



**Сергей Васильевич
СОЛОВКИН,**

старший преподаватель
кафедры
информационного права
и цифровых технологий,
научный сотрудник НОЦ
правового регулирования
в сфере высоких
технологий НИИ
Университета имени
О.Е. Кутафина (МГЮА),
кандидат юридических наук
svsolovkin@msal.ru
125993, Россия, г. Москва,
ул. Садовая-Кудринская, д. 9

Генеративный искусственный интеллект как инструмент манипуляции: правовые вызовы и пределы регулирования¹

Аннотация. Статья посвящена правовому анализу манипулятивного воздействия на сознание и поведение человека, осуществляемого посредством систем генеративного искусственного интеллекта. На основе психологической и правовой доктрины сформулировано авторское определение манипулятивного воздействия с использованием систем ИИ как самостоятельной правовой категории, охватывающей политическую, коммерческую, межличностную и психоэмоциональную сферы. Выявлены ключевые технологические механизмы такого воздействия (большие языковые модели, технологии синтеза аудиовизуального контента (дипфейки), персонализированные рекомендательные алгоритмы) и приведены случаи их манипулятивного применения. Сформулированы предложения по совершенствованию российского законодательства, включающие введение легального определения манипулятивного воздействия с использованием систем ИИ, прямой запрет на разработку и использование систем ИИ с манипулятивным функционалом, а также институциональные механизмы контроля.

Ключевые слова: цифровая трансформация, информационная безопасность, информационное право, цифровое право, право искусственного интеллекта, персональные данные, искусственный интеллект, генеративный искусственный интеллект, манипуляция сознанием и поведением, угрозы искусственного интеллекта

DOI: 10.17803/2311-5998.2026.142.6.040-046

Sergey V. SOLOVKIN,

Senior Lecturer of the Chair of Information Law and Digital Technologies,
Research Fellow of the Research and Education Centre for Legal Regulation in the
Field of High Technologies of the Research Institute,
Kutafin Moscow State Law University (MSAL), Cand. Sci. (Law)
svsolovkin@msal.ru
9, ul. Sadovaya-Kudrinskaya, Moscow, Russia, 125993

Generative Artificial Intelligence as a Tool of Manipulation: Legal Challenges and the Limits of Regulation

Abstract. This article presents a legal analysis of manipulative influence on human consciousness and behavior carried out by means of generative artificial intelligence (GAI) systems. Drawing on psychological and legal

¹ Исследование выполнено в рамках программы стратегического академического лидерства «Приоритет-2030».

doctrine, the author formulates an original definition of AI-assisted manipulative influence as an independent legal category encompassing the political, commercial, interpersonal and psycho-emotional spheres. The article identifies the principal technological mechanisms of such influence and cases of their manipulative deployment. The article puts forward proposals for the reform of Russian law, including the introduction of a statutory definition of AI-assisted manipulative influence, a direct prohibition on the development and use of AI systems with manipulative functionality, and institutional oversight mechanisms.

Keywords: digital transformation, information security, information law, digital law, artificial intelligence law, personal data, artificial intelligence, generative artificial intelligence, manipulation of consciousness and behavior, artificial intelligence threats

Понятие манипулятивного воздействия с использованием систем генеративного искусственного интеллекта

Стремительное распространение систем генеративного искусственного интеллекта (ГИИ) в последние годы фундаментально изменило информационную среду, в которой существует современный человек. Одним из наиболее острых вызовов, обусловленных применением ГИИ, является их манипулятивный потенциал².

Разработка надлежащего правового инструментария противодействия манипулятивным практикам с применением систем ГИИ предполагает прежде всего последовательное доктринальное конструирование соответствующей категории. Что следует понимать под манипулятивным воздействием применительно к системам ГИИ?

В психологической науке понятие манипуляции разработано достаточно детально. Е. Л. Доценко, чья концепция по праву признается классической в отечественной психологии, определяет манипуляцию как «вид психологического воздействия, искусное исполнение которого ведет к скрытому возбуждению у другого человека намерений, не совпадающих с его актуально существующими желаниями»³. Данное определение фиксирует три сущностных признака манипуляции: во-первых, скрытый (латентный) характер воздействия; во-вторых, направленность на изменение внутренних психических состояний адресата; в-третьих, рассогласование между вносимыми установками и исходными намерениями адресата, т.е. воздействие вопреки его подлинной воле.

Рецепция психологической категории манипуляции в правовую науку сопряжена с рядом концептуальных трудностей. В отличие от таких явлений, как обман (введение в заблуждение относительно фактических обстоятельств) или

² Минбалева А. В. Риски и угрозы агентского искусственного интеллекта (agentic AI) // Информационное право. 2025. № 3 (85). С. 10—13.

³ Психология манипуляции: феномены, механизмы и защита. М. : ЧеРо ; Изд-во МГУ, 1997.



принуждение (прямое давление на волю субъекта), манипуляция не предполагает ни ложного утверждения в строгом смысле, ни применения силы или угрозы. Она действует через формально достоверную или нейтральную информацию, избирательно конструируя информационный контекст таким образом, чтобы на править поведение адресата в заданном направлении при сохранении иллюзии его свободного выбора. Именно эта особенность делает манипуляцию наиболее труднодоступной для правового регулирования и одновременно наиболее опасной с точки зрения прав личности.

Применительно к системам ГИИ указанные трудности многократно возрастают. Как отмечает А. А. Симонян, анализируя угрозы применения систем ГИИ, «опасность такой технологии заключается в психологическом давлении на различные группы лиц с целью манипуляции общественным мнением граждан, шантажа, мошенничества или клеветы»⁴. Вместе с тем представляется, что данный перечень не является исчерпывающим: манипулятивный потенциал ГИИ не сводится к конкретным правонарушениям, но проявляется на более глубоком уровне — уровне систематического подавления когнитивной автономии личности.

В этой связи в настоящей статье предлагается следующая авторская дефиниция: манипулятивное воздействие с использованием систем искусственного интеллекта — это скрытое автоматизированное воздействие, осуществляемое системой ИИ либо лицом посредством системы ИИ, на сознание, эмоциональное состояние и (или) поведение физического лица путем формирования, персонализации и (или) распространения информационного контента с намерением изменить убеждения, решения или действия адресата вопреки его подлинным интересам и (или) без его осознанного согласия на такое воздействие, при этом независимо от того, является ли содержание распространяемой информации ложным.

Технологические механизмы манипулятивного воздействия и актуальные угрозы

Представляется необходимым установить, посредством каких именно инструментов современные системы ГИИ реализуют свой манипулятивный потенциал, и оценить, являются ли соответствующие угрозы лишь теоретическими конструкциями или же они уже нашли документальное подтверждение в общественной практике.

Большие языковые модели представляют собой, пожалуй, наиболее многофункциональный инструмент манипулятивного воздействия. Способность этих систем генерировать связный, стилистически адаптированный текст в режиме реального времени открывает возможность для создания персонализированных убеждающих нарративов в масштабе, ранее технически недостижимом⁵. При

⁴ Симонян А. А. Генеративный искусственный интеллект и DeepFake-технологии как угроза информационной безопасности граждан // Вопросы российского и международного права. 2025. Т. 15. № 2А. С. 166.

⁵ Минбалева А. В. Особенности правовой природы агентского искусственного интеллекта // Аграрное и земельное право. 2025. № 8. С. 75—78.

наличии данных о психологическом профиле конкретного пользователя — его политических взглядах, страхах, ценностях, поведенческих паттернах — система способна автоматически конструировать сообщения, точно эксплуатирующие индивидуальные когнитивные уязвимости. Исследователи по изучению человеко-компьютерного взаимодействия зафиксировали, что гиперперсонализированное воздействие систем ГИИ в режиме реального времени способно формировать устойчивые паттерны компульсивного потребления контента, подрывая когнитивную автономию пользователей, причем в особой зоне риска находятся несовершеннолетние и пожилые люди⁶.

Принципиально важным в контексте правового регулирования является то обстоятельство, что ряд исследователей квалифицирует персонализированные манипулятивные интерфейсы, генерируемые системами ГИИ, как так называемые теневые паттерны (dark patterns) нового поколения, т.е. намеренно вводящие в заблуждение либо психологически эксплуататорские элементы пользовательского взаимодействия⁷. При этом, как установлено в ходе многоуровневого бенчмаркингowego исследования 2025 г., системы ГИИ способны к таким формам психологического вреда, как антропоморфизация ИИ с целью формирования у пользователя эмоциональной зависимости, газлайтинг относительно фактических обстоятельств, а также подрыв самовосприятия и оценки реальности⁸. Следовательно, манипулятивный потенциал больших языковых моделей не ограничивается распространением дезинформации, но проявляется на уровне систематической деструкции когнитивных способностей личности.

Технологии синтеза аудиовизуального контента (диффузионные модели, GAN-архитектуры, модели синтеза речи) — то, что в публичном дискурсе принято обозначать термином «дипфейк-технологии», — образуют второй ключевой вектор манипулятивного воздействия. Возможности этих технологий состоят в создании синтетических аудио- и видеоматериалов, воспроизводящих облик, голос и речевые паттерны реально существующих лиц с качеством, практически неотличимым от подлинных записей. Европейский парламент в аналитическом докладе 2025 г. констатирует, что диффузионные модели значительно упростили создание дипфейков, включая материалы, направленные против несовершеннолетних, что порождает качественно новые угрозы достоинству и безопасности личности⁹.

⁶ *Khraishi R.* [et al.] Real-Time Hyper-Personalized Generative AI Should Be Regulated to Prevent the Rise of “Digital Heroin” // NeurIPS 2025 Position Paper. URL: <https://neurips.cc/virtual/2025/loc/san-diego/poster/121952> (дата обращения: 25.03.2026).

⁷ *Guerrini F.* AI-Driven Dark Patterns: How Artificial Intelligence Is Supercharging Digital Manipulation // Forbes. 17.11.2024. URL: <https://www.forbes.com/sites/federicoguerrini/2024/11/17/ai-driven-dark-patterns-how-artificial-intelligence-is-supercharging-digital-manipulation/> (дата обращения: 25.03.2026).

⁸ *Asif A. et. al.* DarkPatterns-LLM: A Multi-Layer Benchmark for Detecting Manipulative and Harmful AI Behavior. arXiv. 2025 // URL: <https://arxiv.org/pdf/2512.22470.pdf> (дата обращения: 25.03.2026).

⁹ Children and Deepfakes. Briefing. PE 775.855. European Parliament Research Service. March 2025 // URL: [https://www.europarl.europa.eu/regdata/etudes/brie/2025/775855/eprs_bri\(2025\)775855_en.pdf](https://www.europarl.europa.eu/regdata/etudes/brie/2025/775855/eprs_bri(2025)775855_en.pdf) (дата обращения: 25.03.2026).



Документально подтвержденные случаи применения указанных технологий в манипулятивных целях уже исчисляются десятками¹⁰. При этом необходимо особо подчеркнуть, что манипулятивное применение систем ГИИ отнюдь не ограничивается политической сферой.

Рекомендательные алгоритмы и системы микротаргетинга, интегрированные с возможностями ГИИ, образуют третий вектор манипулятивного воздействия, отличающийся наибольшей латентностью. В отличие от дипфейков, создание которых требует целенаправленных усилий, алгоритмическая манипуляция осуществляется в фоновом режиме: система автоматически формирует информационную среду конкретного пользователя таким образом, чтобы максимизировать его вовлеченность или направить его поведение в заданную сторону, при этом сам пользователь, как правило, не осознает факта воздействия. Генеративный ИИ многократно повышает эффективность подобных систем, поскольку позволяет создавать не просто персонализированную, но индивидуально сконструированную информационную реальность для каждого отдельного субъекта¹¹.

Рекомендации по совершенствованию российского законодательства

Действующее российское законодательство не обеспечивает надлежащей правовой защиты физических лиц от манипулятивного воздействия с использованием систем ГИИ.

Необходимым основанием для построения системного регулирования является введение в российское законодательство легального определения манипулятивного воздействия с использованием систем ИИ. По мнению автора, данное определение надлежит закрепить в Федеральном законе от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации» в качестве одного из ключевых понятий информационного права, а также отразить в будущих законах в сфере искусственного интеллекта.

Наряду с этим представляется необходимым закрепить понятие «системы ИИ с манипулятивным функционалом», т.е. системы, конструктивные особенности, обучающие данные и (или) алгоритмы которой направлены или систематически приводят к оказанию манипулятивного воздействия в указанном выше смысле. Данная категория необходима для разграничения систем, создающих риск манипуляции в силу своего прямого назначения, и систем, способных быть использованными в манипулятивных целях лишь при определенных условиях

¹⁰ См., например: ВЭФ назвал дипфейки и ИИ главными угрозами для выборов в разных странах // РБК. 11.01.2024. URL: https://www.rbc.ru/technology_and_media/11/01/2024/659fe9ff9a79475f70435f09 (дата обращения: 25.03.2026) ; A fake recording of a candidate saying he'd rigged the election went viral. Experts say it's only the beginning // CNN. 01.02.2024. URL: <https://edition.cnn.com/2024/02/01/politics/election-deepfake-threats-invs> (дата обращения: 25.03.2026) ; Case study: Can we believe what we hear? // URL: <https://www.unesco.org/mil4teachers/en/toolkit-media/indicator-5/activities/ai-disinformation/case-study-2> (дата обращения: 25.03.2026).

¹¹ Asif A. [et. al.] Op. cit.

применения, поскольку правовые последствия для этих двух категорий должны принципиально различаться.

Центральным элементом предлагаемой системы регулирования является нормативный запрет, содержательно аналогичный ст. 5 Регламента ЕС об ИИ, однако адаптированный к российской правовой системе и расширенный с учетом выявленных в настоящей статье пробелов европейского подхода. Предлагается установить следующую систему запретов:

- 1) запрет на разработку и введение в гражданский оборот систем ИИ с манипулятивным функционалом. Лица, осуществляющие разработку систем ИИ на территории Российской Федерации либо предназначенных для использования на территории Российской Федерации, обязаны обеспечить отсутствие в системе функциональных элементов, намеренно направленных на осуществление манипулятивного воздействия в установленном законом смысле. Данный запрет должен подкрепляться обязательной оценкой соответствия системы ИИ на этапе ее сертификации или регистрации;
- 2) запрет на использование систем ИИ с манипулятивным функционалом. Ни один оператор или пользователь системы ИИ не вправе применять ее в целях оказания манипулятивного воздействия на физических лиц; нарушение данного запрета влечет административную или уголовную ответственность в зависимости от степени причиненного вреда и наличия умысла. При этом, следуя логике Регламента ЕС об ИИ, запрет надлежит сформулировать альтернативно — по критерию цели или эффекта, что позволит охватить случаи «непреднамеренной» манипуляции, возникающей вследствие недостаточной осмотрительности разработчика при проектировании системы;
- 3) запрет на использование техник сублиминального воздействия. С учетом стремительного развития нейроинтерфейсных технологий и систем биометрического профилирования представляется целесообразным отдельно запретить применение систем ИИ, воздействующих на сознание человека методами, выходящими за пределы его осознанного восприятия, в том числе посредством встраивания скрытых побудительных стимулов в аудиовизуальный контент;
- 4) в части Федерального закона от 27.07.2006 № 152-ФЗ «О персональных данных» предлагается расширить действие ст. 15, распространив запрет автоматизированного принятия значимых решений на все случаи использования персональных данных субъекта в целях построения его психологического профиля для нужд манипулятивного таргетирования, независимо от того, имеют ли результаты такого воздействия формально юридически значимый характер.

Эффективность запретительных норм определяется не только их содержанием, но и наличием надлежащего механизма реализации и контроля. В этой связи предлагается следующее:

1. Создание реестра запрещенных систем ИИ с манипулятивным функционалом. Включение системы в реестр должно влечь обязательный запрет ее распространения и использования на территории Российской Федерации; иностранные системы, внесенные в реестр, подлежат блокировке.

2. Введение обязательной оценки манипулятивного риска как самостоятельной процедуры в рамках сертификации систем ИИ высокого риска. Данная процедура



должна предшествовать выводу системы на рынок и включать проверку обучающих данных, алгоритмической архитектуры и функциональных возможностей системы на предмет наличия признаков манипулятивного функционала.

3. Формирование многоуровневой системы ответственности, охватывающей все звенья жизненного цикла системы ИИ: разработчика — за создание системы с манипулятивным функционалом; оператора — за введение в оборот и применение такой системы; пользователя — за намеренное использование систем ИИ общего назначения в манипулятивных целях. При этом разработчик и оператор должны нести повышенную ответственность, основанную на принципе строгой ответственности (без необходимости доказывания вины), аналогично ответственности за вред, причиненный источником повышенной опасности (ст. 1079 ГК РФ), тогда как ответственность пользователя строится на общих началах виновной ответственности.

4. Наделение Роскомнадзора дополнительными полномочиями по расследованию фактов манипулятивного воздействия с использованием систем ИИ, включая право на запрос алгоритмического аудита у разработчиков и операторов.

Таким образом, предлагаемая система мер образует взаимосвязанный нормативный комплекс, обеспечивающий правовую защиту физических лиц от манипулятивного воздействия систем ГИИ.

БИБЛИОГРАФИЯ

1. Минбалеев А. В. Особенности правовой природы агентского искусственного интеллекта // Аграрное и земельное право. — 2025. — № 8. — С. 75—78.
2. Минбалеев А. В. Риски и угрозы агентского искусственного интеллекта (agentic AI) // Информационное право. — 2025. — № 3 (85). — С. 10—13.
3. Психология манипуляции: феномены, механизмы и защита. — М. : ЧеРо ; Изд-во МГУ, 1997. — 344 с.
4. Симонян А. А. Генеративный искусственный интеллект и DeepFake-технологии как угроза информационной безопасности граждан // Вопросы российского и международного права. — 2025. — Т. 15. — № 2А. — С. 165—172.
5. Asif A. [et. al.] DarkPatterns-LLM: A Multi-Layer Benchmark for Detecting Manipulative and Harmful AI Behavior. arXiv. 2025 // URL: <https://arxiv.org/pdf/2512.22470.pdf> (дата обращения: 25.03.2026).
6. Guerrini F. AI-Driven Dark Patterns: How Artificial Intelligence Is Supercharging Digital Manipulation // Forbes. 17.11.2024. URL: <https://www.forbes.com/sites/federicoguerrini/2024/11/17/ai-driven-dark-patterns-how-artificial-intelligence-is-supercharging-digital-manipulation/> (дата обращения: 25.03.2026).
7. Khraishi R. [et al.] Real-Time Hyper-Personalized Generative AI Should Be Regulated to Prevent the Rise of «Digital Heroin». NeurIPS 2025 Position Paper // URL: <https://neurips.cc/virtual/2025/loc/san-diego/poster/121952> (дата обращения: 25.03.2026).