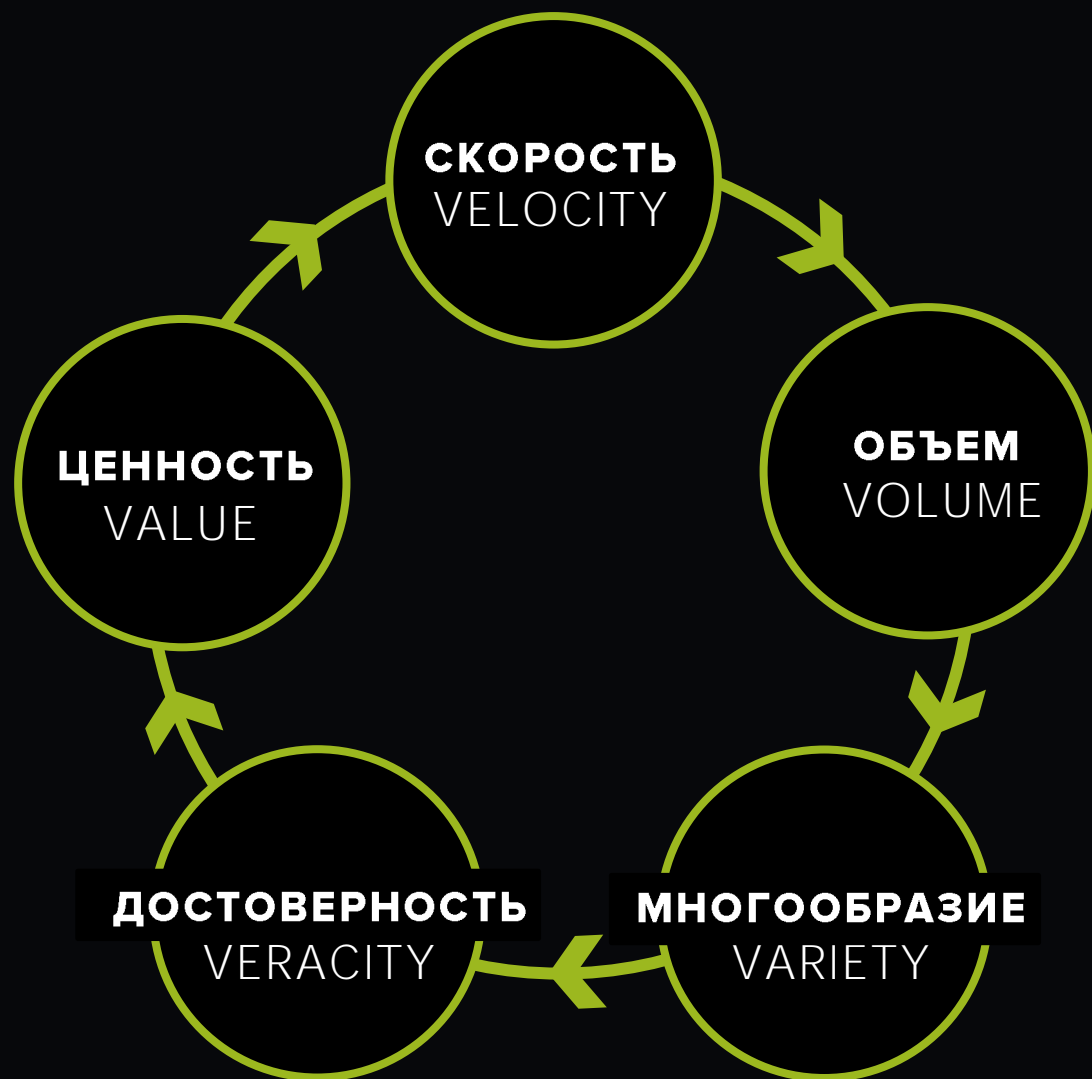
АКАДЕМИЯ
искусственного интеллекта
для школьников



КАКИЕ ДАННЫЕ ЯВЛЯЮТСЯ БОЛЬШИМИ?

БЛАГОДАРЯ РОСТУ ВОЗМОЖНОСТЕЙ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ, ОПИСАННОМУ В ЗАКОНЕ МУРА, ОБЪЕМ ДАННЫХ НЕ МОЖЕТ ЯВЛЯТЬСЯ ТОЧНЫМ КРИТЕРИЕМ ТОГО, ЯВЛЯЮТСЯ ЛИ ОНИ БОЛЬШИМИ.

ГЛАВНОЙ ХАРАКТЕРИСТИКОЙ BIG DATA ЯВЛЯЕТСЯ СТЕПЕНЬ ИХ СТРУКТУРИРОВАННОСТИ И ВАРИАНТОВ ПРЕДСТАВЛЕНИЯ.



СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА

Характеристика	Традиционная база данных	База больших данных
Объем информации	От гигабайт до терабайт	От петабайт до эксабайт
Способ хранения	Централизованный	Децентрализованный
Структурированность данных	Структурирована	Полуструктурирована или неструктурирована
Модель хранения и обработки данных	Вертикальная модель	Горизонтальная модель
Взаимосвязь данных	Сильная	Слабая



Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников

ПРЕИМУЩЕСТВО БОЛЬШИХ ДАННЫХ И АНАЛИТИКИ

Большие данные дают возможность получать более полные ответы, так как они предоставляют больше информации.

Более полные ответы означают, что вы можете быть более уверены в достоверности данных, что обеспечивает абсолютно новый подход к решению задач.



Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников

ИСТОЧНИКИ БОЛЬШИХ ДАННЫХ

ИНТЕРНЕТ

соцсети, блоги, СМИ, форумы,
сайты, Интернет вещей (Internet
of Things, IoT)

КОРПОРАТИВНАЯ ИНФОРМАЦИЯ

транзакции, архивы, базы
данных и файловые хранилища

ПОКАЗАНИЯ ПРИБОРОВ

датчиков, сенсоров,
регистраторов и пр.



Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников

ИСПОЛЬЗОВАНИЕ БОЛЬШИХ данных

БОЛЬШИЕ ДАННЫЕ МОЖНО ПРИМЕНЯТЬ В САМЫХ РАЗЛИЧНЫХ СФЕРАХ ДЕЯТЕЛЬНОСТИ: ОТ ВЗАИМОДЕЙСТВИЯ С КЛИЕНТАМИ ДО АНАЛИТИКИ.

**Машинное
обучение**

**Качество
обслуживания
клиентов**

**Предиктивное
обслуживание**

**Обнаружение
несанкционированного
доступа и соблюдение
нормативных
требований**

**Разработка
продуктов**



Вклад
в будущее
СБЕР



АКАДЕМИЯ
искусственного интеллекта
для школьников



БОЛЬШИЕ ДАННЫЕ
ЭТО БОЛЬШИЕ ВОЗМОЖНОСТИ,
НО И НЕМАЛЫЕ ТРУДНОСТИ