В интернете кончилась правда. Чем поддельные видео грозят миру? (18+)

Источник: https://zen.yandex.ru/media/4pda.ru/v-internete-konchilas-pravda-chem-poddelnye-video-groziat-miru-18-5bf342c22610c300aacd87d4?&from=feed

Виктор Пелевин в романе «Generation П» описал компьютер, подделывающий внешность и действия политиков так, что телезрители не подозревают подвоха. Требуются лишь похожий по комплекции актёр и студия с оборудованием. А дальше хоть на Березовского компромат снимай, хоть на Ельцина. Книга вышла в 1999 году, и тогда это казалось фарсом — в отличие от 2018-го, когда подобное стоит на пороге обыденности. И сейчас нам уже не до смеха. Потому что на фейковом видео могут засветиться не только селебрити, но и ваши мама, подруга или дочка.

ВНИМАНИЕ! СТАТЬЯ СОДЕРЖИТ МАТЕРИАЛЫ, НЕ РЕКОМЕНДОВАННЫЕ К ПРОСМОТРУ НЕСОВЕРШЕННОЛЕТНИМ.

Игра в имитацию

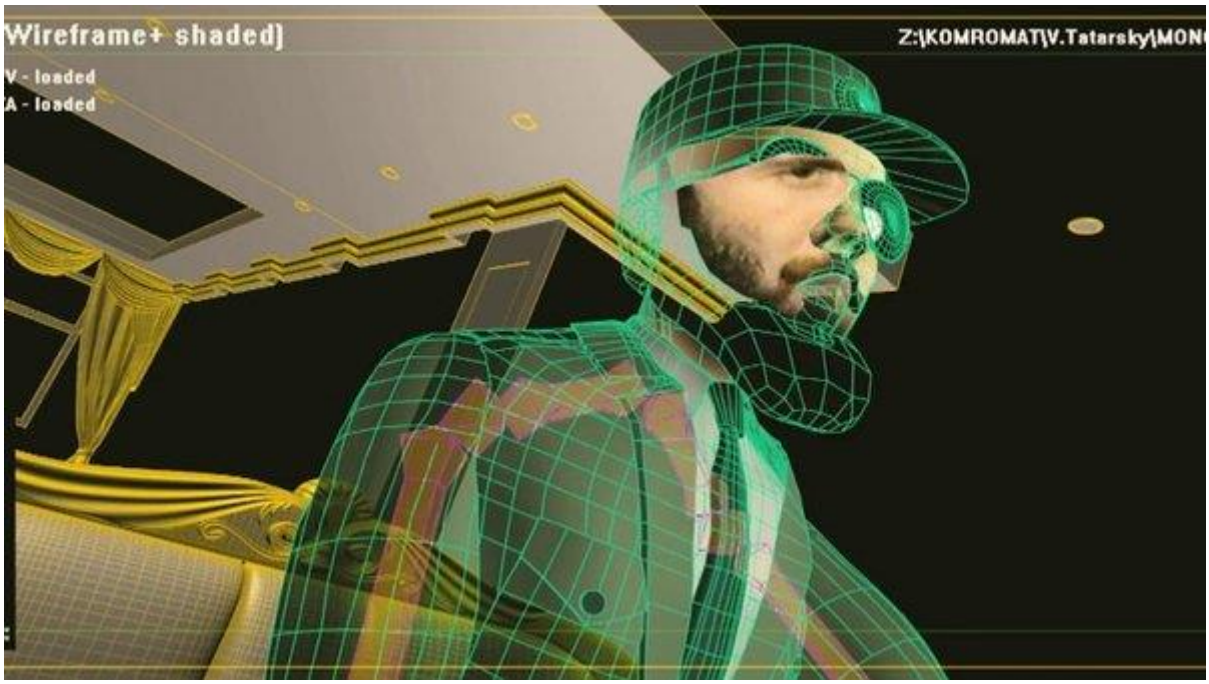
Хао Ли руководит стартапом Pinscreen в Лос-Анджелесе. Цель разработчика — сделать так, чтобы программа на ходу имитировала мимику человека, подстраивая её под чужую речь.

— И что для этого понадобится? Будете записывать меня на видео? — спрашивает репортёр LA Times.

— В том-то и фокус: хватит обычного фото, селфи. Остальное машина придумает сама. Вот, смотрите.

Конечно, ни Пелевин, ни господин Ли не открыли Америку. Исследователи как минимум полвека возятся с компьютерами и программами, пытаясь заставить их по-человечески распознавать, искать и выстраивать образы — достаточно вспомнить когнитрон 1975 года. Эта и подобная ей системы имитируют взаимодействие нервных клеток при помощи математических моделей. Поэтому их, по аналогии с мозгом, ещё называют искусственными нейронными сетями. Как и настоящее серое вещество, машина обучается путём сбора и анализа статистики, а заодно исправляет собственные ошибки. Но есть нюанс: лишь недавно техника достигла такой производительности, чтобы весь этот процесс начал реально менять общество. Отсюда — бум нейросетей в последние пару лет.

Человечество тут же придумало, как заставить сверхумные нейросети генерировать контент. Зачем нужны высокие технологии и несколько дней кропотливых расчётов? Правильно, чтобы приклеить к актрисе из фильма для взрослых физиономию матушки одноклассника. Это же неизбежно, особенно, если авторы упаковали чертовски сложную штуку в интуитивно понятную обёртку. Так даже обычный школьник соорудит нехитрый видеоколлаж. Лишь бы железо было достаточно мощное, чтобы не ждать сутки, пока отрендерится полуминутный ролик.



Кадр из фильма «Generation П» по мотивам романа Пелевина

Фэйс на боди натяну

Ажиотаж вокруг самообучающихся алгоритмов возник не из-за созданной при помощи ИИ живописи, музыки или написанного тем же макаром сценария короткометражки. Нет, скандал разгорелся, когда нейросеть DeepFake научилась использовать лица знаменитостей в роликах с пометкой *18+*. Иногда даже удачно — как в фальшивке с Галь Гадот, звездой фильмов о Чудо-женщине. Сперва анонимный энтузиаст выложил на Reddit приложение FakeApp, превращающее кого угодно в вашу бывшую и не требующее навыков программирования. Затем журналистка Motherboard Саманта Коул прославила это творчество в статье с двусмысленным, но точным заголовком *AI-Assisted Fake Porn Is Here and We're All Fucked*. И тут всё заверте…

С тех пор прошло больше года: подфорум с фейками заблокировали, хостинги изображений, гифок и видео объявили любителям визуального самопала войну. Однако количество роликов продолжает расти. Сегодня оно исчисляется сотнями, и некоторые умельцы даже берут индивидуальные заказы — получается хорошая подработка. А кто рождён

не для денег, тот просто прикалывается, вставляя повсюду лицо актёра Николаса Кейджа: эта мода возникла с подачи завсегдатая Reddit с ником Peter_File. И порой у него получается забавно. Толково обучить нейросеть выходит не у всех, а при ошибках в методах «тренировки» FakeApp порождает чудовищ. Однако факт налицо: из гостиной технология перебралась в спальню. Что дальше?

Кадр из фальшивого видео с Галь Гадот

Казалось бы, недавно интернет будоражили утечки интимных фото звёзд, выходки голливудских кутил попадали с плёнок папарацци на первые полосы газет, но скандальные «сливы» в одночасье потеряли смысл. Потому что сегодня изготовить подделку можно с кем угодно — хоть с соседкой из дома напротив или с коллегой по работе, если у вас есть несколько минут записи с её участием. Либо, как вариант, изменить собственную внешность ради подшучивания над теми, кто истово верует в картинки на мониторе. Киберпанк как он есть.

Посмеялись? Отлично. Теперь представьте, что технологии начнут использовать не только для фильмов со взрослыми, но и впутают сюда детей. Просто вообразите: анонимный шутник приставляет лицо вашей дочурки к обнажённому телу Биаты Ундин — и запускает «потешное» видео гулять по всей школе. Ну как, весёлая перспектива?

Идём дальше. Однажды вы открываете «ВКонтакте», а там — ролик с национальным лидером какого-нибудь государства. И этот национальный лидер говорит, что Муссолини, в общем-то, всё делал правильно. Шок? Ужас? Без сомнения. И невдомёк вам, что в преддверии выборов конкуренты запустили вирусное видео, состряпанное нейросетями. И все поверят. Ведь даже более халтурные коллажи прокатывали в новостях за вещдоки.

Программы вроде FakeApp обучаются, качество продукта растёт. А это значит, что рано или поздно мы очутимся в натуральном информационном аду. Примеры с сюрреалистичными высказываниями политиков покажутся невинной шуткой, когда за нейросети возьмутся террористы. Смотрели жутковатые ролики, где заложники стоят на коленях, а рядом бородачи с автоматами? Лица пленных не скрывают, чтобы потребовать с родных выкуп. А если это будут не террористы, а обычные вымогатели, разыгрывающие постановки для шантажа обеспеченных людей? «*Ваш родственник попал в беду*», версия 2.0, улучшенная. Конечно, если вдуматься и напрячь глаза, подделку разглядеть несложно. Но в стрессовой ситуации логика и мозг отключаются в первую очередь. Зато леденящий ужас окутывает моментально.

Не верь глазам своим

Сейчас будет немного сложно. Слышали о Рене Декарте? Французский философ утверждал, что лучше вообще ничего не знать, чем сомневаться в знаниях, — последнее почти всегда приводит к ложным убеждениям. Со смерти мыслителя прошло без малого четыре века, но его идеи актуальны и сейчас. Мы живём в условиях переполненного информацией мира и многое воспринимаем на веру. Где истина, а где ложь? Разобрать почти невозможно. Но один бастион у нас оставался — видео. «*Это не фото, его так просто не подделаешь*», — думали мы, всецело доверяя слитым роликам голливудских актрис и скандальным записям с людьми, похожими на прокуроров. А теперь, кажется, рухнул и этот — последний — оплот правды.

В интервью изданию Science Friday уже упомянутая журналистка Саманта Коул обрисовала суть проблемы:

Долгое время видео было золотым стандартом правды — возможно, не юридически, но в нашем сознании уж точно. Вы видите что-то на экране, а через пять секунд говорите «О боже, это и впрямь интересно!» и делаете ретвит, который набирает миллион репостов на Facebook. Так всё и происходит. Вот почему реалистичность фейков пугает. С поддельными новостями дела и без того плохи — достаточно вспомнить президентскую гонку 2016 года.

«*Если реальным может быть что угодно, то всё нереально*» — нет, это не вольное прочтение девиза Assassin's Creed, а лозунг, вброшенный защитниками DeepFake от блокировок. Они эпатируют, нагнетают, повергают устои, но в итоге отделаться от мысли об их правоте крайне сложно. Ведь именно так и выглядит прогресс. Только в данном случае мы не меняем водяное колесо на паровой котёл, а доверяем всю тонкую работу компьютеру, который ни черта не смыслит в делах «*этих кожаных ублюдков*».

На этом скриншоте заложники настоящие. Но нейросети позволяют «налепить» на тела любые лица — в том числе и ваших родных

Тут вы, конечно, спросите: а кто мешает черпать данные из авторитетных источников, не обращая внимания на случайные ролики в интернете? Правильно составленная информационная диета до поры до времени спасает от жирного троллинга. Вроде новости про «сгоревшие» холмы с той самой заставки Windows XP или жителя Обнинска, купившего у «*интеллигентных цыган*» монеты-биткоины. Но все ли готовы к подобному аккуратизму, да и кто убережёт сторожей? Ведь попасться на крючок могут в том числе лидеры мнений. Системе брошен вызов — она обязана ответить.

Над проблемой распознавания видеоподделок работает профессор Сивей Лю из университета Олбани. Он ищет программный ключ к истине:

Существует лазейка, которую изучаем мы, а также некоторые эксперты Дартмутского колледжа. Речь о физиологических сигналах вроде колебания в цвете кожи из-за давления. Вы же знаете, что у каждого человека есть сердцебиение, приводящее к притоку и оттоку крови? Так вот, используя оборудование, разработанное исследователями из МИТ, мы можем уловить эти незначительные ритмы, анализировать их и распознать фейк — в нём сигналы окажутся слабее, чем в настоящей записи.

Кроме лиц, нейросети умеют модифицировать погоду

Оно ближе, чем кажется

Думаете, всё это чертовски далеко и не на ваш век? Несколько лет назад точно так же характеризовали попытки нейросетей написать человеческий — во всех смыслах — текст. А теперь проведём эксперимент. Это — отзыв на Tripadvisor к одному из баров. Угадайте, кто его написал — благодарный клиент или бездушная машина? «*Мне нравится это место. Приехал сюда с братом, заказал вегетарианскую пасту — она оказалась вкусной. Пиво тоже отличное, да и обслуживание на высоте. Я рекомендую это место всем, кто ищет, где бы позавтракать*». Вы знаете ответ — здесь постарались нейросети. Мы немного схитрили — перевели отзыв с английского на русский, — но поверьте нам на слово: в оригинале всё выглядит столь же достоверно.

Вообще, обмануть наивного человека совсем не сложно. Эксперты по информатике из Чикагского университета пришли к выводу, что потребители уже сейчас не отличают качественный машинный обзор от «живого». Коэффициент полезности оценок нейросети — 3,15 против 3,28 (в случае отзывов, оставленных рядовыми пользователями и блогерами). Как видите, между ними не пропасть — и эта разница будет только сокращаться.

Конечно, провести нашего читателя проблематично — зря, что ли, вы круглосуточно читаете новости про технологии и гаджеты. Но что будет с мамой или бабушкой, если они разглядят ваше лицо в каком-нибудь ролике аферистов? Да что там бабушка! Эксперимент 4PDA показал: если обработать лицо нейросетями, не все взрослые люди смогут понять, мужское оно или женское.

Строго говоря, проблема не только в поддельных видео или отзывах. Будущее в целом обещает быть мрачным. Да, сегодня мы ещё можем установить истину, но что случится завтра? Ничего хорошего, считает Авив Овадья, главный медиатехнолог Центра контроля за социальными сетями Мичигана. Специалист предсказал «инфопокалипсис» задолго до того, как об этом стали трубить все СМИ. В интервью он высказал мысль, что технические решения лишь откладывают смерть реальности:

Как только появляется система, способная обнаружить фальшивку, кто-то приступает к обучению программы, обманывающей алгоритмы поиска подделок. И пока открыт доступ к методам защиты, атаки будут всё эффективнее. Так что я не считаю технологические решения долгосрочными — это больше напоминает игру в кошки-мышки.

А это значит, что мы вступаем в новую эру — верить нельзя ничему, даже своим глазам. Особенно своим глазам.