Российские ученые выложили в открытый доступ инструмент предотвращения автоматического сбора биометрических данных из открытых аудио источников

Группа исследователей из AIRI, МТУСИ, Сбера и VeinCV выложила в открытый доступ метод защиты персональных голосовых данных в публичном цифровом пространстве. Технология основана на универсальных состязательных патчах (universal adversarial patches, UAP) и призвана блокировать несанкционированный автоматический сбор и анализ голосовых биометрических характеристик из аудиозаписей, доступных для скачивания, включая подкасты, интервью, лекции и онлайн-стримы.

Публично доступные аудиозаписи – богатый источник для автоматического сбора голосовых биометрических данных злоумышленниками или недобросовестными компаниями. Скачивание большого объема доступных голосовых записей приводит к риску кражи голосовой биометрии с целью подделки идентификации, несанкционированного профилирования личности, а также создания подделок голоса (дипфейков).

Предложенный математический аппарат позволяет разрабатывать незаметную цифровую "защиту" (UAP) и встраивать ее прямо в аудиофайл перед публикацией. Решение делает голосовые биометрические данные максимально «неидентифицируемыми» для автоматических систем распознавания и анализа голоса (ASI/V), стремящихся извлечь уникальные характеристики диктора. Патч нарушает работу алгоритмов извлечения голосовых «отпечатков» (speaker embeddings), делая попытки автоматической идентификации или кластеризации голосов по публичным записям бесполезными. Благодаря новой экспоненциальной TV-функции потерь с полной вариацией, патч минимально влияет на воспринимаемое качество звука и соотношение сигнал/шум, сохраняя разборчивость для человеческого восприятия и задач, не связанных с идентификацией говорящего, например, не мешает работе систем распознавания содержания речи, сохраняя полезность аудио для транскрипции, перевода или поиска по ключевым словам. Представленное решение работает на аудиофайлах любой длины от коротких цитат до длинных лекций.

Технология важна для публичных лиц и контент-мейкеров, чьи выступления легко можно найти в интернете. Кроме того, модель позволит представителям бизнес-сообщества и образовательных платформ защищать записи корпоративных вебинаров, конференций и обучающих материалов от сбора данных о выступающих. Провайдерам безопасных коммуникаций разработка пригодится для обеспечения дополнительного уровня защиты пользователей.

«Наша цель — усилить методологию защиты цифрового следа человека в публичном пространстве. В первую очередь — его голосовой идентичности. Разработанное решение позволяет свободно делиться аудиоконтентом, значительно снижая вероятность того, что голос станет сырьем для несанкционированного сбора биометрических данных или создания дипфейков» — подчеркнул Олег Рогов, к.ф.-м.н., руководитель лаборатории Безопасного ИИ SAIL AIRI-МТУСИ.

Исследование принято к публикации в материалах международной конференции Interspeech (рейтинг Core-A) в Роттердаме в 2025 году. Полный текст доступен по ссылке: https://arxiv.org/pdf/2505.19951

Вопросы: pr@airi.net

Институт AIRI — *автономная некоммерческая организация, занимающаяся фундаментальными и прикладными исследованиями в области искусственного интеллекта. На сегодняшний день более 200 научных сотрудников AIRI задействовано в исследовательских проектах Института для работы совместно с глобальным сообществом разработчиков, академическими и индустриальными партнерами.*

Лаборатория безопасного ИИ SAIL (Safe Artificial Intelligence Laboratory) AIRI-МТУСИ *функционирует на базе Московского технического университета связи и информатики с марта 2025 года. В задачи Лаборатории входит подготовка профильных специалистов в области кибербезопасности с применением ИИ, а также проведение фундаментальных и прикладных исследований.*