РОССИЙСКАЯ АКАДЕМИЯ НАУК

ИНСТИТУТ НАУЧНОЙ ИНФОРМАЦИИ ПО ОБЩЕСТВЕННЫМ НАУКАМ
РОССИЙСКАЯ АССОЦИАЦИЯ ПОЛИТИЧЕСКОЙ НАУКИ

Политическая наука 3 2021

POLITICAL SCIENCE (RU)

Москва
2021

ТЕМА НОМЕРА: ЦИФРОВИЗАЦИЯ ПОЛИТИКИ

СОДЕРЖАНИЕ

ИДЕИ И ПРАКТИКА

А.В. ТУРОБОВ, М.Г. МИРОНЮК*

ЭМПИРИЧЕСКАЯ МОДЕЛЬ АНАЛИЗА ДИНАМИКИ АЛГОРИТМИЗАЦИИ (ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА) В СФЕРЕ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ НА ПРИМЕРЕ США[1]

Аннотация. Как меняется система безопасности государства под влиянием технологии искусственного интеллекта (ИИ)? В статье предлагается эмпирическая модель оценивания системы безопасности государства (на примере США) на основе показателя согласованности (консистентности) безопасности, измеряющего то, как государство оценивает угрозы (показатель угроз), а также наличный уровень возможностей для их отражения (показатель возможностей) применительно к технологиям искусственного интеллекта. Показатель согласованности (консистентности) безопасности, по сути, описывает то, как государство способ-

* **Туробов Алексей Владимирович**, научный сотрудник Факультета социальных наук, преподаватель Департамента политики и управления, аспирант аспирантской школы по политическим наукам, Национальный исследовательский университет «Высшая школа экономики» (Москва, Россия), e-mail: aturobov@hse.ru; **Миронюк Михаил Григорьевич**, кандидат политических наук, доцент Департамента политики и управления, Национальный исследовательский университет «Высшая школа экономики» (Москва, Россия), e-mail: mmironyuk@hse.ru

[1] Исследование выполнено при финансовой поддержке РФФИ и ЭИСИ в рамках научного проекта № 20–011–31658. Авторы выражают благодарность анонимным рецензентам за критику, а также В.В. Кабернику (МГИМО МИД России) и Д.В. Стефановичу (ИМЭМО РАН имени Е.М. Примакова) за поддержку и ценные предложения.

 DOI: 10.31249/poln/2021.03.04

но оценивать угрозы (показатель угроз), и отвечает на вопрос, обладает ли государство необходимым уровнем возможностей для их отражения (показатель возможностей).

Разработанная модель (а также концептуализация технологии искусственного интеллекта в контексте сферы обеспечения безопасности) предоставляет свидетельства того, «как» происходят изменения в сфере обеспечения безопасности при помощи эмпирической модели, и представляет собой инструмент для изучения соответствующих изменений и оценки системы обеспечения безопасности государства. Необходимо обозначить ограничение исследования: мы не рассматриваем непосредственные военные приложения в области автоматизации и алгоритмизации (технологии искусственного интеллекта).

Валидизация представленной эмпирической модели проводилась на кейсе США (анализу подлежат восемь временных промежутков, а именно: 1999, 2002, 2006, 2010, 2012, 2015, 2017, 2019 гг.). Примечательно то, как меняются с развитием самой технологии «заинтересованность» государства и определение угроз, а также стремительный рост возможностей технологии искусственного интеллекта (в годы максимального прогресса в вычислительных мощностях и появления новых алгоритмов), а с 2012 г. динамика – более поступательная, так как все новые и новые «открытия» имеют скорее не революционный, а эволюционный характер.

Отличительной чертой модели является ее масштабируемость, которая выражается в возможности замены технологии искусственного интеллекта на любой другой тип цифровых технологий. В результате появляется возможность прослеживать динамику изменений системы безопасности государства, но уже применительно к иным технологиям, а затем проводить сравнительный анализ стран.

Ключевые слова: безопасность; исследования безопасности; модель; искусственный интеллект; оценка угроз; система безопасности государства.

Для цитирования: Туробов А.В., Миронюк М.Г. Эмпирическая модель анализа динамики алгоритмизации (технологии искусственного интеллекта) в сфере обеспечения безопасности на примере США // Политическая наука. – 2021. – № 3. – С. 72–111. – DOI: http://www.doi.org/10.31249/poln/2021.03.04

Введение

Футурологи, публицисты и, наконец, исследователи все больше утверждают (правда, первые скорее «предвидят» и, возможно, стараются предупредить), что технологии, обозначаемые понятием «искусственный интеллект», меняют привычный уклад во внутренней и международной безопасности, трансформируют сферу обеспечения безопасности и, вероятно, требуют иного подхода ко всей системе безопасности.

Но как именно меняется система безопасности государства под влиянием распространения «искусственного интеллекта» (ИИ)?

Некоторые исследователи высказывают опасения, что расширение возможностей обычных вооружений при помощи искусственного интеллекта усугубляет риск непреднамеренной эскалации, вызванной объединением ядерного и стратегического неядерного оружия и возрастающей скоростью войны, тем самым подрывая стратегическую стабильность и увеличивая риск ядерной конфронтации [Johnson, 2020]. Также указывается, что передовые возможности, такие как автономные рои дронов и гиперзвуковое оружие, управляемые ИИ, будут иметь дестабилизирующее воздействие на стратегическом уровне конфликта и более того – значительно увеличат скорость и темп ведения боевых операций [Ibid., p. 29–30]. С прагматической точки зрения, ИИ интегрируется в выполнение боевых задач с целью улучшения знаний об оперативной обстановке, о возможностях противника, а также обеспечивает скорость и точность наступательных и оборонительных вооружений [Davis, 2019]. Специальный выпуск журнала «Диалог о безопасности» (Security Dialogue) за 2017 г. полностью посвящен проблематике ИИ, подводя читателей к мысли о росте неопределенности будущего (как бы людям ни хотелось обратного), причем в сфере безопасности, а также поднимает проблемы принятия решений в сфере обеспечения безопасности в условиях как раз еще большей неопределенности[1] [Amoore, Raley, 2017].

Другая сторона ИИ в сфере обеспечения безопасности рассматривается М. Фергюсоном (через призму проблемы «черного ящика» алгоритмов). Исследователь утверждает, что военное командование должно быть сосредоточено на информировании высших руководителей оборонных ведомств и политиков о том, чего НЕ может сделать ИИ и что может пойти не так [Ferguson, 2019, p. 141]. Иными словами, важно помнить, что сколько возможностей предлагает ИИ, столько же ответственности эти технологии влекут за собой [Ibid., p. 141].

В целом результаты академических и прикладных исследований и в России[2], и за рубежом указывают на определенные из-

[1] Показательно, что неопределенность, по версии авторов, связана именно с алгоритмами, как будто решения, принимаемые людьми, – исключительно рациональны, прозрачны, понятны и существуют в условиях определенности.

[2] Обращает на себя внимание то, что в «Энциклопедии РВСН» на официальном сайте Министерства обороны РФ присутствует статья «Искусственный интеллект в военном деле», которая определяет ИИ как область исследований, но

менения (а также фиксируют обеспокоенность или даже тревогу, связанную с такими изменениями) в сфере безопасности под влиянием технологий ИИ. Однако мало исследований фокусируются на рассмотрении того, КАК именно технологии ИИ проникают в сферу обеспечения безопасности, а также КАКИЕ изменения происходят в самой системе безопасности государства. Практически отсутствует понимание того, какой аналитический инструментарий (метод / подход) позволяет проследить такие изменения с возможностью оценки последствий для системы безопасности. Безусловно, экспертные интервью предоставляют некоторую «почву» для исследований, однако попытки «объективного» взгляда на систему безопасности посредством количественной методологии весьма редки. Не преуменьшая значение экспертного знания, мы стремимся его дополнить более верифицируемыми свидетельствами.

Данная статья предоставляет определенные свидетельства того, КАК происходят изменения в сфере обеспечения безопасности при помощи эмпирической модели (инструмента) изучения изменений и оценки системы безопасности государства – на примере США. В работе предлагается эмпирическая модель оценивания системы безопасности государства на основе показателя согласованности (консистентности) безопасности, определяемой как отношение индикатора возможностей технологий ИИ и оценки угроз. Модель согласованности (консистентности) безопасности показывает, как государство способно оценивать угрозы (показатель угроз) с последующим публичным их определением в официальных документах, и определяет то, обладает ли государство необходимым уровнем возможностей для их отражения (показатель возможностей).

Структуру статьи можно представить следующим образом: сначала дается обзор эволюции сферы обеспечения безопасности с целью «локализации» технологий ИИ в ней. Далее мы определяем понятие «искусственный интеллект» (для целей данного исследования). Затем описывается эмпирическая модель, данные и общий дизайн исследования. Учитывая, что речь идет о «системе безо-

---

при этом отмечает реальность внедрения достижений исследований, причем не только для создания «интеллектуальных систем и образцов вооружения (в частности – их бортовых систем управления) различного назначения». Подробнее см.: https://encyclopedia.mil.ru/encyclopedia/dictionary/details_rvsn.htm?id=13200@morfDictionary.

пасности», мы рассматриваем два ключевых элемента – возможности технологии и оценка угрозы. Далее приводятся результаты применения модели на примере системы обеспечения безопасности США и, наконец, представляются некоторые выводы нашего исследования.

Важное ограничение – мы не фокусируемся на военно-техническом измерении ИИ. Мы не фокусируемся на многолетнем развитии автоматизированных систем (управления) военного назначения, в том числе бортовых систем и (или) систем управления войсками, поддержки принятия решений и т.п. Мы не оцениваем наличие или отсутствие влияния ИИ на концептуальную составляющую военных технологий и ведения войн. Мы фиксируем, что дискуссии на этот счет идут и далеки от завершения, и предложения авторитетных исследователей и практиков (чаще – отставников) включить дискуссию о перспективах ИИ в контексте ядерного оружия в переговоры по разоружению не утихают, а скорее становятся громче.

Эволюция сферы обеспечения безопасности

Сфера обеспечения безопасности не статична. Наполнение самих конструктов «безопасность» и «угроза» развивается, и границы / сферы распространения национальной безопасности также трансформируются. Традиционно обеспечение безопасности рассматривалось в военных категориях [Wright, 1942; Mead Earle, 1944; Bayley, 1975; Paret, 1986; Baldwin, 1995; Parker, 1996]. Угрозами безопасности воспринимались только те угрозы, которые непосредственно находились в плоскости военной готовности / возможностей государств. Именно поэтому исследования войны и мира являются отдельным направлением академических исследований. Первично исследования были направлены на триаду: State Power – War – Strategy.

После Первой мировой войны зрелости достигает концепция коллективной безопасности [Kennedy, 1987], хотя практики коллективной безопасности (в рамках узких альянсов) – это существенно более раннее явление. Коллективная безопасность не идентична международной безопасности. Коллективная безопасность распространяется только на тех акторов (государства), которые взяли на себя от-

ветственность за поддержание определенного уровня безопасности (по международным соглашениям и договорам). После Второй мировой войны [Wolfers, 1952; Buzan, 1991; Baldwin, 1997] концепция безопасности постоянно модернизировалась, так как сам факт мировой войны продемонстрировал несостоятельность действующих ранее подходов к обеспечению безопасности, и приверженность коллективной безопасности великих держав и даже сверхдержав (с созданием соответствующей системы) становится доминирующей практикой, по крайней мере публично.

Для нас существенно, что изучение опыта холодной войны способствовало входу в широкий оборот с 1991 г. теории секьюритизации [Buzan et al., 1998; Balzacq, 2011] и секторального подхода к безопасности [Buzan, 1991; Buzan, Waever, 2003; Brauch et al., 2008; Floyd, Matthew, 2013; Hanlon, Christie, 2016; Neack, 2017], которые обычно атрибутируются Копенгагенской школе безопасности. Данный подход расширил классические представления о секторах безопасности (не только военная сфера, не только суверенитет государства) и включил экономический, экологический, социальный и политический сектора. Несмотря на критику и объективные недостатки заявляемого подхода (см., например, тезисы Парижской школы безопасности), расширение секторов продолжается. Например, информационная безопасность [Nance, Straub, 1988; Alter, Sherer, 2004; Deshmuhk, Ashutosh, 2004; Bulgurcu et al., 2010; Whitman, Mattord, 2011] (информационный сектор) выделяется не просто в отдельный сектор, а пронизывает все сектора безопасности и, по сути, является одним из ключевых. Точно также выделяется кибербезопасность, которая активно «проникает» во все сектора безопасности.

Краткое представление эволюции сферы обеспечения безопасности необходимо для фокусировки понимания безопасности данного исследования. Мы не отождествляем военную безопасность (отсутствие угроз и вызовов военного свойства) с безопасностью государства как таковой. Также следует отметить и специфику рассматриваемого нами государства – США, успешно применяющего логику секьюритизации для сохранения статус-кво в мировой политике.

Соответственно, данное исследование находится в предметном поле исследований безопасности (security studies) и подчинено его логике. Мы умышлено не учитывали какие-то конкретные

военные технологии и / или специфичные системы вооружений и т.д., потому что нас интересует непосредственно «слепок» состояния дел здесь и сейчас – в сфере обеспечения безопасности конкретной страны. Безусловно, все указанные предметные сферы взаимосвязаны, как, собственно, и все в социальных науках. Именно поэтому мы концептуализируем и операционализируем как «безопасность», так и саму технологию ИИ[1], и формируем эмпирическую модель по доступным и верифицируемым данным.

Что такое «искусственный интеллект»?

В рамках данного исследования необходимо (1) установить единообразие в понимании технологий ИИ и (2) концептуализировать технологии ИИ относительно сферы обеспечения безопасности. Решение обеих задач подразумевает первичный анализ академической литературы для получения оптимальной дефиниции с последующей актуализацией определения на основании отчетов (policy reports) и нормативно-правовых актов в сфере обеспечения безопасности.

Существуют разнообразные концепции понимания ИИ, но в широком смысле данный вид цифровых технологий определяют как «интеллектуальные системы со способностью думать и учится» [Russel, Norvig, 2010]. По сути, это разнородный набор инструментов, методов и конкретных алгоритмов [Jarrahi, 2018]. Различные приложения и методы – от нейронных сетей (и моделей глубокого машинного обучения) до распознавания речи и / или образов и генетических алгоритмов, обработка естественного языка и машинное зрение – объединены «зонтичным» понятием технологий ИИ, на что обращает внимание государство (и государственное управление) [Reis, Santo, Melao, 2019]. Искусственный интеллект также определяется как система, способная независимо интерпретировать внеш-

[1] В качестве эпиграфа этого раздела можно привести отрывок небольшой статьи: «По мере того, как нефть вносит существенный вклад в великие геополитические конфликты XX века, политика данных и искусственного интеллекта быстро станет политикой международной безопасности…». Подробнее см.: Winning the AI Revolution for American Diplomacy // Warontherocks. – 2020. – Mode of access: https://warontherocks.com/2020/11/winning-the-ai-revolution-for-american-diplomacy/ (accessed: 01.04.2021).

ние данные и учиться на них для достижения конкретных результатов посредством гибкой адаптации [Kaplan, Haenlein, 2019].

В свою очередь, некоторые исследования указывают, что «искусственного интеллекта не существует» (например, [Galanos, 2018] или [Edwards, 1997], ИИ и роботы являются историческими конструкциями). Рассматривая ИИ (1) как отдельную категорию интеллекта, которая отличается от «естественного» интеллекта и (2) что интеллект является неотъемлемым свойством физических субъектов (например, людей) или объектов (например, роботов) – демонстрируется, что ни первое, ни второе невозможно в рамках привычной социальной философии, так как интеллект – системное явление, а не свойство отдельной единицы [Longino, 2014]. В рамках исследований в социальных науках вполне допустимо существование «зонтичного» понятия технологий ИИ, определяющего совокупность подходов, инструментов и алгоритмов, так как основное «преломление» ИИ – социальное воздействие и социально-политические эффекты. Иными словами, для исследований социальных и политических процессов не имеет значения, насколько «искусственен» или «естественен» интеллект и каков он сам [Vallverdu, 2017], если рассматривать феномены через понимание интеллекта как основного явления с людьми и машинами в качестве его агентов [Galanos, 2018, p. 362].

Несмотря на сложность в понимании и содержательном насыщении технологии ИИ, государства вполне «успешно» вступили в новую «гонку вооружений» в отношении технологий ИИ [Horowitz, 2018; Sharre, 2019; Brose, 2019]. Более того, технологии ИИ по потенциалу и возможностям сравниваются с ядерным оружием [Payne, 2018], и даже ведутся дискуссии о допустимости применения ИИ в области ядерного сдерживания [Johnson, 2020]. Если учесть все разработки и практические результаты технологии ИИ в сферах здравоохранения, логистики, транспорта, коммуникации, экономики и т.д., совсем не удивительно, что правительства разных стран уделяют все больше внимания ИИ. Уже более 20 стран имеют национальные программы в сфере цифровизации и искусственного интеллекта[1], причем одними из последних (по

[1] Building an AI World: Report on National and Regional AI Strategies // Cifar. – 2018. – Mode of access: https://www.cifar.ca/cifarnews/2018/12/06/building-an-ai-world-report-on-national-and-regional-ai-strategies (accessed: 01.04.2021).

времени) в соперничество за первенство в сфере цифровизации и искусственного интеллекта включились США: в январе 2019 г. президентом США Д. Трампом подписана соответствующая национальная программа развития и внедрения ИИ в экономику, сферу безопасности и социальную сферу.

Согласно отчету «Building an AI World»[1], национальные стратегии ИИ включают как минимум восемь направлений (областей) государственной политики: (1) исследования, (2) развитие талантов, (3) навыки и трудовое будущее, (4) индустриализация технологий искусственного интеллекта, (5) этические стандарты ИИ, (6) данные и цифровая инфраструктура, (7) ИИ в правительстве, (8) инклюзивность и социальное благополучие.

Правительства разных стран используют разные подходы для развития технологий ИИ. Общим является то, что ИИ включается в число ключевых факторов развития и конкурентоспособности государств. Что еще более важно, это один из самых быстрорастущих источников международной конкуренции (как между государствами, так и между компаниями).

Тематическое ядро проблематики алгоритмизации и ИИ рассматривает следующие вопросы:

– применение усложненных алгоритмов в публичном секторе политики [Mikhaylov, Esteve, Campion, 2018];

– допустимость и применимость отправления правосудия при помощи алгоритмов [Ang, Goh, 2013; Aletras et al., 2016; Chen, Eagel, 2017; Martin Katz et al., 2017; Liu, Chen, 2018];

– алгоритмический процесс принятия политических решений[2] [Kissell, Malamut, 2005; Zarsky, 2016; Coglianese, Lehr, 2017; Tene, Polonetsky, 2018];

– проблемы «черного ящика» алгоритмов и этико-ценностная проблематика [Keskinbora, 2019; Williams et al., 2016; Martin, 2018; Brożek, Janik, 2019], включая предвзятость (bias), «уравнивание», а именно – невозможность эффективного анализа

[1] Building an AI World: Report on National and Regional AI Strategies. – 2018. – Mode of access: https://www.cifar.ca/cifarnews/2018/12/06/building-an-ai-world-report-on-national-and-regional-ai-strategies (accessed: 01.04.2021).

[2] Perspective on Issues in AI Governance report from Google. – Mode of access: https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf (accessed: 30.04.2021).

уникальных ситуаций; отсутствие морально-нравственных ориентиров [Mittelstadt et al., 2016] и пр.;

– применение алгоритмов в военных целях [Payne, 2018 Ayoub, Payne, 2016; Zegart, Morell, 2019] и / или как инструмент сдерживания (примером может служить дискуссия об использовании ИИ в вопросах мониторинга и развития проектов в области ядерного вооружения и боевого дежурства ядерных систем);

– алгоритмические решения и искусственный интеллект в правоохранительной деятельности [Patil et al., 2014; Lum, Isaac, 2016], в обеспечении социальной безопасности [Devereux, Vincent, 2010] и прогнозировании преступности [Nakaya, Yano, 2010; McClendon, Meghanathan, 2015];

– ИИ как вопрос глобальной безопасности с учетом конкуренции между наиболее влиятельными и технологически продвинутыми странами[1].

Обращаясь непосредственно к специфике понимания ИИ в США[2], в первую очередь отметим инициативу администрации Президента Трампа «Искусственный интеллект для американского народа» (Artificial Intelligence for the American People[3]), которая целенаправленно определяет основные этапы и направления развития технологии. В свою очередь, план федерального участия в

[1] Artificial Intelligence Index 2018 Annual Report. – 2018. – Mode of access: https://hai.stanford.edu/ai-index-2018 (accessed: 05.05.2021); Report «Building an AI World: Report on National and Regional AI Strategies» // Cifar. – 2018. – Mode of access: https://www.cifar.ca/cifarnews/2018/12/06/building-an-ai-world-report-on-national-and-regional-ai-strategies (accessed: 01.04.2021).

[2] Для целей сравнения следует указать определение ИИ в Национальной стратегии развития искусственного интеллекта на период до 2030 г., утвержденной Указом Президента Российской Федерации от 10 октября 2019 г. № 490: «Искусственный интеллект – комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений». – Режим доступа: http://publication.pravo.gov.ru/Document/View/0001201910110003 (дата посещения: 30.04.2021).

[3] Artificial Intelligence for the American People. – 2019. – Mode of access: https://trumpwhitehouse.archives.gov/ai/ (accessed: 30.04.2021).

разработке технических стандартов и связанных с ними инструментов «Лидерство США в области ИИ»[1], разработанный Национальным институтом стандартов и технологий, определяет девять основных направлений стандартов ИИ: (1) Концепции и терминология; (2) Данные и знания; (3) Взаимодействие с людьми; (4) Показатели; (5) Сеть; (6) Методология тестирования производительности и отчетности; (7) Безопасность; (8) Управление рисками; (9) Надежность. Указывается, что «хотя определения ИИ различаются, для целей этого плана технологии и системы ИИ считаются включающими программное обеспечение и / или оборудование, которое может научиться решать сложные проблемы, делать прогнозы или выполнять задачи, требующие человеческого восприятия (например, зрение, речь и прикосновение), восприятие, познание, планирование, обучение, общение или физическое действие. Примеры разнообразны и быстро расширяются. Они включают в себя, помимо прочего, помощников ИИ, системы компьютерного зрения, биомедицинские исследования, системы беспилотных транспортных средств, передовое игровое программное обеспечение и системы распознавания лиц, а также применение ИИ как в информационных технологиях, так и в операционных технологиях»[2].

Отдельно следует отметить «Обзор средств управления некоторыми развивающимися технологиями» Бюро промышленности и безопасности от 19.11.2018[3], которым определяется, что к ИИ относятся такие технологии, как: (1) нейронные сети и глубокое обучение (например, моделирование мозга, прогнозирование

---

[1] U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools: Prepared in response to Executive Order 13859 Submitted on August 9, 2019. – 2019. – Mode of access: https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf (accessed: 30.04.2021).

[2] U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools. Prepared in response to Executive Order 13859 Submitted on August 9, 2019. 2019. – P. 7–8. – Mode of access: https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf (accessed: 30.04.2021).

[3] Review of Controls for Certain Emerging Technologies A Proposed Rule by the Industry and Security Bureau on 11/19/2018. – 2018. – Mode of access: https://www.federalregister.gov/documents/2018/11/19/2018-25221/review-of-controls-for-certain-emerging-technologies (accessed: 24.04.2021).

временных рядов, классификация); (2) эволюция и генетические вычисления (например, генетические алгоритмы, генетическое программирование); (3) обучение с подкреплением; (4) компьютерное зрение (например, распознавание объектов, понимание изображений); (5) экспертные системы (например, системы поддержки решений, системы обучения); (6) обработка речи и звука (например, распознавание речи и производство); (7) обработка естественного языка (например, машинный перевод); (8) планирование; (9) обработка аудио- и видеотехнологии (например, клонирование голоса, дипфейки); (10) облачные технологии ИИ или же (11) наборы микросхем ИИ. В данном обзоре технологии ИИ ставятся в один ряд с машинным обучением и нейронными сетями (как проявлением усложненных математико-статистических алгоритмов).

Относительно понимания ИИ непосредственно в сфере обеспечения безопасности, необходимо обратиться к отчетам Комиссии национальной безопасности по искусственному интеллекту США[1], в частности промежуточному докладу от ноября 2019 г.[2] в разделе «Что мы подразумеваем под “ИИ”?», который определяет, что «ИИ – это способность компьютерной системы решать проблемы и выполнять задачи, которые в противном случае потребовали бы человеческого интеллекта. Технологии искусственного интеллекта развивались на протяжении многих десятилетий, включая распознавание образов, машинное обучение, компьютерное зрение, понимание естественного языка и распознавание речи. Эти технологии используются для расширения возможностей людей и машин, помогая им принимать решения более высокого качества и с большей скоростью. В растущем, но все еще ограниченном наборе областей машины могут достигать показателей, подобных человеческим или превосходящих человеческие, путем анализа больших объемов данных, выявления закономерностей и выполнения массового поиска полезных ответов, оценок и рекомендаций. Эти системы совершенствуются по мере перехода от экспертных систем, основанных на

[1] The National Security Commission on Artificial Intelligence // The National Security Commission on Artificial Intelligence (NSCAI). – Mode of access: https://www.nscai.gov/home (accessed: 24.04.2021).

[2] Interim Report. – 2019. – Mode of access: https://drive.google.com/file/d/153OrxnuGEjsUvlxWsFYauslwNeCEkvUb/view (accessed: 24.04.2021).

явных моделях, к системам машинного обучения, которые могут учиться на опыте и повышать свою производительность, в том числе те, которые могут учиться на достаточно больших и надежных наборах данных. Это системы, предназначенные для решения задач и достижения определенных целей, с компетенциями, которые в некоторых отношениях параллельны когнитивным процессам человека: восприятию, рассуждению, обучению, общению, принятию решений и действию». Таким образом, можно заключить, что власти США (1) определяют технологию ИИ максимально широко (зачастую объединяя с машинным обучением и нейронными сетями), (2) отмечают эволюционную природу технологии (постоянное развитие и усовершенствование) и (3) стремятся типологизировать технологии ИИ.

Более предметное рассмотрение ИИ в контексте непосредственно действий (обороны, безопасности, военных операций) представлено в отчете Исследовательской службы Конгресса «Искусственный интеллект и национальная безопасность» от августа 2020 г.[1], где представлены различные подходы и вызванные таким разнообразием сложности, начиная от систем с когнитивными функциями как у людей и заканчивая автоматизированным вооружением. Однако сам отчет содержит идентичное, с перечисленными выше, широкое определение ИИ, исходя из разнообразия практического (прикладного) применения технологии.

Завершая концептуализацию технологий ИИ, следует обратить внимание на отчет ООН «Милитаризация искусственного интеллекта» (2019)[2], где указывается, что ИИ не является единой технологией, а является скорее совокупностью теорий, методов, технологии и прикладных систем для стимулирования и расширения человеческого интеллекта. Отдельно рассматривается призма практического военного применения и влияния как на систему безопасности, так и на стратегию государств и международную стабильность.

---

[1] Artificial Intelligence and National security, august 2020. – Mode of access: https://fas.org/sgp/crs/natsec/R45178.pdf (accessed: 24.04.2021).

[2] The Militarization of Artificial Intelligence, August 2019. – 2019. – Mode of access: https://www.un.org/disarmament/the-militarization-of-artificial-intelligence/ (accessed: 24.04.2021).

В рамках данного исследования (включая единообразный сбор данных и методологическую валидность) *под технологией искусственного интеллекта будут пониматься алгоритмические и компьютерные системы (в том числе программное обеспечение и / или оборудование), которые, обучаясь, могут решать сложные проблемы, делать прогнозы или выполнять задачи, требующие человеческого восприятия, познавать, планировать, обучаться, общаться или совершать физическое действие, обязательно в сфере обеспечения безопасности или непосредственно в военной сфере*.

Еще раз повторим: в рамках данного исследования мы не фокусируемся на военно-техническом измерении ИИ, на многолетнем развитии автоматизированных систем (управления) военного назначения. Действительно, различные автоматизированные системы в военном деле успешно применяются многие десятилетия. Эти успехи обязаны своим существованием развитию кибернетики и соответствующим исследованиям конца первой половины – середины XX в. Однако нынешний этап после всех «зим» ИИ особый. Впервые в распоряжении и государств, и исследователей, и негосударственных акторов находятся вычислительные возможности и опыт, которых не было прежде. И более совершенные системы самонаведения, слежения и т.п. – это только часть приложения достижений последних десятилетий. Да, ИИ сохраняет специализацию (и, вероятно, будет сохранять), но уже сейчас он вышел за пределы задач дорогих и негибких АСУ, компьютеры «научились» «видеть, слышать, понимать» (при всей условности), справляться с беспрецедентно большими объемами информации. Иными словами, ИИ в сфере безопасности – это (уже) не только и не столько автоматика в сложных системах (кинетического) вооружения и управления войсками, которая по-прежнему вполне надежно обеспечивает скорость и точность ведения боевых действий.

Эмпирическая модель и данные

Формируя математическую модель, наш выбор индикаторов и параметров определяется акцентом на то, *что* поддается измерению [Fioramonti, Kononykhina, 2015, p. 476]. Работа с индикаторами и показателями, в свою очередь, подчинена необходимости соблюдать баланс между *полнотой* и *доступностью* данных [Ibid., p. 477].

Показатель согласованности (консистентности) безопасности (Security Consistency) формируется посредством разницы показателей угроз (Threats) и возможностей технологий ИИ (AI capability) реагировать на такие угрозы. Указанное можно представить в виде блок-схемы.

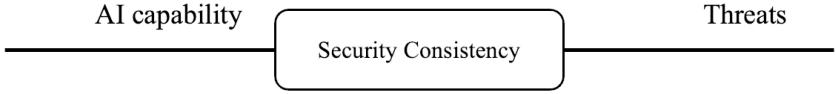

Рис. 1
Блок-схема показателя согласованности (консистентности) безопасности

Показатель согласованности (консистентности) безопасности будет, по сути, отражать то, как государство способно оценивать угрозы (показатель угроз) и обладает ли государство необходимым уровнем возможностей для их отражения (показатель возможностей).

Таким образом, формула (1) вычисления индикатора согалсованности безопасности имеет следующий вид:

$$Security\ Consistency = AI\ Capability - Threats\ Evaluation \quad (1)$$

Индикатор возможностей Искусственного интеллекта (AI capability). Вопросы измерений и оценки возможностей (capability) являются весьма актуальным направлением в политической науке (начиная от составления индексов государственной состоятельности (state capacity) и заканчивая исследованиями организационных возможностей (organizational capabilities)). По сути, все измерения «capability» в литературе представлены или различными регрессионными моделями для выявления связей индикаторов и их влияния на возможности (capabilities), или экспертными интервью [подробнее см. Grant, Verona, 2015, где проводится анализ основных эмпирических исследований и их проблемных зон в области «organizational capabilities»]. Отдельным направлением исследований с 2010-х годов является создание составных индексов «capabilities» на национальном и международном уровнях (например: Global capability index, Composite Index of National Capability).

В рамках данного исследования за основу взят методологический подход, используемый в «Global capability index» [Fioramonti, Kononukhina, 2015] и «Composite Index of National Capability»[1], который, в частности, учитывает расходы и финансирование военной сферы. Оба индекса предполагают формирование единого индикатора по стране, который представляет собой сумму показателей конкретных сфер, деленную на количество данных сфер. Например, «Global capability index» формируется на основе измерений трех областей (dimensions):

– социально-экономическая среда: образование, равенство и гендерное равенство, цифровое участие, инфраструктура коммуникационных технологий;

– социально-культурная среда: доверие, социальная толерантность, участие в коллективных действиях и т.д.;

– среда управления (Governance environment): индивидуальные и коллективные возможности социальной и политической активности, верховенство закона, политический диалог, нормативно-правовая база гражданских объединений и организаций и т.д.

Учитывая большой массив параметров в каждой области, финальный расчет индикатора «capability» представляет собой сумму показателей по областям, с учетом коэффициентов относительно количества каждого индикатора, деленной на три области.

Схожий подход и при измерении «Composite Index of National Capability». Только в данном случае высчитывается шесть параметров (а не три, как с «Global Capability Index»), они суммируются и делятся на количество параметров.

Исходя из концептуальной рамки понимания Искусственного интеллекта и целей создания показателя возможностей Искусственного интеллекта (AI capability), мы выделили четыре области / сферы:

1. Технологическая (Technological): учитывает технологические аспекты Искусственного интеллекта – показатели точности алгоритмов и результаты прохождения тестов (Turing Test, Loverace Test и т.д.), а также применение технологии в государственном управлении и сфере безопасности. Является отражением технологических возможностей технологии.

[1] Composite Index of National Capability // The Correlates of War Project. – Mode of access: https://correlatesofwar.org (accessed: 20.04.2021).

2. Экономическая (Economic environment): учитывает финансирование технологии ИИ в контексте общего военного бюджета. Мы понимаем, что сфера безопасности может иметь различные источники финансирования, включая засекреченные статьи бюджета, финансирование за счет иных статей и разделов и т.п., поэтому вынужденно опираемся на публичные данные о финансовой поддержке технологии в непосредственно в военном бюджете. Демонстрируются финансово-экономические возможности технологии; без финансирования и экономического стимулирования развитие технологии и особенно ее применение в сфере обеспечения безопасности маловероятно.

3. Управление (Governance environment): учитывает количество государственных компаний, связанных с технологией ИИ и существование правовой санкции на использование ИИ в военной сфере. Отражает готовность государства развивать технологию и применять возможности технологии.

4. Социальная (Social environment): учитывает занятость населения в сферах и областях разработки и применения технологии ИИ и количество стартапов, фокусирующихся на технологии. Отражает вовлеченность общественности, а также возможность государства мобилизовать высокопрофессиональные кадры.

Указанные области / сферы с показателями и их краткой характеристикой расчета указаны в табл. 1.

Агрегирование указанных областей / сфер является заключительным этапом формирования индикатора возможностей технологии ИИ (AI capability), на котором «взвешиваются» эти области: Технологическая и Экономическая получают вес «0.25» из 1, а область Управления – вес «0.3» в связи с социально-политической значимостью данной области в вопросах безопасности. По схожим мотивам, Социальная сфера имеет вес «0.2», несмотря на значимость гражданского общества и общественной реакции, в сфере обеспечения безопасности население «знает только то, что государство считает позволительным знать». Иными словами, роль общества в вопросах возможности технологии в сфере безопасности будет наименее значима на ряду с остальными областями / измерениями.

Таблица 1

Измерения возможностей ИИ

Область / сфера (dimension)	**Показатель**	**Краткая характеристика**
Технологическая (0.25)	UT (use of technology)	Применение ИИ в государственном управлении и военной сфере (*да / нет*)
	Test	Существование технологии ИИ, прошедшей тест (Turing Test, Loverace Test e.t.c.) (*да / нет*)
	AA (algorithm accuracy)	Медианный показатель «корректности» / точности алгоритмов[1] (Vision & Image Recognition; Visual Question Answering; Speech Recognition; Image Generation; Language Modelling and Comprehension; Translation; Conversation: Reading Comprehension)
Экономическая (0.25)	MF (military funding)	Военные расходы на ИИ / Общие военные расходы
Управление (0.3)	SC (state companies)	Государственные компании по ИИ (*да / нет*)
	LA (legal authorization)	Правовая санкция на использование ИИ в военной сфере (да / нет)
Социальная (0.2)	JO (job openings)	Занятость – открытые вакансии (Job openings) / количество работоспособного населения
	RS	Стартапы, связанные с искусственным интеллектом

В результате формула (2) вычисления индикатора AI capability имеет следующий вид:

$$AIcapability = \frac{0.25 \cdot \Sigma(UT, TEST, AA) + 0.25 \cdot MF + 0.3 \cdot \Sigma(SC, LA) + 0/2 \cdot \Sigma(JO, RS)}{4} \quad (2)$$

где первые три показателя относятся к Технологической сфере: *UT* – показатель применения / использования ИИ в государственном управлении и военной сфере, *TEST* – показатель наличия технологии ИИ, прошедшей тесты (Turing Test, Loverace Test e.t.c.), *AA* – показатель корректности / точности алгоритмов;

один показатель относится к Экономической сфере: *MF* – показатель финансирования ИИ в военной сфере;

два показателя относятся к области Управления: *SC* – показатель государственных компаний в сфере AI, *LA* – показатель наличия правовой санкции на применение технологии ИИ в военных целях;

[1] Measuring the Progress of AI Research. Project, metrics and datasets // Electronic Frontier Foundation. – Mode of access: https://www.eff.org/ai/metrics (accessed: 20.04.2021).

два показателя относятся к Социальной сфере: *JO* – показатель занятости в сфере ИИ, *RS* – показатель стартапов в сфере ИИ.

Индикатор оценки угроз. Оценке угроз (threat evaluation) посвящен отдельный раздел специализированной литературы, зачастую объединенный единым направлением – Оценка угрозы и распределение оружия (Threat Evaluation and Weapon Assignment – TEWA) (например, [Cocelli, Arkin, 2017; Johansson, Falkman, 2008; Naeem, Masood, 2010; Naseem et al., 2017; Kumar, Tripathi, 2016]). TEWA считается основным компонентом системы противовоздушной обороны (Air Defense system – ADS). В последнее время наиболее распространенными являются модели на основе байесовских сетей (Bayesian networks) [Kumar, Tripathi, 2016], нечеткой логики (fuzzy logic/ fuzzy inference rules) [Naeem, Masood, 2010; Johansson, Falkman, 2008] и систем поддержки принятия решений (decision support system) [Naseem et al., 2017].

TEWA-модели, основанные на байесовских сетях, позволяют преодолевать неопределенности (неполнота информации об объектах; отсутствие информации о состоянии инфраструктуры; вероятности и / или случайности в управлении конкретным вооружением и т.д.) при моделировании. В байесовском подходе переменные TEWA-модели содержат пределы вероятностей, или распределение вероятностей, что позволяет оценивать угрозы даже в случае неполноты данных.

В свою очередь, модели, основанные на правилах нечеткой логики (концепция нечетких множеств), строятся по принципу функций принадлежности (функций членства – membership function). В теории четких множеств члены x универсального множества X являются либо членами, либо не членами множества $A \subseteq X$. Таким образом, значения, присвоенные x, попадают в диапазон, указывая степень членства элемента в (нечетком) наборе, о котором идет речь. Большие значения указывают на более высокую степень членства, в то время как более низкие значения указывают на более низкую степень членства (degree of membership). Иными словами, для конкретного контекста может быть трудно определить четкие границы (показатели / параметры) переменной, поэтому используется функция членства, которая позволяет рассчитывать схожие, по степени членства, показатели переменной. Два важных замечания: (1) оценка членства (membership grades) в правилах нечетких множеств не относится к оценке вероятности

[Johansson, Falkman, 2008], (2) относительно самих правил нечеткой логики отсутствует единство в научной академической среде – часть исследователей в принципе не признают такой подход, считая его слишком абстрактным.

Применение систем принятия решений в TEWA-моделях позволяет учитывать показатели геоинформационных систем (ГИС картрирование уязвимых активов), дополнять модель методами прогнозирования, распределять и оценивать «экономически эффективное назначение оружия» [Naseem et al., 2017, p. 169]. Таким образом, система принятия решений позволяет расширить перечень параметров в модели, а также оценивать дополнительные факторы (например, экономическую целесообразность) при оценке угроз. Сами TEWA-модели с системой принятия решений могут строиться на основе машинного обучения (наиболее популярные модели с древом решений; более продвинутые модели основываются на глубоком обучении, например Tactical Air Combat Decision Support System), теории игр и динамических байесовских сетей.

Отметим, что модели TEWA подразумевают создание полноценной системы реагирования и противодействия угрозам, что выходит за рамки данного исследования. Таким образом, на основе имеющихся исследований по оценке угроз, в том числе с учетом масштабного мета-анализа (156 публикаций исследований TEWA с 1975 по 2016 г.), проведенного Насимом, Шахом, Ханом и др., разработана математическая модель расчета угроз.

Оценка угроз зачастую представлена двумя [Naeem, Masood, 2010] или тремя [Naseem et al., 2017] этапами. Двухэтапная модель подразумевает (1) оценку и ранжирование угрозы, и соответственное (2) назначение оружия (weapon assignment). Трехэтапная модель состоит из (1) оценки восприятия угрозы (threat perception), (2) расчета индекса угрозы (threat index calculation) и соответственно (3) назначения оружия. Каждая модель оценки угроз обязательно учитывает соответствие угрозы защищаемому объекту / активу (defended asset).

Каждый этап включает расчет конкретных характеристик [Naseem et al., 2017]. Так, на этапе *восприятие угрозы* рассчитываются критические параметры конкретного типа вооружения (например, крылатой ракеты): скорость, высота, радиолокационное поперечное сечение / эффективная поверхность рассеяния радио-

локационных волн (Radar cross-section), маневренность (maneuvering capability), угол пикирования (dive angle), атакующий подход (attack approach) и др. Расчет *индекса угрозы* включает характеристики скорости, направления / курса, высоты над уровнем моря, угрозу маневрирования, расстояние от уязвимых точек, угрозу летальности из базы знаний и пр. Этап *назначения оружия* непосредственно учитывет характеристики имеющегося обороняющего вооружения и содержит параметры: (а) угроза назначается оружию на основе индекса угрозы, (б) угроза с самым высоким индексом угрозы (TI) назначается первой, (в) угроза назначается оружию с наибольшей вероятностью убийства.

Обобщенно модель TEWA представляет собой оценку и ранжирование угрозы по перечисленным выше характеристикам и расчет соответствия оцененной угрозы защищаемым объектам / активам, с последующим определением вооружения для обеспечения безопасности объектам и нейтрализации угрозы. Непосредственно оценку угрозы (без распределения вооружения) можно представить в виде блок-схемы, где рассчитывается общее число угроз (при оценке и ранжировании) и выводится показатель соответствия угрозы защищаемому объекту / активу [Naeem, Masood, 2010].

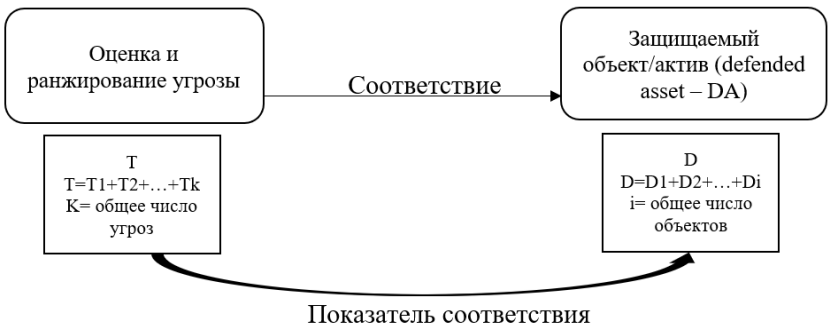

Рис. 2
Блок-схема обобщенной модели TEWA (оценка и ранжирование угроз)

Для целей данного исследования показатели восприятия угрозы (ТР) и характеристика / тип угрозы (ТТ) концептуализированы следующим образом (все характеристики представлены в виде прокси-показателей), указанным в табл. 2.

Таблица 2

Показатели восприятия угрозы и характеристики угроз

Критические параметры, используемые для оценки угроз	
Этапы	**Характеристики**
1	2
Восприятие Угрозы (*на основании стратегий и НПА*)	1. Показатель образования (education – Ed) – без обучения и переподготовки населения, чтобы соответствовать темпам технологических изменений и различным типам угроз, злоумышленники столкнутся с меньшим количеством препятствий при попытке использовать уязвимости ИИ[1]. Образованное население также уменьшит непреднамеренные ошибки. Таким образом, любое государство, указывающее на угрозы ИИ, на уровне политики адаптирует образовательную систему к повышению осознания технологии ИИ. В данном исследовании показатель образования будет рассчитываться исходя из показателей результатов страны по рейтингу PISA исключительно по направлению математики, так как математическое образование является фундаментальным для разработок и применения технологии ИИ. Расчет представляет собой соотношение рейтинга страны в конкретный год к максимально возможному показателю рейтинга[2]. 2. Показатель нормативного регулирования (regulation – Reg) – восприятие угроз со стороны государства отражается в сфере нормативного регулирования. Удельный вес нормативно-правовых актов, закрепляющих (в той или иной мере) угрозы со стороны технологии Искусственного интеллекта будет рассчитываться следующим образом: количество НПА с упоминанием технологии ИИ на общее количество НПА в год. 3. Показатель внутренних патентов (domestic patents – DP) – определение восприятия угроз со стороны государства невозможно без научно-технических исследований и разработок. Для расчета данного показателя будут учитываться только внутригосударственные патенты по тематике «Искусственный интеллект». Показатель будет представлять собой удельный вес количества внутренних патентов по тематике в соотношении с общим количеством патентов в рассматриваемый год.

Продолжение таблицы 2

1	2
Характер / тип угрозы (*на основании исследований, индексов и показателей отчетов, экспертных оценок*)	Данный показатель задумывался как демонстрация конкретного (и вычисляемого) показателя характера / типа угроз. К сожалению, на данный момент отсутствуют устойчивые статистические показатели непосредственно релевантных технологий ИИ[3]. Практически все тчеты и исследования несут «доктринальный» характер, а именно: указывают, что есть какой-то тип угрозы или ИИ обладает определенной характеристикой угрозы, однако никакого статистического или математического выражения нет. В связи с чем принято решение построения численного бинарного показателя: 0 – тип / характер угрозы отсутствовал в стране в рассматриваемом году; 1 – тип / характер угрозы присутствовал (зафиксирован / задокументирован) в стране в рассматриваемом году, согласно следующему перечню типов / характера ИИ угроз[4]: 1) угрозы критической инфраструктуре; 2) киберугрозы / кибератаки при помощи ИИ (ИИ расширяет векторы угроз, уязвимых для кибератак, обнаруживая и эксплуатируя слабые места системы); 3) кампании по дезинформации (в том числе Deepfakes); 4) нарушение прав человека (имеется в виду bias алгоритмов, нарушения персональных данных, угрозы биометрическим данным и т.д.). Соответственно указанному перечню, за каждый конкретный год будет формироваться показатель от «0» до «4». Если в исследуемом году были зафиксирован какой-то тип угроз – выставляется «1», если не было – «0».

[1] Например, AI Using Standards of Mitigate Risk. – Mode of access: https://www.dhs.gov/sites/default/files/publications/2018_AEP_Artificial_Intelligence.pdf (accessed: 20.04.2021).

[2] In each test subject, there is theoretically no minimum or maximum score in PISA; rather, the results are scaled to fit approximately normal distributions, with means for OECD countries around 500 score points and standard deviations around 100 score points. About two-thirds of students across OECD countries score between 400 and 600 points. Less than 2% of students, on average across OECD countries, reach scores above 700 points, and at most a handful of students in the PISA sample for any country reach scores above 800 points. – Mode of access: http://www.oecd.org/pisa/pisafaq/ (accessed: 20.04.2021).

[3] Рассматривались возможности расчетов в каждом типе угроз. Например, а разделе «дезинформация», несмотря на существование Глобального индекса дезинформации, исследования ЕС «Регулирование дезинформации с помощью искусственного интеллекта» и т.д. отсутствует какая либо статистическая база для расчета непосредственной роли / угрозы ИИ. Рассматривалась возможность расчета, исходя из индекса v-dem свободы прессы (например, показатель «freedom of Expression», или показатель «the media» (или даже отдельную субкатегорию «media bias» (v2 mebias)) и устойчивого показателя количества угроз ИИ. Предполагалось, что устойчивый показатель будет выявлен на основании анализа литературы по исследованиям взаимосвязи свободы прессы и количества сгенерированной

дезинформации при помощи ИИ. Однако таких исследований не было найдено (анализ проводился по базам Jstor, Science Direct и EBSCO).

В разделе «киберугроз», несмотря на наличие различных индексов киберугроз (например, Cyber Threat // Impreva. – Mode of access: https://www.imperva.com/cyber-threat-index/ (accessed: 05.05.2021); Cybern Risk // Trend Micro. – Mode of access: https://www.trendmicro.com/en_hk/security-intelligence/breaking-news/cyber-risk-index.html (accessed: 05.05.2021); и т.д.) и кибербезопасности (например, Global Cybersecurity Index v. 4 // The Telecommunication Development Sector (ITU-D). – Mode of access: https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx (accessed: 05.05.2021); IBM X-Force Threat Intelligence // IBM. – Mode of access: https://www.ibm.com/security/data-breach/threat-intelligence (accessed: 05.05.2021); Cybersecurity: Let's get tactical // TechRepublic. – Mode of access: https://www.techrepublic.com/resource-library/whitepapers/cybersecurity-let-s-get-tactical-free-pdf/?ftag=CMG-01-10aaa1b (accessed: 05.05.2021) и т.д.) невозможно выделить отдельно роль / значение ИИ.

Более сложная ситуация (в контексте наличия данных) с иными типами, например отсутствует возможность определить удельный вес / объем потенциально подвергаемых угрозой ИИ элементов критической инфраструктуры и т.д.

[4] Перечень сформирован на основании анализа профильных национальных НПА и отчетов (например, National Security Commission on Artificial Intelligence, Interim Report, 2019. – Mode of access: https://www.nscai.gov/reports (accessed: 05.05.2021); Artificial Intelligence and UK National Security: Policy Considerations, from RUSI, 2020 // RUSI. – Mode of access: https://rusi.org/publication/occasional-papers/artificial-intelligence-and-uk-national-security-policy-considerations (accessed: 05.05.2021); и т.д.), а также международных отчетов (например, Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It, from Harvard Kennedy School. Belfer Center for Science and International Affairs, 2019. – Mode of access: https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf) (accessed: 05.05.2021).

Расчет показателя восприятия угроз (TP), по сути, представляет собой сумму показателей прокси-образования (Ed), прокси-регулирования (Reg) и прокси-внутренних патентов (DP) и определяется по формуле (3) следующим образом:

$$TP = \sum(Ed, Reg, DP) \quad (3)$$

где: *Ed* – показатель образования;

Reg – показатель нормативного регулирования;

DP – показатель внутренних патентов.

Показатель типа / характера угроз (TT) будет принимать значение от «0» до «4» и может быть выражен следующей формулой (4):

$$T \in [0.1\ldots4] \quad (4)$$

Предлагаемый подход к оценке угроз будет основан на оценке Восприятия угрозы (TP) и оценке Типа угрозы (TT) в соотнесении с Защищаемыми объектами (DA). Восприятие угрозы (TP) относительно технологии Искусственного интеллекта основывается на анализе национальных стратегий по безопасности стран и нормативно-правовых актов в сфере безопасности. Тип угрозы (TT), в свою очередь, типологизирован из релевантного теоретического обзора, экспертных оценок, показателей и индексов из международных и национальных отчетов.

Защищаемые объекты / активы (DA) – то, на что, по сути, направлены угрозы технологии Искусственного интеллекта. Учитывая специфику направленности данного исследования, с практической точки зрения невозможно высчитать общее число защищаемых активов / объектов. Поэтому учитываться (численный бинарный показатель) будут все области, упоминаемые в НПА страны как защищаемые. Примерный перечень[1] указан в табл. 3.

Таблица 3

Примерный перечень защищаемых объектов / активов

Область, определяемая НПА	Численный показатель (определение в НПА)
Разведданные	0 – отсутствует; 1 – присутствует.
Данные государственных компаний	0 – отсутствует; 1 – присутствует.
Персональные данные граждан	0 – отсутствует; 1 – присутствует.
Автомобили с автономным управлением (self-driving cars)	0 – отсутствует; 1 – присутствует.
Автономное вооружение (autonomous weapons)	0 – отсутствует; 1 – присутствует.
…	…

Отдельным элементом расчета является показатель значимости / ценности активов / объектов. Кумар и Трипати указывают на высокую роль учета в модели показателя «значения защиты акти-

[1] Примерный перечень продемонстрирован на примере раздела об ИИ из: National Security Strategy of the United States of America, December 2017 // White House. – Mode of access: https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf (accessed: 19.04.2021).

вов (protection value), который назначается лицом, принимающим решения» и «лежит между 0 и 1» [Kumar, Tripathi, 2016, p. 1270]. Такой показатель необходим для учета распределения приоритета угрозы со стороны политических акторов и лиц, принимающих решение в сфере обеспечения безопасности. Предлагаемая модель также будет иметь показатель значимости (PV) и принимать значение от «0» до «1», но с распределением весов. Характеристика оценивания приведена в таблице 4.

Таблица 4

Показатели значимости

Характеристика показателя значимости	**Шкала**	**Веса**	**Обоснование**
Технология ИИ в Стратегии национальной безопасности страны	0 – отсутствует; 1 – присутствует.	0.3	Закрепление технологии в национальной стратегии обеспечения безопасности является наивысшим «признанием» со стороны государства значимости как самой технологии, так и потенциала угроз.
Отдельный государственный орган по проблематике ИИ	0 – отсутствует; 1 – присутствует;	0.3	Если в рамках системы государственных органов в структуре обеспечения безопасности создан специальный орган, посвященный технологии ИИ – мы можем утверждать, что государство определяет высокий приоритет данной технологии (вес – 0.3).
Национальная стратегия по ИИ	0 – отсутствует; 1 – присутствует.	0.2	Национальная стратегия по ИИ, хоть и не имеет прямого отношения к сфере обеспечения безопасности, однако в силу специфики самой технологии в определенной мере будет регламентировать и вопросы безопасности.
Сформулированное определение ИИ в национальной стратегии по цифровой трансформации	0 – отсутствует; 1 – присутствует.	0.1	Отсутствие отдельного регулирования (или, хотя бы, доктринального закрепления намерения в виде национальной стратегии) с фокусом на ИИ отражает не столь сильную «заинтересованность» со стороны государства.
Государственные (или аффилированные с государством) компании по разработке технологии ИИ в военной сфере и / или в сфере безопасности	0 – отсутствует; 1 – присутствует.	0.1	Учет технологических компаний является отражением наличия собственных (не иностранных) технических возможностей (вычислительных мощностей, программного обеспечения и тд.) у государства выявлять угрозы технологии.
Максимальный показатель			**1**

Наиболее важным и одним из ключевых моментов расчета угроз является учет факторов и отдельных категорий. Например, ПИР-центр при создании «Индекса международной безопасности»[1] указывает на приоритет военных факторов перед любыми иными (политические факторы, терроризм, техногенные и природные факторы, экономические факторы): «…Всеобщий политический или экономический кризис можно каким-то образом преодолеть, последствия даже глобальной экологической катастрофы, в том числе и вызванной действиями террористов – пусть и не полностью, но нейтрализовать… Что же касается глобальной ядерной войны – то это явление можно считать полностью необратимым и "летальным" для всего человечества» (стр. 5). Более того, предполагается ранжирование внутри каждой из групп по показателям глобального, региональных и локальных факторов безопасности. В предлагаемом подходе указанные факторы также учтены. Исходя из общей теоретической рамки исследования – секторального подхода анализа сферы обеспечения безопасности, предложенного Копенгагенской школой, вводится показатель – Фактор угрозы (TF). Фактор угрозы нашей модели представлен как отношение оценки каждой страны к общей оценке по пяти секторам безопасности Барри Бузана.

Таблица 5

Показатели угроз

Сфера / сектор безопасности	Описание [Buzan, Waever, de Wilde, 1998]	Шкала оценивания
1	2	3
Политическая	Угрозы суверенности, посягательство на легитимность и властный авторитет	0 – отсутствие; 1 – локальная; 2 – региональная; 3 – международная
Военная	Все военные вопросы определяются как угрозы безопасности (кроме миротворческих целей и ликвидации последствий стихийных бедствий)	0 – отсутствие; 1 – локальная; 2 – региональная; 3 – международная
Экономическая	Угрозы экономической стабильности государства и отдельным элементам экономической системы (например, банковскому сектору)	0 – отсутствие; 1 – локальная; 2 – региональная; 3 – международная

[1] Индекс международной безопасности iSi. Описание и методология расчета // ПИР-Центр. – Режим доступа: http://www.pircenter.org/media/content/files/9/13462438640.pdf (дата посещения: 21.04.2021).

Продолжение таблицы 5

1	2	3
Экологическая	Все вопросы окружающей среды на территории национальных границ государства, в том числе глобальные международные климатические вызовы, касающиеся государства (глобальное потепление, загрязнение, озоновый слой и т.д.)	0 – отсутствие; 1 – локальная; 2 – региональная; 3 – международная
Социальная	Вопросы коллективной идентичности (языковой, культурной, религиозной и т.д.) и баланс идентичности в государстве (например, соотношение различных культур и мультикультуральность)	0 – отсутствие; 1 – локальная; 2 – региональная; 3 – международная

По сути, факторы угрозы (TF) учитываются как удельный вес отношения суммы шкалы оценивания на максимальное количество оценки сфер безопасности. Математическая формула (5) расчетов представлена следующим образом:

$$TF = \frac{\sum(score\ \ per\ \ \text{sec}tor)}{max\ score} \quad (5)$$

Общая формула расчета угроз (6) для целей исследования представлена ниже:

$$TE = \frac{PV \cdot \sum(TP, TT)}{DA \cdot TF} \quad (6)$$

где:

PV – показатель значимости (PV∈ [0; 1]);

TP – показатель восприятия угроз;

TT – показатель характера / типа угрозы (TT ∈[0; 4]);

DA – показатель числа защищаемых объектов (DA = {DA1, DA2, DA3, … DAn});

TF – показатель факторов угроз.

Иными словами, логическое содержание формулы можно представить следующим образом: оценка угроз (*TE*) представляет собой **отношение**

суммы показателя восприятия угроз (*TP – как угрозы представлены на уровне нормативно правовых актов – акт политической воли*) и показателя характера / типа угрозы (*TT – как угрозы представлены в отчетах, релевантной литературе*) умноженной

на показатель значимости (*PV – как лица принимающие решения и политические акторы оценивают важность/значимость угроз*) к произведению показателя числа защищаемых объектов (*DA – объекты / активы, на которые направлены угрозы*) на показатель факторов угроз (*TF – оценка и ранжирование секторов безопасности, к которым относятся угрозы*).

Указанное можно выразить в виде блок-схемы оценки угроз.

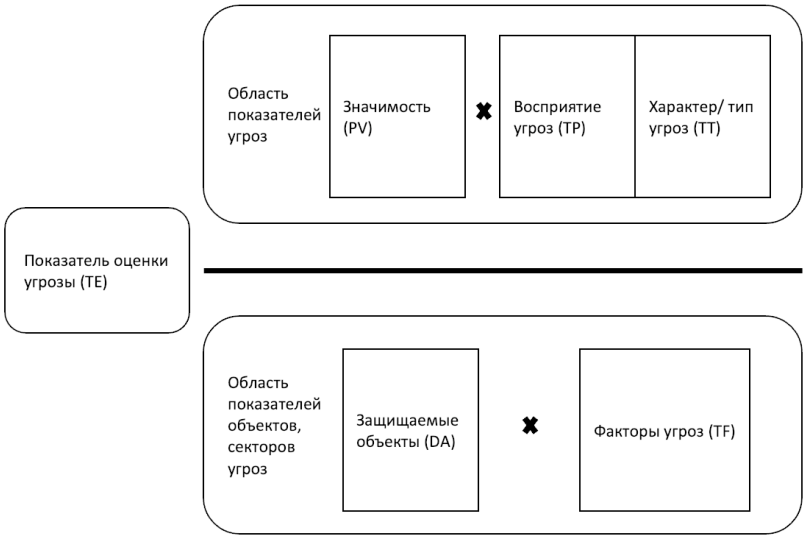

Рис. 3
Блок-схема показателя оценки угрозы

Результаты: динамика применения технологий ИИ на примере США

Валидизация представленной эмпирической модели проводилась на кейсе США. Выбор обоснован тремя основными факторами: (1) доступность (транспарентность) данных о сфере безопасности; (2) специфика стратегии национальной безопасности США, представляющей отдельный интерес; (3) лидирующие позиции по развитию технологий, особенно технологий ИИ, в мире и постоянное развитие (см. исследования DARPA), а также приме-

нение технологий непосредственно в сфере обеспечения безопасности, в том числе за пределами «узких» военных приложений.

Первичной задачей анализа было определение и обоснование временных промежутков, которые будут учитываться в модели. Иными словами, необходимо было выделить конкретные годы, по которым будет проводиться сбор данных. Реализация такой задачи потребовала изучения стратегий национальной безопасности США и нормативно-правовых актов федерального уровня о применении / внедрении технологий во временном охвате более 20 лет. В итоге, анализу подлежат восемь временных промежутков, а именно, года: 1999, 2002, 2006, 2010, 2012, 2015, 2017, 2019. Обоснование относительно анализируемых актов представлено в табл. 6.

Таблица 6

Обоснование дат (временных периодов анализа)

1999	Memorandum on the Use of Information Technology to Improve Our Society[1]	Администрация У. Клинтона
2002	E-government act[2]	Администрация Дж. Буша-мл.
	The National Security Strategy of the United States of America[3]	
	E-Government Strategic Action Plan[4]	
2006	The National Security Strategy March 2006[5]	
	Expanding E-Government Making a Difference for the American People Using Information Technology[6]	
2010	National Security Strategy. May 2010[7]	Администрация Б. Обамы
2012	Digital Government. Building a 21 st Century Platform to Better Serve the American People[8]	
2015	National Security Strategy. February 2015[9]	
2017	National Security Strategy of the United States of America. December 2017[10]	Администрация Д. Трампа
2019	Executive Order on Maintaining American Leadership in Artificial Intelligence[11]	
	Artificial Intelligence for the American People[12]	
	National-AI-Research-and-Development-Strategic-Plan[13]	

[1] Memorandum on the Use of Information Technology To Improve Our Society // Authenticated U.S. Government Informatio GPO. – Mode of access: https://www.govinfo.gov/content/pkg/WCPD-1999-12-27/pdf/WCPD-1999-12-27-Pg2639.pdf (accessed: 05.05.2021).

[2] E-government act // Authenticated U.S. Government Informatio GPO. – Mode of access: https://www.govinfo.gov/content/pkg/PLAW-107publ347/html/PLAW-107publ347.htm (accessed: 05.05.2021).

[3] The National Security Strategy of the United States of America // 2009–2017 Archive of the U.S. Department of State. – Mode of access: https://2009-2017.state.gov/documents/organization/63562.pdf (accessed: 05.05.2021).

[4] E-Government Strategic Action Plan A Road Map for Delivering Services // U.S. Department of Energy. – Mode of access: https://www.hsdl.org/?view& did=444594 (accessed: 05.05.2021).

[5] The National Security Strategy // White House U.S. – Mode of access: https://georgewbush-whitehouse.archives.gov/nsc/nss/2006/ (accessed: 05.05.2021).

[6] Expanding E-Government Making a Difference for the American People Using Information Technology // Homeland Security Digital Library. – Mode of access: https://www.hsdl.org/?view&did=468481 (accessed: 05.05.2021).

[7] The National Security Strategy // The White House President Barack Obama Archive. – Mode of access: https://obamawhitehouse.archives.gov/sites/default/files/rss_viewer/national_security_strategy.pdf (accessed: 05.05.2021).

[8] Digital Government. Building a 21 st Century Platform to Better Serve the American People // Homeland Security Digital Library. – Mode of access: https://www.hsdl.org/?abstract&did=711162 (accessed: 05.05.2021).

[9] The National Security Strategy // The White House President Barack Obama Archive. – Mode of access: https://obamawhitehouse.archives.gov/sites/default/files/docs/2015_national_security_strategy_2.pdf (accessed: 05.05.2021).

[10] The National Security Strategy // The White House President Donald Trump Archive. – Mode of access: https://trumpwhitehouse.archives.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf (accessed: 05.05.2021).

[11] Maintaining American Leadership in Artificial Intelligence // Federal Register. – Mode of access: https://www.federalregister.gov/documents/2019/02/14/2019-02544/ maintaining-american-leadership-in-artificial-intelligence (accessed: 05.05.2021).

[12] Artificial Intelligence for the American People // The White House President Donald Trump Archive. – Mode of access: https://trumpwhitehouse.archives.gov/ai/ (accessed: 05.05.2021).

[13] National-AI-Research-and-Development-Strategic-Plan // National Science & Technology Council. – Mode of access: https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf (accessed: 05.05.2021).

Таким образом, анализируемые периоды охватывают (1) все ключевые даты определения и регулирования стратегии национальной безопасности США (национальные стратегии зачастую определяют основные вектора направленности сферы обеспечения безопасности), (2) основные нормативные документы в сфере технологического развития (начиная от концепции электронного правительства, заканчивая непосредственно регулированием технологий ИИ), (3) последние четыре Администрации, что позволяет прослеживать непосредственную динамику изменений.

После определения временных промежутков сбор данных происходил по единому протоколу по каждому индикатору по каждому году. Финальный датасет, с электронными ссылками на каждый источник и кратким описанием сбора данных, находится в открытом доступе[1].

[1] Data_AL_Security_USA // GitHub. – Mode of access: https://github.com/AITurobov/Data_AI_Security_USA/ (accessed: 29.07.2021).

После расчета[1] модели были получены следующие показатели по параметрам *Возможностей ИИ (AI capability)* и *Оценке угроз (Threat Evaluation)*, а также по искомому показателю согласованности (консистентности) безопасности *(Security Consistency)*. Результаты представлены в табл. 7.

Таблица 7

Результаты расчета модели

Год	Возможности ИИ (AI capability)	Оценка угроз (Threat Evaluation)	Согласованность безопасности (Security Consistency)
1999	0.19625	0	0.19625
2002	0.248425	0	0.248425
2006	0.247	0	0.247
2010	0.9031422	0.1995818	0.7035604
2012	1.563833	0.3844019	1.1794311
2015	1.604973	0.414918	1.190055
2017	1.639961	0.4498543	1.1901067
2019	1.648234	0.9116182	0.7366158

Результаты в графическом виде представлены ниже на рисунке 4.

График позволяет отследить динамику изменений не только самого показателя Согласованности безопасности, но и по каждому значению Оценки угроз и Возможностей ИИ по годам. Примечательно, как меняется с развитием самой технологии «заинтересованность» государства и определение угроз, а также стремительный рост возможностей ИИ (в годы максимального прогресса в вычислительных мощностях и появления новых алгоритмов), а с 2012 г. динамика более поступательная, так как все новые и новые «открытия» в технологиях ИИ имеют уже не революционный, а эволюционный характер.

[1] Все вычисления проводились в бесплатной программой среде для статистический вычислений и графики R (v. 4.0.5).

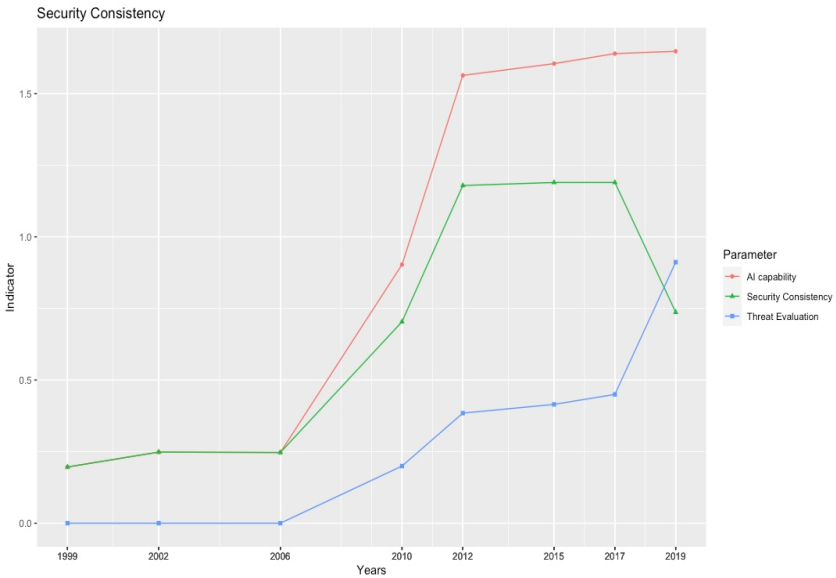

Рис. 4
График Согласованности (консистентности) безопасности, Возможностей ИИ, Оценки угроз (с 1999 по 2019 г.)[1]

Дискуссия и обсуждение

Предлагаемый подход подразумевается как дополнение к существующим и позволяет взглянуть на динамику развития (эволюции и трансформации) системы безопасности конкретного государства под влиянием технологий искусственного интеллекта.

Попытка «объективного» анализа динамики, реализуемая на примере США, демонстрируют некоторые преимущества разработанной модели. Во-первых, мы можем отдельно наблюдать изменения как в официальных оценках угроз (то, КАК и на ЧТО госу-

[1] На оси абсцисс расположены года, на оси ординат – индикаторы. Красная кривая представляет параметр возможностей технологии ИИ, синяя кривая – параметр оценки угроз. Зеленая кривая является иллюстрацией согласованности (консистентности) безопасности.

дарство обращает внимание и определяет как угрозу), а также на возможности технологии ИИ (в контексте публичного восприятия реализуемых исследований, апробации и внедрения различных алгоритмических систем ИИ). Во-вторых, мы можем с учетом временных отрезков анализировать то, как развивалась технология, как она «проникала» в систему безопасности. Например, стремительное развитие технологии ИИ в период с 2006 по 2012 г. (показатель «AI capability» на рис. 1) позволило США инициировать и возглавить технологическую гонку, последствия которой выходят за сугубо научно-исследовательские и технологические пределы. В-третьих, отдельный акцент на оценке угроз позволяет анализировать политическое внимание (отношение) к технологии, что расширяет существующую дискуссию в предметном поле политизации и секьюритизации технологий. В-четвертых, представленная модель обладает интерпретационной мобильностью (гибкостью), учитывая, что собираемые данные представлены из разных областей (социальная, экономическая, политическая) и с большим временным охватом. В-пятых, модель может быть полезна для проведения сравнительных исследований стран. Последующие исследования будут направлены на построение аналогичных моделей для других стран и проведение кросстранового сравнительного анализа как по единому показателю согласованности безопасности, так и по отдельным индикаторам возможностей ИИ и оценки соответсвующих угроз.

Апробация эмпирической модели оценивания системы безопасности государства (на примере США) на основе измерения согласованности безопасности продемонстрировала и некоторые ограничения. Следует отметить, что сама логика согласованности безопасности демонстрирует, что чем меньше показатель, тем более «согласована» система безопасности государства. Иными словами, исходя из математической логики разности (мы вычитаем из индикатора возможностей индикатор угроз), чем больше у государства возможностей и чем больше государство определяет для себя угроз, тем выше вероятность того, что показатель согласованности стремится к нулю. Однако на уровне интерпретации и аргументации необходимо учитывать сложности с оценкой угроз. Например, государство может просто «не видеть» угрозу (умышленно или случайно), но она будет существовать реально. Наоборот, государство может излишне политизировать и секьюритизи-

ровать целые сектора (а акторов – в этих секторах), в результате модель будет демонстрировать высокий индикатор оценки угроз, но фактически эти угрозы не будут иметь реального воплощения. В свою очередь, аналогичные флуктуации могут происходить и с индикатором возможностей ИИ (когда государство «преувеличивает» возможности и т.д.). Также большим ограничением может служить отсутствие данных для некоторых стран. Отдельно следует отметить, что на данный момент сложно говорить о существовании пороговых значений, которые были бы оптимальными для показателя согласованности безопасности. Иными словами, сейчас модель позволяет лишь фиксировать, а не оценивать с позиции успешности / эффективности национальных стратегий развития ИИ, в том числе в приложении к сфере безопасности. Проведение сравнительного анализа с участием большего числа стран позволит выявить лучшие практики и определить границы показателя согласованности безопасности (в сугубо аналитических целях).

Предложенная модель не рассматривает целенаправленно военные приложения технологии ИИ, в том числе по причине того, что сведения о таких приложениях могут иметь закрытый характер (например, для получения и (или) сохранения преимущества в реальном или потенциальном конфликте). Любая модель, в том числе претендующая на отражение сложности реального мира, – это неизбежное упрощение реальности. Использованные в модели параметры описывают не столько реальность («на самом деле»), сколько ее отражение (и понимание), элементы которого содержатся в официальных документах. Отсюда опора на публичные источники, проверяемые и «считаемые» параметры на их основе.

Особо отметим масштабируемость модели: потенциально можно заменить технологию ИИ любой другой цифровой технологией (например, облачные вычисления, технологии связи 5 g и т.д.) и строить эмпирическую модель согласованности (консистентности) безопасности с учетом конкретной технологии. Подразумевается, что сферы, веса и коэфициенты будут сохранены и для другого типа цифровой технологии, однако указанное будет проверено в последующих исследованиях. В результате можно проследить динамику изменений системы безопасности государства, но уже относительно конкретного типа технологии, и также проводить сравнительный анализ стран. Последующие исследования будут сфокусированы на тестировании возможностей модели

и проверке ее интерпретационных возможностей (особенно в сравнительной перспективе).

A.V. Turobov, M.G. Mironyuk*

Empirical model for analysis of the dynamics of algorithmization (artificial intelligence technology) in the field of security by the example of the USA[1]

Abstract. How does the state security system evolve under the influence of the artificial intelligence technology? To answer this question, an empirical model is proposed. The model evaluates the state security system (by the example of the USA) using the security consistency parameter, which estimates how the state perceives threats (indicator of threats) and whether the state has the necessary capabilities to counter them (indicator of capabilities) in relation to the artificial intelligence technology. The model (as well as the conceptualization of the artificial intelligence technology in the context of the security domain) provides evidence of how security transformations occur. It serves as a tool for studying the corresponding changes and assessing the state security system. It is necessary to indicate the limitation of the study: we do not consider direct military applications in the field of automation and algorithms (artificial intelligence technology).

The validation of the empirical model has been undertaken using the case of the USA (eight-time intervals are subject to analysis, namely: 1999, 2002, 2006, 2010, 2012, 2015, 2017, 2019). With the development of the technology itself, the "interest" of the state and the definition of threats, as well as the rapid growth of the capabilities of the artificial intelligence technology (coincides with the years of maximum progress in computing power and the introduction of new algorithms) are growing, and since 2012, the dynamic has been linear, since more new "discoveries" have contributed to evolutionary rather than "revolutionary" growth trajectory.

The developed model is scalable. This feature may be useful in the empirical security studies: the artificial intelligence technology within the model can be replaced with other types of digital technologies (for example, big data, cloud computing or 5 g connection technologies, etc.); thus, empirical models of security consistency under the impact of other technologies can be developed. The approach proposed allows to under-

* **Turobov Aleksey**, HSE University (Moscow, Russia), e-mail: aturobov@hse.ru; **Mironyuk Mikhail**, HSE University (Moscow, Russia), e-mail: mmironyuk@hse.ru

[1] The reported study was funded by RFBR and EISR according to the research project № 20–011–31658. The authors appreciate criticism of anonymous reviewers. The authors thank Vitaly Kabernik (MGIMO University) and Dmitry Stefanovich (IMEMO RAS) for support and advice.

take cross-country comparisons with respect to specific types of digital technologies and their interactions with the security domain.

Keywords: security; security studies; model; artificial intelligence; threat evaluation; national security.

For citation: Turobov A.V., Mironyuk M.G. Empirical model for analysis of the dynamics of algorithmization (artificial intelligence technology) in the field of security by the example of the USA. *Political science (RU)*. 2021, N 3, P. 72–111. DOI: http://www.doi.org/10.31249/poln/2021.03.04

References

Aletras N., Tsarapatsanis D., Preoțiuc-Pietro D., Lampos V. Predicting judicial decisions of the European court of human rights: A natural language processing perspective. *PeerJ computer science*. 2016, Vol. 2, P. 93.

Alter S. Sherer S.*A.* A general, but readily adaptable model of information system risk. *Communications of the association for information systems*. 2004, Vol. 14, Article 1, P. 1–28. DOI: 10.17705/1 CAIS. 01401

Amoore L., Raley R. Securing with algorithms: knowledge, decision, sovereignty. *Security dialogue*. 2017, Vol. 48, Iss. 1, P. 3–10.

Ang R.P., Goh D.H. Predicting juvenile offending: A comparison of data mining methods. *International journal of offender therapy and comparative criminology*. 2013, 57(2), P. 191–207.

Ayoub K., Payne K. Strategy in the age of artificial intelligence. *Journal of strategic studies*. 2016, Vol. 39, P. 793–819.

Baldwin D.A. Security studies and the end of the Cold War. *World politics*. 1995, Vol. 45, P. 117–141.

Baldwin D.A. The concept of security. *Review of international studies*. 1997, Vol. 23, P. 5–26.

Balzacq T. (ed.). *Securitization theory*. London : Routledge. 2011, 272 p.

Buzan B. *People, states and fear. An agenda for international security studies in the Post-Cold War Era*. Brighton : ECPR Press, 1991, 318 p.

Bayley D.H. The Police and political development in Europe. *In:* Tilly C., Ardant G. (eds). *The Formation of National States in Western Europe.* Princeton, NJ : Princeton university press, 1975, P. 328–339.

Brauch H.G., Spring O.Ú., Mesjasz C., Grin J., Dunay P., Behera N.C., Chourou B., Kameri-Mbote P., Liotta P.H. (eds). *Globalization and environmental challenges: rreconceptualizing security in the 21 st century: Vol. 3*. Berlin : Springer Science, Business Media, 2008, 1141 p.

Brose C. The new revolution in military affairs: War's sci-fi future. *Foreign affairs.* 2019, Vol. 98, N 3. Mode of access: https://www.foreignaffairs.com/articles/2019-04-16/new-revolution-military-affairs (accessed: 15.05.2021).

Brożek B., Janik B. Can artificial intelligences be moral agents? *New ideas in psychology*. 2019, Vol. 54, P. 101–106.

Bulgurcu B., Cavusoglu H., Benbasat I. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*. 2010, N 34, P. 523–548.

Buzan B., Wæver O. *Regions and powers*. Cambridge : Cambridge university press, 2003, 598 p. DOI: https://doi.org/10.1017/CBO9780511491252

Chen D.L., Eagel J. Can machine learning help predict the outcome of asylum adjudications? *Proceedings of the ACM Conference on AI and the Law*. 2017, P. 237–240.

Cocelli M., Arkin E. A threat evaluation model for small-scale naval platforms with limited capability. *EEE Symposium Series on Computational Intelligence (SSCI 2016)*. 2017, P. 1–8.

Coglianese C., Lehr D. Regulating by robot: Administrative decision making in the Machine-learning era. *Georgetown law journal*. 2017, 1734 p.

Davis, Zachary. Artificial intelligence on the battlefield: implication for deterrence and surprise. *Institute for national strategic security*. 2019, P. 114–131.

Deshmuhk A. A Framework for Online Internal Controls. *AMCIS August*. 2004, P. 4471–4479.

Devereux S., Vincent K. Using technology to deliver social protection: exploring opportunities and risks. *Development in practice*. 2010, Vol. 20, N 3, P. 367–379.

Edwards P.N. The closed world: computers and the politics of discourse in Cold War America. Cambridge : MIT press, 1997, p. 468.

Ferguson M.P. The digital maginot line: autonomous warfare and strategic incoherence. *Prism*. 2019, Vol. 9, N 2, P. 132–145.

Fioramonti L., Kononykhina O. Measuring the Enabling Environment of Civil Society: A Global Capability Index. *Voluntas*. 2015, Vol. 26, P. 466–487.

Floyd R., Matthew R.A. *Environmental security: approaches and issues. In environmental security: approaches and issues*. London : Routledge, 2012, 320 p.

Galanos V. Artificial intelligence does not exist: Lessons from shared cognition and the opposition to the nature/nurture divide. *IFIP advances in information and communication technology*. 2018, P. 359–373.

Grant R.M., Verona G. What's holding back empirical research into organizational capabilities? Remedies for common problems. *Strategic organization*. 2015, Vol. 13, Iss. 1, P. 61–74.

Buzan B., Waever O., Wilde, J. de. *Security: a new framework for analysis.* London : Lynne Rienner, 1998, 239 p.

Hanlon R.J., Christie K. *Freedom from fear, freedom from want: an introduction to human security.* Torento : University of Torento press, 2016, 288 p.

Horowitz, M.C. Artificial intelligence, international competition, and the balance of power. *Texas national security review*. 2018, Vol. 1, P. 37–57.

Jarrahi M.H. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business horizons*. 2018, Vol. 61, Iss. 4, P. 577–586.

Johansson F., Falkman G.A. Comparison between two approaches to threat evaluation in an air defense scenario. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics).* 2008, Vol. 5285, P. 110–121. DOI: https://doi.org/10.1007/978-3-540-88269-5_11

Johnson J.S. Artificial Intelligence: A Threat to strategic stability. *Strategic studies quarterly*. 2020, Vol. 14, P. 16–39.

Kaplan A., Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*. 2019, Vol. 62, Iss. 1, P. 15–25.

Kennedy D. The move to institutions. *Cardoso law review*. 1987, N 8(5), P. 841–988.

Keskinbora K.H. Medical ethics considerations on artificial intelligence. *Journal of clinical neuroscience.* 2019, Vol. 64, P. 277–282.

Kissell R., Malamut R. Algorithmic Decision-Making Framework. *The Journal of trading*. 2005, Vol. 1, P. 12–21.

Kumar S., Tripathi B.K. Modelling of threat evaluation for dynamic targets using Bayesian network approach. *Procedia technology*. 2016, Vol. 24, P. 1268–1275.

Liu Z., Chen H. A predictive performance comparison of machine learning models for judicial cases. *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017*. 2018, P. 1–6

Longino H. Individuals or populations*?* In: Cartwright N., Montuschi E. (eds). *Philosophy of social science: an introduction.* Oxford : Oxford university press, 2014, P. 102–120.

Lum K., Isaac W. To predict and serve? *Significance*. 2016, Vol. 13, Iss. 5, P. 14–19.

Martin Katz D., Bommarito M.J., Blackman J. A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE. 2017, Vol. 12(4). DOI: https://doi.org/10.1371/journal.pone.0174698

Martin K. Ethical implications and accountability of algorithms. *Journal of business ethics.* 2018, Vol. 160, P. 835–850. DOI: https://doi.org/10.1007/s10551-018-3921-3

McClendon L., Meghanathan N. Using machine learning algorithms to analyze crime data. *Machine learning and applications: an international journal.* 2015, Vol. 2, N 1, P. 1–12.

Mead Earle E. (ed.). *Makers of modern strategy: military thought from Machiavelli to Hitler.* Princeton, NJ : Princeton university press, 1944, p. 951.

Mikhaylov S.J., Esteve M., Campion A. Artificial intelligence for the public sector: Opportunities and challenges of cross-sector collaboration. *Philosophical transactions of the royal society a: mathematical, physical and engineering sciences.* 2018, Vol. 376, Is. 2128. DOI: https://doi.org/10.1098/rsta.2017.0357

Mittelstadt B.D., Allo P., Taddeo M., Wachter S., Florid, L. The ethics of algorithms: mapping the debate. *Big data and society*. 2016, Vol. 3, Iss. 2, P. 1–21.

Naeem H., Masood A. An optimal dynamic threat evaluation and weapon scheduling technique. *Knowledge-based systems.* 2010, Vol. 23, Iss. 4, P. 337–342.

Nakaya T., Yano K. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*. 2010, Vol. 14, Iss. 3, P. 223–239.

Nance William D., Straub Detmar W. An Investigation into the Use and Usefulness of Security software in Detecting Computer Abuse. ICIS 1988 Proceedings. 1988, N 36. Mode of access: http://aisel.aisnet.org/icis1988/36 (accessed: 05.05.2021).

Naseem A., Shah S.T. H., Khan S.A., Malik A.W. Decision support system for optimum decision-making process in threat evaluation and weapon assignment: Current

status, challenges and future directions. *Annual reviews in control.* 2017, Vol. 43, P. 169–187.

Neack L. *National, international, and human security: a comparative introduction.* Lanham, MD : Rowman & Littlefield, 2017, 236 p.

Paret P. (ed.). *Makers of modern strategy from Machiavelli to the Nuclear age.* Princeton, NJ : Princeton university press, 1986, 951 p.

Parker G. *The Military Revolution: Military Innovation and the Rise of the West. 1500–1800: 2 nd ed.* Cambridge : Cambridge university press, 1996, 292 p.

Patil S., Potoglou D., Lu H., Robinson N., Burge P. Trade-off across privacy, security and surveillance in the case of metro travel in Europe. *Transportation research procedia*. 2014, Vol. 1, Iss. 1, P. 121–132.

Payne K. Artificial intelligence: A revolution in strategic affairs? *Survival.* 2018, Vol. 60, P. 7–32.

Reis J., Santo P.E., Melão N. Artificial intelligence in government services: a systematic literature review. *Advances in intelligent systems and computing.* 2019, P. 241–252.

Russell S., Norvig P. *Artificial intelligence a modern approach.* New Jersey *:* Prentice-Hal, 2010, 1152 p.

Sharre *P.* Killer apps: The real dangers of an AI arms race. *Foreign Affairs*. 2019. Mode of access: https://www.foreignaffairs.com/articles/2019-04-16/killer-apps (accessed: 20.04.2021).

Tene O., Polonetsky J. Taming the golem: challenges of ethical algorithmic decision-making. *North Carolina journal of law & Technology.* 2018, Vol. 19, Iss. 1, P. 1–15.

Vallverdu J. The emotional nature of post-cognitive singularities. *In:* Callaghan V. et al. (eds). *The Technological singularity, the frontiers collection*. Germany : Springer-Verlag, Heidelberg, 2017, P. 193–208.

Whitman M.E., Mattord H.J. *Principles of information security*. Boston : Course Technology, 2011, 656 p.

Williams M., Axon L., Nurse J.R.C., Creese S. Future scenarios and challenges for security and privacy. *2016 IEEE 2 nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow, RTSI 2016*. 2016, P. 1–6.

Wolfers A. "National Security" as an Ambiguous Symbol. *Political science quarterly*. 1952, N 67(4), P. 481–502.

Wright Q. *A Study of War.* Chicago : University of Chicago press, 1942, 466 p.

Zarsky T. The Trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science technology and human values*. 2016, P. 118–132.

Zegart A., Morell M. Spies, lies, and algorithms: Why U.S. Intelligence agencies must adapt or fail. *Foreign Affairs*. 2019, N 3. Mode of access: https://www.foreignaffairs.com/articles/2019-04-16/spies-lies-and-algorithms (accessed: 30.04.2021).