

Научная статья

Research article

УДК 316.62

<https://doi.org/10.26907/2079-5912.2026.3.111-121>

© 2026 г.

ОРИГИНАЛЬНАЯ СТАТЬЯ

Презумпция генерации: нейросетевая паранойя и кризис доверия к политическому контенту

Скипин Н.С.

*Институт научной информации по общественным наукам РАН,
117418, Нахимовский проспект, д. 51/21, г. Москва, Российская Федерация*

Аннотация. Стабильное развитие больших языковых моделей в последние годы привело к увеличению количества контента, в том числе политического, сгенерированного такими инструментами. Обилие такого сгенерированного, или проще говоря – нейросетевого контента привело к тому, что практически любой контент потребитель подозревает в генеративном происхождении. Поэтому автор в статье вводит понятие «нейросетевая паранойя» для описания режима такого подозрения, при котором нейросетевое авторство контента трактуется как объяснение по умолчанию и вытесняет презумпцию человеческого авторства. Предпринимается попытка показать, что данный режим отличается от технофобии, поскольку его фокус связан с установлением происхождения контента и условиями публичной ответственности. В контексте политического дискурса предлагается типология проявлений нейросетевой паранойи, которая включает эвристическую, обвинительно-полемическую, конспирологическую, нормативно-регулятивную и рефлексивно-оборонительную формы. Анализируются риски делегитимации высказываний, размывания ответственности и монополизации инфраструктур, отвечающих за доверие, а также обсуждаются способы их смягчения через процедурную прозрачность и логическое разделение истинности содержания и способа происхождения контента. Работа носит концептуально-аналитический характер и опирается на элементы дискурс-анализа и интерпретацию типичных кейсов цифровой коммуникации.

Ключевые слова: нейросетевая паранойя, нейросети, подозрение, авторство, легитимация, доверие

Для цитирования: Скипин Н.С. Презумпция генерации: нейросетевая паранойя и кризис доверия к политическому контенту. *Казанский социально-гуманитарный вестник*. 2026;3(74):111–121.

Presumption of generation: neural network paranoia and the crisis of trust in political content

Skipin N.S.

*Institute of Scientific Information for Social Sciences RAS,
117418, Nakhimovsky Prospekt, 51/21, Moscow, Russian Federation*

Abstract. The steady development of large language models in recent years has led to an increase in the amount of content, including political content, generated by such tools. The abundance of such generated, or more simply, neural network content has led consumers to suspect virtually any content of generative origin. Therefore, the author introduces the concept of "neural network paranoia" to describe this mode of suspicion, in which neural network authorship of content is treated as the default explanation and displaces the presumption of human authorship. An attempt is made to demonstrate that this mode differs from technophobia, as its focus is on establishing the origin of content and the conditions of public accountability. In the context of political discourse, a typology of neural network paranoia manifestations is proposed, which includes heuristic, accusatory-polemic, conspiratorial, normative-regulatory, and reflexive-defensive forms. The risks of delegitimization of speech, erosion of responsibility, and monopolization of trust infrastructures are analyzed, and ways to mitigate them through procedural transparency and the logical separation of the truth of content from its origin are discussed. This work is conceptual and analytical in nature, drawing on elements of discourse analysis and the interpretation of typical digital communication cases.

Keywords: neural network paranoia, neural networks, suspicion, authorship, legitimation, trust

For citation: Skipin N.S. Presumption of generation: neural network paranoia and the crisis of trust in political content. *The Kazan Socially-Humanitarian Bulletin*. 2026;3(74):111–121. (In Russ.)

Введение

Распространение генеративных нейросетевых инструментов изменило не только объем и скорость производства цифрового контента, но и сами условия его восприятия и легитимации. Современные языковые модели позволяют создавать тексты, имитирующие стилистические и содержательные особенности журналистских материалов. Экспериментальные исследования показывают, что аудитория зачастую не способна уверенно отличить такие тексты от созданных человеком (РБК-1). В частности, искусственно сгенерированные новостные сообщения могут восприниматься как сопоставимые по достоверности с журналистскими пу-

бликациями [1], а суждения об авторстве текста нередко основываются на неточных и потенциально манипулируемых эвристических признаках [2].

Аналогичная проблема возникает при восприятии визуального и аудиовизуального контента. Получатели информации не всегда надежно распознают дипфейки и при этом склонны переоценивать собственную способность к их обнаружению [3]. Поэтому расширение возможностей генерации сопровождается ростом значения процедур фактчекинга, установления первоисточника и проверки обстоятельств создания сообщения. В отечественных исследованиях фактчекинга фрагменты, созданные искусственным интел-

лектом, уже рассматриваются как новый объект верификации, требующий учета не только достоверности отдельных сведений, но и технологической специфики производства контента [4].

Политические последствия этой трансформации не сводятся к распространению заведомо ложных сообщений. Экспериментальные данные показывают, что политические дипфейки способны порождать неопределенность относительно истинности увиденного и тем самым снижать доверие к новостям в социальных медиа, даже если аудитория не принимает ложное сообщение за достоверное [5]. Более того, само знание о существовании технологий генерации может переносить подозрение на подлинные материалы. Анализ дискуссий о дипфейках в контексте российско-украинского конфликта показал, что пользователи обозначали как дипфейки реальные видеозаписи, при этом использовали указание на нейросетевое происхождение контента в качестве универсального аргумента дискредитации и включали его в конспирологические объяснения событий [6]. Таким образом, политически значимым становится не только риск принятия искусственного контента за подлинный, но и обратный риск – отказ признавать подлинным контент, созданный человеком.

Существующие исследования рассматривают преимущественно технологии генерации и обнаружения дипфейков, их способность вводить аудиторию в заблуждение, влияние сгенерированных материалов на доверие, а также возможные регулятивные ответы [7]. Вместе с тем недостаточно концептуализирована ситуация, при которой гипотеза о нейросетевом происхождении контента превращается в объяснение по умолчанию и начинает вытеснять анализ содержания сообщения. В этом случае происхождение контента становится самостоятельным предметом политической

интерпретации, а подозрение в использовании нейросети – средством делегитимации автора, института или публичного высказывания.

Для обозначения такого когнитивно-дискурсивного режима в настоящей статье предлагается понятие «нейросетевая паранойя». Цель исследования заключается в его концептуализации и выявлении основных форм проявления в политическом дискурсе.

Материалы и методы

Работа носит концептуально-аналитический характер и сочетает методы концептуального анализа и качественного дискурс-анализа. Материал исследования составили пять публично зафиксированных кейсов 2023–2026 годов, связанных с использованием генеративных нейросетей в политической коммуникации либо с публичным приписыванием политическому контенту нейросетевого происхождения. Кейсы отбирались целенаправленно по следующим критериям: политическая релевантность сообщения; наличие явного указания на предполагаемое или подтвержденное участие нейросетевых инструментов; возможность установить источник и контекст публикации; различие выполняемых подозрением дискурсивных функций. Единицей анализа выступало публичное высказывание о происхождении контента в совокупности с сопровождающим его медийным описанием. Кейсы сопоставлялись по четырем параметрам: объект подозрения, форма приписывания нейросетевого происхождения, способ аргументации и политическая функция высказывания. Выборка имеет иллюстративный, а не статистически репрезентативный характер: ее назначение состоит в построении концептуальной типологии, требующей последующей проверки на расширенном эмпирическом материале.

Прежде необходимо провести методологическое разграничение, т.к. в рамках используемого политико-психологического подхода паранойя рассматривается не как клинический диагноз или форма патологии – автор не обладает необходимой врачебной компетенцией для этого. В данном исследовании «паранойя» является устойчивой формой интерпретации социальной реальности, которая основана на повышенном подозрении, недоверии. При этом необходимо провести еще одну границу – между паранойей и фобией в контексте данного подхода. Так, фобия (опять же в политико-психологическом подходе) описывает преимущественно аффективную реакцию на объект, который воспринимается как опасный, а структура фобии обычно ориентирована на избегание или на защитное дистанцирование (или неучастие [8]). В случае же паранойи первичным является не столько страх как эмоция, сколько интерпретационная установка: стремление реконструировать скрытое происхождение, намерение и субъект действия. Так, емко паранойю можно определить как «всеобъемлющее и необоснованное недоверие к другим» (a pervasive and unwarranted mistrust of others) [9]. Если фобия фиксируется на объекте угрозы и заставляет минимизировать контакт с ним, то паранойя фиксируется на объяснении и стремится развернуть картину скрытых причин, связей и авторства.

Требуется также учитывать эпистемический аспект – фобия может существовать без развитой доказательной схемы, не иметь рациональной (или квазирациональной) системы объяснения страха, тогда как паранойя, как правило, воспроизводит именно схему аргументации, где подозрение становится основанием интерпретации новых случаев. Для нее характерны, во-первых, расширительная атрибуция (в смысле установления скрытого

автора) или «глубинной» причинности (тенденция объяснять неоднозначные явления действием внешнего агента). Во-вторых, приписывание намеренности там, где возможны альтернативные объяснения (случайность, институциональная аффилиация авторов контента, жанровая стандартизация и пр.). В-третьих, самоподдерживаемость аргументации – подозрение часто легче сформулировать, чем опровергнуть. В дискурсивном измерении паранойя проявляется как особый стиль объяснения. В нем критерием убедительности является не верифицируемость утверждения, а его согласованность с исходной установкой на подозрение. В таком виде паранойю можно охарактеризовать как устойчивую логическую конфигурацию, через которую интерпретируются контент, события и действия акторов.

Следует также заметить, что в данном исследовании нейросетевая паранойя не сводится к техно-паранойе и ИИ-паранойе. Техно-паранойя в общем виде фокусируется на технологии как на факторе скрытого контроля, принуждения или подмены человеческой автономии, то есть на технологии как на источнике политически значимого воздействия. ИИ-паранойя, в свою очередь, обычно охватывает более широкий спектр недоверия и страха, которые связаны с искусственным интеллектом как возможной автономной силой, способной замещать человека, управлять поведением, принимать решения или формировать повестку.

Нейросетевая паранойя имеет иной предметный фокус в сравнении с техно- и ИИ-паранойей. Ее объектом является нейросетевое авторство контента как основной аргумент объяснения происхождения контента. Здесь ключевой аспект формулируется как восприятие источника контента – кто выступает автором, можно ли надежно установить происхождение и каковы последствия этого для доверия к пу-

бличной коммуникации. По этой причине нейросетевая паранойя может воспроизводиться и при отсутствии общей технофобной установки. Наблюдатель может не разделять негативной оценки нейросетей как инструмента и, тем не менее, систематически приписывать нейросетевое происхождение значительной части контента.

Поэтому под «нейросетевой паранойей» далее предлагается понимать когнитивно-дискурсивный режим подозрения, в рамках которого нейросетевое происхождение контента трактуется как единственное возможное, что вытесняет презумпцию традиционной (естественной) формы создания контента. При этом стилевые признаки превращаются в индикаторы авторства, а опровержение подозрения затрудняется отсутствием данных о производстве контента или недостаточным доверием к источнику. Нейросетевая паранойя фиксирует современную ситуацию в понимании происхождения контента, в которой авторство и момент создания становятся базовой проблемой публичной коммуникации и ключевым условием легитимации политического высказывания.

Результаты

После концептуализации понятия приведем возможную типологию проявлений нейросетевой паранойи в политическом дискурсе на основании конкретных кейсов. Для этого целесообразно опереться на четыре признака: что именно становится объектом подозрения, каким образом формулируется само приписывание нейросетевого происхождения контента, на чем строится аргументация и какую политическую роль выполняет такое высказывание. В первом случае подозрение может быть направлено на отдельный текст, канал коммуникации или социальный институт, производящий сообщения. Во втором – оно может иметь характер предположения,

обвинения или нормативного суждения. В третьем случае значение имеют способы обоснования, такие как ссылка на стилевые признаки, апелляция к процедурным подтверждениям или конспирологическое достраивание «недостающих звеньев». Наконец, в политическом плане подобные высказывания могут использоваться для делегитимации, мобилизации, оправдания регулирования или переноса ответственности. Исходя из этого, далее можно выделить несколько основных типов проявления нейросетевой паранойи в политическом дискурсе.

Первый тип – **эвристическая атрибуция нейросетевого авторства**. В данном случае подозрение является быстрым и наиболее простым способом классификации контента. Нейросетевое происхождение предполагается преимущественно по стилевым и жанровым признакам: стандартизированная композиция, повторяемые связки, отсутствие индивидуализирующих маркеров, риторическая «ровность». Для этого типа характерно, что атрибуция не обязательно сопровождается тезисом о политическом умысле, или «автор» нейросетевого контента сам заявляет, что использовал инструменты ИИ при его создании (BG). В политическом дискурсе данный тип наиболее заметен в социальных сетях, в комментариях к новостям, а также в медиаобсуждениях, где подозрение в генерации выступает заменой содержательной критики стиля и жанровой стандартизации. Политическое следствие здесь состоит в постепенной нормализации сомнения как базовой установки по отношению к контенту, то есть в расширении реакции недоверия без прямого конфликта вокруг конкретного актора.

Второй тип – **обвинительно-полемическая атрибуция**. Здесь нейросетевое авторство используется как инструмент делегитимации и дискредитации субъекта высказывания

(РБК-2). Атрибуция нейросетевого происхождения приобретает характер обвинения, поскольку связывается с предположением о несамостоятельности, отсутствии компетенции, имитации реальной работы по созданию контента или уходе от ответственности. Для данного типа характерно смещение центра критики, т.к. обсуждение аргументов заменяется указанием на якобы недопустимое происхождение контента. В политическом дискурсе этот тип проявляется как в социальных сетях, так и в других сетевых форматах, где полемика строится вокруг авторства и «подлинности» речи политического оппонента, а также в экспертных комментариях, когда критика аргумента подменяется критикой происхождения контента как такового. Политический эффект состоит в том, что нейросетевое подозрение становится самостоятельным приемом борьбы, формируя новую стратегию делегитимации контента, не требующую развернутого опровержения содержания.

Третий тип – **конспирологическое объяснение и достраивание «неопределенного»**. В данном случае нейросетевое авторство не является конечной точкой объяснения, а используется как промежуточное звено для вывода о скрытом «заказчике», координации и внешнем управлении (Телеграф). Логика аргументации в данном случае может быть такой, что если контент сгенерирован, значит, есть субъект, задающий параметры и сценарии, следовательно, публичная речь является производной от «скрытого центра» или актора. Для этого типа характерны обобщения, перенос подозрения с конкретного контента на каналы и социальные институты, а также включение нейросетевой генерации в более широкие нарративы о манипуляции и управляемости публичной сферы. В политическом дискурсе конспирологический тип чаще проявляется в сегментах медиаиндустрии и социаль-

ных сетей, однако он может проникать и в экспертные обсуждения в форме гипотез о «скриптах» и «единых методичках», которые теперь получают технологическое обоснование через ссылку на генеративные инструменты. Политическое следствие заключается в усилении недоверия [10] к институтам публичной коммуникации и в сдвиге к интерпретациям, где авторство и субъектность систематически подменяются режиссированием политических процессов и высказываний.

Четвертый тип – **нормативно-регулятивная артикуляция нейросетевой паранойи**. Здесь презумпция генерации перерабатывается в язык норм и правил, то есть приобретает институциональную форму. Объектом становятся не отдельный контент, а способы обращения с ним: маркировка, ответственность за публикацию, требования к подтверждению источника, допустимость нейросетевой обработки в официальной и экспертной коммуникации (Коммерсантъ). В отличие от обвинительно-полемического типа, нормативно-регулятивный не обязательно носит конфронтационный характер по отношению к конкретному политическому субъекту. Он ориентирован на нормализацию и на установление процедур, которые должны снизить неопределенность авторства. В политическом дискурсе этот тип проявляется в публичных позициях политиков и государственных акторов, в обсуждении правил для медиа-платформ, а также в экспертных комментариях, где нейросетевое подозрение переводится в плоскость институциональных решений. Политический эффект выражается в усилении процедурного измерения легитимности – подлинность и авторство начинают рассматриваться как предмет регулирования.

Пятый тип – **рефлексивно-оборонительная артикуляция**. В данном случае акторы публичной политики и эксперты учитывают презумпцию ге-

нерации и заранее встраивают в коммуникацию элементы, призванные нейтрализовать подозрение в нейросетевом авторстве. Это проявляется в подчеркнутой демонстрации процесса производства контента, в акцентировании человеческого участия как основания доверия, как в кейсе с поздравлением Президента РФ, когда «черновой» необработанный материал был выложен, вероятно, чтобы показать реальное происхождение данного контента (РИА). В политическом дискурсе данный тип возможен в заявлениях политиков, в публичном описании процедур подготовки и публикации, в демонстрации «живых» форматов. В экспертных комментариях он проявляется как предварительное оговаривание статуса контента и условий его производства. Политическое значение этого типа состоит в том, что нейросетевая паранойя начинает влиять на стратегию создания публичного высказывания. В таком случае актер вынужден не только производить контент, но и предъявлять условия, при которых этот контент должен считаться «подлинным».

Такая типология позволяет зафиксировать, что нейросетевая паранойя в политическом дискурсе не является однородной. При этом один и тот же эпизод может включать элементы нескольких типов.

Обсуждение

Рассмотрение способов нивелирования существующих рисков, связанных с нейросетевой паранойей, требует создания условий ограничения подозрения и снижения его деструктивных последствий для публичной коммуникации, институциональной легитимности и распределения ответственности. В этом смысле корректнее говорить о «смягчении» рисков как о стабилизации доверия и подлинности, а не как о «борьбе» с мнениями или эмоциями аудитории контента.

На уровне дискурсивных практик центральным источником риска является превращение презумпции нейросетевой генерации в универсальную схему интерпретации происхождения контента, вытесняющую анализ его содержания. В таких условиях вопрос происхождения становится самоценным аргументом, а стилевые признаки начинают функционировать как доказательство нейросетевого происхождения контента. Смягчение этого риска связано с нормализацией различий, которые в публичной дискуссии часто оказываются стертыми. Прежде всего требуется объяснительная работа по разделению способа происхождения контента и его истинности содержания. Нейросетевое авторство не является логическим основанием для вывода о недостоверности, так же как человеческое авторство не гарантирует истинности. Когда это различие удерживается, подозрение перестает автоматически превращаться в делегитимирующий аргумент и возвращает критику в плоскость смыслов содержания, аргументации и последствий, а не в плоскость стилистического распознавания.

В рамках политического дискурса значимое место занимает конспирологическое достраивание. Смягчение этого риска связано с требованием разрыва автоматической логической цепочки между генерацией и координацией. Для политологического анализа здесь существенную роль может сыграть методологическая норма, согласно которой факт использования инструмента не является достаточным доказательством централизованного контроля. Это не отменяет возможности политического заказа, но требует отдельной доказательной работы, которую нельзя заменить ссылкой на стилистические признаки или на саму возможность генерации.

Критическим элементом процедурного уровня является понятная

демаркация форм участия нейросетевых инструментов в производстве контента. Нейросети применяются для генерации, редактирования, перевода и стилистической обработки. Именно эта неоднородность практик подпитывает недоверие к контенту, который имеет нейросетевые следы производства. Если авторы не разделяют способы использования инструмента и не артикулируют их, то аудитория получает стимул интерпретировать любой стандартизированный контент как полностью сгенерированный. Проговаривание, как именно использовалась нейросеть – создает условия для снижения подозрения и уменьшает вероятность того, что любой случай обработки будет восприниматься как подмена авторства. При этом речь идет не о запрете использования инструмента. Необходимо фиксировать границы ответственности, четко проговаривая, кто отвечает за смысловое содержание и за окончательную редакцию, как устроен процесс подготовки, каковы правила публикации и т.д.

Смягчение рисков необходимо рассматривать и применительно к экспертному полю, поскольку здесь нейросетевая паранойя может подрывать саму структуру публичной аргументации. Если подозрение в нейросетевом происхождении превращается в универсальный критерий оценки, экспертность начинает определяться не воспроизводимостью аргумента и не качеством источниковой базы, а возможностью предъявить символическое подтверждение «человеческого участия». Более устойчивой альтернативой является усиление тех характеристик экспертного контента, которые делают его верифицируемым, проверяемым независимо от предположений об авторстве (ссылки на свежие новости, первоисточники и пр.). В таком случае даже при наличии подозрения критика возвращается к проверяемо-

сти и к содержательным основаниям, а не к стилистической идентификации.

Нормативно-регулятивная артикуляция нейросетевой паранойи формирует запрос на маркировку контента, но сама маркировка способна либо снижать неопределенность, либо усиливать стигматизацию и тем самым подпитывать подозрение. Поэтому смягчающий эффект маркировки возможен только при нейтральности формулировок и при классификации способов использования нейросетевых инструментов. В противном случае любая маркировка будет прочитываться как признание не-аутентичности, что усиливает делегитимирующую силу подозрения. Требуется учитывать тот факт, что маркировка не должна интерпретироваться как механизм оценки содержания, оставаясь в рамках ответственности автора. Если маркировка начинает выполнять функцию санкции – она становится инструментом политической борьбы и усиливает конфликты вокруг авторства.

Заключение

В условиях неопределенности происхождения контента перестает быть фоновой характеристикой и превращается в самостоятельный критерий его легитимности. Вопрос о том, кто или что создало высказывание, начинает предшествовать вопросу о его содержании. Возникает принципиальная асимметрия: подозрение в нейросетевом происхождении формулируется легко, тогда как его опровержение требует доступа к процессу создания контента и зачастую не может быть безусловным. Тем самым бремя доказательства постепенно переносится на автора, который должен подтверждать не только обоснованность своей позиции, но и сам факт человеческого участия в ее формулировании. Подобный сдвиг делает делегитимацию публичного высказывания сравнительно де-

шевым и многократно воспроизводимым приемом политической борьбы.

Можно предположить, что данная проблема не может быть решена исключительно совершенствованием средств автоматического распознавания. Даже технически надежный инструмент не устраняет интерпретационную установку на подозрение, поскольку она обращена не только к фактическому происхождению контента, но и к предполагаемым намерениям, связям и формам ответственности его создателя. Более того, инфраструктуры удостоверения происхождения сами способны становиться источником политического неравенства. Субъекты, обладающие доступом к сертифицированным платформам и средствам фиксации процесса производства, могут получать преимущество в подтверждении подлинности, тогда как отсутствие подобного подтверждения может начать восприниматься как свидетельство недостоверности. В результате техническая процедура рискует превратиться в механизм распределения права на публичное доверие.

Поэтому перспективная модель регулирования должна исходить не из бинарного разделения контента на «человеческий» и «нейросетевой», а из признания множественности форм участия искусственного интеллекта. Генерация, редактирование, перевод, обработка изображения и техническая помощь предполагают различную степень вмешательства и не могут обозначаться одной универсальной мар-

кировкой. Такая маркировка должна сообщать о способе производства, но не выполнять функцию оценки истинности и не освобождать конкретного субъекта от ответственности за опубликованный материал. Одновременно процедуры подтверждения происхождения должны оставаться оспоримыми, нецентрализованными и доступными различным участникам публичной коммуникации.

Устойчивый ответ на нейросетевую паранойю предполагает последовательное разделение трех вопросов: «истинно ли содержание», «каким способом создан контент» и «кто несет ответственность за его публикацию». Человеческое происхождение само по себе не гарантирует достоверности, а использование нейросетевого инструмента не отменяет возможности содержательно обоснованного высказывания. Поэтому восстановление прежней определенности авторства представляется малореалистичной задачей. Более продуктивной целью становится создание таких процедур коммуникации, при которых неизбежная неопределенность происхождения не разрушает возможность проверки, аргументированного несогласия и установления ответственности. В этом смысле нейросетевая паранойя выступает индикатором перехода от доверия, основанного на предполагаемой идентичности автора, к доверию, основанному на прозрачности и проверяемости процедур.

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.
Conflicts of Interest. The author declares no conflict of interests.

Источники

РБК-1 – Почти 80% участников исследования не отличили ИИ-контент от реального. *РБК*. 13 февраля 2026 г. URL: <https://www.rbc.ru/rbcfreenews/698e2cb79a7947e408795241> (дата обращения: 28.02.2026)

BG – Massachusetts lawmaker asked ChatGPT to write his floor speech. *The Boston Globe*. 25 января 2023 г. URL: <https://www.bostonglobe.com/2023/01/25/nation/massachusetts-lawmaker-asked-chatgpt-write-his-floor-speech/> (дата обращения: 28.02.2026)

РБК-2 – В Госдуме опровергли использование ИИ при написании законопроекта. *РБК*. 24 января 2026 г. URL: <https://www.rbc.ru/politics/24/01/2026/6974a9969a7947c33526548f> (дата обращения: 28.02.2026)

Телеграф – На новогоднем поздравлении был не настоящий Путин? Ис-

кусственный разум разобрал видео и дал ответ. *Телеграф*. 1 января 2026 г. URL: <https://news.telegraf.com.ua/novosti-rossii/2026-01-01/5929514-na-novorichnomu-privitanni-buv-ne-spravzhniy-putin-shtuchniy-rozum-rozibrav-video-ta-dav-vidpovid> (дата обращения: 28.02.2026)

Коммерсантъ – ИИ-контент ждет маркировка: как ГД предлагает защищать пользователей от фейков. *Коммерсантъ*. 15 ноября 2025 г. URL: <https://www.kommersant.ru/doc/8209716> (дата обращения: 28.02.2026)

РИА – Неотредактированное обращение Путина к 8 Марта опубликовали по ошибке. *РИА Новости*. 9 марта 2026 г. URL: <https://ria.ru/20260309/kreml-2079525765.html> (дата обращения: 09.03.2026)

Список литературы / References

1. Kreps S, McCain RM, Brundage M. All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science*. 2022;9(1):104-117. DOI: 10.1017/XPS.2020.37

2. Jakesch M., Hancock J.T., Naaman M. Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences*. 2023;120(11). DOI: 10.1073/pnas.2208839120

3. Köbis N.C., Doležalová B., Soraperra I. Fooled twice: People cannot detect deepfakes but think they can. *iScience*. 2021;24(11):103364. DOI: 10.1016/j.isci.2021.103364

4. Манскова Е.А. Фактчекинг научно-популярного контента: типы недостоверной информации и проблемы верификации. *Медиаскоп*. 2024;4.

Manskova E.A. Fact-checking in popular science media discourse: types of unreliable information and verification difficulties. *Mediascope*. 2024;4. (In Russ.)

5. Vaccari C., Chadwick A. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception,

uncertainty, and trust in news. *Social Media + Society*. 2020;6(1):1–13. DOI: 10.1177/2056305120903408

6. Twomey J., Ching D., Aylett M.P., Quayle M., Linehan C., Murphy G. Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLOS ONE*. 2023;18(10). DOI: 10.1371/journal.pone.0291668

7. Birrer A., Just N. What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media & Society*. 2025;27(12):6819–6838. DOI: 10.1177/14614448241253138

8. Соколов А.В., Бабаджанян П.А. Цифровое неучастие студенческой молодежи. *PolitBook*. 2025;3:76–89. DOI: 10.24412/2227-1538-2025-3-76-89

Sokolov A.V., Babadzhanyan P.A. Digital non-participation of students. *PolitBook*. 2025;3:76–89. DOI: 10.24412/2227-1538-2025-3-76-89 (In Russ.)

9. Liu W. Pandemic paranoia: Toward a reparative practice of the global psyche. *Psychoanalysis, Culture & Society*.

2021;26(4):608–622. DOI: 10.1057/s41282-021-00236-2

10. Евплов Н.А. Оценка методов обнаружения дипфейков в видеоконтенте. *Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки*. 2025;10-2:17–22. DOI: 10.37882/2223-2966.2025.10-2.04

Evplov N.A. Evaluation of methods for deepfake detection in video content. *Modern Science: Actual Problems of Theory and Practice. Series: Natural and Technical Sciences*. 2025;10-2:17–22. DOI: 10.37882/2223-2966.2025.10-2.04 (In Russ.)

Информация об авторе

Скипин Николай Сергеевич, кандидат политических наук, старший научный сотрудник, заведующий отделом, Институт научной информации по общественным наукам Российской академии наук. Scopus Author ID: 57299150000, ORCID: 0000-0003-4267-0099, SPIN-код: 3019-6197; e-mail: skipin@inion.ru

Information about the author

Nikolai S. Skipin, Cand. Sci. (Political Sciences), Senior Researcher, Head of Department, Institute of scientific information for social sciences of Russian Academy of Sciences. Scopus Author ID: 57299150000, ORCID: 0000-0003-4267-0099, SPIN-code: 3019-6197; e-mail: skipin@inion.ru

Поступила в редакцию 09.06.2026
Принята после рецензирования 25.07.2026
Принята к публикации 01.09.2026

Received June 9, 2026
Revised July 25, 2026
Accepted September 1, 2026