Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ВОЗМОЖНОСТИ, РИСКИ И РЕГУЛИРОВАНИЕ

Контекст

Искусственный интеллект способен радикально изменить экономику и общество: ускорить научный прогресс, повысить качество жизни, преодолеть ресурсные ограничения и освободить людей от тяжелого и рутинного труда.

Машины давно стали сильнее и быстрее людей. В ближайшем будущем они могут превзойти человека и в интеллектуальной деятельности. Это создаст новые возможности для развития человечества, но одновременно поставит под угрозу привычные формы занятости, распределения доходов и общественного устройства.

Развитие искусственного интеллекта будет продолжаться, а суперинтеллект может появиться уже в обозримом будущем. Государство должно готовиться к этим изменениям.

1. Безопасность искусственного интеллекта

1.1 Международная система безопасности

Россия, преодолев диктатуру, должна стать страной, которая противостоит диктатуре во всем мире, и последовательно выступает за международный мораторий на социальный скоринг, распознавание лиц с помощью искусственного интеллекта и другие авторитарные практики технологического контроля над населением. Отдельным направлением международной политики должно стать предотвращение угроз, связанных с небезопасным суперинтеллектом.

Несмотря на отсутствие у России реалистичных возможностей самостоятельно в ближайшее время разработать фронтирную модель, страна должна стать лидером в формировании правил безопасности таких систем. России следует поддерживать международные соглашения, предотвращающие создание небезопасного суперинтеллекта, и ограничивать создание наиболее опасных систем до достижения научного консенсуса о безопасности суперинтеллекта.

Необходимо инициировать создание международной организации по контролю безопасности искусственного интеллекта по аналогии с Международным агентством по атомной энергии (МАГАТЭ). Она должна устанавливать стандарты безопасности передовых моделей, проводить независимую экспертизу и контролировать соблюдение международных соглашений.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

Начинать стоит с площадок, которые уже работают. В августе 2025 Генеральная Ассамблея ООН резолюцией A/RES/79/325 учредила Независимую международную научную группу по ИИ из 40 экспертов и Глобальный диалог по управлению ИИ. Россия должна участвовать в обоих механизмах и выдвигать в научную группу независимых исследователей, а не представителей госведомств. Следует также вернуться к процессу саммитов по безопасности ИИ и брать на себя конкретные обязательства. Россия должна соблюдать порядок взаимного уведомления о начале обучения моделей, превышающих согласованный порог вычислений. Сведения об инцидентах с опасными моделями государства должны передавать в международную организацию, описанную выше, и эти правила должны распространяться на Россию наравне с другими участниками.

Если безопасный суперинтеллект будет создан, его преимущества должны быть доступны всему человечеству. Ситуация, где суперинтеллект создается одной страной, может привести к геополитической нестабильности. По аналогии со сдерживанием ПРО России следует продвигать соглашение о равном доступе к суперинтеллекту и его использовании на благо всего человечества.

Необходимо разработать механизмы международного противодействия биотерроризму, кибертерроризму и другим видам преступной деятельности, возможности которых расширяются благодаря ИИ.

1.2 Национальные институты безопасности

Необходимо создать Институт безопасности искусственного интеллекта и специализированную комиссию по образцу существующих международных институтов и органов, контролирующих безопасность гражданской авиации. Новые модели ИИ, включая открытые, должны проходить экспертизу безопасности перед выпуском. Помимо государственной проверки необходим независимый аудит.

Компании-разработчики должны иметь механизмы экстренного прекращения работы опасных моделей. Наличие таких систем необходимо учитывать при оценке безопасности. Необходимо регулярно проводить учения по чрезвычайным ситуациям, связанным с искусственным интеллектом, включая атаки на критическую инфраструктуру и акты биотерроризма.

Также следует ускорить и упростить сертификацию обновлений программного обеспечения критически важных систем, чтобы эти процедуры не мешали оперативному исправлению уязвимостей.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

1.3 Предотвращение опасного использования ИИ

Следует ограничить распространение и дообучение моделей с отключенными защитными механизмами, позволяющих использовать ИИ для разработки биологического, химического и ядерного оружия. Доступ к таким возможностям может предоставляться по специальной лицензии.

Следует ввести механизмы идентификации пользователей (KYC) при доступе к опасным возможностям ИИ и проверки заказов на синтез белков для предотвращения разработки биологического оружия. Однако подобные механизмы не должны использоваться как инструмент цензуры и слежки за населением.

Системы ИИ в будущем смогут самостоятельно воспроизводиться. Это создает угрозу бесконтрольной утечки опасных моделей в общественный доступ. Значит, и меры противодействия должны быть похожи на меры борьбы с биологическими угрозами: физический рубильник, который останавливает утечку, карантинные меры и т.д. Эти меры также не должны использоваться государством в политических целях, например для осуществления цензуры и контроля за населением.

1.4 Открытые модели

Необходима государственная поддержка разработки моделей с открытым исходным кодом (open source) и/или открытыми весами (open weights).

Необходимо делать доступными открытые модели для предотвращения полной зависимости граждан, бизнеса и государства от крупных разработчиков закрытых систем.

1.5 Безопасность и инновации

Регулирование ИИ не должно препятствовать безопасному использованию технологий двойного назначения, таких как защита информационных систем и ядерных объектов и фармацевтика.

Следует снимать барьеры для внедрения инноваций, созданных с помощью искусственного интеллекта, прежде всего в медицине, биоинформатике и фармацевтике.

Следует ускорить внедрение лекарственных препаратов, уже одобренных регулирующими органами других стран.

1.6 Автономное оружие

Россия должна отказаться от разработки и применения атакующих летальных автономных систем, способных самостоятельно выбирать цели и принимать решения о нанесении ударов.

Необходимо установить международные правила ответственности за применение оружия, в котором выбор цели или решение об атаке осуществляется алгоритмом.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

2. Труд, экономика и социальные гарантии

2.1 Право на труд

Развитие ИИ приведет к частичному обесцениванию человеческого труда. Во многих сферах машина станет экономически выгоднее человека. При этом работа останется важным способом самореализации, участия в жизни общества и ощущения собственной полезности. Поэтому необходимо закрепить право на труд и создать механизмы, позволяющие людям сохранять занятость и достойный уровень жизни в условиях автоматизации. Сохранение занятости может быть оправданно даже тогда, когда автоматизация экономически эффективнее.

Люди продолжат ставить цели, принимать решения и руководить сложными процессами. ИИ позволит большему количеству работников заниматься задачами высокого уровня.

Кроме того, автоматизация интеллектуальной деятельности не означает немедленного исчезновения физического труда. Например, ИИ может разработать план модернизации коммунальной системы, но для его реализации по-прежнему потребуются люди.

Государство должно обеспечивать профессиональную переподготовку и помогать людям переходить в новые сферы деятельности.

Любая масштабная трансформация в сфере труда должна происходить с учетом прав работников. Государство должно предотвращать увольнения, которые ведут к крупным социальным последствиям, отслеживать сферы, сокращения в которых наиболее вероятны, и разработать модель оценки рисков и план модерации сокращений в крупных компаниях.

На случай массовой безработицы необходимо формирование специального денежного фонда.

2.2 Новая экономика человеческого труда и общественная деятельность

Подавляющая часть экономики будет полностью автоматизирована, поскольку машины смогут производить товары и услуги лучше и дешевле человека. Одновременно сохранятся сферы, в которых человеческое участие будет обладать самостоятельной ценностью, такие как уход за детьми и престарелыми.

Уже сегодня люди выбирают фермерские продукты вместо товаров массового производства. В будущем аналогичный спрос может возникнуть на самые разные товары и услуги, созданные людьми, даже если машины способны выполнять ту же работу.

Необходимо разработать стратегию развития нишевой экономики человеческого труда, предусматривающую поддержку ремесел, творчества, предпринимательства и других видов деятельности, где важны индивидуальность и непосредственное участие человека.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

Необходимо стимулировать людей заниматься творчеством, спортом, играми, предпринимательством и общественной деятельностью. Следует строить соответствующую инфраструктуру, спонсировать призовые фонды конкурсов и спортивных состязаний и развивать местные сообщества, общественные пространства и локальные праздники. Такие формы взаимодействия помогут сохранять социальные связи и вовлекать людей в общественную жизнь.

В перспективе общественная деятельность может стать основной работой для значительной части населения. Государство должно создавать рабочие места или выделять финансирование в областях науки, культуры, образования, спорта, местного самоуправления, общественной безопасности — роль этих сфер и доля занятых в них должны вырасти. Нужно вкладываться в повышение социального престижа такой работы.

2.3 Сокращение рабочего времени

Рост производительности должен сопровождаться сокращением продолжительности рабочего дня и рабочей недели. Это даст человеку больше времени для самореализации, а также будет способствовать найму большего количества людей.

2.4 Доходы и социальные гарантии

Автоматизация потребует новых механизмов распределения доходов. Экономические преимущества ИИ должны приносить пользу всему обществу.

Главный принцип социальной политики: **«Возможности — тем, кто хочет. Гарантии — тем, кто не может».**

Вместо единого безусловного базового дохода предлагаются к обсуждению другие механизмы, которые помогут избежать деградации общества.

Связь дохода с деятельностью. Доход человека должен оставаться связанным с его деятельностью, включая творчество и общественно полезную работу. Это сохранит стимулы к самореализации и поможет предотвратить саморазрушительное поведение, возникающее из-за отсутствия занятий и жизненных целей.

Дифференцированный базовый доход. Размер базового дохода может зависеть от уровня образования. Такая система должна сохранять материальные стимулы к обучению и интеллектуальному развитию.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

2.5 Финансирование программ поддержки занятости и базового дохода

Следует ввести механизм налогообложения, при котором для каждой компании рассчитывается нормативный фонд оплаты труда (ФОТ). Если фактические расходы компании на оплату труда ниже установленного норматива, разница направляется в фонд, описанный в разделе 2.1. Таким образом, компании получают альтернативу: создавать рабочие места и платить сотрудникам либо перечислять средства в фонд.

Возможны три варианта расчета нормативного ФОТ:

1. **В зависимости от оборота компании.**
2. **В зависимости от фактической прибыли компании.**
3. **В зависимости от размера ИИ-капитала** — количества GPU, вычислительной мощности, потребляемой энергии и других показателей использования ИИ.

При этом важно, чтобы минимальный размер оплаты труда (МРОТ) был выше налога на несозданное рабочее место, чтобы компаниям было выгодно нанимать людей, только если они приносят пользу.

Внедрение такого механизма следует осуществлять с опорой на Международную организацию труда (МОТ) и международные профсоюзные объединения.
Для предотвращения ухода от налогообложения и переноса бизнеса в другие юрисдикции рекомендуется заключение международного соглашения о согласовании принципов налогообложения ИИ.

3. Образование и развитие человека

Один из рисков массового использования ИИ и обесценивания интеллектуального труда — снижение когнитивных способностей человека. Если интеллектуальные задачи будут постоянно выполняться машинами, люди могут утрачивать навыки самостоятельного мышления и решения сложных проблем. Кроме того, у значительного числа людей снизится мотивация к получению образования.

Поэтому государству следует разработать меры стимулирования образования, в том числе финансовые.

Следует реформировать систему образования, исходя из новой экономической и социальной ситуации. Необходимо создать систему непрерывного образования, доступную человеку на протяжении всей жизни. Прежде всего следует учить людей пользоваться искусственным интеллектом.

Развитие интеллектуальных способностей должно стать такой же общественной задачей, как поддержание физического здоровья. Для этого необходимо создать инфраструктуру интеллектуального развития по аналогии со спортивной: образовательные пространства, клубы, соревнования и другие формы интеллектуальной активности.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

4. Искусственный интеллект и демократия

4.1 ИИ в государственном управлении

Искусственный интеллект необходимо использовать для анализа данных, прогнозирования и повышения качества государственного управления. У каждого ведомства могут появиться собственные системы ИИ, помогающие оценивать последствия решений и разрабатывать рекомендации.

Но даже если искусственный интеллект будет управлять отдельными процессами эффективнее человека, право определять цели государственной политики должно оставаться за демократическими институтами. Задача ИИ — находить способы достижения этих целей.

При этом необходимо определить категории решений, которые не могут приниматься искусственным интеллектом самостоятельно и должны оставаться исключительно в компетенции человека.

4.2 Контроль государственных решений

ИИ можно использовать для независимого аудита государственных решений, но важно не допустить ситуации, когда чиновники будут принимать определенные решения исключительно ради положительной оценки алгоритма. Человек должен сохранять возможность прервать автоматизированную процедуру или отклонить рекомендацию ИИ. Такое решение должно сопровождаться публичным обоснованием.

В государственном управлении следует использовать разные модели, учитывающие интересы различных групп населения. Это поможет сопоставлять альтернативные оценки и предотвращать монополию одной системы на определение оптимальной политики.

5. Развитие искусственного интеллекта в России

5.1 Технологическая независимость

России необходимо преодолеть отставание в разработке и применении искусственного интеллекта. Стратегия технологического развития должна сочетать создание собственных решений с доступом к лучшим зарубежным технологиям: полная технологическая автаркия не должна становиться самоцелью.

Необходимо восстановить доступ к современным вычислительным мощностям, передовым моделям и специализированным чипам. Однако даже снятие санкций не приведет к автоматическому получению российскими компаниями, институтами и гражданским обществом фронтирных моделей без ограничений в области кибербезопасности и биотехнологий. Государство должно содействовать получению такого доступа, предоставляя гарантии и обеспечивая аудит — это критически важно для защиты инфраструктуры и проведения научных исследований. Следует стимулировать программы обмена опытом с передовыми ИИ-лабораториями и восстанавливать доверие.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

Параллельно с этим следует поддерживать собственные исследования и открытые модели, чтобы снизить зависимость от отдельных иностранных поставщиков, а также создать условия для возвращения российских специалистов, уехавших из страны, и привлечения иностранных исследователей.

5.2 Вычислительная инфраструктура

Государство должно создать современную инфраструктуру для обработки данных, привлекать иностранные компании для строительства дата-центров, создавать благоприятные условия для инвестиций в эту сферу и обеспечивать развитие необходимой инфраструктуры и доступ к энергоресурсам.

Развитие вычислительных мощностей необходимо сочетать со строительством электростанций на основе возобновляемых источников энергии.

Строительство дата-центров в России может быть крайне выгодным для инвесторов из-за дешевой энергии и обширных территорий с холодным климатом. Малонаселенность этих территорий снизит социальное напряжение. При этом решение об использовании территорий стоит принимать на уровне конкретных муниципалитетов и с учетом интересов местных жителей, включая представителей коренных народов. Следует применять передовые экологические стандарты при строительстве и эксплуатации дата-центров, а также повышать энергоэффективность и сокращать негативное воздействие на окружающую среду.

Отдельным перспективным направлением должны стать робототехника и беспилотный транспорт. Для их развития также следует создать благоприятные условия и специальное регулирование в определенных зонах.

Для стимулирования инвестиций следует отменить законы, которые требуют «преземления» компаний, провести либерализацию законодательства и обеспечить защиту собственника.

6. Правовое регулирование искусственного интеллекта

6.1 Правовой статус ИИ

Необходимо законодательно определить правовой статус искусственного интеллекта и автономных ИИ-агентов. Здесь возможны два подхода, которые не исключают друг друга.

Вариант 1. ИИ-агент — инструмент

Ответственность за действия ИИ-агента несет его владелец. ИИ-агент выступает в качестве инструмента в руках пользователя.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

Вариант 2. Ограниченная правосубъектность ИИ-агента

ИИ-агент может получить статус «электронного лица» по аналогии с коммерческой организацией. Коммерческая организация самостоятельно отвечает по своим обязательствам, но ее собственник в определенных случаях также несет ответственность. Аналогичный принцип может применяться к ИИ-агенту: агент обладает ограниченной самостоятельной ответственностью, а владелец несет субсидиарную ответственность, поскольку определяет условия его деятельности.

6.2 Ответственность за причиненный вред

Главный принцип: робот не должен своими действиями или бездействием допускать причинения вреда человеку. Ответственность за такие действия должна быть распределена между разработчиком, оператором, владельцем и пользователем ИИ в зависимости от того, по чьей вине был нанесен вред. Для систем ИИ повышенного риска (беспилотный транспорт, медицинские системы, управление критической инфраструктурой, автономные агенты, совершающие юридически значимые действия) следует ввести ответственность независимо от вины. Такие системы нужно приравнять к источникам повышенной опасности: оператор или владелец возмещает вред, если не докажет, что он возник вследствие непреодолимой силы или умысла потерпевшего. Сам разработчик конечной системы должен отвечать без вины за вред, причиненный недостатками модели как продукта, по аналогии с ответственностью изготовителя за недостатки товара.

Следует ввести обязательное страхование гражданской ответственности при разработке и использовании искусственного интеллекта.

6.3 Идентификация агентов

Агент при взаимодействии с людьми и другими агентами должен сообщать, что является искусственным агентом.

Для отдельных видов коммуникации следует обеспечить возможность установить, в интересах какого человека или организации он действует. Чтобы избежать рисков авторитарных практик, идентификация должна происходить без контроля государства, возможно с использованием криптографических протоколов.

Для выполнения определенных сложных операций с помощью ИИ может требоваться специальное обучение пользователя. По аналогии с правилами для квалифицированных инвесторов в финансовом регулировании можно вводить ограничения на операции агентов.

6.4 Авторское право

Необходимо реформировать законодательство об интеллектуальной собственности с учетом распространения генеративного ИИ. Контент, созданный искусственным интеллектом, должен подлежать обязательной маркировке.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

Одновременно необходимо расширить возможности использования произведений для развития технологий. Для этого предлагается изменить форму охраны: право запрещать обучение моделей ИИ на произведении должно возникать только после регистрации авторского права в открытом государственном реестре. Если правообладатель зарегистрировал произведение, обучение на нем допускается только с его согласия (opt-in). Если произведение не зарегистрировано, обучение разрешено без согласия правообладателя, а остальные авторские права на него сохраняются в полном объеме. Регистрация, таким образом, служит способом отказа от использования произведения для обучения (opt-out). Сроки охраны следует сократить до минимума, который допускает Бернская конвенция (50 лет после смерти автора вместо нынешних 70), а принцип добросовестного использования (fair use) расширить. Необходимо заменить ОКУПы более справедливыми алгоритмическими моделями распределения авторских гонораров.

Россия должна участвовать в модернизации международных норм авторского права с учетом развития ИИ, в том числе в пересмотре соответствующих конвенций и договоров.

6.5 Персональные данные

Законодательство о защите персональных данных также должно применяться и к системам искусственного интеллекта. Для тренировки моделей искусственного интеллекта на основе персональных данных необходимо требовать явного согласия субъекта. В научных целях можно применять упрощенный порядок работы с такими данными, но они должны быть обезличены.

Россия должна ратифицировать Протокол CETS 223 (Конвенция 108+ о защите персональных данных), присоединиться к Рамочной конвенции Совета Европы об искусственном интеллекте и правах человека, демократии и верховенстве права (CETS № 225), открытой для подписания в сентябре 2024 года. Конвенция открыта и для стран, не входящих в Совет Европы, но им для присоединения нужно приглашение Комитета министров. Поэтому присоединение к ней следует включить в план восстановления отношений России с Советом Европы, а ее требования к прозрачности, оценке рисков и средствам правовой защиты учитывать в национальном законодательстве об ИИ и приватности.

Мораторий на распознавание лиц и социальный скоринг, за который Россия должна выступать на международном уровне, нужно прежде всего ввести у себя. Идентификацию людей в общественных местах с помощью распознавания лиц следует запретить, кроме розыска конкретного человека по решению суда. Московскую систему «Сфера» и ее региональные аналоги нужно вывести из эксплуатации, а собранные ими биометрические базы уничтожить.

Этот документ — плод совместной работы группы экспертов. Это не часть политической программы, а вариант решения проблемы, предложенный участниками форума.

6.6 Свобода информации и ограничения ИИ

Необходимо отменить политические цензурные ограничения, установленные законом и применяемые Роскомнадзором. Ограничения на деятельность систем ИИ должны определяться общим законодательством и не выходить за пределы установленных им запретов. Все государственные ограничения работы моделей должны быть публичными и прозрачными.

Необходимо различать массовые ИИ-сервисы и персональные модели. На массовые сервисы искусственного интеллекта должны распространяться ограничения, аналогичные действующим в отношении средств массовой информации. При этом пользователь должен иметь возможность самостоятельно настраивать собственную персональную модель.

Для повышения достоверности информации необходимо развивать краудсорсинговую проверку фактов и системы коллективных примечаний по аналогии с Community Notes, а также вводить требования об обязательном указании источников в ответах моделей искусственного интеллекта.

НАД ДОКУМЕНТОМ РАБОТАЛИ:

Саркис Дарбинян, киберадвокат, исполнительный директор RKS Global, правовой советник VPN Guild

Р.И., эксперт по ИИ

Наталья Крапива, юристка и экспертка по цифровым правам

Александр Лурье, IT-инженер, предприниматель

Михаил Самин, исполнительный директор AI Governance and Safety Institute