DOI 10.36622/VSTU.2023.26.1.015
УДК 004.056

ЗАЩИТА БИОМЕТРИЧЕСКИХ ДАННЫХ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ОТ СОСТЯЗАТЕЛЬНЫХ АТАК

В.М. Герасимов, М.А. Маслова, Э.И. Халилаева, Н.С. Смирнов

С развитием цифровых технологий, а также повсеместного внедрения искусственного интеллекта в нашу жизнь (к примеру, распознавание голоса или лица пользователя) вошла обязательная защита данных, которая для машинного обучения играет важную роль в безопасности. Для надлежащего и безопасного использования систем искусственного интеллекта при идентификации пользователя необходимо обратить внимание на хранение и передачу данных, на которых будет происходить машинное обучение и выделение признаков для идентификации пользователя.

В статье будет рассмотрен способ безопасного хранения и передачи биометрических отпечатков, при котором может улучшиться качество идентификации пользователя и работоспособность системы искусственного интеллекта. Основная идея заключается в том, что при помощи ключа возможно зашифровать биометрические данные пользователя (фото и аудио). Из обычного голоса и картинки, биометрических данных пользователя, при неверном шифровании получится белый шум.

Для реализации данного метода используется специальное программное обеспечение, которое содержит следующие компоненты: 1) генератор ключей – используемое для создания уникального шифрованного ключа для каждого пользователя; 2) шифровальщик данных – используемое для шифрования биометрических данных пользователя используя генерируемый ключ; 3) анализатор данных – используемое для анализа зашифрованных данных и проверки их на соответствие с идентификатором пользователя. Использование данного метода безопасного хранения и передачи биометрических данных может помочь в улучшении безопасности системы искусственного интеллекта, а также обеспечить более высокую точность идентификации пользователя.

Ключевые слова: безопасность системы ИИ, защита голосовых отпечатков, защита биометрических данных, стеганография, шифрование данных.

.

Введение

В современном мире главная задача информационной безопасности — это сохранение конфиденциальности, целостности и доступности. С приходом в жизнь искусственного интеллекта, соответственно, появилась потребность защищать данные ИИ. Любая система имеет свои «ошибки» и уязвимости, на которые злоумышленники могут повлиять на работу системы не в лучшую сторону. Стоит отметить тот факт, что иногда именно атака на ИИ может привести к катастрофическим последствиям, если не уделять особое внимание её безопасности [1]. Например, медицинские исследования, при которых ИИ может ставить «ложный» диагноз, который может повлиять на жизнь человека [2].

Возможность атаковать систему искусственного интеллекта, а также влиять на неё посредством разных видов атак (см. рис. 1), приводит к тому, что система ИИ не имеет своего первоначального смысла — автоматизация и безопасность жизнедеятельности человека.

Так, например, голосовые отпечатки в современном мире защищаются дорогостоящим, но эффективным методом. С помощью разграничения доступа, на защищённом сервере. Как показывает практика, инсайдеры — главная проблема всех предприятий, а также возможные утечки и атаки («взломы») серверов [3]. На данном промежутке времени, в котором происходит формирование и модернизация вычислительной мощности — возможно и правильно использовать именно подобный подход, но необходимо «идти в ногу со временем».

Основная часть

Для решения следующих задач: обеспечение защиты голосовых данных; обеспечение защиты биометрических данных. Было предложено следующее решение, которое позволяет не только шифровать голосовой отпечаток, но и скрывать его параметры и характеристики. Это преобразование из одного формата данных в другом с дальнейшим шифрованием исходного аудиофайла [4].

Главный принцип данного исследования заключается в том, что при передаче биометрических данных на сервер — они уже зашифрованы. Из-за того, что у нас существует возможность расшифровки при помощи ключа, то при обратной операции не имея ключ — получим белый шум, который является нашим голосом (рис. 1-3).

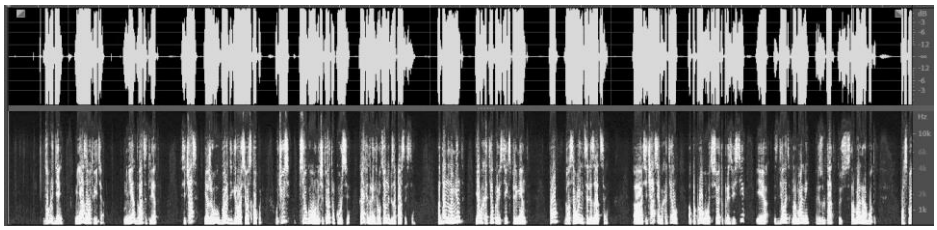

Рис. 1. Пример чистого голосового аудиофайла (со спектрограммой)

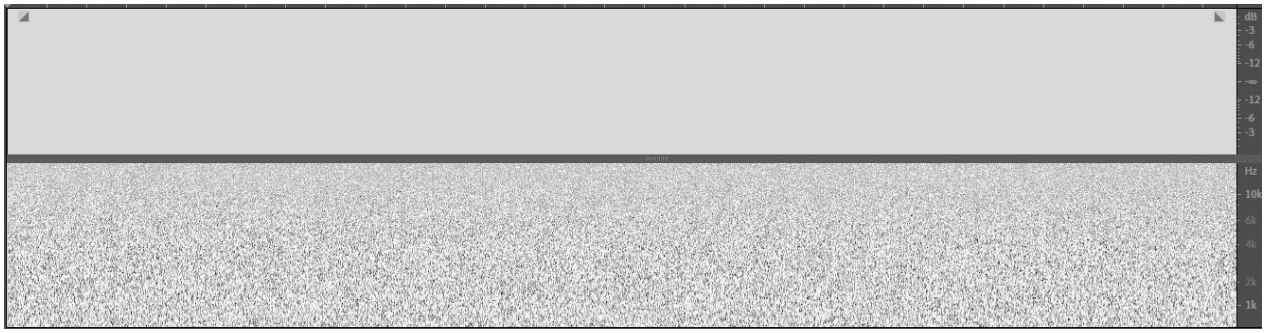

Рис. 2. Пример зашифрованного голосового аудиофайла (со спектрограммой)

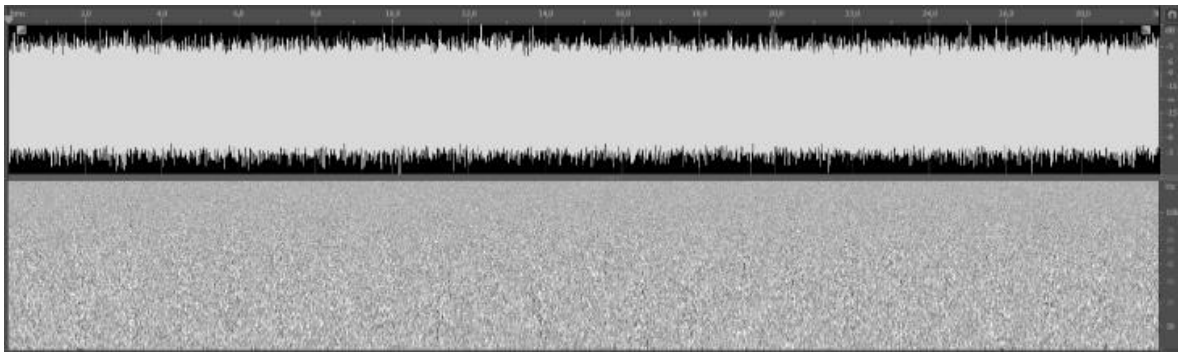

Рис. 3. Пример аудиозаписи белого шума (со спектрограммой)

Состязательные атаки на искусственный интеллект могут привести к функционированию AI в нежелательном или опасном режиме. Такие атаки могут быть вызваны для поражения системы или для получения доступа к важным данным, программам или сервисам. В связи с этим, атаки могут привести к непредсказуемым и нежелательным последствиям для информационной системы.

Например, атака может привести к неверным принятиям решений или к сбоям в работе системы.

Защита биометрических данных от атак данного типа позволяет избежать ряд проблем. Если злоумышленник попытается внедрить или изменить файл, при передаче зашифрованных данных, то при получении системой изменённого файла — произойдет ошибка. Система ИИ не сможет идентифицировать полученный файл как «внутрисистемный». Это значит, что у файла есть свой уникальный формат, при изменении которого файл повреждается и не считывается системой.

Для понимания защиты данных от состязательных атак необходимо определиться какие модели атак существуют и используются в реальной жизни (рис. 4).

При использовании систем ИИ, которые используют распознавание личности по голосу, данная технология будет применяться эффективно. Принцип работы алгоритма [5] следующий (рис. 5):

1) происходит запись голосового отпечатка и кодируется в доступный («расширенный») формат аудиоданных (.wav);

2) конвертация при помощи алгоритма в отличный от аудиофайла формата (.png) с дальнейшим преобразованием в бинарный (обезличенный) вид — данные действия направлены на ликвидацию возможного поиска файла по «тегам» (характеристикам);

3) при помощи ключа происходит симметричное шифрование уже преобразованного аудиофайла, который «замыкает» цикл преобразования файла на всех этапах.

Итого, получаем обезличенный файл, который не только передаётся безопасно, но и храниться на защищённом сервере (рис. 5).

Из-за того, что существует угроза атаки на систему распознавания пользователей, а также возможные способы «имитации», подделки голосовых данных. Данный подход к решению поставленных задач имеет огромное значение с точки зрения усложнения перехвата (нахождения) подобной информации [6, 7]. Соответственно, повышается безопасность системы искусственного интеллекта при распознавании пользователя.

Применение

При помощи данной технологии вероятность получения данных голосовых отпечатков стремиться к нулю. Возможно создание систем, которые будут устойчивы к атакам на ИИ. Злоумышленник при возможности не сможет получить какую-либо информацию, потому что устанавливаются дополнительные рубежи защиты.

Одним из преимуществ SGEC-алгоритма является то, что при изменении зашифрованных данных, алгоритм не сможет его дешифровать, тем самым защищая ИИ от получения нежелательных данных.

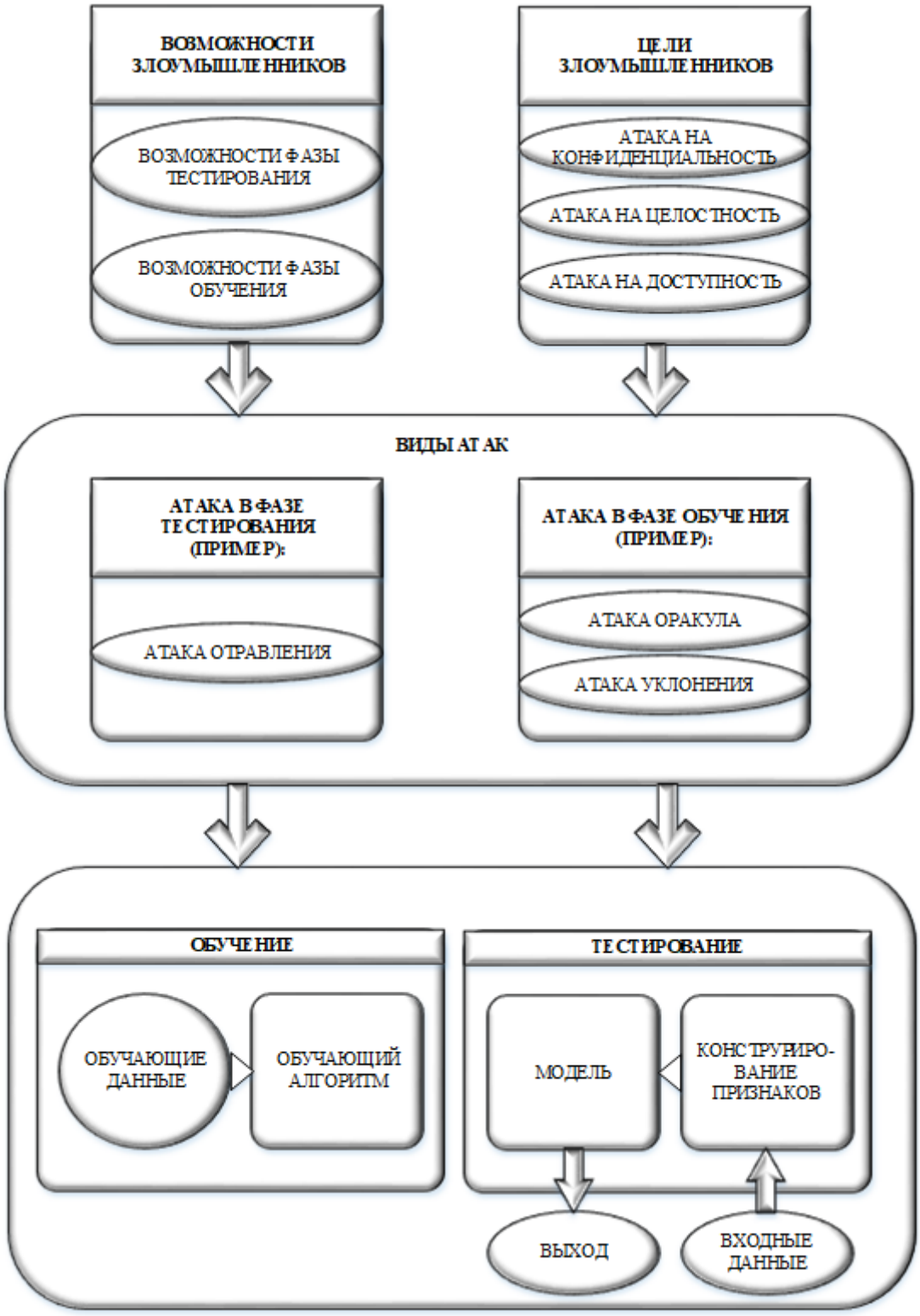

Рис. 4. Пример схемы возможных состязательных атак на ИИ модель

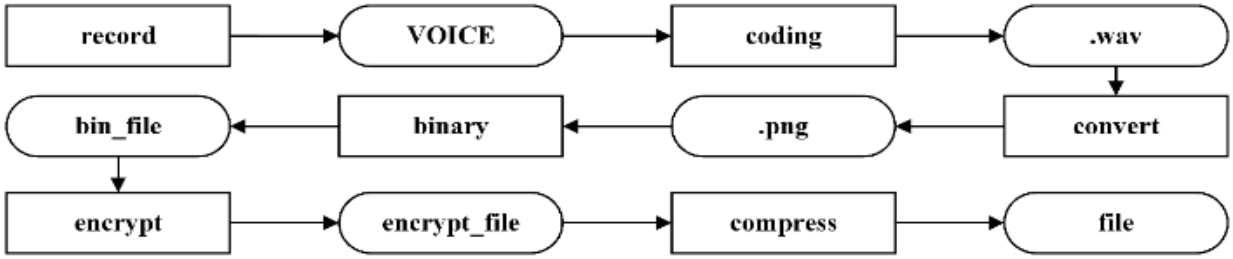

Рис. 5. Пошаговые действия обеспечения безопасности голосового отпечатка

Так, например, согласно исследованию, Cisco [8] под влиянием атак на системы искусственного интеллекта (рис. 6).

– доля руководителей служб безопасности, которые уверены в успехе систем машинного обучения значительно сократилась;

– число заинтересованных в решениях на основе искусственного интеллекта также уменьшилось.

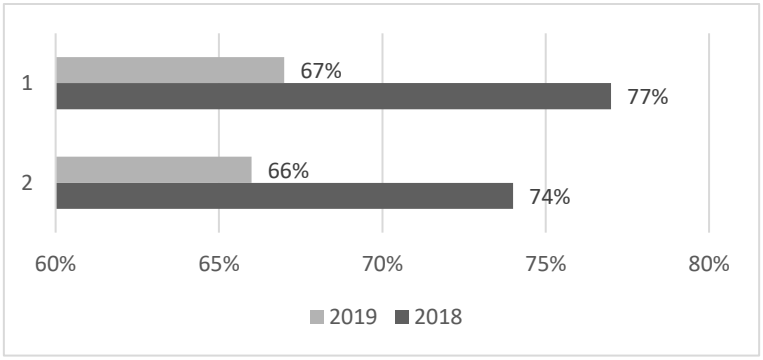

Рис. 6. Влияние атак ИИ на общий интерес общества

При детальном разборе всех возможных атак — можно определить значимость данного решения. По данным Kaspersky существует множество известных атак, а также способы защиты от них (рис. 7). На каждом из этапов необходимо определить возможные угрозы, т.к. это связано с целостностью системы и дальнейшим принятием решения. Например, из-за того, что злоумышленник может изменить обучающую выборку, в дальнейшем это повлияет на конечный результат.

Таким образом, системы, которые как-то взаимодействуют с биометрическими данными — например, распознавание голоса и лица пользователя. Метод SGEC [9] позволяет защищать голосовые данные пользователей на этапе передачи и хранения информации.

При детальном разборе всех возможных атак — можно определить значимость данного решения. По данным Kaspersky существует множество известных атак, а также способы защиты от них (рис. 7).

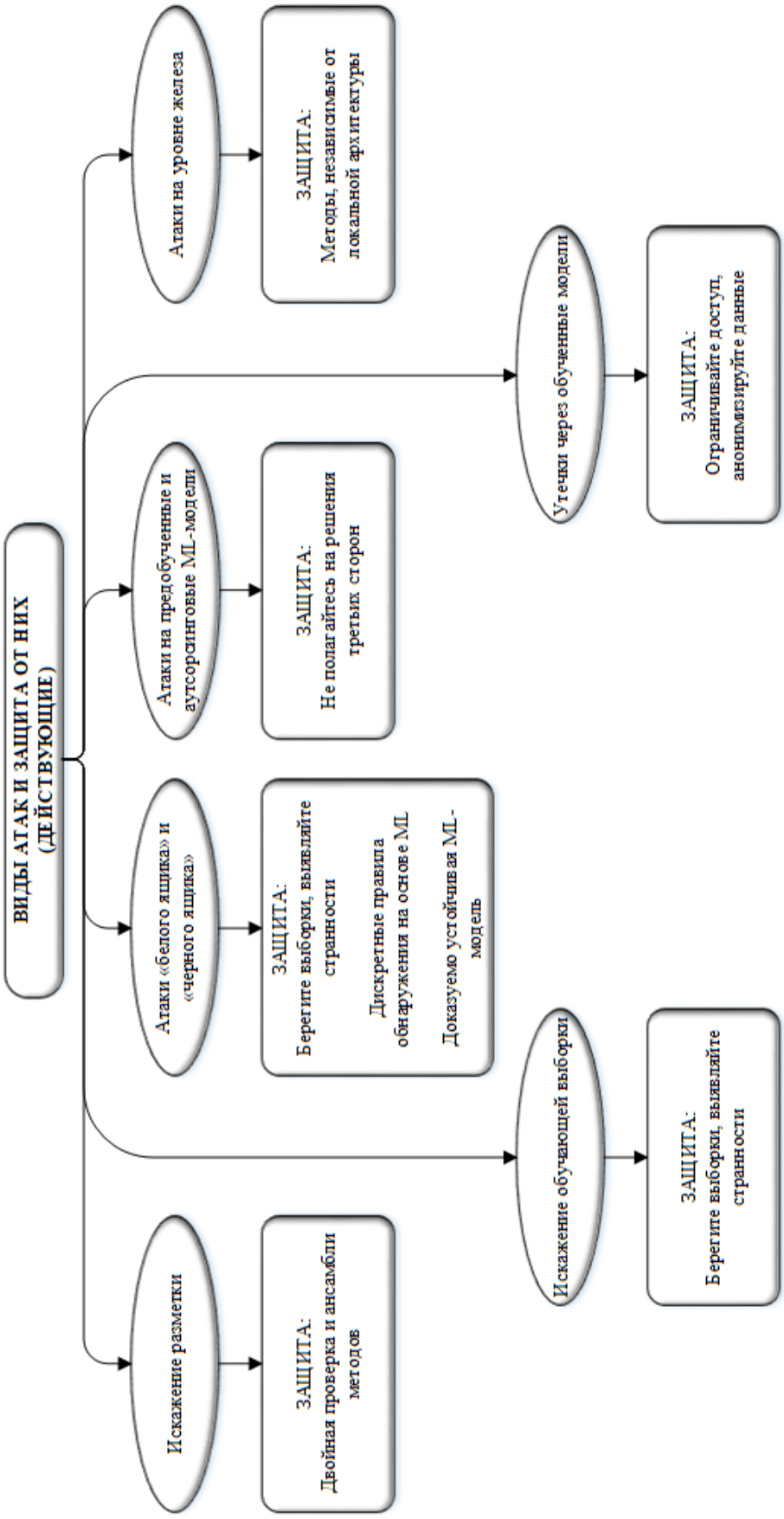

Рис. 7. Виды атак на систему ИИ и защита от них

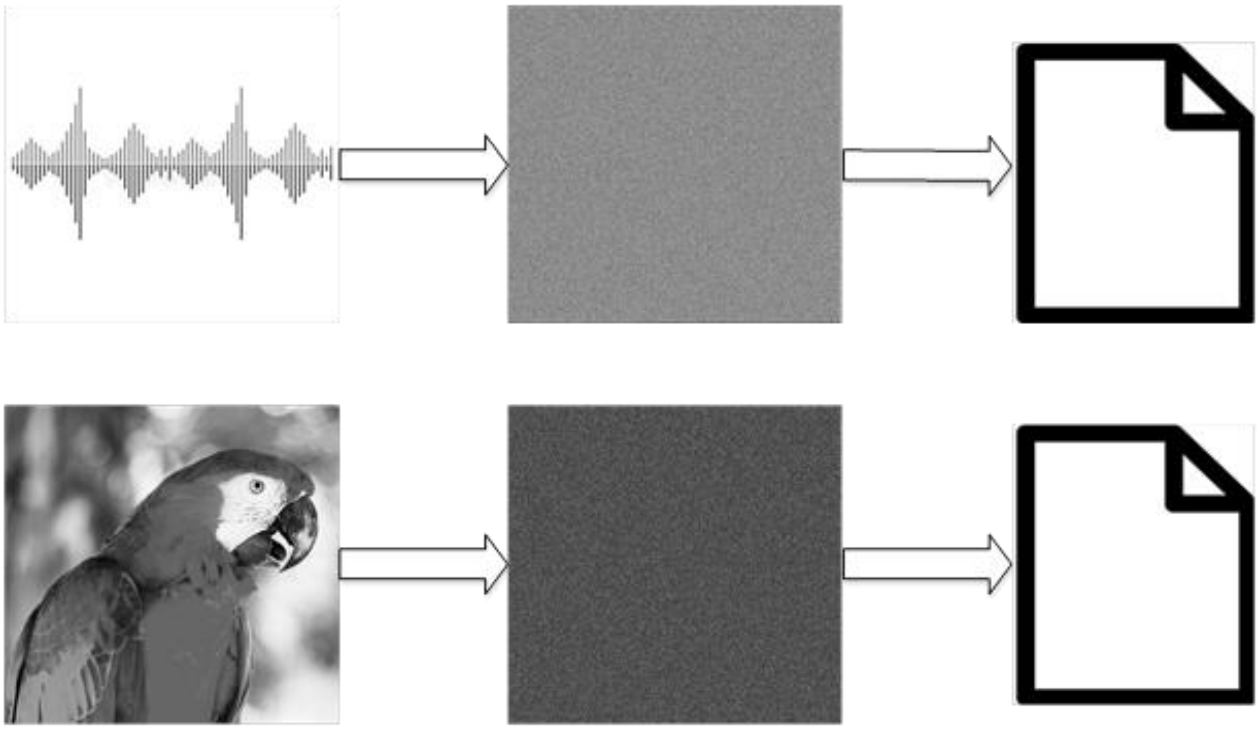

Рис. 8. Работа алгоритма SGEC для защиты биометрических отпечатков системы ИИ

На каждом из этапов необходимо определить возможные угрозы, т.к. это связано с целостностью системы и дальнейшим принятием решения. Например, из-за того, что злоумышленник может изменить обучающую выборку, в дальнейшем это повлияет на конечный результат.

Таким образом, системы, которые как-то взаимодействуют с биометрическими данными — например, распознавание голоса и лица пользователя. Метод SGEC [9] позволяет защищать голосовые данные пользователей на этапе передачи и хранения информации.

Так, например, работа алгоритма выглядит следующим образом (см. рис. 8), который направлен на обеспечение безопасной передачи биометрических данных пользователей, а также для дальнейшего хранения в БД системы ИИ.

Подобный подход позволит защитить как голос, так и изображение от изменений.

Данный алгоритм позволяет предотвратить изменения данных и предотвратить несанкционированный доступ к информации SGEC позволяет проверить и подтвердить подлинность данных, а также защитить их от несанкционированного изменения.

Вывод

Данный метод защиты алгоритма, при котором происходит обезличивание, шифрование при помощи ключа и сокрытия данных в контейнере других форматов позволяет обеспечить безопасность систем ИИ, использующих биометрические данные пользователей — голос и лицо.

При существующих проблемах необходимо тестировать и изобретать новые методы защиты систем ИИ, которые позволят ограничить злоумышленникам доступ к тем или иным данным. Данные методы могут занимать дополнительные ресурсы (временные затраты, а также

вычислительная мощность), но предотвращает множественные попытки атак, которые могут повлечь за собой катастрофические последствия.

Для защиты данных от состязательных атак следует применять несколько мер безопасности, включая применение сильных и уникальных паролей, ограничение доступа к данным только авторизованным пользователям, шифрование данных и проверка логов. Также следует применять правильную архитектуру системы и устанавливать системы обнаружения вторжений для отслеживания попыток атак.

Список литературы

1. Gavrilenko T. V., Gavrilenko A. V. Ошибки машинного обучения искусственных нейронных сетей и использование их для атак на системы искусственного интеллекта // Успехи кибернетики. 2021. Т. 2. №. 3. С. 23-32.

2. Богомолов А. И. Искусственный интеллект и экспертные системы в мобильной медицине / А.И. Богомолов, В.П. Невежин, Г.А. Жданов // Хроноэкономика. 2018. №. 3 (11). С. 17-28.

3. Бородакий Ю. В. и др. Инсайдерология наука о нелегитимности в компьютерной инфосфере // Известия Южного федерального университета. Технические науки. 2008. Т. 85. №. 8. С. 55-64.

4. Костиков, В. А. Преобразование биометрических данных аудиофайла в RGB-изображение для защиты информации на сервере / В. А. Костиков, М. А. Маслова // Проблемы проектирования, применения и безопасности информационных систем в условиях цифровой экономики : Материалы XXI Международной научно-практической конференции, Ростов-на-Дону, 29–30 ноября 2021 года. Ростов-на-Дону: Ростовский государственный экономический университет "РИНХ", 2021. С. 53-58.

5. Герасимов, В. М. Комплексная система защиты биометрического голосового отпечатка от воздействия кибермошенников / В. М. Герасимов // XI Конгресс молодых учёных : сборник научных трудов, Санкт-Петербург, 04–08 апреля 2022 года. – С-Пб: федеральное государственное автономное образовательное учреждение высшего образования "Национальный исследовательский университет ИТМО", 2022. С. 72-76..

6. Герасимов, В. М. Необходимость комплексной системы защиты биометрического голосового отпечатка от воздействия кибермошенников в сети интернет / В. М. Герасимов, М. А. Маслова // Вестник Луганского государственного университета имени Владимира Даля. 2022. № 5(59). – С. 95-102.

7. Герасимов, В. М. Возможные угрозы и атаки на систему голосовой идентификации пользователя / В. М. Герасимов, М. А. Маслова // Научный результат. Информационные технологии. 2022. Т. 7, № 1. С. 32-37.

8. TK Keanini The State of Machine Learning in 2019 URL: https://blogs.cisco.com/security/the-state-of-machine-learning-in-2019 (дата обращения: 18.10.2022).

9. Свидетельство о государственной регистрации программы для ЭВМ № 2022663168 Российская Федерация. SGEC-система "BIOM" для шифрования и сокрытия голосовых данных пользователей на сервере : № 2022662279 : заявл. 27.06.2022 : опубл. 12.07.2022 / В. М. Герасимов, М. А. Маслова ; заявитель Федеральное государственное автономное образовательное учреждение высшего образования «Севастопольский государственный университет».

Национальный исследовательский университет ИТМО
National Research University ITMO

Севастопольский государственный университет
Sevastopol State University

Ростовский государственный экономический университет (РИНХ)
Rostov State University of Economics (RINH)

Поступила в редакцию 20.02.23

Информация об авторах

Герасимов Виктор Михайлович – инженер, студент первого курса магистратуры направления «Безопасность систем искусственного интеллекта» факультета безопасности информационных технологий, Национальный исследовательский университет ИТМО, Санкт-Петербург, e-mail: my.virus.kaspersky@gmail.com.
Маслова Мария Александровна – старший преподаватель кафедры "Информационная безопасность", Севастопольский государственный университет; младший научный сотрудник, Ростовский государственный экономический университет (РИНХ), e-mail: mashechka-81@mail.ru
Халилаева Эмине Илимдаровна – студент первого курса магистратуры кафедры «Информационная безопасность» Института информационных технологий, Севастопольский государственный университет, e-mail: emine.halilaeva@yandex.ru
Смирнов Никита Сергеевич – студент первого курса магистратуры кафедры «Информационная безопасность» Института информационных технологий, Севастопольский государственный университет, e-mail: n@newair.ru

PROTECTING THE BIOMETRIC DATA OF ARTIFICIAL INTELLIGENCE SYSTEMS FROM ADVERSARIAL ATTACKS

V.M. Gerasimov, M.A. Maslova, E.I. Khalilayeva, N.S. Smirnov

With the development of digital technologies, as well as the widespread introduction of artificial intelligence into our lives (for example, voice or face recognition of the user), mandatory data protection has entered, which plays an important role in machine learning in security. For the proper and safe use of artificial intelligence systems in user identification, it is necessary to pay attention to the storage and transmission of data on which machine learning and feature extraction will take place for user identification.

The article will consider a method for the secure storage and transmission of biometric fingerprints, which can improve the quality of user identification and the performance of the artificial intelligence system. The main idea is that using the key it is possible to encrypt the user's biometric data (photo and audio). From the usual voice and picture, the user's biometric data, if the encryption is incorrect, white noise will be obtained.

To implement this method, special software is used, which contains the following components: 1) key generator - used to create a unique encrypted key for each user; 2) data encryptor - used to encrypt the user's biometric data using a generated key; 3) data analyzer - used to analyze encrypted data and check it against the user ID. The use of this method of secure storage and transmission of biometric data can help improve the security of the artificial intelligence system, as well as provide a higher accuracy of user identification.

Keywords: AI system security, voice fingerprints protection, biometric data protection, steganography, data encryption.

Submitted 20.02.23

Information about the author

Viktor M. Gerasimov – engineer, first-year student of the master's program in the field of "Security of artificial intelligence systems" of the Faculty of Security of Information Technologies (SIT), National Research University ITMO, St. Petersburg, Russia, e-mail: my.virus.kaspersky@gmail.com.

Maria A. Maslova – Senior Lecturer of the Department of Information Security, Sevastopol State University, junior researcher Rostov State University of Economics (RINH), e-mail: mashechka-81@mail.ru.

Emine I. Khalilaeva – first-year master's student of the department "Information Security" of the Institute of Information Technologies, Sevastopol State University, e-mail: emine.halilaeva@yandex.ru

Nikita S. Smirnov – a first-year master's student of the department "Information Security" of the Institute of Information Technologies, Sevastopol State University, e-mail: n@newair.ru/