Разработана система распознавания эмоций человека по видео и речи

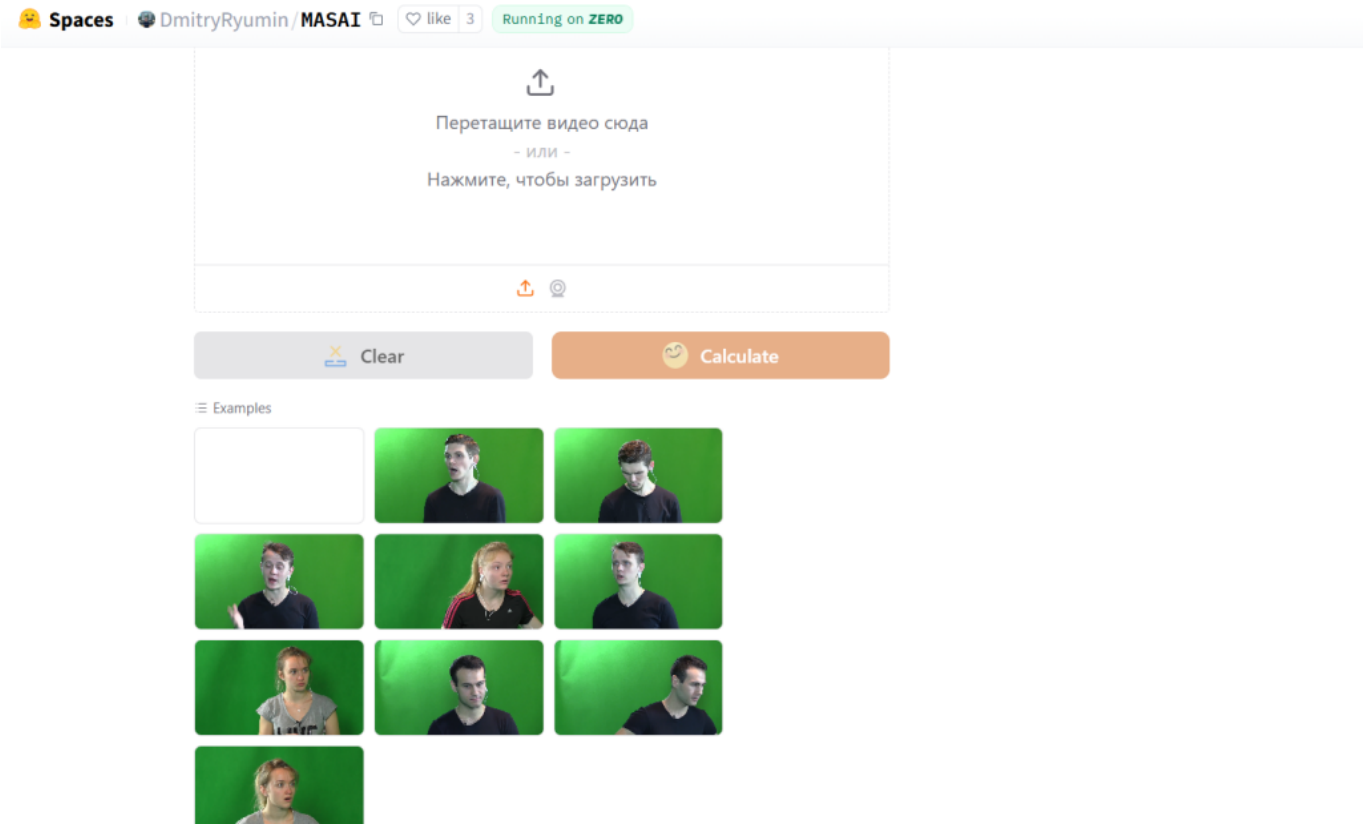
28.08.2025

Исследователи Санкт-Петербургского Федерального исследовательского центра РАН (СПб ФИЦ РАН) создали интеллектуальную программную систему MASAI, которая включает в себя комплекс нейросетей, способных распознавать естественные эмоции и сентимент человека по аудиовизуальной информации (видео, звук и текст).

Точность распознавания эмоций превосходит существующие аналоги и достигает 80%. В перспективе разработка позволит создавать более эффективные цифровые помощники, работающие на принципах эмоционального искусственного интеллекта.

Внедрение искусственного интеллекта (ИИ) в современном мире играет важнейшую роль, поскольку он способствует повышению эффективности, автоматизации и инновационности в различных сферах жизни и деятельности человека — сегодня этот процесс стремительно развивается. ИИ-сервисы позволяют ускорить обработку больших объёмов данных, принимать более точные решения, оптимизировать процессы и создавать новые возможности для бизнеса, науки и повседневной жизни.

Одной из задач для дальнейшего внедрения ИИ-сервисов в жизнь человека — научить искусственный интеллект распознавать весь спектр разнородной информации от пользователя: речь, мимику, жесты, взгляд, поведение, а также текстовую информацию. Подобные системы смогут работать даже в сложных условиях: шума, плохого освещения или когда человек просто молчит или не смотрит в камеру. Более того, ИИ-сервисы смогут распознавать такие сложные эмоциональные проявления, как сарказм или ирония — когда смысл сказанного не совпадает с интонацией или выражением лица. В итоге это позволит повысить качество работы цифровых помощников.



Интерфейс программы MASAI

«В рамках решения этой задачи мы разработали интеллектуальную систему MASAI многомодального анализа видео, звука и текста для распознавания эмоций человека, например, радости, злости, грусти и прочих эмоциональных проявлений. Кроме того, программа может проводить сентимент-анализ информации — определять отношение человека к какому-либо событию — отрицательное, положительное или нейтральное. Взаимодействие с пользователем происходит через дружественный графический интерфейс системы», — поясняет руководитель Лаборатории речевых и многомодальных интерфейсов СПб ФИЦ РАН **Алексей Карпов**.

В основе системы лежит несколько нейронных сетей, которые проходили обучение на более десятка находящихся в свободном доступе для исследователей базах данных, содержащих мультиязычную аудиовизуальную информацию о людях разного возраста и пола из различных стран мира. В частности, при создании базы данных на русском языке принимали участие молодые актёры театральных вузов, которые имитировали различные эмоции.

Программа может распознавать семь базовых человеческих эмоций (радость, грусть, страх, отвращение, удивление, злость и спокойствие), а также положительный, отрицательный и нейтральный сентименты. Учёные используют наиболее распространённую в психологии классификацию эмоциональных состояний человека.

Разработка представляет собой программное приложение, через интерфейс которого можно загрузить и обработать любые аудиовизуальные данные. В результате система автоматически проанализирует информацию и выдаст результат распознавания: эмоции и сентимент. По точности распознавания эмоций приложение СПб ФИЦ РАН превосходит все аналоги; этот показатель составляет около 75-80 % в зависимости от анализируемых данных.

«Система MASAI может быть встроена в различные виды цифровых помощников, которые сейчас используются во многих сферах жизни человека. Например, в телефонных службах скорой помощи или психологической поддержке, где эмоциональный искусственный

интеллект позволит эффективнее реагировать на потребности человека», — объясняет Алексей Карпов.

Сейчас программа распознает эмоции по заранее записанным аудиовизуальным данным, а также может анализировать мимику и речь человека в режиме реального времени. Разработка поддержана грантом РНФ.

Источник: пресс-служба Минобрнауки России.

Новости Российской академии наук в Telegram ➡