

Компьютер и человек – кто кого?

Продолжение.

Начало в №№ 7, 8, 9, 10/2023, №1/2024

«Компьютер делает не то, что вы хотите, а то, что вы ему приказываете»

Автор неизвестен

Арсений Ворошилов, по материалам OpenAI

В рамках этого цикла уже рассматривались как проблемы, связанные с применением генеративного искусственного интеллекта применительно к созданию медиаконтента, так и возможности, которые этот интеллект открывает для творчества. Объемы информации на эту тему растут, как снежный ком, поскольку появляется все больше инициатив, связанных с применением AI для создания медиаконтента. Больше того, эти инициативы воплощаются во вполне конкретных моделях, которые способны на такое, о чем раньше и мечтать было сложно. Одним из лидеров здесь является компания OpenAI, уже хорошо известная своим чат-ботом ChatGPT. Недавно компания представила свою новую AI-модель, способную создавать реалистичные образы на основе текстовых описаний. Степень реалистичности такова, что просто завораживает и даже немного пугает. И уж точно, эта модель заслуживает внимания.

Специалисты OpenAI стали учить свой искусственный интеллект понимать события и явления, происходящие в реальном физическом мире, и на основе этого понимания моделировать их, причем не только в статике, но и в движении. Цель тренировки моделей данным образом заключается в том, чтобы помочь людям решать проблемы, требующие взаимодействия с реальным миром. Именно для этого и была создана модель Sora, которую уже многому научили применительно к созданию видеоизображения на основе текстового описания. Sora способна генерировать видеофрагменты длительностью до минуты, сохраняя в видеоряде визуальное качество и обеспечивая соответствие текстовому описанию, сформированному пользователем.

Чуть выше уже отмечалось, что реалистичность сгенерированных средствами Sora видеофрагментов очень высока. Даже искушенный взгляд не сразу способен разглядеть, что видео – не результат съемки, а плод компьютерных вычислений. Соответственно, есть и обоснованные опасения относительно применения Sora. Поэтому, как уже давно принято в сфере компьютерных технологий, первыми доступ к модели получили специалисты по информационной безопасности, которые должны тщательно изучить все возможности Sora, выяснить, какие риски несет ее применение и какой вред может нанести злоумышленное ее использование.

Параллельно с этими специалистами OpenAI предоставил доступ к модели некоторым художникам, дизайнерам и кинематографистам, чтобы собрать отзывы о том, как можно улучшить модель, сделав ее максимально полезной для творческих профессионалов.

В целом, OpenAI применила стратегию определенной открытости процесса изучения Sora и привлечения к нему людей, не работающих в OpenAI, чтобы дать общественности возможность почувствовать, какие AI-возможности возникают на горизонте.

А каковы же они, эти возможности? На что способна Sora? Она способна генерировать сложные сцены с несколькими персонажами, специфическими типами движения и точными деталями как на самих персонажах и/или объектах «съемки» (условно назовем это так по аналогии с реальной съемкой), так и на фоне. Модель уже научилась понимать не только буквально то, что изложил в своем запросе пользователь, но и то, как это все существует в реальном физическом мире.

Модель характеризуется глубоким пониманием языка, что позволяет ей точно интерпретировать запросы и генерировать впечатляющие персонажи, способные выражать яркие эмоции. Sora даже может сгенерировать видео, состоящее из нескольких сцен (то есть фактически смонтированное из них), и в этих сценах на протяжении всего видеофрагмента сохраняется целостность персонажей и визуального стиля в целом.

Однако почитать на лаврах пока рано. Как уже отмечалось, это только первые шаги Sora, и как у любой новой технологии, у нее есть не только достоинства, но и недостатки. Точнее, слабые места, которые еще предстоит устранить разработчикам. В частности, у Sora периодически возникают проблемы с точной симуляцией физических процессов в сложных сценах, она может не понимать специфических причинно-следственных связей. К примеру, персонаж в одном кадре может откусить кусочек от конфеты, а в следующем кадре конфета снова будет целой, не надкусанной.

Порой модель вводит в заблуждение пространственные подробности текстового описания. Так, она может путать лево и право, а также испытывать затруднения с точным пониманием описания событий, длящихся во времени. Как пример, следование конкретной траектории движения виртуальной камеры.

Все это технологические задачи, которые без сомнения будут решены разработчиками Sora. Уверенность в этом вселяют уже имеющиеся у модели возможности. Но есть и соображения безопасности, над которыми в OpenAI тоже интенсивно работают.

В компании собираются предпринять несколько важных шагов по обеспечению безопасности, прежде чем сделать Sora широкодоступной, включив ее в состав других разработок OpenAI. Как отмечалось выше, команда специа-

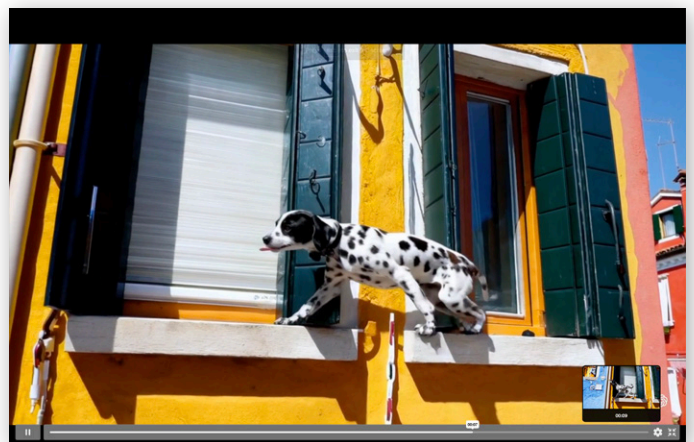
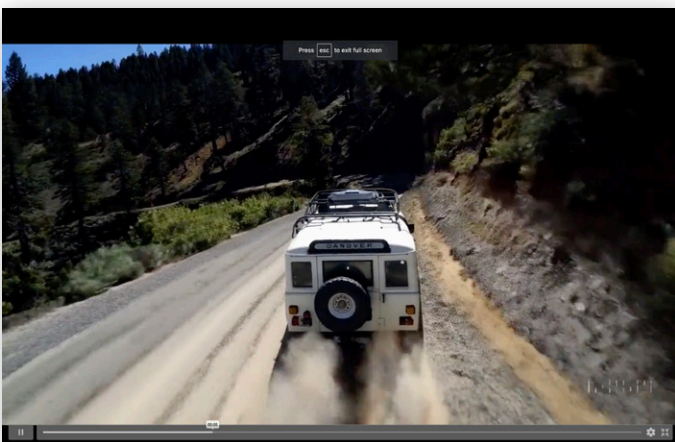
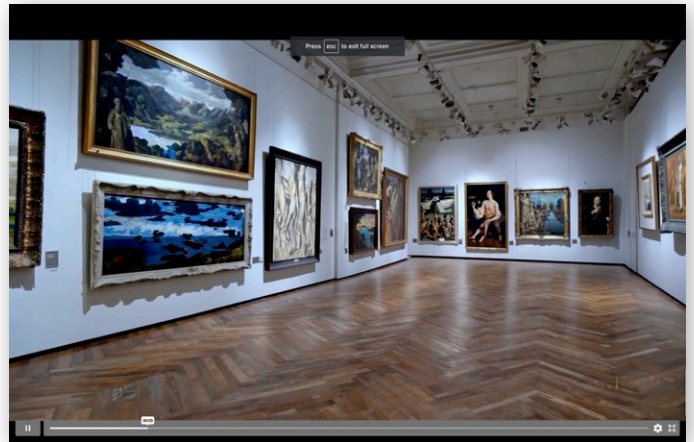
листов по информационной безопасности, специализирующаяся именно на данной тематике, уже занимается детальным изучением всех потенциальных опасностей. В частности, они уделяют пристальное внимание возможному применению Sora для создания дезинформации, ненавистнического и предвзятого контента. Им предстоит всесторонне протестировать модель.

Кроме того, в компании разрабатывают средства, способные помочь в выявлении вводящего в заблуждение контента. Например, это обнаруживающий классификатор, который может проинформировать о том, что данное видео сгенерировано моделью Sora. Есть также планы по включению в генерируемый контент метадан-

ных C2PA. Это будет сделано в будущем, если модель получит зеленый свет на внедрение ее в общий набор AI-средств компании.

В дополнение к разработке новых технологий в процессе подготовки к внедрению модели в практику, компания использует и существующие методы обеспечения информационной безопасности, которые были созданы в OpenAI ранее для решений, в основе которых лежит DALL·E 3. Эти методы вполне применимы и к Sora.

Например, оказавшись в составе решения OpenAI, текстовый классификатор компании проверит вводимые текстовые запросы и отсекает те, что нарушают правила пользования, принятые в компании. В частности, связан-



Кадры из видеофрагментов, сгенерированных Sora

ные с экстремизмом, сексуальным контентом, ненавистническими изображениями, схожестью со знаменитостями, IP других лиц и т. п. Разработаны также надежные классификаторы изображения, которые используются для оценки кадров каждого сгенерированного видео, чтобы обеспечить его соответствие принятым в компании правилам. Только после этого готовое видео становится доступным для пользователя.

В планах OpenAI – привлечь к работе политиков, педагогов и художников по всему миру, чтобы понять их проблемы и определить положительные примеры использования этой технологии. Несмотря на интенсивные исследования и испытания, пока нельзя предсказать ни все достоинства и выгоды, которые несет людям данная технология, ни то, как люди будут злоупотреблять ею. Вот почему в компании верят, что именно отзывы о реальном применении модели критически важны для создания и внедрения максимально безопасных AI-систем, что, несомненно, будет происходить.

Отдельного рассмотрения заслуживает и методика изучения модели. Sora представляет собой диффузионную модель, которая генерирует видео, начиная с того, которое выглядит как статический шум, постепенно трансформируя, поэтапно удаляя шум, и количество этапов велико.

Sora способна генерировать целые видеоролики сразу или делать уже сгенерированные видеофрагменты длиннее. Сложная задача обеспечения неизменности объекта в кадре, даже когда он временно исчезает из поля зрения, была решена за счет предоставления модели возможности предвидеть множество кадров одновременно.

Аналогично GPT-моделям, в Sora применена архитектура трансформатора, открывающая широкие возможности масштабирования.

Видеоролики и изображения представлены как коллекция меньших единиц данных, которые называются патчами. Каждый из них сродни токену в GPT. За счет унификации представления данных появляется возможность тренировки диффузионных трансформаторов на более широком спектре визуальных данных по сравнению с тем, что было возможно раньше, в том числе с охватом различных вариантов разрешения и форматов изображения.

Sora базируется на проведенном ранее изучении моделей DALL·E и GPT. В ней используется техника повторного захвата из DALL·E 3, которая предусматривает генерирование очень наглядных подписей для видеоданных, применяемых для тренировки модели. Как результат, модель способна следовать текстовым инструкциям пользователя, генерируя более реалистичное видео.

В дополнение к способности генерировать видео исключительно с опорой на текстовые инструкции, модель может взять существующее статичное изображение и на его основе сгенерировать видео, анимируя содержимое изображения с точностью и вниманием к малейшим деталям. Sora также способна взять уже готовое видео и сделать его длиннее либо добавить пропущенные кадры. И как здесь не задуматься о реставрации поврежденных кинофильмов или документальных видеоматериалов?

В общем, Sora служит основанием для моделей, которые могут понимать реальный мир и моделировать его. Именно эта способность, как считают в компании OpenAI, станет важнейшей вехой на пути к всеобъемлющему искусственному интеллекту – AGI (Artificial General Intelligence).

Продолжение следует

НОВОСТИ

Phantom S711

Компания Vision Research выпустила новую компактную высокоскоростную камеру Phantom S711, относящуюся к категории систем машинного зрения и предназначенную для применения в сложных приложениях. Камера обладает пропускной способностью 7 Гпк/с (56 Гбит/с) и набором полезных функций. Благодаря применению технологии CXPoF (CXP-over-Fiber) камера обеспечивает передачу данных по двум оптическим кабелям со скоростью до 56 Гбит/с или по одному кабелю там, где достаточно 40 Гбит/с.

Предусмотрены возможность выделения двух областей из общего кадра и совместимость с объективами EOS. Съемка может вестись со скоростью до 7360 кадр/с в разрешении 1280×800, подача опорного сигнала и управление фокусом осуществляются дистанционно. Разрядность обработки изображения может быть 8 и 12 бит на выбор пользователя, чувствительность камеры находится на квантовом

уровне, экспозиция составляет 1 мкс, а при установке опции FAST – 300 нс.

При минимальном разрешении видеозаписи, которое составляет 128×16, скорость съемки достигает 249080 кадр/с. Изображение формируется сенсором CMOS разрешением 1280×800, размер пикселя – 20 мкм, размер сенсора по диагонали – 30,2 мм, разрядность формируемых сенсором данных – 12 бит.

Камера не содержит встроенной памяти, так что данные с ее выходов записываются

во внешние кадровые буферы или на видеорекодеры.

В стандартной комплектации Phantom S711 комплектуется байонетом F, на который можно установить объективы типов F и G, в качестве опции допускается замена байонета на C и Canon EOS. В комплект поставки объектив не входит.

Для питания камеры требуется источник 12...32 В (блок питания входит в комплект), от которого она потребляет 52 Вт. Размеры камеры – 125×125×160 мм, масса – 2,4 кг.

