Новосибирский государственный университет
*НАСТОЯЩАЯ НАУКА

Ученые НГУ разработали отечественную платформу для управления массивами данных

Первыми ее протестируют участники проекта «Школьники — научные волонтеры». Ребята со всей страны смогут не только исследовать природу своей малой родины, но и формировать датасеты с фото- и аудиоматериалами, описаниями, метриками и данными о локации объектов живой природы.

Новосибирск, 03 сентября 2024 года: В рамках гранта, направленного на поддержку исследовательских центров искусственного интеллекта, научные сотрудники Центра искусственного интеллекта и Лаборатории аналитики потоковых данных и машинного обучения Механико-математического факультета Новосибирского государственного университета создают решение, обладающее всеми преимуществами ИИ для сферы строительства и обустройства городской среды. Над разработкой базового программного обеспечения — фреймворка КАППА «Управление датасетами и моделями», — работает интернациональная команда разработчиков, к которой привлечены специалисты из России, Перу, Индии, Вьетнама, Китая, Казахстана, а также группа философов, которые формируют нормы этичной разработки алгоритмов искусственного интеллекта.

— *Разрабатываемый нами фреймворк предназначен для поддержки управления различными наборами данных. На его основе можно создать приложения для различных применений. Первое отладочное применение фреймворка — проект «Школьники – научные волонтеры», поскольку в его рамках школьники в соответствии с этическими принципами создают научный набор данных, который впоследствии можно использовать для обучения искусственного интеллекта. Мы планируем и самих ребят привлекать для перекрестной проверки, чтобы повысить качество данных. Когда оно будет достаточно высоким, мы сможем обучать искусственный интеллект при обработке данных решать те же задачи, что и школьники. Но делать это мы будем не для того, чтобы избавить юных исследователей от выполнения этой работы и переложить ее на искусственный интеллект. Для нас важно показать им, как на основе полученных ими данных создаются алгоритмы искусственного интеллекта, предназначенные для решения тех задач, которые они ранее решали сами. Мы хотим достичь важной педагогической цели — обучения школьников основам применения и обучения искусственного интеллекта, исходя из понимания сути алгоритма*, — рассказал заведующий Лабораторией **Евгений Павловский**.

Идея проекта, находящегося на стыке двух наук – биологии и искусственного интеллекта, принадлежит заведующему Лабораторией аналитики потоковых данных и машинного обучения Механико-математического факультета НГУ, ведущему научному сотруднику Центра искусственного интеллекта НГУ Евгению Павловскому. Данный проект призван пополнить знания школьников о природе малой родины, пробудить в них интерес к изучению окружающего мира, научить применять научный подход, мотивировать к исследовательской деятельности через познание методов искусственного интеллекта. Своей идеей ученый поделился с директором регионального Фонда «Образование» Майей Гичгелдиевой. Данная организация занимается проектами по вовлечению школьников в исследовательскую и инженерную деятельность, организует программы подготовки наставников проектной деятельности школьников. Было решено подать грантовую заявку, которая получила поддержку Фонда президентских грантов.

В настоящее время проект «Школьники – научные волонтеры» вышел на новый этап — более 1300 школьников и педагогов со всей страны прошли регистрацию в специально разработанной базе данных, которую они будут пополнять результатами своих наблюдений за живой природой. Данный проект вошел в число победителей конкурса Фонда президентских грантов и реализуется при содействии Правительства Новосибирской области. Его организатором выступает региональный Фонд «Образование». Цель проекта — профессиональная ориентация школьников в научно-исследовательской деятельности через систему научного волонтерства. Научными партнерами проекта стали Институт систематики и экологии животных СО РАН и Институт почвоведения и агрохимии СО РАН.

Более сотни образовательных учреждений из 30 регионов страны, получив письма от организаторов, заявили о готовности участвовать в данном волонтерском проекте. Среди них – школы Санкт-Петербурга, Новосибирской, Омской, Томской, Калининградской, Московской, Нижегородской, Тверской и Челябинской областей, Красноярского, Ставропольского и Приморского краев, а также других регионов России.

Для участия в проекте были сформированы команды из педагогов и учеников, где на одного наставника приходилось несколько школьников. Обучение научных волонтеров проводилось в формате онлайн-лекций на платформе SYNCWOIA. Участников проекта обучали ведению записей, определению видов животных, растений, грибов и почв, правильному ведению видео- и аудиозаписей. Также ребята получили новые знания по проведению исследований и работе в сфере информационных технологий — с большими данными, нейросетями и компьютерным зрением.

С марта этого года группы научных волонтеров работали на местности. Каждая команда выбирала несколько природных площадок, всесторонне исследовала их и собирала материал о растениях, животных, грибах и почвах, подробно фиксируя данную информацию. Теперь им предстоит сформировать собственные датасеты с фото-, аудио-, текстовыми описаниями, метриками и данными о площадке в общей базе данных. Важно, чтобы данные не обладали внутренней противоречивостью и соответствовали принятым нормам научных наблюдений, поэтому они будут верифицированы учеными. После присвоения данным статуса «проверено» или «верифицировано» всю информацию объединят в научную базу данных, на которой и будет обучаться модели искусственного интеллекта.

В середине сентября школьники перейдут к защите своих проектов, а создаваемые на основе их наблюдений за природой базы данных пополнятся новой информацией. Будет произведена работа по повышению качества собранных данных. Специалисты, задействованные в проекте, снабдят базу данных дополнительными функциями, которые позволят взглянуть на уже внесенную информацию, делая отбор по территориям и другим признакам. Непосредственно работа по обучению искусственного интеллекта с участием школьников начнется в сентябре 2024 года.