Вестник МГИМО-Университета. 2021. С. 1–25
DOI 10.24833/2071-8160-2021-olf1

ОБЗОРНАЯ СТАТЬЯ

Влияние технологий искусственного интеллекта на международные отношения

А.Д. Коробков

Московский государственный институт международных отношений (университет) МИД России

Автоматизация человеческой деятельности растёт с каждым годом последние десять лет благодаря всё более передовым ИИ-алгоритмам, увеличению количества больших данных и вычислительных мощностей. Уже сегодня технологии искусственного интеллекта применяются в повседневной жизни, автоматизируя процессы, ранее выполнявшиеся людьми. Более того, есть мнение, что данные технологии в какой-то момент могут достичь когнитивных способностей человека и даже превзойти их. В этой связи появляются вопросы о том, как будет выглядеть будущее человечества в связке с ИИ-технологиями. Если раньше большинство исследований носили скорее технический характер, то последние три-четыре года появилось много научных исследований с точки зрения дисциплин, в которых возможно влияние искусственного интеллекта. Международные отношения не стали исключением. Цель статьи – обзор научных публикаций, в которых авторы задаются вопросом: как технологии искусственного интеллекта влияют на международные отношения? В рамках статьи было рассмотрено шесть книг и восемь статей. Автор приходит к выводу, что ИИ-технологии оказывают серьёзное влияние на международные отношения в социально-экономическом и политическом плане. Социально-экономической аспект включает в себя последствия автоматизированного капитализма на мировую политику, рост безработицы, появление «безнадежного» класса, поляризацию общества и другое. Международно-политический аспект подразумевает влияние ИИ-технологий на стратегическую стабильность, ядерное сдерживание и кибервойны. Статья приоткрывает верхушку айсберга влияния ИИ-технологий на международные отношения, и предлагает несколько вопросов для обсуждения.

Ключевые слова: искусственный интеллект, большие данные, автоматизация, международные отношения, капитализм, стратегическая стабильность, киберпространство, США, Китай

Сегодня прорывы в ИИ-технологиях приковывают внимание по всему миру, в том числе международных акторов. Использование технологий искусственного интеллекта во всевозможных сферах кардинально меняет их природу, создавая абсолютно новые возможности для стран и их населе-

УДК 327.82
Поступила в редакцию: 18.09.2021
Принята к публикации: 20.11.2021

ния. Различные узкие функции, задачи и действия могут уже выполняться на уровне человеческих способностей и выше, иногда снижая надобность человека. Стремительное развитие в этой сфере и их активное применение повышает уровень неопределённости в международном сообществе. Появляются вопросы: чем может обернуться использование ИИ-технологий в сферах, имеющих значение для человеческих жизней на локальном, региональном и международных уровнях; как обнаруживать, управлять и обезвреживать вовремя эти риски? Более того, появляется вопрос о том, каким может быть будущее для человечества, в котором ИИ-технологии смогут соревноваться с человеком или стать умнее.

Цель статьи – попытаться ответить на вопрос, как технологии искусственного интеллекта влияют на международные отношения? Сегодня возможным ответить уже в какой-то мере есть, поскольку за последние три-четыре года появилось много литературы, которая рассматривает ИИ-технологии не через технологическую призму, а через призму процессов, имеющих отношение к международным отношениям, мировой политике, экономике и социологии.

Под технологиями искусственного интеллекта в статье подразумеваются два направления — общий искусственный интеллект (ОИИ) и узкоспециализированный (УИИ). УИИ активно используется во многих отраслях: от браузеров (например, «Яндекса»), до управления автомобилями и самолетами. ОИИ ещё не существует, поскольку его когнитивные способности должны повторять человеческие. ОИИ должен уметь обучаться и адаптироваться в условиях неопределённости не хуже, чем человек. Также в статье упоминается термин «большие данные» (англ. Big Data). Считается, что это маркированные или неструктурированные данные объёмом больше 150 ГБ в сутки, но единого критерия нет. Большие данные используются для обучения ИИ-алгоритмов и обработки информации при помощи ИИ-алгоритмов для анализа, статистики, прогнозирования и принятия решений. Благодаря появлению массивов больших данных, ИИ-технологии сделали большой скачок в развитии за последние десять лет.

Для анализа литературы в обзорной статье был использован метод сравнения в повествовании. Для понимания влияния ИИ-технологий на мир, общество и международные отношения были использованы материалы: «Атлас искусственного интеллекта: власть, политика и планетарные издержки от искусственного интеллекта» (Crawford 2021), «Сверхдержавы искусственного интеллекта: Китай, Кремниевая долина и новый мировой порядок» (Lee 2018.), «Искусственность белого превосходства: политика и идеология в искусственном интеллекте» (Katz 2020), «Инновационное государство» (Noveck 2021), «Сдерживание и разубеждение в киберпространстве» (Nye 2017), «Конец контроля над вооружениями?» (Brooks 2020), «Новые технологии и стратегическая стабильность» (Chyba 2020), «Бесчеловечная сила» (Kjøsen, Steinhoff, Dyer-

Witheford 2019), «Лживые СМИ: ИИ и социальная жизнь после теста Тьюринга» (Natale 2021), «Грядущая революция в области искусственного интеллекта (ИИ): ее влияние на общество и компании» (Makridakis 2017), «Замена бюрократов автоматизированными колдунами?» (Bell 2021), «Власть камеры: доказательства, контроль, конфиденциальность и большие аудиовизуальные данные» (Fan 2019). Для понимания актуальных концепций работы ИИ-технологий были взяты статьи «Объяснимый искусственный интеллект (XAI): концепции, таксономия, возможности и проблемы на пути к ответственному искусственному интеллекту» (Arrieta et al. 2020) и «Аргументация в искусственном интеллекте» (Bench-Capon, Dunne 2007).

Революция ИИ-технологий и будущее человечества

Автор «Грядущая революция в области искусственного интеллекта (ИИ): её влияние на общество и компании» (Makridakis 2017) задаётся вопросом, приведут ли скачки в ИИ-технологиях к такому же влиянию и последствиям, как в своё время промышленная и цифровая революции. В статье утверждается, что прогресс в ИИ-технологиях можно считать глобальной революцией, которая повлияет на наше общество и жизнь, в частности, в течение следующих пятнадцати лет сильнее, чем промышленная и цифровая революции вместе взятые.

Во-первых, будет серьёзным влияние ИИ-технологий на компании и трудоустройство. В этой связи появятся группы тесно взаимосвязанных компаний и организаций с принятием решений на основе больших данных, что приведёт к усилению глобальной конкуренции между компаниями. Во-вторых, люди смогут покупать товары и получать услуги из любой точки мира, используя интернет, и пользоваться неограниченными дополнительными преимуществами благодаря искусственному интеллекту.

В этой связи главный вопрос, которым задаётся автор, как будет выглядеть сосуществование людей и технологий искусственного интеллекта. Так называемые оптимисты считают, что люди смогут интегрироваться в ИИ-технологии, которые будут помогать им принимать верные решения. Например, как пишет автор, принимать быстрое решение о лучшем ходе в шахматах. Пессимисты считают, что ИИ-технологии станут умнее нас, принимая любые решения за нас. Например, в какой ресторан пойти или на какую работу устроиться.

Макридакис считает, что у человечества есть все шансы максимизировать пользу от ИИ-технологий и при этом не стать ее заложником. Во-первых, по его мнению, риски для человечества понятны. В статье говорится о возможном социальном недовольстве из-за уменьшения количества рабочих мест и росте неравенства между безработными, трудоустроенными и владельцами ИИ-технологий. Во-вторых, имеется достаточно времени, чтобы придумать

меры противодействия. Макридакис пишет, что в случае с индустриальной и цифровой революциями, уровень безработицы и неравенство уменьшились. Однако он отмечает, что быстрое внедрение ИИ-технологий по миру может действительно создать огромную безработицу. По мнению автора, главное своевременно определять риски и придумывать решения.

Конкуренция среди компаний уже ускоряется благодаря ИИ-технологиям. Быстрое внедрение технологий одними и отставание других может привести к большому разрыву в конкуренции между странами. Что касается новых возможностей ИИ-технологий для людей, то уже сегодня ИИ-алгоритмы упрощают жизнь. Например, автопилот в машинах, персонализированная подборка в приложении для знакомств и так далее. Учитывая, что это носит глобальный характер, можно констатировать вероятность дальнейшей трансформации глобального капитализма, что не может не сказаться на мировой политике.

Автор приводит список потенциальных рисков, но, в целом, смотрит достаточно оптимистично на будущее. С его точки зрения у человечества достаточно времени, чтобы избежать безработицы и не стать рабами ИИ-технологий. Но как именно выглядит взаимодействие человека с ИИ сегодня на повседневной основе, и какие последствия это влечет?

Искусственный интеллект через призму социологии

В «Лживые СМИ: ИИ и социальная жизнь после теста Тьюринга» (Natale S. 2021) Натал рассматривает развитие технологий искусственного интеллекта с точки зрения социального взаимодействия человека с ними. То есть то, как человек реагирует на взаимодействие и как последствия таких реакций формируют человеческое поведение с этими технологиями. Важно отметить, что в данном труде автор рассматривает только технологии коммуникативного искусственного интеллекта. Например, такие как «Сири» или «Алекса».

Одна из главных мыслей книги – идея о так называемом «заурядном обмане» (англ. *banal deception*), которому подвергается человек при взаимодействии с коммуникативными ИИ-технологиями. Этот термин может показаться отрицательным на первый взгляд, но это не так. Положительный пример, когда человек оборачивается при ощущении опасности. Автор отмечает, что такого рода обман норма для человека. Именно заурядный обман является одним из важнейших элементов взаимодействия человека с коммуникативными ИИ-технологиями. Такого рода обман норма для человека. Автор описывает пять критериев заурядного обмана: заурядный обман происходит на повседневной основе; он всегда приносит пользу субъекту; обман настолько постоянен, что воспринимается как данность; обман подразумевает активное взаимодействие с субъектом, а не пассивное вовлечение; наконец, заурядный обман может быть «запрограммирован» на какие-либо цели своим создателем.

Натал пишет, что тест Тьюринга, по сути, пройден, если у компьютера получается ввести в заблуждение человека своими ответами до такой степени, что он начинает воспринимать компьютер как человека. В этом есть признаки заурядного обмана. Наше восприятие играет большую роль, а восприятием можно манипулировать. Автор приводит пример первого в мире чат-бота с функциями психолога «Элайза». Сначала пользователям казалось магией общение с компьютерной программой, которая как настоящий психолог задавала вопросы. Как только становилось понятно, как она работает, то магия исчезала. Этот чат-бот имел все признаки заурядного обмана, поэтому для человека он воспринимался не как чат-бот, а человекоподобный психолог. Ровно то "же самое происходит сегодня с нами при взаимодействии с различного рода коммуникативными ИИ-технологиями, такими как «Сири» или «Алекса». Каждый день мы подвергаемся заурядному обману хотя и осознанно. Мы понимаем, что это компьютер, но воспринимаем совсем иначе. Это достигается за счёт разных характеристик, которыми можно наделить машину. Например, человеческий голос, пол, имя, поведение, харизма и другое. Большинство голосовых помощников («Сири», «Алекса», «Алиса», «Маруся») имеют женский голос. Используя эти технологии, сам пользователь принимает активную роль, создавая значимость для себя. Например, обращаясь по имени, учитывая пол. Учитывая, что коммуникативные ИИ-технологии активнее применяются с каждым годом, проникая буквально в каждую семью, они явно будут менять наши жизни.

Автор приходит к выводу, что теперь, осознавая это, важно понять какие последствия влечёт заурядный обман, потому что он является самой мягкой формой, которая может быть превращена в осознанный обман.

Главный вывод из статьи в том, что коммуникативный искусственный интеллект моделирует поведение пользователя благодаря самому пользователю. Это не значит, что эта технология такая же умная, как пользователь-человек. Просто пользователь наделяет её смыслом за счёт заурядного обмана и это меняет восприятие и отношение к технологии самого пользователя. Это взаимодействие меняет нашу природу взаимодействия и восприятия этой технологии. Становясь более совершенными, ИИ-технологии меняют нас всё быстрее и необратимей. Также важно отметить, что одно дело заурядный обман, но совсем другое дипфейки. В случае дипфейков человек обманывается неосознанно. Учитывая это и гуманизацию ИИ-технологий в повседневной жизни, важно понять границы для своевременного обнаружения дефектов. Дипфейки – это уже вызов для международного сообщества из-за достаточно открытого доступа к такого рода технологиям. Существует много фейковых аудио- и видеопримеров в интернете с личностями влиятельных политиков. Благодаря информационно-коммуникационным технологиям, распространение таких видео может привести к мощному дезинформационному воздействию на локальном, региональном или даже международном уровне.

Однако, единственные ли это проблемы, которые создают ИИ-технологии для человечества? Эту тему раскрывает автор следующей публикации.

«Сколько стоит торжество искусственного интеллекта?»

Главный исследовательский вопрос книги «Атлас искусственного интеллекта: власть, политика и планетарные издержки от искусственного интеллекта» (Crawford 2021): как создаются технологии искусственного интеллекта, и к каким проблемам это может привести? В книге рассматриваются процессы, задействованные в создании и развитии любого рода технологий искусственного интеллекта: от материалов для производства до больших данных и космоса. Основной тезис – сегодня технологии искусственного интеллекта являются экстрактивными и служат узкой группе лиц для обеспечения и расширения их власти.

Главными захватывающими идеями публикации являются природные источники, трансформация труда и космоса.

Под природными источниками автор подразумевает то, из каких природных ископаемых создаются технологии на основе искусственного интеллекта. Литий является одним из ключевых для жизнеспособности ИИ-технологий. Однако главная проблема в уроне от добычи природных ископаемых в будущем и огромном использовании энергии для реализации технологий искусственного интеллекта. В случае добычи главная угроза в том, что ископаемые, которые формировались миллионы лет, извлекаются и используются молниеносно. Средняя продолжительность использования «айфона» – 4.7 лет. В тоже время сегодня 70% опасных отходов – электронные отходы. Для обучения и поддержания работы ИИ-технологий требуется большое количество энергии. Досконально неизвестно, сколько энергии тратят такие корпорации как Apple, Microsoft и другие на это. Это влечёт немалый выброс углекислого газа, заявляет Кроуфорд.

Труд человека меняется и будет еще сильнее меняться благодаря слиянию труда людей с ИИ-технологиям. Автор пишет, что для людей это скорее не добровольное, а принудительное сотрудничество. Сегодняшние тренды напоминают подход к труду, появившийся во время индустриальной революции. Труд человека был разделён, чтобы отвечать нуждам станков. Очень похожее происходит с ИИ-технологиями, только в данном случае работникам нужно наблюдать, оценивать и регулировать процессы за ИИ-технологиями. Автор приводит пример Бабиджа, учёного, придумавшего механизацию работы. В его понимании компания должна выглядеть по аналогии с вычислительной системой. Только так можно добиться идеальной производительности труда. С его точки зрения в производственной цепи роль работника вторична и часто отрицательна из-за человеческого фактора. Большинство технологических корпораций, особенно «Амазон», организовывают труд именно так.

Ещё одна важная мысль из книги повествует о роли ИИ-технологий в колонизации и добыче ископаемых в космосе. Сегодня такие миллиардеры как Джеф Безос или Илон Маск участвуют в космической гонке. Причиной этому является убеждение, что человечество должно продолжать экспансию вовне, иначе оно будет вынуждено сильно экономить на потреблении. Например, цель компании Blue Origin (владелец Джеф Безос) в краткосрочной перспективе собирается запускать суборбитальные ракеты, но в долгосрочной колонизация запланирована и добыча ископаемых. В 2015 г. Сенат США принял акт о коммерческом использовании астероидов для частных компаний. Если ранее считалось, что всё, что получится извлечь из космоса, будет общественным достоянием, то теперь это может стать достоянием частных компаний.

Важный вывод, что развитие ИИ-технологий обходится дорого для Земли. Как и в случае изменения климата, истощённые литьевые шахты и другие шахты нужных минералов, отравляющие своими отходами окружающую территорию, могут стать серьёзной региональной и международной проблемой. Учитывая, что литий – важнейший дефицитный элемент для ИИ-технологий, с ним может произойти то же, что и с нефтью. Он и другие минералы станут своего рода отдельной проблемой в рамках или вне энергетической безопасности в межгосударственных отношениях. Это несёт гораздо более опасные риски, чем описывает Натал. Однако его книга фокусируется исключительно на коммуникативных ИИ-технологиях. Кроме того, ясно, что технологии искусственного интеллекта внесут большие коррективы в социальную составляющую, меняя уклад человеческого труда. Какие сложности это может создать для международной миграции, усиления неравенства и поляризации среди стран и их населения, остаётся только предполагать. К слову в этом смысле статья соглашается с мнением Макридакиса. Такие проблемы требуют космополитического подхода для их решения. Стремление завоевать космос благодаря ИИ-технологиям и ресурсам ТНК свидетельствует об усилении роли негосударственных акторов и о дальнейшей приватизации мировой политики. Также это может иметь определённый эффект на стратегическую стабильность, в которой космическая отрасль чрезвычайно важна для равновесия.

В конце своей книге Кроуфорд пишет, что ИИ-технологии служат их создателям (корпорациям, государству, университетам), расширяя их власть горизонтально и вертикально. Как работают ИИ-технологии в авторитарной стране, рассказывает следующая статья.

Как сочетаются авторитарная сверхдержава и ИИ-технологии?

Книга «Сверхдержавы искусственного интеллекта: Китай, Кремниевая долина и новый мировой порядок» (Lee 2018) повествует о развитии ИИ-технологий в Китае в связке с анализом в США. Стоит отметить её уникальность, поскольку она написана человеком, имеющим степень доктора в этих технологиях и опыт

работы в данной сфере как в Америке, так и в Китае. В книге есть две главные идеи: сегодня китайская внутренняя политика по развитию ИИ-технологий весьма эффективна; и искусственный интеллект не приведёт к общему искусственному интеллекту.

После того, как в 2015 г. американская ИИ-технология «Альфа Го» (англ. Alpha Go) выиграла у лучшего китайского игрока в традиционную игру го, китайское государство было шокировано. Это, по словам автора, вызвало эффект «спутника», когда СССР и США конкурировали за космос. Сегодня Китай стал одним из лидеров в ИИ-технологиях благодаря четырём факторам. Во-первых, большой импульс дало население и активное использование смартфонов, которые агрегируют массивы больших данных для быстрого обучения ИИ-алгоритмов. Во-вторых, был сделан фокус на стимулирование большого количества учёных в сфере больших данных и ИИ-технологий. В-третьих, важную роль сыграли предприниматели, создавая жёсткую конкуренцию и, как следствие, сильные технологические решения. Наконец, открытая политика доступа к данным по сравнению с западными странами, позволяет Китаю конкурировать с более передовыми решениями США, но с меньшим количеством данных. Именно авторитарная природа Китая даёт возможность успешно воплощать такую политику касательно данных.

Другая важная мысль автора заключается в утверждении, что ОИИ и, как следствие, сингулярность недостижима. Тут автор вводит в противовес термину «эпоха открытий» термин «эпоха воплощения». Под «эпохой воплощения» автор подразумевает наше время, когда мало фундаментальных открытий, но много открытий в прикладной науке. Открытия в технологии искусственного интеллекта автор относит именно к прикладным наукам. С точки зрения автора в такую эпоху не может произойти сингулярности и ОИИ, так как все открытия заточены на улучшение и углубление существующих технологий. В этой связи разного рода кризисы не будут связаны с ИИ-технологиями, но они будут триггерами. Например, появление трактора автоматизировало труд и повлияло на снижение количества людей для выполнения работы.

Также автор говорит про нарастание неравенства и поляризацию в обществе. Различные автоматизации создадут класс людей, которые не будут иметь добавочную ценность для общества. Они, по сути, будут исключены из новой формы капитализма из-за низкой эффективности. Те, кто сможет найти смысл жизни, будут адаптироваться. Остальным придётся вести безнадёжное существование.

Автор ясно позиционирует ИИ-технологии, как серьёзное стратегическое конкурентное преимущество для сверхдержав, таких как Китай и США. В этой связи он преподносит это как гонку за преимущество в политической организации мира, в который только один актор останется победителем. Однако в журнале *Foreign Affairs* вышла статья «За пределами гонки вооружений искусственного интеллекта. Америка, Китай и угрозы мышления с нулевой

суммой»[1]. Авторы раскритиковали подход к вопросу гонки между державами с точки зрения игры с нулевой суммой. Это так не работает. В статье они сетует на раздутый в книге уровень развития ИИ-технологий и переоцененную ценность больших данных в Китае.

Что касается автоматизации человеческого труда, то тут Ли, Кроуфорд и Макридакис солидарны касательно ИИ-технологий. Однако, если Макридакис и Кроуфорд менее пессимистичны, то Кай-Фу Ли пишет про появление безнадёжного класса людей без работы и смысла жизни. Безнадёжное существование ведёт, прежде всего, к болезням, что нарушает принципы биополитики Мишеля Фуко, которые успешно применяются многими развитыми странами мира.

В эпоху больших данных открытый доступ к ним даёт огромное преимущество для развития технологий, отмечает Ли. У кого больше данных, тот может обучать алгоритмы быстрее, несмотря на их изощрённость. Китай может позволить это в любом случае в большей степени, чем западные страны. Кроуфорд на этот счёт поднимает два важных вопроса. Во-первых, какие могут быть следствия неверной маркировки больших данных для обучения ИИ-алгоритмов. Во-вторых, насколько опасно стремление государства к плотному взаимодействию с корпорациями, агрегирующими эти данные, в первую очередь слежения за населением в целях безопасности. Действительно, доступ к большим данным даёт возможность обнаруживать опасность быстрее, предсказывать риски и угрозы. С другой стороны, оказаться под колпаком «большого брата» у индивидуалистического общества явно нет.

Тем не менее некоторые исследователи, считают что агрегация и использование больших данных внутри государства пойдёт только на пользу.

Как использовать ИИ-технологии, большие данные полиции во благо общества?

«Власть камеры: доказательства, контроль, конфиденциальность и большие аудиовизуальные данные» (Fan 2019) – книга о последствиях повсеместного использования видео- и аудиозаписей полицией США. На первый взгляд может показаться, что книга не имеет никакого отношения к искусственному интеллекту. Главный тезис книги – использование больших данных и ИИ-технологий во благо общества.

Автор приходит к выводу, что анализ больших данных с помощью ИИ-технологий на основе видеозаписей полицейских может лучше предотвращать ущерб, чем традиционные методы; может способствовать благополучию и защите сотрудников и общества. Более того, большие данные вместе с прогнозной аналитикой смогут открыть новые горизонты для эффективности по-

[1] Zwetsloot R. et al. 2018. Beyond the AI Arms Race: America, China, and the Dangers of Zero-Sum Thinking. *Foreign Affairs*. URL: https://www.foreignaffairs.com/reviews/review-essay/2018-11-16/beyond-ai-arms-race

лицейских в предотвращении конфликтов. Например, ИИ-технологии смогут осуществлять поиск и анализ на основе голоса, тона, выбора языка, сканирования лица и других критериев. Эта способность выявлять закономерности, поможет своевременно предотвращать вред со стороны полицейских управлений и укрепит гражданские свободы. Фан подчеркивает, что при правильном использовании большие данные и ИИ-технологии можно достичь эффективного результата в полицейской деятельности.

Однако автор указывает и на минусы. В их числе риск ущерба конфиденциальности личности; отказы от раскрытия данных для общественности; вводящие в заблуждение или предвзятые аудиовизуальные данные и другие. Тем не менее Фан не считает, что это может привести к ситуации наподобие «большого брата».

Вопрос с большими данными в США стоит остро. После огромной утечки данных из Facebook и скандала с Cambridge Analytics правительство США внимательно изучает этот вопрос. Тем не менее общегосударственное состояние законодательство фрагментировано: в отличие от единого закона GDPR в Европе, законы разнятся от штата к штату. Это естественным образом создает риски, прежде всего, для кибератак недружественных акторов. Более того, Кроуфорд отмечает, что наблюдается неэтичное использование больших данных. ИИ-алгоритмам скармливают всё, что есть вне зависимости от того, «полезно» ли это для них с точки зрения нашего общества или нет. В отличие от Фан, она считает, что у государства может появиться соблазн вести постоянно тотальную слежку за населением. Ли на примере Китая показывает, что чрезмерно открытая политика по отношению к большим данным, так или иначе, ведёт к манипуляции населением.

Тем не менее, другие исследователи также солидарны касательно применения ИИ-технологий в государственном секторе.

ИИ-технологии и государственное и муниципальное управление

Автор «Инновационного государства» (Noveck 2021) описывает использование ИИ в государственном и муниципальном управлении. Он ссылается на собственный опыт работы в муниципальной администрации Нью-Джерси, где, действуя сообща и используя преимущества новых технологий сбора, анализа и визуализации информации, было продемонстрировано, как бюрократия может быть эффективной, а не громоздкой и невосприимчивой для граждан.

Автор приходит к выводу, что ведомствам необходимо использовать большие данные и связанные с ними технологии машинного обучения и прогнозной аналитики. Роль ведомств в данном случае играют отдельные сотрудники и целые подразделения муниципальных и государственных органов. Такие подходы к анализу данных помогут ведомствам понять проблемы, которые они решают, более эмпирически и разработать более гибкую политику и услуги. Такие ин-

струменты обработки данных также можно использовать для повышения эффективности взаимодействия с гражданами, помогая ведомств разбираться в больших объёмах информации и приглашать к конструктивному участию разнообразных граждан, которые никогда ранее не участвовали в общественной жизни.

Опыт, описанный автором в публикации и относящийся к работе в административных органах, особенно актуален на сегодняшний день в связи с муниципальной работой в условиях пандемии COVID-19: «Совершенно необходимо изменить то, как мы работаем в правительстве. Кризис показал, насколько плохо административное государство справляется с новыми вызовами».

Автор подчёркивает, что главной проблемой сейчас не является ИИ, который сможет соперничать с человеком. Скорее, главный риск заключается в том, что административные органы не смогут интегрировать технологии в себя. Сейчас есть возможность использовать ИИ-технологии для усиления демократии и реформации административного государства. Однако это требует заметного времени из-за недостаточной политической воли чиновников и их начальников, и трудностей согласования. Автор подчёркивает, что если Вашингтон хочет добиться серьёзных результатов, то нужно провести многократные реформы государственного сектора.

Проанализировав данную публикацию, можно сделать важный вывод: в настоящий момент технологии уже позволяют собирать и обрабатывать информацию, достаточную для генерации вполне обоснованных управленческих административных решений. Это совпадает с точкой зрения Фан, которая показывает ценность такого подхода для полиции.

В чем может быть главная опасность применения ИИ-технологий?

В статье Белла (Bell 2021) поднимаются вопросы применения ИИ в «бюрократическом» управлении социальными системами. В частности, автор обращает внимание на возможность искусственного интеллекта позволить агентствам решать некоторые проблемы. Во-первых, непоследовательное принятие решений и отклонение от политики, связанное с проявлением дискреционных полномочий чиновниками низшего уровня. Во-вторых, неточность агентских правил.

Вместе с тем искусственный интеллект обладает и двумя угрожающими основным ценностям административного права характеристиками: непрозрачностью и не интуитивным характером его корреляций. По мнению автора, искусственный интеллект имеет отрицательные последствия для сложившихся норм. Вероятно, это ограничит наиболее перспективные возможности использования искусственного интеллекта административными органами.

При этом точка зрения Белла представляется реальной. Действительно, искусственный интеллект может давать собственную оценку о тех или иных

субъектах, а именно гражданах, организаций и т.д. Это проблема касается не только государственного и муниципального управления, но и любых других активностей, в которых уже используются УИИ или только планируется. Страшно представить, например, если такая технология применяется в военном деле для принятия решений о ликвидации объекта противника, но при этом никто не может объяснить её логику принятия решений. Это поднимает большое количество вопросов для международного сообщества касательно применения ИИ-технологий в разных сферах. Поэтому важно, чтобы ИИ-алгоритмы были прозрачными с понятной логикой для человека. Иначе это создаст много вопросов о доверии, регулировании и этичном использовании искусственного интеллекта по всему пиру. Тяжело доверять и регулировать то, что непонятно. В этом смысле на первый взгляд позитивная взгляд на применение УИИ Новека и Фан блекнет.

Как пишет Кроуфорд, очень важно понимать, как данные понимаются и интерпретируются ИИ-алгоритмами. Это будет иметь огромное последствия для общества. С точки зрения Кроуфорд, сегодня сбор больших данных не «благотворительная» практика, коей она считалась ранее. Сегодня это действия, нацеленные на властвование и защиту тех, кто получает наибольшую прибыль, избегая при этом ответственности за его последствия. Именно это так или иначе показывает Ли на примере Китая. Фан считает, что при бережном использовании больших данных проблем возникнуть не должно. Но стоит взглянуть на состояние законодательства в разных штатах США касательно данных и конфиденциальности, как начинают появляться вопросы. Особенно учитывая прецеденты утечек больших данных из больших корпораций. Поэтому остаётся открытым вопрос о персональной конфиденциальности. Кто и как имеет право использовать данные без ведома человека? Главное, как правильно защитить эти данные от злоумышленников, которые могут использовать их для нанесения ещё большего ущерба населению? Учитывая, что в США только в двух штатах были приняты полноценные меры по данным, это представляется большой проблемой.

Что если ИИ-алгоритмы помимо отсутствия прозрачности имеют предвзятости по отношению к людям, расам или полу? Что если действительно важные решения будут систематически приниматься во вред определённым группам людей, только потому что так устроена технология? Этот вопрос поднимает следующий автор.

Как ИИ-технология стала инструментом власти белой расы?

В книге «Искусственность белого превосходства: политика и идеология в искусственном интеллекте» (Katz 2020) автор рассматривает искусственный интеллект с точки зрения политики и идеологии. Однако главный акцент книги заключается в утверждении, что ИИ-технологии служат белому расовому превос-

ходству, капитализму и империализму. Тут автор приводит пример Саида и его книгу про ориентализм, суть которой в том, что ориентализм стоит оценивать через призму западного доминирования. Поскольку ИИ-технологии продукт Запада, то автор верит, что они должны рассматриваться конкретно с точки зрения порядка, построенного на превосходстве белой расы.

В качестве главного аргумента автор приводит то, что ИИ-технологии с самого начала создавались и нацелены служить представителям белой расы, а именно, белым мужчинам с университетским образованием, которые чаще всего нацелены на расширение в той или иной форме империализма и капитализма. Это касается как белых учёных и военных, которые дали старт этим технологиям, так и представителей кремниевой долины, ускоряющих развитие капитализма за счёт ИИ-технологий. Данные технологии очень гибки в использовании и приспосабливаются к любой форме белого превосходства.

С точки зрения автора в сфере технологии искусственного интеллекта присутствуют фундаментально ложные утверждения. Одно из них в том, что ИИ-технологии могут соревноваться с человеческим интеллектом или превзойти его в ближайшее время. Катц пишет, что если внимательнее почитать исследования ведущих специалистов, станет ясно, что это невозможно в ближайшее время. Это утверждение раздуто новостными ресурсами.

В конце автор размышляет о возможности альтернативных парадигм. Он задаётся вопросом: может ли быть реализована альтернатива искусственного интеллекта как попытка понять самих себя, наш разум и поведение в цифровом измерении? Автор пишет, что вероятной альтернативой может быть так называемое «воплощенное сознание». Эта теория утверждает, что планирование и принятие решений происходит «снизу вверх» при помощи взаимодействия с миром, нежели при помощи непонятных и сложных внутренних алгоритмов, которые работают «сверху внизу».

Из этой книги можно сделать вывод о том, что, действительно, ИИ-технологии могут быть подвержены белому расизму. Этому свидетельствует большое количество новостей. Другие исследователи также открыто предупреждают об этом в научной среде. В 2021 в *Nature* вышла статья «Размышления об ИИ в 2020 году», отражающая мнение о сегодняшнем состоянии и дальнейшем развитии ИИ-технологий семнадцати учёных. Почти все они так или иначе обеспокоены отсутствием чёткой этики в ИИ-технологиях. Именно это создает проблемы с предвзятостью и т.д. Если ИИ-алгоритмы действительно подвержены таким предвзятостям с точки зрения расы, пола и других критериев, то аргументы Фан и Новека про активное использование ИИ-технологий становятся опасны для общества. Это в первую очередь опасно для американского общества по причине этнического разнообразия. В этом смысле Бел прав, что указывает на отсутствие прозрачности в принятии решений УИИ. Однако, например, учитывая передовое состояние ИИ-технологий в Китае, описанные Кай-Фу Ли, сложно наверняка утверждать, что искусственный интеллект по всему миру на-

правлен на распространение капитализма и империализма исключительно белой расы.

Что касается возможности существования ОИИ, то Катц, как и Ли, отрицают это. Ли считает, что это невозможно, так как мы живем в эпоху «воплощений». Катц приводит перечень мнений исследователей искусственного интеллекта, консенсус которых гласит о низких шансах достижения когнитивных способностей у ОИИ человеческого уровня на нашем веку. Макридакис в отличие от них считает, что это возможно.

Возможно ли как-то управлять ИИ-технологиями, не допуская вышеописанного, расскажет следующая статья.

Что такое «объяснимый искусственный интеллект», и в чём его важность?

В данной работе в фокусе находятся технологическая составляющая ИИ-технологий. В ней обозреваются разные публикации про «объяснимый искусственный интеллект». Главная проблема машинного обучения и любых других видов ИИ-технологий в том, что зачастую неясна их логика принятия цепочки решений, выводов и так далее. То есть даже создатели алгоритма не могут объяснить цепочку, которая привела алгоритм к тому или иному решению. Это вызывает большие опасения. Особенно, когда дело касается решений в отраслях, имеющих влияние на людей, где требуется транспарентность для пресечения несчастных случаев в будущем. Например, применения ИИ-технологий в медицине, транспорте и т.д. «Объяснимый искусственный интеллект» даёт ответ на вопросы о логике работы алгоритмов и делает их прозрачными.

Однако, что ещё важно выделить из этой статьи – это концепция ответственного искусственного интеллекта. Она подразумевает использование ИИ-технологий, учитывая следующие принципы: справедливость, транспарентность, гуманизация, конфиденциальность и безопасность. Все эти принципы также должны быть применимы к любым третьим лица, которые тоже должны быть ответственны.

Авторы приходят к выводу, что нужно добиться правильного понимания потенциальных возможностей и угроз, несмотря на очевидные преимущества «объяснимого искусственного интеллекта». Понимание работы алгоритма должно решаться совместно, учитывая конфиденциальность, справедливость и ответственность. Ответственное внедрение и использование ИИ-технологий по всему миру может быть реализовано при совместном изучении принципов, перечисленных выше.

«Объяснимый искусственный интеллект» и концепция ответственного искусственного интеллекта действительно полезны и важны для обсуждения развития нормативно-правовой базы в мире, потому что на их основе можно создать ясные и понятные для всех правила игры. Можно отделять «плохой»

алгоритм от «хорошего». Если нельзя объяснить, как тот или иной алгоритм принял решение, то будет тяжело сотрудничать на региональном или международном уровне в любой сфере, поскольку никто не может объяснить толком, почему было принято то или иное решение ИИ-технологиями.

В публикации «Аргументация в искусственном интеллекте» авторы (Bench-Capon, Dunne 2007) рассказывают о применении аргументации в искусственном интеллекте. Много подходов в ИИ-технологии пришло из логики, подход аргументирования дал один из небольших результатов. Аргументация – один из ключевых методов создания «необъяснимого искусственного интеллекта».

Эта публикация – важное дополнение для статьи про объяснимый искусственный интеллект (XAI). Именно подход аргументации может шаг за шагом показать, как ИИ-алгоритм принимает решение. Он даёт ясные объяснения в ИИ-технологиях касательно принятия решений. Благодаря аргументации можно создавать алгоритмы с объяснимым искусственным интеллектом в различных областях. Например, робототехника, медицина, безопасность, юриспруденция, безопасность. На основе тандема объяснимого искусственного интеллекта и аргументации, в отличие от других подходов в ИИ-технологиях, представляется возможным обсуждение и создание прозрачных правовых норм для участников международного сообщества.

Подход объяснимого искусственного интеллекта совместно с аргументацией подразумевает, что даже если есть риски предвзятости в ИИ-технологиях, то их можно будет увидеть. Это может дать сильный импульс в сторону воплощения этики в развитии технологий искусственного интеллекта. Хоть Катц утверждает, что, поскольку эти технологии создаются внутри расистского общества, то повлиять на это тяжело. Однако Кроуфорд пишет про демократизацию ИИ-технологий. Концепция ответственных ИИ-технологий с её принципами о транспарентности, справедливости, гуманизации как раз нацелена на решение таких проблем. В этой связи аргумент Катца про расистское общество, если и имеет силу, то только в настоящем. Более того, это значит, что можно решить проблему отсутствия транспарентности при использовании ИИ-технологий, про которую пишет Белл. Соответственно, их использование ведомствами в целом реалистично, как пишут Новек и Фан.

Тем не менее Кроуфорд и Катц настаивают на том, что ИИ-технологии являются эксклюзивным инструментом власти прежде всего для капиталистов, которые стремятся преумножить добавленную стоимость любыми методами.

Влияние ИИ-технологий на капитализм

Главный вопрос книги «Бесчеловечная сила» (Kjøsen, Steinhoff, Dyer-Witheford 2019): что технологии искусственного интеллекта значат для развития капитализма? Отвечая на него, авторы сначала рассматривают сложившиеся

точки зрения про дальнейшее развитие ИИ-технологий, затем освещают самые главные вызовы.

Авторы рассматривают три основные точки зрения на этот счёт: минималисты, максималисты и абиссальные (англ. *abyssists*). Минималисты считают, что никакого кризиса с появлением ИИ-технологий не будет и его потенциал раздут. Как и в прошлом, разрушение устаревших отраслей и профессий влекло к появлению новых. Авторы книги соглашаются с этой точкой зрения в краткосрочной перспективе использования ИИ-технологий. Но они подчёркивают, что представители данного направления недооценивают стремительное развитие искусственного интеллекта и сопутствующих факторов, а именно, ежегодное снижения затрат на вычислительные мощности и стремительный рост больших данных. Максималисты считают, что искусственный интеллект приведёт к концу капитализма и появлению универсального базового подхода. Автоматизация приведёт к разрушению заработной платы и к так называемому роскошному коммунизму с универсальным базовым доходом. Авторы считают, что в данном случае искусственный интеллект с оттенками социализма и коммунизма вполне возможен, так как кризис труда — это кризис капитализма. Но позитивного исхода авторы не видят. Представители последнего направления считают, что как первые, так и вторые заблуждаются. Они уверены, что искусственный интеллект совершенствует капитализм, создавая опасность для рабочих и всего населения планеты. Именно благодаря ИИ-технологиям совершенная форма капитализма станет возможной. С этой точкой зрения авторы соглашаются.

Главный вызов с точки зрения настоящего – это разложение рабочего класса из-за автоматизации труда и фейковые новости. Выборы в США тому хороший пример. Авторы книги в целом соглашаются с этой точкой зрения. Но подчёркивают, что представители данного направления недооценивают стремительное развитие искусственного интеллекта. В недалеком будущем ИИ-автоматизации значительно снизят количество человеческого труда, что вызовет цикличность волны безработицы. Однако главной проблемой будет не столько безработица, сколько отсутствие власти над искусственным интеллектом и автоматизацией капитала, которые начинают жить своей жизнью. Авторы считают, что искусственный интеллект с оттенками социализма и коммунизма вполне возможен. В данном случае кризис труда – это кризис капитализма. Финальная точка развития искусственного интеллекта выглядит для авторов, как гибрид человека и общего искусственного интеллекта. Эти существа заменят любой человеческий элемент в отношении труда и капитала. В данном случае, с точки зрения марксистов, как раз должно произойти уничтожение капитализма. Однако авторы труда так не считают. Они верят в возможность становления капитализма среди этих существ, которые смогут производить и продавать добавочную ценность друг другу. Более того, противостояние между эксплуататорами и эксплуатируемыми машинами может продолжиться.

По итогу авторы приходят к трём главным выводам в своей книге. Во-первых, восстания против капитала на основе искусственного интеллекта уже начались. Например, забастовки работников ИТ-корпораций (Google, Facebook и т.д.), демонстрации в Кремниевой долине против милитаризации ИТ-технологий, движения против предвзятых алгоритмов и т.д. Все они направлены на переосмысление автоматизации за счёт ИИ-технологий. Во-вторых, стоит отказаться от идеи, что искусственный интеллект приведёт к социализму и мировому базовому доходу, даже в случае полного контроля социал-демократами. Важно понимать, что это предложение лоббируется для стабилизации капитализма, основанного на олигополии ИТ-корпораций. Наконец, авторы утверждают, что исключительно капиталистическая природа развития ИИ-технологий должна быть изменена и должна регулироваться обществом вместе. Например, решения про развитие той или иной области ИИ-технологий должны приниматься вместе открыто. В любом случае они приходят к выводу, что нас ждут либо техно-капиталистическая, либо социал-коммунистическая сингулярности. Они не совместимы.

Важные выводы, которые можно сделать в рамках международных отношений: ИИ-технологии влияют на глобальный капитализм, а, следовательно, на мировую политику. Если раньше капитализм преследовал географическую, а затем информационную экспансию, то теперь благодаря ИИ-технологиям будет происходить оптимизация капитала. Про отрицательное влияние глобального капитализма много пишут неомарксисты и критическая школа. На основе статьи можно сделать вывод, что капитализм, основанный на искусственном интеллекте, только усилит классовость государств центра, полупериферии и периферии. Безработица явно может привести к миграционным волнам и давлению на межгосударственные отношения. При возникновении общего искусственного интеллекта сложно представить, как будут выглядеть межгосударственные отношения.

Мнение авторов касательно возможности общего искусственного интеллекта совпадает с позицией Макридакиса, на первый взгляд, но противоречит Катцу и Ли. Однако стоит отметить, что все авторы воздерживаются от точных временных рамок возникновения ОИИ и скорее предполагают, когда он не появится. Например, Катц, ссылаясь на разных специалистов в этой области, пишет, что ОИИ не появится на нашем веку. Ли указывает, что в нашу эпоху это невозможно. Однако эпоха может длиться несколько десятилетий, а может и несколько тысячелетий, как эпоха мезолита.

В чем Кроуфорд, Катц, Кай-Фу Ли, Макридакис, Белл, Новек и авторы данной публикации единогласны так или иначе – ИИ-технологии занимают центральную роль в труде, что влечёт отрицательные последствия для трудоустройства по всему миру.

Ещё один важный пункт, по которому соглашаются авторы книги, Кроуфорд и Катц – эти технологии служат инструментом силы на данный момент и

несут пользу исключительно для их владельцев, нежели чем для всего общества, в частности, и во всем мире. Кйосен, Стейнхоф, Даер-Вифорд предлагают решать эту проблему через общественное регулирование. Кроуфорд также пишет про похожий подход, но подчёркивает, что из-за централизованной природы ИИ-технологий их очень сложно демократизировать. Это то же самое, что предложить демократизацию производства оружия в мире для поддержания мира, пишет автор.

Через призму взаимодействия сторон рассматривается контроль над ИИ-технологиями в рамках межгосударственных военно-политических отношений, а именно в киберпространстве.

Как ИИ-технологии сказываются на стратегической стабильности?

В статье «Новые технологии и стратегическая стабильность» (Chyba 2020) Чиба рассматривает как новые технологии, в том числе ИИ-технологии, влияют на стратегическую стабильность в мире. Автор анализирует влияние с трёх разных углов: темп развития и скорость распространения технологии; влияние технологии на безопасность; и потенциал влияния технологии на кризисное принятие решений.

С точки зрения темпа развития и распространения, ИИ-технологии точно будут иметь серьёзный эффект на стратегическую стабильность благодаря быстрому развитию и широкому применению по всему миру. Технологии искусственного интеллекта автор относит к так называемым «стимулирующим технологиям», то есть тем, которые дают импульс к развитию другого рода технологий. Поэтому, как отмечает автор, все ядерные державы активно развивают данное направление. Однако, как отмечает автор, тут всё зависит от применения. Главные проблемы с такого рода технологиями заключаются в отслеживании и контроле, поскольку в отличие от тестирования и запуска ракет, например, источник вируса достаточно тяжело отследить. Автор приходит к выводу, что в случае ИИ-технологий будет тяжело создать эффективный режим контроля.

ИИ-технологии явно будут иметь сильный эффект на сдерживание и безопасность. Автор приводит пример значительно более эффективного спутникового слежения благодаря слиянию обработки больших данных и ИИ-технологии. Такая комбинация технологий может позволять шпионить за стратегическими объектами чуть ли не в реальном времени. Например, отслеживать нахождение и передвижение межконтинентальных баллистических ракет здесь и сейчас. Наличие таких данных даёт возможность нанести первыми удар по стратегическим целям для причинения непомерного урона противнику. Это вызывает вопросы о том, стоит ли в таком случае делать передвижными комплексы МБР, и др. Эта идея может показаться безумной, но на горизонте двадцати лет она возможна. Свидетельством тому появление с технологической точки зрения похожего проекта «Старлинк», цель которого к 2025 г. – создать глобальную спут-

никовую систему из двенадцати тысяч спутников для обеспечения интернета в любом месте мира. Более того, уже есть частные компании, например, Planet Lab, у которой триста действующих спутников следит за планетой в целях разведки и безопасности. Конечно, сейчас ни одна из компаний не способна обеспечить тот уровень слежки, о котором шла речь выше, но тем не менее. Такие вызовы для России и Китая, вероятно, могут внести дестабилизацию в стратегическую стабильность.

Что касается принятия кризисных решений, то тут автор делает акцент на повышение скорости принятия разного рода решений в случае интеграции ИИ-технологий. Будут ли эти решения приниматься в конце концов людьми или частично машинами, в любом случае обработка и анализ данных будет происходить благодаря ИИ-технологиям. Главная проблема в том, что люди начинают слепо доверять предоставленной информации. Примерно так же как водитель навигатору. Конечно, принятие решений в такой форме может быстро привести к ядерной эскалации. Автор снова подчёркивает, что контроль такого рода технологий крайне непрост.

Чиба приходит к выводу, что самый эффективный метод контроля – активный обмен мнениями с союзниками и партнёрами для определения и обсуждения путей ядерной эскалации и возможных превентивных или смягчающих мер.

Статья ясно показывает большой и глубокий спектр проблем, которые могут вызвать ИИ-технологии для равновесия стратегической стабильности. Одним из главных инструментов для регулирования стратегической стабильности в случае ядерного оружия были постоянное межгосударственное взаимодействие и закрепление результатов на основе многосторонних и двусторонних договоров. Вероятно, что в случае с ИИ-технологиями они будут играть важную роль. Такой подход контроля схож с точками зрения Кроуфорд, Кйосен, Стейнхоф и Даер-Вифорд.

Как контролировать ИИ-технологии в рамках межгосударственных отношений?

Главный фокус статьи «Конец контроля над вооружениями»? (Brooks 2020) на причинах предполагавшегося на тот момент потенциального развала СНВ-3. Автор предлагает расширенную концепцию контроля над вооружениями, которая должна включать формы совместного снижения рисков и предлагает новые меры по предотвращению непреднамеренной эскалации кризисов. Данная концепция подразумевает взаимодействие и сотрудничество по всем возможным ситуациям, а не только в рамках договоров. Автор рассуждает, прежде всего, про ядерное оружие и немного про контроль киберпространства. Именно эта часть имеет главную ценность в рамках статьи. Ведь важно понимать, что ИИ-технологии активно применяются в данной отрасли.

Линтон пишет, что последовательные повторяющиеся кибератаки могут быть восприняты, как предпосылка для ядерной атаки. Неверно оцененные риски и неэффективное кризисное управление являются главной проблемой. Сдерживание в данном случае сильно не поможет. Идеальным сценарием было бы официальное или неофициальное обсуждение участников эскалации конфликта. Важно, чтобы были подобраны верные представители с двух сторон.

Касательно опасных воздействий кибератак, с точки зрения автора, сторонам следует создать постоянную группу киберэкспертов, которая будет собираться каждые шесть месяцев для обсуждения возможных вторжений и обнаружений. Благодаря этой группе каждая сторона сможет определить, что, по их мнению, будет свидетельствовать о возможной подготовке к первому удару. Поскольку обе стороны заинтересованы в предотвращении эскалации кризисов, у них нет стимула быть неискренними в обмене. Если одна из сторон будет обеспокоена, то требуется созвать всех на дискуссию на дипломатическом, военном и других высоких уровнях для прояснения ситуации. Цель обычных совещаний в значительной степени состоит в том, чтобы эксперты сторон были знакомы с мышлением и подходом друг друга и, таким образом, были более эффективными в предотвращении конфликтов.

Динамика киберпространства на стыке с международными отношениями сегодня широко обсуждается, так как эта сфера действительно имеет большое влияние. В этой связи отдельно стоит отметить возрастающую роль УИИ, которые смогут автоматизировать кибератаки, делая их эффективней и опасней. Линтон, как и Чиба, пишет про межгосударственное взаимодействие касательно главных вызовов, которые несут эти технологии. Однако Линтон даёт более развёрнутое предложение о том, как это может выглядеть.

Сдерживание в киберпространстве

Хотя в данной статье автор не углубляется в технологии искусственного интеллекта в рамках киберпространства, важно понимать, что там они широко применимы.

В статье «Deterrence and Dissuasion in Cyberspace» (Nye 2017) Джозеф Най приходит к выводу, что сдерживание в киберпространстве скорее возможно, чем невозможно. Во-первых, потому что проблема с установлением источника кибератаки не так уже велика. В целом, установить источник чаще всего возможно, несмотря ни на что. Най отмечает, что в случае с киберсферой сдерживание выглядит иначе, чем с ядерным оружием. В случае ядерного оружия легко определить, чья ракета и откуда она летит. Во-вторых, несмотря на разнородность угроз, их влияние становится всё яснее и яснее. Най пишет, что в случае киберпространства угроза скорее сравнима не с ядерным оружием, а с повседневной преступностью внутри государства. С такого рода угрозами для безопасности странам приходится бороться постоянно.

Най приходит к выводу, что со временем передовые военные, разведывательные и другие службы смогут взламывать большинство систем безопасности, но правильно применённые оборонительные и наступательные киберметоды смогут сделать ситуацию равновесной. Если раньше ядром концепции сдерживания было ядерное оружие, то сейчас становится ясно, что в случае киберпространства нужно взять за основу базу, но адаптировать методы с учётом сегодняшней динамики.

Харкнетт дал исчерпывающий ответ на статью Ная в журнале International Security (Harknett, Nye 2017). Он считает, что в случае киберсдерживания исследователям и политикам стоит полностью уйти от парадигмы сдерживания, сложившейся на основе сдерживания ядерного. Это концепция не подходит для ведения верной политики и объяснения агрессии в киберпространстве. В качестве аргумента он приводит объяснения Ная, который в статье указывал, что источник кибератаки тяжело определить, поэтому сдерживание посредством наказания невозможно. Сдерживание посредством недопущения фактически нереализуемо, так как кибератаки больше похожи на борьбу с повседневной преступностью. Харкнетт считает, что требуются совершенно новые подходы вне парадигмы сдерживания. Важность создания новой парадигмы очевидна, поскольку, по мнению Харкнетта, такие страны как Россия и Китай воспринимают это поприще как стратегически важное для поддержания их позиции в мироустройстве. Использование ИИ-технологий даст им очередное преимущество. Харкнетт приходит к выводу, что в случае киберпространства, где присутствует постоянное взаимодействие и конъюнктурные изменения, акторы постоянно будут создавать разного рода вызовы, обеспечивая своего рода безопасность друг другу. Безопасность будет достигнута тогда, когда каждый будет понимать свои слабые стороны и способы их эксплуатации.

Най в ответ на комментарии Харкнетта написал, что нет надобности сильно отходить от концепции классического сдерживания. Най считает, что скорее требуется адаптировать существующие типы сдерживания к нынешней динамике, но не отказываться от них вовсе.

ИИ-технологии определённо будут оказывать влияние на сдерживание в киберпространстве. Во-первых, с точки зрения автоматизации атак. Они будут быстрее и эффективней. Но в тоже время стоит отметить, что в рамках кибербезопасности уже используются ИИ-технологии для повышения уровня безопасности. УИИ делает систему более эффективной для обнаружения атак, но, с другой стороны, создаёт разного рода уязвимости для атакующих, как показывает статья «Cyber Security Meets Artificial Intelligence: a Survey». Это значит, что киберсдерживание в ближайшее время будет меняться в той или иной степени, чтобы найти своё оптимальное состояние, о чём пишут Харкнетт и Най. Это явно скажется на стратегической стабильности, как пишет Чиба.

Чиба, Линтон, Кйосен, Стейнхоф, Даер-Вифорд, Кроуфорд, Катц, Макридакис в той или иной степени подразумевают, что в связи со стремительным раз-

витием ИИ-технологий, огромным потенциалом и угрозами требуется начать обсуждать на глобальном уровне, как правильно управлять рисками и какие должны быть нормы применения.

* * *

Так как именно влияют технологии искусственного интеллекта на международные отношения? На этот вопрос тяжело ответить в форме статьи, поскольку ИИ-технологии применимы повсюду. Но, основываясь на прочитанной литературе, можно выделить две основных сферы влияния: социально-экономическая и международно-политическая.

ИИ-технологии имеют социально-экономическое влияние на международные отношения. Во-первых, поскольку внедрение ИИ-технологий носит глобальный характер, то уже происходит влияние автоматизации капитализма на мировую политику. Про влияние на капитализм пишут Кйосен, Стейнхоф, Даер-Вифорд, Кроуфорд, Катц и Макридакис. Если раньше капитализм преследовал географическую, а затем информационную экспансию, то теперь благодаря ИИ-технологиям будет происходить оптимизация капитала. Кроуфорд пишет, что эти технологии по инерции привели к экспансии в космосе. Капитал растёт, как никогда, но без внешней экспансии он так или иначе перестанет расти. Завоевание космоса и добыча ископаемых там может решить эту проблему. Во-вторых, человеческий труд будет автоматизироваться, что приведёт к повышению уровня безработицы, поляризации общества, в том числе на международном уровне. С этим соглашаются Макридакис, Кроуфорд, Ли, Кйосен, Стейнхоф, Даер-Вифорд. Последние четыре автора отмечают, что это создаёт класс людей, которые не могут дать добавочной стоимости автоматизированному капитализму. Ли считает, что эти люди обречены на безжизненное существование. Кйосен, Стейнхоф и Даер-Вифорд пишут о возможности введения базового дохода и появления роскошного коммунизма. Но, сами не считают такой вариант лучшим. Если смотреть на краткосрочное будущее ИИ-технологий с точки зрения неомарксизма, в частности, Валлерстайна, можно говорить про усиление классовости государств. Активное применение ИИ-технологий в передовых странах и слабое развитие или полное отсутствие в развивающихся странах тому доказательство. В-третьих, как становится ясно из книги Кроуфорд, ресурсы для создания ИИ-технологий, например, литий, ограниченны и быстро расходуются. Более того, обучение и поддержание этих технологий несёт выделение углекислого газа. Станет ли литий новой нефтью, а сервера новыми коровами, пока до конца неясно. Но риски для международного сообщества с точки зрения безопасности цифровой инфраструктуры и климатических последствий явно есть. В-четвёртых, отдельной темой являются вопросы использования больших данных в связке с ИИ-технологиями государством и конфиденциальность. Фан и Новек считают, что это сильно увеличит эффективность государственных ведомств. Однако Белл и Кроуфорд предостерегают от этого

по разным причинам. Белл, так как часто непонятна логика принятий решений алгоритмами, что делает её непрозрачной и непонятной. Кроуфорд пишет про этическое использование больших данных алгоритмами, а не чем попало. Однако, как показывают две статьи про подход «объяснимого искусственного интеллекта», аргументации и ответственного искусственного интеллекта, алгоритмы можно сделать прозрачными. Наконец, возможное появление ОИИ приведёт к сингулярности. Макридакис, Кйосен, Стейнхоф и Даер-Вифорд считают, что это возможно. Катц и Ли считают, что в ближайшее время невозможно, при этом не указывает точных сроков. Конечно, на этот вопрос сложно ответить, но большинство экспертов считает, что ОИИ появится до конца нынешнего столетия. Такое фундаментальное изменение очевидно скажется на всём, в том числе на международных отношениях. Как именно эта сфера будет выглядеть, остаётся только догадываться.

ИИ-технологии имеют международно-политическое влияние. Во-первых, эта технология воспринимается и используется ИИ-технологию для власти, о чем пишут Ли, Кроуфорд и Катц. Во-вторых, эта технология будет иметь воздействие на стратегическую стабильность, по мнению Чибы. Он приводит пример спутниковой слежки на основе ИИ-алгоритмов, которая в будущем может помочь отследить местонахождение МБРов и нанести точный удар, лишая противника ответного удара.

В-третьих, ИИ-технологии применимы в киберпространстве, что делает их неотъемлемой частью международных отношений. Про использование этих технологий коротко пишет Харкнетт. Что касается контроля и регулирования ИИ-технологий, то тут Чиба и Линтон соглашаются в том, что этот процесс должен проходить посредством постоянного обмена информацией между союзниками и враждующими сторонами. Харкнетт считает, что в случае киберсдерживания, исследователи должны придумать абсолютно новую парадигму, так как ядерное сдерживание не подходит. Най не согласен, так как считает, что нужно адаптировать имеющиеся подходы и развивать их дальше в соответствии с динамикой. Наконец, ИИ-технологии, применяемые в космической отрасли, которая также является частью стратегической стабильности. Крауфорд пишет про гонку космических компаний миллиардеров. Это явно свидетельствует об усилении ТНК, как акторов на международной арене. В то же время это подразумевает дальнейшие процессы приватизации власти.

ИИ-технологии, как и информационные технологии, сформировались на основе информационного общества и явно тесно связаны с международными отношениями и мировой политикой. Создание и применение этих технологий уже сегодня предмет противоречий на международном, региональном и даже локальном уровнях. В обзорной статье поднимается только верхушка айсберга касательно проблем и вызовов в сфере ИИ-технологий, поскольку все охватить невозможно.

Об авторе:

Андрей Дмитриевич Коробков – аспирант кафедры международных отношений и внешней политики России МГИМО МИД России, 119454, Москва, проспект Вернадского, 76.
E-mail: korobkov.a@inno.mgimo.ru

Конфликт интересов:
Автор заявляет об отсутствии конфликта интересов.

Благодарности:
Исследование выполнено в рамках проекта «Посткризисное мироустройство: вызовы и технологии, конкуренция и сотрудничество" по гранту Министерства науки и высшего образования РФ на проведение крупных научных проектов по приоритетным направлениям научно-технологического развития (Соглашение № 075-15-2020-783)

UDC 327.82
Received: September 18, 2021
Accepted: November 20, 2021

The Impact of Artificial Intelligence Technologies on International Relations

A.D. Korobkov
DOI 10.24833/2071-8160-2021-olf1

MGIMO University

Abstract: The automation of human activity has been growing every year for the past ten years thanks to AI algorithms, increasing the amount of big data and computing power. Today, artificial intelligence technologies are used in everyday life, automating processes previously performed by humans. Moreover, it is believed that these technologies can reach human cognitive abilities and even surpass them at some point. In this regard, questions arise what the future of humanity will look like in conjunction with AI technologies. If earlier, most of the research was more technical, then in the last three to four years, there has been much scientific research in terms of disciplines in which the influence of artificial intelligence is possible. International relations are no exception. The purpose of the article is to review publications on how artificial intelligence technologies affect international relations? Within the framework of the article, six books and eight articles were considered. The author concludes that AI technologies seriously impact international relations in socio-economic and political terms. The socio-economic aspect includes the consequences of automated capitalism on world politics, the rise in unemployment, the emergence of a "hopeless" class, the polarization of society, and more. AI technologies influences strategic stability, nuclear deterrence, and cyber warfare.

Keywords: artificial intelligence, Big Data, automation, international relations, capitalism, strategic stability, cyberspace, USA, China

About the author:

Andrey D. Korobkov – Postgraduate Student of the Department of International Relations and Foreign Policy of Russia, MGIMO, Ministry of Foreign Affairs of Russia, 76 Vernadskogo Ave., Moscow, 119454. E-mail: korobkov.a@inno.mgimo.ru

Conflict of interests:
The author declares the absence of conflicts of interests.

Acknowledgements:
The study was carried out within the framework of the project "Post-crisis world order: challenges and technologies, competition and cooperation" under a grant from the Ministry of Science and Higher Education of the Russian Federation to conduct large research projects in priority areas of scientific and technological development (Agreement No. 075-15-2020-783)

References:

Arrieta A.B. et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion.* Vol. 58. P. 82-115.

Bell B.W. 2021. Replacing Bureaucrats with Automated Sorcerers? *Daedalus*. 150(3). P. 89–103.

Bench-Capon T.J.M., Dunne P.E. 2007. Argumentation in Artificial Intelligence. *Artificial Intelligence.* 171(10–15). P. 619-641. DOI: 10.1016/j.artint.2007.05.001

Brooks L.F. 2020. The End of Arms Control? *Daedalus*. 149(2). P. 84–100.

Chyba C.F. 2020. New Technologies & Strategic Stability. *Daedalus*. 149(2). P. 150–170.

Crawford K. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press.

Fan M.D. 2019. *Camera Power: Proof, Policing, Privacy, and Audiovisual Big Data.* Cambridge University Press.

Harknett R.J., Nye J.S. 2017. Is Deterrence Possible in Cyberspace? *International Security.* 42(2). P. 196–199.

Katz Ya. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence.* Columbia University Press.

Kjøsen A.M., Steinhoff J., Dyer-Witheford N. 2019. *Inhumane Power.* Pluto Press.

Lee Kai-Fu. 2018. *AI Super-Powers: China, Silicon Valley and the New World Order.* Houghton Mifflin Harcourt.

Makridakis S. 2017. The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms. *Futures*. Vol. 90. P. 46-60.

Natale S. 2021. Deceitful Media: Artificial Intelligence and Social Life after the Turing Test. Oxford University Press DOI: 10.1093/oso/9780190080365.001.0001

Noveck B.S. 2021. The Innovative State. *Daedalus*. 150(3). P. 121–142.

Nye J. 2017. Deterrence and Dissuasion in Cyberspace. *International Security.* 41(3). P. 44–71.