Инновационная система синхронизирует аудио с любым видеоконтентом.

Исследователи Google DeepMind добились значительных успехов в области создания видео, представив технологию V2A (Video-to-Audio). Их новая разработка способна автоматически генерировать звуковое сопровождение для видеороликов, изначально созданных без звука.

V2A работает на основе машинного обучения и объединяет анализ видеопикселей с текстовыми подсказками. Это позволяет программе создавать реалистичные звуковые эффекты, музыку и даже диалоги, соответствующие происходящему на экране. Технология совместима с другими разработками DeepMind, такими как модель генерации видео Veo. Совместная работа V2A и Veo позволит создавать полноценные видеоролики с синхронизированным звуком и визуальными эффектами.

Запрос: Кинематографический, триллер, фильм ужасов, музыка, напряжение, атмосфера, шаги по бетону

Помимо озвучивания «немого кино», V2A может применяться к различным типам архивных материалов, включая старые фильмы и видеозаписи, что открывает широкие возможности для творческой обработки и реставрации архивного видео.

V2A способна создавать неограниченное количество звуковых дорожек для одного и того же видео. Пользователь может влиять на результат с помощью текстовых подсказок. «Положительные подсказки» направляют модель на создание определенных звуков, а «отрицательные» – наоборот, исключают нежелательные звуковые элементы. Такая гибкость управления позволяет авторам экспериментировать с разными вариантами озвучки и выбирать наиболее подходящий.

V2A выделяется на фоне существующих решений для озвучивания видео тем, что способна работать с «сырыми» видеопикселями, а текстовые подсказки являются необязательным элементом. Кроме того, система избавляет от необходимости вручную синхронизировать звук и изображение, что обычно требует кропотливой подгонки различных элементов (звуков, визуальных эффектов и таймингов).

Однако технология все еще находится в стадии разработки, и исследователям предстоит решить ряд задач. Например, качество генерируемого звука напрямую зависит от качества исходного видео. Артефакты или искажения в видео могут привести к заметному снижению качества звука.

Все права защищены

Еще одна область совершенствования – синхронизация речи с движениями губ. V2A пытается генерировать речь на основе текстовых расшифровок и синхронизировать ее с движениями губ персонажей. Однако модель генерации видео, с которой сопряжена V2A, может не учитывать эти расшифровки. Это приводит к неестественной синхронизации, когда движения губ не совпадают с произносимыми словами.

Google подчёркивает важность безопасности и прозрачности при разработке своих технологий. Компания собирает отзывы от ведущих создателей и кинематографистов, чтобы совершенствовать V2A. Кроме того, Google использует инструмент SynthID для нанесения водяных знаков на созданный искусственным интеллектом контент, чтобы предотвратить возможные злоупотребления.

Прежде чем сделать технологию V2A доступной для широкой аудитории, она пройдёт тщательные проверки и тестирование. Первоначальные результаты показывают, что эта технология имеет большой потенциал для создания оживленных и реалистичных видео.

Все права защищены