Microsoft представила искусственный интеллект, имитирующий любой человеческий голос

Автор: Administrator
12.01.2023 07:53

Разработка получила название VALL-E. Она может имитировать тембр и манеру речи, прослушав голос всего три секунды.

Microsoft назвала своё достижение «языковой моделью нейронного кодека». VALL-E создавалась на основе EnCodec (звуковой кодек, использующий методы машинного обучения). В отличие от других синтезаторов речи, которые используют преобразование форм сигналов, решение от Microsoft проводит анализ, как именно звучит человек, разбивает эту информацию на отдельные сегменты и использует обучающие алгоритмы, чтобы сопоставить информацию из своих баз данных с тем, как этот голос будет звучать, если ИИ произнесёт другие фразы.

На сайте проекта можно ознакомиться с множеством примеров работы VALL-E , которые поделены на 4 колонки. В разделе Speaker Prompt можно прослушать оригинальную трехсекундную запись голоса, в Ground Truth — фраза целиком, Baseline приводит пример обычного синтезатора речи, в колонке VALL-E представлен результат работы новой технологии Microsoft.
VALL-E обучали на основе библиотеки LibriLight, содержащей 60 000 часов англоязычной речи более чем от 7000 человек.