

Договоры со мною, робот!

тест редакции

Арсений Ворошилов

Применение алгоритмов искусственного интеллекта – AI (Artificial Intelligence) в медиаиндустрии становится все более широким и систематическим. AI-системы используются для решения разных задач, включая фильтрацию и сортировку контента, первичную обработку текстов, поиск данных в архивах, автоматизированное описание контента и т. д.

Одним из наиболее распространенных, а также наиболее доступных вариантов применения AI являются преобразование речи в текст и текста в речь. Первое упрощает журналистам расшифровку интервью, а второе дает возможность сформировать звуковую дорожку в случаях, когда пользователь не хочет либо не способен начитать ее собственным голосом, либо когда необходимо сделать это на языке, который пользователь освоил недостаточно хорошо либо не освоил вовсе.

Сегодня в Интернете достаточно много различных онлайн-сервисов, заявленных как сервисы преобразования текста в речь. Поэтому в редакции журнала MediaVision возникла мысль протестировать некоторые из таких сервисов, чтобы сформировать более четкую картину того, что доступно пользователям.

Первое, что сразу бросилось в глаза, – сервисы есть бесплатные, условно бесплатные и платные, как, в общем, и в любом сегменте рынка онлайн-сервисов. Под условно бесплатными я понимаю сервисы, которые вводят ограничения на ассортимент возможностей, например, делают доступными не все голоса, лимитируют объем текста и т. д.

И второе, что нельзя не заметить, это довольно бедные возможности применительно к русскому языку. Что, опять же, характерно для широкого спектра программного обеспечения. Дело в том, что хотя русский язык и является одним из самых распространенных на планете, а также входит в число официальных языков ООН, рынок России настолько невелик, что производителям ПО и поставщикам сервисов просто не рентабельно делать русскоязычную локализацию своих решений. Увы, придется с этим смириться. Исключение составляют в основном сервисы, изначально разработанные русскоязычными программистами для русскоязычной же аудитории. Такие тоже есть, но речь о них пойдет в одной из следующих статей.

А ниже хочу рассказать об опыте знакомства с сервисом Amazon Polly. Для меня это был наиболее очевидный стартовый вариант, поскольку и Amazon Web Services относится к числу лидеров в сфере применения AI, и сам этот сервис является бесплатным.

Что нужна регистрация, это, думаю, объяснять не надо. Ну а дальше переходим к работе с Polly. Для этого требуется войти в консоль управления и выбрать сервис Polly. Дальше все очень просто и понятно. Нужно выбрать язык и голос. Для русского языка все не просто скучно, а очень скучно – всего два варианта голоса: мужской и женский. Для сравнения, при выборе английского языка голосов девять – шесть женских и три мужских.

Теперь, собственно, к предмету. Пока текст имеет ярко выраженный бытовой характер, не содержит каких-либо специальных терминов или иностранных слов, все выглядит довольно неплохо. Конечно, слышен некоторый «машинный акцент», но в целом все вполне прилично.

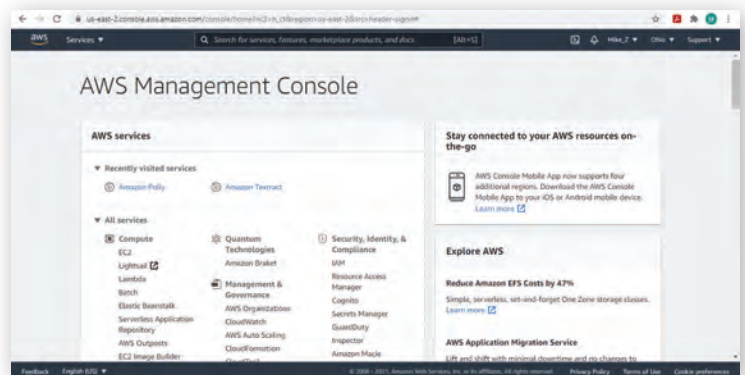
А вот если в тексте есть нечто, хоть немного выходящее за пределы базового русского лексикона, начинаются проблемы. Причем складывается впечатление, что робот выучил русский язык не вполне, а английский при этом подзабыл. Пример: слово services робот произносит как «серви-чез». Откуда это взялось в недрах облака AWS, понять сложно.

Возникают проблемы и с ударениями в словах, и пока принудительно поставить ударение в слове возможности нет. Можно лишь расставить акценты в предложении, да и то, применив базовые операции SSML-программирования. Аналогичные операции придется выполнить и для

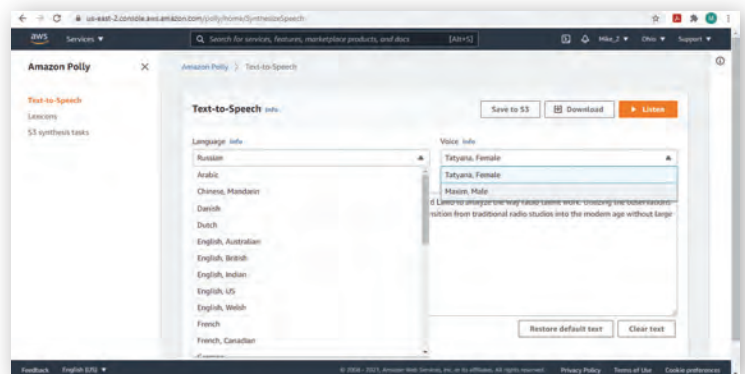
того, чтобы, к примеру, добавить паузу, обозначить слово на отличном от выбранного языка, использовать какое-то особое произношение и т. д. Опций много, что хорошо, но придется освоить азы SSML-программирования. И все равно при преобразовании текста в речь именно на русском языке могут возникнуть проблемы, причем порой довольно неожиданные.

А вот на английском, что вполне ожидаемо, все очень хорошо, почти оптимально. Немного портит идиллию довольно сильно выраженное механическое звучание речи. Но если срочно требуется сформировать звуковую дорожку на правильном английском, а рядом нет носителя языка с хорошо поставленной речью, то AWS Polly – то, что нужно.

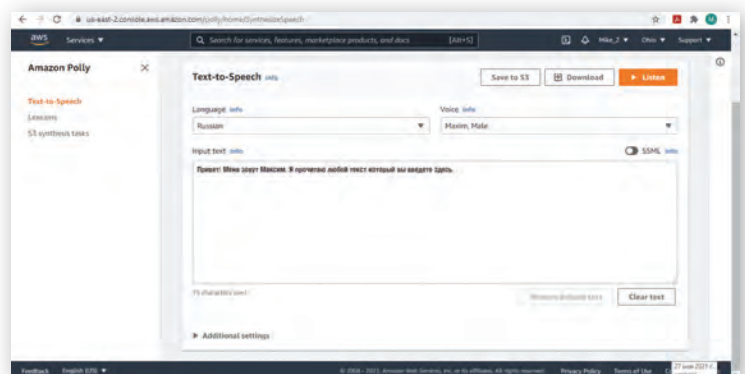
Продолжение следует



Консоль управления AWS



Выбор языка и голоса



Интерфейс ввода текста для преобразования в речь