Чек-лист по деплою нейронного ABR

Pensieve · Comyco · Kairos · ABRL · Fugu · SODA — когда нейронный ABR едет в продакшен, а когда нет

Четыре формы деплоя, где нейронный ABR выигрывает

Форма	Когда выигрывает
Большой корпус трасс	**У оператора сотни тысяч реальных user session-трасс (Netflix, YouTube, Disney+, Meta, Amazon, крупные MVPD). Меньшие операторы редко проходят data-планку без партнёрства с вендором.**
Характеризуемая сеть	**Проводной residential broadband, fixed wireless, характеризованные мобильные сети. Гетерогенная мобильность (спутник, 4G↔5G hand-offs, публичный Wi-Fi) плохо обобщается из коробки.**
Подавление переключений отслеживается	**Приложения, где продуктовая команда измеряет switches-per-minute как brand quality, выигрывают непропорционально — Comyco, Kairos, SODA все подавляют переключения агрессивнее простых правил.**
Команда для поддержки модели	**Production-деплои нуждаются в retraining-конвейере, A/B инфраструктуре, observability-инструментах и детерминированном fallback. Без них lifecycle-стоимость затмевает QoE-выигрыш.**

Четыре production failure-mode

Failure-mode	Симптом и фикс
Несоответствие распределения трасс	**Политика, обученная на одном сетевом распределении, работает на другом. Pensieve проиграл Fugu в Puffer именно по этой причине. Фикс: переобучить in situ на собственных трассах деплоя.**
Стоимость обучения и операц. нагрузка	**RL-обучение хрупкое; reward shaping — искусство; свежие корпуса могут требовать тысячи GPU-часов. Imitation learning (Comyco, Fugu) дешевле, но всё ещё нужен expert и поддерживаемый пайплайн.**
Непрозрачность в отладке	**Нейронная политика — чёрный ящик. Стройте observability-инструменты — saliency maps, decision logging, counterfactual replays — до деплоя, а не после первой клиентской эскалации.**
Несоответствие reward-функции	**Политика максимизирует reward, на котором обучается. Engagement-shaped reward (switches, TTFF, time-to-1080p) обгоняет QoE-shaped reward в продакшене. SODA делает этот аргумент явно.**

Пять рычагов настройки, реально двигающих QoE

Рычаг	Как настроить
Метод обучения	**RL, когда нет эксперта (Pensieve, ABRL). Imitation learning, когда offline-эксперт существует (Comyco, Fugu). Imitation — дефолт 2026.**
Где сидит модель	**Чистая политика (Pensieve, ABRL, Comyco), когда команда может поглотить непрозрачность. Обученный предсказатель в MPC (Kairos, Fugu) для production-отладки. Гибрид — дефолт 2026.**
Формирование reward	**Академический дефолт: качество − λs·switch − λr·rebuffer. Production: добавить TTFF, time-to-1080p, viewing-duration. Переобучайте при изменении весов.**
Каденс переобучения	**Ежемесячно для стабильных проводных сетей; еженедельно для быстрых мобильных; непрерывное lifelong learning (L-Comyco), где пайплайн позволяет.**
Стратегия fallback	**Обязательный детерминированный fallback на BOLA или MPC, когда политика возвращает патологическое действие. Логируйте инциденты, A/B засекайте дрейф, переобучайте или откатывайте.**

Три production-деплоя, документированные на масштабе

Деплой	Архитектура	Production-результат
Facebook ABRL (2020)	RL-политика. 30M+ сессий/неделю.	**+1,6% средний битрейт · −0,4% stall. arXiv 2008.12858.**
Stanford Puffer / Fugu (NSDI 2020)	MPC + in-situ предсказатель. 8 131 ч · 3 719 пользователей.	**0,13% stall ratio · обгоняет Pensieve в поле. nsdi20-paper-yan.**
Amazon Prime Video SODA (SIGCOMM 2024)	Smoothness-optimised DP с QoE-гарантиями.	**−88,8% переключений · +5,91% длит. просмотра.**