

# Вячеслав ДЕГТЯРЕВ:

## «Перспективы ИИ зависят от того, удастся ли компаниям перейти от точечных пилотов к системному внедрению»



### – Расскажите о подразделении K2 «НейроТех»: на чем вы специализируетесь?

– Сегодня бизнес активно внедряет ИИ, и все чаще возникает разрыв между классической и ИИ-инфраструктурой. Внутренним ИТ-командам не всегда хватает экспертизы, чтобы подобрать под ИИ-задачи «железо», софт и выстроить эксплуатацию. K2Тех, опираясь на 19-летний опыт строительства и сопровождения суперкомпьютерных кластеров, запустила самостоятельное подразделение K2 «НейроТех».

K2 «НейроТех» – ИИ-интегратор полного цикла. Мы проектируем, строим, масштабируем и сопровождаем ИИ-инфраструктуру и кластеры, а также занимаемся консалтингом и обучением команд. Подразделение закрывает задачи в области LLMOps, MLOps и ИИ-кибербезопасности, обеспечивая управление моделями и защиту данных.

Конкуренция в сфере ИИ набирает обороты, несмотря на риски и регуляторные препятствия, которыми сопровождается внедрение технологии. Создание, масштабирование и развитие ИИ-инфраструктуры требует соответствующих компетенций. Это одно из условий перехода от экспериментов и пилотных проектов к промышленной эксплуатации решений. О вызовах, с которым сталкиваются предприятия на этом пути, подходах к внедрению ИИ-инструментов рассказал **Вячеслав ДЕГТЯРЕВ**, руководитель по развитию продуктовых решений «K2 НейроТех».

### – Какие продукты формируют портфель K2 «НейроТех»?

– В портфеле – индивидуальные проекты под ключ, готовые программно-аппаратные комплексы для быстрого развертывания корпоративного ИИ и высокопроизводительных вычислений в локальном контуре: ПАК-AI, ПАК-ML, ПАК-НРС, а также комплексная ML-платформа в облаке K2 Cloud. Эти решения объединяют GPU-кластеры, системы хранения, сертифицированные операционные системы и инструменты для управления жизненным циклом ИИ-моделей.

### – На чем вы сейчас фокусируетесь в развитии K2 «НейроТех»?

– Мы стремимся быть связующим звеном между ИИ и бизнесом, адаптируя ПАКи под задачи промышленности, финсектора и других отраслей. Важный фокус – партнерства с российскими вендорами ПО и оборудования. Мы тестируем новые отечественные решения, чтобы благодаря вариативной архитектуре ПАКов заказчики могли строить ИИ-инфраструктуру на российском стеке в соответствии с регуляторными нормами и без потери гибкости.

### – Если смотреть на рынок в целом, какие тенденции вы считаете ключевыми?

– Первая тенденция – переход от пилотов к масштабированию: у компаний уже есть успешные эксперименты с ИИ, но перевести их в промышленную эксплуатацию удастся не всегда. Основной барьер – отсутствие выстроенной инфраструктуры, данных и процессов под ИИ. Вторая – рост спроса на инфраструктуру: бизнес интересуется не только выбором моделей, но и вычислительные мощности, управление данными и MLOps-платформы. Третья – ужесточение регуляторики и требований к безопасности, что стимулирует развитие on-premise и гибридных решений.

### – С какими вызовами чаще всего сталкиваются ваши заказчики и отрасль?

– Главный вызов – дефицит GPU и специалистов, которые умеют строить и эксплуатировать ИИ-инфраструктуру: LLMOps- и MLOps-инженеров, архитекторов, платформенные команды. Второй – качество и разрозненность данных: во многих компаниях данные существуют фрагментарно и противоречат друг другу, в результате ИИ масштабирует хаос. Третий вызов – регуляторная неопределенность:



Выступление на конференции «ИИПром»-2026

формируются требования к маркировке, локализации данных, ответственности за результаты работы ИИ, и бизнес ждет платформ, которые обеспечат такое соответствие «из коробки».

сейчас компании переходят от отдельных технологий к запросу на платформы и процессы. Бизнес перестает воспринимать инфраструктуру как железо и набор GPU. Вместе с заказчиками

**– Если смотреть вперед на 5–7 лет, что вы считаете основным драйвером роста российского ИИ-рынка?**

– Перспективы зависят от того, удастся ли компаниям перейти от точечных пилотов к системному подходу с выстроенной инфраструктурой и общей стратегией. По нашим прогнозам, к 2030 г. до 75 % вычислительных затрат в сфере ИИ придется на инференс – эксплуатацию моделей, а не обучение. Это изменит архитектуру ЦОДов: на первый план выйдут эффективность и управляемость ресурсов. Рынок движется к сервисным моделям – GPU-as-a-Service, гибридным экосистемам, где облако используется для экспериментов, а локальный контур – для чувствительных данных, и к отечественным платформенным решениям, обеспечивающим технологический суверенитет и соответствие регуляторным требованиям. ■

**К 2030 г. до 75 % вычислительных затрат в сфере ИИ придется на инференс – эксплуатацию моделей, а не обучение.**

**– Как в этой картине меняются ожидания бизнеса от ИИ-инфраструктуры?**

– Раньше ИИ-инфраструктура создавалась под точечные запросы и ручное сопровождение,

мы создаем ИИ как устойчивый внутренний сервис с понятной экономикой, политиками доступа, соблюдением требований ИБ и предсказуемыми сроками запуска новых сценариев.