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

ПСИХОЛОГИЧЕСКАЯ БЕЗОПАСНОСТЬ В УСЛОВИЯХ ЭВОЛЮЦИИ КИБЕРБУЛЛИНГА: ФЕНОМЕН ДИПФЕЙК-ТРАВЛИ

Аннотация*: в статье рассматривается проблема кибербуллинга и ее изменения в условиях развития технологий искусственного интеллекта, в частности, появление дипфейк-буллинга. Описаны различные аспекты кибербуллинга, приводится классификация кибербуллинга, а также анализируются риски, связанные с использованием дипфейков для распространения ложной информации, манипулирования общественным мнением и совершения мошеннических действий.*

Ключевые слова*: кибербуллинг, дипфейк, искусственный интеллект, буллинг, контент, компроментирующие изображения.*

Актуальность проблемы кибербуллинга существенно возрастает в контексте развития технологий искусственного интеллекта. В частности, чат-боты, порождающие новую форму цифровой травли – дипфейк-буллинг. Данная ситуация обуславливает необходимость модернизации образовательных систем, в рамках которой требуется целенаправленное формирование у школьников компетенций в области цифровой гигиены, медиаграмотности и психологической устойчивости для противодействия принципиально новым формам цифрового насилия.

В 1970-х годах шведский психолог Дэн Олвеус вместе с соавторами впервые ввел в научный обиход понятие «буллинг» после проведения первого лонгитюдного анализа подростковой агрессии в школьной среде совместно с коллегами.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

Далее они определили понятие травли (bullying от анг. bully – хулиган, драчун, задира, грубиян, насильник, bullying – запугивать, издеваться, тиранизировать) [10].

В 1997 году Бил Белси [2] вывел определение кибербуллинга. По его мнению, кибербуллинг – это использование информационных и коммуникативных технологий, например, электронной почты, мобильного телефона, личных интернет-сайтов, для намеренного, неоднократного и враждебного поведения лица или группы, направленного на оскорбление других людей. Ученый выделил главное различие буллинга и кибербуллинга, обозначив современные технологии, как средство травли жертв. В остальном же, а именно: враждебное поведение и неоднократные попытки травли – схожи с обычным буллингом.

Если раньше классический буллинг можно было относительно легко классифицировать, то сегодня проявления уже его новой формы – кибербуллинга, стали сложнее, масштабируемее и психологически опаснее. Анализ последних инцидентов позволяет выделить 10 основных современных вида кибербуллинга [4]:

− *флейминг* (англ. flaming – воспламенение) – наиболее бурно развивающаяся форма кибербуллинга, которая начинается с оскорблений и перерастает в быстрый эмоциональный обмен репликами, обычно публично, реже в частной переписке;

− *киберхарассмент* (англ. harassment – притеснение). Обычно выражается в повторяющихся оскорбительных сообщениях жертве, от которых она чувствует себя морально уничтоженной, которым она не может ответить по причине страха или невозможности идентифицировать преследователя;

− формой харассмента является *троллинг*. Кибертролли (cyber trolls) публикуют негативную, вызывающую тревогу информацию на вебсайтах, страницах социальных сетей, даже на мемориальных страницах, посвященных умершим людям, провоцируя сильную эмоциональную реакцию;

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

− специфическую форму харассмента осуществляют так называемые *гриферы* (griefers) – игроки, целенаправленно преследующие других игроков в многопользовательских онлайн-играх;

− *кибersталкинг* (cyberstalking; от англ. to stalk – преследовать, выслеживать) – использование электронных коммуникаций для преследования жертвы через повторяющиеся вызывающие тревогу и раздражение сообщения, угрозы противозаконных действий или повреждений, жертвами которых могут стать получатель сообщений или члены его семьи;

− *секстинг* (sexting, от англ. sex – секс и text – текст) – это рассылка или публикация фото- и видеоматериалов с обнаженными и полуобнаженными людьми;

− *клевета* (диссинг – denigration) – обнародование обманчивой, унижающей, обвиняющей информации;

− *хеппислеппинг* (video recording of assaults/happy slapping and hopping) – хулиганское нападение на прохожего группой подростков, во время которого один из хулиганов снимает происходящее на видеокамеру мобильного телефона.

− *социальная изоляция* (*бойкот*) – избегание жертвы, яркое нежелание общаться, исключение из электронных групп и/или бесед.

В январе 2025 года было опубликовано исследование [9] «Российские школьники: о карьере, патриотизме, буллинге и кибербезопасности». В исследовании приводятся следующие статистические данные о кибербуллинге в подростковой среде: каждый пятый (19%) опрошенный школьник сообщил о том, что подвергался кибербуллингу. Среди случаев кибербуллинга, в 54% агрессия проявлялась в форме троллинга и оскорблений, 40% респондентов сталкивались с угрозами, 31% подвергались клевете, а 26% и 23% сталкивались с вымогательством и публикацией «отфотошопленных» фотографий в социальных сетях соответственно.

Проблема кибербуллинга обостряется с усовершенствованием уже существующих технологий и появлением новых. Таким образом новоиспеченной угрозой для российских школьников стало появление нейросетей и искусствен-

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

ного интеллекта (далее – ИИ). Если раньше для создания нелицеприятной фотографии порнографического или иного оскорбляющего характера необходимо было обладать особыми умениями и знаниями в сфере фотошопа, то сейчас для генерации подобной картинки достаточно составить текст, в контексте использования нейросетей – промт (промт – запрос, по которому нейросеть генерирует ответ (от англ. prompt -»подсказка») [8] и отправить его в ИИ-бот, который на основе отправленного фото и промта сгенерирует изображение, порочащее личность автора фото. Это приводит к появлению новой формы клеветы в контексте кибербуллинга – созданию *дипфейков*.

Подростковый кибербуллинг с использованием порнографических дипфейков становится глобальной проблемой, о чем свидетельствует доклад Stanford Cyber Policy Center (США) [6]. В докладе подчеркивается, что создание нейросетями порнографических изображений сверстников подростками представляет собой серьезную проблему, к которой оказались не готовы школы, полиция и действующее законодательство.

Существует несколько определений понятия «дипфейк». Ян Гудфеллоу выдвинул подобное определение этого феномена: Deepfake (Дипфейк, от Deep iearning – глубинное изучение и Fake – подделка) является синтезом изображения, основанным на искусственного интеллекте [3]. Исследователь М. Вестерлунд предлагает следующее определение: «Дипфейки – гиперреалистичные видео, в которых используется искусственный интеллект (ИИ), чтобы изобразить, как кто-то говорит и делает то, чего никогда не было» [5].

Первый зафиксированный дипфейк был опубликован в 2017 году анонимным пользователем платформы Reddit, который загрузил ролики, содержащие порнографические ролики с лицами известных актрис (Тейлор Свифт, Скарлетт Йохансонн) [1]. По имени пользователя стали называть все дальнейшие подобные инциденты. Существуют два основных типа дипфейков: видеодипфейки, манипулирующие внешностью, и аудиодипфейки, манипулирующие голосом. Зачастую они используются совместно для создания убедительного, но ложного контента [7].

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

Видеодипфейки это наиболее распространенный тип, использующий ИИ для замены лица или тела человека в реально существующих видеоматериалах. Примеры включают в себя создание фальшивых заявлений от имени известных личностей или подмену лиц в порнографических материалах.

Аудиодипфейки представляют собой технологию, позволяющую манипулировать голосами, чтобы они звучали как чужие. Широко используется в мошеннических схемах, например, для вымогательства денег у родителей под видом звонка от ребенка, попавшего в беду.

Дипфейки представляют собой серьезную и быстро развивающуюся угрозу в контексте кибербуллинга и онлайн-безопасности детей. Простота создания дипфейков, их высокая степень реалистичности и широкая распространенность в интернете делают детей и подростков особенно уязвимыми для различных форм онлайн-вреда. Актуальным является проведение мероприятий по информированию и осведомлению о дипфейках, разработка инструментов для их выявления и внедрение мер, направленных на защиту детей и подростков от негативных последствий использования данной технологии.

Список литературы

1. Albahar M., Almalki J. Deepfakes: Threats and countermeasures systematic review // Journal of Theoretical and Applied Information Technology. 2019. Vol. 97. No. 22. Pp. 3242–3250.

2. Belsey B. The World’s First Definition of «Cyberbullying» / Mr. Belsey // «Always On? Always Aware!» [Electronic resource]. – Access mode: http://www.cyberbullying.ca/ (date of request: 15.10.2025).

3. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu. Generative Adversarial Networks // Advances in Neural Information Processing Systems. 2014. No. 3 (11).

4. Kowalski R.M. Bullying in the digital age: A critical review and meta-analysis of cyberbul-lying research among youth / R.M. Kowalski, G.W. Giumetti, A.N. Schroeder [et al.] // Psychol. Bull. 2014. Vol. 140 (4). Pp. 120–137.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

5. Mika Westerlund The Emergence of Deepfake Technology: A Review // Technology Innovation Management Review. 2019. No. 9 (11). Pp. 39–52.

6. Pfefferkorn Riana, Grossman Shelby, Liu Sunny. AI-Generated Child Sexual Abuse Material: Insights from Educators, Platforms, Law Enforcement, Legislators, and Victims // Stanford University Libraries [Electronic resource]. – Access mode: https://purl.stanford.edu/mn692xc5736 (date of request: 18.10.2025).

7. What is a Deepfake? [Electronic resource]. – Access mode: https://www.internetmatters.org/ru/resources/what-is-a-deepfake/ (дата обращения: 19.10.2025).

8. Кириллов А. Что такое промпт для нейросети и как его составлять / А. Кириллов, Н. Низамова, А. Павлова // Яндекс Практикум. Блог: сайт [Электронный ресурс]. – Режим доступа: https://practicum.yandex.ru/blog/chto-takoe-prompt-dlya-neyroseti-kak-sozdat/ (дата обращения: 18.10.2025).

9. Новое исследование МП Аналитики – «Российские школьники: о карьере, патриотизме, буллинге и кибербезопасности» // Михайлов и партнёры: сайт [Электронный ресурс]. – Режим доступа: https://m-p.ru/media/novoe-issledovanie-mp-analitiki-rossijskie-shkolniki-o-karere-patriotizme-bullinge-i-kiberbezopasnosti/ (дата обращения: 15.10.2025).

10. Янова Н.Г. От буллинга к антибуллингу: школьные программы профилактики агрессии / Н.Г. Янова. – Барнаул: Принт-Экспресс, 2021. – 180 с.

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)