В AIRI разработали ИИ-модель, которая мыслит образами

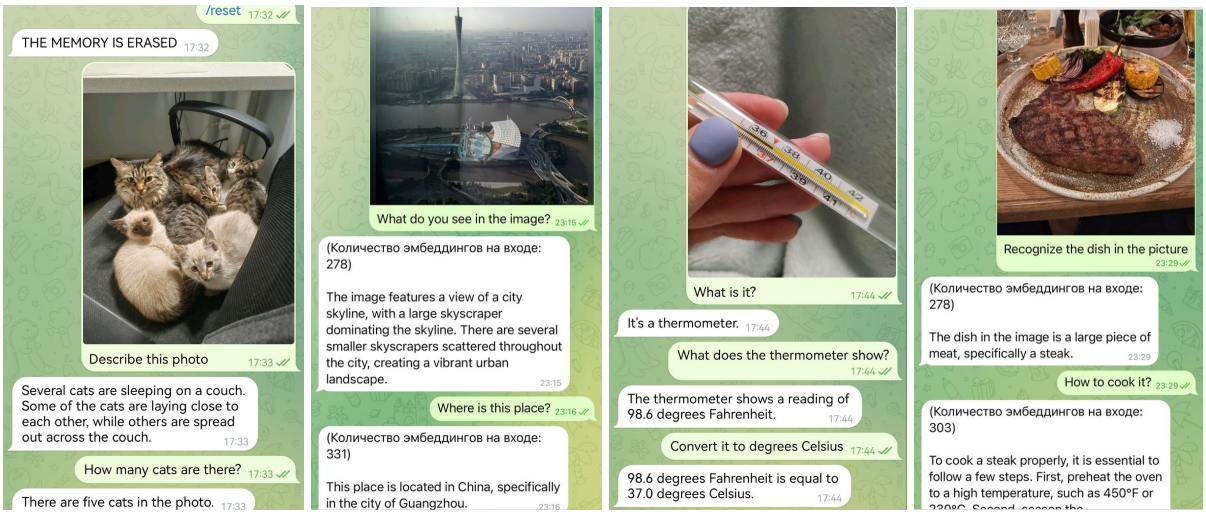
Новая мультимодальная диалоговая модель OmniFusion способна анализировать, описывать и отвечать на вопросы по изображениям, поддерживая непрерывный диалог с пользователем. Об этом 22 ноября на конференции AI Journey 2023 рассказал доктор физико-математических наук, CEO Института искусственного интеллекта AIRI Иван Оселедец.

Анализируя предоставленные пользователем изображения, OmniFusion точно распознаёт расположенные на них объекты, их количество, цвет и положение в пространстве. Модель способна не просто описать картинку, но и ответить на сопутствующие вопросы, а также использовать извлечённую информацию в ходе диалога с человеком. Например, она может распознать сфотографированное блюдо и предложить рецепты его приготовления, найти ответ на логическую задачу или графический ребус, а также объяснить мем.

В модели OmniFusion используется очень интересный способ объединения разных модальностей — картинок и текста — без дорогостоящего обучения «с нуля». Исследователи правильным образом построили энкодеры и дообучили уже существующую языковую модель понимать изображения.

Обучением модели занималась научная группа FusionBrain Института AIRI при участии ученых из Sber AI и SberDevices.

В ходе обучения OmniFusion использовали датасеты, составленные из картиночно-текстовых диалогов, а также вопросов с ответами по картинкам.



Как работает OmniFusion

Среди уже существующих в мире аналогов модели можно выделить 2 наиболее производительных решения: модель LLaVA и модель GPT-4V от OpenAI, которая ранее была интегрирована в сервис ChatGPT. Модель OpenAI закрыта для сторонних разработчиков. Сравнение OmniFusion с открытой моделью LLaVA на основе 10 различных бенчмарков показало, что качество OmniFusion не уступает, а в ряде случаев даже превосходит

конкурента, несмотря на то, что в основе OmniFusion лежит намного более «легкая» языковая модель. В основе LLaVA лежит языковая модель с 13 млрд. параметров, в то время как языковая модель в основе OmniFusion содержит всего 7 млрд. Это значит, что модель более экономичная и быстрая.

«Сейчас модель стабильно работает на английском языке и обучается грамотному владению русским, чтобы стать доступной пользователям, а наша команда готовит научную публикацию о процессе создания OmniFusion. Общение с помощью изображений – это новый уровень взаимодействия ИИ-модели с человеком, более естественный и привычный для каждого из нас формат коммуникации. Мы будем продолжать активно развивать модель и добавлять в неё новые модальности»

Иван Оселедец, доктор физико-математических наук, CEO Института искусственного интеллекта AIRI

Вопросы: pr@airi.net

Об AIRI

Научно-исследовательский Институт искусственного интеллекта AIRI — автономная некоммерческая организация, занимающаяся фундаментальными и прикладными исследованиями в области искусственного интеллекта. На сегодняшний день более 90 научных сотрудников AIRI задействовано в исследовательских проектах Института для работы совместно с глобальным сообществом разработчиков, академическими и индустриальными партнерами.