Федоровская Дарья Александровна
студентка
Научный руководитель
Бобровникова Наталия Сергеевна
старший преподаватель
ФГБОУ ВО «Тульский государственный
педагогический университет им. Л.Н. Толстого»
г. Тула, Тульская область

ПРОФИЛАКТИКА ФЕНОМЕНА «ДИПФЕЙК (DEEPFAKE)» КАК ФАКТОР ОБЕСПЕЧЕНИЯ ЦИФРОВОЙ БЕЗОПАСНОСТИ

Аннотация*: в статье рассматривается проблема кибербуллинга, усугубляемая распространением феномена «дипфейк», который используется кибер-агрессорами для нанесения ущерба репутации жертве. Подчеркивается важность профилактических мер, включая повышение осведомленности о дипфейках, развитие критического мышления и разработку инструментов для их обнаружения. Автором анализируются признаки, по которым можно идентифицировать сгенерированный контент, и предлагаются стратегии противодействия, такие как соблюдение правил кибербезопасности и проверка информации из альтернативных источников.*

Ключевые слова*: дипфейк, профилактика дипфейков, кибербуллинг, буллинг, компрометирующие изображения.*

В цифровом обществе, характеризующимся стремительным развитием информационных технологий, проблема кибербуллинга приобретает новые, более сложные формы, связанные с созданием и распространением дипфейков. Дипфейки, синтезированные видео- или аудиозаписи, представляют собой один из инструментов кибербуллинга, позволяющий злоумышленникам создавать заведомо ложные материалы, компрометирующие репутацию автора оригинала видео, распространять дезинформацию и манипулировать общественным мнением.

Отмечая рост случаев применения технологии дипфейков для реализации стратегий травли в онлайн-среде, возникает острая необходимость в разработке и внедрении комплексных мер профилактики, направленных на повышение осведомленности, развитие критического мышления, создание технических инструментов обнаружения и совершенствование нормативно-правовой базы.

Кибербуллинг – это современное понятие, которое обозначает психологическое насилие в сети. Первичным же было понятие «буллинг» (от англ. «bullying» – хулиган, насильник), обозначающее насилие, запугивание, направленное на то, чтобы вызвать у жертвы страх и тем самым подчинить себе [1]. Несмотря на то, что термин «буллинг» относительно нов, он описывает давнее, укоренившееся явление – проявление жестокости в социальной среде.

Существенный вклад в изучение феномена буллинга внесли многочисленные зарубежные работы таких исследователей как Д. Олвеус, К. Салмивалли, П.К. Смит, Д. Лэйн и др., которые позволили расширить понимание его проявлений и идентифицировать определяющие факторы. Так, например, американский исследователь Дэвид Иден Лэйн дает следующее определение буллинга: «Буллинг – это длительное физическое или психическое насилие со стороны одного человека или группы в отношении отдельного ученика, который в данной негативной ситуации не может за себя постоять [6]».

Согласно результатам опроса «Лаборатории Касперского», каждый десятый ребенок получал оскорбляющие его сообщения в Интернете либо сталкивался с активной травлей в виде специально созданных для этого пабликов и чатов. О том, что в подобной ситуации оказывались их друзья и знакомые, сообщили 18% детей [3]. Одной из новейших форм онлайн-агрессии является травля с использованием дипфейков – технологии, позволяющей с помощью искусственного интеллекта создавать синтезированный медиаконтент, подменяя одного человека другим для создания компрометирующих ситуаций. Согласно данным исследования «Лаборатории Касперского» так же было выяснено, что 1% детей и 2%

взрослых уже сталкивались с кибербуллингом, включающим создание дипфейков с их участием, и, вероятно, эта проблема будет лишь обостряться по мере развития технологий ИИ.

Термин «дипфейк» впервые появился в 2017 году на платформе Reddit [2] и первоначально обозначал использование искусственного интеллекта для реалистичной модификации лиц на изображениях. Однако, с развитием технологии, дипфейки стали включать в себя как изображение, так и звук, создаваемые одновременно или по отдельности. Современные дипфейк-технологии позволяют создавать аудио- и видеоматериалы, имитирующие голос и/или изображение другого человека, и они стали доступны широкому кругу пользователей благодаря простым приложениям для смартфонов (например, FaceApp, Zao, ReFaceApp). Несмотря на положительное применение дипфейков в киноиндустрии, образовании, здравоохранении и развлечениях, их доступность для обычных пользователей повышает риск неправомерного использования этой технологии. Таким образом, дипфейки представляют собой значительную угрозу, потенциал которой требует детального изучения. Ключевую роль в снижении негативного влияния дипфейков играет профилактика, которая должна носить комплексный характер и охватывать различные уровни [5].

Профилактика распространения дипфейков должна носить многоуровневый характер, включающий просветительскую деятельность, направленную на повышение медиаграмотности, и внедрение технологических решений, позволяющих выявлять и блокировать компроментирующий контент. Проанализируем практические способы для распознавания дипфейков и их признаки, а также рассмотрим стратегии активной защиты как элементы комплексной профилактической программы.

Для повышения вероятности идентификации дипфейков и снижения рисков, связанных с их распространением, рекомендуется учитывать следующие признаки сгенерированного контента, проявляющиеся при анализе аудио- и видеоматериалов, вызывающие сомнения в их достоверности [4]:

−область лица более размыта, чем другие участки изображения или видео;

−освещение лица отличается от освещения остального пространства;

−область по периметру лица имеет измененный оттенок кожи в сравнении с другими его участками;

−брови, нос, губы, глаза или все лицо периодически дублируются или «пикселизируются»;

−мимика и движения кажутся неестественными;

−речь собеседника постоянно прерывается;

−интонации собеседника меняются часто и неестественно;

−фоновые звуки не соответствуют пространству, в котором находится собеседник.

При оценке достоверности аудио- и видеоматериалов рекомендуется учитывать совокупность перечисленных критериев и руководствоваться собственными ощущениями, возникающими при просмотре. Субъективное ощущение несоответствия может являться дополнительным аргументом в пользу сомнения в подлинности ролика.

Российский сервис для проведения онлайн-конференций Контур.Толк рекомендует придерживаться следующих практических стратегий для противодействия технологии дипфейков [7]:

− *соблюдайте правила кибербезопасности*. Антивирусы, двухфакторная аутентификация, сложные неповторяющиеся пароли, постоянное обновление ПО – это база кибербезопасности, без которой данные под угрозой;

− *ограничьте доступ к соцсетям*. Закройте доступ к своим социальным сетям от незнакомых пользователей. Лучший вариант – выкладывать как можно меньше данных о себе в открытые источники и на сторонние сайты;

− *проверяйте людей по второму каналу связи*. Если вам поступила необычная просьба, например передать конфиденциальные данные или перевести деньги, свяжитесь с человеком другим способом, чтобы подтвердить информацию. Например, если написали в мессенджере, позвоните по сотовому телефону или напишите в соцсети.

Список литературы

1. ABBYY Lingvo Англо-русский словарь общей лексики «Lingvo Universal» // Словариум. Словари для компьютеров и смартфонов: сайт [Электронный ресурс]. – Режим доступа: https://dic.1963.ru/121 (дата обращения: 02.11.2025).

2. Kugler Matthew B., Pace Carly. Deepfake Privacy: Attitudes and Regulation. 116 Nw. U. L. Rev. 611 (2021).

3. Большинство детей, рассказавших родителям об онлайн-травле, справились с проблемой благодаря этому [Электронный ресурс]. – Режим доступа: https://www.kaspersky.ru/about/press-releases/bolshinstvo-detej-rasskazavshih-roditelyam-ob-onlajn-travle-spravilis-s-problemoj-blagodarya-etomu?ysclid=mb4wqgiy9y739595527 (дата обращения: 02.11.2025).

4. Власов Е. Как не стать жертвой дипфейка – 8 советов для защиты личной и корпоративной репутации / Е. Власов [Электронный ресурс]. – Режим доступа: https://rb.ru/columns/deepfake-defence/ (дата обращения: 04.11.2025).

5. Ефремова М.А. Дипфейк (deepfake) и уголовный закон / М.А. Ефремова, Е.А. Русскевич // Вестник Казанского юридического института МВД России. – 2024. – Т. 15. №2 (56). – С. 97–105. – DOI 10.37973/VESTNIKKUI-2024-56-13. – EDN LXAWLM

6. Лэйн Д.А. Школьная травля (буллинг) // Детская и подростковая психотерапия / Д.А. Лэйн. – СПб., 2001 – 438 с.

7. Теплоухов Р. Как распознать дипфейк / Р. Теплоухов [Электронный ресурс]. – Режим доступа: https://kontur.ru/talk/spravka/80808-raspoznat_dipfeyk#header_80808_5 (дата обращения: 04.11.2025).