Woosh – много звука из всего

Арсений Ворошилов, по материалам Sony AI

Сегодня AI-генераторами звука уже никого не удивишь. Их много, применяются они довольно широко и для решения всевозможных задач, начиная с преобразования текста в речь до создания различных не речевых звуков, таких как звук работающего мотора, шум ветра, течение воды, музыка и др.

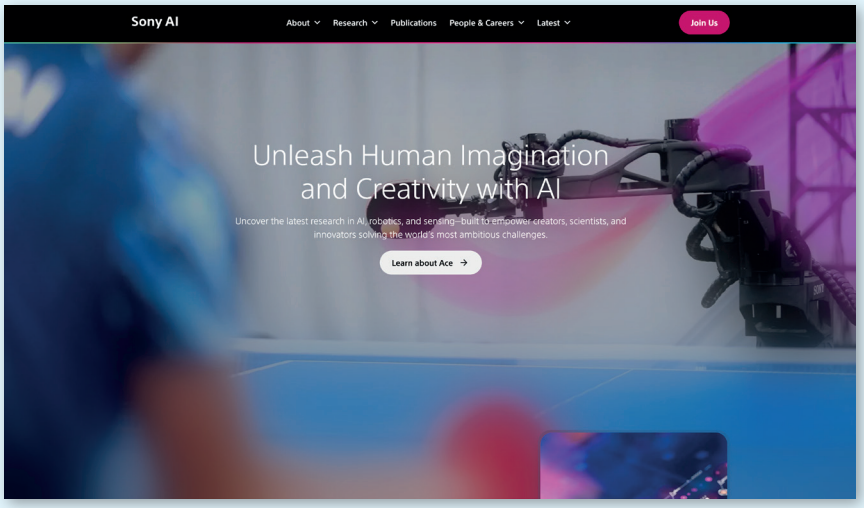
Тем не менее, многим профессионалам медиаиндустрии, специализирующимся в сфере аудио, а значит, обладающим тренированным искушенным слухом, недостаточно того качества и разнообразия, которое обеспечивают нынешние AI-генераторы. Во всяком случае, именно к такому выводу пришли специалисты Sony AI, посчитав, что профессиональное создание аудио с помощью искусственного интеллекта требует фундаментально иных данных и управления, чем те, что применяются в AI-генераторах общего применения. А дальше – от слов к делу. В Sony AI разработали и выпустили базовую модель, созданную специально для генерации звуковых эффектов.

Команда Sony AI одновременно обучила две версии своей AI-модели. Частная модель, оптимизированная для получения результата студийного качества, обучалась на библиотеках лицензированного профессионального звука, таких как Pro Sound Effects и BOOM. Модель с такой же архитектурой прошла обучение на общедоступных наборах данных и была сделана доступной для исследовательского сообщества. Она получила название Woosh в честь одного из наиболее известных звуковых эффектов, используемых в компьютерных играх и кинофильмах.

AI-генератор Woosh, предназначенный для рабочих процессов, применяемых в игровой индустрии, кинематографе и интерактивных медиа, способен решать две генеративные задачи: преобразовывать текст в звук, генерируя звуковой эффект на основе текстового описания, и создавать звук на основе видео, генерируя его непосредственно на базе видеопоследовательности. Во втором случае предусмотрена возможность дополнительной текстовой подсказки для получения более точного результата. Дизайнеры звука для игр и кино, работающие больше с визуальным контентом, а не с абстрактными описаниями, вероятно, найдут вторую функцию модели более полезной и подходящей для себя.

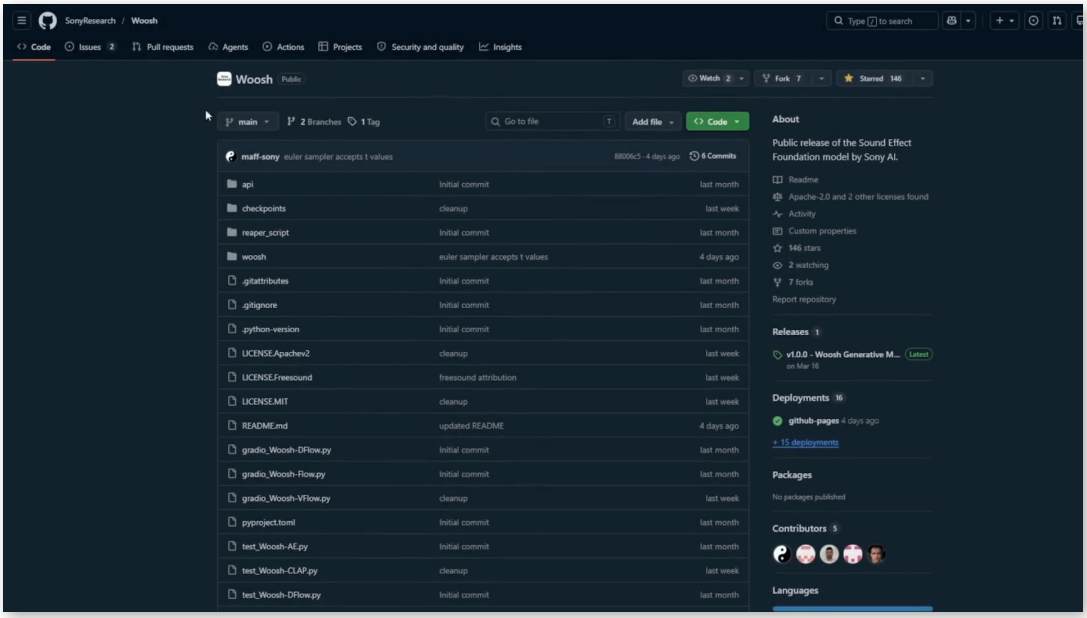
Нужно отметить, что между общедоступными наборами аудиоданных и профессиональными SFX-библиотеками есть существенная разница, на которую указала группа разработчиков генератора. Общедоступные наборы аудиоданных могут изобиловать окружающим звуком и накладывающимися друг на друга шумами, не очень четко выделяясь на этом фоне. К примеру, записанный на улице звук работающего автомобильного мотора может смешиваться с шумом ветра, шелестом листьев, звуками дорожного движения и др. В отличие от этого, профессиональные SFX-библиотеки содержат специально записанные звуки, тщательно отредактированные, снабженные точной маркировкой и разметкой в соответствии с тем, как профессионалы выполняют поиск желаемого аудио.

Частная модель Woosh существенно превосходит общедоступные альтернативы в плане данных о профессиональных звуковых эффектах. Так отмечается в отчете Sony AI. Там же говорится, что общедоступная модель превосходит сравнимые модели с открытым исходным кодом по имеющимся в общем доступе показателям. Тестирование на FoleyBench – первой крупномасштабной платформе, предназначенной для оценки генерации аудио на основе видео по методу Foley, показало, что Woosh решает задачу генерации видео по аудио лучше по качеству и показате-



Основанная в апреле 2020 года, Sony AI работает на стыке технических исследований и человеческой изобретательности. Здесь исследуют искусственный интеллект, предназначенный не для замены человеческого духа, а для того, чтобы AI действовал как катализатор творчества.

Компания видит свою миссию в анализе «умных» систем, которые работают в гармонии с людьми, дополняя природные таланты художников, ученых и творцов, чтобы обогатить общество и раздвинуть границы его коллективного творчества. Коллектив компании исследует это будущее в сотрудничестве с провидцами со всего мира, приглашая к партнерству тех, кто одержим переосмыслением ландшафта человеческого развития.



Пользовательский интерфейс Woosh

лям семантического выравнивания, используя при этом меньше исходных параметров.

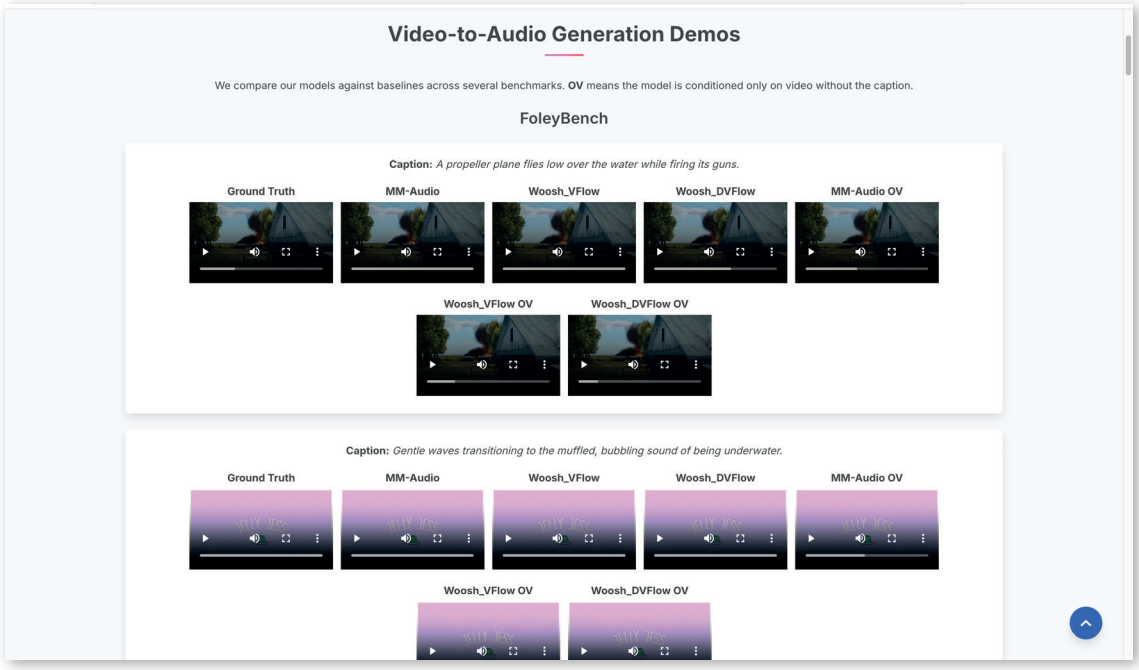
Команда дополнительно занималась разработкой подключаемого программного модуля для DAW с запланированной поддержкой генерации вариаций, возможности сопряжения сгенерированного аудио с уже существующим звуком и персонализации. Как отметили представители Sony AI, этот модуль может быть тесно интегрирован в эти конвейеры, рабочие процессы и наборы инструментов так, чтобы звукорежиссеры могли работать более эффективно.

По мере совершенствования экосистемы планируется расширить управление генератором. В планах добавление более точного временного контроля, морфинга (преобразования звука в другой на основе семантического описания результата), генерация точных циклов и персонализация на базе одного образца аудио или небольшого числа образцов. Все эти возможности отражают тот уровень точности творческого контроля, который необходим профессионалам.

Полностью осознавая споры вокруг AI и его потенциального влияния на ситуацию с рабочими местами, команда Sony AI ставит перед собой цель понять, где искусственный интеллект может выступать инструментом поддержки человеческого творческого процесса. Элементы управления, встроенные в программный модуль, а также решение обучать модель на лицензированных профессионально отобранных библиотеках, а не на аудиоданных, собранных из открытых источников, подтверждают приверженность компании созданию помощника для человека, а на замены ему.

В Sony AI заявляют, что лицензирование отражает продуманную стратегию. Публичный релиз носит некоммерческий характер и предназначен для демонстрации возможностей технологии, при этом коды инференса и веса модели доступны сообществу для исследований и экспериментов. В то же время закрытая – частная – модель, обученная на лицензированных студийных данных высокого качества, ориентирована на коммерческое применение. В целом же публичный релиз призван подготовить почву для профессиональной модели, которая уже разрабатывается. Уровень производительности у нее существенно выше, в чем и заключается смысл для профессионалов.

В завершение нужно отметить, что AI-генератор типа Woosh может оказаться очень полезным в самых разных ситуациях – от создания оригинального звука для сгенерированного с помощью AI видео до замены исходного реального аудио или его обогащения дополнительными элементами и эффектами.



Тестирование Woosh на платформе FoleyBench