ПАЛІТАЛОГІЯ

УДК 32.019.5

Владислав Олегович Калишук
аспирант 2-го года обучения каф. политологии
Белорусского государственного университета
Vladislav Kalishuk
Post-graduate the 2nd at the Department of Political Science
at the Belarusian State University
e-mail: rristar@mail.ru

НЕДОСТОВЕРНАЯ ИНФОРМАЦИЯ В ПОЛИТИЧЕСКОЙ КОММУНИКАЦИИ: МЕТОДЫ И ТЕХНОЛОГИИ ПРОТИВОДЕЙСТВИЯ

На основе анализа современной научной литературы предпринята попытка комплексного анализа методов и технологий противодействия распространению дезинформции в информационном поле. Для достижения цели работы были выявлены наиболее распространенные виды недостоверной информаци – фейковые новости (и их подвиды) и дипфейки. Предложена классификация конкретных способов противодействия недостоверной информации в зависимости от видовой принадлежности. Сделан вывод о том, что современные методы противодействия находятся на начальном этапе своего технологического развития, а для стимулирования их дальнейшего совершенствования требуется участие государства в форме комплексного правового регламентирования правоотношений, возникающих в результате распространения недостоверной информации, и финансирования.

Ключевые слова: *фейки, недостоверная информация, противодействие, методы, технологии.*

Unreliable Information in Political Communication: Methods and Technologies of Counteraction

In the article, based on the analysis of modern scientific literature, an attempt was made to comprehensively analyze the methods and technologies for countering the spread of disinformation in the information field. To achieve the goal of the work, the most common types of inaccurate information were identified: fake news (as well as their subtypes) and deepfakes. A classification of specific methods of counteracting inaccurate information, depending on the species, was proposed. It was concluded that modern methods of counteraction are at the initial stage of their technological development, and to stimulate their further improvement, the participation of the state is required in the form of a comprehensive legal regulation of legal relations arising from the dissemination of inaccurate information and funding.

Key words: *fakes, false information, counteraction, methods, technologies.*

Введение

Роль информационных технологий в мире продолжает расти. Они активно включаются практически во все сферы деятельности личности и общества. Так, например, по данным Internet World Stats за декабрь 2019 г., количество пользователей Интернета в мире составило 4,5 млрд человек [1], т. е. около 59 % населения всей планеты, а по информации Digital 2020 – We are social за январь 2020 г., число активных пользователей социальных медиа насчитывало 3,8 млрд человек, т. е. около 49 % населения Земли [2]. Информационные технологии внедряются в государственное управление, что проявляется в развитии «электронного правительства» по всему миру [3] и стремлении государственных органов включаться в процесс коммуникации на базе социальных медиа.

Растет значение информации как таковой. Защищенность, скорость распространения и достоверность становятся важнейшими ее характеристиками. Изучение каждой из этих характеристик имеет огромное значение, однако, по нашему мнению, особую актуальность представляет проблема достоверности информации, т. к. с распространением Интернета и особенно социальных медиа, количество фейковой информации значительно увеличилось, как и разме-

Научный руководитель – О. Е. Побережная, кандидат политических наук, доцент кафедры политологии Белорусского государственного университета

ры реального и потенциального ущерба от ее распространения. Специалисты из Балтиморского университета и компании CHEQ оценивают «стоимость» фейковых новостей в 2019 г. в 78 млрд долл., а прогнозируемый ущерб, который будет причинен их распространением составит около 9 млрд долл. в области здравоохранения, 17 млрд долл. из-за финансовой дезинформации, 9 млрд. долл. вследствие имиджевых потерь, 3 млрд долл. по причине трат на обеспечение безопасности и 400 млн долл. из-за распространения фейковой информации в области политики [4].

Существует масса видов и форм фейковой информации: слухи, мисинформация, дезинформация и т. д. Однако наибольшие опасения у специалистов в области информационной безопасности вызывают фейковые новости и дипфейки (более подробно их сущность будет раскрыта ниже). Опасность конкретно этих видов недостоверной информации обусловлена в первую очередь их широчайшим деструктивным потенциалом. В частности, они могут применяться для следующих целей: воздействие на избирательные кампании, разжигание ненависти между этнонациональными группами, стимулирование межгосударственных конфликтов. Вместе с тем это далеко не все варианты использования данных видов фейковой информации. Более того, эти технологии активно используются на практике. Например, во время предвыборной кампании США Дональда Трампа вирусные фейковые новости привлекали на Facebook больше внимания, чем реальные новости [5, с. 100], а в создании фейковых новостей и со стороны Д. Трампа, и со стороны Х. Клинтон принимали участие около 19 млн ботов [6].

Данная проблема актуальна и для Республики Беларусь, т. к. около 74 % населения страны являются пользователями Интернета, а около 40 % – активные пользователи социальных медиа. Страна занимает 33 место в рейтинге «электронного участия» [3]. Беларусь является активным участником глобального информационного пространства, однако при этом находится в сфере влияния информационных полей Российской Федерации. В информационном пространстве Республики транслируется огромное количество разноплановой информации из различных источников.

В связи с этим возникает необходимость разработки методов выявления и противодействия фейковым новостям и дипфейкам. В отечественной и российской научной литературе данная проблема освещается достаточно скудно, хотя и отмечается ее актуальность. Ключевым прикладным исследованием, предлагающим конкретную методику выявления фейковых новостей, является работа специалистов из Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики А. О. Третьяковой, О. Г. Филатова, Д. В. Жука, Н. Н. Горлушкина, А. А. Пучковской «Метод определения русскоязычных фейковых новостей с использованием элементов искусственного интеллекта» [5]. В зарубежной научной литературе данная тематика значительно более проработана, однако конкретных универсальных и эффективных прикладных методик противодействия фейкам на сегодняшний день предложено не было. В основном ученые акцентируют внимание на разработке базовых принципов выявления фейковых новостей и дипфейков, пытаются создать некую основу для дальнейших исследований и методик.

В данной статье мы попытаемся описать, систематизировать существующий прикладной опыт. Таким образом, цель статьи – проанализировать основные методы и технологии борьбы с недостоверной информацией в политической коммуникации и предложить рекомендации по совершенствованию системы противодействия ее распространению. Объект исследования – недостоверная информация в политической коммуникации, предмет исследования – методы и технологии борьбы с недостоверной информацией.

Основная часть

Проблема распространения недостоверной информации не нова и, вероятно, существовала на протяжении практически всего развития человечества. Первые упоминания о намеренном распространении фейковых новостей относятся к XII–XIII вв. и связываются с завоеваниями Чингисахана. Дипфейки – это новейшая технология на

пике научно-технического прогресса, которая появилась только в 2017 г. [5, с. 100].

В качестве причин стремительного распространения фейковых новостей и дипфейков можно указать следующие:

1) широкое внедрение социальных медиа; они охватывают огромные аудитории по всему миру, являются одним из основных источников получения информации об окружающей действительности и позволяют без серьезных материальных затрат распространять информацию;

2) заинтересованность субъектов политического процесса и бизнеса в распространении фейковой информации, вытекающая из высокой степени окупаемости технологий распространения фейковой информации (в особенности фейковых новостей).

Вместе с тем следует отметить, что фейковые новости и дипфейки можно использовать в конструктивном ключе – для создания развлекательного контента и в медицинских целях.

Определимся с дефинициями. Трактовок термина «фейковые новости» довольно много, однако общепризнанным в Западной и отечественной литературе является понятие, предложенное Европейской комиссией и предполагающее два подхода к определению сущности «фейкньюз»: широкий (новости, охватывающие весь спектр недостоверной информации) и узкий (новости, содержащие недостоверную информацию, заложенную в них умышленно) [7, с. 3–4; 5, с. 100]. Дипфейки (deepfakes) – это технология создания медиаконтента, базирующаяся на «глубоком машинном обучении» и позволяющая синтезировать гиперреалистичные изображения, как правило, путем накладывания изображения одного человека на видео и/или фото с изображением другого человека [8, с. 1].

Теперь необходимо определить место фейковых новостей и дипфейков в структуре видов недостоверной информации. Как считают американские специалисты в области информационной безопасности Ксинь Жоу и Резы Зафарани, в зависимости от закладываемого в фейковые новости смысла можно выделить фейковые новости негативного характера и фейковые новости без ярко выраженного негативного характера [7, с. 4]. К фейковым новостям негативного характера относятся злонамеренные фейковые новости и дезинформация. К фейковым новостям без ярко выраженного негативного характера относятся сатирические новости. На наш взгляд, можно дополнительно выделить фейковые новости с неизвестным характером интенции: ошибочные (ложные) новости, мисинформацию и слухи.

Таким образом, мисинформация, слухи, сатирические новости – это одновременно виды недостоверной информации и инструменты создания фейковых новостей (в зависимости от способа их применения). Например, слух может распространяться среди населения, а затем использоваться в качестве основы для создания фейковой новости или быть включен в ее структуру. По нашему мнению, дипфейки являются высокотехнологичным видом недостоверной информации и, как правило, выступают в качестве инструмента создания фейковых новостей.

Более того, в позиции К. Жоу и Р. Зафарани есть неточность: не указывается место (где, в чем?) дипфейков, столь распространившихся на сегодняшний день. Следуя логике К. Жоу и Р. Зафарани, можно выделить следующие виды фейковых новостей на базе интенции их создателя: злонамеренные фейковые новости и случайные фейковые новости, а также сатирические фейковые новости. Наибольшую угрозу представляют злонамеренные фейковые новости, однако нельзя не учитывать и случайные фейковые новости, которые не только доносят в фоновом режиме до потребителей новостей недостоверную информацию, но и могут создавать эффект «волны», тем самым усиливая негативное воздействие первого вида фейковых новостей.

В свою очередь дипфейки бывают следующих видов: фейковые видео, фейковые аудио (относительно новый вид) и фейковые изображения.

Перейдем к обзору методов выявления фейковой информации. Можно выделить три базовых подхода к *выявлению фейковых новостей* [7, с. 4; 6]:

1) анализ знаний (смысла текста);

2) анализ стиля написания текста;

3) анализ способов распространения.

Подход к выявлению фейковых новостей, основанный на анализе смысла текста. В основе данного подхода лежит т. н. метод проверки фактов/факт-чекинга.

Проверка фактов направлена на оценку достоверности новостей путем сравнения информации, извлеченной из подлежащего проверке новостного контента (например, утверждений или заявлений), с общеизвестными фактами (т. е. истинными знаниями) [7, с. 7; 6].

Существует два вида факт-чекинга – *ручной и автоматический*. Массовое применение ручного факт-чекинга не является эффективным, хотя и дает более качественные результаты. Стоит отметить, что ручная проверка фактов может быть включена в автоматическую. Ручной факт-чекинг можно разделить на экспертный и краудсорсинговый.

Экспертный ручной факт-чекинг опирается на экспертов (факт-чекеров), которые действуют в малых группах. В качестве плюсов применения данной методики можно отметить хорошую управляемость и точные результаты. В качестве минусов – стоимость и плохую масштабируемость (речь идет об увеличении объемов данных). Как правило, данный вид проверки фактов осуществляется на базе различных экспертных интернет-платформ: PolitiFact, The Washington, Post Fact Checker, FactCheck, Snopes, TruthOrFiction, FullFact и т. д. [7, с. 7–8; 6].

Краудсорсиногвый факт-чекинг нацелен на проверку данных обычными людьми, которые выступают в качестве экспертов. Основным его плюсом является широкая масштабируемость. В качестве основных минусов можно выделить низкую точность и необходимость дополнительной обработки результатов проверки (устранение противоречий в заключениях, выявление предвзятости), низкая управляемость. Данный вид ручного факт-чекинга не распространен. В качестве примера платформы, осуществляющей данный вид проверки, можно указать Fiskkit. Ключевыми проблемами данного вида факт-чекинга как технологии выявления и противодействия фейковой информации в целом и фейковым новостям в частности являются:

1) уязвимость к атакам ботов (исходя из сущности данного метода платформа должна обеспечивать свободный доступ к инструментарию широкому кругу пользователей),

2) скорость выявления фейков даже потенциально большим количеством пользователей (каким, например, располагает Wikipedia) не сможет соответствовать скорости генерации фейковых новостей, например, нейросетями.

Хотя специалисты в области информационной безопасности прогнозируют дальнейшее развитие данного метода на базе крупных корпораций, таких как Google, Facebook, Twitter, применение этой технологии, по нашему мнению, возможно только на данном временном промежутке, т. к. в дальнейшем указанные нами проблемы только усугубятся и эффективное применение факт-чекинга на базе краудфайндинга станет возможным только в комплексе с другими мерами.

Автоматический факт-чекинг. Ключевым преимуществом данного метода является то, что он решает проблему масштабируемости, т. к. в гораздо меньшей степени ограничивается объемом обрабатываемой информации. Основным его недостатком выступает низкий уровень точности результатов. Автоматический факт-чекинг базируется на информационном поиске и обработке естественного языка. В рамках данного метода под «знанием» понимаются «тройки» (наборы данных о субъекте, предикате и объекте), которые извлекаются из обрабатываемой информации. Тройки представляются в виде «графа знаний», т. е. как информационный блок, содержащий заголовок, описание страницы и/или дополнительную информацию. Субъект и объект в рамках троек являются «сущностями», это значит обладают определенным состоянием и поведением, имеют определенные свойства (атрибуты), и операции над ними (методы) представлены в виде узлов. В свою очередь, предикаты представляют собой наборы ребер и представлены в виде отношений. Процесс автоматической проверки фактов может быть разделен на два этапа: извлечение фактов (построение базы знаний) и проверка фактов (сравнение знаний). В рамках первого этапа производится сбор т. н. «диких данных» (сбор необработанных данных в виде текста, табличных данных, структурированных страниц и постов реальных людей, которые содержат реляционную информацию и могут быть использованы для извлечения знаний различными экстракторами). Сбор данных может производиться из одного достоверного источника

(при этом снижается охват и полнота извлекаемых знаний) или из множества открытых источников (при этом снижается качество извлекаемых знаний, т. к. объединяются различные источники с разнородной информацией). После извлечения фактов производится их обработка путем выполнения следующих задач [7, с. 11]:

1. Дедупликация, т. е. выявление всех упоминаний, которые относятся к одному и тому же объекту в базе знаний (или в нескольких базах знаний). В основном проверка фактов связана с отношением реляционных объектов и требует дорогостоящих парных вычислений подобия. Как правило, для решения данной задачи используются методы индексации.

2. Слияние знаний с целью обработки противоречивых данных. Зачастую для решения этой задачи для фактов устанавливается система поддерживающих (опорных) ценностей-маркеров либо используются методы машинного обучения.

3. Регистрация времени с целью удаления устаревших знаний. Для этого в основном используется тип составного значения (Compound Value Type). Стоит отметить важность данной задачи, т. к. своевременность – это одна из основополагающих характеристик, определяющих эффективность метода выявления фейковых новостей, однако при этом конкретных разработок в этом направлении довольно мало.

4. Оценка достоверности. Она базируется на контролируемом обучении и статистических выводах.

5. Установление связей между извлекаемыми фактами с целью выявления новых фактов. Для решения данной задачи используются следующие методы:

1) модели скрытых признаков (предполагается, что существование троек базы знаний является условно независимым, учитываются скрытые особенности и параметры),

2) модели признаков графов (предполагается, что существование троек является условно независимым),

3) модели случайного поля (предполагается, что существующие тройки имеют локальные взаимодействия).

После выполнения перечисленных выше задач на основе полученных «очищенных» знаний формируется база знаний.

Этап сравнения фактов заключается в сравнении извлеченных из проверяемого новостного содержимого (из троек) фактов с фактами, хранящимися в построенной или существующей базе знаний или графе знаний с истинным знанием. Как правило, стратегия проверки фактов для тройки заключается в оценке возможности существования ребер с меткой «предикат» от узла с меткой «субъект» до узла, представляющего объект в графе знаний [7, с. 11].

Таким образом, автоматический факт-чекинг является довольно перспективным, хотя и обладает рядом проблем и требует усовершенствования в части повышения скорости проверки новостей. В данном случае наиболее эффективным решением видится формирование динамических баз знаний. Важна так же и проблема точности получаемых результатов. Одним из возможных вариантов повышения качества результатов автоматического факт-чекинга может выступить его комбинирование с ручным краудфандинговым факт-чекингом или с ручным экспертным факт-чекингом в особо сложных случаях, т. е. путем создания своего рода второго уровня проверки. В целом это может позволить повысить уровень точности и сохранить приемлемый уровень оперативности (хотя и недостаточный с учетом повышения скорости распространения информации в целом и недостоверной информации в частности).

Подход к выявлению фейковых новостей, основанный на стиле. Исследования на основе стиля нацелены на оценку *интенции*, заложенной в содержании новостей, а именно в определении, есть ли намерение ввести потребителей контента в заблуждение. Разработки в области обнаружения фейковых новостей на основе стиля находятся на ранней стадии своего развития.

Общая стратегия обнаружения обмана на основе стилей заключается в использовании вектора признаков, представляющего стиль содержимого данной информации в рамках машинного обучения, для прогнозирования того, является ли информация недостоверной (т. е. проблемой классификации) и насколько она недостоверна (проблема регрессии). Ключевым отличием классификации от регрессии является выполняемая ими задача: задача классификации – получение категориального ответа на

основе набора признаков, а задача регрессии – это прогноз на основе выборки объектов с различными признаками. До настоящего времени в большинстве исследований использовались методики обучения «под наблюдением»: формировался набор признаков с соответствующими метками (ложный против правдивого).

В целом данная стратегия полностью применима к выявлению фейковых новостей на основе стиля. Однако необходимо учитывать, во-первых, размытость тематики фейковых новостей (экономика, политика, образование и т. д.) порождает необходимость междоменного, межъязыкового или межпредметного анализа, а во-вторых, высокая степень латентности и необходимость разработки более тонких маркеров.

Стоит отметить и значимость дополнения контролируемого обучения полууправляемым обучением, т. к.

а) количество и размер доступных наборов данных, содержащих помеченные (поддельные и оригинальные) новостные статьи, ограниченны,

б) трудно создать некий набор данных «золотого стандарта»,

в) ручная маркировка практически не масштабируется.

Можно резюмировать, что основной проблемой данного подхода к определению фейковых новостей является неуниверсальность классификаторов. В качестве основного плюса стоит указать потенциальную возможность выявления злонамеренных фейковых новостей. В целом данная методика хорошо дополняет подход, базирующийся на знаниях, и его можно использовать, например, в качестве третьего уровня проверки информации, если брать за основу автоматизированный факт-чекинг.

Более того, потенциал у данного способа гораздо выше, чем у ручного факт-чекинга, т. к. проблема масштабируемости последнего фактически является нерешаемой. На наш взгляд, если в рамках стилистического подхода будет достигнут уровень точности, сравнимый с ручными видами факт-чекинга, то путем комбинации стилистического подхода и автоматического факт-чекинга появится возможность в значительной степени нивелировать проблемы точности и своевременности последнего.

Подход к выявлению фейковых новостей, основанный на анализе способов распространения информации. Данный подход акцентирует внимание на механизме распространения фейковых новостей. Допустимо выделить два типа методов: каскадные и сетевые [7, с. 22].

Под каскадом фейковых новостей понимается древо, или древовидная структура, иллюстрирующая распространение определенной фейковой новостной статьи в социальной сети пользователей. «Корневой» узел каскада представляет пользователя, который первым опубликовал фейковые новости. Другие узлы в каскаде представляют пользователей, которые опубликовали новость после ее публикации «родительскими» узлами, к которым они подключены через «ветви». Каскад может быть представлен с позиции количества шагов (т. е. скачков), которые прошли фейковые новости (каскад на основе «прыжка»), или времени, когда они были опубликованы (временной каскад). Каскадные методы подразделяются на два вида: 1) основанные на выявлении каскадного сходства (сравнение каскадов с использованием ядер графов) и 2) основанный на каскадном представлении (поиск информативных представлений, которые могут быть использованы как функции в контролируемой структуре обучения) (данный метод не является автоматическим). Как альтернатива: можно проводить репрезентативное обучение, которое часто достигается посредством глубокого обучения [9, с. 22–23].

Сетевые методы выявления фейковых новостей направлены на создание гибких сетей трех видов: однородных, гетерогенных, иерархических.

Однородные сети – это сети, содержащие один тип узла и один тип ребра. Типичная однородная сеть – это сеть позиций, где узлами являются посты, связанные с новостями пользователей, а ребра представляют поддерживающие или противоположные отношения между каждой парой постов [9, с. 24]. Выявление фейковых новостей в рамках данного метода сводится к оценке достоверности сообщений, связанных с новостями, которые в дальнейшем могут рассматриваться как проблема оптимизации графов.

Гетерогенные сети имеют несколько типов узлов или ребер. Для такой сети используется гибридная структура с тремя основными компонентами: внедрение и представление сущностей, моделирование отношений и обучение под наблюдением. В иерархических сетях различные типы узлов и ребер образуют отношения «набор – подмножество) (т. е. иерархию). В таких сетях проверка новостей также превращается в задачу оптимизации графов [9, с. 24].

Основными проблемами подхода в целом являются отсутствие автоматизированности большинства методов (проблема масштабируемости и своевременности), а также сложность разграничивания злонамеренных и случайных видов фейковых новостей. Первая проблема потенциально может быть решена при использовании, например, глубокого обучения. Вторая проблема может быть решена путем комбинирования описываемого подхода с подходом выявления фейковых новостей, основанном на стиле. Однако данный подход является довольно эффективным, т. к. выявление создателей фейков и дальнейшая их блокировка или формирование баз «недостоверных» авторов, ресурсов и т. д. позволяют устранять проблему в зародыше. Критерием определения эффективности применения данного подхода в будущем станет возможность его автоматизации.

Методы выявления дипфейков. Для выявления дипфейков возможно применение очень широкого круга междисциплинарных методов. Австралийские специалисты Т. Т. Нгуен, К. М. Нгуен, Д. Т. Нгуен и С. Нахаванди разделили все известные на данный момент методы выявления дипфейков в зависимости от области применения, т. е. на методы выявления дипфейков видео, методы выявления дипфейков изображений и комбинированные методы [8, с. 12].

Методы выявления дипфейков видео:

1. Проверка «моргания». Используются долгосрочные рекуррентные сверточные сети, чтобы узнать временные схемы моргания глаз. Как правило, частота моргания в дипфейках намного меньше, чем у реальных людей.

2. Использование пространственно-временных функций. Временные расхождения между кадрами исследуются с использованием рекурсивных кортикальных сетей, которые объединяют сверточную сеть DenseNet и стробированные рекуррентные элементарные ячейки.

3. Использование внутрикадровых и временных несоответствий. В рамках данного метода сверточные нейронные сети применяются для извлечения характеристик уровня кадра, которые распространяются в долгой краткосрочной памяти для создания дескриптора последовательности.

4. Использование артефактов на изображении лиц. Артефакты обнаруживаются с использованием моделей сверточных нейронных сетей (в частности, VGG16, ResNet50 (101 или 152)), основанных на выявлении несоответствий разрешения искривленной области лица и окружающей области.

5. Проект «MesoNet» – две глубокие сети (Meso-4 и MesoInception-4), представленные для изучения дипфейков уровне мезоскопического анализа. Точность, полученная для наборов данных DeepFake и FaceForensics, равна 98 %.

6. Использование различий в текстуре лица, отсутствующих отражений и деталей в области глаз и зубов. Логистическая регрессия и нейронная сеть используются для классификации.

7. Анализ диаграмм шума светочувствительных датчиков цифровых камер, появляющихся из-за их заводских дефектов, а именно сравнение на предмет различия в шаблонах между подлинным и фейковыми видео [8, с. 12–13]:

Методы выявления дипфейков изображений:

1. Предварительная обработка в сочетании с глубокой сетью, которая предполагает модификацию обобщающей способности моделей для обнаружения изображений, сгенерированных генеративно-состязательной сетью путем удаления низкоуровневых функций фейковых изображений через поиск сходства уровней пикселей между фейковыми и реальными изображениями.

2. Извлечение дискриминантных функций, используя метод «мешков слов», их обработка методом опорных векторов, «случайным» или многослойным персептроном для двоичной классификации: оригинал – подделка.

3. «Парное обучение». Предполагает двухфазную процедуру: извлечение признаков с использованием «общей сети фейковых особенностей» на основе архитектуры сиамской сети и классификацию с использованием сверточных нейронных сетей [8, с. 13].

Комбинированные методы:

1. Метод анализа позы головы. Он основан на извлечении особенностей дипфейков путем использования 68 ориентиров области лица. Извлеченные функции классифицируются методом опорных векторов.

2. «Капсульная криминалистика». Признаки, извлеченные сверточной нейронной сетью VGG-19, поступают в капсульную сеть для классификации. Алгоритм динамической маршрутизации используется для маршрутизации выходов трех сверочных капсул в две выходные капсулы (одну для поддельных, а другую – для реальных изображений) через ряд итераций [8, с. 12].

По мнению норвежских специалистов в области информационной безопасности А. Ходабахса, Киран Б. Раджа, Р. Рагхавендра, из всего многообразия методов выявления дипфейков наиболее эффективными методами являются:

1) метод на основе текстур (локальных двоичных шаблонов),

2) методы на основе сверточных нейронных сетей, т. е. CNN (AlexNet, VGG19, ResNet50, Xception, GoogLeNet/Inceptionv3) [10, с. 3]. Более того, вышеназванные ученые протестировали данные методы на сформированной ими же базе данных «Fake Faces in the Wild» и отметили, что как текстовые дескрипторы, так и методы глубокого обучения *на данном этапе своего развития не способны решить проблему выявления дипфейков*. Основная причина этого заключается в неприменимости классификаторов к разным наборам данных, что порождает большое количество неточных данных. Предлагается дополнительно тестировать все разработки в области выявления дипфейков по нескольким наборам данных. Перспективными видятся дальнейшие исследования и внедрение мультимодальных сигналов [10, с. 6].

Главной особенностью методов выявления дипфейков является практически *полная автоматизированность всех методик*. Стоит отметить и невозможность применения человеческого ресурса в дополнительной обработке результатов проверок, потому что с уже достигнутым качеством дипфейков вероятность распознания дипфейка человеком находится на грани случайности.

Заключение

Таким образом, можно указать следующие проблемы, которые существуют на данный момент в области выявления фейковых новостей и дипфейков.

Во-первых, эффективность существующих технологий выявления недостоверной информации в целом довольно низкая. Отсутствуют технологии, которые позволяют идентифицировать фейковую информацию даже с удовлетворительным уровнем точности. В целом противостояние методов противодействия и создания фейковой информации напоминает противостояние снаряда и брони. Стоит отметить, что первые в значительной степени опережают последних (особенно это касается дипфейков). Развитие технологий их создания значительно опережает методы их выявления. При этом можно констатировать, что методы выявления фейковой информации находятся на начальном этапе своего развития и в целом развиваются недостаточно быстро в отличие от методов их создания и распространения. На наш взгляд, это обосновано высоким уровнем монетизации технологий создания фейковой информации, т. к. ими заинтересованы обычные граждане, бизнес-структуры и даже субъекты политического процесса. В то же время заинтересованность в разработке методов выявления и противодействия фейковой информации только начинает расти. Усугубляет ситуацию и то, что большинство перспективных методик противодействия находится в общем доступе и подробно описано, что дает преимущества лицам, заинтересованным в разработке методов распространения фейков.

Во-вторых, существует системная проблема неуниверсальности методов обнаружения как фейковых новостей, так и дипфейков, т. е. конкретные методы могут демонстрировать высокую точность для одного набора данных и низкую для другого. При этом разработчики зачастую игнорируют этот факт и отмечают высокую эффективность своих методик, проводя проверку только на одном-двух наборах данных.

В качестве перспективных направлений развития и совершенствования методов выявления фейковых новостей и дипфейков можно указать:

1) создание комплексных систем выявления недостоверной информации, базирующихся на комбинировании различных технологий;

2) увеличение уровня автоматизации с целью повышения уровня масштабируемости и своевременности.

Проблема распространения фейковой информации будет только нарастать, а актуальность разработки методов ее выявления повышаться. Как уже было отмечено, системы борьбы с фейками развиваются достаточно медленно. В связи с этим возникает необходимость стимулирования данного процесса уже сейчас, чтобы в дальнейшем нивелировать возможные негативные последствия. Ключевую роль в этом может сыграть государство, в первую очередь, через создание правовой основы для создания эффективной системы выявления и противодействия распространению недостоверной информации. Именно нормативное регулирование (в части определения дефиниций, направлений противодействия, установления уголовно-административной ответственности) должно выступить в качестве первого шага на пути создания конкретных технологических решений.

Базируясь на исследовании российского юриста, директора Глобального центра правовых исследований П. Ровдика «Инициативы по борьбе с поддельными новостями в выбранной стране», в котором были сравнены подходы к правовому регулированию фейковых новостей в 15 странах, можно сделать вывод, что на данный момент мировое сообщество только приходит к осознанию потенциальной опасности данного явления, т. к. в большинстве стран фейковые новости в контексте права находятся только на стадии концептуализации [11].

Условно можно выделить два подхода к регулированию какого-либо потенциально опасного явления: либо через включение его в существующую правовую систему (например, в Уголовный кодекс; в большинстве случаев этого оказывается достаточно), либо через принятие специальных нормативных актов и параллельную имплементацию в существующее законодательство.

Как показывает опыт Великобритании, Бразилии, Канады, Франции, Японии и Никарагуа, недостаточно простого включения дефиниции фейковых новостей в существующее законодательство или простой криминализации их создания и распространения [11, с. 7, 11, 29, 52, 78, 100]. Необходимо разрабатывать и внедрять комплексную систему правового регулирования проблемы распространения фейковой информации через принятие специального законодательства. Однако, базируясь на опыте, например, Израиля и Кении, имеет смысл сделать акцент на максимальной однозначности и точности применяемой терминологии, т. к. субъектами распространения фейковых новостей может выступать невероятно широкий спектр физических и юридических лиц. Помимо этого, из-за распространенности случайной фейковой информации особые сложности могут появиться в квалификации субъективной стороны правонарушений в рамках как уголовного, так и административного права или даже в рамках обоснования вины субъекта как условия возникновения гражданско-правовой ответственности.

На наш взгляд, в развитие статей 33 и 34 Конституции Республики Беларусь [12], а также п. 34 Концепции национальной безопасности [13] допустимо:

1) выделить в отдельное направление противодействие распространению фейковой информации в рамках Концепции информационной безопасности [14];

2) разработать и принять Закон «О противодействии распространению недостоверной информации», который определит понятие и конкретные виды недостоверной информации, а также предусмотрит перечень органов, которые будут ответственны за организацию системы противодействия и мониторинга распространения недостоверной информации;

3) включить в Уголовный и Административный кодексы статьи, предусматривающие ответственность за распространение недостоверной информации.

Наиболее эффективно будет создать Департамент по противодействию распространению недостоверной информации при Министерстве внутренних дел Республики

Беларусь. Это обусловлено в первую очередь тем, что МВД обладает необходимым кадрово-материальным ресурсом для достижения данной цели, а ряд его структурных подразделений на данный момент выполняет функции схожие с теми, которые будет выполнять создаваемый Департамент.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ / REFERENCES

1. Internet world stats [Electronic resource] // Internetworldstats.com. – Mode of access: https://www.internetworldstats.com/stats.htm. – Date of access: 18.09.2021.

2. Digital 2020 – We are social [Electronic resource] // Wearesocial.com. – Mode of access : https://wearesocial.com/digital-2020. – Date of access: 18.09.2021.

3. Rejting stran mira po urovniu razvitija eliektronnogo pravitiel'stva [Eliektronnyj riesurs] // gtmarket.ru. – Riezhim dostupa: https://gtmarket.ru/ratings/e-government-survey/info. – Data dostupa: 16.09.2021.

4. Study Finds' Fake News' Has Real Cost: $78 Billion [Electronic resource] // Mediapost.com. – Mode of access: https://www.mediapost.com/publications/article/343603/study-finds-fake-news-has-real-cost-78-billion.html. – Date of access: 20.09.2021.

5. Mietod opriedielienija russkojazychnykh fejkovykh novostiej s ispol'zovanijem eliemientov iskusstviennogo intielliekta / A. O. Triet'jakov [i dr.] // International Journal of Open Information Technologies. – 2018. – № 8. – S. 99–105.

6. Aldwairi, M. Detecting fake news in social media networks / M. Aldwairi, A. Aldwairi // Conference: The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2018). – Leuven (Belgium), 2018.

7. Zhou, X. Fake news: a survey of research, detection methods, and opportunities [Electronic resource] / X. Zhou, R. Zafarani // Researchgate.net. – Mode of access: https://www.researchgate.net/-publication/329388190_Fake_News_A_Survey_of_Research_Detection_Methods_and_Opportunities. – Date of access: 20.09.2021.

8. Deep learning for deepfakes creation and detection [Electronic resource] / T. N. Thanh [et al.] // researchgate.net. – Mode of access: https://www.researchgate.net/publication/336055871_Deep_Learning_for_Deepfakes_Creation_and_Detection. – Date of access: 20.09.2021.

9. Twitter nachinajet markirovat' i udaliat' fejkovyje foto i vidieo s 5 marta [Eliektronnyj riesurs] // Esquire.ru. – Riezhim dostupa: https://esquire.ru/articles/154513-twitter-nachnet-markirovat-i-udalyat-feykovye-foto-i-video-c-5-marta/. – Data dostupa: 20.03.2021.

10. Fake Face Detection methods: can they be generalized? / A. Khodabakhsh [et al.] // Conference: Bio-Sig. – Darmstadt(Germany), 2018.

11. Roudik, P. Initiatives to Counter Fake News in Selected Countries [Electronic resource] / P. Roudik. – Mode of access: https://www.loc.gov/law/help/fake-news/counter-fake-news.pdf. – Date of access: 20.09.2021.

12. Konstitucija Riespubliki Bielarus' 1994 goda : s izm. i dop., priniatymi na riesp. riefierien-dumakh 24 nojab. 1996 g. i 17 okt. 2004 g. – Minsk : Nac. centr pravovoj inform. Riesp. Bielarus'. – 2014. – 62 s.

13. Koncepcija nacional'noj biezopasnosti Riespubliki Bielarus', utv. Ukazom Priezidienta Riespubliki Bielarus' ot 9 nojab. 2010 g. 575 [Eliektronnyj riesurs] // Konsul'tantPlius. Bielarus' / OOO «YurSpektr», Nac. centr pravovoj inform. Riesp. Bielarus'. – Minsk, 2021.

14. Koncepcija informacionnoj biezopasnosti Riespubliki Bielarus', utv., Sovietom Biezopasnosti Riesp. Bielarus' ot 18 marta 2019 g. № 1 [Eliektronnyj riesurs] // Konsul'tantPlius. Bielarus' / OOO «YurSpektr», Nac. centr pravovoj inform. Riesp. Bielarus'. – Minsk, 2021.

Рукапіс паступіў у рэдакцыю 26.10.2021