Учёные СПб ФИЦ РАН обучили нейросеть выявлять дипфейки, используемые злоумышленниками в видеозвонках

03.10.2025

Исследователи Санкт-Петербургского Федерального исследовательского центра Российской академии наук (СПб ФИЦ РАН) создали прототип программного обеспечения, который способен автоматически выявлять дипфейки, созданные с помощью технологии фейс-свап – когда на оригинальную фотографию человека накладывается поддельное лицо. В будущем разработка может быть встроена в системы защиты видеозвонков, которые активно используют кибермошенники.

Одной из серьезных проблем функционирования цифровой среды является сфера мошенничества в интернете, в частности, в социальных сетях и мессенджерах. Она характеризуется низкой раскрываемостью преступлений и, как следствие, постоянным ростом ущерба как для отдельных граждан, так и для государственных и частных учреждений. Например, в России за 2024 год от деятельности киберпреступников жители страны потеряли около 300 млрд рублей.

В последние годы мошенники стали активно использовать сервисы искусственного интеллекта, и одной из популярных технологий является фейс-свап – технология замены лица на фото (или видео), чтобы создать поддельный контент, оригинальность которого пользователю крайне сложно определить «на глаз».

Часто мошенники используют фейс-свап в видеозвонках для маскировки своей настоящей внешности, которая может мгновенно изменяться, ретушироваться и улучшаться в реальном времени. Это позволяет им выдавать себя за других людей: жертвы доверяют друзьям и родственникам больше, чем незнакомцам, что облегчает обман в целях вымогательства денег, кражи личных данных и других манипуляций. Подмена лица затрудняет распознавание мошенничества, так как визуальный контакт кажется действительно достоверным.

«Мы разработали программу, которая, используя нейросеть, может автоматически распознать фейс-свап, созданный с применением технологии дипфейк. В отличие от обычного визуального восприятия, когда человеку очень сложно заметить хорошо сделанную подделку, система анализирует множество скрытых параметров изображения и выявляет характерные искажения», — рассказывает руководитель Международного центра цифровой криминалистики СПб ФИЦ РАН Андрей Чечулин.

На первом этапе проекта учёные создали базу данных с лицами, созданными по технологии фейс-свап. Для создания этих изображений они использовали технологии искусственного интеллекта, которые находятся в свободном доступе – именно ими, как правило, пользуются киберпреступники (реже создают какие-то оригинальные сервисы). Всего исследователи собрали более 30 тысяч картинок.

На основе собранной базы данных проводилось обучение нейросети. Ученые сконцентрировали его на целом комплексе признаков фейс-свапа, среди которых изменение геометрии головы, искажение отдельных элементов лица, а также цветовых оттенков кожи и окружения.

«Наши эксперименты на фотографиях, созданных с помощью фейс-свап, показали высокую точность распознавания. На основе нашей разработки создается библиотека, которую разработчики систем видеосвязи смогут встраивать в мессенджеры. И пользователь будет автоматически получать информацию о вероятности подделки собеседника», — отмечает старший научный сотрудник Лаборатории интегрированных систем автоматизации СПб ФИЦ РАН Алексей Кашевник.

Проект учёных СПб ФИЦ РАН поддержан грантом Фонда содействия инновациям.