23 ноября 2021 г.

Искусственный интеллект в контексте информационной безопасности

Президиум РАН, заслушав и обсудив доклады академика РАН Соколова И.А. «Безопасность технологий искусственного интеллекта» и академика РАН Аветисяна А.И. «Кибербезопасность в контексте искусственного интеллекта», выступления академика РАН Лекторского В.А. «Человек и системы искусственного интеллекта», академика РАН Каляева И.А. «Некоторые аспекты искусственного интеллекта, требующие научного обоснования», доктора физико-математических наук, профессора РАН Воронцова К.В. «Технологии искусственного интеллекта и безопасность информационного пространства» (Московский государственный университет имени М.В. Ломоносова), заместителя министра цифрового развития, связи и массовых коммуникаций Российской Федерации, президента Академии криптографии Российской Федерации доктора физико-математических наук Шойтова А.М. «Проблемы безопасности искусственного интеллекта в Российской Федерации и исследования, ведущиеся в Академии криптографии», директора Департамента перспективных технологий «Лаборатории Касперского» Духвалова А.П. «Безопасность искусственного интеллекта», отмечает, что за последнее десятилетие, достигнут существенный прогресс в создании новых методов искусственного интеллекта (ИИ), в первую очередь, - машинного обучения. Для многих прикладных задач

разработаны алгоритмы, достигающие качества, достаточного для промышленного применения. Технологии искусственного интеллекта получили преимущественное распространение в областях, для которых доступно большое количество данных. Растет применение машинного обучения и ИИ в науке, особенно в биологии и медицине. Наблюдается увеличение количества работ по предсказанию свойств лекарственных препаратов, анализу геномных данных, обработке и анализу биомедицинских изображений.

Внедрение технологий ИИ требует учета ряда возникающих рисков и новых угроз. Современный ИИ включает целый ряд технологий, состоящий из методов и алгоритмов, платформ машинного обучения, а также инфраструктурных решений для их поддержки (облачные системы, специализированные аппаратные системы и др.). Помимо классических уязвимостей программного обеспечения, технологии ИИ являются источниками новых типов ошибок и уязвимостей, которые предоставляют иные возможности для проведения атак злоумышленниками. Это глобальный вызов, поиск ответа на который начался лишь 3-4 года назад. Работа в этом направлении будет восходящим трендом в ближайшее десятилетие, поскольку ответы на эти вызовы требуют соответствующих инструментов и фундаментальных результатов. Очень важно сейчас исследовать математические аспекты безопасности ИИ и применять криптографию.

В связи с развитием ИИ возникают новые проблемы правового, мировоззренческого, философского и этического свойств. Каков правовой статус цифровых субъектов и роботов, что такое цифровой двойник человека, может ли двойник быть наделен теми же правами и обязанностями, что и реальный человек? Эти вопросы активно обсуждаются в теории права и в философии. Развитие ИИ меняет отношение и к традиционным вопросам о свободе воли, о личности как таковой: что такое «я», что такое сознание и может ли им обладать искусственный интеллект? Возникают и этические проблемы:

бесконтрольные разработки могут стать угрозой для человечества, не случайно введены понятия «дружественный (friendly) искусственный интеллект», «доверенный (trustworthy) искусственный интеллект», то есть доброжелательный по отношению к человеку.

В рамках реализации Национальной стратегии развития ИИ на период до 2030 года, утвержденной Указом Президента Российской Федерации от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» и Федерального закона от 24 апреля 2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве», и внесении изменений в статьи 6 и 10 Федерального закона от 27 июля 2006 г. № 152-ФЗ «О персональных данных» запущен ряд программ поддержки развития технологий ИИ. В частности, в 2021 году запущенна программа поддержки исследовательских центров в сфере ИИ. В рамках данной инициативы, на основе конкурсных процедур, было отобрано шесть организаций, ставших опорными точками развития технологий ИИ в Российской Федерации (Опорные центры ИИ РФ). Одним из таких центров реализуется программа исследований по направлению доверенного ИИ. Объективно возникает необходимость: активизации взаимодействия РАН с ведущими отечественными и международными организациями в области развития ИИ как ключевого направления сохранения информационной безопасности; формулирования и продвижения в практику научно обоснованного определения «искусственного интеллекта»; разработки и утверждения научно обоснованной методики оценки доверия к ИИ; разработки и внедрения совместно с Росстандартом и техническим комитетом № 164 по стандартизации ИИ, иными регуляторами стандарта оценки доверия системам ИИ.

Президиум РАН ПОСТАНОВЛЯЕТ:

1. Принять к сведению информацию, представленную в докладах и выступлениях.

2. Научному совету РАН по методологии искусственного интеллекта и когнитивных исследований до 1 июня 2022 г. разработать (совместно с заинтересованными организациями) методику оценки доверия к «искусственному интеллекту».

3. Отделению математических наук РАН (академик РАН Козлов В.В.) и Отделению нанотехнологий и информационных технологий РАН (академик РАН Красников Г.Я.) до 1 марта 2022 г. подготовить предложения по созданию Научного совета РАН по вопросам безопасности информационных технологий.

4. Отделению нанотехнологий и информационных технологий РАН (академик РАН Красников Г.Я.) до 1 марта 2022 г. подготовить и представить в президиум РАН проекты обращений в:

4.1. Министерство науки и высшего образования Российской Федерации с предложением создать на базе научных организаций (институтов РАН), находящихся под научным и научно-методическим руководством РАН, не менее пяти новых лабораторий по обсуждаемому направлению исследований с финансированием в рамках государственных заданий;

4.2. Министерство экономического развития Российской Федерации с предложением создать на базе региональных научных или образовательных организаций высшего образования не менее трех исследовательских центров в области искусственного интеллекта для реализации собственных программ по обсуждаемому направлению исследований;

4.3. Российский научный фонд с предложением разработать программу поддержки малых научных групп для реализации собственных проектов (не менее десяти проектов в год) по обсуждаемому направлению исследований;

4.4. Федеральное агентство по техническому регулированию и метрологии (Росстандарт) и технический комитет № 164 по стандартизации ИИ с предложением по разработке проекта стандарта оценки доверия к системам ИИ.

5. Рассмотреть возможность активизации на базе научных организаций (институтов РАН), находящихся под научным и научно-методическим руководством РАН, международного научного сотрудничества и диалога между учеными по широкому спектру направлений, представляющих взаимный интерес по вопросам ИИ.

6. Контроль за выполнением настоящего постановления возложить на вице-президента РАН академика РАН Козлова В.В.

Президент РАН
академик РАН А.М. Сергеев

И.о. главного ученого секретаря
президиума РАН
член-корреспондент РАН А.А. Макоско