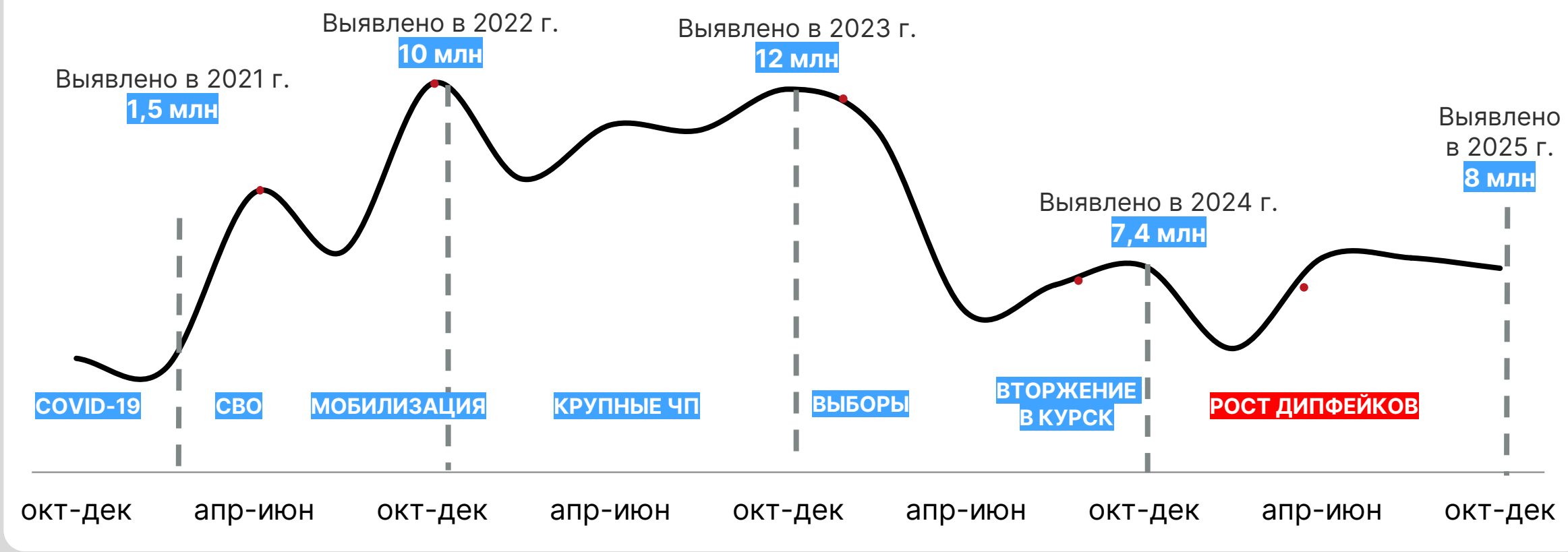
ОПЫТ ПРОТИВОДЕЙСТВИЯ АНО "ДИАЛОГ РЕГИОНЫ" ГЕНЕРАТИВНОМУ КОНТЕНТУ:

ТРЕНДЫ, ИНСТРУМЕНТЫ И ВЫЗОВЫ

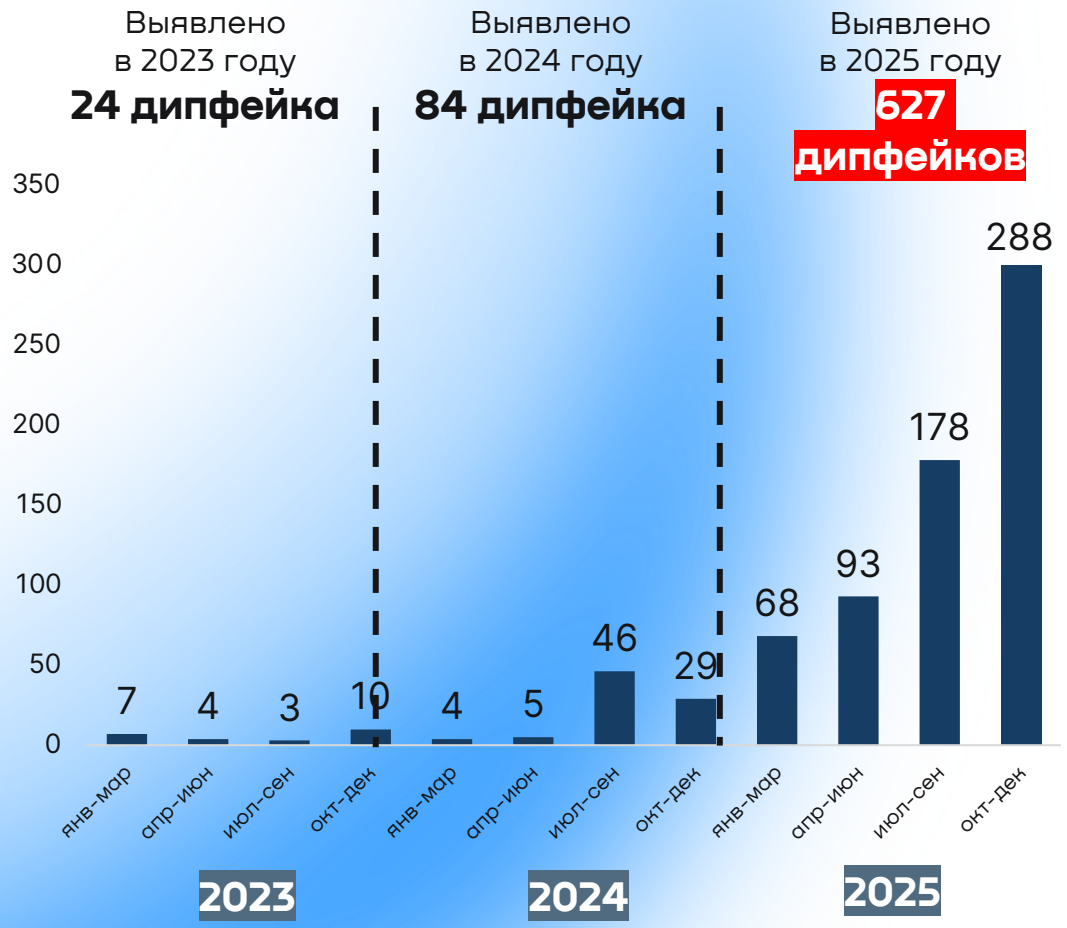
ДИАЛОГ

2026

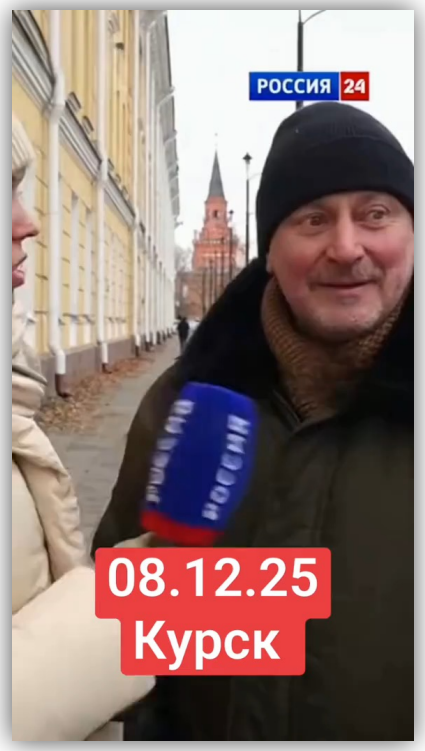
ДИНАМИКА ФЕЙКОВ В РОССИЙСКОМ СЕГМЕНТЕ ИНТЕРНЕТА



ДИПФЕЙКИ



Рост доли контента диффузионных моделей, генерирующих дипфейки без видео-доноров.



ДИПФЕЙК: Житель Курска резко критикует ОИВ в эфире канала «Россия 24»

МЕХАНИКИ СОЗДАНИЯ ДИПФЕЙКОВ В 2025 ГОДУ

- Большая часть дипфейков **(81%)** создается с применением технологий FaceSwap, LipSync или их сочетания.
- В 2025 году **17%** дипфейков созданы диффузионными моделями на основе промта без видео-донора. В 2024 году подобная механика не использовалась
- **2% дипфейков** создано с помощью наложения сгенерированного аудио на подлинную видеозапись.

КАК ВЫСТРОЕНА РАБОТА С ФЕЙКАМИ

СБОР ДАННЫХ

Система автоматического мониторинга СМИ и соцсетей

81 000 СМИ

2,4 млрд аккаунтов в соцсетях

ОБРАЩЕНИЯ В ЧАТ-БОТЫ

ЗЕФИР

(транскрибирование, мониторинг видео, дипфейки)

> 8000 источников

МАРШРУТИЗИ-РОВАНИЕ

ИНФОРМАЦИОННАЯ СИСТЕМА ПРОТИВОДЕЙСТВИЯ ФЕЙКАМ

МЕДИЙНАЯ ОТРАБОТКА

СМИ И ЭКСПЕРТЫ

ТАСС · LENTA.RU · Школа экономики и менеджмента РАНХиГС · 360 · URA.RU · Комсомольская правда

ЛАПША медиа

СОЦСЕТИ

IT-РЕШЕНИЯ ПО БОРЬБЕ С ФЕЙКАМИ

МОДУЛИ «ЕДИНОЙ СИСТЕМЫ» АНО «ДИАЛОГ РЕГИОНЫ»

ИАС «FAKECHECKER»

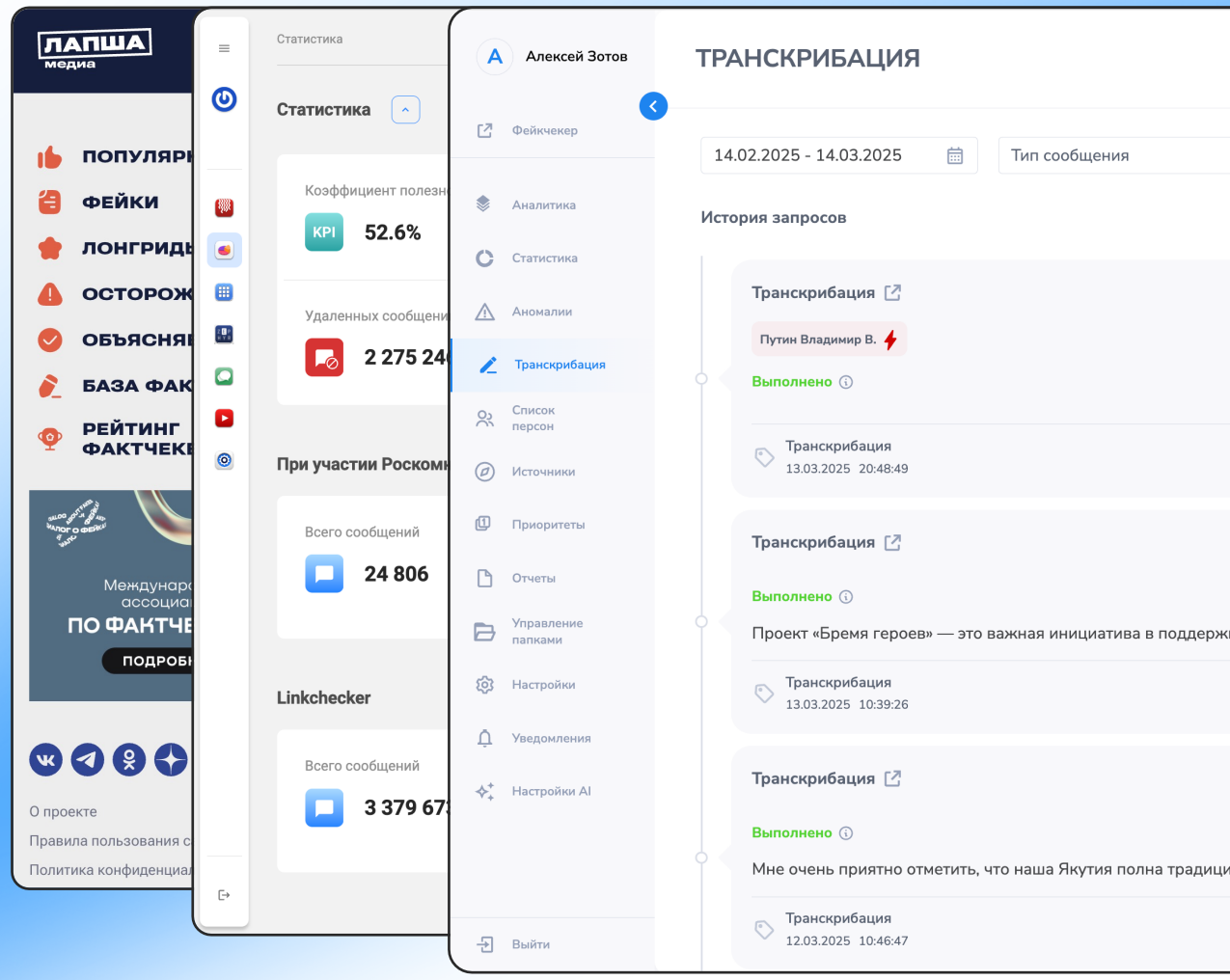
Дашборд для хранения, агрегирования и полного цикла обработки сообщений, содержащих недостоверные данные.

ИС «ЗЕФИР»

система для верификации и аналитики сгенерированного контента.

Проект «Лапша Медиа»

Публичный бренд - фактчекинговые сервисы для информирования граждан о недостоверной информации, в том числе ее опровержения.

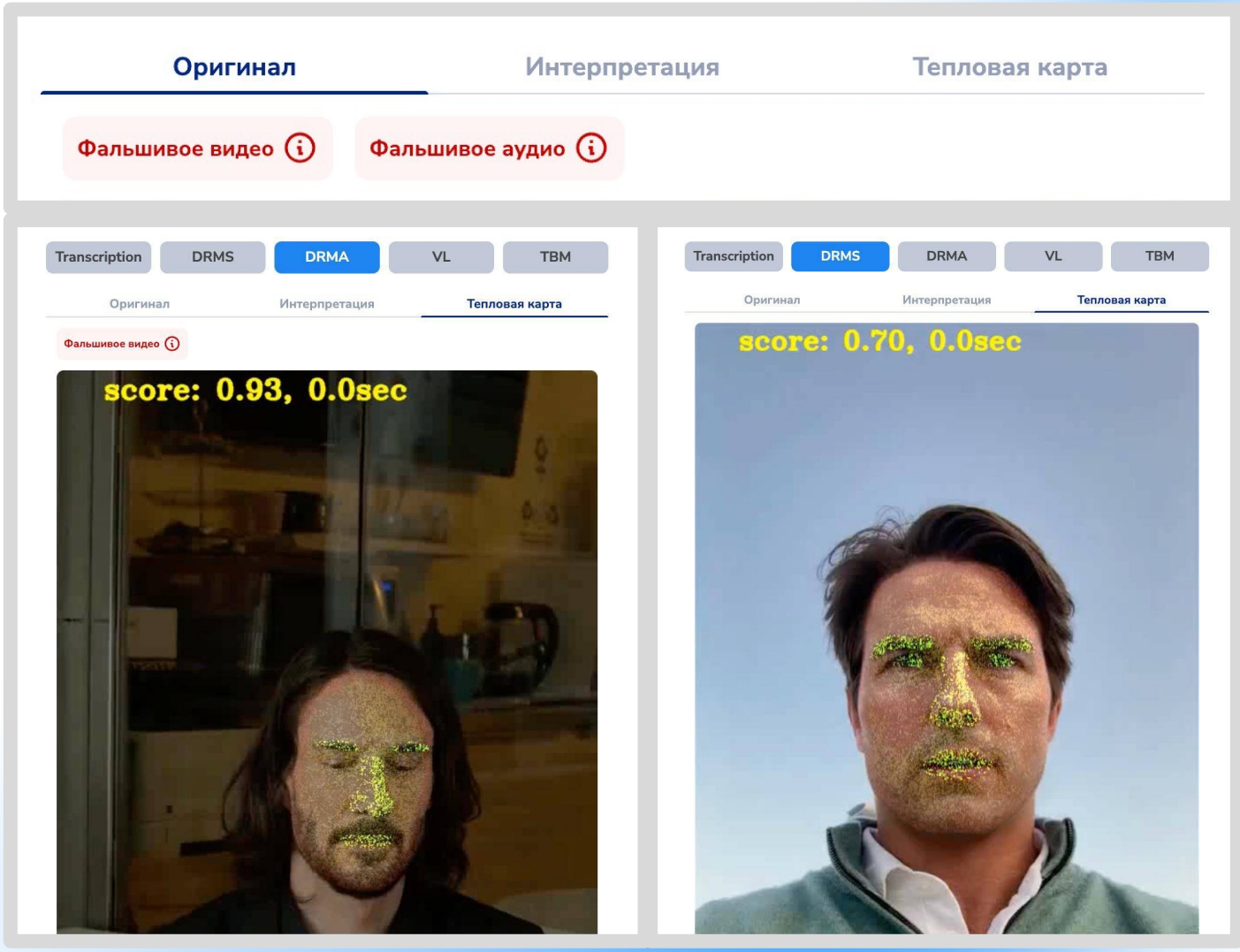


«ЗЕФИР» – СИСТЕМА МОНИТОРИНГА АНО «ДИАЛОГ РЕГИОНЫ»

Анализа (выявление)
и мониторинг сфабрикованного аудиовизуального контента, в том числе дипфейков

Решение на основе сочетания моделей машинного обучения

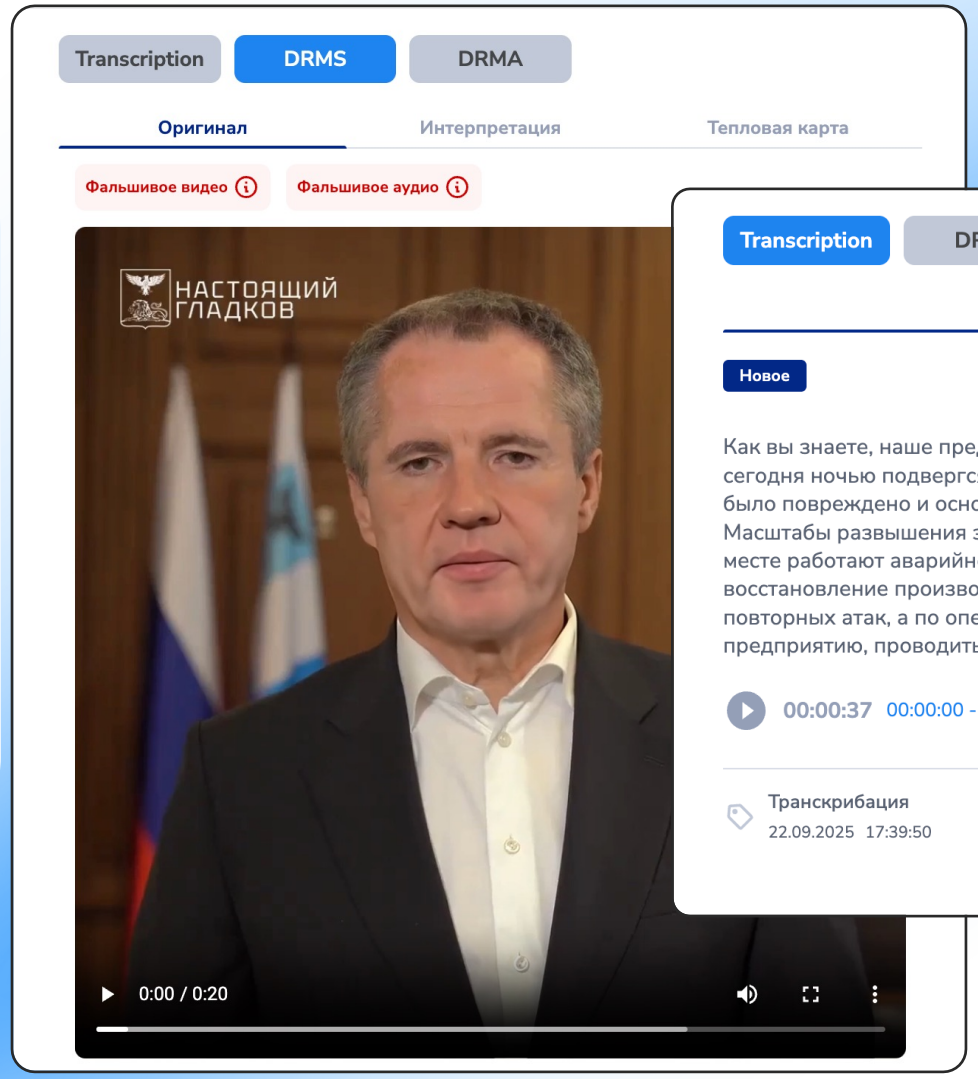
Комплексная оценка материала: транскрибирование и перевод аудио, поиск дубликатов и первоисточников, детекция персон на видео- и аудиозаписях



«ЗЕФИР»

ВЫЯВЛЕНИЕ И АНАЛИЗ ГЕНЕРАТИВНОГО КОНТЕНТА

- **5 моделей** детекции дипфейков
- Более **70 обучений** моделей за 2 года
- **1 модель** детекции генерации голоса
- **1 модель** транскрибации
- **Интерпретация и визуализация** — тепловые карты и оценка применяемых в атаке моделей
- **< 5 минут** — SLA проверки контента



«ЗЕФИР» и «ДАШБОРД ФЕЙКОВ»

ЗАДАЧИ НА 2026 ГОД

- **Развитие систем** по детекции полностью синтетического контента
- **Актуализация** текущих моделей детекции новых моделей генеративного ИИ
- **Интеграция ключевых решений** по детекции в единый сервис верификации недостоверной информации и генеративных фейков
- **Внедрение AI-решений** на отечественные и дружественные платформы для модерации контента
- **Создание экспертно-технологического объединения** в целях развития детекции ИИ (НИИ, корпорации, международные партнеры)

ЭФФЕКТИВНОСТЬ ТЕХНОЛОГИЧЕСКИХ РЕШЕНИЙ ПО ДЕТЕКЦИИ

будет критически снижаться без:

- **Повышения информированности и медиаграмотности пользователей;**
- **Развития критического мышления и навыков фактчекинга;**
- **Без учета прогнозов тенденций медиапотребления и медиапроизводства;**
- **Баланса стимулирования развития ИИ и обеспечения безопасности граждан**

ДИАЛОГ