

Пройдет ли Резолюция тест Тьюринга в новых реалиях?

Ирина Кохтюлина, ученый секретарь АНО «Национальный институт исследований глобальной безопасности», к.полит.н. Анатолий Смирнов, член НАМИБ, президент АНО «Национальный институт исследований глобальной безопасности», д.и.н., профессор МГИМО МИД России



*В 1950 году один из пионеров в области вычислительной техники и криптографии, английский учёный Алан Тьюринг в статье под названием «Может ли машина мыслить?» (англ. *Computing Machinery and Intelligence*) описал процедуру для определения момента, когда машина сравнивается в плане разумности с человеком.*

Интерпретация теста Тьюринга следующая: человек - судья взаимодействует с одним компьютером и одним человеком. На основании ответов на вопросы он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Задача компьютерной программы — ввести человека-судью в заблуждение, заставив сделать неверный выбор. При этом участники теста не видят друг друга». Если человек-судья не может сказать определённо, кто из собеседников является человеком, то считается, что машина прошла тест.

Тест Тьюринга в модернизированной форме используется для оценки новых поколений компьютеров, а также как критерий адаптации новейших информационно-коммуникационных технологий (ИКТ) к новым реалиям. При этом большинство экспертного сообщества считает технологии ИИ не как новое направление научно-технической революции, а как эволюцию ИКТ.

□

В условиях стремительного внедрения систем ИИ в жизнь человека, общества и государства Генеральная Ассамблея ООН 21 марта 2024 года приняла на консенсусной основе первую в истории резолюцию о продвижении “безопасных и заслуживающих доверия” систем искусственного интеллекта (ИИ), предложенную США в соавторстве с 123 государствами – членами. [\[1\]](#)

Немного истории

Согласно Уставу ООН главная ответственность за поддержание международного мира и безопасности возложена на Совет Безопасности ООН. Великобритания как действующий в июле 2023 г. председатель Совбеза ООН приняла решение провести 18 июля 2023 г. заседание на тему «Искусственный интеллект: возможности и риски для международного мира и безопасности».

Мининдел Великобритании Клеверли на заседании Совбеза обратился к генсеку ООН Антониу Гутерришу «Ваше присутствие сегодня подчеркивает важность рассматриваемой темы» [\[2\]](#). Генсек ООН, по его словам, уже шесть лет пристально наблюдает за развитием ИИ. Он подчеркнул, что не ожидал столь сильного влияния ИИ

на устойчивое развитие и социальный порядок и признался, что «был поражен и впечатлен» радикальным прорывом в возможностях ИИ, который проявился в ChatGPT, а также других подобных сервисах. Согласно прогнозам к 2030 г. ИИ может принести мировой экономике от \$10 трлн до \$15 трлн дохода.

Относительно «возможностей и рисков» Гутерриш заявил, что даже сами разработчики не знают, куда приведет их этот потрясающий технологический прорыв. Генсек ООН отметил, что *злонамеренное использование систем ИИ преступными и террористическими организациями вплоть до государственного уровня может вызвать смерти и разрушения ужасающих масштабов.*

В этом контексте он поддержал призывы ряда стран к созданию нового органа ООН по вопросам ИИ, по аналогии с МАГАТЭ или ИКАО.

Клеверли согласился и предложил провести в Лондоне глобальный саммит, посвященный безопасности и регулированию ИИ, в работе которого примут участие ученые, а также представители руководства стран–лидеров на рынке технологических компаний и «ключевых государств».

Саммит по безопасности ИИ в Блетчли–парке

В разрекламированном саммите по безопасности ИИ в Блетчли–парке 1–2 ноября 2023 года, приняли участие около 100 представителей правительств, ведущих компаний, занимающихся ИИ, гражданского общества и научных кругов 28 стран и ЕС. Россия не была приглашена.

Представители правительств 12 стран подписали достаточно обтекаемую Декларацию Блетчли, а также провели обсуждение на восьми круглых столах достаточно дежурных тем ^[3]. Важным для Великобритании стал запуск первого в мире Института безопасности ИИ, задачей которого является тестирование безопасности новых типов ИИ ^[4].

Институт призван тестировать новые типы ИИ до и после их выпуска. Цель – выявление потенциально опасных возможностей у моделей ИИ, включая изучение всех рисков, от социального вреда, такого как предвзятость и дезинформация, до самой экстремальной угрозы: полной потери человечеством контроля над ИИ.

Запустив Институт безопасности ИИ, Великобритания стала позиционировать себя как лидера в области безопасности ИИ и уже договорилась о сотрудничестве с американским институтом безопасности ИИ и с правительством Сингапура [\[5\]](#) о взаимодействии в области тестирования безопасности ИИ.

При этом, по замыслу коллективного Запада *Институт безопасности ИИ* должен стать носителем международного стандарта. В преддверии выхода новых мощных моделей ИИ, возможности которых, возможно, еще не полностью поняты, его первой задачей будет быстрое внедрение процессов и систем для их тестирования перед запуском, включая модели с открытым исходным кодом. На саммите достигнута договоренность, что переговоры по безопасности ИИ продолжатся в Южной Корее через полгода, а через год – на следующем саммите во Франции.

Нетрудно видеть, что коллективный Запад использовал уже апробированную методику обхода универсальных международных организаций, в первую очередь ООН.

В данном случае представители 12 государств–подписантов декларации приняли «правила» глобальной безопасности технологий ИИ в целях обеспечения технологического доминирования с далеко идущими геополитическими амбициями.

Основные положения Резолюции

Не без влияния критики итогов саммита в Блетчли–парке [\[6\]](#) коллективный Запад в лице США внес на рассмотрение Генассамблеи ООН Резолюцию «Использование возможностей безопасных, защищенных и надежных систем искусственного интеллекта для устойчивого развития», которая была консенсусом принята A/78/L.49 [\[7\]](#)

Резолюция состоит из 13 пунктов, в первом из них содержится призыв преодолеть разрыв в сфере ИИ и иных цифровых технологий между странами и в них. Тем самым подтверждена актуальность проблематики использования потенциала ИИ в достижении 17 целей устойчивого развития (ЦУР).

Документ ссылается на солидную базу резолюций ООН: 77/320 от 25 июля 2023 года о влиянии стремительного технического прогресса на достижение целей и выполнение задач в области устойчивого развития, 78/132 от 19 декабря 2023 года об использовании информационно-коммуникационных технологий в целях устойчивого развития, 78/160 от 19 декабря 2023 года об использовании научно-технических достижений и новаторства в целях устойчивого развития, 78/213 от 19 декабря 2023 года о поощрении и защите прав человека в контексте цифровых технологий, 77/211 от 15 декабря 2022 года о праве на неприкосновенность частной жизни в цифровую эпоху, 70/125 от 16 декабря 2015 года об общем обзоре хода осуществления решений Всемирной встречи на высшем уровне по вопросам информационного общества, на все итоговые документы Всемирной встречи на высшем уровне по вопросам информационного общества, включая Женевскую декларацию принципов [\[8\]](#), Женевский план действий [\[9\]](#), Тунисское обязательство [\[10\]](#)

и Тунисскую программу для информационного общества [\[11\]](#)

, а также на Декларацию о праздновании семьдесят пятой годовщины ООН, содержащуюся в ее резолюции 75/1 от 21 сентября 2020 года и др.

В резолюции учтены меры, предпринимаемые Международным союзом электросвязи (МСЭ) в партнерстве с 40 органами ООН, по организации платформы «Искусственный интеллект во благо», в том числе:

- Проведение ежегодной встречи на высшем уровне и запуск репозитория ИИ МСЭ в целях выявления практических видов применения ИИ для содействия достижению ЦУР.

- Принятие 23 ноября 2021 года Генеральной конференцией ЮНЕСКО Рекомендации об этических аспектах ИИ [\[12\]](#) и плана ее осуществления, включающего Методологию оценки готовности и Оценку этического воздействия, а также проведение Глобального форума по этическим аспектам ИИ.

- Принятие «Руководящих принципов предпринимательской деятельности в аспекте прав человека: осуществление рамок ООН в отношении “защиты, соблюдения и средств правовой защиты”» [\[13\]](#).

- Деятельность Управления Верховного комиссара ООН по правам человека в области ИИ.

Документ также ссылается на доклад Генсека ООН «Дорожная карта по цифровому сотрудничеству» [14] и учреждение Канцелярии Посланника Генсека по вопросам технологий для координации ее осуществления, а также учреждение Генсеком Консультативного органа высокого уровня по ИИ с участием многих заинтересованных сторон [15] и представление промежуточного доклада Консультативного органа, опубликованного 21 декабря 2023 года, и его заключительного доклада летом 2024 года в преддверии «Саммита будущего» ООН, который будет организован 22–23 сентября 2024 г. [16]

Проблема обеспечения безопасности ИИ

В резолюции значительное место уделено проблемам обеспечения безопасности ИИ. Так, в частности, определяется, что, жизненный цикл безопасных, защищенных и надежных систем ИИ в невоенной сфере включает этапы предварительного проектирования, проектирования, разработки, оценки, тестирования, внедрения, использования, продажи, закупки, эксплуатации и вывода из эксплуатации.

При этом системы ИИ названы антропоцентричными, надежными, объяснимыми, этичными, инклюзивными, в полной мере уважающими, поощряющими и защищающими права человека и международное право, обеспечивающими конфиденциальность, ориентированными на устойчивое развитие и ответственными. Системы ИИ способны ускорить и стимулировать прогресс в достижении ЦУР во всех его трех компонентах: экономическом, социальном и экологическом на сбалансированной и комплексной основе, а также способствовать цифровым преобразованиям, содействовать миру, преодолевать цифровой разрыв между странами и внутри стран, равно как и защищать осуществление прав человека и основных свобод.

В резолюции признается, что злонамеренное проектирование, разработка, внедрение и использование систем ИИ, например без соблюдения надлежащих мер безопасности или в нарушение норм международного права, создает риски, которые могут помешать достижению ЦУР, расширить цифровой разрыв между странами и внутри стран; усилить структурное неравенство и предвзятость, привести к дискриминации, подорвать целостность информации и доступ к ней, отрицательно сказаться на защите, продвижении и осуществлении прав человека и основных свобод,

В этой связи подчеркивается необходимость укрепления потенциала и оказания технической и финансовой помощи развивающимся странам, а также поддержки их эффективного, равноправного и значимого участия и представительства в международных процессах и форумах по управлению системами ИИ. При этом признается, что управление системами ИИ является формирующимся направлением деятельности и что необходимо продолжать обсуждение возможных подходов к управлению, которые были бы уместными, основанными на международном праве, совместимыми, гибкими, адаптируемыми, инклюзивными, учитывающими различные потребности и возможности как развитых, так и развивающихся стран и служили бы на благо всех по мере развития этой технологии и понимания ее сути.

Что заявляет Резолюция?

В резолюции, в частности, заявляется [\[17\]](#) :

1. Решимость преодолеть разрыв между странами и внутри стран в области ИИ и других цифровых технологий;
2. Содействие развитию безопасных, защищенных и надежных систем ИИ для ускорения прогресса в деле полной реализации Повестки дня в области устойчивого развития на период до 2030 года [\[18\]](#) , способствуя преодолению разрыва между странами и внутри стран в области ИИ и других цифровых технологий;
3. Призыв к государствам-членам и различным заинтересованным сторонам из всех регионов и стран, включая частный сектор, международные и региональные организации, гражданское общество, СМИ, научные и исследовательские учреждения, технические сообщества и отдельных лиц разрабатывать и поддерживать нормативно-правовые и организационно-управленческие подходы и рамки, относящиеся к безопасным, защищенным и надежным системам ИИ и создающие благоприятную экосистему.
4. Принимать меры по налаживанию сотрудничества с развивающимися странами и оказанию им помощи в обеспечении инклюзивного и равноправного доступа к преимуществам, создаваемым благодаря цифровому преобразованию и безопасным, защищенным и надежным системам ИИ, в том числе путем:
5. Права человека и основные свободы должны уважаться, защищаться и поощряться на протяжении всего жизненного цикла систем ИИ. Все государства-члены и другие заинтересованные стороны призываются воздержаться от использования систем ИИ, которые невозможно эксплуатировать в соответствии с международным правом прав человека или которые создают неоправданные риски для осуществления прав человека. Особенно для тех, кто находится в уязвимом положении, и подтверждает, что те права, которыми люди обладают вне интернета, должны быть также защищены и в интернете, в том числе на протяжении всего жизненного цикла

систем ИИ;

6. Рекомендует всем государствам-членам, в соответствии с их национальными приоритетами и обстоятельствами и при реализации их различных национальных нормативно-правовых и организационно управленческих подходов и рамок, а также другим заинтересованным сторонам содействовать развитию безопасных, защищенных и надежных систем ИИ путем:

7. Данные имеют основополагающее значение для разработки и функционирования систем ИИ. Справедливое, инклюзивное, ответственное и эффективное управление данными, совершенствование сбора данных, их доступности и соответствующей инфраструктуры и использование цифровых общественных благ имеют важнейшее значение для задействования потенциала безопасных, защищенных и надежных систем ИИ в интересах устойчивого развития а также для обеспечения большей согласованности и совместимости подходов к продвижению надежных трансграничных потоков данных для безопасных, защищенных и надежных систем ИИ;

8. Обсуждение изменений в области управления ИИ международное сообщество призывается продолжать осуществляемые в интересах всех сторон усилия, направленные на содействие инклюзивным исследованиям, систематизации и анализу вопросов, касающихся потенциальных видов воздействия и применения систем ИИ для ускорения достижения всех 17 ЦУР и информировать о том, как разрабатывать, продвигать и внедрять эффективные, совместимые на международном уровне меры безопасности, стратегии, стандарты и инструменты для проектировщиков, разработчиков, специалистов по оценке и внедрению, пользователей и других заинтересованных сторон в целях обеспечения безопасности, защищенности и надежности систем ИИ. При этом подчеркивается необходимость того, чтобы правительства, частный сектор, гражданское общество, международные и региональные организации, научные и исследовательские учреждения, технические сообщества и все другие заинтересованные стороны продолжали работать вместе, сообразно обстоятельствам;

9. Призывает частный сектор придерживаться применимых международных и национальных законов и действовать в соответствии с принятыми ООН «Руководящими принципами предпринимательской деятельности в аспекте прав человека: осуществление рамок Организации Объединенных Наций в отношении “защиты, соблюдения и средств правовой защиты”». При этом признается важность обеспечения более широкого и справедливого доступа к преимуществам безопасных, защищенных и надежных систем ИИ, а также необходимость расширения сотрудничества, в том числе между государственным и частным секторами и гражданским обществом, научными и исследовательскими учреждениями и техническими сообществами и внутри них, в целях обеспечения и поощрения справедливых, открытых, инклюзивных и недискриминационных условий для предпринимательства, экономической и коммерческой деятельности, конкурентных экосистем и торговых площадок на протяжении всего жизненного цикла безопасного, защищенного и надежного ИИ; а также призывает государства-члены разрабатывать стратегии и нормативные акты, способствующие развитию конкуренции в области безопасных, защищенных и надежных систем ИИ и связанных технологий.

10. Призывает специализированные учреждения, фонды, программы, другие структуры, органы и учреждения системы ООН продолжать оценивать и усиливать

принимаемые ими меры реагирования в целях использования возможностей и решения проблем, создаваемых системами ИИ, на совместной, скоординированной и всеохватной основе, задействуя соответствующие межучрежденческие механизмы, в том числе путем проведения исследований, систематизации и анализа потенциального воздействия и видов применения в интересах всех сторон; подготовки докладов о достигнутом прогрессе и проблемах в решении вопросов; и сотрудничества с развивающимися странами и оказания им помощи в наращивании потенциала, обеспечении доступа и совместного использования преимуществ безопасных, защищенных и надежных систем ИИ для достижения всех 17 ЦУР.

11. Согласно резолюции 76/307 от 8 сентября 2022 года и решению 77/568 от 1 сентября 2023 года *о порядке проведения и сфере охвата Саммита будущего и в этой связи ожидает разработки глобального цифрового договора*

;

12. Ожидает проведения Генассамблей в 2025 году общего обзора прогресса, достигнутого со времени проведения Всемирной встречи на высшем уровне по вопросам информационного общества;

13. Признает, что система ООН, в соответствии со своим мандатом, вносит уникальный вклад в достижение глобального консенсуса в отношении безопасных, защищенных и надежных систем ИИ, что соответствует международному праву, в частности Уставу ООН, Всеобщей декларации прав человека и Повестке дня в области устойчивого развития на период до 2030 года, в том числе путем поощрения международного сотрудничества и содействия вовлечению, участию и представленности развивающихся стран в обсуждениях

Выводы

1. Проблема обеспечения безопасности использования систем ИИ несомненно актуальна. Об опасности неконтролируемого развития ИИ предупреждали в США еще в марте 2023 г. эксперты и представители IT-отрасли, в том числе бизнесмен Илон Маск и сооснователь Apple Стив Возняк. [19] Выступая перед принятием Резолюции, Линда Томас-Гринфилд, посол и постоянный представитель США при ООН, отметила, что *страны мира "решили управлять ИИ и не допустить, чтобы он управлял людьми"*

2. Резолюция рассматривает только невоенный дискурс безопасности использования систем ИИ. На фоне галопирующей гонки ИИ вооружений во многих странах мира данная резолюция явно не отвечает требованиям новых реалий. *В силу этого Резолюция не проходит тест Тьюринга.*

3. В Резолюции США не осмелились поднять тему применимости существующих норм международного права к системе ИИ, как это изложено в выступлениях официальных представителей США и в известных трудах, в том числе в обоих изданиях «Таллинского руководства», подготовленных Центром передового опыта НАТО в Таллине. Как

известно, позиция России и её партнеров, изложенная в концепции Конвенции ООН об обеспечении международной информационной безопасности противоположна, ибо нацелена на необходимость заключения государствами в рамках ООН юридически обязательного многостороннего международного договора, который будет обеспечивать решение задачи предотвращения и урегулирования межгосударственных конфликтов в глобальном информационном пространстве, способствовать исключительно мирному использованию ИКТ [20] и создавать основу для сотрудничества государств в данных целях.

[21]

4. Вызывает удивление включение в п.11. Резолюции Генассамблеи *ожидание разработки глобального цифрового договора (ГЦД)* согласно

□

резолюции 76/307 от 8 сентября 2022 года и решению 77/568 от 1 сентября 2023 года *о порядке проведения Саммита будущего* □

(Нью-Йорк, 22-23 сентября 2024 г.). Анализ существующей версии ГЦД показывает ее аморфность, наличие в тексте множества противоречий.

[22]

5. Обеспечение безопасности и необходимость международного сотрудничества в сфере ИИ подчеркнута в обновленной Национальной стратегии развития искусственного интеллекта в России до 2030 года. [23] В документе предложены конкретные шаги по обеспечению безопасности систем ИИ, их регулированию и международного сотрудничества, в т.ч. с учетом опыта КНР, ЕС, ОЭСР и других стран и организаций.

6. В рамках председательства Российской Федерации в объединении БРИКС в 2024 г. Президент России утвердил *перечень поручений по итогам конференции «Путешествие в мир искусственного интеллекта»*, состоявшейся 24 ноября 2023 г., в котором, в частности, предусмотрено сотрудничество в формате БРИКС

[24]

. Правительству Российской Федерации совместно с МИДом России поручено: обеспечить включение вопросов, касающихся формирования этических стандартов в области ИИ, сбалансированного регулирования и научно–технического сотрудничества в указанной области, в повестку дня заседаний объединения БРИКС в рамках председательства России в объединении в 2024 г.

7. При всей значимости сотрудничества России по проблематике безопасности технологий ИИ в БРИКС, важное значение имеет продолжение взаимодействия с партнерами из ЕАЭС, ОДКБ, СНГ, Союзного государства, ШОС и других организаций, в том числе на международных площадках.

[1] <https://news.un.org/en/story/2024/03/1147831>

[2] Искусственный интеллект: возможности и риски для международного мира и безопасности – Совет Безопасности, 9381-е заседание // UN Web TV URL: <https://webtv.un.org/ru/asset/k1j/k1ji81po8p>
(дата обращения: 17.02.2024).

[3] The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023 // Gov.UK URL: <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>

[4] Prime Minister launches new AI Safety Institute // Gov.UK URL: <https://www.gov.uk/government/news/prime-minister-launches-new-ai-safety-institute>
(дата обращения : 17.02.2024).

[5] Позднее и с аналогичным институтом в Японии и в Сингапуре. – *Прим. авт.*

[6] См. Обновленная Национальная стратегия развития искусственного интеллекта до 2030 года: международные аспекты. Статья. Смирнов А.И., Кохтюлина И.Н. Журнал. Международная жизнь. <https://interaffairs.ru/news/show/44643> (Дата обращения: 25.03.2024).

[7] Резолюция ГА ООН A/78/L.49 Использование возможностей безопасных, защищенных и надежных систем искусственного интеллекта для устойчивого развития. URL: <https://documents.un.org/doc/undoc/ltd/n24/065/94/pdf/n2406594.pdf?token=yEq2KiGFWwzO97wA91&fe=true> (Дата обращения: 25.03.2024).

[8] См. A/C.2/59/3, приложение.

[9] Там же

[10] См. A/60/687

[11] Там же

[12] Рекомендации по этике искусственного интеллекта. ЮНЕСКО Цифровая библиотека. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (дата обращения: 23.03.2024).

[13] A/HRC/17/31, приложение.

[14] <https://documents.un.org/doc/undoc/gen/n20/102/53/pdf/n2010253.pdf?token=HDB4nsAFpxvPFsYNsO&fe=true>

[15] От России в личном качестве участвует директор Центра искусственного интеллекта, зав.кафедрой МГИМО МИД России. Абрамова А.В.. обсуждение 28 ФЕВРАЛЯ 2024, МГИМО <https://aicentre.mgimo.ru/aiab-interim-report-seminar>

[16] The Multistakeholder High–Level Advisory Body on Artificial Intelligence (AI) // United Nations URL: https://odin.mgimo.ru/upload/2023/10/AIAB_Onepager_231022_113938.pdf

[17] В целях более объективного восприятия приводится изложение, близкое к тексту

[18] <https://documents.un.org/doc/undoc/gen/n15/291/92/pdf/n1529192.pdf?token=Qzywa2Vt>

[914k8SWg4r&fe=true](#)

[19] <https://tass.ru/ekonomika/17395655> (Дата обращения: 23.03.2024).

[20] ИИ – это эволюция ИКТ

[21] <http://www.scrf.gov.ru/media/files/file/P7ehXmaBUDOOAAcATW2Rwa3yNK1bNAWI9.pdf> (Дата обращения: 23.03.2024).

[22] https://interaffairs.ru/virtualread/ia_rus/32024/files/assets/downloads/publication.pdf (Дата обращения: 23.03.2024).

[23] Указ Президента Российской Федерации от 15.02.2024 № 124 «О внесении изменений в Указ Президента Российской Федерации от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» и в Национальную стратегию, утвержденную этим Указом». URL: <http://publication.pravo.gov.ru/document/0001202402150063> (дата обращения: 23.03.2024). См. Смирнов А.И., Кохтюлина И.Н. Журнал. Международная жизнь. <https://interaffairs.ru/news/show/44643> (Дата обращения: 23.03.2024).

[24] Перечень поручений по итогам конференции «Путешествие в мир искусственного интеллекта» // Сайт Президента России URL: <http://kremlin.ru/acts/assignments/orders/73282> (дата обращения 06.03.2024).

[Читайте другие материалы журнала «Международная жизнь» на нашем канале Яндекс.Дзен.](#)

Подписывайтесь на наш Telegram – канал: <https://t.me/interaffairs>

Ключевые слова: [ООН](#) [ИКТ](#) [искусственный интеллект](#) [кибертехнологии](#)