



Экосистема данных: как  
ИИ и сотрудничество  
бизнеса/науки/регулятора  
формируют будущее

# Какие факторы являются ключевыми для развития исследований на основе больших и альтернативных в России?

- **Свободный доступ к данным** при соблюдении требований законодательства и правообладателей данных,
- Наличие полноценной линейки **эффективных инструментов и методологий** для анализа данных,
- **Партнерство** с разработчиками аналитического инструментария и правообладателями данных для анализа, основанное на взаимном экономическом (коммерческом) интересе.



# Что сдерживает развитие этого направления (развитие исследований на основе больших и альтернативных данных в России) в нашей стране?



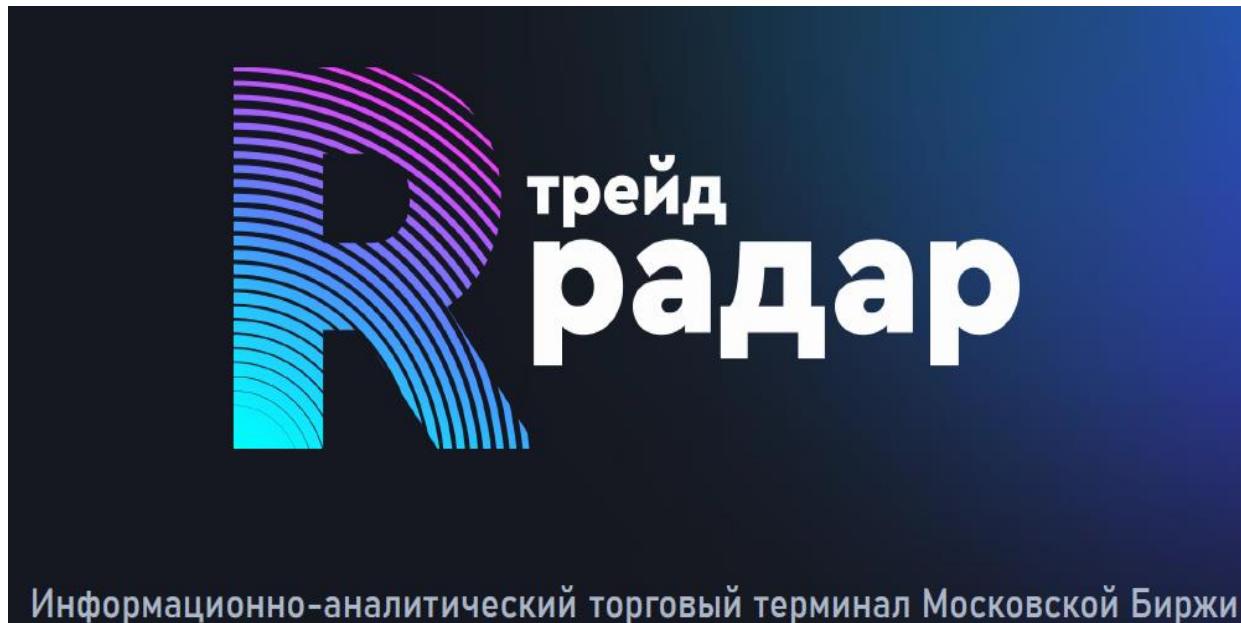
Развитие направления тормозится из-за:

- **Данные - как "закрытый клуб"** (платные сервисы, запреты на иностранные инструменты, сложности с доступом).
- **Мало «грамотных» людей** (не хватает специалистов, которые умеют обрабатывать большие данные).
- **Недостаток "топлива" для роста** (финансирование проектов слабое, ресурсы ограничены).
- **«Старые» технологии** (отсутствие облачных платформ и современных решений).
- **«Платформы» вразнобой** (отрасли не цифровизированы, данные в разных форматах).
- **Правила не определены** (нет чётких стандартов для работы с данными и ИИ)
- **Все по своим дорожкам** (государство, бизнес и наука не достаточно плотно сотрудничают)



# Какие форматы партнерств между научным сообществом, бизнесом и регуляторами могли бы ускорить развитие этого направления?

- **Стажировки на бирже** - студенты участвуют в исследованиях и тестировании сервисов (например, Трейд Радар, AlgoPack).
- **Лаборатории с реальными инструментами** - бизнес помогает вузам оснастить классы терминалами, учебными серверами и архивными данными (опыт с Финансовым университетом).
- **Конкурсы-хакатоны** - студенты и учёные решают аналитические задачи от бизнеса



# На какие примеры из зарубежного опыта стоит обратить внимание?



В качестве успешного примера из зарубежного опыта давайте рассмотрим партнёрство компаний **Nasdaq** и **Google** в облаке для их анализа.

**Nasdaq** и **Google** объединились, чтобы создать экосистему для анализа больших данных, включая финансовые, рыночные и альтернативные источники (например, соцсети, IoT\*).

Их сотрудничество строится на интеграции данных **Nasdaq** (рыночные данные, индексы, аналитика) с мощной инфраструктурой **Google Cloud**, что позволяет пользователям получать доступ к инструментам машинного обучения, обработки и визуализации данных.



\*IoT (Internet of Things) - это сеть физических устройств, машин, датчиков и бытовой техники, которые подключены к интернету и могут собирать, передавать и анализировать данные. Например, умные часы, которые отслеживают здоровье, или холодильник, который сам заказывает продукты, когда они заканчиваются.

# Как сделать доступ к данным удобным для исследователей?



В настоящее время Мосбиржа рассматривает возможность реализации **«Песочницы данных» на базе ИСС (API Московской Биржи)** - это экосистема для автоматизации обработки, анализа и монетизации рыночных данных, предоставляемых через ИСС (Информационно-статистический сервер).

Платформа позволит исследователям (например, аналитикам, трейдерам или компаниям) не только получать первичные данные о ценах, объемах торгов и других показателях, но и трансформировать их в более сложные, специализированные наборы данных с помощью AI-агентов и интеграционных инструментов (iPaaS).

Интересно?



# Какие направления в работе с данными являются наиболее перспективными?

- **Анализ взаимосвязей** между поведением финансовых рынков и новостным фоном в прошлых периодах с возможностью прогнозирования поведения рынков в будущем на фоне анонсов ожидаемых новостей или мнений экспертов, размещенных в публичном пространстве.
- **Анализ стратегий** успешных инвесторов с возможностью автоследования.



# Какие источники альтернативных данных особенно интересны для анализа?

МОЕХ GROUP

## 1. Социальные сети и публичные коммуникации

Например, *Twitter, Reddit, Telegram* - анализ настроений и трендов для прогноза реакции рынка.

## 2. Геопространственные данные

Спутники, GPS, датчики - отслеживание активности в реальном мире (трафик, строительство, загруженность магазинов).

## 3. Финансовые ордерлоги и алгоритмы трейдеров

История заявок/сделок на бирже, стратегии успешных инвесторов.

## 4. Публичные и открытые базы

*Google Trends, патенты, социальные метрики (подписчики, лайки)*.

## 5. IoT и промышленные датчики

Энергопотребление, производственные мощности, логистика.

## 6. Данные из других отраслей

Ритейл (посещаемость магазинов), медицина (заболеваемость), образование (курсы ИИ).



# Как развитие ИИ технологий может поменять требования к данным необходимым для исследований?



- **Расширение типов данных:** переход от структурированных таблиц к неструктурированным данным (текст, изображения, аудио, видео, IoT-сенсоры).
- **Данные в реальном времени:** необходимость потоковой обработки и анализа «живых» данных для актуальных прогнозов.
- **Качество и чистота:** повышенные требования к очистке, нормализации и валидации данных (принцип «мусор на входе - мусор на выходе»).
- **Синтетические данные:** рост использования сгенерированных данных для обучения моделей при дефиците реальных.
- **Междисциплинарность:** интеграция данных из разных областей.
- **Масштабируемость:** потребность в инфраструктуре для обработки растущих объёмов данных.
- **Объяснимость:** требование к интерпретации решений ИИ («прозрачные» модели для критических сфер).