

## 1 — Три AI-системы (оценивайте отдельно)

|               |                                                                               |
|---------------|-------------------------------------------------------------------------------|
| <b>Подбор</b> | Рекомендации: кого/что показать. Two-tower retrieval + ранжирование → облако. |
| <b>Защита</b> | Trust & safety: воронка + верификация. Гибрид: облако + устройство + люди.    |
| <b>Медиа</b>  | Живые и генеративные медиа: блюр, AR, шумоподавление, субтитры → устройство.  |

## 2 — Две границы, управляющие сборкой

|                   |                                                                                     |
|-------------------|-------------------------------------------------------------------------------------|
| ГРАНИЦА МОДЕРАЦИИ | AI оценивает; решают система + человек → стройте строгость по тяжести.              |
| ГРАНИЦА ПЛАТФОРМЫ | охват + несовершеннолетние = регулируемая платформа → обязанности задаёт аудитория. |

## 3 — Где работает каждая система + бюджет сэмплирования live-видео

Облако ..... подбор/ранжирование + тяжёлые классификаторы + хеш-матч известного.  
Устройство .... живые видеофункции + захват селфи верификации; приватно, real-time.  
Комната людей сложные решения + апелляции (апелляция — закон ЕС). Заложите в бюджет.  
Бюджет:  $1 \text{ ч} \times 30 \text{ fps} = 108 \text{ } 000 \text{ кадров}$ . Каждый кадр  $\sim \$108/\text{ч}$ ; 1 кадр / 2 с  $\sim \$1,80/\text{ч}$  (в 60 раз).

## 4 — Воронка модерации + купить / собрать / построить

ВОРОНКА: хеш-матч известного + ML-классификаторы (новое) + жалобы + люди + апелляция.  
Купить API ..... дни; Hive/Rekognition/Azure/Sightengine; плата за элемент; БЕЗ CSAM-матча.  
Собрать своё ..... недели-месяцы; открытые модели + PhotoDNA/PDQ + движок политик; вы владеете.  
Построить ..... месяцы; только когда trust & safety — это отличие. Для большинства рано.

## 5 — Соответствие + безопасность — до публичного запуска (контекст, не юр. совет)

- ☐ ЕС (DSA): уведомление-реагирование, обоснование, апелляции, отчёты о прозрачности (штраф до 6% оборота).
- ☐ Великобритания (OSA): «высокоэффективные» проверки возраста + борьба с незаконным (надзор Ofcom).
- ☐ США (TAKE IT DOWN): удалять заявленные интимные изображения / дипфейки за 48 часов (надзор FTC).
- ☐ EU AI Act ст. 50: машиночитаемая маркировка AI-сгенерированных медиа (с августа 2026).
- ☐ Проверка человека: оценка возраста по лицу + документ для спорных; liveness-селфи; согласие на биометрию.
- ☐ Специалист по CSAM (напр. Thorn Safer) на каждом пути; авто-блок по хешу и отчёт в NCMEC.