

В пресс-службе Минобрнауки РФ сообщили, что в Санкт-Петербурге завершена разработка системы распознавания речи на ливвиковском наречии карельского языка. Проект реализовали специалисты Федерального исследовательского центра РАН. Новая технология может лечь в основу онлайн-переводчика, который будет работать с использованием искусственного интеллекта и переводить устную речь в реальном времени.

Карельским языком владеют около 30 тысяч человек по всему миру, из них 25 тысяч — в России. Однако в быту его регулярно используют лишь 5–7 тысяч человек. Это делает язык уязвимым для исчезновения.

На первом этапе система будет работать как компьютерное приложение, распознающее речь через микрофон. Со временем ее планируют адаптировать для мобильных устройств, чтобы расширить круг пользователей. Все собранные данные будут храниться на сервере и использоваться для дальнейшего обучения нейросетей.

Сложность проекта заключалась в малом количестве доступных аудио- и текстовых материалов. Команда использовала корпуса карельского и вепсского языков, а также собрала собственные записи.