Источник: “Народная газета” – 2024-12-06

Границы дозволенного

Джинн XXI века — искусственный интеллект — уже вырвался из бутылки. И хотя история ИИ как научного направления началась еще в середине XX века, а множество предпосылок для его зарождения было сформировано и вовсе задолго до этого, самое интересное, похоже, только начинается. Или страшное? В духе научной фантастики, пророчившей неизбежность доминирования ИИ над миром. Вот и Президент Беларуси Александр Лукашенко на недавнем саммите ОДКБ в Астане высказал обеспокоенность по поводу перспектив неконтролируемого использования возможностей искусственного интеллекта, который все чаще задействуют для планирования и решения военных задач: «Имея способность к самообучению, этот инструмент может погубить человечество, если его выпустить из-под контроля». Сегодня «взращиванием» искусственного интеллекта занимаются крупные исследовательские центры в разных странах. Как обстоят дела с ИИ в Беларуси? Об этом поговорили с заведующим лабораторией анализа биомедицинских изображений Объединенного института проблем информатики Национальной академии наук Эдуардом Снежко.

Александр Лукашенко:

— С одной стороны, современные технологии создают тысячи новых возможностей и перспектив. С другой стороны, они порождают множество рисков и угроз — фейки, дезинформация, атаки на критическую инфраструктуру.

28 ноября 2024 года, на саммите ОДКБ в Астане.

На третьей волне

— Эдуард Витальевич, насколько быстро развивается искусственный интеллект и в каких сферах он чаще всего используется?

— Методы искусственного интеллекта начали развиваться довольно давно. Но около десяти лет назад произошли качественные изменения в ряде направлений (объем данных, аппаратное обеспечение и математический аппарат), в результате которых технологии при решении некоторых задач сравнялись или превзошли по точности результаты, предоставляемые людьми, не говоря уже о скорости. Например, это было показано в некоторых задачах распознавания текста и изображений.

Сегодня технологии ИИ используются в самых разных сферах: машиностроении для проектирования различных сложных систем, медицине, банковской сфере, моделировании лекарственных препаратов, управлении автотранспортом и так далее. Нейросети становятся все более доступными для широкого круга пользователей. Поэтому для некоторых людей открывается направление для поиска уязвимостей в системах на основе ИИ, что влечет необходимость разработки методов защиты от подобных воздействий.

— Какими разработками в сфере ИИ заняты ученые Объединенного института проблем информатики НАН?

— И наш институт, и Национальная академия наук в целом давно занимаются интеллектуальными системами. В частности, в нашей лаборатории разработан и зарегистрирован в НЦИС веб-сервис LungExpert, который позволяет анализировать радиологические изображения грудной клетки. Например, на изображениях КТ можно автоматически локализовать и количественно оценить параметры образований шести типов: легочный инфильтрат, ателектаз, каверны и др.

Чтобы обучить нейросеть детектировать признаки проявления патологий на медицинских изображениях, мы работали со многими медицинскими центрами, совместно определяли задачи, затем передавали им свои разработки на апробацию.

Александра Янкович. Границы дозволенного

Источник: “Народная газета” – 2024-12-06

Отзывы были положительные. В настоящее время хотели бы прийти к широкому использованию наших разработок профильными специалистами. Однако организационно этот процесс небыстрый: есть ряд этапов, которые еще нужно пройти, — создание инфраструктуры, соответствие существующих и разрабатываемых нормативных актов и так далее.

Также в нашем институте есть лаборатория, занимающаяся поиском и проектированием химических соединений, которые потенциально могут стать активными компонентами лекарственных препаратов для конкретных заболеваний.

Не стоит слепо доверять

— Надежна ли информация, которую для нас генерирует нейросеть?

— Нейросети подвержены определенного вида атакам. Поэтому важно понимать, как их подготовить или модифицировать, чтобы они были устойчивы к угрозам такого типа. Ведь последствия могут быть самые разные. Если будет атакована наша нейросеть, анализирующая медицинские данные, то результат может быть искажен. Грубо говоря, больного человека она идентифицирует как здорового. Хорошо, если медик обратит на это внимание. А если упустит?

Еще одним аспектом, на который стоит обращать внимание в работе с методами ИИ, — генеративные нейросети могут уверенно давать ответ на любой вопрос без возможности сопоставлять смысловое содержание с реальностью. Будет ли такой ответ соответствовать действительности? Нейросеть спроектирована таким образом, что не ответит вам «недостаточно данных или нет информации», а будет генерировать результат без того смысла, который мы хотели бы ожидать. Не говоря уже о том, что фотографии, аудио- и видеозаписи, тексты, которые ранее считались более-менее надежным и достоверным источником информации, сегодня приходится дополнительно проверять. Существуют инструменты, позволяющие определять, сгенерировано нечто или нет, но они не так распространены для массового использования. Да, некоторые задачи нейросети решают хорошо, даже лучше, чем человек, так как они обобщают большие базы данных и выделяют ключевые для решения ряда задач признаки об объектах из этих баз. Но не стоит слепо доверять им, итоговое решение должно быть за человеком.

Задать рамки

— Тема правового регулирования применения искусственного интеллекта на слуху во всем мире. В Беларуси тоже планируется разработать проект закона о технологиях ИИ. Назрела необходимость?

— Необходимость в таком законе на международном уровне назрела давно, его проект разрабатывался несколько лет, сейчас подошли к финальной стадии. В чем суть? У нас заключен договор в рамках СНГ о разработке модельного закона о технологиях искусственного интеллекта. Цель — общая, рамочная унификация законодательства стран, входящих в СНГ.

Развитие искусственного интеллекта — это тренд, который нельзя игнорировать. Он сегодня генерирует миллиарды долларов в разных странах. Везде свои подходы. В целом они делятся на два типа: все запрещено, кроме того, что разрешено, и наоборот — все разрешено, кроме того, что запрещено. Последний принцип применяется в США и Китае. И уже видно, насколько они ушли вперед в развитии ИИ. А вот суровое законодательство в Евросоюзе, особенно касающееся обмена данными в разных областях, напротив, стало тормозом в развитии входящих в него стран. В итоге две трети компаний перебазировались из ЕС в Великобританию, где соответствующее законодательство мягче. В нашем случае, я думаю, будут учтены разные подходы реализации с учетом опыта других стран.

Александра Янкович. Границы дозволенного

Дать команду

— Может ли ИИ выйти из-под контроля человека, как это произошло в прошлом году с дроном во время учений ВВС США?

— Насколько я помню, в данном случае в целевую функцию дрона внесли поощрение и наказание. Поощрение давалось за близость к цели, а когда оператор специально препятствовал достижению цели, система его неким образом якобы опознала как помеху. Этот случай стоило бы более внимательно изучить. К слову, это не единичный случай, когда ИИ ошибался. Появляются периодически новости об ошибках, в том числе и трагических, автопилота наземного транспорта, и это никого не удивляет.

Здесь, вероятно, есть недочеты на уровне разработки. То есть важно создать надежные алгоритмы и системы безопасности, которые в некой степени гарантируют, что ИИ будет выполнять только заданные и одобренные команды. Это подразумевает проведение разносторонних тестов и проверок для обнаружения и предотвращения возможных уязвимостей.