



Институт развития  
интернета

Еженедельный правовой дайджест  
Интернет и информационные  
технологии  
Выпуск № 142  
28.04.2023

## Оглавление

<b>Российская Федерация</b> .....	3
Раздел 1. Акты и проекты актов Правительства .....	3
Раздел 2. Практические кейсы .....	3
<b>Зарубежные страны</b> .....	6
Раздел 1. Зарубежное нормотворчество.....	6
Раздел 2. Зарубежные практические кейсы.....	12
<b>Исследования, экспертные мнения, позиции государственных органов, организаций</b> .....	17
<b>ChatGPT-исследования</b> .....	22

## РОССИЙСКАЯ ФЕДЕРАЦИЯ

### Раздел 1. Акты и проекты актов Правительства

#### О системе «Моя школа»

Минцифры России опубликовало [проект постановления](#) Правительства Российской Федерации «О внесении изменений в постановление Правительства Российской Федерации от 13 июля 2022 г. № 1241».

Предусматривается предоставление посредством системы «Моя школа» государственных и муниципальных услуг (сервисов) в сфере образования.

В реализации проекта предусматривается участие на безвозмездной основе ООО «Компания ВК».

### Раздел 2. Практические кейсы

#### О штрафе за удаление «Википедией» противоправного контента

Мировой судья судебного участка № 422 Таганского района города Москвы вынес постановления о привлечении к административной ответственности компании Wikimedia Foundation, Inc по делам № [05-0517/422/2023](#) и № [05-0519/422/2023](#).

За удаление противоправного контента по требованию Роскомнадзора компании по части 2 статьи 13.41 КоАП РФ назначены штрафы на сумму 800 000 рублей и 2 000 000 рублей соответственно.

#### О Совете цифровых платформ

Компании Яндекс, OZON, Wildberries, HeadHunter, СберМаркет, Avito, Газпром нефть (Профессионалы 4.0), YouDo, объявили о [создании](#) Совета цифровых

платформ при Российском союзе промышленников и предпринимателей (РСПП) и подписали Хартию о принципах развития платформенной занятости в России.

Цель компаний – обмениваться эффективными практиками и создавать благоприятные условия сотрудничества для всех, кто ежедневно предоставляет и использует цифровые сервисы.

В Хартии перечислены основные принципы взаимодействия участников платформенной экономики. Компании готовы информировать исполнителей о социальных гарантиях, обеспечивать для всех исполнителей равные условия доступа на платформы, расширять географию присутствия платформ и в постоянном диалоге создавать другие благоприятные условия для развития сегмента.

Также участники совета готовы вырабатывать дополнительные рекомендации и подходы к регулированию платформенной занятости и приглашают другие компании присоединиться к объединению.

### **Об исключении Spotify из перечня Закона «О приземлении»**

Компания Spotify AB исключена Роскомнадзором из [перечня](#) иностранных лиц, осуществляющих деятельность в сети «Интернет» на территории Российской Федерации.

Основанием для исключения стала посещаемость ресурса, составляющая менее 500 тысяч российских пользователей в сутки.

### **О Стратегии отрасли связи**

Заместитель Председателя Правительства Дмитрий Чернышенко провёл очередную стратегическую [сессию](#) по разработке Стратегии развития отрасли связи до 2035 года.

В ходе сессии Минцифры России было поручено до 1 июня 2023 года организовать общественное обсуждение проекта Стратегии отрасли связи.

Кроме того, Минцифры России, Минфину России и Минэкономразвития России поручено рассмотреть предложения по финансовому обеспечению реализации Стратегии отрасли связи.

Среди основных целей развития стратегии были определены обретение контроля над инфраструктурой связи России, достаточного для обеспечения национального суверенитета и устойчивого развития страны, а также развитие конкурентоспособных, современных и качественных услуг связи для граждан, бизнеса и госорганов.

### **О блокировке стримеров**

Роскомнадзор внёс в реестр запрещённой информации страницы аккаунтов 15 стримеров и блогеров, нарушающих законодательство о рекламе азартных игр, в Twitch, Telegram и Kick.

Нарушителями были признаны:

- Николай Лебедев (n3koglai)
- Егор Крид;
- Юрий Хованский;
- Александр Зубарев;
- Виктор Нигрини (generalqw);
- Олег Горбенко (indian0ch);
- Альберт Васильев (kyivstoner);
- Игорь Лысаковский (rxnexus);
- Ильнур Хафизов (followkudes);
- Арен Зурабян (versuta);
- Игорь Сайганов (AltaOda);
- Кирилл и Александр Волынкины (farbizzbat9);



- Андрей Бурим (mellstroy);
- Гусейн Гасанов.

## Об уголовной ответственности за распространение «фейков» по телефону

Перовский районный суд г. Москвы **вынес** приговор по делу № **01-0042/2023** по обвинению Веделя С.В. в совершении преступления, предусмотренного пунктом «д» части 2 статьи 207.3 Уголовного кодекса Российской Федерации.

Как установлено судом, обвиняемый в ходе телефонных разговоров с коллегами публично распространял заведомо ложные сведения об использовании Вооруженных сил Российской Федерации.

Суд признал подсудимого виновным и назначил ему наказание в виде 7 лет лишения свободы с лишением права занимать должности в системе органов внутренних дел на 4 года.

## ЗАРУБЕЖНЫЕ СТРАНЫ

### Раздел 1. Зарубежное нормотворчество

#### О регулировании криптовалют в ЕС

Европейский парламент **одобрил правила** отслеживания переводов криптоактивов.

Действие «правила движения» (travel rule) распространяется на криптоактивы – информация об источнике актива и его получателе должна будет передаваться вместе с транзакцией.

Под действие регулирования подпадают операции по переводу активов стоимостью свыше 1000 евро с криптокошельков частных лиц при взаимодействии

с кошельками, управляемыми поставщиками услуг в области криптоактивов (криптопровайдерами).

Новые правила не затрагивают переводы между частными лицами без привлечения криптопровайдера или сделки между криптопровайдерами, действующими от собственного лица.

Европарламент также обрил новые единые правила, касающиеся надзора, защиты потребителей и природоохранных мер, в отношении криптоактивов, включая криптовалюты ([Markets in Crypto-assets, MiCA](#)).

В числе прочего предусматривается, что Европейское управление по надзору за рынком ценных бумаг (European Securities and Markets Authority, ESMA) создаст реестр криптопровайдеров-нарушителей, действующих на территории ЕС без соответствующих полномочий.

Правила также требуют от крупных поставщиков услуг раскрывать данные о потребляемой энергии для снижения негативного воздействия на окружающую среду.

## О регулировании конфиденциальности в Интернете в США

В Конгресс США [внесен](#) закон о конфиденциальности в интернете ([Online Privacy Act, OPA](#)), регламентирующий правила обработку данных пользователей Интернета.

Ключевые положения законопроекта включают в себя следующее:

1) Закрепление прав пользователя: законопроект предоставляет каждому гражданину право доступа, исправления или удаления своих данных. Он также создает новые права, такие как «право на непостоянство», которое позволяет пользователям решать, как долго компании могут хранить свои данные.

2) Установление четких ограничений и обязательств для компаний: законопроект сводит к минимуму объем данных, которые компании собирают,

обрабатывают, раскрывают и поддерживают, а также запрещает компаниям использовать данные дискриминационным образом. Кроме того, компании должны будут получить согласие от пользователей, изложенное в понятной форме.

3) Создание Агентства цифровой конфиденциальности (Digital Privacy Agency, DPA): законопроект учреждает независимое агентство, возглавляемое директором, который назначается президентом и утверждается Сенатом на шестилетний срок. DPA будет обеспечивать защиту конфиденциальности и расследовать злоупотребления.

4) Усиление правоприменения: законопроект наделяет генеральных прокуроров штатов полномочиями по обеспечению соблюдения положений законопроекта и позволяет частным лицам назначать некоммерческие организации для представления их интересов в коллективных исках.

5) Установление федерального минимума защиты конфиденциальности: законопроект устанавливает федеральный минимальный уровень защиты конфиденциальности для всех граждан, позволяя штатам усиливать меры защиты или реагировать на изменения в технологиях и государственной политике.

6) Поддержка исследований и разработок в области конфиденциальности: законопроект предписывает Национальному институту стандартов и технологий (National Institute of Standards and Technology, NIST) создать систему управления рисками конфиденциальности и проводить исследования, связанные с уменьшением риска конфиденциальности. Кроме того, NIST будет выделять премии высшим учебным заведениям или некоммерческим организациям для поддержки исследований, связанных с технологиями сохранения конфиденциальности.

## О защите детей в социальных сетях в США

Американские сенаторы внесли на рассмотрение законопроект о защите детей в социальных сетях ([The Protecting Kids on Social Media Act](#)).



Законопроект предусматривает получение согласия на использование детьми социальных сетей у их родителей.

Вводится запрет на использование алгоритмом для предоставления контента несовершеннолетним, а также устанавливается минимальный возраст для использования социальных сетей – 13 лет.

Кроме того, предусматривается запуск пилотной программы по разработке нового возрастного идентификатора для предоставления доступа к платформам, ответственным за которую будет Минторг США.

### **О запрете TikTok в Монтане**

Губернатор штата Монтана, США отложил подписание законопроекта, запрещающего TikTok на всех устройствах в штате, который был одобрен обеими палатами на прошлой неделе.

Отказ подписать законопроект в текущем виде связан с необходимостью, по мнению губернатора, внести изменения в законопроект, расширив запрет, не применяя его конкретно к TikTok, а распространяя его на все приложения социальных сетей, которые передают данные американских граждан недружественным странам.

### **О цифровых рынках, конкуренции и потребителях в Великобритании**

Министерство науки, инноваций и технологий и Министерство бизнеса и торговли Великобритании представили законопроект о цифровых рынках, конкуренции и потребителях ([Digital Markets, Competition and Consumers Bill](#)).

Законопроект:

- создает новый режим для повышения конкуренции на цифровых рынках путем наделения Управления по конкуренции и рынкам полномочиями и обязанностями по регулированию конкуренции на этих рынках;

- обновляет полномочия по расследованию и обеспечению соблюдения законодательства о конкуренции;
- обновляет и расширить полномочия по расследованию и применению законодательства о защите прав потребителей и разрешению потребительских споров;
- предоставляет потребителям защиту в отношении недобросовестных коммерческих практик, таких как «ловушки подписок» и предоплаты.

### О регулировании социальных сетей в Нью-Йорке

В Сенат штата Нью-Йорк, США внесен [законопроект](#) о регулировании деятельности социальных сетей.

Законопроект предусматривает, что компании социальных сетей не могут разрешать несовершеннолетнему жителю Нью-Йорка иметь или регистрировать аккаунт в социальных сетях, если несовершеннолетний не имеет такого права в соответствии с любым другим положением законодательства штата или федерального закона.

Кроме того, законопроект устанавливает, что компании социальных сетей должны проверять возраст владельца существующего или нового аккаунта в штате Нью-Йорк, и, если владелец существующего или нового аккаунта - несовершеннолетний, подтверждать наличие согласия несовершеннолетнего, как это требуется для:

- нового аккаунта, в момент, когда такой житель штата Нью-Йорк регистрирует аккаунт;
- владелец аккаунта в штате Нью-Йорк не предоставил подтверждение возраста, как того требует законопроект, в течение 14 календарных дней после попытки получить доступ к аккаунту.



Если владелец аккаунта не выполнил требования по проверке возраста, компания социальных сетей должна отказать в доступе к аккаунту по истечении требуемого периода времени и до тех пор, пока не будут выполнены все требования по проверке возраста.

Кроме того, компании социальных сетей должны:

- запретить прямые сообщения между аккаунтом и любым другим пользователем, который не связан с аккаунтом через дружбу;
- не показывать аккаунт в результатах поиска любого пользователя, который не связан с аккаунтом через дружбу;
- запрещать показ любой рекламы в аккаунте;
- не собирать и не использовать любую личную информацию из постов, контента, сообщений, текста или действий по использованию аккаунта, кроме информации, необходимой для соблюдения и/или проверки соответствия закону, которая включает в себя такую информацию, как имя родителя или опекуна, дату рождения;
- запретить использование целевых или предлагаемых групп, услуг, продуктов, постов, аккаунтов или пользователей в аккаунте.

Также законопроект предусматривает ограничение доступа для несовершеннолетних жителей штата Нью-Йорк в период с 22:30 до 06:30.

Кроме того, введены механизмы, позволяющие родителям или опекунам получить доступ к учетной записи несовершеннолетнего для просмотра всех сообщений, сделанных в соцсетях, а также всех ответов и сообщений, отправленных владельцу учетной записи несовершеннолетнего.

### **Об использовании данных несовершеннолетних для рекламы**

В Палату представителей штата Северная Каролина, США внесен законопроект об алгоритмическом контроле социальных медиа в ИТ ([НВ 644](#)).



Законопроект устанавливает обязательства для операторов платформ социальных сетей.

Также законопроектом предусмотрены специальные меры защиты для несовершеннолетних пользователей, требуя, чтобы данные пользователя использовались в алгоритмических рекомендациях только в том случае, если пользователь был уведомлен и дал согласие на использование данных таким образом.

Кроме того, законопроект требует, чтобы платформа была полностью функциональной для пользователей без необходимости давать согласие на использование их данных в алгоритмических рекомендациях.

Установлено, что политика конфиденциальности платформы социальных сетей должна быть доступна на веб-сайте и должна раскрывать, как данные пользователя будут использоваться платформой, в краткой и легкой для понимания форме, которая передает истинное намерение использования данных менее чем в 250 словах.

## **Раздел 2. Зарубежные практические кейсы**

### **О блокировке Telegram в Бразилии**

Суд в Бразилии вынес решение об ограничении работы мессенджера Telegram.

Основанием послужило непредоставление запрашиваемых правоохранительными органами данных об администраторах и членах неонацистских групп, которые могут быть причастны к нападению на школу, в ходе которого погибли четыре человека.

Письмо о приостановке работы мессенджера направлено в телефонные компании Vivo, Claro, Tim и Oi, а также Google и Apple, отвечающие за магазины приложений Playstore и App Store.

Также суд повысил размер ежедневного штрафа для Telegram за непредоставление данных со 100 000 реалов до 1 000 000 реалов (~16,1 млн руб.)

## **О запрете слияния Microsoft и Activision Blizzard в Великобритании**

Управление по конкуренции и рынкам (СМА) Великобритании [заблокировало](#) сделку между Microsoft и Activision Blizzard.

Регулятор пришёл к выводу о том, что сделка изменит будущее быстрорастущего рынка облачных игр, что приведет к сокращению инноваций и меньшему выбору для британских геймеров в ближайшие годы.

Microsoft заключила сделку на сумму 68,7 миллиарда долларов по покупке Activision, одного из самых популярных издателей видеоигр в мире, в январе 2022 года.

СМА провела углубленный обзор сделки в сентябре 2022 года, а в феврале 2023 года предварительно установила, что слияние может сделать Microsoft еще сильнее в облачных играх, подавив конкуренцию на этом растущем рынке.

## **О призыве к расследованию в отношении ChatGPT в ЕС**

Европейская организация по защите прав потребителей (BEUC) опубликовала открытое [письмо](#) к Европейской комиссии с просьбой провести расследование в отношении систем генерации текста, в том числе ChatGPT.

В письме отмечается, что генеративные системы искусственного интеллекта предназначены для имитации человеческих речевых паттернов, поведения и эмоций, они по своей сути являются манипулятивными и непредсказуемыми. Они могут создавать убедительный контент, который выглядит правдивым и надежным, но часто фактически неверен.

По мнению BEUC такие системы также могут убедить потребителей предпринять действия, которые они в противном случае не предприняли бы.

Как отмечает организация, использование таким систем создаёт риски введения потребителей в заблуждение в результате создания ложного впечатления, недостоверной рекламы, непредоставлении необходимой информации, а также может способствовать нарушению прав детей.

### **Об авторском праве искусственного интеллекта в США**

Верховный суд США отказался рассматривать жалобу учёного Стивена Талера на отказ Управления по патентам и товарным знакам США выдавать патенты на изобретения, созданные его ИИ-системой DABUS.

Ранее суд нижестоящей инстанции, а также регулятор пришли к выводу о том, что патенты могут быть выданы только изобретателям-людям, а система не может считаться законным создателем двух изобретений.

### **Об иске Epic Games к Apple в США**

Апелляционный суд девятого округа США оставил в силе [решение](#) нижестоящего суда по иску компании Epic Games, Inc. (разработчик игры Fortnite) к компании Apple Inc.

Суд первой инстанции ранее постановил, что Apple не должна запрещать разработчикам включать в их приложения внешние ссылки, специальные кнопки и иные способы оповещения пользователей о других методах оплаты, кроме App Store.

В остальной части требований истца было отказано.

В частности, суд не признал Apple монополистом, однако, указал, что компания прибегает к антиконкурентным действиям.

## О выплате Google компенсации диффамацию в Канаде

Верховный суд канадской провинции Квебек вынес [решение](#) по иску гражданина к компании Google в связи с неудалением из поисковой выдачи ссылок на недостоверные порочащие сведения.

Истец более 10 лет добивался от компании удаления ссылок на ресурсы, содержащие ложные обвинения в педофилии, в результате которых истец лишился карьеры.

Суд обязал Google выплатить 500 000 долларов США в качестве компенсации морального вреда, указав, что, отказывая истцу в удалении ссылок на порочащие его материалы, компания руководствовалась ложной интерпретацией законов Канады.

## О проверке ChatGPT в ФРГ на соответствие требованиям GDPR

Регуляторами в ФРГ инициировано расследование в отношении практики конфиденциальности OpenAI и соответствия Общему регламенту ЕС о защите данных (GDPR).

Цель проверки – выявить, проводила ли OpenAI оценку воздействия на защиту данных и находятся ли риски защиты данных под контролем.

## О коллективном иске к брокеру данных Kochava

Брокер данных – владелец мобильного приложения Kochava подал окружной суд штата Айдахо ходатайство об отклонении коллективного иска, поданного пользователями смартфонов из штатов Калифорния и Вашингтон, которые утверждают, что компания продала их геолокационные данные.

Компания заявила, что пользователи не могут утверждать, что им был нанесен ущерб в результате предполагаемой продажи, и даже если бы они могли это сделать, компания утверждает, что не нарушила закон штата.

В иске утверждается, что Kochava продает точные геолокационные данные пользователей вместе с идентификаторами мобильной рекламы, которые могут быть использованы для отслеживания потребителей в определенных местах.

## **О санкциях против поставщиков облачных услуг в США**

Сенаторы США направили [письмо](#), призывающее правительство ввести санкции против технологических компаний.

Письмо направлено министру финансов, министру торговли и государственному секретарю с призывом ввести санкции против Huawei Cloud и Alibaba Cloud.

В частности, письмо призывает правительство использовать доступные инструменты для дальнейшего расследования и принятия решительных мер в отношении этих поставщиков облачных услуг.

## **О перечне очень крупных онлайн-платформ и поисковых систем в рамках DSA**

Европейская комиссия опубликовала [перечень сервисов](#), которые квалифицируются в соответствии с Законом о цифровых услугах (DSA) как очень крупные онлайн-платформы (VLOPs) и очень крупные онлайн-поисковики (VLOSEs).

В перечень вошли: Alibaba AliExpress; Amazon Store; Apple AppStore; Booking.com; Facebook; Google Play; Google Maps; Google Shopping; Instagram; LinkedIn; Pinterest; Snapchat; TikTok; Twitter; Wikipedia; YouTube; Zalando; Bing; Google Search.

Теперь у компаний будет четыре месяца на выполнение обязательств по DSA. Помимо прочего, указанные организации должны:

- гарантировать пользователям право отказаться от использования



рекомендательных систем, основанных на профилировании;

- воздерживаться от показа рекламы, основанной на конфиденциальных данных пользователей, таких как этническое происхождение или политические взгляды;
- перепроектировать свои системы для обеспечения высокого уровня конфиденциальности, безопасности и безопасности несовершеннолетних.

## ИССЛЕДОВАНИЯ, ЭКСПЕРТНЫЕ МНЕНИЯ, ПОЗИЦИИ ГОСУДАРСТВЕННЫХ ОРГАНОВ, ОРГАНИЗАЦИЙ

### Об отношении россиян к ИИ

Компания Rambler&Co опубликовала результаты [опроса](#) 150 тысяч читателей об их отношении к применению ML-моделей для создания контента.

Больше двух третей (68%) россиян считают, что с помощью ИИ удастся полностью исключить орфографические и пунктуационные ошибки в текстах.

Каждый десятый (10%) уверен, что со временем нейросети смогут быстрее сообщать аудитории экстренные новости.

Столько же (10%) рады за журналистов, у которых будет дополнительное время на расследования и репортажи вместо написания рутинных заметок. 9% видят преимущества ИИ в улучшении визуализации материалов, 3% – в работе персональных рекомендаций.

Четверть россиян (25%) доверили бы языковым моделям писать заголовки вместо авторов, 23% – объяснительные материалы и тематические карточки с инструкциями, 16% – новости спорта. Каждый десятый (10%) считает, что ИИ хорошо справится с постами для соцсетей, 8% – с оперативными новостями.

Респонденты также поделились своими опасениями насчёт быстрого развития нейросетей в медиа. Так, 36% участников опроса переживают, что ИИ легко

взломать, а 27% считают, что внедрение «цифровых помощников» приведет к потере журналистских навыков.

21% уверены, что с нейросетями все тексты станут одинаковыми, а потому неинтересными, 16% допускают, что применение языковых моделей может существенно снизить уровень грамотности авторов.

Отвечая на вопрос о дальнейшем развитии ИИ, 38% респондентов считают его необходимым и закономерным процессом, но с ограничениями. 22% настроены крайне негативно и предлагают запретить использование языковых моделей в журналистике, в то время как 21%, напротив, видят в них будущее СМИ. 19% остановились бы в той точке развития, которая есть сейчас, чтобы глубже изучить технологию.

## О развитии технологий ИИ

Роскомнадзор и Главный радиочастотный центр совместно с аналитическим центром MINDSMITH и компанией «Ростелеком» [представили исследование «Инструменты ИИ в руках злоумышленников – классификация угроз и способы противодействия»](#).

В исследовании определены ключевые технологии с использованием искусственного интеллекта, задействованные в мониторинге Интернета и обеспечении безопасности его пользователей.

На основе анализа научных публикаций, патентов и инвестиционных проектов, а также обсуждения с экспертами выделены 12 групп технологий, соответствующих интересам профильных организаций по мониторингу и обеспечению безопасности интернет-среды:

- обнаружение дипфейков,
- определение контекста происходящего на видео,
- автоматизация мониторинга и модерации контента,

- распознавание лиц,
- извлечение смысла из текста,
- поддержка проверки фактов,
- распознавание символики,
- извлечение и анализ метаданных,
- распознавание эмоций,
- поддержка решений при информационных атаках,
- генерация контента,
- рекомендация контента.

Эксперты пришли к следующим выводам:

1) системы приближаются к проверке фактов в реальном времени и будут интегрированы в соцсети;

2) на данный момент полностью положиться на ИИ в проверке фактов нельзя, можно лишь автоматизировать часть процессов;

3) системы для фактчекинга находятся на раннем этапе развития и будут массово внедрены только через 3–5 лет;

4) внедрение автоматизированной проверки фактов приведет к тому, что каждая новость будет размечена специальными тегами на крупных платформах.

## О пользователях игровых сервисов

Ресурс GameDiscoverCo опубликовал [данные](#) о пользователях игровых сервисов PlayStation, Xbox, и Steam.

В топ-5 стран по процентному соотношению пользователей PlayStation входят:

- 1) США – 37,2%;
- 2) Великобритания – 7,6%;
- 3) Испания – 7,2%;
- 4) Франция – 6,4%;



5) Германия – 5,1%.

Россия занимает 13 место (2,1% пользователей).

Топ-5 стран по количеству пользователей платформы Xbox выглядит следующим образом:

1) США – 40,6%;

2) Великобритания – 7,9%;

3) Бразилия – 6,9%;

4) Франция – 5,8%;

5) Германия – 5,0%.

Россия занимает 8 место (3,7% пользователей).

В топ-5 стран по пользователям сервиса Steam (предназначенного для персональных компьютеров), по результатам оценки 25 млн аккаунтов вошли:

1) США – 13,7%;

2) Китай – 12,3%;

3) Россия – 9,7%;

4) Бразилия – 4,9%;

5) Германия – 3,6%.

## **О влиянии ИИ на производительность труда**

Учёные из Стэнфордского университета и Массачусетского технологического института представили [исследование](#) воздействия генеративного искусственного интеллекта на живую рабочую силу.

Эксперты пришли к выводу, что самые слабые сотрудники могут извлечь выгоду из внедрения ИИ, который не даёт ощутимого преимущества сильным, а наоборот, выравнивает поле.

Исследователи изучали работу 5 179 специалистов службы поддержки компании из списка Fortune 500, занимающейся разработкой ПО для предприятий.

Все они должны были знать продукт и иметь опыт работы с ним, обладать навыками решения проблем и способностью вести диалог с разозлёнными клиентами.

Некоторым из испытуемых был предоставлен доступ к последней версии одной из больших языковых моделей Open AI. Учёные оценивали производительность работников исходя из того, насколько быстро они решали проблемы клиентов, и сколько обращений обрабатывали в течение часа.

До внедрения ИИ самые эффективные работники компании находили решения в два раза быстрее, чем их коллеги.

С началом эксперимента у тех, кто использовал AI-помощника, производительность выросла на 14% по сравнению с теми, кто работал самостоятельно.

При этом самые неопытные и наименее квалифицированные участники эксперимента получили наибольшую выгоду от AI-инструментов и смогли выполнять свою работу на 35% быстрее.

Отработавшие два месяца новички, получив в напарники искусственный интеллект, сравнивали свои показатели с теми, кто работал шесть месяцев, но не использовал ИИ-помощника.

### **Об использовании дипфейков в мошеннических целях**

По данным [исследования](#), проведенного компанией Regula, 37% организаций столкнулись с мошенничеством с использованием синтезированного голоса, а 29% стали жертвами поддельных видео.

80% компаний воспринимают поддельные биометрические данные, такие как глубоко подделанный голос или видео, как реальную угрозу, а 91% американских компаний считают ее растущей угрозой.

Regula утверждает, что для борьбы с этой проблемой организациям следует внедрять тщательную проверку личности и биометрические проверки. Эти проверки

включают в себя расширенную проверку документов, перекрестную проверку информации и иных атрибутов субъекта данных.

В исследовании также подчеркивается сложность мер по борьбе с мошенничеством и проверке личности, которые должны принимать предприятия.

26% малых и 38% крупных компаний в 2022 году пережили более 50 атак, связанных с мошенничеством с использованием личных данных, с убытками до 480.000 долларов США.

## ChatGPT-исследования

*Материал подготовлен ChatGPT  
и не является позицией АНО «ИРИ»*

### Текстовые дипфейки

#### *Что такое текстовые дипфейки?*

Текстовые дипфейки (deepfake text или синтетический текст) представляют собой текст, который был сгенерирован алгоритмами искусственного интеллекта так, чтобы выглядеть так, как если бы он был написан человеком.

Эта технология использует методы глубокого обучения (deep learning) для анализа и воспроизведения паттернов человеческого письма, таких как стиль, тон и структура.

Полученный текст можно использовать для создания поддельных новостных статей, постов в социальных сетях, электронных писем и других типов цифрового контента, авторами которого кажутся реальные люди.

Текстовые дипфейки также могут использоваться для выдачи себя за физических лиц и создания поддельных документов, таких как банковские выписки или юридические контракты.

Текстовые дипфейки становятся все более изощренными, из-за чего людям становится трудно отличить подлинный текст от синтетического. Это вызывает беспокойство по поводу потенциального воздействия текстовых подделок на общество, включая их использование для распространения дезинформации, совершения мошенничества и манипулирования общественным мнением.

В результате растет интерес к разработке методов обнаружения и предотвращения использования текстовых дипфейков, а также к пропаганде ответственного и этичного использования искусственного интеллекта при генерации текстов.

### ***С какими рисками сопряжено использование текстовых дипфейков?***

Текстовые дипфейки могут представлять ряд рисков, в том числе:

1. *Дезинформация.* Текстовые фейки могут использоваться для создания и распространения ложной информации, которая может иметь значительные последствия в реальном мире.

2. *Мошенничество.* Текстовые подделки могут использоваться для того, чтобы выдавать себя за физических лиц и создавать поддельные документы, такие как банковские выписки, юридические контракты или даже посты в социальных сетях, с целью получения финансовой выгоды.

3. *Кибербезопасность.* Текстовые дипфейки могут использоваться для создания фишинговых электронных писем или других видов кибератак, в которых используются убедительные формулировки, чтобы обманом заставить людей раскрыть конфиденциальную информацию.

4. *Ущерб репутации.* Текстовые подделки могут быть использованы для нанесения ущерба репутации отдельных лиц или организаций путем распространения ложной или порочащей информации.

5. *Политическая манипуляция.* Текстовые дипфейки могут использоваться для манипулирования общественным мнением и влияния на выборы путем создания

поддельных новостных статей или постов в социальных сетях, поддерживающих определенную политическую повестку дня.

6. *Юридические последствия.* Текстовые подделки могут привести к юридическим последствиям, таким как диффамация, нарушение авторских прав или даже уголовные обвинения, если они используются в мошеннических целях.

7. *Потеря доверия.* Распространение текстовых подделок может подрвать доверие общественности к онлайн-контенту, что приводит к общему чувству подозрительности и недоверия.

В целом, текстовые дипфейки представляют значительный риск для отдельных лиц, организаций и общества в целом, поскольку способствуют распространению ложной информации, мошенничеству и другим неэтичным практикам.

Важно разработать стратегии по выявлению и предотвращению их использования, а также поощрять ответственное и этическое использование текста, сгенерированного с помощью искусственного интеллекта.

### ***Какие меры позволят снизить риски использования текстовых дипфейков?***

Вот некоторые меры по снижению рисков текстовых дипфейков:

1. *Образование и осведомленность:* информирование отдельных лиц и организации о рисках и последствиях текстовых дипфейков, обучение тому, как распознавать их и реагировать на них.

2. *Верификация и проверка фактов:* внедрите системы верификации и проверки фактов, которые могут обнаруживать и предотвращать распространение текстовых дипфейков. Эти системы должны быть способны проверять подлинность источника и содержания информации.

3. *Цифровая аутентификация:* использование технологии цифровой аутентификации для проверки подлинности онлайн-контента. Эти технологии могут



включать цифровые подписи или водяные знаки.

4. *Анализ текста*: разработка инструментов и технологий, которые могут анализировать язык и структуру текста для выявления потенциальных дипфейков. Эти инструменты должны быть способны обнаруживать аномалии в языке и стиле написания.

5. *Законодательство и нормативные акты*: разработка законодательства, которое будет регулировать создание и распространение поддельных текстов. Оно должно предусматривать штрафы за неправильное использование дипфейков.

6. *Сотрудничество*: содействие сотрудничеству между технологическими компаниями, академическими институтами и правительственными учреждениями в разработке и внедрении решений для обнаружения и предотвращения текстовых дипфейков.

7. *Этические рекомендации*: разработка этических рекомендаций по созданию и использованию дипфейков. Эти руководящие принципы должны касаться таких вопросов, как согласие, точность, указание авторства и конфиденциальность.

8. *Медиаграмотность*: пропаганда медиаграмотности и навыков критического мышления среди людей, чтобы помочь им идентифицировать сгенерированный текст и реагировать на него.

В целом, снижение рисков текстовых дипфейков требует многостороннего подхода, который включает образование, технологии, регулирование и сотрудничество. Реализуя эти меры, мы можем уменьшить влияние текстовых дипфейков на отдельных лиц, организации и общество в целом.

### ***Примеры текстовых дипфейков.***

Вот несколько более конкретных примеров текстовых дипфейков.

*Новостные статьи, генерируемые GPT-3*: GPT-3 – это мощная языковая модель, которая может генерировать текст, похожий на человеческий. Исследователи продемонстрировали, как GPT-3 может быть использован для

создания убедительных новостных статей, содержащих ложную информацию. Например, исследовательская группа под названием OpenAI создала поддельную новостную статью о ядерной катастрофе, которая смогла обмануть некоторых читателей, заставив их думать, что это реально.

*Фишинговые электронные письма, генерируемые искусственным интеллектом:* инструменты на базе искусственного интеллекта можно использовать для создания фишинговых электронных писем, имитирующих стиль письма надежного отправителя, такого как банк или платформа социальных сетей. Эти электронные письма могут быть адаптированы к конкретным получателям, что делает их более убедительными. Например, фишинговое электронное письмо, сгенерированное искусственным интеллектом, может быть адресовано получателю по имени и содержать сведения об его учетной записи, чтобы оно выглядело более легитимным.

*Поддельные отзывы о ресторанах:* Исследователи показали, как сгенерированные искусственным интеллектом отзывы можно использовать для манипулирования онлайн-рейтингами ресторанов или других предприятий. Например, инструмент под названием GROVER может генерировать убедительные отзывы о ресторанах, которые трудно отличить от реальных отзывов, написанных людьми.

*Дипфейки общественных деятелей:* В то время как большинство людей ассоциируют дипфейки с манипулируемыми изображениями или видео, также возможно создавать текстовые дипфейки, которые выдают себя за общественных деятелей. Например, исследователи использовали GPT-2 для генерации твитов, имитирующих стиль письма Дональда Трампа, которые могут быть использованы для распространения ложной информации или создания путаницы в социальных сетях.

*Научные статьи, созданные с помощью искусственного интеллекта:*  
Исследователи продемонстрировали, как искусственный интеллект может быть использован для создания убедительных научных работ, которые являются плагиатом из существующих работ. Например, группа исследователей создала инструмент под названием SCIGen, который мог генерировать статьи по информатике, которые были приняты на конференциях и в журналах несмотря на то, что содержали бессмысленный текст.

### ***Примеры реализации рисков использования текстовых дипфейков.***

Вот несколько реальных примеров рисков, связанных с использованием текстовых дипфейков.

*Дезинформация:* Текстовые дипфейки могут быть использованы для создания убедительных новостных статей, постов в социальных сетях и другого письменного контента, распространяющего ложную информацию или дезинформацию по существу. В 2020 году группа исследователей создала текстогенерирующую модель искусственного интеллекта, которая смогла писать убедительные фейковые новостные статьи о пандемии COVID-19. Исследователи использовали эту модель, чтобы подчеркнуть потенциальные риски использования текста, сгенерированного искусственным интеллектом, для распространения ложной информации. (Источник: <https://www.technologyreview.com/2020/06/04/1002474/ai-fake-text-gpt-2-dangerous/> )

В 2020 году на PRNewswire был опубликован ложный пресс-релиз, в котором утверждалось, что Walmart начнет принимать Litecoin в качестве формы оплаты. Пресс-релиз был текстовой подделкой, и хотя он был быстро опровергнут, это вызвало краткий скачок цены лайткоина. (Источник: <https://www.bbc.com/news/technology-58323858>)

*Фишинговые мошенничества:* Текстовые подделки могут использоваться в фишинговых атаках или для распространения вредоносного ПО. В 2020 году текстовая подделка была использована в фишинг-атаке на британскую

энергетическую фирму. Злоумышленник использовал текстовую подделку, чтобы выдать себя за генерального директора компании и убедить сотрудника перевести деньги на мошеннический счет. (Источник: <https://www.bbc.com/news/technology-53432590>)

*Киберзапугивание:* Текстовые подделки могут использоваться для создания поддельных постов в социальных сетях или сообщений, которые могут быть использованы для киберзапугивания или домогательств. В 2019 году группа исследователей создала текстогенерирующую модель искусственного интеллекта, которая смогла генерировать поддельные обзоры Yelp, которые было трудно отличить от реальных отзывов. Исследователи отметили, что этот тип технологии может быть использован для киберзапугивания или для нанесения ущерба репутации бизнеса. (Источник: <https://www.businessinsider.com/ai-is-being-used-to-fake-reviews-on-yelp-and-amazon-2019-10>)

Эти примеры иллюстрируют некоторые риски, связанные с использованием текстовых подделок, включая возможность распространения дезинформации, совершения мошенничества и участия в киберзапугивании или домогательствах.

### ***Комментарий АНО «ИРИ»:***

*Представленная GPT информация о текстовых дипфейках в целом отвечает действительности и информирует пользователей разной степени осведомленности о существующей проблематике.*

*Риски, связанные с использованием дипфейков, определены корректно и отвечают действительности.*

*Интересны примеры из реальной жизни, которые сервис определил как текстовые подделки.*

*Поскольку ChatGPT генерирует текст без ссылок на используемые источники, у сервиса было запрошено, откуда была взята информация о реальных*

*случаях текстовых дипфейков. Некоторые из ссылок оказались недействительными и найти первоисточник не удалось.*

*Так, из всех представленных сервисом ссылок работают только ссылки на статьи Гровера и Вашингтонского университета. Остальные ссылки некорректны, хотя приведенные сервисом примеры действительно имели место быть и проверяются другими достоверными источниками.*

*Однако имеются некоторые ошибки в ряде фактов, а именно использование текстового дипфейка в фишинговой атаке на британскую энергетическую фирму в 2020 году. Во-первых, такая атака произошла в 2019 году. Во-вторых, использовался не текстовый дипфейк, а подделанный голос.*

*Ложный пресс-релиз об принятии Litcoin сетью Walmart был опубликован не в 2020 году, а в 2021.*





Институт развития  
интернета

[iri.rf](http://iri.rf)