

«Яндекс» занялся разработкой новой нейросетевой модели под названием SpeechGPT, хотя компания пока официально не объявила об этом. Эта модель, как сообщается, будет мультимодальной, что означает, что она способна обрабатывать и текстовую, и звуковую информацию, а также выполнять разные задачи, связанные с текстом и звуком.

Представители «Яндекса» пояснили, что они уже работают над мультимодальностью в своих сервисах, таких как ассистент «Алиса», однако они не дали ответа на вопрос о модели SpeechGPT. Некоторые сервисы «Яндекса» уже умеют обрабатывать данные разного типа: например, ассистент «Алиса» позволяет общаться голосом и текстом, а в сервисе Yandex SpeechSense появилась возможность анализировать текстовые сообщения.

Однако эксперты отмечают, что есть разница между мультимодальным пользовательским опытом и самой мультимодальностью модели. В отличие от мультимодальных моделей, такие как GPT-4o, от OpenAI, которые могут работать сразу с разными видами контента без задержек, текущие подходы часто требуют перевода данных из одного формата в другой, что занимает больше времени.