Чек-лист: конвейер аналитики обучения

Чек-лист на одну страницу: схема событий, выборка, выбор LRS, модель склада данных в российском контуре и меры приватности по 152-ФЗ, которые нужно заложить с первого дня.

А. Сбор – сначала спроектируйте схему событий

- [ ] Выберите стандарт заранее: xAPI (гибкий) или Caliper (вузы)
- [ ] Задайте глаголы и стабильные URL-id для каждой активности до плеера
- [ ] Внедрите xAPI Video Profile: собирайте played-segments и progress
- [ ] Один registration на попытку; учащийся – по account-id, не по email

Б. Транспорт – пакеты, выборка, устойчивость к сбоям сети

- [ ] Отправляйте утверждения массивом; пакет атомарен – проверяйте до отправки
- [ ] Выборка: глаголы действий + ping progress каждые ~10 секунд
- [ ] Не отправляйте покадровые события time-update в продакшн
- [ ] Локальная очередь, отправка после восстановления связи; statement-id исключает дубли

В. Хранилище – взять готовый LRS или написать свой

- [ ] Берите готовый LRS, соответствующий стандарту (open source в РФ), если масштаб не требует своего
- [ ] Проверьте соответствие xAPI 2.0 (IEEE 9274.1.1-2023) и наличие нужных ресурсов
- [ ] LRS – неизменяемый журнал, а не аналитическая база
- [ ] LRS, склад и резервные копии – только в РФ: зарубежные базы для ПДн запрещены (ст. 18 ч. 5)

Г. Склад и приватность – модель и защита

- [ ] Загружайте сырые данные, затем ELT в звёздную схему (например, ClickHouse в РФ)
- [ ] Заранее агрегируйте тяжёлые played-segments в сводные таблицы по когортам
- [ ] Минимизируйте данные при сборе; соответствие id и личности храните отдельно
- [ ] Заранее задайте сроки хранения и путь удаления по actor
- [ ] Подрядчик – по поручению оператора (ст. 6); план уведомления РКН за 24 часа