

Сам себе AI-режиссёр: популярная нейросеть научилась создавать анимационные видеоролики

13.10.2023



Недавно возможность создавать видеоролики по текстовому описанию в режиме анимации появилась у пользователей нейросети Сбера Kandinsky 2.2. По одному текстовому описанию генерируется 4-секундное видео с выбранным эффектом анимации, с частотой 24 кадра в секунду и разрешением 640 x 640 пикселей. Синтез одной секунды видео в среднем занимает около 20 секунд.

Генерация анимационных видеороликов работает в тестовом режиме и доступна самым активным пользователям нейросети, которые получают приглашение в ближайшее время. До конца года оценить новые возможности нейросети смогут абсолютно все. Те пользователи, которые хотят попробовать новые возможности модели сейчас, могут оставить заявку на получение доступа [в Telegram-боте](#).

Для того чтобы сгенерировать видео в режиме анимации, нужно описать текстом то, что хочется увидеть. Далее бот предложит на выбор 16 вариантов анимации сцены, а после этого нейросеть сгенерирует анимированный видеоролик. Также доступна генерация составных сцен: пользователь может ввести несколько текстовых описаний (до трёх), затем выбрать для каждого свою механику анимации, а после этого модель создаст «мини-фильм».

«С момента выхода модели Kandinsky 2.2 пользователи уже сгенерировали более 50 млн изображений. Теперь у них появилось ещё больше возможностей для творчества совершенно бесплатно.

Сегодняшний запуск видеофункции в режиме анимации – это важный шаг в развитии нашей нейросети и для всей глобальной индустрии мультимодальных моделей искусственного интеллекта. Мы продолжим и дальше совершенствовать Kandinsky, и в следующих обновлениях качество будет только улучшаться», – рассказал первый заместитель Председателя Правления Сбербанка Александр Ведяхин.

Основой для синтеза видео является модель генерации изображений по текстовым описаниям нейросети. Для расширения её возможностей были реализованы разные виды анимации изображений, что позволило перемещать объекты, приближать и отдалять их, оживлять статику всеми возможными способами. В основе всех режимов анимации лежат функции `image2image` (перерисовка изображения по текстовому описанию) и `inpainting/outpainting` (дорисовка части изображения внутри и за его пределами), которые уже были реализованы в базовой модели.

Нейросеть разработали и обучили исследователи Sber AI совместно с учёными из Института искусственного интеллекта AIRI на объединённом датасете Sber AI и компании SberDevices.