



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ

«*Российская Академия Наук*»

ПРЕЗИДИУМ
ПОСТАНОВЛЕНИЕ

12 марта 2024 г.

Москва

№ 42

Технологии искусственного
интеллекта в изучении и
поддержке человека

Президиум РАН, заслушав и обсудив доклады академика РАН Ушакова Д.В. «Технологии искусственного интеллекта в изучении и поддержке человека», академика РАН Хабриевой Т.Я. и профессора РАН, доктора юридических наук Черногора Н.Н. (Институт законодательства и сравнительного правоведения при Правительстве Российской Федерации) «Проблемы правовой идентификации искусственного интеллекта», академика РАН Соколова И.А. «Искусственный интеллект и анализ данных», академика РАН Аветисяна А.И. «Искусственный интеллект в гуманитарной сфере. Угрозы и возможности», академика РАН Бетелина В.Б. «Научные проблемы обеспечения суверенитета в области технологий искусственного интеллекта» и выступления академика РАН Каляева И.А., академика РАО Галажинского Э.В., академика РАО Зинченко Ю.П., профессора РАН Нестика Т.А., заместителя министра промышленности и торговли Российской Федерации Шпака В.В., статс-секретаря – заместителя министра цифрового развития, связи и массовых коммуникаций Российской Федерации Руденко А.В., директора Департамента законопроектной деятельности Правительства Российской Федерации кандидата юридических наук Степанова П.В., председателя Арбитражного суда Московского округа кандидата юридических наук Кресса В.В., отмечает, что темы, связанные с развитием искусственного интеллекта, затрагивают наивысшие достижения

человечества, а быстрый прогресс последних лет в этой сфере имеет существенное значение для наук о человеке. Благодаря методам машинного обучения возникла возможность разрабатывать принципиально новые модели поведения человека, основанные на больших объемах данных, в том числе – регистрации явлений реальной жизни. В прикладном плане на основе машинного обучения началось развитие совокупности технологий, связанных с автоматической оценкой свойств человека, его состояний, предсказанием возможных действий, а также оценкой эффективности методов воздействия.

Эти технологии могут быть использованы для помощи людям в различных областях: в сфере образования, профессиональной деятельности, труде в сложных условиях, личной жизни, психологическом консультировании, психотерапии и т.д. Существует большой потенциал коммерциализации этих разработок: они потенциально широко востребованы населением как средство обеспечения психологического комфорта, роста, здоровья и безопасности, но они также имеют емкий рынок в таких сферах народного хозяйства, как образование, здравоохранение, сфера услуг, труд в напряженных и экстремальных условиях и т.д. Эти разработки, несомненно, связаны с обеспечением безопасности государства. Все это свидетельствует о том, что развитие технологий искусственного интеллекта в сфере наук о человеке составляет одну из наиболее перспективных и приоритетных областей современного научно-технологического развития.

В то же время научно-технологический прогресс в этой сфере требует приложения существенных усилий и решения ряда проблем. Важной задачей является преодоление дисциплинарных барьеров и организация взаимодействия между научными и образовательными учреждениями, относящихся к разным областям – гуманитарным наукам, информационным технологиям, математике и т.д. Междисциплинарность должна обеспечиваться на уровне государственных заданий на НИР и НИОКР. Важную роль в поддержке междисциплинарного взаимодействия должны играть российские научные и научно-технологические фонды, а также комплексные программы и проекты, предполагающие создание высокотехнологичного продукта. Продуктивность междисциплинарного взаимодействия требует подготовки специалистов, владеющих

междисциплинарными компетенциями, что должно учитываться на всех уровнях высшего образования и повышения квалификации.

Развитие исследований в сфере искусственного интеллекта, в том числе направленного на изучение и поддержку человека, требует инфраструктуры, связанной с созданием вычислительных мощностей, а также формированием баз данных для машинного обучения, доступных для исследователей. Необходимо обеспечить максимально полное использование вычислительных мощностей страны для создания необходимых условий проведения приоритетных работ. Необходим доступ специалистов к данным для машинного обучения, в том числе в сфере наук о человеке, с учетом защиты персональных данных и иных требований безопасности.

Условием успешного развития в области искусственного интеллекта для изучения и поддержки человека в нашей стране является взаимодействие научно-исследовательских центров с индустриальными партнерами, в качестве которых выступают высокотехнологичные компании, компании-потребители технологий искусственного интеллекта в сфере человеческого фактора, а также малые инновационные предприятия, и с государственными заказчиками. Необходимо установить наиболее плотный контакт и обмен между научно-исследовательскими центрами, имеющими наиболее продвинутые фундаментальные разработки, и российскими индустриальными партнерами.

Внедрение технологий искусственного интеллекта в сфере человека предполагает учет ряда возникающих рисков и новых угроз. Эти технологии, наряду с применением в целях повышения психологического благополучия, увеличения производительности труда, личностного роста и обеспечения психологической безопасности, могут быть использованы для причинения вреда, нанесения психологического ущерба, манипуляций, введения в заблуждение, систематического воздействия на население вплоть до психологических войн.

Современный искусственный интеллект (ИИ) создал, помимо классических уязвимостей программного обеспечения, также новые типы ошибок и уязвимостей, которые предоставляют возможность проведения атак злоумышленниками. Эта проблема становится особенно значимой в контексте

обеспечения национального технологического суверенитета. В настоящее время происходит широкое использование зарубежных полуэмпирически построенных искусственных нейронных сетей (ИНС) как основы технологий ИИ и для промышленности, и для ФОИВ. Для этих ИНС отсутствуют доказательства их сходимости и устойчивости, что, собственно, и является непосредственной причиной «... отсутствия понимания того, как искусственный интеллект достигает результата». То есть, по сути дела, означает угрозу неустраняемого отсутствия гарантий штатного функционирования систем ИИ на основе таких ИНС. В настоящий момент именно отсутствие документов стратегического характера, нацеленных на достижение технологической независимости от зарубежных технологий в области вычислительных и информационных технологий, включая и технологии ИИ, препятствует устранению этой угрозы.

В связи с развитием ИИ возникают новые проблемы правового, мировоззренческого, философского и этического свойств. Правовое регулирование достижений в области регулирования ИИ не всегда поспевает за развитием новых технологий. В российском законодательстве ориентиры развития ИИ обозначены только в стратегических документах. Между тем, в пункте 19 Национальной стратегии развития искусственного интеллекта на период до 2030 года защита прав и свобод человека обозначена в числе основных принципов развития и использования технологий ИИ. Данный принцип включает: обеспечение защиты гарантированных российским и международным законодательством прав и свобод человека, в том числе права на труд, и предоставление гражданам возможности получать знания и приобретать навыки для успешной адаптации к условиям цифровой экономики. На защиту прав человека направлен и принцип безопасности, предполагающий защиту человека от умышленного причинения вреда в результате использования ИИ.

В целом развитие технологий искусственного интеллекта, направленных на изучение и поддержку человека, представляет собой важную задачу отечественной науки, для решения которой требуются приложить существенные усилия, нацеленные как на создание условий продуктивной работы специалистов, так и на избежание возможных рисков.

Заслушав изложенную в докладах и выступлениях информацию и обсудив ее, президиум РАН считает целесообразным активизировать работу в сфере искусственного интеллекта, направленного на изучение и поддержку человека, и ПОСТАНОВЛЯЕТ:

1. Принять к сведению информацию, представленную в докладах.

2. Отделению общественных наук РАН, Отделению нанотехнологий и информационных технологий РАН и Отделению математических наук РАН активизировать совместную работу по формированию научно-технической политики в сфере создания технологий искусственного интеллекта, направленных на изучение и помощь человеку, учету рисков, сопутствующих этим технологиям, и юридическому, социологическому, философскому и психологическому сопровождению внедрения этих технологий. Привлечь к этой работе представителей высокотехнологичных компаний, стартапов и потенциальных заказчиков создаваемых технологий.

3. Отделению общественных наук РАН совместно с Отделением нанотехнологий и информационных технологий РАН и Отделением математических наук РАН до 12 апреля 2024 г. подготовить и представить в установленном порядке в президиум РАН проекты обращений в:

3.1. Министерство науки и высшего образования Российской Федерации с предложением стимулировать проведение междисциплинарных исследований искусственного интеллекта, направленных на изучение и поддержку человека, научное освоение социального взаимодействия в высокотехнологизированной сфере его обитания в рамках государственных заданий для научно-исследовательских институтов и высших учебных заведений, а также комплексных проектов. Эти исследования должны быть направлены на разработку фундаментальных основ создания и использования технологий искусственного интеллекта в социальной практике, а также на выявление для человека, общества и государства рисков и создание правовых, этических, психологических и социологических механизмов смягчения этих рисков;

3.2. Министерство науки и высшего образования Российской Федерации, Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации, Министерство промышленности и

торговли Российской Федерации и Министерство экономического развития Российской Федерации с предложением по созданию инфраструктуры для сбора и хранения данных для машинного обучения в сфере поведения человека, а также для взаимодействия между научно-исследовательскими центрами, вузовской наукой и индустриальными партнерами;

3.3. Российский научный фонд и Фонд содействия развитию малых форм предприятий в научно-технической сфере с предложением о подготовке целевых грантовых программ, направленных на поддержку фундаментальных исследований в сфере искусственного интеллекта для изучения и поддержки человека, обеспечения его безопасности и создание в этих целях малых инновационных компаний;

3.4. Министерство науки и высшего образования Российской Федерации с предложением разработать программы междисциплинарной аспирантуры и магистратуры для подготовки квалифицированных кадров, способных проводить научно-исследовательские и опытно-конструкторские работы в сфере искусственного интеллекта, направленные на изучение, поддержку и защиту человека в условиях его взаимодействия с технологиями искусственного интеллекта.

4. Отделению математических наук РАН совместно с Отделением нанотехнологий и информационных технологий РАН до 12 апреля 2024 г. разработать и представить в установленном порядке в президиум РАН во исполнение поручений Президента Российской Федерации В.В. Путина от 17 января 2024 г. № Пр-83 по итогам Конференции «Путешествие в мир искусственного интеллекта» 24 ноября 2023 г. проект обращения в Правительство Российской Федерации с предложением:

4.1. создать Репозиторий открытых доверенных решений искусственного интеллекта (далее – Репозиторий) для депонирования результатов исследований при реализации совместных проектов с исследователями из дружественных стран, а также формирования и развития международных сообществ доверенных решений в области искусственного интеллекта. Такой Репозиторий должен обеспечивать: поддержку непрерывной разработки ИИ-решений с требуемым уровнем доверия (в том числе CI/CD инструменты); поддержку отбора и развития базовых, открытых

проектов мирового уровня; апробацию лучших практик и конвейеров создания доверенных систем ИИ с учетом специфик различных прикладных областей;

4.2. инициировать запуск совместных (двух- и многосторонних) проектов с участием дружественных стран в области систем искусственного интеллекта, в том числе: развитие открытых проектов мирового уровня (как отечественных, так и международных), критически необходимых для обеспечения технологической независимости РФ и дружественных стран, создания и развития отраслевых больших генеративных моделей, создания и развития систем ИИ с заданным уровнем доверия (гранты РНФ, Фонд содействия инновациям, в рамках Межгосударственных соглашений и др.) с широким вовлечением филиалов российских ВУЗов за рубежом и славянских университетов в странах ЕАЭС, СНГ, БРИКС с обязательным использованием экосистемы, создаваемой в рамках Репозитория, разработку регуляторных механизмов, обеспечивающих полноценную реализацию полученных при выполнении проектов результатов.

5. Считать целесообразным создать при президиуме РАН Научный совет РАН для решения научных проблем обеспечения суверенитета страны в области вычислительных и информационных технологий, включая технологии ИИ.

6. Контроль за выполнением настоящего постановления возложить на вице-президента РАН академика РАН Панченко В.Я. и вице-президента РАН академика РАН Макарова Н.А.

Президент РАН
академик РАН Г.Я. Красников
Протокольный
отдел
Главный ученый секретарь
президиума РАН
академик РАН М.В. Дубина