Проект 21-514-92001 ВАОН_а

Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в Северо-Восточной Азии (двухгодичный проект)

Актуальность проекта

Технологии искусственного интеллекта (ИИ), при всей своей высокой значимости для общественного развития, поднимают угрозы национальной и международной информационно-психологической безопасности (ИПБ) на новый уровень. Экономические трудности, социальные противоречия, политические конфликты, активизация военно-политического присутствия США и некоторых других стран НАТО в регионе, создают объективную основу для злонамеренного использования искусственного интеллекта (ЗИИИ) в Северо-Восточной Азии (СВА). На этом фоне резко возросли темпы развития технологий ИИ в странах региона, особенно в Китае, Японии и Южной Корее, что, при всей прогрессивности этих достижений, создает и новые вызовы ИПБ, требующие своевременного ответа со стороны государственных и негосударственных, национальных и международных структур и институтов.

Основные полученные результаты и их практическое значение

Итоговый отчет по НИР по проекту содержит:

- оценку существующих внутри- и межгосударственных конфликтов, наличия агрессивных антисоциальных акторов в СВА;
- анализ угроз злонамеренного использования ИИ для дестабилизации ИПБ;
- обзор существующих и прогноз будущих угроз злонамеренного использования ИИ по дестабилизации ИПБ в СВА с учетом политической ситуации и межгосударственных отношений, конфликтов в регионе;
- анализ значения политической ситуации в СВА и угроз злонамеренного использования ИИ для дестабилизации ИПБ в контексте национальной безопасности России;
- исследование методов нейтрализации угроз злонамеренного использования ИИ и обеспечения ИПБ в СВА, в том числе методов, применяемых государственными и общественными институтами;
- оценку значения изученного опыта для поддержания национальной безопасности России.

Установлено, что сохраняющийся между государствами СВА конфликтный потенциал и развивающееся межгосударственное соперничество, наличие противоречий у отдельных стран СВА с внерегиональными странами, существование «болевых» точек в развитии стран СВА (социально-экономические противоречия, вопросы территориальной целостности, другие факторы) обуславливают возможность использования против стран СВА инструментов высокотехнологичного информационно-психологического воздействия как внутренними, так и внешними антисоциальными акторами.

Среди наиболее распространенных форм и методов ЗИИИ в определении повестки дня в странах СВА отмечаются злонамеренное использование ботов, ранкинга и деранкинга, анализа настроений эмоционального ИИ, глубоких фейков (дипфейков), прогнозной аналитики и др. инструментов двойного назначения.

Результаты исследований в рамках настоящего проекта подтверждают, что новые угрозы возникают из-за преимуществ наступательных и оборонительных информационно-психологических операций с использованием ИИ. Эти преимущества растут с количественным и качественным совершенствованием традиционных механизмов производства, доставки и управления информацией; новыми возможностями информационно-психологического воздействия на людей.

В частности, эти преимущества могут включать более высокие:

(1) *объем* информации, который может быть сгенерирован для дестабилизации противника;

(2) *скорость* генерации и распространения информации;

(3) *возможности* для получения и обработки данных;

(4) *эффективность* прогнозной аналитики;

(5) *возможности* обеспечения процесса принятия решений;

(6) *новые способы* обучения людей;

(7) *силу* интеллектуального и эмоционального воздействия генерируемой информации на целевые аудитории;

(8) *уровень мышления с новыми качественными характеристиками* за счет создания общего и сильного искусственного интеллекта, а также за счет дальнейшего развития киборгизации человека.

Основываясь на качественной и, частично, количественной оценке данных из открытых первичных и вторичных источников сделан вывод, что преимущества 1-6 уже достигнуты и продолжают расти в ряде важных аспектов, хотя и не во всех, качественно превосходя человеческие возможности. В то же время все возможности узкого (слабого) ИИ, как правило, все еще находятся под контролем человека. Преимущества 7 практически не реализовано; но недавние успехи в решении широкого круга задач ИИ (модель Gato (создана в 2022 г. компанией DeepMind) способна одновременно решать свыше 600 разных задач и, в 450 из них, она превосходит способности экспертов-людей), а также формировании высокой эмоциональной убедительности ИИ (см., например, модель эмоционального ИИ Emma компании Ziva Dynamics), могут получить широкое распространение за счет количественного и качественного улучшения существующих технологий уже в обозримом будущем. Преимущество 8 требует фундаментальных научных прорывов и новых технологических решений. Данный перечень преимуществ использования ИИ в информационно-психологическом противоборстве не является исчерпывающим и меняется в процессе развития технологий ИИ.

Нарастание угроз ЗИИИ в СВА имеет прямое отношение к проблемам национальной безопасности России и Вьетнама. Поскольку часть России входит в этот регион, а Вьетнам соседствует с ним, оба государства имеют там важных партнеров и ярко выраженные интересы к экономическому и политическому сотрудничеству.

Как показал анализ угроз ЗИИИ, уже сегодня наносится заметный ущерб информационно-психологической безопасности, по крайней мере, ряду ведущих стран СВА. Это неизбежно наносит ущерб их политической и экономической стабильности. Потеря стабильности у соседей и партнеров России и Вьетнама неизбежно создаст серьезные проблемы и для последних.

Нельзя исключить использование ИИ для дискредитации внешнеполитических партнеров стран СВА (включая Россию и Вьетнам), для провоцирования межгосударственных конфликтов и противоречий. Здесь уже используется арсенал разных средств информационно-

психологического воздействия с привлечением ИИ в рамках технологий формирования информационной повестки дня.

Проведенный анализ открытых первичных и вторичных источников из всех стран региона (не были получены достаточные для обоснованных выводов данные из КНДР) показал отсутствие сформированного и научно обоснованного развернутого системного представления об угрозах ЗИИИ в контексте информационно-психологической безопасности. Наиболее далеко в понимании угроз ЗИИИ продвинулся Китай, но и там, пока, не встречается системное выделение угроз ЗИИИ национальной и международной информационно-психологической безопасности в качестве самостоятельного предмета изучения. Примечательно, что китайские СМИ (на основе данных Sputnik) отмечают вклад участников проекта в выделении угроз ЗИИИ информационно-психологической безопасности в качестве самостоятельного предмета изучения. Другие страны СВА демонстрируют более низкий уровень защиты от ЗИИИ в этой сфере.

ЗИИИ на сегодня стало значимым, но еще не главным инструментом подрыва информационно-психологической безопасности в СВА, уступая по результативности в целом хорошо известной и изученной совокупности форм и методов традиционной пропаганды (хотя и она все чаще не обходится без ИИ в качестве вспомогательного инструмента). Однако в недалеком будущем, за счет количественного и качественного совершенствования возможностей ИИ, дальнейшего его внедрения в различные сферы общественной жизни, ЗИИИ может выйти на первый план. Соответственно, Россия и Вьетнам должны активно развивать ИИ и стратегическое видение его использования с учетом растущих угроз ЗИИИ.

Представление результатов исследования в России и за рубежом

Комплексный анализ угроз ЗИИИ ИПБ и соответствующих мер противодействия в контексте СВА был представлен в рамках грантового проекта в России и за рубежом впервые (судя по открытым источникам).

Полученные результаты *были апробированы* на крупных международных научных конференциях широкого и отраслевого профиля, всего участники проекта с российской стороны представили *38 докладов (2 совместно с вьетнамскими коллегами)* на международных и российских научных конференциях, научных семинарах, прошедших в гибридном формате в России, Вьетнаме, Индии, Португалии, Польше, Франции (Первый Санкт-Петербургский конгресс исследователей международных отношений «Глобальные и региональные вызовы в меняющемся мире», 3-я Европейская конференция по влиянию искусственного интеллекта и робототехники, 9-я Европейская конференция по проблемам социальных сетей и др.). Было сформировано 5 профильных секций/ панелей на международных конференциях.

Выступление ведущего научного сотрудника Института актуальных международных проблем Дипломатической академии МИД РФ, д.и.н., проф. Е.Н. Пашенцева на четвертой международной конференция «Информация и коммуникация в цифровую эпоху: явные и неявные воздействия» 8 июня 2022 г. в Ханты-Мансийске.

Выступление Выступление ведущего научного сотрудника Института Европы, д.п.н., Базаркиной Д.Ю. 25 ноября 2021 г. в Институте актуальных международных проблем Дипломатической академии МИД РФ на круглом столе на тему «Злонамеренное использование искусственного интеллекта и международная информационно-психологическая безопасность».

Удалось привлечь в число экспертов для участия в международных опросах по проблеме угроз ЗИИИ ИПБ в СВА видных специалистов, чей h-index в Scopus доходит до 52 (профессор

Дон Донгхи Шин, https://www.scopus.com/authid/detail.uri?authorId=16242424900). Таким образом, результаты исследований по тематике проекта по значимости находятся на мировом уровне.

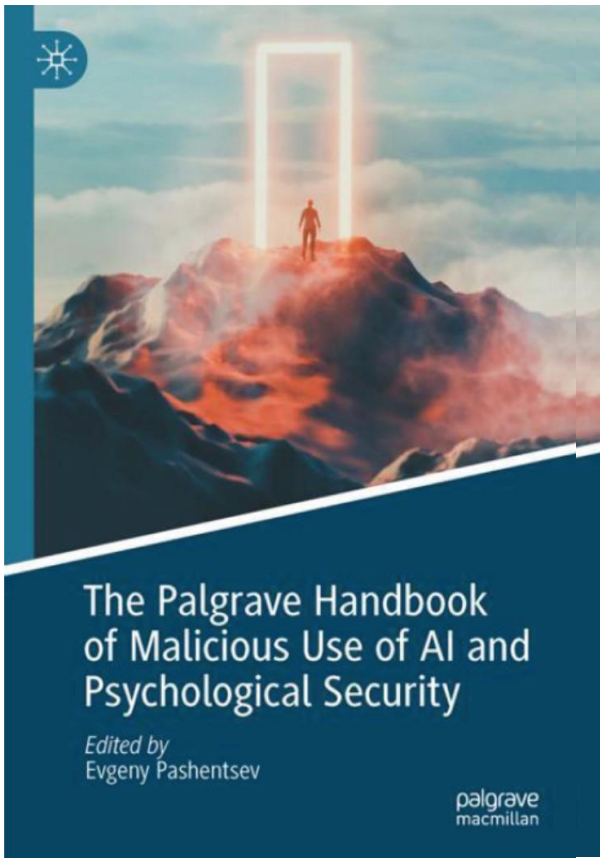
Российскими участниками проекта было опубликовано 29 статей в рецензируемых изданиях (из них 4 в соавторстве с вьетнамскими коллегами, 3 на русском, одна на вьетнамском), 3 публикации (главы международной монографии) приняты к печати. 6 статей проиндексированы в WoS/Scopus, три главы международной монографии будут направлены на индексацию в WoS/Scopus, подавляющее большинство других публикаций проиндексированы в РИНЦ или готовятся к индексации.

Взаимовыгодное сотрудничество с вьетнамскими коллегами

Задачи комплексного рассмотрения угроз ЗИИИ ИПБ были выдвинуты и получили рассмотрение в многочисленных публикациях российских участников настоящего проекта еще до начала его осуществления. Затем последовали первые совместные публикации с вьетнамскими коллегами, что логично подвело к разработке и выдвижению совместного проекта**.** В результате сотрудничества с участниками проекта с вьетнамской стороны во главе с д-ром А. Фаном, директором Центра японских исследований, а в дальнейшем заместителем генерального директора Института Индии и Юго-Западной Азии ВАОН, его коллегами по ВАОН, российскому коллективу удалось получить более комплексный анализ ЗИИИ в СВА. Знания и экспертиза вьетнамского участника проекта, кандидата технических наук Н. Дама добавили совместным исследованиям как более детальную проработку технической специфики злонамеренного использования искусственного интеллекта, так и региональную перспективу из государства, соседствующего с Северо-Восточной Азией. Ключевым совместным достижением стало включение в совместные исследования набора примеров регионального применения искусственного интеллекта, конкретизирующего формы его злонамеренного использования. Как отметил д-р Фан во вступительном слове ко второму докладу по итогам международного опроса экспертов: «Исследовательский проект “Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в Северо-Восточной Азии” (21-514-92001), финансируемый Российским фондом фундаментальных исследований (РФФИ) и Вьетнамской академией общественных наук (ВАОН), позволил получить важные результаты о влиянии развития индустрии искусственного интеллекта на экономические, социальные и политические конфликты в Северо-Восточной Азии. Это важное исследование для оценки ЗИИИ». Стоит отметить участие в отдельных аспектах реализации проекта вьетнамских коллег не из числа его формальных участников, которые приняли участие в экспертных опросах, подготовке главы в международную монографию, что, значительно расширило базу взаимовыгодного сотрудничества российских и вьетнамских исследователей по данной важной проблеме.

Внедрение результатов исследования

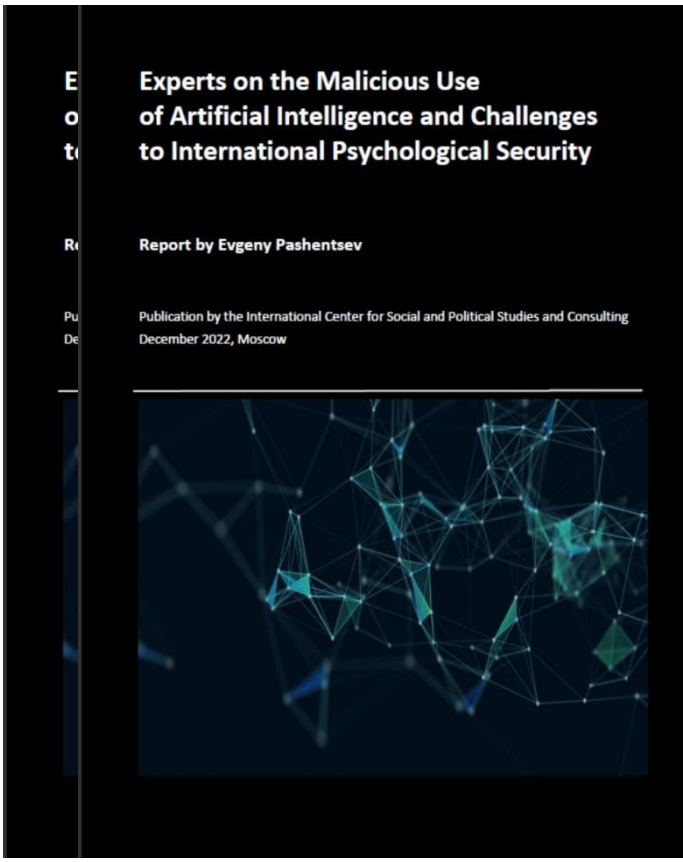
Результаты совместных исследований удалось внедрить в работу международной организации: коллективом проекта были предоставлены рекомендации для итогового документа конференции “Accelerating Actions and Promoting Digital Wellness (DW) in the context of Artificial Intelligence (AI)”, организованной под эгидой Межправительственной программы ЮНЕСКО «Информация для всех» (Хайдарабадская декларация). Часть рекомендаций, сделанных в докладах Е. Н. Пашенцевым и Д. Ю. Базаркиной, вошла в финальный вариант Хайдарабадской декларации: http://cdltr.uohyd.ac.in/downloads/



На основе исследовательских наработок, полученных в ходе реализации проекта, его участниками с российской стороны были подготовлены три главы по ЗИИИ в СВА со ссылкой на поддержку РФИИ (не считая др. глав) для научного коллективного труда «The Palgrave Handbook of Malicious Use of AI and Psychological Security», с участием специалистов из 11 стран, включая Беларусь, Бельгию, Великобританию, Вьетнам, Индию, Канаду, Китай, Россию, Румынию, США, Францию, под редакцией руководителя настоящего совместного проекта с российской стороны https://www.amazon.com/Palgrave-Handbook-Malicious-Psychological-Security/dp/3031225511 https://link.springer.com/book/9783031225512.

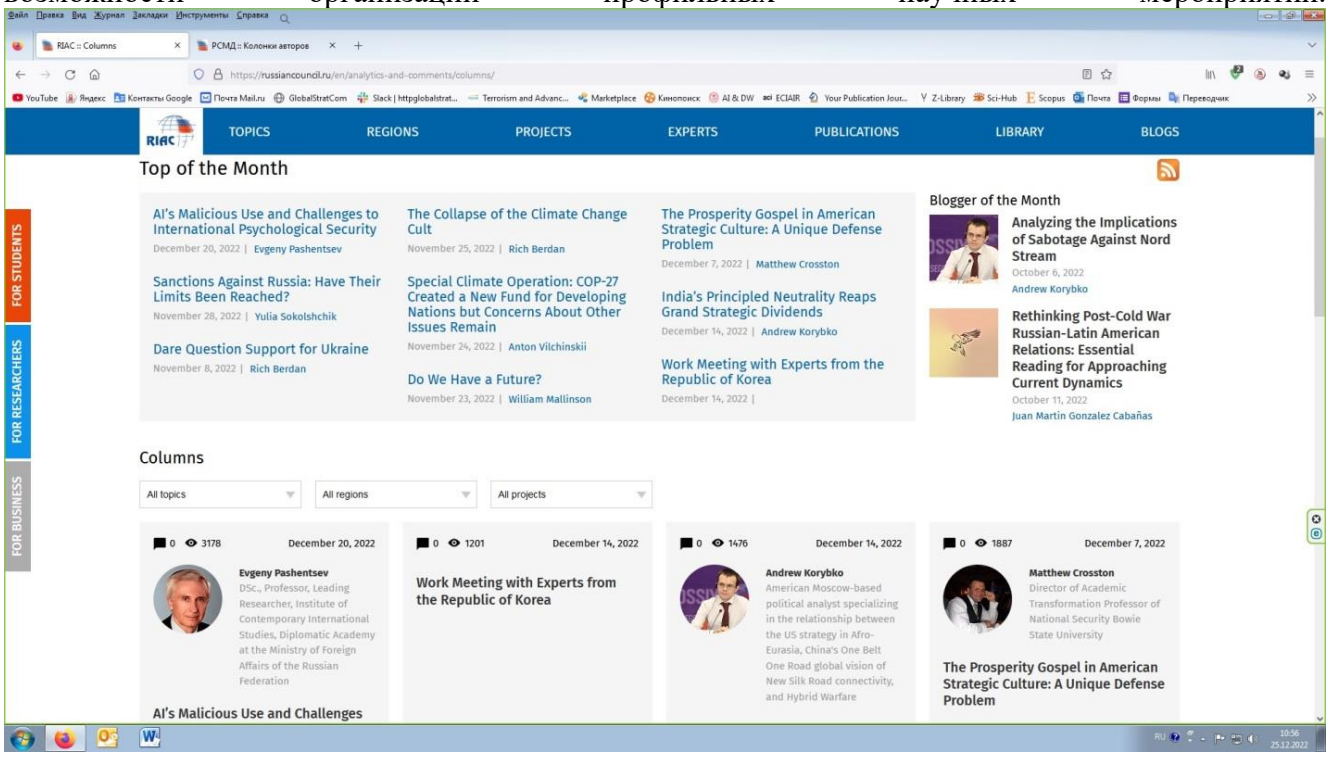
Выпуск книги планируется Palgrave Macmillan в рамках популярной серии “The Palgrave Handbooks” в июне 2023 г.

Результаты НИР по проекту, в том числе, доклады по итогам экспертных опросов (*Pashentsev, 2021 and 2022* в которых в общей сложности приняло участие 27 исследователей из 14 стран, выход сборников по итогам международных семинаров молодых исследователей были широко освещены в СМИ и на сайтах профильных научных и научно-практических структур в России и других странах (к итоговому отчету приложено соответствующее приложение с гиперссылками). Вводная статья о выходе доклада по итогам второго экспертного опроса на английском языке вышла на сайте РСМД в декабре 2022 г. и стала лидером среди публикаций в колонке экспертов за ноябрь-декабрь всего за 5 дней с момента размещения.



Доклад на английском языке был подробно освещен на научных и новостных порталах Италии, Румынии, Великобритании. Это повышает возможность обращения специалистов к результатам НИР и непосредственно к коллективу для дальнейшего сотрудничества. Таким образом, растут возможности создания

исследовательских коллабораций, в том числе привлечения молодых исследователей, а также возможности организации профильных научных мероприятий.



В ходе подготовки и проведения уже упомянутых экспертных опросов и итоговых докладов по ним сформирована международная экспертная база, которая на начало января 2023 г. включает свыше 150 специалистов из 22 стран мира. В рамках проведения экспертного опроса и подготовки книги *"The Palgrave Handbook of Malicious Use of AI and Psychological Security"* установлен ряд коллабораций со специалистами в изучаемой области с опытом профильных исследований в США. В частности, экспертные оценки предоставили профессор журналистики и политических коммуникаций Бангорского университета (Великобритания) Виан Бакир и профессор в области проблем цифровизации Бангорского университета (с опытом работы в КНР) Эндрю МакСтэй; профессор Университета Зайда в Абу-Даби, заслуженный профессор Министерства образования Южной Кореи Донгхи Шин; группа исследователей по главе с профессором медиа, этики и технологий Азиатско-Тихоокеанского университета «Рицумейкан» (Япония) Питером Мантелло, группа исследователей во главе с профессором в области международных отношений Центра американских исследований Фуданьского университета (КНР) Кихонг Цай, а также группа исследователей во главе с Арвиндом Гуптой – главой и соучредителем Digital India Foundation, который возглавлял кампанию в социальных сетях для премьер-министра Н. Моди во время выборов 2014 года, за что был удостоен звания цифрового лидера года и премии "Прорыв пути". Такое сотрудничество российских и зарубежных специалистов из разных стран по все более острой проблеме, имеющей глобальный охват и свое социально-политическое измерение в нынешней крайне сложной и опасной международной обстановке, трудно недооценить и оно должно быть продолжено в самых разных формах.