

**ТЕХНОЛОГИИ DEEPFAKE: АЛГОРИТМИЧЕСКИЕ ОСНОВЫ,
СОЦИОТЕХНИЧЕСКИЕ РИСКИ И МЕТОДЫ ВЕРИФИКАЦИИ В ЦИФРОВОЙ СРЕДЕ****У.Д. Бык, X класс****Научный руководитель – Н.Н. Пасовец, учитель математики****Государственное учреждение образования «Средняя школа № 10 г. Пинска»**

В современных условиях глобальной цифровизации и стремительного развития систем искусственного интеллекта человечество столкнулось с беспрецедентным вызовом в области информационной безопасности. Одним из наиболее значимых и потенциально опасных явлений последних лет стала технология синтеза медиаконтента, получившая название «дипфейк». Данный термин охватывает широкий спектр методов машинного обучения, направленных на создание высокореалистичных подделок изображений, аудиозаписей и видеороликов, на которых действия или слова объекта не соответствуют действительности. Актуальность глубокого изучения данной проблемы продиктована тем, что дипфейки постепенно эволюционируют из инструмента индустрии развлечений в мощное оружие дезинформации, киберпреступности и политических манипуляций. Мы попытались разобраться с технической архитектурой дипфейков, установить их деструктивный потенциал и как им противостоять.

Технологический фундамент дипфейков опирается на архитектуру генеративно-сопоставительных нейросетей, которые в процессе обучения имитируют сложные биологические системы обработки информации. Суть процесса заключается в непрерывном взаимодействии двух независимых алгоритмов, один из которых выполняет роль создателя контента, а другой – критического контролера. В ходе миллионов итераций система обучается извлекать мельчайшие особенности человеческой мимики, такие как движение глазных яблок, микронапряжение лицевых мышц при произнесении определенных звуков и даже едва заметные изменения тона кожи, связанные с кровотоком. Помимо сопоставительных сетей, в последние годы активно применяются диффузионные модели, которые позволяют достигать невероятной детализации текстур, делая синтетическое изображение практически неотличимым от реальной фотографии. Эти процессы требуют значительных вычислительных мощностей, однако с развитием облачных технологий создание качественного дипфейка стало доступно даже пользователям без глубоких технических навыков.

Рассматривая сферы применения дипфейков, нельзя игнорировать их дуальную природу. В конструктивном русле технология открывает революционные возможности для кинематографа, позволяя проводить цифровую реставрацию классических произведений или осуществлять «омоложение» актеров для ретроспективных сцен без использования сложного грима. В образовании дипфейки могут использоваться для создания интерактивных лекций от лица исторических деятелей, что значительно повышает вовлеченность учащихся. Тем не менее, деструктивные аспекты технологии вызывают гораздо большее беспокойство. Злоумышленники используют синтез голоса для совершения мошеннических операций, имитируя звонки от доверенных лиц или руководителей. Более того, феномен дипфейков порождает так называемую «кризисную неопределенность», когда само наличие возможности подделки видео позволяет реальным преступникам отрицать подлинность улик, ссылаясь на то, что представленные материалы являются продуктом работы нейросетей.

Проблема обнаружения дипфейков сегодня является одной из приоритетных задач для специалистов по кибербезопасности. Несмотря на высокую точность алгоритмов, синтетические медиа всё еще обладают рядом характерных дефектов. При внимательном анализе видеоряда можно заметить неестественную динамику моргания, отсутствие синхронизации между движением губ и звуковыми волнами, а также визуальные артефакты в области границ волос и аксессуаров. Технические методы детекции включают в себя спектральный анализ частот, который выявляет цифровой шум, нехарактерный для линз реальных камер. Однако гонка вооружений между создателями фейков и разработчиками систем защиты продолжается: как только появляется новый метод обнаружения, нейросети обучаются обходить его, что требует постоянного обновления защитных протоколов и внедрения криптографических методов подтверждения подлинности контента, таких как цифровые подписи на уровне аппаратного обеспечения камер.

В заключение следует отметить, что технология Deepfake представляет собой закономерный этап развития информационных технологий, который несет в себе как огромные риски, так и значительный созидательный потенциал. Главным инструментом защиты в долгосрочной перспективе станет не только развитие автоматизированных систем детекции, но и повышение уровня медиаграмотности населения. Умение критически оценивать входящую информацию, проверять источники и осознавать возможности современных нейросетей должно стать базовым навыком каждого пользователя в цифровом пространстве. Правовое регулирование данной сферы в Республике Беларусь и на международном уровне также требует адаптации, чтобы обеспечить надежную защиту личности от несанкционированного использования её биометрических данных и предотвратить подрыв основ информационной безопасности общества.

Особого внимания заслуживает стремительное развитие технологий синтеза речи, которые зачастую рассматриваются как сопутствующий, но не менее опасный компонент видео-дипфейков. Клонирование голоса (Voice Cloning) сегодня достигло такого уровня, что нейросети достаточно всего тридцати секунд аудиозаписи оригинала, чтобы воспроизвести индивидуальный тембр, интонации и даже характерные дефекты речи конкретного человека. В отличие от видеоряда, аудио-контент воспринимается человеческим мозгом более доверительно, так как слух менее критичен к мелким искажениям, чем зрение. Это создает идеальные условия для совершения преступлений в сфере социальной инженерии. Например, в банковском секторе фиксируются атаки, при которых мошенники используют синтезированный голос высокопоставленного сотрудника для подтверждения несанкционированных платежей. Технически это реализуется через рекуррентные нейронные сети и алгоритмы WaveNet, которые моделируют звуковую волну на атомарном уровне, учитывая даже естественные паузы на вдох и специфическую ритмику речи субъекта.

Рассматривая психологическое воздействие дипфейков на социум, необходимо выделить феномен «информационной усталости». Когда плотность синтетического контента в сети достигает критической массы, у рядового пользователя притупляется способность к верификации данных. Возникает эффект тотального недоверия, который парадоксально играет на руку создателям фейков. В условиях, когда «любое видео может быть подделкой», реальные доказательства военных преступлений, коррупционных схем или экологических катастроф могут быть девальвированы простой фразой о том, что это продукт работы ИИ. Таким образом, дипфейки атакуют не только конкретную личность, чей образ был украден, но и саму структуру общественного договора, базирующуюся на существовании объективной, проверяемой истины. Это ставит перед разработчиками информационных систем задачу создания не просто «детекторов лжи», а полноценной экосистемы верифицированного контента.

Перспективным направлением в борьбе с деструктивным влиянием дипфейков является внедрение концепции «контролируемого происхождения» (Provenance). Ведущие технологические консорциумы уже работают над протоколами, которые позволят камерам смартфонов и профессиональному оборудованию автоматически подписывать каждый снятый кадр криптографическим ключом на аппаратном уровне. В такой модели при загрузке видео в социальные сети система автоматически проверяет цепочку подписей: если видео подвергалось глубокой нейросетевой обработке, цифровая подпись нарушается, и контент помечается как «синтетический». Это перекладывает бремя доказательства с пользователя на создателя контента, заставляя последнего подтверждать подлинность своих материалов. Однако внедрение таких стандартов требует глобального консенсуса между производителями «железа», разработчиками софта и государственными регуляторами.

Нельзя забывать и о роли образования в формировании устойчивого к манипуляциям общества. Современная школа должна адаптировать учебные программы по информатике, смещая акцент с простого освоения офисных приложений на изучение основ машинного обучения и алгоритмической грамотности. Учащиеся должны понимать, как работают рекомендательные системы и какие возможности открывают генеративные модели. Только понимая внутреннюю «кухню» создания дипфейка, человек сможет развить интуитивное недоверие к контенту, который выглядит слишком идеальным или провокационным. В этом контексте международные конференции, подобные проводимой в ПолесГУ, становятся важнейшей площадкой для обмена опытом между молодежью, которая завтра станет основными потребителями и создателями цифрового мира.

Список использованных источников

1. Бранденбургер, А. Бизнес – это игра. Конкурентное сотрудничество и теория игр / А. Бранденбургер, Б. Нейлбафф; пер. с англ. А. Аккерманн. – 2-е изд. – Москва: Альпина Паблишер, 2012. – 350 с. – ISBN 978-5-9614-1782-9.
2. Дегтерев, Д. А. Теоретико-игровой подход в маркетинге: учебное пособие / Д. А. Дегтерев. – Москва: Издательство Российского университета дружбы народов, 2014. – 272 с. – ISBN 978-5-209-05893-7.
3. Замолоцких, В. С. Применение теории графов для анализа социальных сетей: учебное пособие / В. С. Замолоцких, В. Г. Сидоренко. – Москва: Российский университет транспорта (МИИТ), 2020. – 73 с.
4. Кацов, И. Искусственный интеллект на предприятии. Теория и практика для маркетинга, управления поставками и производством / И. Кацов. – Санкт-Петербург: Питер, 2023. – 384 с. – ISBN 978-5-4461-2296-7.
5. Огарков, А. Интеллектуальный маркетинг / А. Огарков. – Москва: РИПОЛ классик, 2024. – 320 с. – ISBN 978-5-386-15246-8.