



УПРАВЛЯЯ НЕИЗБЕЖНЫМ: ПРАВОВЫЕ ПРИОРИТЕТЫ РАЗВИТИЯ ПОЛИТИЧЕСКИХ ИНСТИТУТОВ В ЭПОХУ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. В статье рассматриваются системные конституционные вызовы, возникающие в связи с распространением искусственного интеллекта (ИИ) в политической сфере. Автор отмечает, что ИИ трансформирует институты демократии, создавая риски подмены публичного диалога непрозрачными алгоритмами, алгоритмической дискриминации, микроманипулирования информацией и концентрации власти у обладателей технологий. Особое внимание уделяется угрозе социогуманистической парадигме в условиях выбора между классическим гуманизмом и трансгуманизмом. В качестве ответа на эти вызовы предлагается система из восьми правовых принципов регулирования ИИ, включая государственное стратегическое руководство, принцип «человека в контуре», антропологический приоритет, цифровое равенство и управляемую прозрачность. В контексте избирательного процесса анализируются такие специфические риски, как использование микротаргетинга, «темной рекламы», дипфейков и ботов, что ведет к угрозе честности выборов и манипуляции волей избирателей. Делается вывод о необходимости развития национальных моделей искусственного интеллекта, законодательного регулирования, включая аудит алгоритмов и запрет манипулятивных технологий на выборах для обеспечения суверенитета и легитимности политических институтов в цифровую эпоху.

Ключевые слова: искусственный интеллект, нейросеть, конституционные вызовы, политические институты, принципы правового регулирования ИИ, выборы и манипуляции, социогуманистическая парадигма.

MANAGING THE INEVITABLE: LEGAL PRIORITIES FOR THE DEVELOPMENT OF POLITICAL INSTITUTIONS IN THE ERA OF ARTIFICIAL INTELLIGENCE

Abstract. This article examines the systemic constitutional challenges arising from the spread of artificial intelligence (AI) in the political sphere. The author notes that AI is transforming democratic institutions, creating risks of substituting public dialogue with

КАБЫШЕВ Сергей Владимирович — профессор кафедры конституционного и муниципального права Московского государственного юридического университета имени О.Е. Кутафина, председатель комитета Государственной Думы Федерального Собрания Российской Федерации по науке и высшему образованию, кандидат юридических наук, доцент, заслуженный юрист Российской Федерации, г. Москва

opaque algorithms, algorithmic discrimination, micromanipulation of information, and the concentration of power in the hands of technology holders. Particular attention is paid to the threat to the socio-humanistic paradigm in the context of the choice between classical humanism and transhumanism. In response to these challenges, a system of eight legal principles for regulating AI is proposed, including state strategic guidance, the «human in the loop» principle, anthropological priority, digital equality, and managed transparency. Specific risks, such as the use of microtargeting, «dark advertising», deepfakes, and bots, are analyzed in the context of the electoral process, which threaten the integrity of elections and the manipulation of voter will. The article concludes that it is necessary to develop national AI models and legislative regulation, including algorithm audits and bans on manipulation technologies in elections, to ensure the sovereignty and legitimacy of political institutions in the digital age.

Keywords: *artificial intelligence, neural network, constitutional challenges, political institutions, principles of AI legal regulation, elections and manipulation, socio-humanistic paradigm.*

Искусственный интеллект (или нейросеть) уже превратился в инфраструктурную технологию, которая не просто влияет на все ключевые сферы государственной деятельности, экономики и социальной жизни, но и определяет, по существу, новую политическую эпоху. Доступ к соответствующим технологиям становится всевозрастающим фактором, который вместе с новыми, доселе невиданными информационными возможностями, включая огромный потенциал повышения уровня и качества знаний для стратегического управления и принятия политических решений, способен существенным образом повлиять на изменение предпочтений избирателей [1; 2] и провоцировать серьезные отступления от конституционных принципов с разрушительными последствиями.

Конституционные вызовы технологий искусственного интеллекта носят системный, синергетический и экзистенциальный характер. Здесь можно обозначить (разумеется, лишь пунктирно и не в исчерпывающей мере) такие проблемы, как:

подмена демократических институтов волеобразования народа, основанных на общественном диалоге, управлением посредством рекомендаций непрозрачных алгоритмов, анализирующих большие данные, и слепая вера в способность таких алгоритмов решать самые сложные социальные и политические проблемы (что ведет к упрощению проблем, отказу от демократического обсуждения ценностей и целей развития, а сами алгоритмы делает не просто средством, но и в известном смысле «источником» власти, проистекающим из уникальной «экспертной» легитимности); данная проблема лишь усугубляется риском фрагментации общества под воздействием рекомендательных алгоритмов, управляющих социальными сетями, на изолированные группы с разными «реальностями», что может ставить под сомнение социальное единство и целостность народа как носителя суверенитета и единственного (и единого) источника власти;

многомерное технологически обусловленное неравенство (радикально различающиеся условия реализации прав и свобод в зависимости от доступа к технологиям являются новым вектором социального расслоения и создают предпосылки для фактической маргинализации граждан, не интегрированных в цифровую среду) и алгоритмическая дискриминация (различие в обращении на основе предвзятых данных);

микроманипулирование информационным потоком, подавление одних тем и продвижение других посредством алгоритмов и симулирования поддержки либо протеста (через автоматизированные боты, ИИ-агенты), либо создания убедительных «фейков» (что подрывает саму основу свободного формирования убеждений и информированного выбора);

смещение и концентрация ресурсов власти (основанной на доступе к большим данным вычислительных ресурсов и аналитическим возможностям) у обладателей алгоритмов (что может нарушать баланс в системе публичной власти, в частности, существенно ослабляя потенциал законодательных и судебных органов).

Вместе с тем мета-угрозой, стоящей над всеми остальными, является вызов социогуманистической парадигме, имеющей основополагающее значение не только для развития конституционализма и права вообще, но и для самой цивилизации, которая в этой парадигме разворачивается вокруг и для человека. В.Д. Зорькин, обнажая самую суть проблем социально-правового развития в условиях новой технологической реальности, предлагает сосредоточить внимание на мировоззренческой стороне вопроса: на какой ценностно-концептуальной основе допустимо и оправданно выстраивать правовые режимы распространения новейших технологий, — на основе трансгуманизма, который, связывая совершенство человека (включая этический аспект) с уровнем технологического развития, не просто стремится к синтезу человека и машины, но и очеловечивает, а в итоге обожествляет машину, либо же на основе классического гуманизма, воплощающего традиционные духовно-нравственные ценности и не допускающего разделения личности и совести, бездушного релятивизма морали и ответственности?

Именно в этой дилемме, по мнению Зорькина, заключается главный, коренной вызов искусственного интеллекта для человечества и его права, которое оказывается под угрозой сохранения человечности, нравственного начала, может превратиться в систему умозрительно-фикционных алгоритмов, сводящих личность к определенным рефлексам, поведенческим реакциям, и в этом смысле «освобождающих» ее от собственных естественных ограничений, позволяющих ей приобретать («выбирать») любые искусственные формы [3, с. 257–277].

Капитуляция перед трансгуманизмом ведет к двуединым последствиям, состоящим в «очеловечивании» машины и деперсонализации (алгоритмизации, расчеловечивании) человека. Технологии искусственного интеллекта, особенно в увязке с био- и психометрией и нейротехнологиями могут

ставить под вопрос саму автономию человеческой личности и ее достоинство. Речь идет о редукции юридического и политического понимания человека как свободного, разумного и целостного субъекта к прогнозируемому «профилю», что дегуманизирует общественные отношения и политическую сферу, в то время как развивающиеся в сторону всеохватности технологические системы мониторинга и контроля способны превратить общественное пространство в паноптикум, где человек оказывается в положении объекта постоянной оценки, социального скоринга, а не свободного субъекта.

* * *

В поисках ответов на фундаментальные вызовы, связанные с распространением искусственного интеллекта, необходимо опираться на взвешенный, сбалансированный общий подход к формированию правовых режимов для соответствующих технологий, избегая алармизма и самонадеянной беспечности. Можно предложить (в порядке обсуждения) авторскую точку зрения на ряд основных принципов, служащих своеобразным методологическим каркасом для обеспечения устойчивого развития технологий искусственного интеллекта на основе защиты интересов человека, общества и государства в цифровой сфере.

1. *Принцип государственного стратегического руководства и обеспечения технологической целостности в сфере искусственного интеллекта.* Российская Федерация, признавая искусственный интеллект стратегическим направлением развития, должна осуществлять комплекс суверенных прерогатив и мер, направленных на формирование и защиту полного национального технологического цикла (от фундаментальных исследований и подготовки кадров до создания элементной базы, программного обеспечения, данных и рынков сбыта), активное формирование регулятивных, экономических и инфраструктурных условий для ускоренного развития и безопасного применения искусственного интеллекта, достижение и удержание лидерских позиций на основе создания конкурентоспособных, безопасных и этичных технологий. Данным принципом охватывается также суверенный контроль в отношении критически значимых технологий нейросетей в жизненно важных сферах (национальная безопасность, оборона, критическая информационная инфраструктура, энергетические системы и т.п.) и формирование «фильтров национальной идентичности», обеспечивающих соответствие работы алгоритмов принципам и ценностям, лежащим в основе российского общества, его традиционным духовно-нравственным идеалам.

2. *Принцип гарантированной защиты пострадавшего и повышенной ответственности* [4, с. 37–52; 5, с. 2–5]. Это подразумевает законодательное закрепление по крайней мере статуса передовых моделей искусственного интеллекта (наиболее развитые модели, обладающие высоким уровнем автономного функционирования, которые сейчас демонстрируют способности экспертного уровня в разных дисциплинах: биология, химия, математика и др.),

как источника повышенной опасности, что должно предполагать, вероятно, солидарную ответственность, в частности, разработчика (создатель алгоритма, модели), оператора, владельца (лицо, использующее систему). Очевидно, что сбой возможен на каждом этапе технологической цепи, а для пострадавшего крайне затруднительно с технической стороны обосновать его причину. Стоит подумать над формированием специального государственного или отраслевого фонда страхования или компенсаций (с возможностью регрессных требований к участникам соответствующей цепочки). Здесь важно исключить такие максималистские подходы, которые приведут к «бегству» наукоёмких компаний в юрисдикции с более мягким регулированием или сверхконсервативному дизайну алгоритмов (например, отказу диагностической системы ставить редкий диагноз).

3. *Принцип «человека в контуре» в критических решениях*: обязательное сохранение за человеком окончательного решения в социально значимых вопросах, затрагивающих его жизнь, здоровье и свободу.

4. *Принцип антропологического приоритета*. Любая система искусственного интеллекта должна строиться на том, что ее целью, мерилom и ограничителем является человек, — его автономия, достоинство, свобода и благополучие. Технология не имеет собственной ценности вне служения этим целям. Данный принцип противостоит узкотехнократической парадигме, где эффективность, скорость и экономическая выгода, предоставляемые за счет алгоритма, могут превращаться в самоцель. Дизайн и интерфейс названных систем должны создаваться для усиления возможностей человека (образно говоря, для со-, а не автопилотирования), предоставлять ему лучшее ситуационное понимание, устранять рутину, обрабатывать большие данные, но оставлять за ним смысловое лидерство, творчество и ответственность, что подразумевает также применение проектных решений, исключающих иллюзию истинности машины (результаты работы алгоритма должны сопровождаться пояснением основных факторов, визуализацией неопределенности и возможностью для эксперта дать оценку принципам и методам принятия решений). Вместе с тем не должны допускаться разработка и применение систем искусственного интеллекта, которые используют, в том числе, персональные и поведенческие данные для скрытого воздействия на пользователя путем эксплуатации известных когнитивных искажений с целью подрыва автономии его воли, принятия решений в ущерб его очевидным интересам или в условиях, когда пользователь находится в особо уязвимом положении, о котором известно системе.

5. *Принцип цифрового равенства и запрета дискриминации* предполагает в том числе обязательность регулярного независимого аудита на предвзятость соответствующих технологий, принимающих решения в публичной и социально значимой сфере, а также запрет на использование систем, демонстрирующих необъяснимые и неустраняемые различия в обращении.

6. *Принцип уважения достоинства личности и невмешательства в частную сферу*: запрет на массовое психографическое профилирование для скрытого манипулирования поведением, ограничение использования эффективных технологий (распознавание эмоций) и социального скоринга в публичной сфере, сохранение пространства для «неоцифрованной» жизни.

7. *Принцип управляемой прозрачности в зависимости от риска*: для систем высокого риска должна обеспечиваться возможность гражданина на получение содержательного объяснения алгоритмического решения, повлиявшего на его права и интересы, в форме, понятной человеку, а также обязанность разработчиков предоставлять регулирующим органам доступ к алгоритмам и данным для проверки.

8. *Принцип технологической нейтральности и опережающего регулирования*: ориентация на регулирование свойств и рисков систем, а не конкретных технологий (которые меняются и устаревают), и создание специальных правовых режимов для тестирования инноваций в реальных и контролируемых условиях.

Остановлюсь подробнее на некоторых основных рисках для политико-правовых отношений, включая сферу выборов, которые требуют надлежащего учета в правовом регулировании. Здесь важно иметь в виду, что сама практика применения технологий искусственного интеллекта становится всё более обширной и влиятельной и, как свидетельствует анализ в том числе международного опыта, во многих случаях речь идет о попытках навредить конкурентам или организации выборов, при том, что само происхождение соответствующего контента зачастую установить невозможно [6].

Поскольку функционирование нейросетей определяется алгоритмами, которые формируются и обучаются в рамках конкретно заданного социокультурного и исторического контекста, применение неоригинальных (ненациональных) моделей искусственного интеллекта в политической сфере может ставить под угрозу российскую общегосударственную идентичность, формировать представления и взгляды и побуждать к действиям (либо бездействию), которые основаны на ценностях и подходах, не соответствующих российским политико-правовым традициям, идеалам и реалиям. Экспертные исследования показывают, что многие существующие модели нейросетей склонны воспроизводить набор предпочтений, совпадающий с ценностями США и Западной Европы, особенно в вопросах прав личности, гендерной политики и социальной справедливости. Отмечается тенденция моделей рассматривать западные нормы как универсальные и глобально применимые. Есть конкретные исследования, которые демонстрируют, что в ситуациях морального выбора нейросети опираются на западные философские традиции: утилитаризм и либеральную этику [7; 8, р. 971–1000; 9].

В связи с этим при внедрении искусственного интеллекта в политическую сферу важно ориентироваться на развитие собственных моделей и формирование системы доверенных технологий, что становится важнейшим

условием обеспечения суверенитета. Отечественные модели позволят учитывать национальный контекст, специфику российского права, приоритеты государственной политики и особенности социальной структуры.

* * *

Нейросеть способна кардинально трансформировать само понятие свободы выбора и таким образом привести к подрыву свободного, осознанного и добровольного политического волеизъявления. Так, технологии микро-таргетинга, основанные на анализе личных данных, позволяют создавать персонализированную пропаганду, формирующую у избирателя искусственную информационную реальность («смысловые галлюцинации») и эффективно управляющую его мотивацией и убеждениями. Применение в этих целях так называемой темной рекламы (dark ads), которая видна только конкретному пользователю или узкой, специально подобранной аудитории, недоступна для всеобщего наблюдения и включает в себя динамическое изменение контента (алгоритмы тестируют тысячи различных вариантов заголовков, изображений, текстов, выявляют предпочтительный, наиболее эмоционально заряженный и эффективный формат, и демонстрируют его остальной аудитории), превращает политическое взаимодействие в скрытое одностороннее навязывание, которое практически невозможно распознать (и оценить на предмет достоверности) отдельному гражданину и проблематично контролировать уполномоченным органам.

В этих же целях могут применяться технологии так называемой «армии ботов», когда искусственный интеллект управляет сетями фейковых аккаунтов, имитирует общественное мнение, «раздувает» скандалы и создает видимость массовой поддержки или осуждения, манипулируя общественной повесткой. Такое системное воздействие, направленное на то, чтобы склонить к определенному выбору, по сути может выполнять и ту роль, которую играет предвыборная агитация, но с учетом характера, предметного охвата информации его крайне сложно отследить и идентифицировать в таком качестве.

При этом технологии позволяют создавать поддельные заявления кандидатов, компрометирующие материалы или искаженные репортажи с избирательных участков, которые практически невозможно отличить от реальных. В этих условиях правда может не поспевать за ложью. Но такие материалы могут использоваться для дискредитации оппонентов, подавления явки или провокаций социальной напряженности.

В этом плане представляет интерес опыт Китая, где была установлена обязанность разработчиков алгоритмов, которые «оказывают влияние на общественное мнение» или «способны мобилизовать общество», регистрировать алгоритм. Регистрация алгоритма — это специальная процедура, в рамках которой сведения об алгоритме вносятся в специальный реестр, поддерживаемый и контролируемый уполномоченным государственным органом. То есть это такая форма контроля за алгоритмами, которая напрямую может

воздействовать на общественно-политическую ситуацию в государстве: если платформа использует технологию рекомендаций по персонализированному алгоритму для распространения информации, то она должна создать модель рекомендаций на основе установленных ограничений, запретов {включая недопустимость распространения слухов; принятие мер для предотвращения распространения информации, которая использует преувеличенные названия (при этом содержание серьезно не соответствует названию), ориентирована на раздувание скандалов, размолвок, проступков, содержит ненадлежащее комментирование стихийных бедствий, направлена на продвижение вульгарного или китчевого контента} (ст. 6, 7, 12 Положения об управлении экосистемой онлайн-информационного контента, 2019¹).

Положение об управлении алгоритмическими рекомендациями в информационных службах Интернета (2021) прямо запрещает провайдерам алгоритмических рекомендательных сервисов использовать алгоритмы для регистрации поддельных учетных записей, манипулирования учетными записями пользователей или предоставления ложных реакций («лайков»), комментариев или пересылаемых сообщений; использовать алгоритмы для вмешательства в представление информации (например, путем ее блокирования), создания избыточных рекомендаций, манипулирования списками топового контента или результатами поиска, а также контролировать поисковые запросы или выборки, влиять на общественное мнение в Интернете или уклоняться от надзора и управления (ст. 14)².

Распространение искусственного интеллекта выдвигает проблему формирования технологического ценза на выборах, что является угрозой принципам равенства и справедливости выборов. Финансовые и технологические ограничения порождают равноуровневый доступ к возможностям искусственного интеллекта, которые дают несправедливые преимущества. Вместе с тем возникает возможность создания полностью цифровых кандидатов-аватаров или гиперреалистических копий реальных политиков, которые могут быть «настроены» под запросы любой аудитории. Это стирает грань между реальным человеком с его убеждениями и управляемым программным продуктом, вводя избирателей в заблуждение. В итоге возникает совершенно иной ландшафт политического взаимодействия.

* * *

В конечном счете следует говорить о такой системной и всеобъемлющей конституционной угрозе, как утрата честности, добросовестности на выборах, что может привести к масштабному кризису их политико-правовой

¹ Положение об экологическом управлении информационным контентом в Интернете от 20 декабря 2019 года // Государственная канцелярия интернет-информации КНР. URL: https://www.cac.gov.cn/2019-12/20/c_1578375159509309.htm (дата обращения: 25.11.2025).

² Положение об управлении алгоритмическими рекомендациями в информационных службах Интернета (2021) // China Law Translate. URL: <https://www.chinalawtranslate.com/en/algorithms/> (дата обращения: 25.11.2025).

эпистемологии (когда уже невозможно доверять тому, что видишь и слышишь). Вне зависимости от конкретного типа применяемой избирательной системы выборы неразрывно связаны с человеком как носителем ценностей, морально-политических установок и убеждений. Как следует из решений Конституционного Суда Российской Федерации, правовая демократия, чтобы быть устойчивой, нуждается в эффективных правовых механизмах, способных охранять ее от злоупотреблений публичной властью, легитимность которой во многом основывается на доверии общества. Это предполагает установление повышенных требований к репутации лиц, занимающих публичные должности, во имя того, чтобы у граждан не возникало сомнений в их морально-этических и нравственных качествах³. Между тем при распространении искусственного интеллекта выборы рискуют превратиться в скрытую войну обладателей алгоритмов, где побеждает не тот, кто убедительнее, а тот, кто лучше манипулирует, материально богаче и изощреннее в обходе моральных норм.

Обеспечение легитимности выборов в новых реалиях требует не только технологических решений (например, детекторов дипфейков), но и законодательного контроля, не уповая только на мягкое (этическое) регулирование, чтобы добиться прозрачности алгоритмов, «видимости» (маркировки и архивирования) синтезированного контента, запрета на применение манипулятивных скрытых технологий, а также обеспечения роста цифровой грамотности избирателей. Было бы оправданным ставить вопрос о введении обязательного «аудита» алгоритмов, используемых в избирательном процессе, чтобы любая система на основе искусственного интеллекта, применяемая в ходе выборов, проходила предварительную оценку в том числе на предмет объяснимости (возможности понять логику принятия решений) и отсутствия дискриминационных смещений. С учетом существующего зарубежного опыта (Южная Корея, отдельные штаты США — Калифорния, Техас и др.) требует обсуждения вопрос запрета (и, возможно, криминализации) злонамеренного использования дипфейков на выборах.

Список литературы

1. Lin H., Czarnek G., Lewis B., White J.P., Berinsky A.J., Costello T., Pennycook G., Rand D.G. Persuading voters using human–artificial intelligence dialogues // Nature. 2025. Vol. 648. URL: <https://www.nature.com/articles/s41586-025-09771-9> (дата обращения: 30.03.2025).
2. Hackenburg K., Tappin B.M., Hewitt L., Saunders E., Black S., Lin H., Fist C., Margetts H., Rand D.G., Summerfield C. The levers of political persuasion

³ См. напр.: Постановление Конституционного Суда РФ от 10.10.2013 № 20-П «По делу о проверке конституционности подпункта «а» пункта 3.2 статьи 4 Федерального закона «Об основных гарантиях избирательных прав и права на участие в референдуме граждан Российской Федерации», части первой статьи 10 и части шестой статьи 86 Уголовного кодекса Российской Федерации в связи с жалобами граждан Г.Б. Егорова, А.Л. Казакова, И.Ю. Кравцова, А.В. Куприянова, А.С. Латыпова и В.Ю. Синькова» // СПС КонсультантПлюс.

- with conversational artificial intelligence // *Science*. 2025. Vol. 390, № 6777. URL: <https://www.science.org/doi/10.1126/science.aea3884> (дата обращения: 30.03.2025).
3. Зорькин В.Д. Лекции о праве и государстве. СПб.: Конституционный Суд Российской Федерации, 2024.
 4. Суханов Е.А. О гражданско-правовых проблемах цифровизации // *Вестник гражданского процесса*. 2025. № 1.
 5. Василевская Л.Ю. Деликтная ответственность за причинение вреда искусственным интеллектом: проблемы и перспективы развития // *Гражданское право*. 2025. № 4.
 6. Устинович Е.С. Генеративный искусственный интеллект в избирательных процессах 2024 года в мире: дезинформационные кампании и онлайн-тролли // *Социальная политика и социальное партнерство*. 2024. № 3.
 7. Munker S. Cultural Bias in Large Language Models: Evaluating AI Agents through Moral Questionnaires // URL: <https://arxiv.org/html/2507.10073v16> (дата обращения: 30.11.2025).
 8. Peters U., Carman M. Cultural Bias in Explainable AI Research: A Systematic Analysis // *Journal of Artificial Intelligence Research*, 2024, № 79. URL: <https://arxiv.org/abs/2507.10073> (дата обращения: 30.11.2025).
 9. Tao Y., Viberg O., Baker R.S., Kizilcec R.F. Cultural bias and cultural alignment of large language models. *PNAS Nexus*. 2024. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11407280> (дата обращения: 30.11.2025).