ПРОГРАММНЫЙ КОМПЛЕКС ОБНАРУЖЕНИЯ КОНТЕНТА, СГЕНЕРИРОВАННОГО НЕЙРОСЕТЯМИ

Бугаев И.А. (ВКА), Пеньков Д.В. (ВКА)

Научный руководитель – кандидат военных наук Бугаев И.А.
(Военно-космическая академия имени А.Ф.Можайского)

Введение. В современном информационном поле высокая степень соперничества за влияние на общественное сознание часто приводит к применению различных видов контента, сгенерированного нейросетями [1]. Одной из целей такого контента может являться компрометация или дискредитация какой-либо публичной или влиятельной личности, что может повлечь за собой финансовые, репутационные или даже политические последствия. В контексте информационной безопасности потенциальная компрометация видео или аудио материалов несет угрозу целостности информации о человеке. Предлагаемый программный комплекс позволяет решать актуальные проблемы в области обнаружения контента, сгенерированного нейросетями [2].

Основная часть. Программный комплекс для обнаружения контента, сгенерированного нейросетями, состоит из следующих элементов: модель обнаружения контента «audio-deepfake-detection», графический интерфейс для работы пользователя (GUI), система управления базой данных (СУБД). В технологической основе программного комплекса лежит нейронная сеть audio-deepfake-detection, специализирующаяся на обнаружении сгенерированных аудиоматериалов [3]. Данная нейросеть использует современную модель, предназначенную для извлечения признаков из mel-спектрограмм и классификации аудиодипфейков. Она состоит из следующих слоёв: слой свёрточной обработки (предназначен для извлечения характеристик из mel-спектрограммы), слой максимального объединения (выполняет понижающую дискретизацию для уменьшения размеров), блок пакетной нормализации (производит нормализацию для стабилизации обучения модели), выпадающий слой (предотвращает переобучение модели), слой глобального усреднения (объединяет карты признаков для получения общей информации), классификатор (выполняет классификацию с помощью функции сигмоиды). Данный инструмент обладает более высокой точностью обнаружения поддельного контента среди доступных решений. Для работы программе требуется предварительно обученная модель для обнаружения подделок голоса конкретного диктора. Это является отличительной особенностью программного комплекса, так как большинство существующих решений используют для обнаружения поддельного контента общие признаки и артефакты аудиозаписи, не учитывающие индивидуальные особенности говорящего. К преимуществам audio-deepfake-detection также относятся: открытый исходный код, отсутствие лицензии для использования, отсутствие требования подключения к Интернету, кроссплатформенность, повышенная точность обнаружения дипфейков.

Ключевым модулем для работы программного комплекса является реляционная база данных на основе СУБД PostgreSQL для хранения аудиозаписей высказываний конкретной личности [4]. Она состоит из следующего набора сущностей, который позволил разделить и разметить данные по ряду признаков и улучшить качество обучения модели: person_id (SERIAL, PK) —идентификатор личности; audio_files_id (SERIAL, PK) —идентификатор аудиофайла; personality_id (INTEGER, FK → personalities.id) — внешний ключ отношения аудиозаписи к личности; параметры duration (INTEGER) — длительность в секундах, sample_rate (INTEGER) — частота дискретизации, bitrate (INTEGER) — битрейт, created_at (TIMESTAMP) — дата загрузки.

Пользователь программного комплекса взаимодействует с простым графическим интерфейсом, реализованным с помощью библиотеки PyQt5 на языке программирования

Python, что также позволяет реализовать кроссплатформенность [5]. Для начала работы необходимо загрузить обучающую выборку из аудиозаписей голоса человека, под которого производится обучение модели, для чего пользователь вводит путь к папке с данными. Далее запускается процесс обучения, который в зависимости от множества факторов (объема и качества данных, вычислительной мощности оборудования, заданного количества эпох и др.) может длиться от нескольких минут до часов. По окончании обучения модель готова к обнаружению сгенерированного контента.

Для обеспечения масштабируемости и простоты управления программный комплекс был реализован в среде контейнеризации Docker [6]. Такой подход позволил: изолировать программу от операционной системы, обеспечить масштабируемость, сократить время на развёртывание, снизить зависимость от инфраструктуры. Основой для создания Docker-образа является Dockerfile, в котором описаны необходимые зависимости, такие как TensorFlow, audio-deepfake-detection, PyTorch, PyQt5, pip3 и системные пакеты ffmpeg, git. Образ включает все нужные инструменты для обработки аудио и работы с нейросетями.

Выводы. Таким образом, разработанный ПК позволил решить следующие проблемы:

1. Обеспечение изолированности базы данных для обучения и защиты персональных данных от утечки за счет использования инструмента генерации не требующего связи с глобальной сетью.

2. Повышение точности идентификации поддельного контента путем обучения модели на данных голоса конкретного человека.

Кроме того, программный комплекс помогает автоматизировать и ускорить процесс обнаружения контента, сгенерированного нейросетями, а также упростить установку и настройку используемых средств обнаружения. В перспективе развития программного комплекса предлагается использовать для развёртывания оркестратор контейнеров Kubernetes, что дополнительно упростит управление системой и повысит её отказоустойчивость.

Список использованных источников:

1. Статья о проблемах дипфейков [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/dipfeyki-sovremennoe-ponimanie-podhody-k-opredeleniyu-harakteristiki-problemy-i-perspektivy>
2. Методы синтеза речи. – Р. Taylor, Text-to-Speech Synthesis, Cambridge University Press, New York, NY, USA, 1st edition, 2009.
3. Страница проекта audio-deepfake-detection [Электронный ресурс]. – URL: <https://github.com/sksmta/audio-deepfake-detection?tab=readme-ov-file>
4. Проектирование БД Postgress [Электронный ресурс]. – URL: <https://habr.com/ru/companies/otus/articles/451042/>
5. Документация PyQt5 [Электронный ресурс]. – URL: <https://python-scripts.com/pyqt5>
6. Документация Docker [Электронный ресурс]. – URL: <https://docs.docker.com/manuals/>