

Baidu, которую иногда называют китайским аналогом Google, в 2013 году открыла исследовательскую лабораторию, посвященную искусственному интеллекту в Кремниевой долине. Теперь компания представила свои разработки в сфере синтеза речи. Результатом работы стала система преобразования текста в речь, названная Deep Voice, которая может научиться говорить всего за несколько часов с минимальным вмешательством человека.

В прошлом году DeepMind от Google также совершила значительный прорыв в этой области. Компания представила нейронную сеть, которая учится говорить, слушая реальную речь и сравнивая ее с текстом расшифровки. После обучения система WaveNet смогла синтезировать речь на основе изученного текста. Осенью прошлого года она научилась говорить уже на двух языках.

Deep Voice — это усовершенствованный корпорацией Baidu WaveNet. Нейросеть использует методы глубокого обучения, чтобы преобразовать текст в минимальные воспринимаемые единицы звука, называемые фонемами. Затем она использует синтез речи для воспроизведения этих звуков. На каждом этапе процесса применяется глубокое обучение, а потому нет необходимости в том, чтобы в дальнейшем система была настроена человеком. Возьмем, к примеру, слово «hello». ИИ Baidu сначала определяет границы фонемы следующим образом: (молчание НН), (НН, ЕН), (ЕН, L), (L, OW), (OW, молчание). Затем он вводит их в систему синтеза речи, которая «произносит» слово.

Google также не сидит без дела, а продолжает развитие своего продукта — компания планирует увеличить штат DeepMind в 2,5 раза, а также переманивает к себе лучших специалистов из Оксфорда и Кембриджа. Искусственный интеллект от Google называют лидером в своей области, и компания не собирается сдавать позиций.