Фам Ван Куонг, студент,
Краснодарское высшее военное училище
Pham Van Cuong, student,
Krasnodar Higher Military School

Научные руководители:
Шейн Сергей Леонидович, Преподаватель,
Краснодарское высшее военное училище
Shein Sergey Leonidovich, Lecturer,
Krasnodar Higher Military School

Кубенко Егор Георгиевич, Старший преподаватель,
Краснодарское высшее военное училище
Kubenko Yegor Georgievich, Senior Lecturer,
Krasnodar Higher Military School

ОБМАН ЗРЕНИЯ И СЛУХА: КАК DEEPFAKES СТАНОВЯТСЯ ОРУЖИЕМ В КИБЕРШПИОНАЖЕ И ДЕЗИНФОРМАЦИИ
OCULAR AND AUDITORY DECEPTION: HOW DEEPFAKES BECOME WEAPONS IN CYBERESPIONAGE AND DISINFORMATION

Аннотация. Статья показывает, что deepfake стал практическим инструментом кибершпионажа и дезинформации, поскольку позволяет убедительно подделывать лицо и голос. В статье рассматриваются ключевые риски: целевой обман сотрудников, компрометация официальных коммуникаций, шантаж, влияние на выборы и подрыв доверия к видеодоказательствам.

Abstract. The article argues that deepfakes have become a practical instrument of cyber espionage and disinformation, as they enable highly convincing falsification of both facial identity and voice. It examines key risks, including targeted deception of employees, the compromise of official communications, blackmail, electoral interference, and the erosion of trust in video evidence.

Ключевые слова: Дипфейк, кибершпионаж, дезинформация, социальная инженерия, цифровая криминалистика, эрозия доверия.

Keywords: Deepfakes, cyber-espionage, disinformation, social engineering, digital forensics, trust erosion.

Эпоха цифровых технологий подарила человечеству невиданные ранее возможности для коммуникации, творчества и доступа к информации. Однако та же самая технологическая мощь породила и принципиально новые угрозы, стирающие грань между реальностью и вымыслом. Речь идет о глубоких подделках, или Deepfakes – синтетических медиа, созданных с помощью искусственного интеллекта (ИИ) и машинного обучения, которые с пугающей точностью подделывают внешность, голос и действия реальных людей [2]. Если изначально эта технология вызывала интерес как курьез или инструмент для развлечения, то сегодня она быстро превращается в мощное оружие в арсенале кибершпионов и пропагандистов, способное подрывать национальную безопасность, манипулировать общественным мнением и разрушать доверие к фундаментальным институтам общества.

Технологическая основа: Как создаются Deepfakes.

Чтобы понять масштаб угрозы, необходимо разобраться в том, что стоит за этим явлением. Deepfakes – это не просто продвинутый фотошоп. В их основе лежат генеративно-состязательные сети (GANs – Generative Adversarial Networks).

Принцип работы GANs:

Генератор: Одна нейронная сеть (генератор) создает поддельное изображение или видео из случайного шума. Изначально результат далек от идеала.

Дискриминатор: Вторая нейронная сеть (дискриминатор) получает на вход как реальные изображения, так и подделки от генератора. Ее задача – определить, какое из них фальшивое.

Процесс обучения: Генератор и дискриминатор непрерывно состязаются. Генератор учится обманывать дискриминатор, создавая все более правдоподобные подделки. Дискриминатор, в свою очередь, становится все лучше в их распознавании. Этот цикл продолжается до тех пор, пока генератор не начинает производить контент, который дискриминатор не может отличить от настоящего [1, 2].

Для создания видео с заменой лица используются автоэнкодеры. Система обучается на большом массиве данных (тысячах изображений и часов видео) двух людей - «донора» и «цели». Она изучает их мимику, особенности лица, освещение и ракурсы, а затем накладывает мимику и речь «донора» на лицо «цели». Аналогичные технологии существуют и для клонирования голоса, где ИИ анализирует уникальные тембровые характеристики человека и может синтезировать любую речь, которую этот человек якобы произносит.

Доступность этих технологий стремительно растет. Появляются коммерческие сервисы и открытые исходные коды, позволяющие создавать убедительные Deepfakes даже неподготовленным пользователям, что многократно увеличивает потенциальный круг злоумышленников.

Deepfakes в сфере кибершпионажа: Новая парадигма шпионажа.

Кибершпионаж всегда был битвой за информацию, но Deepfakes добавляют в него новый, психологический, уровень.

1. Целевые фишинговые атаки (Spear Phishing 2.0): Традиционные фишинговые письма могут быть обнаружены благодаря бдительности сотрудников. Но представьте себе видеозвонок в Zoom или получение голосового сообщения, где ваш начальник или высокопоставленный чиновник отдает вам четкий и срочный приказ: «Переведите эти средства на счет X», «Отправьте конфиденциальные документы по этому адресу», «Откройте вложение – это критически важно». Реалистичность такого Deepfake делает традиционные меры безопасности бесполезными. У сотрудника не возникает сомнений в подлинности указания, что приводит к масштабным финансовым и информационным потерям. Известны реальные случаи, когда с помощью клонирования голоса мошенники похищали крупные суммы денег, выдавая себя за руководителей компаний [5].

2. Компрометация дипломатических и военных каналов: Deepfakes могут быть использованы для создания фальшивых заявлений политических или военных лидеров. Вброс такого видео в момент напряженности может спровоцировать международный кризис, сорвать переговоры или даже стать «казусом белли» – формальным поводом к войне. Противник может создать ролик, где министр обороны другой страны якобы отдает приказ о нападении, что вызовет незамедлительную и, возможно, ошибочную ответную реакцию.

3. Шантаж и вербовка агентуры: Кибершпионы могут создавать компрометирующие материалы с участием целевых лиц – политиков, военных, сотрудников спецслужб или корпораций. Даже если жертва знает, что видео фальшивое, угроза его публикации и последующего репутационный ущерб может быть использована для шантажа и принуждения к сотрудничеству. Это мощный инструмент для вербовки агентов влияния.

Deepfakes как оружие дезинформации: Подрыв основ общества.

Если в кибершпионаже цели точечные, то в дезинформации – массовые. Здесь Deepfakes используются для манипуляции общественным сознанием в беспрецедентных масштабах.

1. Влияние на избирательные процессы: Выборы – ключевой момент уязвимости для демократических обществ. Deepfake-видео, опубликованное в решающий момент предвыборной гонки, может оказаться «убийственным аргументом». Оно может изображать кандидата, признающегося в коррупции, произносящего расистские или оскорбительные высказывания. Даже если факт фальсификации будет быстро установлен, ядовитое семя сомнения будет посеяно. Последующее опровержение, как правило, имеет меньший охват и резонанс, чем сам фейк (эффект «ложной истины»). Это подрывает легитимность всего избирательного процесса и победы конкретного кандидата.

2. Разжигание социальной розни: Deepfakes – идеальный инструмент для стравливания социальных, этнических или религиозных групп. Можно создать видео, где представитель одной группы оскорбляет или призывает к насилию против другой. В условиях напряженности такой контент, распространяемый через социальные сети и мессенджеры, может спровоцировать реальные столкновения и беспорядки, дестабилизируя обстановку внутри страны.

3. Эрозия доверия к объективной реальности (эффект «ликвидации правды»): Это, пожалуй, самая опасная и долгосрочная угроза. Массовое распространение как высококачественных, так и примитивных Deepfakes ведет к тому, что люди перестают доверять любым видеодоказательствам. Возникает среда тотального скептицизма, где любое неудобное или разоблачительное реальное видео можно списать на «еще один фейк». Это парализует общественную дискуссию, делает невозможным привлечение к ответственности по видеодоказательствам и разрушает саму идею общей, объективной реальности. Политики и преступники могут использовать эту тактику, отрицая подлинность реальных компрометирующих записей, заявляя: «Это Deepfake!» [4].

Выявление и противодействие: Технологическая гонка вооружений.

Борьба с Deepfakes превратилась в технологическую гонку между создателями подделок и разработчиками систем детекции.

Методы обнаружения Deepfakes:

Анализ артефактов: ИИ ищет микроскопические несоответствия, которые не видит человеческий глаз: неестественное моргание, размытые границы между лицом и фоном, артефакты в области зубов, несовершенная симуляция физики волос.

Анализ физиологических сигналов: Алгоритмы проверяют соответствие видео физиологическим процессам, например, ритму сердцебиения, который проявляется в микродвижениях кожи на лице (фотоплетизмография).

Анализ эмоционального соответствия: Система проверяет, соответствуют ли мимика и жесты произносимому тексту и эмоциональному контексту.

Криминалистический анализ цифровых отпечатков: Каждая камера и программа для редактирования оставляет уникальные «отпечатки» в метаданных файла. Детекторы ищут несоответствия в этих данных [3].

Нетехнологические меры противодействия:

Цифровая грамотность и медиаобразование: Обучение населения критическому восприятию информации, проверке источников и базовым признакам подделки.

Правовое регулирование: разработка законов, криминализирующих создание и злонамеренное распространение Deepfakes с целью причинения вреда. В разных странах этот процесс идет с разной скоростью.

Создание «цифровых водяных знаков» и сертификация контента: разработка стандартов, при которых аутентичное видео с момента создания получает криптографическую подпись, подтверждающую его подлинность.

Межведомственное и международное сотрудничество: обмен информацией между правоохранительными органами, IT-компаниями и государствами для отслеживания кампаний по дезинформации.

Однако проблема в том, что технологии детекции всегда отстают от технологий создания. Как только находится способ выявления одной уязвимости, создатели Deepfakes находят способ ее обойти.

Заключение: Гонка без финишной черты

Феномен Deepfakes знаменует собой наступление новой эры, где доверие к аудиовизуальной информации, столетие бывшей незыблемым доказательством, оказалось под вопросом. Эта технология стала стратегическим оружием, позволяющим влиять на фондовые рынки, избирательные процессы и международные отношения с минимальными затратами и рисками.

Борьба с этой угрозой не может быть исключительно технологической. Это комплексный вызов, требующий консолидации усилий специалистов в области технологий и информационной безопасности, законодателей, журналистов, социологов и общества в целом. Необходимо укреплять правовые рамки, инвестировать в образование и развивать критическое мышление. В противном случае мы рискуем оказаться в мире, где «увидеть - уже не значит поверить», а где любая истина может быть оспорена, а любая ложь - представлена как истина. Будущее информационной безопасности и стабильности общества зависит от нашей способности адаптироваться к этой новой, тревожной реальности, где обман зрения и слуха становится оружием массового поражения доверия.

Список литературы:

1. Goodfellow et al., “Generative Adversarial Networks,”. URL: https://arxiv.org/abs/1406.2661
2. Verdoliva, “Media Forensics and DeepFakes: an overview,”. URL: https://arxiv.org/abs/2001.06564
3. Rössler et al., “FaceForensics++: Learning to Detect Manipulated Facial Images,”. URL: https://arxiv.org/abs/1901.08971
4. Brennan Center, “Deepfakes, Elections, and Shrinking the Liar’s Dividend,” 23 Jan 2024. URL: https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend
5. C2PA, “Content Credentials: C2PA Technical Specification.” URL: https://c2pa.org/specifications/specifications/2.3/specs/C2PA_Specification.html.