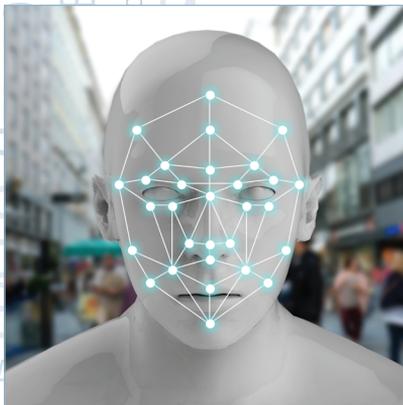
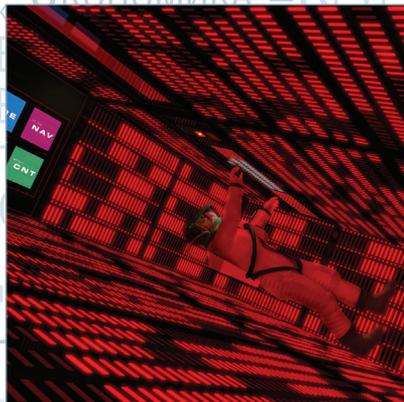




**РОССИЙСКИЙ СОВЕТ
ПО МЕЖДУНАРОДНЫМ
ДЕЛАМ**



РАБОЧАЯ ТЕТРАДЬ

**МЕЖДУНАРОДНЫЕ И СОЦИАЛЬНЫЕ
ПОСЛЕДСТВИЯ ИСПОЛЬЗОВАНИЯ
ТЕХНОЛОГИЙ ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА**

№ 44 / 2018

РОССИЙСКИЙ СОВЕТ ПО МЕЖДУНАРОДНЫМ ДЕЛАМ

МОСКВА 2018

УДК 004.8:327(100)
ББК 32.813+66.4(0)
М43

Российский совет по международным делам

Редакционная коллегия:

Главный редактор:

докт. ист. наук, член-корр. РАН **И.С. Иванов**

Авторский коллектив:

С.В. Карелов, М.В. Карлюк, канд. техн. наук **А.Г. Колонин, Н.М. Маркоткин**,
докт. естественных наук **Д.Р. Шефтелович**

Выпускающие редакторы:

канд. полит. наук **И.Н. Тимофеев**, канд. полит. наук **Т.А. Махмутов**,
канд. полит. наук **А.Ю. Толстухина, Н.М. Маркоткин, О.А. Пылова, И.С. Сорокина, А.Л. Тесля**

Международные и социальные последствия использования технологий искусственного интеллекта. Рабочая тетрадь №44 / 2018 [С.В. Карелов, М.В. Карлюк, А.Г. Колонин, Н.М. Маркоткин, Д.Р. Шефтелович]; Российский совет по международным делам (РСМД). – М.: НП РСМД, 2018 – 60 с. – Авт. и ред. указаны на обороте тит. л.

ISBN 978-5-6040388-4-0

В центре анализа настоящей работы – возможное влияние ИИ и смежных с ним технологий (машинное обучение, автономные аппараты) на международные отношения и различные сферы общественной жизни. В публикациях также рассматриваются этико-юридические аспекты применения технологий ИИ. Рабочая тетрадь Российского совета по международным делам (РСМД) включает в себя аналитические материалы, подготовленные специалистами в сферах искусственного интеллекта, машинного обучения, автономных систем, а также юристов и социологов. Публикация призвана внести вклад в общественную дискуссию по вопросам искусственного интеллекта и возможных последствий применения данной технологии.

Высказанные в рабочей тетради мнения отражают исключительно личные взгляды и исследовательские позиции авторов и могут не совпадать с точкой зрения Некоммерческого партнерства «Российский совет по международным делам».

Полный текст рабочей тетради опубликован на интернет-портале РСМД. Вы можете скачать его и оставить свой комментарий к материалу по прямой ссылке – russiancouncil.ru/paper44

© С.В. Карелов, М.В. Карлюк, А.Г. Колонин, Н.М. Маркоткин, Д.Р. Шефтелович, 2018
© Составление, оформление, дизайн обложки. НП РСМД, 2018

Содержание

Введение	4
Искусственный интеллект: время слабых	6
Искусственный интеллект – благо или угроза для человечества?	15
ИИ-национализм и ИИ-национализация	22
Дивный новый мир без работы	42
Этические и правовые вопросы искусственного интеллекта	49
Об авторах	56
Российский совет по международным делам	57

Введение

Искусственный интеллект (ИИ), еще недавно существовавший лишь в научной фантастике, на сегодняшний день является одной из наиболее перспективных и быстро развивающихся технологий. Технологии ограниченного или «слабого» искусственного интеллекта уже активно используются в самых разных областях, от мобильных телефонов и бытовой электроники до военной продукции.

В то же время, найдется мало сфер научного знания, настолько мифологизированных, как ИИ. Благодаря образам из популярной культуры, искусственный интеллект часто предстает в СМИ едва ли не ящиком Пандоры, открыв который человечество неизбежно окажется на грани апокалипсиса. Хотя подобный аларизм представляется беспочвенным, на повестке дня уже сегодня состоит разработка «сильного» искусственного интеллекта, способного к самостоятельному принятию осознанных управленческих решений.

Перспектива создания подобной технологии в значительной степени ставит под вопрос не только текущую систему глобального разделения труда, но и существующий мировой порядок и архитектуру международной безопасности. В условиях обострения противоречий и кризиса доверия между великими державами существует реальная опасность возникновения новой гонки вооружения в сфере технологий искусственного интеллекта. При этом, если контроль за оружием массового уничтожения – ядерным, химическим и биологическим – регулируется соответствующими международными договорами и конвенциями, то разработка военных технологий искусственного интеллекта на сегодняшний день остается в «серой зоне» международного права.

В то же время, на сегодняшний день очевидно, что искусственный интеллект будет все активнее использоваться в военной и гражданской сфере, в том числе в вопросах обеспечения стратегической стабильности. События получить новое более совершенное оружие и вырваться вперед в технологической гонке слишком велик, чтобы страны могли всерьез учитывать соображения общегуманитарного характера.

Одной из проблем, стоящих перед исследователями, является большая понятийная и терминологическая путаница, существующая в сфере ИИ. Данная рабочая тетрадь систематизирует существующий объем знаний о технологиях искусственного интеллекта и оценивает их перспективы их развития.

В центре анализа – возможное влияние ИИ и смежных с ним технологий (машинное обучение, автономные аппараты) на международные отношения и различные сферы общественной жизни. В публикациях также рассматриваются эτικο-юридические аспекты применения технологий ИИ.

Рабочая тетрадь Российского совета по международным делам (РСМД) включает в себя аналитические материалы, подготовленные специалистами в сферах искусственного интеллекта, машинного обучения, автономных систем, а также юристов и социологов. Материал призван внести вклад в общественную дискуссию по вопросам искусственного интеллекта и возможных последствий применения данной технологии.

Искусственный интеллект: время слабых

Шефтелович Д.Р.

Под понятием «искусственный интеллект» (ИИ), как правило, понимают компьютерную программу, т.е. *алгоритм*, способную на решение задач, которые в состоянии решить мозг взрослого человека. В Международном терминологическом словаре по искусственному интеллекту¹ понятие «искусственный интеллект» определяется как область знаний, рассматривающая разработку технологий, позволяющих вычислительным системам действовать таким образом, которое напоминает разумное поведение, в том числе, поведение человека. Стоит обратить внимание на то, что это феноменологическое определение, оставляющее детализацию понятий «интеллект», «разумное поведение» на откуп философии; в условиях дефицита знаний о мозге и когнитивном аппарате биологических систем понятие ИИ не поддаётся более точной, математической формализации.

На текущий момент под исследованиями в области искусственного интеллекта сейчас принято понимать алгоритмическое решение проблем, требующих когнитивного напряжения². Такими проблемами были (и отчасти остаются) игры: шахматы или Го, распознавание почерка, машинный перевод, творческая деятельность. В общественном сознании каждая из этих проблем первоначально казалась последней преградой к созданию «настоящего» ИИ, способного заменить человека во всех сферах его деятельности. В действительности оказалось, что, во-первых, научить компьютер играть в шахматы гораздо проще, чем научить его играть в футбол, и, во-вторых, даже зная, как научить компьютер играть в футбол, сложно перенести этот опыт на проблему машинного перевода. Поэтому после начального энтузиазма научное сообщество разделило понятие ИИ на категории *сильного* и *слабого* ИИ. Под *слабым искусственным интеллектом* понимается алгоритм, способный решать узкоспециализированные проблемы (вроде игры в шахматы), *сильный искусственный интеллект* же менее ограничен в спектре задач; в идеале сильный искусственный интеллект способен как минимум на то же, на что способен взрослый человек (например, делать логические выводы и планировать действия). Вместо понятия *сильного ИИ* в литературе также используется *общий искусственный интеллект* (Artificial General Intelligence, AGI)³.

¹ Raynor W. Jr International Dictionary of Artificial Intelligence. London, United Kingdom: Global Professional Publishing. 2008.

² Boulanin V., Verbruggen M., SIPRI Mapping the development of autonomy in weapon systems. SIPRI, Solna: SIPRI. 2017.

Floreano D., Mattiussi C. Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies. The MIT Press. 2008.

³ Yudkowsky E. Levels of Organization in General Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg. 2007. pp. 389–501.

Необходимо отметить, что классификация алгоритмических проблем на проблемы сильного и слабого ИИ не является исторической константой; так, на начало 2016 г. считалось, что для игры в Го нужен сильный ИИ. После впечатляющей победы алгоритма AlphaGo⁴ над действующим чемпионом мира в марте 2016 г. игра в Го стала проблемой слабого ИИ.

К концу 2010-х годов можно выделить следующие магистральные направления исследований в области алгоритмов искусственного интеллекта.

Машинное обучение и распознавание образов

Поиск закономерностей в данных, например, классификация объектов на фотографии по категориям «фон», «человек», «автомобиль», «здание», «растение».

Планирование и логический вывод

Доказательство логических утверждений и планирование действий для достижения определённой цели при знании логических закономерностей, позволяющих эту цель достичь. Пример: синтез данных из сенсоров для оценки дорожной ситуации и адекватного управления автомобилем.

Экспертные системы

Систематизация данных вместе с логическими связями, отображение знаний, ответы на семантические запросы вроде «Какова доля цен на энергоносители в стоимости производства самолёта МС-21?».

Текущее состояние и перспективы

Проблема создания сильного ИИ на данный момент не решена, и научное сообщество существенно охладело к теме ИИ после так называемой «зимы искусственного интеллекта», наступившей в середине 1980-х из-за завышенных ожиданий. Это привело, с одной стороны, к разочарованию потенциальных пользователей и к недостаточной производительности программных комплексов, с другой⁵. Начиная с середины 2000-х годов удалось добиться значительных успехов в более специализированных задачах. В первую очередь это обусловлено постоянным развитием вычислительной техники: если в 2001 г. пределом возможностей был «флагманский» процессор Intel Pentium III, выполнявший 1,4 миллиарда арифметических операций в секунду, то, спустя десять лет, в 2011 году актуальный на тот момент процессор Intel Core i5 выполнял уже 120 миллиардов операций в секунду, т.е. почти в 100 раз больше⁶. Во многом драйвером роста стала индустрия ком-

⁴ Silver D. et al. Mastering the game of Go with deep neural networks and tree search // Nature. 2016. (529). pp. 484–503.

⁵ Рассел С., Норвиг П. Искусственный интеллект: современный подход. Москва: Вильямс. 2007.

⁶ Intel Corporation. Intel® Core i5-2500 Desktop Processor Series // Intel Corporation. 2011.

URL: https://www.intel.com/content/dam/support/us/en/documents/processors/corei5/sb/core_i5-2500_d.pdf
Dongarra J.J. Performance of various computers using standard linear equations software // Netlib. 2014.
URL: <http://www.netlib.org/benchmark/performance.pdf>

пьютерной графики и видеоигр: в погоне за всё более реалистичной графикой видеокарты из периферийного оборудования превратились в мощные вычислительные системы, способные не только обрабатывать графические данные, но и выполнять произвольные (пусть и не очень сложные) параллельные вычисления. Так, в том же 2011 году пиковая производительность видеокарты Radeon HD 6970 составляла 675 миллиардов операций в секунду⁷.

Вторым важным фактором развития отрасли стала оцифровка и ручная классификация текстов, фотографий и звуковых записей, а также построение всеобъемлющих баз знаний в цифровом формате. Благодаря ставшим доступными большим объёмам качественно классифицированных данных стало возможно тренировать алгоритмы машинного обучения на больших выборках. Таким образом, точность классификации выросла⁸, а машинный перевод превратился из грубого инструмента в систему широкой применимости. Так, известный большинству пользователей переводчик компании Google работает за счёт огромной базы *параллельных текстов*, т.е., текстов идентичного значения на разных языках, дополняемой пользователями, что позволяет обучать систему и улучшать качество перевода прямо во время работы⁹.

В сумме оба фактора позволили решать за приемлемое время всё более сложные задачи, сократив время обработки накопленных массивов данных. На начало 2018 г. искусственный интеллект добился нескольких ярких успехов, освещённых СМИ, проник в различные отрасли экономики. Однако было бы ошибкой считать ИИ «серебряной пулей», средством, способным решить все проблемы человечества.

Успехи ИИ

Экспертные системы и распознавание текстов: своя игра IBM Watson

Пожалуй, наибольший общественный резонанс произвела экспертная система IBM Watson, объединившая в себе огромный массив знаний, т.е., данных с семантическими связями между ними, и способность обрабатывать запросы на английском языке к этой базе знаний. В 2011 году впечатляющим успехом IBM Watson была победа над действующими чемпионами в телевикторине «Jeopardy!»¹⁰. Этот успех позволил IBM успешно выйти на рынок экспертных систем, наглядно продемонстрировав способности Watson к обработке и структурированию информации.

⁷ Geeks3D. AMD Radeon and NVIDIA GeForce FP32/FP64 GFLOPS Table // Geeks3D. 2014.

URL: <http://www.geeks3d.com/20140305/amd-radeon-and-nvidia-geforce-fp32-fp64-gflops-table-computing/>

⁸ Saarikoski J. et al. On the Influence of Training Data Quality on Text Document Classification Using Machine Learning Methods // Int. J. Knowl. Eng. Data Min. 2015. № 2 (3). pp. 143–169.

⁹ DeNero J., Uszkoreit J. Inducing Sentence Structure from Parallel Corpora for Reordering // Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2011.

¹⁰ Также известна в России под названием «Своя игра» – прим. авт.

Нейросети: Победа в Го с Google DeepMind

Во многом выиграла от повышения производительности вычислений технология искусственных нейронных сетей, известная ещё с середины 1940-х годов¹¹. Современные технологии позволяют обучать большие нейросети, а от размера сети прямо зависит сложность решаемых задач. Характерной чертой этих систем *глубинного обучения* является структура сети, позволяющая распознавать сначала локальные детали входных данных (вроде разницы в цвете между соседними пикселями), а по ходу обработки и их глобальные свойства (вроде линий и форм). Построив и натренировав нейросеть с подобной архитектурой, инженеры Google смогли сконструировать компьютерный алгоритм игры в Го, неожиданно выигравший у чемпиона мира¹². Схожих результатов другие группы исследователей добились, научив нейросети играть в компьютерные игры вроде StarCraft II¹³.

Практическое применение ИИ

На начало 2018 года машинное обучение, распознавание образов и автономное планирование вышли из исследовательских лабораторий в мир коммерческих приложений. Первыми пользователями новой технологии стали военные по обе стороны железного занавеса, начиная с 1950-х годов заинтересовавшиеся в решении задач планирования¹⁴. Практически одновременно с ними задачи планирования стали решать экономисты. Приведём несколько примеров применения алгоритмов искусственного интеллекта.

- IBM предлагает услуги упомянутого выше Watson различным отраслям: медикам для диагностики симптомов пациента и рекомендации терапии, юристам для классификации конкретных ситуаций в соответствии с правовыми нормами, железнодорожникам для оценки усталости составов и путей¹⁵.
- В области медицины распознавание образов позволяет опознать и классифицировать органы для планирования хирургических процедур¹⁶.
- Онлайн-магазины пользуются механизмами машинного обучения для более точной рекомендации товаров регулярным клиентам¹⁷.
- Автономные роботизированные музейные гиды¹⁸ способны проводить экскурсии и отвечать на вопросы посетителей по теме экспозиции.

¹¹ McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in nervous activity // The Bulletin of Mathematical Biophysics. 1943. № 4 (5). pp. 115–133.

¹² Silver D. et al. Mastering the game of Go with deep neural networks and tree search // Nature. 2016. (529). pp. 484–503.

¹³ Vinyals O. et al. StarCraft II: A New Challenge for Reinforcement Learning // CoRR. 2017.

¹⁴ Keneally S.K., Robbins M.J., Lunday B.J. A Markov decision process model for the optimal dispatch of military medical evacuation assets // Health Care Management Science. 2016. № 2 (19). pp. 111–129.

Цыгичко В.Н. Модели в системе принятия военно-стратегических решений в СССР. Москва: Империиум Пресс. 2005.

¹⁵ AI Stories // IBM. 2018. URL: <https://www.ibm.com/watson/ai-stories/index.html>

¹⁶ Alpers J. et al. CT-Based Navigation Guidance for Liver Tumor Ablation // Eurographics Workshop on Visual Computing for Biology and Medicine. 2017.

¹⁷ Aggarwal C.C. Recommender Systems. Heidelberg: Springer International Publishing, 2016.

¹⁸ Diallo A.D., Gobe S., Durairajah V. Autonomous Tour Guide Robot using Embedded System Control // Procedia Computer Science. 2015. (76). pp. 126–133.

- В военной сфере уже сейчас можно увидеть зачатки автономного принятия решений: комплексы ближней противовоздушной и противоракетной обороны исключают человека из цепочки принятия решений из-за низкой скорости реакции человеческих операторов, а современные противокорабельные ракеты способны распределять цели между ракетами в залпе в зависимости от их важности.

Перспективы применения ИИ

Ожидается, что в 2020-х годах ИИ сможет решать ещё больше задач. Ниже перечислим некоторые из них и ориентировочный прогресс научно-исследовательских и опытно-конструкторских работ.

Автономные транспортные средства

Наиболее сложной задачей для автономных транспортных средств считается участие в дорожном движении в связи с ограниченной свободой манёвра и множеством возможных ситуаций, возникающих на дороге. В 2018 г. самым высокоразвитым решением на рынке считается автопилот Audi A8, способный автономно двигаться в условиях пробок на автострадах¹⁹. В марте 2017 г. представители BMW обещали к 2021 году представить автомобиль, способный без вмешательства водителя доставлять пассажиров в место назначения²⁰.

Военное применение

Самым перспективным направлением для военных считается возможность автоматического распознавания и сопровождения целей для роботизированных платформ, и, как следствие, автономное принятие решения об их поражении²¹. На конец 2017 года в странах «первого эшелона» проводится ряд научно-исследовательских работ по этим темам, в первую очередь в отношении наземных, надводных и подводных комплексов.

Роботизированные хирургические процедуры

Ожидается, что хирургия, в том числе и микрохирургия, требующая точного вмешательства в тело пациента, сможет быть автоматизирована в ближайшем будущем. В 2016 г. был представлен в качестве демонстратора технологий робот STAR²², способный проводить операции на мягких тканях. При достижении успехов в распознавании образов можно ожидать роботизации всех хирургических процедур удаления тканей, что позволит сделать хирургию более доступной и снизить нагрузку на лечащий персонал.

¹⁹ Ross P.E. The Audi A8: the World's First Production Car to Achieve Level 3 Autonomy // IEEE Spectrum. 2017. URL: <https://spectrum.ieee.org/cars-that-think/transportation/self-driving/the-audi-a8-the-worlds-first-production-car-to-achieve-level-3-autonomy>

²⁰ Reuters. BMW says self-driving car to be level 5 capable by 2021 // Reuters. 2017. URL: <https://www.reuters.com/article/us-bmw-autonomous-self-driving/bmw-says-self-driving-car-to-be-level-5-capable-by-2021-idUSKBN16N1Y2>

²¹ Boulanin V., Verbruggen M., RIKEN. Op. cit.

²² Shademan A. et al. Supervised autonomous robotic soft tissue surgery // Science Translational Medicine. 2016. № 337 (8). p. 337.

Нерешённые проблемы

Даже для кажущихся всемогущими компьютерных алгоритмов существуют нерешаемые и плохо решаемые проблемы. В частности, по очевидным причинам сложно решить задачи, трудно поддающиеся формализации, например, «написать роман» или «выбрать самую красивую фотографию». Но даже при наличии математически точной формулировки нет гарантии успешного решения. Причиной этому может быть сложность математического моделирования проблем более низкого уровня (например, моделирование движения при обучении робота играть в футбол), сложность самой проблемы (например, для логического вывода и доказательства математических утверждений нет алгоритмов существенно лучше полного перебора возможных логических цепочек²³), огромное количество параметров и неточность наблюдаемого мира (например, при игре в футбол), и недостаточная производительность вычислительных систем по сравнению с человеческим мозгом. В конце концов, взаимодействие 1×10^{11} человеческих нейронов не поддается алгоритмической симуляции; на начало 2018 года наибольшим успехом в этой области была симуляция $1,7 \times 10^9$ нейронов с 2400-кратным замедлением в 2013 году²⁴; при этом нет уверенности, что симуляция даже нужного количества клеток мозга позволит воссоздать мозговую деятельность в компьютерной модели.

Отдельно стоит упомянуть недостатки машинного обучения. Как правило, машинное обучение предполагает наличие уже классифицированных тренировочных данных, в которых компьютерный алгоритм находит закономерности. Из-за недостаточных данных в работе алгоритма возможны ситуации, когда данные на входе не принадлежат ни к одному из тренированных классов; распознавание нового феномена на входе и автоматическое создание нового класса объектов считается как минимум нетривиальной задачей²⁵; её можно легко усложнить, добавив в условие активное обучение во время работы классификатора и временное изменение распознаваемых классов. Вторым существенным недостатком машинного обучения считается сверхчувствительность: так, распознавание лиц можно «перехитрить», надев кажущиеся безобидными очки²⁶. В отдельных случаях можно добиться неправильной классификации фотографии путём невидимых человеку изменений: так, после кажущейся несущественной манипуляции панда на фотографии может классифицироваться как обезьяна²⁷.

В то время, как компьютеры успешно решают «сложные» задачи вроде символических и численных вычислений и обыгрывают гроссмейстеров

²³ Холкрофт Д., Мотвани Р., Ульман Д. Введение в теорию автоматов, языков и вычислений. Москва: Вильямс. 2008.

²⁴ RIKEN, Forschungszentrum Jülich. Largest neuronal network simulation achieved using K computer // RIKEN. 2013. URL: http://www.riken.jp/en/pr/press/2013/20130802_1/

²⁵ Scheirer W.J., Jain L.P., Boulton T.E. Probability Models for Open Set Recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI). 2014. № 11 (36).

²⁶ Sharif M. et al. Adversarial Generative Nets: Neural Network Attacks on State-of-the-Art Face Recognition // CoRR. 2017.

²⁷ Goodfellow I., Shlens J., Szegedy C. Explaining and Harnessing Adversarial Examples // International Conference on Learning Representations. 2015.

в шахматы, в списке нерешённых проблем много задач, отличающихся от-носительно простой формулировкой: классификация «неизвестных» образ-зов без тренировки на заранее классифицированных образцах (например, распознавание яблок на фотографии, если известны только классы «виш-ня» и «груша»), моторика, рассуждения от «здорового смысла». В литерату-ре этот феномен известен как «парадокс Моравека»²⁸. Во многом это ото-бражает человеческое восприятие: создававшиеся миллионами лет эволю-цией способности, присутствующие у любого взрослого человека, кажутся само собой разумеющимися, в то время, как математические проблемы вроде нахождения кратчайшего пути на карте кажутся неестественными и их решение малоочевидным.

Мифы и легенды ИИ

За полвека исследований тема искусственного интеллекта обросла мас-сой мифов и ложных представлений как о возможностях технологий, так и об их недостатках. Так, в области машинного перевода зачастую можно встретить утверждение о низкокачественных автоматических переводчи-ках 1960-х годов, превративших фразу «дух бодр, плоть слаба» в «водка хороша, но мясо протухло» при переводе с английского языка на англий-ский, минуя русский. Изначально этот пример был упомянут в качестве об-разца неправильного перевода человеком, вооружённым лишь словарём, грамматикой и излишне буйным воображением; достоверных свиде-тельств такой работы систем машинного перевода нет²⁹.

В целом можно выделить несколько мифов, искажающих общественное представление об искусственном интеллекте.

ИИ – это магия

Если не принимать во внимание известное изречение Артура Кларка, со-гласно которому «любая достаточно развитая технология неотличима от магии», современные исследования в области ИИ произрастают из обла-стей математики, робототехники, статистики и информатики; из вышеска-занного в этой статье следует, что любой программный или программно-аппаратный комплекс, реализующий ИИ, в первую очередь, решает несколь-ко математических проблем, зачастую основываясь на заранее собранных и классифицированных данных.

ИИ – это компьютерный мозг

В современной культуре прочно засело представление об искусственном интеллекте как, в первую очередь, антропоморфном механизме, во мно-гом похожем на человеческий мозг. В реальности даже нейронные сети – это математическая абстракция, довольно далеко ушедшая от биологиче-ского механизма; другие модели вроде цепей Маркова вообще не имеют

²⁸ Moravec H. Mind Children: The Future of Robot and Human Intelligence. MA: Harvard University Press. 1990.

²⁹ Hutchins J. «The whicky was invisible», or Persistent myths of MT // MT News International. 1995. (11). pp. 17–18.

реальных биологических аналогов. Во многом, применяемые алгоритмы не нацелены на моделирование сознания, а специализированы на решении одной узко формализованной проблемы.

ИИ имеет лишь несколько сценариев применения

В привитом культурой представлении ИИ редко занят чем-то кроме планирования действий в интересах крупных корпораций либо государств. Однако упомянутые выше возможности позволяют применять ИИ-технологии в интересах широких групп населения: применяющиеся на практике алгоритмы доступны каждому, а в спектр задач машинного обучения входят вполне обыденные задачи вроде оценки потребностей в ремонте таксопарка и планирования оптимального использования операционных залов.

Опасности ИИ

В апокалиптических сценариях научно-фантастических произведений искусственный интеллект, как правило, в какой-то момент решает уничтожить человечество, а учёные не в состоянии этому помешать либо не осознают эту опасность. На практике в научном сообществе вопрос опасности сверх-интеллекта обсуждается достаточно давно³⁰. Основной опасностью считается неправильное целеполагание «сильному» ИИ, располагающему существенными вычислительными и материальными ресурсами, и не принимающему во внимание интересы людей. Цитируемая выше книга Бострома предлагает возможные решения этой проблемы.

Реальность сильного ИИ

Завышенные либо заниженные ожидания на успехи научно-технического прогресса влекут за собой прогнозы, в которых сильный ИИ либо станет доступным уже завтра (в крайнем случае – в следующем году), либо принципиально недостижим. Фактически создание сильного искусственного интеллекта сложно спрогнозировать, так как оно зависит от успешного решения нескольких инженерных и математических задач неизвестной сложности. Исторический пример решения теоремы Ферма, появившегося спустя 350 лет после её формулировки, не позволяет с уверенностью прогнозировать сроки решения задач такого масштаба.

Международные тенденции

Расширение границ применимости ИИ привело к повышенному интересу военных и околовоенных кругов к возможностям автономных систем. Помимо НИОКР по упомянутым выше темам, в международном сообществе разгорелась дискуссия о возможном ограничении или даже запрете робототехнических комплексов. Наибольшую известность получила кампания «Stop Killer Robots»³¹, активно требующая полного запрета на разработку автономных боевых систем по этическим причинам. Однако в этой связи

³⁰ Bostrom N. Superintelligence: Paths, Dangers, Strategies. Oxford, UK: Oxford University Press, 2014.

³¹ Campaign to Stop Killer Robots. URL: <https://www.stopkillerrobots.org/>

стоит упомянуть не только боевых роботов, но и системы классификации, влияющие на решение о применении силы исключительно по метаданным, даже не обращая внимания на содержание писем подозреваемых³².

Общественное внимание к автономным боевым системам привело к переговорам в рамках *Конвенции ООН о запрещении или ограничении применения конкретных видов обычного оружия, которые могут считаться наносящими чрезмерные повреждения или имеющими неизбирательное действие*, под действие которой подпадают противопехотные мины и ослепляющие лазеры. Однако на начало 2018 года дипломатический процесс не привёл к каким-либо взаимным обязательствам³³. Во многом это обусловлено сложностью в определении «автономной боевой системы», наличием комплексов (в первую очередь ПВО и ПРО), удовлетворяющих возможному определению, и нежеланием отказываться от перспективной технологии³⁴. Тем не менее, было бы излишним упрощением ограничивать аргументы за контроль над автономными вооружениями морально-этическими аспектами. Во многом в пользу ограничений говорят аспекты стратегической стабильности³⁵: во-первых, применение автономных систем может привести к неконтролируемой эскалации военных конфликтов с непредсказуемыми последствиями; во-вторых, сложно контролировать экспорт используемых технологий: если экспорт крылатой ракеты очевиден и её дальность ограничена физическими параметрами, то контроль такого уровня над программным кодом невозможен: область применения алгоритмов не ограничена одной лишь военной сферой, а ограничение исследований в области с настолько широким спектром применения, вплоть до запрета публикаций, неизбежно натолкнётся на сопротивление научного сообщества, во многом живущего за счёт международной кооперации и крупных конференций.

³² Экс-глава ЦРУ и АНБ: Мы убиваем людей, используя метаданные // RT на русском. 2014.
URL: <https://russian.rt.com/article/31734>

³³ The United Nations Office at Geneva. CCW Meeting of High Contracting Parties. 2017.
URL: <https://www.unog.ch/80256EE600585943/httpPages/A0A0A3470E40345CC12580CD003D7927?OpenDocument>

³⁴ Boulanin V., Verbruggen M., Op. cit.

³⁵ Altman J., Sauer F. Autonomous Weapon Systems and Strategic Stability // Survival. 2017. № 5 (59). pp. 117–142.

Искусственный интеллект – благо или угроза для человечества?

Колонин А.Г.

Основные проблемы развития и потенциальные угрозы искусственного интеллекта

Несмотря на множество современных достижений в нейробиологии, на сегодняшний день пока никто точно не знает, как устроен естественный интеллект. Соответственно, точно так же никто не знает, как именно создать искусственный интеллект. Существует ряд известных проблем, требующих решения для его создания и разные мнения по поводу приоритетности достижения тех или иных решений. Например, руководитель международных проектов по созданию искусственного интеллекта с открытым кодом OpenCog и SingularityNET Бен Герцель считает, что все необходимые технологии для создания общего ИИ в принципе уже разработаны, необходимо только соединить их некоторым правильным образом для получения такой синергии, результатом которой станет возникновение общего ИИ³⁶. Другие эксперты настроены более скептически, полагая, что необходимо принципиальное решение многих проблем, которые будут перечислены ниже. Также сильно варьируются экспертные оценки срока возникновения сильного ИИ – от десятка до нескольких десятков лет.

Возникновение систем даже просто автономного или адаптивного, а тем более общего или сильного ИИ связывается с несколькими угрозами разного масштаба, которые актуальны уже сегодня. Среди них можно отметить:

- создание так называемых «автономных систем смертоносных вооружений» – Lethal Autonomous Weapons Systems (LAWS), например, дронов для заказных убийств;
- новый виток гонки вооружений, при котором государства будут совершенствовать уровни интеллекта автономных средств поражения и уничтожения;
- интеллектуальная система, не обязательно боевая, но и промышленная или бытовая, способна не только к целенаправленному действию, но и к сознательному целеполаганию, при том что автономная постановка целей системы может привести к постановке целей, противоречащих целям человека;
- исключение из материального производства подавляющей части населения за счет еще большей автоматизации, что может приводить к еще

³⁶ Ben Goertzel, Cassio Pennachin, Nil Geiswiler, Engineering General Intelligence, Part 1: A Path to Advanced AGI via Embodied Learning and Cognitive Synergy (Atlantis Thinking Machines), Atlantis Press. 2014.
URL: <https://www.amazon.com/Engineering-General-Intelligence-Part-Cognitive/dp/9462390266>

большему социальному расслоению, снижению эффективности «социальных лифтов» и увеличению массы «лишних людей» с соответствующими социальными последствиями.

Сложность контроля за системами искусственного интеллекта на сегодняшний день обусловлена, в частности, «закрытостью» существующих прикладных решений на основе «глубоких нейронных сетей», которые не позволяют не только верифицировать правильность принятия решений перед их исполнением, но даже по факту проводить анализ решения, которое было принято машиной. Решению этой проблемы сейчас посвящено как новое направление «объяснимый искусственный интеллект» (Explainable Artificial Intelligence, EAI), так и новый интерес к интеграции ассоциативного (нейро-сетевое) и символического (основанного на логике) подходов к проблеме.

В августе 2018 г. в Праге на площадке Чешского Технического Университета одновременно прошли конференции, посвященные искусственному интеллекту человеческого уровня³⁷, общему искусственному интеллекту³⁸, биологически-вдохновленным когнитивным архитектурам³⁹, а также нейро-символьным технологиям⁴⁰. На конференциях были представлены доклады ведущих специалистов компаний и учреждений, лидирующих в сфере исследовательских разработок в области искусственного интеллекта (ИИ): Microsoft, Facebook, DARPA, MIT, Good AI. В этих докладах были обозначены как текущее состояние разработок в области ИИ, так и стоящие перед обществом нерешенные проблемы, а также угрозы, возникающие в ходе дальнейшего развития этой технологии. В этом обзоре я постараюсь кратко обозначить основные проблемы и угрозы, а также возможные пути противостояния этим угрозам.

Однако, прежде всего, необходимо уточнить значение некоторых терминов, которые обычно употребляются совместно с ИИ в различных контекстах: *слабый или специализированный ИИ*, *автономный ИИ* (Autonomous AI), *адаптивный ИИ* (Adaptive AI), *общий ИИ* (Artificial General Intelligence, AGI), *сильный ИИ* (Strong AI), *ИИ человеческого уровня* (Human-Level AI), *ИИ сверхчеловеческого уровня* (Super-human AI).

Слабый или специализированный ИИ представлен всеми без исключения существующими решениями и предполагает способность автоматизации решения одной конкретной задачи, будь то игра в Go или распознавание лиц на видеорекамерах. При этом отсутствует возможность самостоятельного обучения другим задачам без перепрограммирования человеком.

Автономный ИИ предполагает возможность системы функционировать долгое время без участия оператора. Например, он позволяет дрону, обору-

³⁷ Human Level AI. The Joint Multi-Conference on Human-Level Artificial Intelligence // HLAИ. 2018. URL: <https://www.hlai-conf.org/>

³⁸ Conference Series on Artificial General Intelligence // AGI. 2018. URL: <http://agi-conference.org/>

³⁹ Biologically Inspired Cognitive Architectures Society // BICA Society. 2018. URL: <http://bicasociety.org/>

⁴⁰ Neural-Symbolic Integration. Workshop series on Neural-Symbolic Learning and Reasoning // Neural Symbolic Integration. 2018. URL: <http://dasefab.cs.wright.edu/nesy/>

дованному солнечными батареями, совершить многодневное путешествие с Елисейских полей на Красную площадь или в обратном направлении, самостоятельно выбирая как маршрут, так и места для промежуточных посадок для подзарядки аккумуляторов, избегая при этом всевозможные препятствия.

Адаптивный ИИ предполагает способность системы адаптироваться к новым условиям, приобретая знания, не закладываемые при создании. Например, позволить системе поддержания диалогов на русском языке самостоятельно осваивать новые языки и применять их знание в разговоре, попадая в новую языковую среду или на основе изучения учебных материалов для этих языков.

Общий ИИ предполагает настолько высокую адаптивность, что обладающая им система может быть использована в самых различных видах деятельности при соответствующем обучении. Обучение может быть как самостоятельным, так и направленным (с помощью инструктора). В этом же смысле в противопоставление слабому или специализированному ИИ также часто употребляется сильный ИИ.

ИИ человеческого уровня предполагает уровень адаптивности сравнимый с человеческим, то есть система способна осваивать те же самые навыки, что и человек в сопоставимые сроки обучения.

ИИ сверхчеловеческого уровня предполагает еще более высокую адаптивность и скорость обучения. Таким образом, система может обучиться тем знаниям и способностям, которые человеку в принципе не под силу.

Принципиальные проблемы создания сильного ИИ

Возникновение сильного искусственного интеллекта вполне закономерно в рамках общего эволюционного процесса, как закономерно и возникновение молекул из атомов, клеток из молекул, организмов из клеток, выделение специализированных клеток в центральную нервную систему, возникновение социальных структур, развитие речи, письменности и в конечном итоге – информационных технологий. Закономерность нарастающей сложности информационных структур и способов организации в процессе эволюции хорошо показана Валентином Турчиным⁴¹. Если не произойдет гибель человеческой цивилизации, то такая эволюция будет неизбежной, и в самой долгосрочной перспективе это будет спасением человечества, постольку только не биологические формы существования информации смогут пережить неизбежную со временем гибель Солнечной системы и смогут сохранить во Вселенной информационный код нашей цивилизации.

При этом важно осознавать, что для того чтобы построить сильный искусственный интеллект, не обязательно понимать, как устроен естественный, так как необязательно понимать, как летает птица, чтобы сделать ракету.

⁴¹ В. Ф. Турчин, Феномен науки: Кибернетический подход к эволюции. М.: ЭТС. 2000. 368 с.
URL: <http://refal.net/turchin/phenomenon/>

Очевидно, это будет сделано рано или поздно тем или иным способом или, возможно, несколькими способами.

В качестве принципиальных проблем⁴², решение которых еще предстоит для создания общего или сильного ИИ, большинство экспертов выделяют следующие:

- **Быстрое обучение** (few-shot learning) – необходимость построения систем, обучающихся на небольшом объеме материала в отличие от существующих систем глубокого обучения, требующих большие объемы специально подготовленного обучающего материала.
- **Сильная генерализация** (strong generalisation) – создание технологий распознавания ситуаций, в которых распознаваемые объекты встречаются в условиях, отличных от тех, в которых они встречались в использованном для обучения материале.
- **Генеративные или генерирующие модели обучения** (generative models) – разработка технологий обучения, когда объектом запоминания являются не признаки объекта распознавания, а принципы его формирования, что может позволить отражать более глубокие сущности распознаваемых объектов и осуществлять более быстрое обучение и более сильную генерализацию.
- **Структурированное обучение и предсказание** (structured prediction and learning) – развитие технологий обучения на основе представления объектов обучения в виде многослойных иерархических структур, где более низкоуровневые элементы определяют более высокоуровневые, что может оказаться альтернативным решением проблем быстрого обучения и сильной генерализации.
- Решение проблемы **катастрофического забывания** (catastrophic forgetting) – присущего большинству существующих систем, которые, будучи изначально обучены на одном классе объектов и затем до-обучены распознаванию на новом классе объектов, теряют способность распознавать объекты первого класса.
- Достижение **возможности инкрементального обучения** (incremental learning), предполагающего способность системы накапливать знания и совершенствовать свои возможности постепенно, не теряя при этом полученные ранее знания, но приобретая новые знания применительно к системам диалогового общения на естественном языке. Идеальным является прохождение «младенческого теста Тьюринга»⁴³ (Baby Turing Test), в случае которого система должна продемонстрировать возможность постепенного освоения языка от уровня младенца до уровня взрослого человека.

⁴² З. М. Iklé, A. Franz, R. Rzepka, B. Goertzel, B. (Eds.), Artificial General Intelligence, 11th International Conference, AGI 2018, Prague, Czech Republic. 2018. URL: <https://www.springer.com/us/book/9783319976754>

⁴³ B.H. Partee, S. Peters, R. Thomason. (Eds.), Report of Workshop on Information and Representation. Washington. D.C.. March 30 through April 1. 1985. URL: <https://files.eric.ed.gov/fulltext/ED261533.pdf>

- Решение проблемы **сознания** (consciousness) – предполагает формирование проверенной рабочей модели сознательного поведения, обеспечивающего эффективное прогнозирование и целенаправленное поведение за счет формирования «внутренней картины мира» в рамках которой возможен поиск оптимальных стратегий поведения по достижению поставленных целей без фактического взаимодействия с реальным миром, что существенно повышает безопасность, проверки гипотез, а также повышает скорость и энергетическую эффективность этой проверки, тем самым создавая возможность самообучения живой или искусственной системы в «виртуальном мире» собственного сознания. С прикладной точки зрения проблема сознания имеет две стороны. С одной – создание систем ИИ, обладающих сознанием, позволит резко повысить их эффективность. С другой – появление сознания у таких систем вызывает как дополнительные риски, так и вопросы этического плана, поскольку такие системы в какой-то момент смогут быть по уровню самосознания приравнены к самому человеку с вытекающими из этого последствиями в правовом поле.

Потенциальные угрозы ИИ

Возникновение систем даже просто автономного или адаптивного, а тем более общего или сильного ИИ связывается с несколькими угрозами разного масштаба, актуальными уже сегодня.

Во-первых, угрозу для человека может представлять интеллект не обязательно сильный, общий, человеческого или сверх-человеческого уровня, так как достаточно иметь автономную систему, оперирующую большими объемами информации с большими скоростями. На её основе могут быть созданы так называемые «автономные системы смертоносных вооружений»⁴⁴ – Lethal Autonomous Weapons Systems (LAWS), простейший пример которых – дроны для заказных убийств, печатаемые на 3D-принтерах как в массовых масштабах, так и небольшими партиями в кустарных условиях.

Во-вторых, угрозу для государства может представлять ситуация, когда другое государство (потенциальный противник) получает вооружения с более адаптивным, автономным и общим искусственным интеллектом с повышенной скоростью реакции и предсказательной способностью.

В-третьих, угрозу для всего мира представляет вытекающая из предыдущей угрозы ситуация, когда государства вступают в новый виток гонки вооружений, совершенствуя уровни интеллекта автономных средств поражения и уничтожения – как это было предсказано Станиславом Лемом несколько десятков лет назад⁴⁵.

⁴⁴ Metro. Killer robots are fast becoming a reality – we must stop this from happening if we want to stop a global AI arms race // Metro. 2018.

URL: <https://metro.co.uk/2018/09/02/killer-robots-are-fast-becoming-a-reality-we-must-stop-this-from-happening-if-we-want-to-stop-a-global-ai-arms-race-7903717/?ito=cbsshare>

⁴⁵ Станислав Лем, Мир на Земле, антология Операция «Вечность». М.: Мир. 1988.

AI Safety Research: A Road to Nowhere. 2016.

URL: <https://medium.com/@petervoss/ai-safety-research-a-road-to-nowhere-f1c7c20e8875>

В-четвертых, угрозой для любой стороны представляет любая, не обязательно боевая, но и промышленная или бытовая интеллектуальная система с определенной степенью автономности и адаптивности, способная не только к целенаправленному действию, но и к сознательному целеполаганию, при том что автономная постановка целей системы может привести к постановке целей, противоречащих целям человека и людей, а возможностей достижения этих целей у системы будет намного больше, в силу её более высокого быстродействия, большего объема обрабатываемой информации и большей предсказательной способности. К сожалению, масштабы именно этой угрозы сообществом не вполне изучены и осознаны.

В-пятых, угрозой для общества представляет переход к новому уровню развития производственных отношений в капиталистическом (либо тоталитарном) обществе, когда более малочисленная часть населения получает возможность контролировать материальное производство, исключая из него подавляющую часть населения за счет еще большей автоматизации, что может приводить к еще большему социальному расслоению, снижению эффективности «социальных лифтов» и увеличению массы «лишних людей» с соответствующими социальными последствиями.

Наконец, угрозой для человечества в целом может представлять автономизация глобальных вычислительных систем обработки данных, распространения информации и принятия решений на основе глобальных сетей, поскольку скорость распространения информации в таких системах и масштаб воздействия может приводить к непредсказуемым с позиций имеющегося опыта и существующих моделей управления социальным явлениям. Например, внедряемая система социального кредита⁴⁶ в современном Китае является уникальным экспериментом цивилизационного масштаба с непонятными на сегодняшний день последствиями.

Сложность контроля за системами искусственного интеллекта на сегодняшний день обусловлена, в частности, «закрытостью» существующих прикладных решений на основе «глубоких нейронных сетей», которые не позволяют не только верифицировать правильность принятия решений перед их исполнением, но даже по факту проводить анализ решения, которое было принято машиной. Решению этой проблемы сейчас посвящено как новое направление «объяснимый искусственный интеллект»⁴⁷ (Explainable Artificial Intelligence, EAI), так и новый интерес к интеграции ассоциативного (нейросетевого) и символьного (основанного на логике) подходов к проблеме.

Пути противостояния угрозам

Представляется безусловно необходимым принятие следующих мер для предотвращения катастрофических сценариев развития технологий ИИ и их применения.

⁴⁶ Большой брат 2.0. Как Китай строит цифровую диктатуру // Московский центр Карнеги. 2017.
URL: <https://carnegie.ru/commentary/71546>

⁴⁷ Explainable Artificial Intelligence (XAI) // 2017. DARPA/I2O.
URL: <https://www.darpa.mil/attachments/XAIProgramUpdate.pdf>

- Международный запрет «автономных систем смертоносных вооружений» (LAWS) и разработка и внедрение мер международного контроля за его исполнением.
- Государственная поддержка работ, направленных на решение обозначенных проблем, в особенности «объяснимого искусственного интеллекта» и интеграции различных подходов, а также изучения принципов работы создания механизмов целеполагания с целью получения эффективных средств программирования и контроля интеллектуальных систем, когда средством программирования будут не правила, а ценности, а контролироваться должны не действия, а цели.
- Демократизация доступа к технологиям и методам ИИ, например, за счет реинвестирования доходов от внедрения интеллектуальных систем в массовое обучение вычислительным и когнитивным технологиям, а также создание решений ИИ с открытым кодом и разработка мер стимулирования открытия кодов существующими «закрытыми» системами ИИ. Например, проект Aigents⁴⁸ направлен на создание персональных агентов ИИ для массовых пользователей, работающих автономно и не подверженных централизованной манипуляции.
- Регламентирование на межгосударственном уровне открытости алгоритмов ИИ, протоколов работы распределенных систем обработки данных и принятия решений на их основе с возможностью независимого аудита как международными и государственными органами, так и частными лицами. Одной из инициатив в этом направлении является создание платформы и экосистемы приложений ИИ с открытым кодом SingularityNET⁴⁹.

⁴⁸ Aigents // Aigents. URL: <https://aigents.com/>

⁴⁹ SingularityNET // SingularityNET. <https://singularitynet.io/>

ИИ-национализм и ИИ-национализация

Карелов С.В.

ИИ как геополитический фактор

Технологическое неравенство и раньше сильно влияло на мировую политику и экономику. Самые технологически продвинутые государства становились самыми экономически успешными, обретали неоспоримое военное превосходство и начинали диктовать свою волю менее развитым странам.

В XX в. две мировые войны и немыслимое ускорение научно-технического прогресса еще более усилили это влияние. А с приближением третьей декады XXI в. лавинообразный рост возможностей машинного обучения заставил говорить о спектре IT технологий, объединенных метафорическим обозначением «Искусственный интеллект» (ИИ) как о ключевом факторе экономического, геополитического и военного могущества ближайших десятилетий.

2018 год стал переломным. Ранее основное внимание СМИ, общества и политиков при обсуждении будущих угроз и вызовов развития ИИ было направлено:

- на грядущее вытеснение людей из многих профессий когнитивными агентами ИИ (роботами и программами),
- на правовые и этические проблемы автономных ИИ-устройств (например, беспилотных авто),
- на проблемы кибербезопасности,
- на пугающе манящую перспективу то ли далекого, то ли не очень «бунта машин».

Однако в 2018 г. всё вдруг начало меняться. Перечисленные выше угрозы и вызовы не то чтобы исчезли, но потускнели на фоне происходящего осознания политиками и военными двух **глобальных трансформационных трендов**, которые можно условно назвать

- ИИ-национализм
- ИИ-национализация

Появление этих трендов выразилось в том, что многие развитые страны при создании своих национальных ИИ-стратегий синхронно стали менять отношение к двум базовым принципам, ранее казавшимся незыблемыми:

- вместо всестороннего международного сотрудничества, глобального разделения труда, внедрения открытых платформ и взаимного перетекания

талантов, делается ставка на **ИИ-национализм**, декларирующий главной целью национальных ИИ-стратегий приоритет экономических и военных интересов своей страны;

- вместо распространенного в развитых странах разделения государства и бизнеса, взят курс на **ИИ-национализацию** – интеграцию ресурсов государства и частных компаний, выравнивание скоростей внедрения ИИ-инноваций и рефокусировка стратегических целей на получение государством экономических, геополитических и военных преимуществ на международной арене.

Усиления этих трендов способствует смещению государственных приоритетов в развитых странах **от экономики к геополитике**. И если это продолжится, то в самом недалеком будущем мир довольно сильно изменится.

Краткий техно-геополитический прогноз

Во-первых, **поменяются геополитические и военные доктрины** большинства развитых стран. Произойдет кардинальная трансформация многих экономико-политических процессов, объединяемых понятием **глобализация**. Не исключено, что глобализация, резко ускорившаяся вслед за взрывной динамикой развития технологий после Второй мировой войны, ИИ-технологиями будет вообще остановлена. И тогда на смену глобализации придет **глобальная балканизация**.

Во-вторых, не менее сильно скажется **изменение модели сосуществования государства и бизнеса**. Интеграция целей и ресурсов государства и частного бизнеса для достижения военного превосходства, скорее всего, приведет ко всемирному торжеству **авторитарно-демократической модели** интеграции интересов частного бизнеса и государства, уже взятой на вооружение (в прямом и переносном смысле) Китаем.

В результате главными факторами международных отношений станут:

ИИ-неоколониализм во взаимоотношениях стран-лидеров и стран-аутсайдеров в ИИ-технологиях;

Гонка ИИ-вооружений между странами-лидерами, определяющая и направляющая дальнейшее развитие ИИ-технологий.

У такой гонки возможны 2 варианта окончания:

- Вариант «Большая война»**, в результате которой следующая война будет вестись уже камнями и палками.
- Вариант «AI Сингулярность»** (или «приход лесника»), при котором все ускоряемое гонкой вооружений развитие ИИ-технологий породит **«сильный» ИИ**, а тот поступит как лесник в анекдоте, устранивший обе соперничающие стороны за ненадобностью.

Первый вариант видится куда более вероятным. Однако и второй вариант не исключен.

Важно отметить, что с какого-то момента тренды ИИ-национализма и ИИ-национализации станут самораскручивающимися. Другими словами, их поддержание и укрепление, как при всякой гонке вооружений, уже не будет зависеть от степени прогресса ИИ. И даже если прогресс ИИ окажется куда скромнее, чем предполагалось, потребуется немало времени и усилий для преодоления гигантской инерции обоих трендов.

2018 – год великого AI перелома

После окончания холодной войны США располагали практически непревзойденным статусом сверхдержавы. Важнейшим фактором в этом было ее непревзойденное военно-технологическое превосходство. Тем не менее, технологии, которые лежали в основе обороны американских военных в прошлом, такие как высокоточное оружие, распространились по миру вследствие глобализации и передачи технологий. В результате страны – конкуренты США развили собственные возможности, которые все больше бросают вызов военному превосходству США.

Министерство обороны США, чтобы сохранить и расширить свое военное преимущество в будущем, поставило на ИИ технологии, потенциальное применение которых в военной области весьма разнообразно: от повышения эффективности логистических систем к более чувствительным задачам, таким как автоматическое управление и контроль в современных системах вооружения. Стратегия национальной обороны 2018 г. предполагает, что ИИ, скорее всего, изменит характер войны. И потому, по словам⁵⁰ замминистра обороны Патрика Шэнахэна, Соединенные Штаты «должны стремиться к применению ИИ с дерзновением и готовностью».

Есть основания предположить, что прошлогодний призыв Эрика Шмидта⁵¹, бывшего руководителя Google и Alphabet, к осознанию того факта, что «Момент спутника» в области ИИ уже наступил», наконец дошел до правительства США.

Министр обороны **Джеймс Мэттис** призвал президента Трампа создать национальную стратегию развития технологий искусственного интеллекта не только для правительства США, но и для «всей страны». Письмо Мэттиса президенту включало предложение создать президентскую комиссию, способную «вдохновить усилия всей страны на то, чтобы США стали мировым лидером не только в вопросах обороны, но и в глубокой трансформации всех аспектов жизни человечества».

⁵⁰ Memorandum: Establishment of the Joint Artificial Intelligence Center (JAIC). 2018.
URL: https://admin.govexec.com/media/establishment_of_the_joint_artificial_intelligence_center_osd008412-18_r....pdf

⁵¹ Our Artificial Intelligence 'Sputnik Moment' Is Now: Eric Schmidt & Bob Work // Breaking Defense. 2017.
URL: <https://breakingdefense.com/2017/11/our-artificial-intelligence-sputnik-moment-is-now-eric-schmidt-bob-work/>

В ответ на все эти призывы,

- в «Summary of the 2018 White House Summit on Artificial Intelligence for American Industry» заявлено что администрация Трампа «уделяет **приоритетное внимание финансированию** фундаментальной исследовательской и вычислительной инфраструктуры ИИ, машинного обучения и автономных систем»;
- а 31 июля 2018 г. Белый дом выпустил меморандум для глав департаментов и агентств №М-18-22 о бюджетных приоритетах R&D на 2019–2020⁵², в котором назвал ИИ **первым среди трех высших национальных технологических приоритетов** (2-й – квантовая информатика, а 3-й – суперкомпьютинг) и обязал Административно-бюджетное управление (OMB) совместно с Управлением науки и технологической политики (OSTP) обеспечить **высший бюджетный приоритет** этих направлений для всех федеральных агентств в 2019–2020.

Так уж получилось, что ощущение «момента Спутника» испытало в 2018 году не только правительство США.

Число стран, продекларировавших национальную стратегию развития ИИ, выросло по сравнению с 2017 г. сразу более чем в три раза.

В 2017 г. энтузиасты ИИ (Япония, Канада и Сингапур) открыли марафон подписания национальных ИИ-стратегий. Следом за ними, сразу орудиями главного финансового калибра, вступил один из двух претендентов на мировое лидерство в области ИИ – Китай.

Ну а в 2018 г., не дожидаясь конца года, гораздо проще сказать, кто еще не опубликовал свою национальную стратегию в области ИИ, поскольку Британия, Германия, Франция и еще дюжина стран свои стратегии уже представили.

Краткое описание деталей и специфики каждой из нацстратегий можно прочесть в отчете, подготовленном Politics + AI под названием «An Overview of National AI Strategies»⁵³. Как и следовало ожидать, разные страны обозначили различные цели и подходы в своих национальных ИИ-стратегиях.

Отличались также объем, уровень проработки и детализации документов, подготовленных по поручению правительств различных стран.

Самой лаконичной (25 стр.) и конкретной выглядит национальная стратегия Японии⁵⁴, включающая базовое описание 3-х фаз создания к 2030 г. национальной системы ИИ as a service.

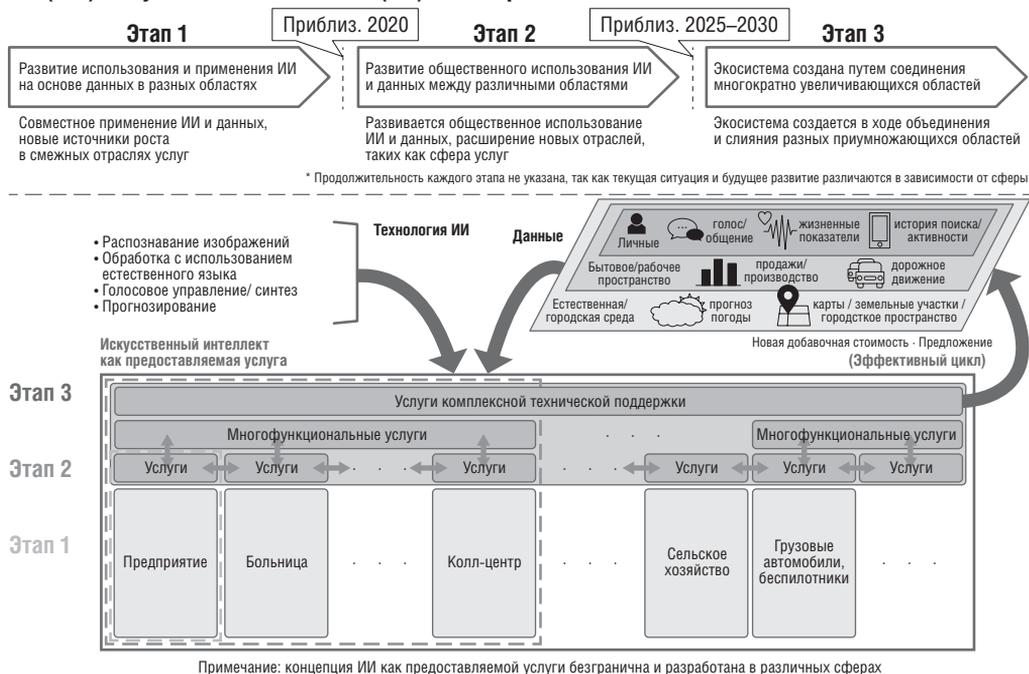
⁵² Memorandum: FY 2020 Administration Research and Development Budget Priorities // The White House. 2018.
URL: <https://www.whitehouse.gov/wp-content/uploads/2018/07/M-18-22.pdf>

⁵³ Politics + AI. An Overview of National AI Strategies.
URL: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>

⁵⁴ Strategic Council for AI Technology. Artificial Intelligence Technology Strategy. 2017.
URL: <http://www.nedo.go.jp/content/100865202.pdf>

(1-1) Искусственный интеллект (ИИ) Этапы развития

[Приложение 1]



Примечание: концепция ИИ как предоставляемой услуги безгранична и разработана в различных сферах

Три фазы создания к 2030 г. национальной системы ИИ as a service в Японии

К этой базовой схеме приложен набор из дорожных карт по трём приоритетным направлениям: продуктивность, медицина и здоровье, мобильность.

В дополнение разработан план интеграции технологий ИИ с технологиями основных отраслей экономики по трем приоритетным направлениям (продуктивность, медицина и здоровье, мобильность). Три фазы этой интеграции показывают уровень планируемого технологического прогресса и социальных изменений.

Самым детальным и тщательно проработанным выглядит подход Великобритании, подготовившей следующий состав документов:

- 180-страничный отчет «AI in the UK: ready, willing and able?»⁵⁵, опубликованный администрацией Палаты Лордов в апреле 2018 г. и во многом отталкивающийся в своих идеях от 77-страничного отчета профессора Computer Science в университете Саусхэмптона Dame Wendy Hall и CEO компании BenevolentAI доктора Jerome Pesenti «Growing the Artificial Intelligence Industry in the UK»⁵⁶.

⁵⁵ Strategic Council for AI Technology. AI in the UK: ready, willing and able? // UK Parliament. 2018. URL: <https://publications.parliament.uk/pa/d201719/dselect/ldai/100/100.pdf>

⁵⁶ W. Hall, J. Pesenti. Growing the Artificial Intelligence Industry in the UK URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

- Два тома письменных и устных свидетельств специалистов и экспертов в подтверждение положений отчета (1-й том содержит 223 письменных свидетельства, изложенных на 1 581 стр.⁵⁷, 2-й содержит 57 устных свидетельств, полученных в ходе 22 сессий и изложенных на 423 стр.⁵⁸)
- 40-страничный ответ Правительства на отчет и свидетельства, представленные Палатой Лордов⁵⁹ Парламенту Великобритании, Secretary of State for Business, Energy and Industrial Strategy by Command of Her Majesty

Что же объединяет все 20+ национальных видений будущего ИИ, на сегодняшний день уже сформулированных государствами в самых разнообразных формах: от планов и дорожных карт (как у Японии) до «судебных решений», вменяющих своему правительству конкретные поручения (как у Великобритании)?

Внимательно изучив весь этот корпус документов, опубликованных по состоянию на конец августа 2018 г, автор выделяет два ключевых тренда, явственно проявляющихся во всех «нацстратегиях»:

1. Курс на ИИ-национализм
2. Курс на ИИ-национализацию

Осталось рассказать, что под этим понимается.

ИИ-национализм

Первый источник ИИ-национализма – сверхвысокие ожидания

Термин ИИ-национализм стал популярным после июньской публикации Яном Хогартом одноименного эссе⁶⁰. В эссе «AI Nationalism» это понятие определяется так: «Новый вид геополитики, набирающий силу в развитых странах вследствие наблюдаемого в последние годы бурного прогресса ИИ».

ИИ порождает **новый вид нестабильности** на национальном и международном уровнях, вынуждая правительства развитых стран действовать, чтобы не оказаться среди проигравших в этой новой для мира конкуренции за обладание ИИ-превосходством.

⁵⁷ Strategic Council for AI Technology. Collated Written Evidence Volume // UK Parliament. 2018.

URL: <https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Written-Evidence-Volume.pdf>

⁵⁸ Strategic Council for AI Technology. Collated Oral Evidence Volume // UK Parliament. 2018.

URL: <https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Oral-Evidence-Volume.pdf>

⁵⁹ The Secretary of State for Business, Energy and Industrial Strategy by Command of Her Majesty. Government response to House of Lords Artificial Intelligence Select Committee's Report on AI in the UK: Ready, Willing and Able? // UK Parliament. 2018.

URL: <https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Government-Response2.pdf>

⁶⁰ I.Hogarth, AI Nationalism // Ian Hogarth. 2018.

URL: <https://www.ianhogarth.com/blog/2018/6/13/ai-nationalism>

Эта конкуренция уникальна и не похожа ни на что в прошлом, включая конкуренцию за обладание ядерной бомбой и межконтинентальными ракетами.

Уникальность ИИ определяется тремя факторами – экономическим и двумя военными.

1. Универсальность инструментария ИИ, как средства повышения эффективности практически во всех постиндустриальных отраслях и видах деятельности (наиболее близкий пример такой универсальности – внедрение электричества повсюду);
2. Базирующееся на прошлом военном опыте предположение о высочайшем потенциале ИИ для революционных прорывов,
 - как в создании абсолютно новых военно-технологических комплексных классов решений (типа комплекса способов снижения заметности боевых машин в радиолокационном, инфракрасном и других областях спектра обнаружения – стелс-технологий),
 - так и в построении на основе ИИ кардинально более совершенных систем военной осведомленности и управления военной логистикой и непосредственно боем на всем театре военных действий (включая смену привычных парадигм для тех или иных видов войск, как, например, произошло с превращением авианосца из транспорта для самолетов разведки и наведения палубной артиллерии в плавающий аэродром, супер-эффективно решающий самостоятельные военные задачи).
3. Предполагаемая возможность (еще не доказанная, но принимаемая многими военными всерьез) решения проблемы ядерного сдерживания в свою пользу (по старой ковбойской мудрости, что главное – умение выстрелить быстрее противника).

Названные факторы в значительной мере гипотетические. Они нисколько не отражают реальные возможности уже достигнутого уровня ИИ-технологий, сколько являются их экстраполяцией в ближайшее будущее, при условии сохранения текущих темпов развития ИИ-технологий.

Иными словами, все три фактора, превращающие ИИ в фактор гипотетического обретения превосходства на международной арене, являются пока что лишь ожиданиями военных и политиков.

Но это вовсе не мешает им утверждать, что мир находится на пороге **новой сингулярности – военной**. То, насколько серьезно к этому относятся лидеры гонки в области ИИ – США и Китай, иллюстрирует отчет CNAS «Battlefield Singularity»⁶¹. Поражает не только логика и аналитика отчета, но и гигантский корпус из 334 американских и китайских источников, на которые он ссылается.

⁶¹ E.B.Kania. Battlefield Singularity // CNAS. 2017.

URL: <https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power>

Второй источник ИИ-национализма – технологическая запутанность

Беспрецедентно высокая важность ИИ может сделать политику в этой области ключевым элементом государственной политики. Однако, на данном пути возникает крайне неприятная проблема т.н. **технологической запутанности** (Technological entanglement), присущей ИИ-технологиям, как никаким другим.⁶²

Под технологической запутанностью понимается возникающее и усиливающееся в процессе глобализации тесное и многогранное сплетение интересов и ресурсов международных компаний в разработке элементов технологий, имеющих двойное назначение (мирное и военное).

В результате технологической запутанности понятие технологического суверенитета практически исчезает. Даже несомненный технологический лидер в области ИИ – США в условиях технологической запутанности оказывается в крайне затруднительном положении.

Как могут США сохранять лидерство в ИИ, предотвращая утечку своих технологий зарубежным конкурентам, если:

- технологическим локомотивом США является частный бизнес;
- международные гиганты частного бизнеса, такие как Alphabet, в рамках глобализации располагают свои исследовательские лаборатории по ИИ в главном конкуренте в этой области – Китае и нанимают на работу тысячи китайских специалистов;
- китайские инвесторы, среди которых крупнейшие ИИ-компании Китая, владеют существенными долями во многих самых перспективных в области ИИ компаниях Кремниевой долины;
- китайские студенты и аспиранты тысячами учатся в лучших университетах США.

Последним и решающим витком технологической запутанности стало провозглашение Китаем национальной **стратегии «военно-гражданского сплава»** интересов и ресурсов частного бизнеса и государственных структур Китая – National Strategy for military-civil fusion⁶³ (军民融合).

Сложившуюся в результате технологической запутанности ситуацию резюмировал⁶⁴ президент Information Technology and Innovation Foundation **Роберт Аткинсон** в статье от 26 июля 2018 г. в National Review.

⁶² Technological entanglement. Cooperation, competition and the dual-use dilemma in artificial intelligence. The Australian Strategic Policy Institute // ASPI. 2018.

URL: <https://www.aspi.org.au/report/technological-entanglement>

⁶³ G.Levesque, M.Stokes. Blurred Lines: Military-Civil Fusion and the “Going Out” of China’s Defense Industry // Pointe Bello. 2016.

URL: https://static1.squarespace.com/static/569925bfe0327c837e2e9a94/t/593dad0320099e64e1ca92a5/1497214574912/062017_Pointe+Bello_Military+Civil+Fusion+Report.pdf

⁶⁴ R.D. Atkinson. Who Lost China? // National Review. 2018.

URL: <https://www.nationalreview.com/magazine/2018/08/13/us-china-relations-who-lost-them/>

- Китай развернул обширную практику «инновационных меркантилистов», которая направлена на несправедливое получение преимуществ китайскими производителями. К ним относятся:
 - требование о том, чтобы иностранные компании для доступа к китайскому рынку передавали свои технологии китайским фирмам;
 - кража иностранной интеллектуальной собственности;
 - манипулирование технологическими стандартами;
 - массовые государственные субсидии;
 - поддержанные правительством приобретения иностранных предприятий.
- У США сейчас только одна важная цель в области торгово-технологической политики: положить конец инновационно-меркантилистской политике Китая, которая угрожает национальной и экономической безопасности США.

Таким образом, крайне высокие ожидания небывалой отдачи от ИИ, способной обеспечить превосходство страны на международной арене, наряду с все усугубляющейся технологической запутанностью, стали основными источниками подпитки техно-националистической повестки ряда стран, число которых увеличивается.

Потенциальные последствия этого – всевозможные протекционистские действия государств для поддержки своих «национальных чемпионов» в области ИИ, ограничения (а в развитии и возможный запрет) на передачу патентов, открытую публикацию исследований и экспорт ИИ технологий, а также ограничения (а в развитии и возможный запрет) на проведение M&A сделок, свободный переток инвестиций и конечно же талантов, как при обучении, так и в профессиональной деятельности.

Кто в авангарде ИИ национализма

По состоянию на конец августа 2018 г., о своем намерении следовать курсу ИИ-национализма, в той или иной форме, объявили: Китай, США, Индия, Великобритания и Франция. Эти же 5 стран объявили о разработке **национальных военных программ** на базе ИИ.

- Китай – в рамках национального Плана 13-й пятилетки⁶⁵, программы «Artificial Intelligence 2.0» Плана Научно-технического и инновационного развития 13-й пятилетки⁶⁶, 3-летнего Action Plan for Promoting Develop-

⁶⁵ Чжунхуа жэньминь гунхэго гоминь цзинци хэ шэжуй фачжэнь ди шисань гз унянь гуйхуа ганьяо (Программа 13-го пятилетнего плана социально-экономического развития КНР) // Сайт Госсовета КНР. 2016. URL: http://www.gov.cn/xinwen/2016-03/17/content_5054992.htm (на китайском)

⁶⁶ Гоуяоань гуаньюй иньфа «шисанью» гоця кэци чуансинь гуйхуа дэ тунчжи (Уведомление Госсовета КНР о государственном плане научно-технических инноваций «13-й пятилетки») // Сайт Госсовета КНР. 2016. URL: http://www.gov.cn/zhengce/content/2016-08/08/content_5098072.htm (на китайском)

ment of a New Generation Artificial Intelligence Industry (2018–2020)⁶⁷ – и все это под зонтиком национальной стратегии «военно-гражданского сплава»⁶⁸.

- США – в рамках Стратегии национальной безопасности⁶⁹, Стратегии национальной обороны⁷⁰ и Меморандума замминистра обороны от 27 июня 2018 г. о создании Объединенного центра искусственного интеллекта Министерства обороны (JAIC)⁷¹.
- Индия – в рамках Национальной стратегии развития ИИ, подготовленной для правительства Индии аналитическим центром NITI Aayog⁷² и «Дорожной карты развития ИИ для обеспечения обороны и безопасности Индии»⁷³, подготовленной «Tata Sons» по поручению Министерства обороны.
- Франция – в рамках⁷⁴ (1) объявленного министром французских вооруженных сил Флоренсом Парли резкого увеличения расходов страны на ИИ для развития будущих систем вооружений (1,83 млрд долл.); (2) парламентского «отчета Виллани»⁷⁵, фиксировавшего, что «ИИ теперь является центральным политическим принципом обеспечения безопасности», «поддержки господства над нашими потенциальными противниками» и «поддержки нашей позиции по отношению к союзникам»; (3) дорожной карты изучения возможностей ИИ для вооружений и ее первого этапа – проекта Man-Machine Teaming (ММТ)⁷⁶.
- Великобритания надеется в области ИИ стать частью военных программ США и в рамках этой кооперации решением Министра обороны в мае 2018 г. открыла AI Lab⁷⁷ – единый исследовательско-внедренческий центр ИИ при Defence Science and Technology Laboratory (Dstl).

⁶⁷ Three Year Action Plan Focuses On Next generation Artificial Intelligence // USITO. 2018.

URL: <http://www.usito.org/news/three-year-action-plan-focuses-next-generation-artificial-intelligence>

⁶⁸ G. Levesque, M. Stokes. Blurred Lines: Military-Civil Fusion and the “Going Out” of China’s Defense Industry // Pointe Bello. 2016.

URL: https://static1.squarespace.com/static/569925bfe0327c837e2e9a94/t/593dad0320099e64e1ca92a5/1497214574912/062017_Pointe+Bello_Military+Civil+Fusion+Report.pdf

⁶⁹ National Security Strategy of the United States of America // The White House. 2017.

URL: <https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf>

⁷⁰ A Summary of the 2018 National Defense Strategy of The United States of America. US Department of Defense. 2018.

URL: <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>

⁷¹ Memorandum: Establishment of the Joint Artificial Intelligence Center (JAIC). Deputy Secretary of Defense. 2018.

URL: https://admin.govexec.com/media/establishment_of_the_joint_artificial_intelligence_center_osd008412-18_r....pdf

⁷² Strategy for Artificial Intelligence // The National Institution for Transforming India (NITI Aayog). 2018.

URL: http://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

⁷³ AI Task Force Hands Over Final Report to RM // PIB. 2018.

URL: <http://pib.nic.in/newsite/PrintRelease.aspx?relid=180322>

⁷⁴ J. Turner. Intelligent design: inside France’s €1.5bn AI strategy // Army Technology. 2018.

URL: <https://www.army-technology.com/features/intelligent-design-inside-frances-e1-5bn-ai-strategy/>

⁷⁵ C. Villani. For a Meaningful Artificial Intelligence // AI For Humanity. 2018.

URL: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

⁷⁶ Man-Machine Teaming (ММТ). 2018. URL: <https://man-machine-teaming.com/>

⁷⁷ UK launches new artificial intelligence lab at Dstl // Army Technology. 2018.

URL: <https://www.army-technology.com/news/uk-launches-new-artificial-intelligence-lab-dstl/>

В качестве главных векторов курса ИИ-национализма все названные страны, так или иначе, продекларировали следующие **три политики**.

1. Политика, гарантирующая сохранение экономических и военных преимуществ первопроходцев (first mover advantages) ИИ, которыми эти страны себя, не без оснований, считают.
2. Политика препятствования копированию новых ИИ-технологий, и, в первую очередь, тех, что легко воспроизводимы любыми странами, обладающими аналогичным технологическим уровнем.
3. Политика ослабления стимулов международной торговли, автоматически ведущих к распространению ИИ-технологий по миру.

Особо отмечается «готовность действовать» всем арсеналом средств государственного регулирования, быстро и решительно разрабатывая новые правила и жестко пресекая любые попытки подрыва другими странами покунуться на национальный технологический суверенитет в области ИИ.

ИИ-национализация

Страх отстать в мировой гонке за превосходство в сфере ИИ породил тренд ИИ-национализма. Однако этим дело не ограничивается.

Когда выяснилось, что проводимая Китаем политика интеграции целей и ресурсов частного бизнеса и государства (Military-Civil Fusion) позволила Китаю в считанные годы практически догнать ранее недостижимого мирового лидера в области ИИ США, стало ясно, что и остальным странам ничего не остается, как пойти тем же путем.

Вместо привычного для развитых стран разделения государства и бизнеса, заговорили о преимуществах **ИИ-национализации** – интеграции ресурсов государства и частных компаний, выравнивания скоростей внедрения инноваций в области ИИ и рефокусировки их стратегических целей на получение государством экономических, политических и военных преимуществ на международной арене.

Повестка ИИ-национализации разделяется:

- на повестку лидеров гонки ИИ,
- и повестку аутсайдеров гонки ИИ.

Повестка лидеров гонки ИИ – США и Китая

Некоторые эксперты приводят аргументы, что Китай пока что чуть ли не вдвое отстает в области ИИ от США⁷⁸.

⁷⁸ Deciphering China's AI Dream. The context, components, capabilities, and consequences of China's strategy to lead the world in AI // Future of Humanity Institute, University of Oxford. 2018.
URL: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

Главный фактор в сфере ИИ	Прогу показатель(и)	Китай	США
Техническое оснащение	Доля мирового рынка производства полупроводников (2015)	4%	50%
	Финансирование производителей чипов для Программируемых Пользователем Вентильных Матриц (ППВМ) (2017)	\$34.4 млн. (7.6% мирового рынка)	\$192.5 млн. (42.4% мирового рынка)
Данные	Пользователи мобильных телефонов (2016)	1.4 млрд. (20.0% мирового рынка)	416.7 млн. (5.5% мирового рынка)
Исследование и алгоритмы	Количество экспертов в области ИИ	39 200 (13.1% мирового рынка)	78 700 (26.2% мирового рынка)
	Процентная доля проведенных конференций Ассоциацией по развитию ИИ (2015)	20.5% мирового рынка	48.4% мирового рынка
Коммерческий сектор ИИ	Доля компаний (%) в сфере ИИ	23%	42%
	Общий объем инвестиций в компании в сфере ИИ (2012-2016)	\$2.6 млрд. (6.6% мирового рынка)	\$17.2 млрд. (43.4% мирового рынка)
	Общий объем капиталовложений в ИИ стартапы (2017)	48% мирового рынка	38% мирового рынка
Индекс потенциальных возможностей ИИ	Среднее значение четырех средних прогу показателей	$(5.8+20+16.8+25.9)/4 = 17$	$(46.2+5.5+37.3+41.1)/4 = 33$

Расчет Индекса ИИ-потенциала по методике из исследования⁷⁹ показывает почти двойное отставание Китая от США.

Однако, комплексный анализ гражданских и военных прорывных технологий Китая⁸⁰, результаты слушаний в комиссии United States-China Economic and Security Review Commission⁸¹ и её итоговый отчет Конгрессу США⁸², а также новейший отчет по углубленному анализу вооружений Китая⁸³ зафиксировали примерный паритет в развитии ИИ-технологий США и Китая.

⁷⁹ Deciphering China's AI Dream. The context, components, capabilities, and consequences of China's strategy to lead the world in AI // Future of Humanity Institute, University of Oxford. 2018.

URL: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

⁸⁰ Planning for Innovation: Understanding China's Plans for Technological, Energy, Industrial, and Defense Development // US-China Economic and Security Review Commission. 2016.

URL: [https://www.uscc.gov/sites/default/files/Research/Planning for Innovation-Understanding China's Plans for Tech Energy Industrial and Defense Development072816.pdf](https://www.uscc.gov/sites/default/files/Research/Planning%20for%20Innovation-Understanding%20China%27s%20Plans%20for%20Tech%20Energy%20Industrial%20and%20Defense%20Development072816.pdf)

⁸¹ China's Advanced Weapons. Hearing Before the U.S.-China Economic and Security Review Commission // US-China Economic and Security Review Commission. 2016.

URL: [https://www.uscc.gov/sites/default/files/transcripts/China's Advanced Weapons.pdf](https://www.uscc.gov/sites/default/files/transcripts/China%27s%20Advanced%20Weapons.pdf)

⁸² 2017 Report to Congress of the U.S.-China Economic and Security Review Commission» Executive Summary and Recommendations // US-China Economic and Security Review Commission. 2017.

URL: [https://www.uscc.gov/sites/default/files/annual_reports/2017 Executive Summary and Recommendations_1.pdf](https://www.uscc.gov/sites/default/files/annual_reports/2017%20Executive%20Summary%20and%20Recommendations_1.pdf)

⁸³ China's Advanced Weapons Systems. Jane's by IHS Markit for U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2018.

URL: [https://www.uscc.gov/sites/default/files/Research/Jane's by IHS Markit_China's Advanced Weapons Systems.pdf](https://www.uscc.gov/sites/default/files/Research/Jane%27s%20by%20IHS%20Markit_China%27s%20Advanced%20Weapons%20Systems.pdf)

ТЕКУЩЕЕ СОСТОЯНИЕ ТЕХНОЛОГИЧЕСКОЙ КОНКУРЕНЦИИ МЕЖДУ США И КИТАЕМ В ДЕВЯТИ ОТРАСЛЯХ

Лидирует США	Примерный паритет	Лидирует Китай
<ul style="list-style-type: none">• Биотехнологии• Нанотехнологии• Облачная обработка данных• Коллаборативная робототехника	<ul style="list-style-type: none">• Искусственный интеллект• Квантовая информатика• Высокопроизводительные вычислительные системы	<ul style="list-style-type: none">• Эксафлопсные вычисления• Коммерческие дроны

Примечание: факторы, определяющие лидерство по технологическим показателям, включают в себя: число компаний; доля на мировом рынке; уровень финансирования научно-исследовательских и конструкторских разработок; количество заявок на патенты; количество публикаций в престижных журналах; число упоминаний для каждой публикации. Статус лидерства по технологическим показателям может быть изменен в силу преобразований в государственной политике или новых прорывов в научно-исследовательских и конструкторских разработках.

Источник: Данный рейтинг основывается на показателях, полученных в марте 2017 года, в ходе слушаний Комиссии по следующим вопросам: достижения Китая в сфере технологий следующего поколения и технологий двойного назначения; научных исследованиях; обсуждениях с представителями власти, академического сообщества, отраслевыми экспертами, а также изучении материалов из открытых источников и аналитической работе.

При рассмотрении девяти наиболее прорывных классов технологий двойного применения эксперты Конгресса США отнесли ИИ к группе из 3-х классов технологий, по которым между США и Китаем примерный паритет⁸⁴.

Выводы этих экспертов⁸⁵ учитывают, в отличие от весьма качественного, но гражданского отчета⁸⁶, три особых фактора:

1. В долгосрочной стратегии развития вооружений Китая автономные беспилотные системы и ИИ обладают высшим приоритетом⁸⁷ развития.
2. Особенности китайской дорожной карты развития ИИ, отдающей высокий приоритет созданию «интеллектуального (智能化) оружия»⁸⁸, включив его в перечень 4-х критических технологических областей «strategic frontier» (战略前沿), в которых ставится цель превзойти американских военных, используя стратегию «leapfrog development» (跨越发) – перепрыгивания через несколько стадий развития.
3. Намерение Китая использовать ИИ для ускорения совершенствования т.н. «козырного оружия»⁸⁹ – «accelerate the development of shashoujian

⁸⁴ China's Advanced Weapons. Hearing Before the U.S.-China Economic and Security Review Commission // US-China Economic and Security Review Commission. 2016.

URL: <https://www.uscc.gov/sites/default/files/transcripts/China%27s%20Advanced%20Weapons.pdf>

⁸⁵ Ibid.

⁸⁶ Deciphering China's AI Dream. The context, components, capabilities, and consequences of China's strategy to lead the world in AI // Future of Humanity Institute, University of Oxford. 2018.

URL: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

⁸⁷ Press Release. U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2018.

URL: https://www.uscc.gov/sites/default/files/China%27s%20Advanced%20Weapons_PR.pdf

⁸⁸ E. B. Kania. Quest for an IA Revolution in Warfare // The Strategy Bridge. 2017.

URL: <https://thestrategybridge.org/the-bridge/2017/6/8/-chinas-quest-for-an-ai-revolution-in-warfare>

⁸⁹ E. Kania. China's trump Card // Prospect Magazine. 2017.

URL: <https://prospectmagazine.co.uk/other/chinas-trump-card>

(杀手锏) armaments», по иронии судьбы, переводимого на английский как Trump Card.

Но речь вовсе не о президенте Трампе, а о т.н. «шатсёшианском вооружении». Термин шатсёшиан (杀手锏) может быть переведен по-разному: в переводе на английский – это «козырь» или «булава убийцы», а на русский – «киллер». Термин ссылается на китайскую легенду, в которой шатсёшиан использовался для неожиданного выведения из строя более сильного врага с помощью хитрого трюка (типа пращи, которой Давид поразил Голиафа). В свое время председатель Компартии Китая Цзян Цзэминь так использовал термин шатсёшиан: «За что враг больше всего боится, то мы и должны атаковать». С тех пор военная доктрина Китая установила приоритетное развитие «козырного оружия» – шатсёшиан, хитроумно бьющего в самые болезненные уязвимости противника. Ну а ИИ, как оказалось, очень при таком подходе полезен.

Военные США прекрасно осведомлены о ставке Китая на «козырное оружие». Как сказал Роберт Ворк⁹⁰, бывший замминистра обороны США с 2014 по 2017 гг., «китайская стратегия победы – это битва за разрушение систем (systems destruction warfare), поскольку в ней основное внимание уделяется электронно-парализующему воздействию на командование и управление войсками противника, а не физическое уничтожение танков, кораблей и самолетов».

Наиболее простым и очевидным примером может быть использование коллективного интеллекта роя микродронов⁹¹ для вывода из строя авианосцев⁹².

Другой пример – интеллектуализация ракетного оружия. Здесь перед ИИ ставится цель «качественного расширения возможностей ракет по восприятию внешней информации (感知) и принятию на этой основе решений (决策) в целях итогового выполнения (执行) миссий, «осознавая» (认知) происходящее и проявляя способности к обучению»⁹³.

По большому счету, как сказал⁹⁴ недавно один из высоких чинов Пентагона: «Первая нация, развернувшая на поле боя электромагнитное импульсное оружие, способное отключить вражеские системы управления, изменила бы лицо войны. И далеко не очевидно, что в этой гонке победят Сое-

⁹⁰ S. J. Freedberg Jr. US Must Hustle On Hypersonics, EW, AI: VCJCS Selva & Work // Breaking Defense. 2018. URL: <https://breakingdefense.com/2018/06/us-must-hustle-on-hypersonics-ew-ai-vjcs-selva-work/>

⁹¹ С. Карелов. Сбылось предсказание Станислава Лема о коллективном интеллекте роя роботов // Medium. Малоизвестное интересное на стыке науки, технологий, бизнеса и общества - содержательные рассказы, анализ и аннотации. URL: <https://clck.ru/EWHaQ>

⁹² E.B.Kania. 杀手锏 and 跨越发展: Trump Cards and Leapfrogging. // The Strategy Bridge. 2017. URL: <https://thestrategybridge.org/the-bridge/2017/9/5/-and-trump-cards-and-leapfrogging>

⁹³ Chinese Advances in Unmanned Systems and the Military Applications of Artificial Intelligence – the PLA's Trajectory towards Unmanned, "Intelligentized" Warfare. Testimony before the U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2017. URL: https://www.uscc.gov/sites/default/files/Kania_Testimony.pdf

⁹⁴ P. Apps. Commentary: Western armies are losing their high-tech edge // Reuters. 2018. URL: <https://www.reuters.com/article/us-apps-military-commentary/commentary-western-armies-are-losing-their-high-tech-edge-idUSKBN1JV2LN>

диненные Штаты». И речь здесь идет ни о чем ином, как объединении возможностей электромагнитного импульсного оружия и ИИ.

Подробное описание всего ландшафта потенциальных угроз от приоритетного внедрения Китаем ИИ-технологий в передовые системы вооружений см. здесь⁹⁵.

Пока что США и Китай идут ноздря в ноздю в области ИИ. Но немало экспертов считают, что максимум через декаду⁹⁶ китайский авторитаризм одержит убедительную победу⁹⁷ над американской демократией за счет абсолютного доминирования в объеме собираемых данных. Уже сейчас у Китая собрано вчетверо больше данных, чем у США. И этот разрыв растет.

Как сказал один из самых осведомленных в вопросах ИИ-противостояния Китая и США специалистов и мой бывший коллега по «Silicon Graphics» Кай-Фу Ли: «На самом деле в ИИ существует лишь одна фундаментальная инновация – глубокое обучение. И всё, что делается сейчас в области ИИ – лишь подстройка глубокого обучения под нужды конкретных прикладных областей».

А для глубокого обучения нужно как можно больше данных. И тот, кто обладает большими данными, скорее всего, уже выиграл соревнование. Хотя, как говорил Смок Беллью, «никто не проиграл, пока никто не выиграл», и США так просто не сдастся в **гонке ИИ-вооружений**.

Пока же в ближайшем будущем просматривается дуополия. Но два мировых лидера в области ИИ – США и Китай – при этом имеют диаметрально противоположные стратегии:

- Китай – продолжать делать «ЭТО» во что бы ни стало
- США – всячески препятствовать Китаю делать «ЭТО»

Детальное описание, что такое «ЭТО», содержится в отчете⁹⁸. Ну а вкратце речь идет о следующем.

A. Использование протекционистских мер по защите своего рынка от импорта и конкуренции в области ИИ.

B. Спонсируемое государством незаконное или «околозаконное» приобретение интеллектуальной собственности в области ИИ посредством: фи-

⁹⁵ Implications of China's Military Modernization. Testimony before the U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2018.

URL: https://www.uscc.gov/sites/default/files/Nurkin_Written%20Testimony.pdf

⁹⁶ Press Release. U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2018.

URL: https://www.uscc.gov/sites/default/files/China%27s%20Advanced%20Weapons_PR.pdf

⁹⁷ T. C. Linn. Race to develop artificial intelligence is one between Chinese authoritarianism and U.S. democracy // San-Francisco Chronicle. 2018.

URL: <https://www.sfchronicle.com/opinion/openforum/article/Race-to-develop-artificial-intelligence-is-one-13189380.php>

⁹⁸ How China's Economic Aggression Threatens the Technologies and Intellectual Property of the United States and the World // White House Office of Trade and Manufacturing Policy. 2018.

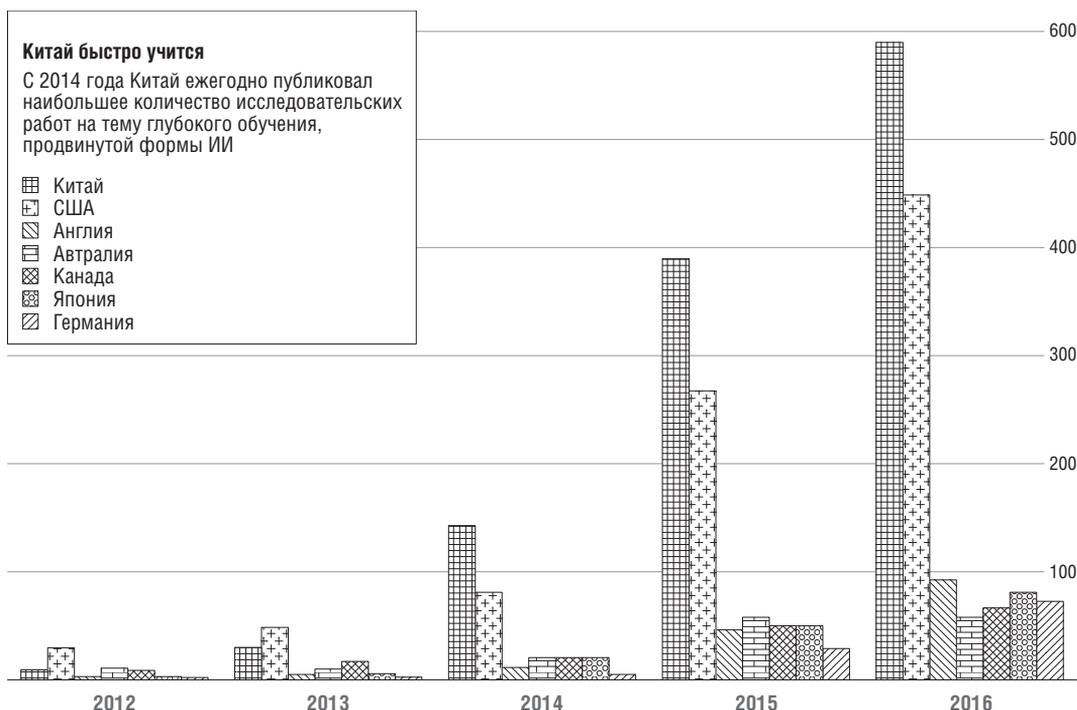
URL: <https://www.whitehouse.gov/wp-content/uploads/2018/06/FINAL-China-Technology-Report-6.18.18-PDF.pdf>

зической кражи, кибер-шпионажа, нарушения экспортного законодательства США, контрафакции и пиратства.

- C. Навязывание или принуждение компаний из США передавать ИИ-технологии китайским компаниям, как правило, в обмен на ограниченный доступ к китайскому рынку.
- D. Выкачивание ценной открытой информации непосредственно из первоисточников научных инновационных идей (университеты, лаборатории, исследовательские центры), а также через скупку талантов из всех смежных областей: бизнес, финансы, экспертиза.
- E. Господдержка инвестиционной деятельности китайских компаний в перспективные ИИ-технологии за пределами Китая.

Повестка аутсайдеров гонки ИИ

Реальное положение таково, что в аутсайдерах AI гонки оказались все, кроме ее лидеров – США и Китая. И хотя технологически развитые страны, типа Франции и Германии или Индии и Южной Кореи, казалось бы, несопоставимы со странами третьего мира по уровню развития ИИ-технологий, – и тем, и другим уготована участь превратиться в **ИИ-колонии** стран лидеров.



*Абсолютное лидерство США и Китая по числу научных публикаций по теме глубокого обучения.
Источник: Digital Transformation Monitor*

В условиях **ИИ-неоколониализма**, вместо золота и серебра страны-колониаторы в XXI веке получают их заменитель – большие данные для обучения своих ИИ-технологий. А странам-колониям остается только уповать на цивилизаторский дух ИИ-неоколониалистов и их финансовую помощь.

Однако, и здесь, как и в гонке ИИ двух лидеров, «никто не проиграл, пока никто не выиграл».

Например, европейские страны пытаются хоть что-то сделать, дабы избежать участи колоний, поставляющих США большие данные.

Признавая уже свершившийся проигрыш США и Китаю в инфраструктуре и масштабах ИИ-программ, европейские специалисты характеризуют свои возможности в области ИИ, как «маленькие капли блеска» и предаются мечтаниям⁹⁹ «как мы берем эти капли блеска и объединяем их вместе».

«Особый аутсайдер». Позиции России в гонке ИИ

Вопрос о потенциале и перспективах России несколько сложнее простого отнесения её к аутсайдерам гонки.

Во-первых, теоретически существует сценарий, в котором аутсайдеры могли бы получить преимущество перед лидерами в области ИИ для военных целей. Этот сценарий остался за границами статьи, поскольку в ней не ставилась задача рассмотреть все сценарии, а лишь наиболее вероятные из них. Но, тем не менее, упомянуть сценарий, благоприятствующий аутсайдерам, необходимо, поскольку он прорабатывается и анализируется весьма авторитетными экспертами. Этот сценарий рассматривался в только что вышедшем осеннем тематическом выпуске журнала «Foreign Policy's Fall 2018»¹⁰⁰, посвященном будущему войны (The Future of War). В статье Michael C. Horowitz «The Algorithms of August»¹⁰¹, в анализе перспектив такого сценария, Россия выделена в качестве «особого аутайдера» – потенциально способного, подобно Китаю, составить конкуренцию в борьбе за лидерство.

Во-вторых, сочетание традиционной российской асимметричности ответов на геополитические вызовы с все еще существующим немалым научно-технологическим наследием СССР могут иметь вполне серьезные последствия на стыках технологий¹⁰² ИИ с новыми классами вооружений (от гиперзвука до систем электронного подавления сигналов). А ведь есть еще стык ИИ с квантовыми вычислениями¹⁰³ и кое-что еще, что может принципиально изменить расклад военных сил ближайшего будущего.

⁹⁹ J. Helmer. The geopolitics of AI raises trust issues for Learning and HR // Lumesse Learning. 2018.
URL: <https://www.lumesselearning.com/ai-trust-issues/>

¹⁰⁰ J. Tepperman. The Future Of War // Foreign Policy. 2018.
URL: <https://foreignpolicy.com/2018/09/12/the-future-of-war-editors-note-fall-2018/>

¹⁰¹ M.C. Horowitz. The Algorithms of August // Foreign Policy. 2018.
URL: <https://foreignpolicy.com/2018/09/12/will-the-united-states-lose-the-artificial-intelligence-arms-race/>

¹⁰² S.J. Freedberg Jr. US Must Hustle On Hypersonics, EW, AI: VCJCS Selva & Work // Breaking Defense. 2018.
URL: <https://breakingdefense.com/2018/06/us-must-hustle-on-hypersonics-ew-ai-vcjcs-selva-work/>

¹⁰³ B. Thomas-Noone. US playing catch-up as technology advantage erodes // ASPI. The Strategist. 2018.
URL: <https://www.aspistrategist.org.au/us-playing-catch-up-as-technology-advantage-erodes/>

Так что с Россией всё не так однозначно. И это заслуживает отдельного экспертного анализа, поскольку даже лучшие из существующих работ на эту тему (например, «Влияние технологических факторов на параметры угроз национальной и международной безопасности, военных конфликтов и стратегической стабильности»¹⁰⁴) названных вопросов касаются лишь вскользь.

К чему ведет ИИ-национализация?

При всем различии повесток гонки ИИ лидеров и аутсайдеров, они предполагают одни и те же шаги, ведущие, по сути, к ИИ-национализации по китайской схеме.

Во-первых, это создание у себя в странах единых гражданско-военных комплексов по разработке ИИ-технологий.

Конечно, гиганты ИИ-бизнеса будут всячески сопротивляться этому. Например, 1 июня 2018 г. Google объявил, что не возобновит свой контракт с поддержкой военной инициативы США под названием Project Maven. Этот проект является первой оперативно развернутой в Ираке системой глубокого обучения ИИ для классификации изображений, передаваемых с военных беспилотников. Решение Google не продлевать контракт с военными последовало после того, как примерно 4 000 из 85 000 сотрудников Google подписали петицию о запрете участия Google в создании «технологий ведения войны».

Ответ последовал незамедлительно.

Уже 6 июня 2018 г. была опубликована статья¹⁰⁵ Грегори Аллена, эксперта «Центра новой американской безопасности», озаглавленная «Исследователи ИИ должны помочь с некоторыми военными разработками». В статье сформулирован новый «этический императив» для коммерческих ИИ-компаний:

«Этическим выбором для ИИ-инженеров Google было бы участие в проектах национальной безопасности, а не их избегание»

О призывах сотрудников Google отказаться от участия в Project Maven, автор пишет:

«Такие призывы создают большую моральную опасность. Внедрение передовых ИИ-технологий в армию столь же неизбежно, как когда-то было с электричеством. Но этот переход чреват этическими и технологическими рисками. Однако, участие в этом процессе талантливых исследователей, и в том числе, таких компаний, как Google, абсолютно необходимо, чтобы помочь военным оставаться на правильной стороне этических норм».

¹⁰⁴ Влияние технологических факторов на параметры угроз национальной и международной безопасности, военных конфликтов и стратегической стабильности. Под ред. Академика РАН А.А.Кокوشина. М.: Издательство Московского университета. 2017.

¹⁰⁵ G.C. Allen. AI researchers should help with some military work // Nature. 2018.
URL: <https://www.nature.com/articles/d41586-018-05364-x>

На помощь Google, по сути, обвиненной в том, что свои этические принципы ее сотрудники ставят выше национальной безопасности, поспешили коллеги по ИИ-индустрии. В июле Илон Маск, соучредители DeepMind, учредитель Skure и несколько известных ИИ-профессионалов призвали¹⁰⁶ коллег по отрасли подписать обещание не создавать «летальное автономное оружие» на основе ИИ-технологий, иначе известное как «роботы-убийцы» (что это, см. здесь¹⁰⁷).

В августе уже 116 известных ИИ-деятели и специалистов подписали письмо¹⁰⁸ с призывом к ООН¹⁰⁹ запретить летальное автономное оружие. В своем заявлении группа заявляет, что развитие такой технологии приведет к «третьей революции в войне», которая могла бы равняться изобретению пороха и ядерного оружия.

И пока ООН молчит, ответ поступил¹¹⁰ от старшего советника НАТО Сандро Гайкена, отметившего, что такие инициативы в высшей степени самодовольны и могут дать авторитарным государствам асимметричное преимущество.

«Если эти наивные хиппи – разработчики из Кремниевой долины не понижают уровня угроз, то ЦРУ должно их заставить понять» – сказал¹¹¹ Гайкен.

«Заставить понять» нужно не только гигантов ИИ-индустрии типа Google, но и множество перспективных стартапов, работающих на переднем крае ИИ-разработок.

«Стартапы должны быть встроены в крупные корпоративные структуры, иметь доступ к требуемым данным и создавать высококачественные ИИ» – считает¹¹² Гайкен.

Эта же логика распространяется и на отдельных специалистов по AI – все они должны работать на решение задач нацбезопасности.

«Существуют огромные различия в том, как талант можно использовать в авторитарных и демократических системах. Военное командование и управляющие экономикой авторитарных стран могут заставить граждан, экспертов и ученых работать на военных. Там, если вам нужны очень хорошие мозги, вы можете просто заставить любых специалистов работать на вас» – объясняет¹¹³ Гайкен.

¹⁰⁶ J. Vincent. Elon Musk, DeepMind founders, and others sign pledge to not develop lethal AI weapon systems // The Verge. 2018.

URL: <https://www.theverge.com/2018/7/18/17582570/ai-weapons-pledge-elon-musk-deepmind-founders-future-of-life-institute>

¹⁰⁷ Initial Reference Architecture of an Intelligent Autonomous Agent for Cyber Defense. ARL US Army Research Laboratory // Cornell University Library. 2018. URL: <https://arxiv.org/ftp/arxiv/papers/1803/1803.10664.pdf>

¹⁰⁸ Killer robots: World's top AI and robotics companies urge United Nations to ban lethal autonomous weapons // Future Of Life. 2017.

URL: <https://futureoflife.org/2017/08/20/killer-robots-worlds-top-ai-robotics-companies-urge-united-nations-ban-lethal-autonomous-weapons/>

¹⁰⁹ J. Vincent. Elon Musk and AI leaders call for a ban on killer robots // The Verge. 2017.

URL: <https://www.theverge.com/2017/8/21/16177828/killer-robots-ban-elon-musk-un-petition>

¹¹⁰ T. Upchurch. How China could beat the West in the deadly race for AI weapons // Wired. 2018.

URL: <https://www.wired.co.uk/article/artificial-intelligence-weapons-warfare-project-maven-google-china>

¹¹¹ Ibid.

¹¹² Ibid.

¹¹³ Ibid.

Трудно представить подобную картину в США:

- ИИ-стартапы, работающие на гигантские ИИ-корпорации;
- Гигантские ИИ-корпорации, работающие на военных;
- ИИ-специалисты, по требованию военных работающие там, где скажут (как «шарашки» при Сталине).

Но ведь так было в СССР. И сейчас это тоже так в Китае...

Смогут ли власти США внедрить подобную практику у себя, зависит, в конце концов, лишь от уровня угроз. Ведь любая демократия кончается там, где объявляется высокий уровень угроз для национальной безопасности.

А многие влиятельные лица в США уже вполне уверены, что ИИ-война с Китаем уже началась. Осталось лишь убедить в этом большинство американцев. Что, после избрания Трампа, уже не кажется такой уж фантастикой.

Дивный новый мир без работы

Маркоткин Н.М.

Что приходит нам в голову, когда мы думаем о грядущем повсеместном внедрении роботов и технологий искусственного интеллекта (ИИ)? Уходящие за горизонт очереди за пособием по безработице? Дроны «Скайнета», бороздящие небо над выжженными трущобами? Или же наоборот, тотальная праздность и коммунизм, обеспеченный трудом механических рабов? Реальность, скорее всего, будет не такой яркой, однако это не повод не задуматься о социальных последствиях происходящих на наших глазах технологических изменений.

Революция на марше

Мир стремительно вступает в стадию IV промышленной революции, среди ключевых компонентов которой роботизация, развитие био- и нанотехнологий, 3D печати, Интернета вещей, генетики и технологий искусственного интеллекта¹¹⁴. Грядущие технологические изменения будут иметь прямые последствия для целого ряда существующих профессий и полностью трансформируют рынок труда как минимум в развитых странах.

Высокий темп происходящих перемен (достаточно сказать, что 10 из наиболее востребованных в 2010 г. специальностей не существовали в 2004 г.¹¹⁵) затрудняет прогнозирование их воздействия на общество. В этой связи оценки экспертов и международных организаций варьируются от оптимистических¹¹⁶, до алармистских¹¹⁷. Однако даже если отбросить радикальные сценарии, можно с уверенностью говорить о том, что в обозримом будущем нас ждет фундаментальная перестройка глобальной экономики, сравнимая с той, что произошла в XVIII–XIX вв., во времена I промышленной революции.

Согласно докладу Всемирного экономического форума (ВЭФ) «Будущее профессий»¹¹⁸, 65% учащихся сегодня в начальных классах будут иметь абсолютно новые, не существующие ныне профессии. К аналогичным выводам пришла компания McKinsey, в докладе¹¹⁹ которой подчеркивает-

¹¹⁴ П. Марш. Новая промышленная революция. Потребители, глобализация и конец массового производства. М.: Издательство Института Гайдара. 2015.

¹¹⁵ Shift Happens // The Guardian. 2010.

URL: <https://www.theguardian.com/science/punctuated-equilibrium/2010/oct/20/2>

¹¹⁶ M. Annunziata and S. Biller. The Future of Work, GE Discussion Paper, General Electric. 2014.

¹¹⁷ When machines can do any job, what will humans do? // EurekAlert. 2016.

URL: https://www.eurekalert.org/pub_releases/2016-02/ru-wmc021016.php

¹¹⁸ The Future of Jobs // World Economic Forum. 2016.

URL: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

¹¹⁹ A Future that Works: Automation, Employment, and Productivity // McKinsey&Company. 2017.

URL: [https://www.mckinsey.com/-/media/McKinsey/Global Themes/Digital Disruption/Harnessing automation for a future that works/MGI-A-future-that-works_Full-report.ashx](https://www.mckinsey.com/-/media/McKinsey/Global%20Themes/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works_Full-report.ashx)

ся, что при текущем уровне развития технологий 60% профессий содержат 30% функционала, который может быть автоматизирован. М. Осборн и К. Фрей из Оксфордского университета дают еще более пессимистичный прогноз. Согласно результатам их исследования¹²⁰, 47% рабочих мест в США рискуют быть автоматизированными в течение 20 лет.

Кого заменят роботы

Какие же профессии находятся в зоне риска? В первую очередь речь идет, конечно же, о неквалифицированном рутинном труде. В исследовании М. Осборна и К. Фрея в число тех, кто наиболее уязвим, попали клерки, наборщики данных, библиотекари, операторы станков, сантехники, специалисты по продажам, наладчики оборудования и другие.

Согласно оценкам ВЭФ¹²¹, с 2015 по 2020 гг. сокращение рабочих мест в наибольшей степени коснется офисных профессий (4,91%) и производственного сектора (1,63%). В пределах 1% также должна сократиться занятость в таких сферах, как дизайн, индустрия развлечений, строительство и продажи. В свою очередь, наиболее значительный прирост рабочих мест предсказуемо ожидается в сфере компьютерных технологий (3,21%), архитектурных и инженерных специальностях (2,71%) и менеджменте (чуть менее 1%).

Под угрозой автоматизации в среднесрочной перспективе также предсказуемо окажутся профессии, связанные с транспортом. Развитие технологии самоуправляемых автомобилей способно радикальным образом изменить рынок как пассажирских, так и в особенности грузовых перевозок. Только в США в индустрии дальних грузовых перевозок заняты 8,7 млн человек¹²². Если же учитывать весь бизнес, завязанный на дальнотойщиках (мотели, придорожные кафе и т. д.), то это число может возрасти до 15 млн, то есть порядка 10% от трудоспособного населения страны. Едва ли не более значимыми могут стать сокращения и в секторе пассажирских перевозок, общественного транспорта. Велика вероятность того, что самоуправляемые технологии в ближайшем будущем будут также внедрены в морских грузовых перевозках¹²³. По мере развития технологий искусственного интеллекта тяжелые времена ожидают¹²⁴ также юристов, учителей, шахтеров, менеджеров среднего звена, журналистов и представителей целого ряда других профессий.

¹²⁰ New study shows nearly half of US jobs at risk of computerisation // University of Oxford. 2013.

URL: <http://www.eng.ox.ac.uk/about/news/new-study-shows-nearly-half-of-us-jobs-at-risk-of-computerisation>

¹²¹ The Future of Jobs // World Economic Forum. 2016.

URL: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

¹²² D. M. Gerstein. Will a Robot Take My Job? // The National Interest. 2018.

URL: <http://nationalinterest.org/feature/will-robot-take-my-job-25444>

¹²³ Rolls-Royce: автономные суда появятся раньше самолетов и автомобилей // Pro-Arctic.ru. 2018.

URL: <http://pro-arctic.ru/08/05/2018/technology/31900>

¹²⁴ A Future that Works: Automation, Employment, and Productivity // McKinsey&Company. 2017.

URL: https://www.mckinsey.com/~media/McKinsey/Global Themes/Digital Disruption/Harnessing automation for a future that works/MGI-A-future-that-works_Full-report.ashx

Можно говорить о том, что основная занятость постепенно переместится из сферы услуг в другие секторы экономики, многие из которых еще только предстоит создать. Впрочем, подобная перспектива не уникальна – она скорее является подтверждением революционного характера происходящих перемен. До I индустриальной революции более 70% населения¹²⁵ было занято в сельском хозяйстве, сегодня же этот показатель в развитых странах колеблется в районе нескольких процентов¹²⁶. Доля занятых в промышленности росла вплоть до середины XX в., однако в результате Цифровой революции на сегодняшний день она снизилась до 24% в ЕС и 19% в США (в России этот показатель составляет 27%)¹²⁷. При этом, хотя рабочих становится меньше, объем производства продолжает стабильно расти¹²⁸. Теперь же, похоже, пришла очередь автоматизации услуг.

Золотой век инженеров и психиатров?

Среди профессий, которые с наименьшей вероятностью пострадают в краткосрочной перспективе, чаще всего называют те, которые связаны с интеллектуальным трудом или подразумевают непосредственный персональный контакт с клиентом. К примеру, в исследовании Оксфордского университета¹²⁹ к наиболее защищенным от автоматизации профессиям отнесены различные виды врачебных и психологических специальностей, а также тренеры, социальные работники, программисты, инженеры, представители высшего управленческого звена и творческих профессий.

Иными словами, лучшим образом подготовлены к новым условиям те, чья работа требует творческого подхода и не сводится к выполнению определенных отработанных комбинаций. В этом плане, говоря, к примеру, об инженерах, важно уточнить, что в «безопасности» находятся в первую очередь инженеры-проектировщики, в то время как инженеры-эксплуатационники, наоборот, находятся в зоне риска.

Автоматизацию креативных профессий сдерживают три ключевых фактора¹³⁰. Для успешного выполнения их задач искусственный интеллект должен обладать восприятием и возможностью оперировать материальными объектами (осознанием), творческим интеллектом и социальным интеллектом. Текущий уровень технологического развития действительно не позволяет решить эти задачи. Однако по мере развития технологий «сильного»

¹²⁵ G. Clark. The Agricultural Revolution and the Industrial Revolution: England, 1500–1912 // University of California. 2002. URL: <http://faculty.econ.ucdavis.edu/faculty/gclark/papers/prod2002.pdf>

¹²⁶ D. M. Gerstein. Will a Robot Take My Job? // The National Interest. 2018.

URL: <http://nationalinterest.org/feature/will-robot-take-my-job-25444>

¹²⁷ Employment in industry (% of total employment) (modeled ILO estimate). International Labour Organization, ILOSTAT database // The World Bank. 2018. URL: <https://data.worldbank.org/indicator/SL.IND.EMPL.ZS>

¹²⁸ D. M. Gerstein. Will a Robot Take My Job? // The National Interest. 2018.

URL: <http://nationalinterest.org/feature/will-robot-take-my-job-25444>

¹²⁹ New study shows nearly half of US jobs at risk of computerisation // University of Oxford. 2013.

URL: <http://www.eng.ox.ac.uk/about/news/new-study-shows-nearly-half-of-us-jobs-at-risk-of-computerisation>

¹³⁰ R.W. Lucky. Are Engineers Designing Their Robotic Replacements? // IEEE Spectrum. 2016.

URL: <https://spectrum.ieee.org/at-work/tech-careers/are-engineers-designing-their-robotic-replacements>

ИИ¹³¹ спектр доступных ему работ будет неуклонно расширяться. Он позволит расширить пределы автоматизации¹³², которые уже были достигнуты при существующих технологиях, и даст компьютерам возможность принимать управленческие решения и даже, возможно, заниматься творческой деятельностью. Поэтому нельзя исключать, что в среднесрочной или долгосрочной перспективе машины смогут успешно заменить не только инженеров и менеджеров, но и писателей и художников. Тем более что прецеденты успешного написания ИИ художественных текстов уже существуют¹³³.

Таким образом, вполне возможно, что большей части трудоспособного населения предстоит в обозримом будущем снова сесть за парту. Однако проблема заключается в том, что никто толком не знает, чему именно им придется учиться. По некоторым оценкам¹³⁴, вплоть до 85% профессий, которые будут востребованы в 2030 г., на сегодняшний день еще не существуют. Текущие же системы образования даже в развитых странах еще не адаптировались¹³⁵ к новой реальности.

Что же будет с родиной и с нами

Сегодня у большинства исследователей не возникает сомнений в том, что развитые страны, так или иначе, успешно приспособятся к переменам (что не исключает вероятности возникновения социального напряжения и роста имущественного неравенства¹³⁶). Новые технологии помогут создать дополнительные рабочие места¹³⁷ взамен исчезнувших, как это произошло в недавнем прошлом¹³⁸ после стремительного развития Интернета. При этом предполагается, что новые специальности будут более творческими и высокооплачиваемыми.

Постепенно на рынке труда установится новое равновесие. Изменится и природа промышленности. Развитие автоматизации и 3D печати позволит создавать локальные эффективные производства¹³⁹, ориентированные под конкретные нужды потребителей. Это будет способствовать возвраще-

¹³¹ Д. Шефтелович. Искусственный интеллект: время слабых // РСМД. 2018.

URL: <http://russiancouncil.ru/analytcs-and-comments/analytcs/iskusstvennyy-intellekt-vremya-slabых/>

¹³² The Future of Jobs // World Economic Forum. 2016.

URL: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

¹³³ N. Shoemaker. Japanese AI Writes a Novel, Nearly Wins Literary Award // BigThink.com. 2016.

URL: <http://bigthink.com/natalie-shoemaker/a-japanese-ai-wrote-a-novel-almost-wins-literary-award>

¹³⁴ Emerging Technologies' Impact on Society & Work in 2030 // Dell Technologies. 2017.

URL: https://www.delltechnologies.com/content/dam/delltechnologies/assets/perspectives/2030/pdf/SR1940_IFTforDellTechnologies_Human-Machine_070517_readerhigh-res.pdf

¹³⁵ D. Shewan. Robots will destroy our jobs – and we're not ready for it // The Guardian. 2017.

URL: <https://www.theguardian.com/technology/2017/jan/11/robots-jobs-employees-artificial-intelligence>

¹³⁶ J. Bessen. How Technology Has Affected Wages for the Last 200 Years // Harvard Business Review. 2015.

URL: <https://hbr.org/2015/04/how-technology-has-affected-wages-for-the-last-200-years>

¹³⁷ M. Annunziata, S. Biller. The Future Of Work // General Electric. 2014.

URL: http://files.publicaffairs.geblogs.com/files/2014/04/AM_IL_FOW_WhitePaper_FINAL-1.pdf

¹³⁸ A Future that Works: Automation, Employment, and Productivity // McKinsey&Company. 2017.

URL: https://www.mckinsey.com/~media/McKinsey/Global Themes/Digital Disruption/Harnessing automation for a future that works/MGI-A-future-that-works_Full-report.ashx

¹³⁹ L. Fratocchi. Is 3D Printing an Enabling Technology? // Springer. 2017.

URL: https://link.springer.com/chapter/10.1007%2F978-3-319-58883-4_5

нию части производства из развивающихся стран в развитие (т.н. решоринг).

В свою очередь, для государств третьего мира последствия автоматизации могут быть гораздо более разрушительными. Доля рабочих мест, связанных с рутинным трудом, в развивающихся странах с 1995 по 2012 гг. снизилась¹⁴⁰ на 8%. Решоринг же способен в ближайшей перспективе существенно ускорить этот процесс. Поскольку доля людей, занимающихся низкоквалифицированным трудом, в странах с низким и средним уровнем доходов населения значительно выше, рост безработицы в них грозит стать крупной глобальной проблемой. Дополнительно усугубляет ситуацию неразвитость институтов защиты труда в подобных странах.

Необходимо отметить, что подобные риски характерны и для России. Несмотря на значительно более высокий уровень образования ее граждан по сравнению с развивающимися странами, российскую экономику трудно назвать высокотехнологичной. Значительная часть работоспособного населения занимается рутинным низкоквалифицированным трудом, при этом его производительность по-прежнему остается низкой. В настоящее время Россия существенно отстает¹⁴¹ по этому показателю от развитых стран (от США – более чем в два раза), а по некоторым оценкам, и от среднемирового уровня. Причем речь не только о фабричном труде – многомиллионная армия бюрократов и клерков также находится под угрозой сокращения в результате цифровизации.

Другая «мина замедленного действия» в российской экономике – отсталая промышленность и упадок отечественной инженерной школы. В настоящее время вузы выпускают преимущественно инженеров-эксплуатационников, ориентированных на обеспечение работы станков и машин. При этом даже существующий ограниченный инновационный потенциал российских инженеров не востребован¹⁴² российской промышленностью.

В то же время нельзя исключать, что в ближайшем будущем в России будет запущена массовая программа по роботизации и внедрению технологий ИИ. Тем более что она прекрасно сочетается с целями модернизации и цифровизации¹⁴³ национальной экономики, неоднократно озвученными российским руководством. Из-за отсутствия сильного профсоюзного движения и распространенности гибридных и «серых» форм занятости автоматизация труда может иметь в России гораздо более жесткие социальные последствия, нежели в западных странах. Причем вполне возможно, что в силу «догоняющего» характера подобной модернизации Россия бу-

¹⁴⁰ C.B. Frey, E. Rahbari. Do labor-saving technologies spell the death of jobs in the developing world? // Brookings.edu. 2016.

URL: https://www.brookings.edu/wp-content/uploads/2016/07/Global_20160720_Blum_FreyRahbari.pdf

¹⁴¹ Производительность труда в Российской Федерации // Аналитический центр при Правительстве Российской Федерации. 2017. URL: <http://ac.gov.ru/files/publication/a/13612.pdf>

¹⁴² Воспроизводство инженерных кадров: вызовы нового времени // УРФУ. 2015.

URL: http://elar.ufrfu.ru/bitstream/10995/32709/1/978-5-7996-0000_2015.pdf

¹⁴³ Президент подписал Указ «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года» // Президент России. 2018.

URL: <http://kremlin.ru/events/president/news/57425>

дет внедрять более примитивные технологии, чем в развитых государствах. Подобный сценарий, в частности, остроумно описал в своей статье¹⁴⁴ главный редактор журнала «Россия в глобальной политике», член РСМД Федор Лукьянов.

Спасти рядового работника

Способы снижения социальных последствий автоматизации труда уже давно находятся в центре дискуссии о IV промышленной революции и развитии технологии ИИ. Среди возможных мер обсуждается, в частности, «налог на роботов». Один из сторонников этой идеи – основатель компании Microsoft Билл Гейтс, который предложил¹⁴⁵ обложить труд роботов подоходным налогом и социальными выплатами, чтобы замедлить темпы автоматизации. «Сейчас, если работник-человек получает 50 тыс. долл., работая на фабрике, то его доход облагается налогом, и вы получаете подоходный налог, социальный налог и другие выплаты. Если эту же работу делает робот, то мы могли бы подумать о том, чтобы обложить его труд на том же уровне» – заявил он в интервью интернет-изданию Quartz. По его мнению, полученные от подобных выплат средства должны использоваться правительствами для создания систем социального обеспечения, ориентированных на тех, кто лишился работы в результате автоматизации.

Первой страной, прибегшей к этой мере, стала Южная Корея, которая ввела¹⁴⁶ непрямой «налог на роботов» в августе 2017 г. Введение аналогичного налога ранее также обсуждалось Европейским союзом, однако этот пункт предложения представителя Прогрессивного альянса социалистов и демократов Мади Дельво был отвергнут Европарламентом¹⁴⁷ по причине того, что подобная мера может замедлить развитие инноваций. При этом саму резолюцию, которая, в частности, предполагает наделение роботов статусом юридических лиц, парламент одобрил.

Смягчить эффект от роста безработицы и неравенства может также универсальный базовый доход. Сторонником его введения, наряду с другими бизнесменами и экспертами¹⁴⁸, является Илон Маск¹⁴⁹. В то же время отсутствие возможности найти работу и реализовать себя может стать значимой общественной проблемой. Кроме того, значительный объем безработицы, даже в условиях отсутствия бедности, может способствовать марги-

¹⁴⁴ Ф. Лукьянов. Россия 2035: грани возможного // Коммерсантъ. 2017.
URL: <https://www.kommersant.ru/doc/3384073>

¹⁴⁵ Билл Гейтс предложил обложить налогами труд роботов // РБК. 2017.
URL: <https://www.rbc.ru/business/18/02/2017/58a7ee769a7947f6e0503587>

¹⁴⁶ В. Vigiariolo. South Korea 'robot tax' is no tax at all; it's a warning of looming automation crisis // TechRepublic. 2017.
URL: <https://www.techrepublic.com/article/south-korea-robot-tax-is-no-tax-at-all-its-a-warning-of-looming-automation-crisis/>

¹⁴⁷ European parliament calls for robot law, rejects robot tax // Reuters. 2017.
URL: <https://www.reuters.com/article/us-europe-robots-lawmaking-idUSKBN15V2KM>

¹⁴⁸ С. Weller. Giving people free money could be the only solution when robots finally take our jobs // Business Insider. 2016.
URL: <http://www.businessinsider.com/basic-income-could-be-the-only-solution-in-a-robot-economy-2016-4>

¹⁴⁹ С. Weller. Elon Musk doubles down on universal basic income: 'It's going to be necessary' // Business Insider. 2017. URL: <http://www.businessinsider.com/elon-musk-universal-basic-income-2017-2>

нализации населения и росту преступности – ведь работу потеряют в первую очередь низкоквалифицированные сотрудники, которые вряд ли будут заниматься исключительно йогой и саморазвитием в свое перманентно свободное время.

Среди возможных способов, которые могут смягчить последствия грядущей перестройки мировой экономики, также называют изменение природы занятости¹⁵⁰. Технологические изменения и расширение доступа к Интернету позволяют все большему количеству людей работать дистанционно. Таким образом, часть тех, кто потеряет работу, сможет найти себя в новой экономике без необходимости смены места жительства.

Существует мнение¹⁵¹, что автоматизация в перспективе не только не снизит, но, возможно, и увеличит общее количество рабочих мест за счет ускорения темпов развития экономики. Одним из примеров того, как роботизация не привела к сокращению персонала, является компания Amazon, которая увеличила количество роботов¹⁵², занятых на своих складах с 1400 до 45 тыс., сохранив при этом прежний уровень занятости. Отмечается также, что автоматизация становится все более необходимой в связи с сокращением населения трудоспособного возраста (в первую очередь в развитых странах)¹⁵³.

Необходимо отметить, что все перечисленные меры носят ограниченный характер и вряд ли соответствуют масштабам изменений, которые принесет IV промышленная революция. Чтобы избежать массовой безработицы и социальной нестабильности, государствам в краткосрочной перспективе придется разработать комплексные стратегии по адаптации населения к новым условиям. Весьма вероятно, что потребуются разработка программ массового переобучения граждан новым специальностям.

Россия здесь не исключение, скорее наоборот, нашей стране особенно необходимо уже ближайшем будущем реформировать систему образования, в первую очередь технического. Не менее важно разработать адресные программы поддержки для различных групп населения, которые наиболее уязвимы для автоматизации и цифровизации. Представляется, при их создании целесообразно было бы использовать существующий опыт смягчения социальных последствий закрытия фабрик в российских моногородах. Если же продолжать двигаться текущим, инерционным путем, мы рискуем оказаться в положении своеобразной резервации для вчерашних технологий со стремительно маргинализирующимся населением.

¹⁵⁰ D. M. Gerstein. Will a Robot Take My Job? // The National Interest. 2018.

URL: <http://nationalinterest.org/feature/will-robot-take-my-job-25444>

¹⁵¹ M. Annunziata, S. Biller. The Future Of Work // General Electric. 2014.

URL: http://files.publicaffairs.geblogs.com/files/2014/04/AM_IL_FOW_WhitePaper_FINAL-1.pdf

¹⁵² S. Kessler. The optimist's guide to the robot apocalypse // QUARTZ. 2017.

URL: <https://qz.com/904285/the-optimists-guide-to-the-robot-apocalypse/>

¹⁵³ Global growth: Can productivity save the day in an aging world? // McKinsey & Company. 2015.

URL: [https://www.mckinsey.com/-/media/McKinsey/Global Themes/Employment and Growth/Can long term global growth be saved/MGI_Global_growth_Full_report_February_2015.pdf](https://www.mckinsey.com/-/media/McKinsey/Global%20Themes/Employment%20and%20Growth/Can%20long%20term%20global%20growth%20be%20saved/MGI_Global_growth_Full_report_February_2015.pdf)

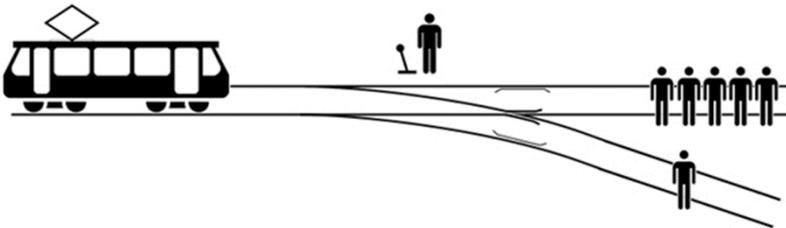
Этические и правовые вопросы искусственного интеллекта

Карлюк М.В.

Этика и право неразрывно связаны в современном обществе, и многие правовые решения вытекают из восприятия тех или иных этических проблем. Искусственный интеллект добавляет новое измерение в данные вопросы. Системы, использующие технологии искусственного интеллекта, становятся всё более автономными в смысле сложности задач, которые они могут выполнять, их потенциального влияния на мир и уменьшающейся способности человека понимать, предсказывать и контролировать их функционирование. Большинство людей недооценивает реальную автономность таких систем. Они могут учиться на собственном опыте и осуществлять действия, которые не были задуманы их создателями. Этим обусловлен ряд этических и правовых затруднений, которые будут затронуты в данной статье.

Этика и искусственный интеллект

Существует довольно известный эксперимент – проблема вагонетки. Он поднимает ряд важных этических вопросов, имеющих непосредственное отношение к искусственному интеллекту. Представьте, что неуправляемая вагонетка несётся по рельсам, а на ее пути к рельсам привязаны пять человек. Вы стоите возле рычага, с помощью которого можно переключить стрелку, и вагонетка повернёт и поедет по другому пути, где к рельсам привязан один человек. Переключите ли Вы стрелку?



Источник: Wikimedia.org

Однозначного ответа на данный вопрос нет. Более того, вариаций ситуации, в которой приходится принимать подобное решение, очень много¹⁵⁴. К тому же разные социальные группы дают разные ответы. Так, например,

¹⁵⁴ См. напр. D. Edmonds, *Would You Kill the Fat Man? The Trolley Problem and What your Answer Tells Us about Right and Wrong*, Princeton University Press. 2013.

буддистские монахи преимущественно готовы¹⁵⁵ пожертвовать жизнью одного человека, чтобы спасти пятерых даже в более сложном варианте проблемы вагонетки.

Что касается искусственного интеллекта, такая ситуация может возникнуть, например, на дороге, по которой движется беспилотный транспорт, в случае, если авария непреодолима. Возникает вопрос, чья жизнь должна быть в приоритете – пассажиров, пешеходов или ни тех, ни других? В Массачусетском технологическом университете был даже создан специальный сайт¹⁵⁶, посвящённый этой проблеме, где пользователь может попробовать на себе различные сценарии и выбрать, как стоило бы поступить в той или иной ситуации.

Возникают также вопросы, что позволено делать с правовой точки зрения? На основе чего делать выбор? И кто в итоге будет виноват? Данная проблема уже получила отклики от компаний и регуляторов. Так, представители компании «Мерседес» прямо сказали, что их машины будут отдавать приоритет пассажирам, на что в Министерстве транспорта Германии был сразу же дан ответ¹⁵⁷, предвосхищающий будущее регулирование, – что делать такой выбор на основе ряда критериев будет неправомерно и в любом случае производитель будет нести ответственность.

Другие страны могут выбрать иной путь. Возьмём, к примеру, китайскую Систему социального рейтинга¹⁵⁸, в соответствии с которой каждому гражданину присваивается рейтинг в зависимости от законопослушности, полезной общественной деятельности и др. Обладатели низких рейтингов будут подвергаться санкциям. Что помешает руководству Китая заставить производителей беспилотного транспорта предусмотреть правило, что в случае непреодолимой аварии жертвовать необходимо человеком с наименьшим рейтингом? За счёт, например, технологий распознавания лиц и доступа к соответствующим базам данных, вполне возможно определить и сравнить рейтинги потенциальных жертв.

Основные проблемы, стоящие перед правом

Но правовые проблемы ещё глубже, особенно в случае роботов. Система, которая учится на информации полученной из внешнего мира, может действовать таким образом, который создатели не могли предсказать¹⁵⁹, а предсказуемость критична для современных правовых подходов. Более того, такие системы могут действовать независимо от создателей или

¹⁵⁵ A.C. Madrigal. If Buddhist Monks Trained AI // The Atlantic. 2017.
URL: <https://www.theatlantic.com/technology/archive/2017/06/how-do-buddhist-monks-think-about-the-trolley-problem/532092/>

¹⁵⁶ Moral Machine // URL: <http://moralmachine.mit.edu/>

¹⁵⁷ M. Brown. Mercedes's Self-Driving Cars Will Kill Pedestrians Over Drivers // Inverse. 2016.
URL: <https://www.inverse.com/article/22204-mercedes-benz-self-driving-cars-ai-ethics>

¹⁵⁸ А. Гордеев. Цифровая диктатура: как в Китае вводят систему социального рейтинга // РБК. 2016.
URL: <https://www.rbc.ru/business/11/12/2016/584953bb9a79477c8a7c08a7>

¹⁵⁹ Asaro P., "From Mechanisms of Adaptation to Intelligence Amplifiers: The Philosophy of W. Ross Ashby", in Wheeler M., Husbands P., and Holland O. (eds.) *The Mechanical Mind in History*, Cambridge, MA: MIT Press. pp. 149–184

операторов, что усложняет задачу по определению субъекта ответственности. Данные характеристики ставят проблемы предсказуемости и способности действовать самостоятельно, но не быть юридически ответственным¹⁶⁰.

Варианты регулирования многочисленны, в том числе на основе уже существующих норм. Так, технологии, использующие искусственный интеллект, можно регулировать как объекты авторских прав либо как имущество. Однако сложности возникают, если учесть, в частности, способность к автономному действию помимо воли создателя, собственника или владельца. В связи с этим можно применить нормы, регулирующие особый вид имущества – животных, поскольку последние также способны к автономным действиям. В российском праве к животным применяются общие правила об имуществе (ст. 137 ГК РФ), поэтому ответственность будет наступать по ст. 1064 ГК РФ: вред, причинённый личности или имуществу гражданина, подлежит возмещению в полном объёме лицом, причинившим вред, которым является собственник автономного агента.

Предложения о применении правового режима животных высказываются¹⁶¹, однако они имеют некоторые ограничения. Во-первых, применение законодательства по аналогии недопустимо в рамках уголовного права. Во-вторых, данные нормы созданы в первую очередь для домашних животных, которые не должны причинять вред при обычных обстоятельствах. Для более развитых систем существуют предложения провести аналогию с дикими животными, поскольку для последних существуют более жёсткие правила¹⁶². Однако и здесь возникает вопрос, как сделать такое разделение в связи с особенностями искусственного интеллекта, указанными выше. Более того, жёсткие правила могут привести к замедленному внедрению технологий искусственного интеллекта в связи с большими и непредсказуемыми рисками ответственности для создателей.

Распространённым предложением является применение к таким системам норм о юридических лицах¹⁶³. Поскольку юридическое лицо является искусственно сконструированным субъектом права¹⁶⁴, то роботов также можно наделять аналогичным статусом. Право может быть достаточно гибким и наделять практически любого правами. Оно также может и ограничить в правах. Например, исторически рабы не обладали практически никакими правами и вообще являлись имуществом. Можно также наблюдать и противоположную ситуацию, когда объекты без явных признаков способности к действию наделяются правами. Даже сегодня существуют примеры необычных объектов, признающихся юридическими лицами, – как в развитых,

¹⁶⁰ Asaro P. The Liability Problem for Autonomous Artificial Agents // AAAI Symposium on Ethical and Moral Considerations in Non-Human Agents, Stanford University, Stanford, CA. p. 191.

¹⁶¹ Архипов В.В., Наумов В.Б. О некоторых вопросах теоретических оснований развития законодательства о робототехнике: аспекты воли и правосубъектности // Закон. 2017. № 5. С. 167.

¹⁶² Asaro P. The Liability Problem for Autonomous Artificial Agents. p. 193.

¹⁶³ Архипов В.В., Наумов В.Б. Указ. соч. С. 164.

¹⁶⁴ См. напр. Winkler A. We the Corporations: How American Business Won Their Civil Rights. Liverlight, 2018. См. описание здесь: <https://www.nytimes.com/2018/03/05/books/review/adam-winkler-we-the-corporations.html>

так и в развивающихся странах. В Новой Зеландии в 2017 г. был принят закон¹⁶⁵, который признал статус юридического лица за рекой Уонгануи. В законе указывается, что данная река является юридическим лицом и обладает всеми правами, полномочиями и обязанностями юридического лица. Таким образом, данный закон трансформировал реку из собственности в юридическое лицо, что расширило границы понимания относительно того, что может являться собственностью, а что нет. В Индии Верховный Суд в 2000 г. признал юридическим лицом основной священный текст¹⁶⁶ сикхов – Гуру Грантх Сахиб.

Но даже если не рассматривать экстремальные варианты, а привести в пример обычные компании, правовые системы предусматривают гражданско-правовую, а в некоторых случаях и уголовно-правовую¹⁶⁷ ответственность юридических лиц. Не определяя, что компании (или государства) обладают свободой воли или интенцией, или что они могут действовать преднамеренно или осознанно, считается возможным признавать их юридически ответственными за определённые действия. Таким же образом, не обязательно приписывать интенцию или свободу воли роботам, чтобы признавать их ответственными за то, что они делают.

Однако аналогия с юридическими лицами проблематична, поскольку, по сути, концепция юридического лица нужна для быстрого и эффективного осуществления правосудия, но действия юридических лиц всегда восходят к действиям индивида или группы людей, даже если невозможно их точно определить¹⁶⁸. Другими словами, правовая ответственность компаний и похожих образований связана с действиями, выполняемыми их представителями или работниками. Более того, в случае уголовной ответственности, когда она признаётся за юридическими лицами, она возможна только при условии, что физическое лицо, совершившее действие от имени юридического лица, определено¹⁶⁹. Действия же систем на основе искусственного интеллекта не обязательно будут прямо восходить к действиям человека.

Наконец, к таким системам можно применять правовые нормы об источниках повышенной опасности. В соответствии с п. 1 ст. 1079 ГК РФ юридические лица и граждане, деятельность которых связана с повышенной опасностью для окружающих (использование транспортных средств, механизмов и др.), обязаны возместить вред, причинённый источником повышенной опасности, если не докажут, что вред возник вследствие непреодо-

¹⁶⁵ Te Awa Tupua (Whanganui River Claims Settlement) Act 2017 // Parliamentary Council Office. New Zealand. 2017. URL: <http://www.legislation.govt.nz/act/public/2017/0007/latest/whole.html>

¹⁶⁶ T. Padmanabha Rao. Guru Granth Sahib, a juristic person: SC // The Hindu. 2000. URL: <https://www.thehindu.com/2000/04/03/stories/01030005.htm>

¹⁶⁷ В странах англосаксонской системы права, в странах ЕС, некоторых странах Ближнего Востока. Также, на постсоветском пространстве такая ответственность предусмотрена в Грузии, Казахстане, Молдове, Украине. В России таковая не предусмотрена, хотя и обсуждается.

¹⁶⁸ Brožek B., Jakubiec M. On the legal responsibility of autonomous machines // Artificial Intelligence Law. 2017. № 25(3), pp. 293–304.

¹⁶⁹ Khanna V.S. Corporate criminal liability: what purpose does it serve? // Harvard Law Review. 1996. № 109. pp. 1477–1534.

лимой силы или умысла потерпевшего. Проблема заключается в разграничении, какие системы могут быть отнесены к источникам повышенной опасности. Этот вопрос похож на вышеуказанную проблему разграничения домашних и диких животных.

Национальное и международное регулирование

Многие страны активно создают правовые условия для развития технологий, использующих искусственный интеллект. Так, в Южной Корее ещё с 2008 г. существует Закон¹⁷⁰ «О содействии развитию и распространению умных роботов». Закон направлен на улучшение качества жизни и развитие экономики путём разработки и продвижения стратегии устойчивого развития индустрии умных роботов, а правительство каждые пять лет разрабатывает основной план обеспечения эффективного достижения данной цели.

Здесь хотелось бы остановиться на двух недавних примерах – Франции, которая заявила об амбициях стать европейским и мировым лидером в сфере искусственного интеллекта; и Европейском союзе, в котором предложены продвинутое нормы регулирования умных роботов.

Франция

В конце марта 2018 г. президент Франции Э. Макрон презентовал¹⁷¹ национальную стратегию в сфере искусственного интеллекта. Франция планирует инвестировать 1,5 млрд евро в течение пяти лет для поддержки исследований и инноваций в данной сфере. Стратегия основана на рекомендациях, сделанных в отчёте¹⁷², подготовленном под руководством французского математика и депутата национального собрания Франции Седрика Виллани. Причём сделан выбор направить стратегию на конкретные четыре сектора: здравоохранение, транспорт, окружающую среду и оборону, и безопасность. Это сделано для того, чтобы использовать потенциал сравнительных преимуществ и сфер компетенций с фокусом на секторах, в которых компании смогут играть ключевую роль на глобальном уровне, а также из-за их важности для общественного интереса и др.

В целом даны семь ключевых предложений, одно из которых представляет особый интерес для целей статьи – сделать искусственный интеллект более открытым. Действительно, алгоритмы закрыты и в большинстве случаев являются коммерческой тайной. Однако алгоритмы могут быть предвзятыми, например, в процессе самообучения впитать стереотипы, существующие в обществе либо переданные разработчиками, и на их основе принимать решения. Таким образом уже принимаются судебные решения.

¹⁷⁰ Закон о содействии развитию и распространению умных роботов №9014 от 28.03.2008, с последующими изменениями и дополнениями // Исследовательский центр проблем регулирования робототехники и искусственного интеллекта Робоправо. 2008. URL: http://robopravo.ru/zakon_iuzhnoi_koriel_2008

¹⁷¹ AI For Humanity. URL: <https://www.aiforhumanity.fr/>

¹⁷² С. Villani. For a Meaningful Artificial Intelligence // AI For Humanity. 2018. URL: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

В США подсудимый был приговорён¹⁷³ к продолжительному сроку заключения на основе информации, полученной от алгоритма, оценивающего возможность повторного преступления¹⁷⁴. Подсудимый безуспешно оспаривал использование алгоритма для принятия такого решения, поскольку не были предоставлены критерии оценки, являющиеся коммерческой тайной. Французская стратегия предлагает развить прозрачность алгоритмов и возможностей по их проверке, а также определить этическую ответственность работающих в сфере искусственного интеллекта, создать консультативные комитеты по этике и т. д.

Европейский союз

В ЕС первым шагом в направлении регулирования вопросов искусственного интеллекта стала Резолюция Европейского Парламента 2017 г. «Нормы гражданского права о робототехнике»¹⁷⁵. Ещё в 2015 г. в Европейском Парламенте была создана рабочая группа по правовым вопросам, связанным с развитием робототехники и искусственного интеллекта в ЕС. Резолюция не является обязательным документом, но дает ряд рекомендаций Европейской Комиссии для возможных действий в этом направлении, причём не только касательно норм гражданского права, но и этических аспектов робототехники.

Резолюция определяет «умного робота» как робота, получающего автономию через использование сенсоров и/или взаимосвязь с окружающей средой; при этом робот имеет по меньшей мере минимальную физическую поддержку, адаптирует своё поведение и действия в соответствии с условиями среды, при этом у него отсутствует жизнь с биологической точки зрения. Предлагается создать систему регистрации продвинутых роботов, которая управлялась бы Агентством ЕС по робототехнике и искусственному интеллекту. Данное агентство также предоставляло бы техническую, этическую и регулятивную экспертизу по робототехнике. Что касается ответственности, предлагаются два варианта: либо объективная ответственность (не требующая вины), либо подход риск-менеджмента (ответственность лица, которое могло минимизировать риски). Ответственность должна быть пропорциональной реальному уровню указаний, которые отдаются роботу и уровню его автономности. Правила ответственности могут быть дополнены обязательным страхованием для пользователей роботов, и компенсационным фондом для выплаты компенсации в случае отсутствия страхового полиса, покрывающего риск.

Для решения этических вопросов резолюция предлагает два кодекса поведения: Кодекс этики для разработчиков робототехники и Кодекс комитетов

¹⁷³ M. Smith. In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures // The New York Times. 2016.

URL: <https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html>

¹⁷⁴ Ibid.

¹⁷⁵ Нормы гражданского права о робототехнике. Резолюция Европарламента от 16 февраля 2017 года // Исследовательский центр проблем регулирования робототехники и искусственного интеллекта Робоправо. 2017. URL: <http://robopravo.ru/uploads/s/z/6/g/z6gj0wkwhv1o/file/oQeHTCnw.pdf>

по этике научных исследований. Первый, в частности, предлагает четыре этических принципа: 1) «делай благо» (роботы должны действовать в интересах людей); 2) «не навреди» (роботы не должны причинять вред человеку); 3) автономия (взаимодействие человека с роботами должно быть добровольным); 4) справедливость (выгоды, получаемые от деятельности роботов, должны быть распределены справедливо).

* * *

Таким образом, примеры, приведённые в статье, показывают, среди прочего, как общественные ценности влияют на отношение к искусственному интеллекту и его юридическому оформлению. Поэтому наше отношение к автономным системам, будь то роботы или что-то другое, наше переосмысление их роли в обществе и положения среди нас, может иметь трансформационный эффект. Правосубъектность определяет то, что важно для общества, и позволяет решить, является ли «что-то» ценным и надлежащим объектом для обладания правами и обязанностями.

В связи с особенностями искусственного интеллекта звучат предложения о непосредственной ответственности определённых систем¹⁷⁶. В соответствии с такими взглядами, не существует фундаментальных причин, почему автономные системы не должны нести ответственность за свои действия. Однако открытым остаётся вопрос о необходимости либо желательности введения такой ответственности (во всяком случае на данном этапе). Это в том числе связано с этическими проблемами, указанными в статье. Возможно, сделать программистов или пользователей автономных систем ответственными за действия таких систем будет более эффективным. Однако это же может замедлить инновации. Поэтому необходимо продолжить поиски оптимального баланса.

Для этого важно ответить на многие вопросы. Например, какие цели мы преследуем, развивая технологии искусственного интеллекта? Чьи интересы должны преобладать? Насколько эффективно это будет? Ответ на эти вопросы поможет предотвратить ситуации подобно той, которая случилась в России XVII в., когда животное (а именно, козёл) было отправлено в ссылку в Сибирь за совершённые им действия¹⁷⁷.

¹⁷⁶ Hage J. Theoretical foundations for the responsibility of autonomous agents // Artificial Intelligence Law. 2017. № 25(3). pp. 255–271.

¹⁷⁷ U. Pagallo. The Laws of Robots. Cries, Contracts, and Torts. Springer. 2013. p. 36.

Об авторах

Карелов Сергей Владимирович – основатель и Chief Technology Officer компании Witology, председатель Лиги независимых IT-экспертов

Карлюк Максим Владимирович – Научный сотрудник Института права и развития ВШЭ-Сколково, Национальный исследовательский университет Высшая школа экономики

Колонин Антон Германович – канд. техн. наук, преподаватель, Новосибирский государственный университет, основатель проекта «Aigents», архитектор проекта «SingularityNET»

Маркоткин Николай Михайлович – программный менеджер по связям со СМИ и правительственными структурами РСМД, аспирант Федерального научно-исследовательского социологического центра РАН

Шефтелович Дмитрий Родионович – докт. естественных наук (Doktor der Naturwissenschaften), научный сотрудник Дортмундского технического университета, специалист в области автономных систем принятия решений

Российский совет по международным делам

Российский совет по международным делам (РСМД) – некоммерческая организация, ориентированная на выработку практических рекомендаций российским организациям, министерствам и ведомствам, задействованным во внешнеполитической деятельности.

РСМД объединяет усилия экспертного сообщества, органов государственной власти, бизнес-кругов и гражданского общества с целью повысить эффективность внешней политики России.

Наряду с аналитической работой РСМД ведет активную образовательную деятельность с целью сформировать устойчивое сообщество молодых профессионалов в области внешней политики и дипломатии.

Совет выступает в качестве активного участника публичной дипломатии, представляя на международных площадках российское видение в решении ключевых проблем глобального развития.

Члены РСМД – это ведущие представители внешнеполитического сообщества России: дипломаты, бизнесмены, ученые, общественные деятели и журналисты.

Президент РСМД, член-корреспондент РАН Игорь Иванов занимал пост министра иностранных дел РФ в 1998–2004 гг. и секретаря Совета Безопасности РФ в 2004–2007 гг.

Генеральным директором Совета является Андрей КОРТУНОВ. В 1995–1997 гг. Андрей КОРТУНОВ занимал должность заместителя директора Института США и Канады РАН.

Российский совет по международным делам

**МЕЖДУНАРОДНЫЕ И СОЦИАЛЬНЫЕ ПОСЛЕДСТВИЯ
ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

Верстка – В. Пучков

Источник фото на обложке:

справа сверху: кадр из кинофильма «2001: A Space Odyssey»,
1968, MGM, Stanley Kubrick Productions

слева внизу: Vostock Photo

справа внизу: Vostock Photo

Формат 70x100 $\frac{1}{16}$. Печать офсетная.

Печ. л. 3.75. Тираж 350 экз.

Для заметок

Для заметок



РОССИЙСКИЙ СОВЕТ ПО МЕЖДУНАРОДНЫМ ДЕЛАМ (РСМД)
119180, Москва, ул. Большая Якиманка, дом 1
Тел.: +7 (495) 225 6283
Факс: +7 (495) 225 6284
E-mail: welcome@russiancouncil.ru
www.russiancouncil.ru