ИТ

Тематическое приложение к газете **Коммерсантъ**

Информационные технологии

Среда 31 марта 2021 №55 (7017 с момента возобновления издания) **kommersant.ru**

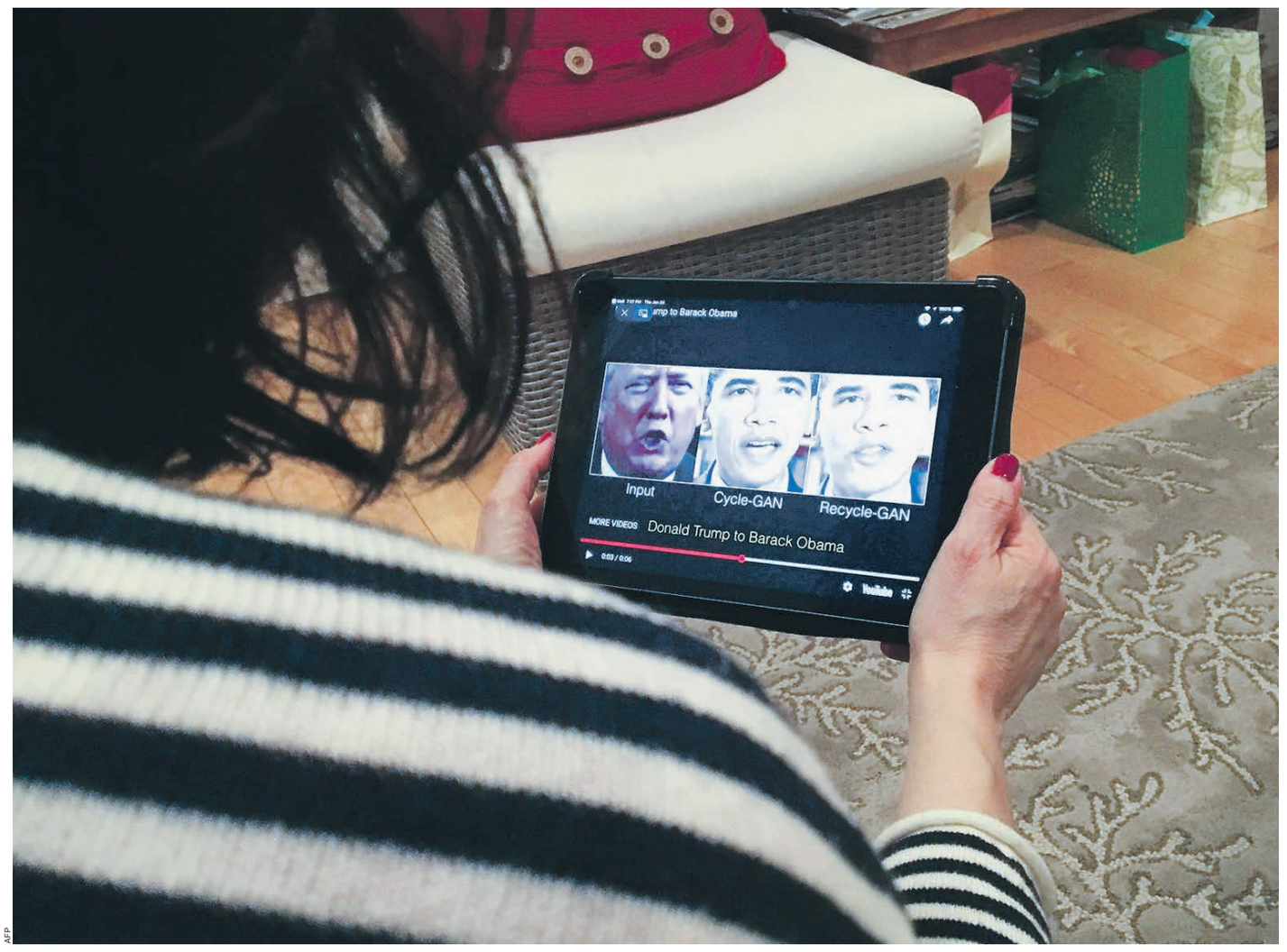
17 | Почему видеоигры становятся главной технологической платформой

18 | К чему уже привел рост популярности цифрового искусства

18 | Как смена бизнес-модели делала стартапы мировыми гигантами

Почувствуйте разницу

Распространение deepfake-видео в социальных сетях растет лавинообразно. Если совсем недавно можно было наблюдать несколько вирусных роликов-поддделок за год, то сегодня они появляются чуть ли не ежедневно. Один из последних примеров — фейковый, то есть поддельный, Том Круз, показывающий фокусы и играющий в гольф в социальной сети TikTok. Пользователи не сразу раскусили подлог, несмотря на то что название аккаунта — deeptomcruise — прямо указывало на шутку. Дипфейки пришли в рекламу, снимаются в кино и устраивают политические баттлы в соцсетях. Однако из розыгрышей и развлечений deepfake могут перерасти в реальную проблему, предупреждают эксперты.



AFP

Сгенерированные с помощью нейросетей видео не оставляют другого выбора, кроме как критически относиться к информации в интернете

— феномен —

Слово «deepfake» родилось из двух понятий: «deep learning» — «глубинное обучение нейросетей» — и «fake» — «фейк, подделка»). Deepfake-видео создаются при помощи переноса лица и мимики реальной персоны в новое видео с актером. Например, берется ряд видео с настоящим Бараком Обамой, на котором обучается нейросеть, и снимается новое видео с актером, на которого новое лицо переносится как живая маска. Подробные инструкции о том, как это сделать, есть на ресурсах для разработчиков, таких как GitHub.

Полученное видео можно редактировать и подтягивать к более реалистичному результату, но на сегодняшний день, как говорят эксперты, результат все еще далек от идеала. Фейки выдают себя недостаточно проработанной мимикой, задержками, артефактами. Однако технологии развиваются, а значит, через некоторое время, весьма короткое, видео на основе GAN (Generative Adversarial Network — генеративно-состязательные нейросети) станут куда более правдоподобными. Отличить подделку от оригинала будет под силу разве что с помощью специальных цифровых инструментов, но не «на глаз».

Дипфейки среди нас

«Deepfake — понятие массовой культуры, за которым стоит комплекс сложных алгоритмов, основанных на GANs. Генеративные нейронки способны решать множество интересных задач: от создания виртуальных ассистентов до реконструкции снимков далеких галактик. Очевидно, что позитивный контекст применения у них впечатляющий»,— объясняет CEO стартапа Deepcake Мария Чмир. При этом риски, к сожалению, соизмеримы с пользой, продолжает она: с повышением качества генерации — а этот процесс неизбежен и заложен в методологию GAN — вероятность заполнения медиа манипулятивным контентом будет неизбежно расти.

Российский стартап Deepcake помогал «Сберу» в создании нашумевшего ролика с Жоржем Милославским. Команда генерировала образ для видео, а голос героя создавался в собственной «Лаборатории AI» «Сбера». Сейчас стартап создает образовательный контент для ряда крупных заказчиков и работает над проектами для кинорынка. «Три ключевых кейса применимости дипфейков в рекламе и кино: производственная необходимость — актер заболел, у него параллельный проект, границы закрыты, и выехать на съемки он или она не могут — сейчас по понятным причинам таких запросов у нас много; желание скорректировать физические параметры — омолодить или состарить героя; „приглашение“ в проект персонажа, который не доступен для съемок — наши алгоритмы отлично справились с этой задачей: в проекте „Сбера: мы сымитировали образ Жоржа Милославского поверх лица дублера для рекламной кампании»,— рассказывает Мария Чмир.

Еще один свежий пример, получивший вирусное распространение в соцсетях,— сервис Deep Nostalgia в приложении MyHeritage. Нейросеть Deep Nostalgia «оживляет» фото людей — пользователи массово стали использовать ее, чтобы посмотреть на своих родственников со старинных фотографий в движении. Завораживающее зрелище — ожившие фото и портреты, как из фильмов про Гарри Поттера. Не удивительно, что людям захотелось испытать этот эффект на родных лицах.

Впрочем, это не первый подобный прием: осенью прошлого года выходцы из России Александр Гунин и Леонид Шейба выпустили приложение Puppets.World, которое анимировало не только фото, но и картины. Вирусным стал ролик с двигающейся Моной Лизой, затем к ней присоединились звезды и культовые персонажи. Приложение также быстро набрало популярность за счет вирального эффекта в соцсетях.

Быстро завоевало мировую популярность и приложение Reface от украинских разработчиков. Оно, например, позволяет переносить собственное лицо в видео или GIF со знаменитостями. Пользователям понравилась возможность посмотреть на себя в клипах Шакиры, Бейонсе и прочих звезд. В основе — все так же технологии GAN.

Упоминавшийся уже фейк Тома Круза для серии роликов в TikTok создал американский разработчик Крис Уме совместно с актером Майлзом Фишером. Сам Уме в одном из интервью говорил, что старается делать забавный контент, но также хочет донести до людей понимание, как развиваются технологии и почему стоит сомневаться и проверять источники видео. На момент создания этого материала аккаунт deeptomcruise был переведен в приватный режим, то есть ролики не доступны другим пользователям.

Множество дипфейков и в TikTok: от безобидных пародий до роликов с участием известных оппозиционеров и представителей действующей власти. При этом в соцсети нет правил, касающихся непосредственно дипфейков — TikTok использует единые политики относительно любого видеоконтента. Блокироваться и удаляться могут видео, нарушающие правила сообщества: те, где пользователи сознательно выдают себя за других личностей без какого-либо указания на это, а также те, где используется вводящий в заблуждение контент, нарушающий правила сообщества. Аккаунт фейкового Тома Круза, как и другие подобные аккаунты, соцсеть расценивает как пародийные, не блокирует их и не применяет каких-либо санкций.

Остерегайтесь подделок

В «Сбере», который первым из крупных российских брендов использовал дипфейк в рекламе, к самой теме относятся двояко. «Развитие технологий синтеза изображений с использованием GAN открывает обширные перспективы для целого ряда индустрий, в первую очередь работающих с изображениями: медиа и рекламы, индустрии игр и развлечений, виртуальных ассистентов, киноиндустрии и т. д. Но существуют понятные риски и уже имеющие место случаи, когда такие технологии, широко известные как deepfake, наносили ущерб репутации голливудских звезд или политиков. Один из самых ярких примеров — deepfake с 44-м президентом США Бараком Обамой, в котором он якобы оскорбляет 45-го президента Дональда Трампа»,— напоминает первый зампред правления Сбербанка Александр Ведяхин.

С этим согласен руководитель направления перспективных технологий Microsoft в России Владислав Шершульский: «Могут ли у технологий создания квазиреалистичных образов быть благие цели? Безусловно. О рекламе и даже художественных фильмах с участием знаменитых артистов прошлого сказано уже много. А разве каждый студент не мечтал бы лично присутствовать на лекциях авторов любимых классических учебников? Теперь это возможно»,— говорит он. При этом, отмечает эксперт, интернет активно пополняется роликами, снятыми без согласия правообладателей и эксплуатирующими неразборчивость зрителя. «Как будут развиваться дела в этой области? Не думаю, что технологии добавят много нового в степень реалистичности дипфейков — уже сегодня человеку довольно трудно заметить признаки манипуляции. Дальнейший прогресс в технологиях дипфейков будет направлен на то, чтобы обманывать технические средства их выявления. Это противостояние будет продолжаться»,— указывает он.

Как всегда, чем эффективнее технология, чем более впечатляющую пользу она может принести, тем опаснее ее использование в неэтичных целях. Искусственный интеллект и машинное обучение не стали исключением. На одной чаше весов — повышение качества медицинского обслуживания, создание более безопасной жизненной среды, избавление от тяжелого и опасного труда, новые развлечения. На другой — перекладывание ответственности на компьютеры, алгоритмическая дискриминация и, конечно, дипфейки.

Различные манипуляции с изображениями существовали еще до появления компьютеров, указывает эксперт по кибербезопасности в «Лаборатории Касперского» Дмитрий Галов. С развитием технологий этот процесс стал относительно проще и доступнее. Не удивительно, что на него обратили внимание и злоумышленники. Эксперт напоминает, как в 2019 году преступники использовали голосовой deepfake: управляющий директор британской энергетической компании был ограблен на €220 тыс. (около $240 тыс.). Он отправил эти деньги фирме-поставщику из Венгрии, потому что его босс, глава материнской фирмы в Германии, несколько раз подтвердил ему эту инструкцию, писал The Wall Street Journal. «При этом важно понимать, что сами по себе технологии deepfake вреда не несут, большую роль играет то, с какой целью ими пользуются люди»,— говорит господин Галов.

Взять под контроль

По словам Александра Ведяхина, вопрос о том, кто должен контролировать использование Deepfake, остается открытым. «Государства уже начинают искать подходы к их регулированию. Например, в США в штате Калифорния уже принят закон о запрете распространения политических deepfake, связанных с выборами президента. В 2019 году в России была принята Национальная стратегия развития технологий ИИ, а в прошлом году — Концепция развития регулирования отношений в сфере технологий ИИ и робототехники. Оба документа закладывают ориентиры в развитии регулирования для правомерного использования ИИ, в том числе для регулирования deepfake,— указывает господин Ведяхин.— Точно можно сказать, что регулирование deepfake может и должно базироваться на этических и правовых принципах использования технологий ИИ. У себя в „Сбере“ мы уже сделали первые шаги в этом направлении и еще в конце 2020 года приняли документ об этических принципах ИИ». Документ опубликован 1 марта, его можно найти в открытом доступе.

В феврале Microsoft вместе с Adobe, Arm, BBC, Intel и Truepic создали «Коалицию за достоверность и аутентичность контента» (Coalition for Content Provenance and Authenticity). Она будет разрабатывать открытые стандарты и инструменты, которые, как надеются в компаниях, позволят каждому убеждаться в качестве получаемой информации и существенно затруднят ее искажение, рассказывает Владислав Шершульский.

По его словам, вопрос с регулированием дипфейков укладывается в общие вопросы ответственных подходов к использованию технологий искусственного интеллекта, к которым относится обучение нейросетей. «Добиться ответственного использования ИИ нельзя в одиночку. Это задача и компаний-разработчиков, и потребителей, и государства, и общества,— продолжает он.— Мы в Microsoft создали специальный внутренний комитет AETHER по этике ИИ, который может прекратить любой, даже самый выгодный проект, если есть риск нежелательных побочных эффектов; сформулировали этические принципы ответственного обращения с ИИ; разработали политики, регламенты, процедуры и инструменты, позволяющие нам заранее выявлять и избегать большинства нежелательных ситуаций».

Поскольку одного универсального способа борьбы с deepfake быть не может, юридические и инженерные команды активно работают над множеством подходов и инструментов. Так, в сентябре 2020 года Microsoft представила Video Authenticator — программу, оценивающую шансы того, что фото или видео подверглось манипуляциям. Также Microsoft вместе с Facebook, Partnership on AI и учеными из Cornell Tech, MIT, Оксфордского университета, Калифорнийского университета в Беркли, Университета Мэриленда, Колледж-парка и Университета в Олбани-Суни участвовали в инициативе Deepfake Detection Challenge, направленной на совместную разработку ПО с открытым кодом для обнаружения deepfake.

Нельзя не учитывать, что если дипфейк создается командой злоумышленников, они могут натренировать свои инструменты обманывать подобные программы. Поэтому, как в истории с вирусами и антивирусами, появляется новый рынок антидипфейков, которые должны дорабатываться и развиваться. Пока рынок софта, призванного распознавать дипфейки и бороться с ними, сложно оценить в деньгах — как правило, этот блок входит в общие затраты компаний на машинное обучение. Однако эксперты ResearchAndMarkets предрекают этому рынку быстрое развитие в ближайшие пять лет.

Не только видео

Говоря о создании фейковых цифровых персон, нельзя забывать о том, что при помощи нейросетей сегодня генерируется не только видео, но и голос. Пример выше от «Лаборатории Касперского» показывает, что использование голосовых фейков может привести к серьезным мошенничествам. Однако возможности применения технологий синтеза голоса гораздо шире. И они быстро стали очевидны в аудиосоцсети ClubHouse.

Нашествие пародистов и фейковых аккаунтов в ClubHouse началось буквально сразу, как только соцсеть стала привлекать внимание российских пользователей. Было ли это проблемой раньше — неизвестно. Зато пользователи из российского сегмента интернета уже повстречали в голосовой соцсети «Аллу Пугачеву» и других фейковых артистов. А недавно в одну из популярных «комнат» заходил Александр Лукашенко. Синтезированный, конечно.

Как рассказал „Ъ“ основатель и гендиректор стартапа Parodist Владимир Свешников, это был подготовленный эксперимент. Разработчики заранее заготовили несколько фраз «президента Белоруссии». «Когда нам дали слово, мы эти фразы успешно проиграли. Фразы были шутливые, людям очень понравилось, многие смеялись»,— рассказал представитель Parodist. На момент появления фейка в «комнате» было около 2 тыс. слушателей. В теории, имея в распоряжении подобную технологию, можно сделать и что-то более серьезное, рассуждает Владимир Свешников: «Есть огромное количество кейсов, где мы коммуницируем голосом. Звонки, радио, голосовые сообщения, ClubHouse, подкасты. Во всех этих каналах может быть использована данная технология».

Существует немало вариантов использования синтеза голоса во благо. Наряду с пародией технология может быть полезна для кинематографа, озвучивания аудиокниг, редактирования подкастов, подчеркивает господин Свешников: «Думаю, в ближайшие два-пять лет большинство стран мира так или иначе начнут регулировать эту область. Анализируя инициативы, которые сейчас есть в США, отмечу, что логика законодателя движется в сторону закрепления ответственности за введение людей в заблуждение с помощью этой технологии, но сохраняется возможность ее использования в художественных и пародийных произведениях, так как полный запрет будет противоречить принципам свободы слова».

Сколько стоит дипфейк

Почему deepfake бурно распространяется и активно обсуждается именно сейчас? Студии компьютерной графики уже не первый десяток лет умеют создавать искусственную картинку, которую трудно отличить от оригинала. Ответ на поверхности: раньше это было долго и дорого. Автоматизация процесса делает создание дипфейков быстрым и доступным.

«До недавнего времени подобное редактирование всегда подразумевало достаточно большой уровень компетенций. Сейчас, чтобы сделать, например, видео с deepfake, не требуется высокого уровня владения инструментом редактуры, нужно лишь немного разбираться в программировании, и то не всегда,— рассуждает Дмитрий Галов.— С другой стороны, deepfake требуют большого количества данных. Иными словами, чтобы изменить лицо человека на видео, нужно много фотографий или видео этого человека. Нельзя сказать, что это по-настоящему массовая угроза. Однако технологии развиваются и отличить deepfake действительно становится все труднее».

На YouTube есть видео, демонстрирующее возможности Deepfake, оно набрало почти 1,3 млн просмотров. На видео пародист Джим Мескимен «примеряет» лица 20 знаменитостей: Джорджа Клуни, Арнольда Шварценеггера и других. Создатель видео, известный под ником Sham00k, рассказал, что на производство ушло более 250 часов. Для создания дипфейков было отснято 1,2 тыс. часов материала и обработано 300 тыс. изображений, в процессе создано около терабайта данных. То есть работа эта кропотливая и под силу далеко не каждому.

Но есть и более простые задачи: примеры с приложениями выше показывают, насколько быстро технология становится массово-доступной.

Сегодня создание «полезного» фейка для рекламного ролика или фильма под силу не только крупным компаниям и киностудиям, но и маленьким продакшенам, и даже одиночкам. Цена производства зависит от сложности задачи, указывает Мария Чмир. «Вилка широкая: от десятков тысяч до сотен тысяч рублей, стоимость зависит от набора факторов,— говорит она.— Материал может быть коротким, но насыщенным обилием разнообразных планов с точки зрения цвета, света, динамики, „крупностей“ и ракурсов. Для корректной работы алгоритмов и экономии времени заказчика мы работаем с такими сценами сепарированно и параллельно, что влияет на объем затрачиваемых ресурсов: и человеческих, и технологических».

Для работы с дипфейками требуются определенные навыки, но нанимать много разработчиков и приобретать дорогостоящее оборудование сегодня необязательно, говорит основатель Deepcake. При необходимости небольшая студия или одиночка-фрилансер может арендовать необходимое оборудование, например видеокарты с мощными графическими процессорами, а не покупать их. Это значит, что порог входа в этот «бизнес» гораздо ниже, чем думают многие.

Марина Эфендиева