

**Уральский государственный экономический университет**  
**ФОРМАЛИЗАЦИЯ ИНФОРМАЦИИ И Big Data**

02.03.03 -Математическое обеспечение и администрирование информационных систем,  
направленность (профиль) -разработка и администрирование информационных систем

<http://vikchas.ru>

**Лекция 8. Тема «Как БФМ ИИ устраняют  
неопределенность и сложность Big Data»**

**Часовских Виктор Петрович**  
доктор технических наук, профессор кафедры  
ШИиКМ, ФГБОУ ВО «Уральский государственный  
экономический университет

Екатеринбург 2026

## Введение

Современный мир производит данные с колоссальной скоростью. Финансовые транзакции, медицинские снимки, логистические маршруты, тексты, изображения, видео, данные сенсоров, социальные сети, IoT-устройства — всё это формирует то, что принято называть **Big Data**, или большими данными.

Однако наличие большого объема данных само по себе не означает наличие знания. Главная проблема Big Data заключается не только в масштабе, но и в том, что данные часто:

- неполные;
- противоречивые;
- шумные;
- неструктурированные;
- быстро меняющиеся;
- поступают из разных источников;
- имеют скрытые закономерности.

Именно здесь на первый план выходят **большие фундаментальные модели ИИ**, которые позволяют извлекать смысл из сложных и неопределенных информационных сред.

## Что такое Big Data

Big Data обычно описывают через несколько характеристик, известных как **5V**.

### 1 Volume — объем

Данные создаются в огромных масштабах: терабайты, петабайты и эксабайты информации.

Пример:

крупная торговая сеть ежедневно собирает данные о покупках, перемещении товаров, работе складов, действиях пользователей на сайте и в мобильном приложении.

### 2 Velocity — скорость

Данные поступают в реальном времени или почти в реальном времени.

Пример:

банковская система должна мгновенно анализировать транзакции, чтобы выявлять мошенничество.

### **3 Variety — разнообразие**

Данные бывают структурированными и неструктурированными:

- таблицы;
- тексты;
- изображения;
- аудио;
- видео;
- графы;
- логи;
- геоданные.

### **4 Veracity — достоверность**

Данные могут содержать ошибки, дубликаты, пропуски, искажения и ложные сигналы.

### **5 Value — ценность**

Главная цель анализа Big Data — получить практическую пользу: прогноз, решение, автоматизацию, оптимизацию или новое знание.

## **В чем состоит неопределенность Big Data**

Неопределенность в больших данных возникает по нескольким причинам.

### **1 Неполнота данных**

Не вся информация доступна. Например, в медицинской карте пациента могут отсутствовать важные анализы, но решение всё равно нужно принимать.

### **2 Шум и ошибки**

Данные могут быть искажены из-за человеческого фактора, технических сбоев или особенностей сбора информации.

### **3 Неоднозначность**

Один и тот же текст, сигнал или поведение пользователя может иметь разные интерпретации.

Например, фраза «банк закрылся» может означать:

- финансовая организация прекратила работу;
- отделение банка завершило рабочий день;
- пользователь говорит о речном берегу в другом контексте.

#### **4 Скрытые зависимости**

Связи между событиями не всегда очевидны. Например, спрос на товар может зависеть не только от цены, но и от погоды, сезона, новостей, социальных трендов и поведения конкурентов.

#### **5 Изменчивость среды**

Данные быстро устаревают. Модель, хорошо работавшая год назад, может хуже работать сегодня из-за изменения рынка, поведения пользователей или внешних условий.

### **В чем состоит сложность Big Data**

Сложность Big Data проявляется не только в объеме, но и в структуре.

#### **1 Высокая размерность**

Объекты могут описываться тысячами или миллионами признаков. Например, пользователь онлайн-сервиса характеризуется историей покупок, кликами, поисковыми запросами, временем активности, устройством, геолокацией и другими параметрами.

#### **2 Неструктурированность**

Большая часть данных существует в форме текста, изображений, видео и аудио. Такие данные трудно анализировать классическими методами.

#### **3 Разнородность источников**

Для решения одной задачи нужно объединять данные из CRM, ERP, социальных сетей, сенсоров, документов и внешних API.

#### **4 Нелинейность связей**

В реальности зависимости между переменными редко бывают простыми. Малое изменение одного фактора может приводить к большим последствиям.

## 5 Масштаб вычислений

Обработка Big Data требует распределенных систем, облачных платформ, GPU/TPU и специальных архитектур хранения.

### Что такое БФМ ИИ

**Большие фундаментальные модели ИИ** — это модели, обученные на масштабных и разнообразных наборах данных, которые могут быть адаптированы к множеству задач.

К ним относятся:

- большие языковые модели;
- мультимодальные модели;
- модели компьютерного зрения;
- генеративные модели;
- модели для анализа временных рядов;
- эмбединговые модели;
- трансформеры для текста, изображений, аудио и табличных данных.

Ключевая особенность фундаментальных моделей — они не создаются под одну узкую задачу с нуля, а формируют универсальное представление данных, которое затем можно использовать для разных сценариев.

### Основной принцип: переход от данных к представлениям

Классический подход часто требует ручного проектирования признаков. Эксперт должен заранее определить, какие параметры важны.

Например, для анализа клиентов нужно вручную задать:

- частоту покупок;
- средний чек;
- давность последней покупки;
- категорию товаров;
- канал взаимодействия.

БФМ ИИ работают иначе. Они автоматически строят **векторные представления**, или **эмбеддинги**, которые отражают скрытый смысл объектов.

Текст, изображение, пользователь, товар или документ превращаются в числовой вектор. Если объекты похожи по смыслу, их векторы оказываются близки в многомерном пространстве.

Это позволяет:

- находить похожие объекты;
- классифицировать данные;
- выявлять аномалии;
- строить рекомендации;
- связывать данные разных типов;
- снижать размерность;
- устранять неоднозначность за счет контекста.

## **Как БФМ ИИ уменьшают неопределенность**

### **1 Использование контекста**

Фундаментальные модели анализируют не отдельный признак, а контекст.

Например, слово «мышь» может означать животное или компьютерное устройство. Модель определяет значение по окружению:

- «мышь подключена к ноутбуку»;
- «мышь живет в подвале».

В Big Data это особенно важно, потому что данные часто фрагментарны и неоднозначны.

### **2 Вероятностное мышление**

БФМ не всегда дают один жесткий ответ. Они оценивают вероятности разных вариантов.

Например, при анализе медицинских данных модель может показать, что один диагноз вероятен на 70%, второй — на 20%, третий — на 10%.

Это позволяет принимать решения с учетом риска, а не создавать иллюзию полной определенности.

### **3 Обнаружение скрытых паттернов**

Фундаментальные модели способны выявлять зависимости, которые трудно обнаружить вручную.

Например:

- необычные комбинации симптомов;
- скрытые признаки мошенничества;
- ранние сигналы отказа оборудования;
- изменение настроений клиентов;
- нестандартные маршруты логистики.

### **4 Работа с неполными данными**

Модели могут восстанавливать недостающий контекст или делать выводы при неполной информации.

Например, если в клиентском профиле отсутствует часть данных, модель может использовать похожие поведенческие паттерны других клиентов.

Важно понимать: модель не «знает» пропущенные факты наверняка, но может оценить наиболее вероятные варианты.

### **5 Фильтрация шума**

БФМ ИИ умеют выделять значимые сигналы среди большого количества нерелевантной информации.

Например, при анализе отзывов клиентов модель может отделить эмоциональный шум от конкретных проблем продукта: доставки, качества, цены или сервиса.

## **Как БФМ ИИ уменьшают сложность**

### **1 Автоматическое извлечение признаков**

Одна из главных возможностей фундаментальных моделей — автоматическое выделение признаков из данных.

Для текста это могут быть:

- тема;
- тональность;
- намерение пользователя;
- факты;

- сущности;
- причинно-следственные связи.

Для изображений:

- объекты;
- сцены;
- дефекты;
- медицинские признаки;
- изменения во времени.

Для бизнес-данных:

- сегменты клиентов;
- сценарии поведения;
- риски;
- прогнозные индикаторы.

## **2 Снижение размерности**

Модель сжимает огромные массивы информации в компактные представления. Например, документ на несколько страниц можно представить вектором из нескольких сотен или тысяч чисел.

Это делает возможным:

- быстрый поиск;
- кластеризацию;
- визуализацию;
- сравнение;
- дедупликацию;
- построение рекомендательных систем.

## **3 Унификация разных типов данных**

Мультимодальные модели могут объединять текст, изображение, звук, видео и табличные данные.

Например, в медицине можно анализировать одновременно:

- жалобы пациента;
- результаты анализов;

- снимки МРТ;
- историю болезни;
- врачебные заключения.

В промышленности:

- показания датчиков;
- видеонаблюдение;
- техническую документацию;
- журналы обслуживания;
- сообщения операторов.

#### **4 Масштабируемость**

Большие модели хорошо интегрируются с распределенными вычислениями, облачной инфраструктурой и потоковой обработкой данных.

Они позволяют строить системы, которые работают с большими объемами информации не вручную, а автоматически.

#### **5 Обобщение на новые задачи**

Фундаментальные модели можно дообучать, настраивать или использовать через промпты для разных задач.

Одна и та же базовая модель может применяться для:

- анализа документов;
- поддержки клиентов;
- прогнозирования спроса;
- поиска аномалий;
- генерации отчетов;
- интеллектуального поиска;
- автоматического контроля качества.

## **Основные технологии БФМ ИИ в Big Data**

### **1 Трансформеры**

Трансформеры используют механизм внимания, который позволяет модели определять, какие части данных важны для текущего вывода.

Это особенно полезно для длинных текстов, временных рядов, последовательностей событий и мультимодальных данных.

## **2 Эмбединги**

Эмбединги превращают сложные объекты в числовые векторы. Благодаря этому можно выполнять семантический поиск и сравнение данных.

Пример:

поиск не только по точным словам, а по смыслу. Запрос «проблемы с оплатой» может найти документы, где написано «платеж не проходит» или «карта отклонена».

## **3 RAG-системы**

RAG, или Retrieval-Augmented Generation, объединяет генеративную модель с поиском по базе знаний.

Система сначала находит релевантные документы, а затем формирует ответ на их основе.

Это снижает риск галлюцинаций и помогает работать с актуальными корпоративными данными.

## **4 Дообучение и адаптация**

Модель можно адаптировать под конкретную предметную область:

- юриспруденцию;
- медицину;
- финансы;
- промышленность;
- образование;
- государственное управление.

## **5 Графы знаний**

Графы знаний позволяют формализовать связи между сущностями: людьми, организациями, событиями, продуктами, документами.

Совместное использование графов и фундаментальных моделей помогает лучше объяснять связи и строить более надежные выводы.

# **Практические примеры**

## **1 Финансы**

БФМ ИИ применяются для:

- выявления мошенничества;
- анализа транзакций;
- оценки кредитного риска;
- мониторинга новостей;
- автоматизации комплаенса;
- обработки обращений клиентов.

Модель может объединять транзакционные данные, историю клиента, тексты заявок, новости и поведенческие сигналы.

## **2 Медицина**

В медицине модели помогают:

- анализировать снимки;
- обрабатывать медицинские записи;
- выявлять риски заболеваний;
- сопоставлять симптомы;
- подбирать клинические рекомендации.

Они уменьшают неопределенность, объединяя разрозненные данные пациента в единую картину.

## **3 Промышленность**

На производстве БФМ ИИ помогают:

- прогнозировать поломки оборудования;
- анализировать сенсорные данные;
- выявлять дефекты по изображениям;
- оптимизировать обслуживание;
- анализировать техническую документацию.

## **4 Ритейл**

В торговле модели используются для:

- прогнозирования спроса;
- персонализации рекомендаций;

- анализа отзывов;
- оптимизации запасов;
- динамического ценообразования.

## **5 Государственное управление**

ИИ может помогать в:

- анализе обращений граждан;
- выявлении социальных проблем;
- прогнозировании нагрузки на инфраструктуру;
- обработке нормативных документов;
- мониторинге чрезвычайных ситуаций.

### **Ограничения и риски**

Важно понимать: БФМ ИИ не устраняют неопределенность полностью. Они ее уменьшают и делают управляемой.

#### **1 Галлюцинации**

Генеративные модели могут создавать правдоподобные, но неверные ответы. Поэтому в критических задачах необходимы проверка, источники и контроль экспертов.

#### **2 Смещения в данных**

Если обучающие данные содержат предвзятость, модель может ее воспроизводить.

#### **3 Непрозрачность**

Большие модели часто сложно интерпретировать. Это особенно важно в медицине, финансах и праве.

#### **4 Конфиденциальность**

Работа с большими данными может затрагивать персональные и коммерчески чувствительные сведения.

#### **11.5 Стоимость**

Обучение и эксплуатация больших моделей требуют значительных вычислительных ресурсов.

---

## **Как правильно внедрять БФМ ИИ в Big Data-проекты**

Эффективное внедрение требует не только модели, но и архитектуры.

### **1 Определить бизнес-задачу**

Нужно начинать не с технологии, а с вопроса: какую неопределенность мы хотим уменьшить?

Например:

- снизить риск мошенничества;
- сократить время обработки документов;
- повысить точность прогнозов;
- улучшить качество обслуживания;
- обнаружить причины отказов оборудования.

### **2 Подготовить данные**

Необходимы:

- очистка;
- нормализация;
- разметка;
- дедупликация;
- управление качеством;
- защита персональных данных.

### **3 Выбрать модельный подход**

Возможны варианты:

- готовая модель через API;
- локальное развертывание;
- дообучение;
- RAG;
- гибридная архитектура;
- ансамбль моделей.

### **4 Встроить контроль качества**

Нужно измерять:

- точность;
- полноту;
- устойчивость;
- объяснимость;
- скорость;
- стоимость;
- безопасность.

## **5 Обеспечить участие человека**

В критически важных сценариях должен использоваться подход human-in-the-loop: модель предлагает решение, а человек проверяет и утверждает его.

### **Главный вывод**

БФМ ИИ устраняют неопределенность и сложность Big Data не магически, а за счет нескольких механизмов:

1. преобразования сложных данных в смысловые представления;
2. анализа контекста;
3. вероятностной оценки вариантов;
4. обнаружения скрытых закономерностей;
5. автоматического извлечения признаков;
6. объединения разных типов данных;
7. работы с неполной и шумной информацией;
8. масштабируемой обработки больших массивов;
9. адаптации к новым задачам;
10. интеграции с базами знаний и экспертными системами.

Иными словами, большие фундаментальные модели превращают Big Data из хаотичного массива информации в систему, с которой можно работать: искать, объяснять, прогнозировать и принимать решения.

## **Заключение**

Big Data создает одновременно огромные возможности и огромные проблемы. Чем больше данных, тем выше сложность, неоднозначность и риск ошибочных выводов.

БФМ ИИ становятся инструментом, который помогает справляться с этими вызовами. Они не заменяют полностью аналитиков, инженеров, врачей, управленцев или исследователей, но усиливают их способность видеть закономерности, работать с неопределенностью и принимать более обоснованные решения.

Будущее анализа данных, вероятнее всего, будет гибридным: большие модели, классическая аналитика, предметные эксперты, графы знаний, правила, статистика и автоматизированные системы будут работать вместе.

Главная задача — не просто собрать больше данных, а превратить их в надежное, проверяемое и полезное знание.