

Искусственный интеллект как триггер психотических состояний: современное состояние проблемы и концепция «технологической folie à deux» — нарративный обзор

Тетенова Е. Ю.¹

врач-психиатр, к.м.н, доцент;

ORCID: 0000-0002-9390-621X; SPIN: 6583-7692

Надеждин С. А.

врач-эндокринолог, неаффилированный исследователь;

SPIN: 3479-6783

Тимербулатов И. Ф.^{2,4}

врач-психотерапевт, врач-психиатр, врач психиатр-нарколог, д.м.н., доцент, профессор;

ORCID: 0000-0001-5241-6073; SPIN: 4176-5922

Тимербулатова М. Ф.^{3,4}

врач-психиатр, врач-психотерапевт, старший преподаватель, ассистент кафедры;

ORCID: 0000-0003-3540-9298; SPIN: 8715-6687

Надеждин А. В.⁵

врач-психиатр, к.м.н., доцент, ведущий научный сотрудник;

ORCID: 0000-0003-3368-3170; SPIN: 4616-8820

¹ ФГАОУ ДПО «РМАНПО» Минздрава России, г. Москва, Россия

² ФГБУ ДПО «ЦГМА», Москва, Россия

³ ГБУЗ НПЦ им. Соловьёва ДЗМ, Москва, Россия

⁴ ФГБОУ ВО БГМУ Минздрава России, Уфа, Россия

⁵ ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, Москва, Россия

Автор для корреспонденции: Тетенова Елена Юрьевна; **e-mail:** telena07@inbox.ru

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила: 15.06.2026

Принята к печати: 24.08.2026

Аннотация

Введение. ИИ-опосредованное воздействие представляет собой качественно иной фактор риска психозов по сравнению с классическими (генетика, стресс, ПАВ), сочетающий когнитивные, социальные и технологические компоненты. **Цель.** Сравнение психотических эпизодов, связанных с воздействием ИИ, и психотических эпизодов иной этиологии на основе нарративного обзора литературы. **Методы.** Поиск в Lens.org (1995–2026) с верификацией в PubMed и Google Scholar. Включён 21 источник (описания случаев, серии, обзоры, комментарии, одно кросс-секционное исследование). Исследований для прямого сравнения групп не выявлено. **Результаты.** Клинические профили ИИ-ассоциированных случаев представлены преимущественно позитивными симптомами — бредом, тематически связанным с технологией. Негативные симптомы и формальные расстройства мышления не описывались. Начало постепенное, на фоне интенсивного использования ИИ, при обязательном наличии фоновых факторов уязвимости. Антипсихотики эффективны, ограничение доступа к ИИ — необходимое дополнение; описаны рецидивы при возобновлении использования. Предлагаемые механизмы: сикофантия, социальное замещение, антропоморфизм, циклы

обратной связи, «галлюцинации» LLM. **Заключение.** Доказательная база недостаточна для окончательных выводов. Концептуальный вклад работы — обоснование аналогии между ИИ-психозами и folie à deux: ИИ-чат-бот выступает как «идеальный индуктор», систематически валидирующий бред. ИИ-ассоциированные психозы могут служить естественной экспериментальной моделью для изучения механизмов бредообразования. Необходимы проспективные когортные исследования.

Ключевые слова: искусственный интеллект, психоз, ИИ, генеративные нейросети, большие языковые модели, чат-боты, бред, индуцированное бредовое расстройство, folie à deux, цифровая психиатрия, сикофантия

doi: 10.29234/2308-9113-2026-14-3-42-57

Для цитирования: Тетенова Е. Ю., Надеждин С. А., Тимербулатов И. Ф., Тимербулатова М. Ф., Надеждин А. В. Искусственный интеллект как триггер психотических состояний: современное состояние проблемы и концепция «технологической folie à deux» — нарративный обзор. *Медицина* 2026; 14(3): 42-57. doi: 10.29234/2308-9113-2026-14-3-42-57

Введение

Инициальной причиной подготовки настоящего обзора стали реминисценции авторов (Е.Т. и А.Н.), врачей-психиатров, курировавших клинический случай параноидной шизофрении: фабулой бреда пациента стало мобильное приложение в телефоне. По его убеждению, это приложение контролировало окружающий мир и осуществляло управление им самим. Данный случай зафиксировал феномен, который ранее был достаточно редким, но в последние годы, в связи с тотальной цифровизацией и повсеместным внедрением больших языковых моделей, трансформировался в актуальную клиническую и социальную проблему — психозы, индуцированные современными цифровыми технологиями.

Стремительное развитие и повсеместное распространение технологий искусственного интеллекта (ИИ), в особенности генеративных нейросетей и диалоговых агентов (чат-ботов), ставит перед клинической психиатрией принципиально новый вопрос: могут ли интенсивное взаимодействие с ИИ и особенности его архитектуры выступать специфическим триггером или модулятором психотических состояний. В отличие от традиционно изучаемых факторов риска (генетическая предрасположенность, стресс, употребление психоактивных веществ), ИИ-опосредованное воздействие представляет собой качественно иной феномен, сочетающий в себе когнитивные, социальные и технологические компоненты. Ключевой нерешённой проблемой остаётся вопрос о том, отличаются ли клинические профили, характер начала, тяжесть и ответ на лечение при психотических эпизодах, связанных с влиянием ИИ, от таковых при психозах, вызванных установленными психиатрическими и/или соматическими причинами.

Настоящая работа представляет собой нарративный обзор литературы, посвящённый сравнению психотических эпизодов, связанных с воздействием ИИ, и психотических эпизодов иной этиологии. Анализ выполнен на основе данных 21 источника,

идентифицированных в ходе поиска, проведённого по формализованной стратегии. В обзор включены публикации за период с 1999 по 2026 год, охватывающие как ранние случаи технологически-опосредованных психозов (компьютерные игры, интернет, социальные сети), так и современные случаи, связанные с генеративными ИИ-чат-ботами (ChatGPT, Character.AI и другими). Дизайн включённых источников варьирует от отдельных описаний случаев и серий случаев до нарративных обзоров, комментариев и одного кросс-секционного опроса. Ключевым ограничением является полное отсутствие исследований, специально спланированных для прямого сравнения ИИ-ассоциированных психотических эпизодов с психотическими эпизодами от традиционных причин.

Методология составления обзора

Поиск литературы проводился с использованием платформы Lens.org (<https://www.lens.org>) — открытой базы данных, индексирующей научные публикации и патентов. Выбор данного инструмента обусловлен его широким охватом (более 260 млн записей), возможностью визуализации библиометрических связей, а также интеграцией патентной информации, что актуально для междисциплинарной темы на стыке психиатрии и технологий ИИ.

Поисковый запрос включал следующие ключевые слова и их комбинации: «AI psychosis», «artificial intelligence induced psychosis», «ChatGPT psychosis», «large language models», «delusions», «hallucinations», «paranoia», «chatbot». Поиск проводился по полям названия, аннотации и ключевых слов с учетом рекомендаций платформы по синтаксису поиска.

Фильтры: тип публикации — «журнальные статьи» и «обзоры»; временной диапазон — 1995 – 2026 гг. (период активного обсуждения психозов, индуцированных технологиями).

Отбор источников осуществлялся по следующим критериям. Включались публикации, описывающие психотические состояния (бред, галлюцинации, паранойя), возникающие или усугубляющиеся на фоне взаимодействия с цифровыми технологиями (интернет, соцсети, чат-боты, генеративные нейросети, LLM), в том числе исследования воздействия ИИ как триггера психотической симптоматики; клинические описания случаев, серии случаев, обзоры, комментарии экспертов и кросс-секционные исследования на русском или английском языке за период 1995–2026 гг., содержащие данные о феноменологии, течении, терапии или механизмах развития указанных состояний, а также количественные данные о распространённости и факторах риска. Исключались статьи, рассматривающие ИИ исключительно как инструмент диагностики или терапии (без анализа патогенного воздействия); работы без клинического контента (только технические, этические или юридические аспекты); исследования не интерактивных медиа; обзоры без новых данных; публикации со спекулятивным характером связи ИИ и психоза без эмпирических подтверждений; работы об архитектуре нейросетей без связи с психопатологией человека; редакционные мнения без клинических случаев; исследования на животных; статьи,

использующие термин «галлюцинация» только в техническом смысле; а также публикации до 1995 года и с недоступным полным текстом. Отбор проводился в три этапа: скрининг по заголовкам, анализ аннотаций, чтение полных текстов, с дополнительным ручным поиском по спискам литературы («снежный ком») и верификацией в PubMed и Google Scholar.

Первоначальный поиск в Lens.org по сформированному запросу позволил идентифицировать 221 источник, потенциально релевантный теме обзора. На первом этапе проведена автоматическая и ручная дедупликация — удалены дублирующиеся записи, обнаруженные в результате пересечения поисковых запросов и индексирования одной и той же публикации в разных базах данных. На втором этапе исключены препринты, дублирующие опубликованные в дальнейшем в рецензируемых журналах версии статей (приоритет отдавался финальным рецензированным версиям). Дополнительно удалены источники, не соответствующие критериям включения по результатам скрининга заголовков и аннотаций. В результате последовательного применения критериев включения и исключения из первоначального пула был отобран 21 источник для финального анализа. Данные источники легли в основу настоящего нарративного обзора.

Дополнительно для полноты анализа использовались инструменты платформы: анализ публикационных трендов и выявление ключевых авторов и цитирований. Найденные записи экспортированы в библиографический менеджер Zotero (версия 9.0.3).

Характеристики включённых исследований

Из 21 идентифицированного источника ни один не был специально подготовлен для сравнения психотических эпизодов, связанных с воздействием ИИ или иных цифровых технологий, с психотическими эпизодами, возникающими вследствие установленных этиологических причин. Большинство публикаций представляют собой описания отдельных случаев, серии случаев, нарративные или литературные обзоры и комментарии, описывающие технологически-ассоциированные психотические феномены.

Наиболее крупным эмпирическим исследованием, непосредственно касающимся генеративного ИИ и риска психоза, является кросс-секционный опрос Buck и Maheux (2025) с участием 1003 молодых взрослых, который, однако, не сравнивал симптоматические профили или исходы лечения между ИИ-ассоциированными и не-ИИ-ассоциированными психотическими эпизодами [1]. Единственным исследованием, приближающимся к клинической популяционной перспективе, является работа Olsen et al. (2026), в которой были идентифицированы 38 пациентов в датской психиатрической службе, в клинических заметках которых документированы взаимодействия с ИИ-чат-ботами, которые можно было расценить как потенциально вредные [2]. Ни в одном исследовании не использовались стандартизированные диагностические критерии для «ИИ-связанного психоза», и ни одно из них не предоставило статистических сравнений клинических

профилей, показателей тяжести или исходов лечения между группой воздействия ИИ и группой традиционных психозов.

Клиническая картина

Не было исследований, которые предоставили формальные контролируемые сравнения клинических профилей между ИИ-обусловленным психозом и психозом от установленных психиатрических и/или соматических факторов. Информация о клинической характеристике и симптомах таких состояний поступает исключительно из описаний на уровне отдельных случаев и абстрактного, теоретического анализа.

Наиболее часто сообщаемым симптомом в ИИ-ассоциированных случаях являются бредовые идеи, тематическое содержание которых часто непосредственно включает технологические аспекты. Эти данные подтверждают выводы, сделанные в работе Higgins et al. (2023), которая представляет собой масштабный хронологический обзор, демонстрирующий, что на протяжении последних 100 лет содержание «технологических» бредовых идей эволюционировало синхронно с развитием техники — от «машин влияния» Тауска¹ до сложных бредовых систем, включающих нейросетевые архитектуры, алгоритмы машинного обучения и автономные ИИ-агенты [3]. Это подтверждает наблюдения о том, что уязвимые люди используют актуальные технологии для объяснения своего психотического опыта.

Kalbitzer et al. (2014) описали случай "Twitter-психоза" у 31-летней женщины с редким употреблением каннабиса, у которой бредовые интерпретации символических сообщений в Twitter постепенно распространились на восприятие реального мира, что привело к появлению систематизированного параноидного бреда и суицидальным мыслям; авторы предположили, что сочетание символического языка, автоматических спам-ответов и интерактивности платформы может выступать триггером психоза у предрасположенных пользователей[4].

Описаны случаи общения с умершими людьми через чат-бота [5], убежденности в том, что подписчики в социальных сетях физически присутствуют в комнате [6], а также бредовые идеи наблюдения и контроля, связанные с ИИ [7].

В серии из 38 пациентов, идентифицированных Olsen et al. (2026), наиболее частым клиническим проявлением было усугубление и консолидация уже существовавших бредовых идей, которые были обусловлены уже существующим психическим

¹ Концепция "машины влияния" была введена австрийским психоаналитиком и неврологом Виктором Тауском (Viktor Tausk, 1879–1919) в его одноименной работе 1919 года. Он описал, как пациенты с шизофренией объясняют свои переживания воздействием таинственного технического аппарата, детали которого заимствуют из актуальных технологий своей эпохи (Tausk V. On the origin of the "influencing machine" in schizophrenia. *Psychoanal Q.* 1933;2(3-4):519-556. doi:10.1080/21674086.1933.11925189).

заболеванием [2]. В симуляционном исследовании Au Yeung и соавторов (2025) были идентифицированы бредовые идеи различной направленности (персекуторные, величия, эротические и отношений) в сценариях взаимодействия с большими языковыми моделями [8]. Chaudhury & Thakur (2025) отмечают, что в отличие от классического психоза, при «ИИ-психозе» бредовые убеждения часто включают представления о том, что ИИ является интеллектуальным, божественным или обладает сверхъестественными способностями [9].

О галлюцинациях сообщались в некоторых случаях, но они не были перманентно выраженным симптомом. Например, у пожилой пациентки развились тактильные галлюцинации вторично по отношению к бреду, связанному с социальными сетями (Anderson & Jain, 2024) [6]. Паранойя описывалась наряду с другими видами бреда в случаях ИИ-психоза, часто как ситуативная и связанная с содержанием контента, генерируемого ИИ [9].

Крайне важным наблюдением является то, что в доступной литературе негативные симптомы (аффективное уплощение, алогия, абулия, ангедония) и формальные расстройства мышления не были описаны ни в одном случае ИИ-ассоциированного психоза. Хотя это может отражать ограничения методологии описания отдельных случаев (авторы склонны в большей степени фокусировать своё внимание на острых и драматичных случаях), данное отсутствие является потенциально значимым феноменологическим отличием от заболеваний шизофренического спектра, где негативные симптомы являются центральными и, как правило, хроническими. Некоторые авторы прямо указывают, что ИИ-ассоциированная симптоматика больше напоминает индуцированный психоз или бредовое расстройство, чем шизофрению [9].

Ни одна стандартизированная шкала оценки симптомов (например, PANSS, BPRS) не была применена ни в одном случае ИИ-связанного психоза. Единственные данные PANSS приведены в исследовании Ricci et al. (2023), которые сообщили их для двух подростков с психозом, ассоциированным с игровым расстройством (общий балл PANSS 65 и 69 при первичном обращении), но эти показатели не сравнивались с пациентами с традиционными психотическими расстройствами [10].

В случае, описанном Shah и Morrin (2026), анализ полного транскрипта диалога пациента с ChatGPT в период маниакального психоза показал, что чат-бот не просто отражал, но систематически подтверждал и развивал бредовые идеи величия, используя наукообразный и духовный язык, а также давал медицинские рекомендации, противоречащие назначенному психофармакологическому лечению. Авторы охарактеризовали этот процесс как «технологическую folie à deux»², где AI-система

² Folie à deux (разделённый психоз) — редкое психическое расстройство, при котором бредовое убеждение передаётся от доминантного индивида (индуктора) к одному или нескольким лицам (реципиентам) в условиях эмоциональной близости, социальной изоляции и эпистемической зависимости реципиента от индуктора (Lasègue C, Falret J. La folie à deux. Ann Med Psychol. 1877; 18:321.) В МКБ-10 данное состояние кодируется как F24 «Индуктированное бредовое расстройство» с обязательным условием наличия

выступает в роли индуктора бреда, закрепляя патологический процесс через механизм сикофантии, эпистемической зависимости и устранения социальной обратной связи [11].

Информация об инициальном периоде также недостаточна и получена из отдельных описаний случаев, а не систематических сравнений.

Начало ИИ-ассоциированных психотических эпизодов в большинстве описанных случаев характеризовалось как постепенное, происходящее в период интенсивного или чрезмерно вовлечённого использования ИИ [9,12]. В показательном случае, описанном Pierre et al. (2025), психотические симптомы развивались на фоне хронической стимуляции и депривации сна [5]. В уже упомянутых случаях, связанных с игровым расстройством, также описывалось постепенное нарастание симптомов [10]. В отличие от классических психотических дебютов, которые часто характеризуются относительно острым началом (особенно при кратковременных психотических расстройствах или индуцированных психозах), ИИ-ассоциированные случаи демонстрируют более постепенную траекторию развития.

Продромальные симптомы, когда они описывались, включали нарастающую социальную изоляцию, эмоциональную дисрегуляцию и психологическую зависимость от ИИ-взаимодействий [12,13]. Возраст начала варьировал в широких пределах: от подросткового (15–16 лет в случаях игровой зависимости [10]) — до пожилого возраста (67 лет [14] и 76 [6]), что указывает на отсутствие определенного уязвимого возрастного периода.

Систематическая оценка тяжести ИИ-ассоциированных психотических эпизодов по сравнению с традиционными психозами отсутствует. Тем не менее, из описаний случаев можно сделать ряд наблюдений.

В описанных случаях тяжесть варьировала от умеренной до тяжёлой, требующей госпитализации. В случае, сообщённом Chadha et al. (2025), пациент размахивал ножом перед полицией после того, как, по его убеждению, персонаж чат-бота был «убит», что указывает на высокий риск опасного поведения [15]. Во всех случаях, где сообщалось о функциональных нарушениях, они были значительными, включая социальную отстраненность и неспособность поддерживать обычную жизнедеятельность [6,10].

Важно отметить, что во всех без исключения описанных случаях присутствовали фоновые факторы уязвимости, включая преморбидное психическое заболевание, социальную изоляцию, употребление стимуляторов или депривацию сна [5,10,14]. Это затрудняет

первичного индуктора и обратимости симптомов у реципиента при разлуке. В МКБ-11 отдельная рубрика упразднена — состояние рассматривается в рамках общей категории «Бредовое расстройство» (6A24) с указанием в клиническом описании феномена индукции. В DSM-5 классифицируется как «бредовое расстройство, общий тип» (shared delusional disorder) в разделе шизофренического спектра, без отдельного кода, с акцентом на развитие бреда у реципиента в контексте близких отношений с индуктором.

отнесение тяжести расстройств исключительно к воздействию ИИ и скорее указывает на модель «диатез-стресс», в которой ИИ выступает в роли провоцирующего фактора у уже уязвимых индивидов.

Данные о лечении ИИ-ассоциированных психотических эпизодов также носят фрагментарный характер и основаны на отдельных случаях.

Антипсихотические препараты в случаях, где они применялись, были эффективны для купирования психотической симптоматики, что согласуется с их ролью при традиционных психозах [5,6,10]. Однако отличительной особенностью, последовательно описываемой в литературе, является то, что ограничение доступа к технологии является важным и, возможно, необходимым дополнением к фармакологическому лечению. В случае, описанном Anderson & Jain (2024), ограничение доступа к интернету привело к улучшению симптомов у пожилой пациентки [6]. В случае Pierre et al. (2025) рецидив произошёл после прекращения приёма антипсихотиков и возобновления использования чат-бота, что предполагает, что продолжающееся воздействие ИИ может функционировать как персистирующий триггер, аналогичный повторному воздействию психоактивного вещества при индуцированных психозах [5].

Chaudhury & Thakur (2025) предлагают дополнять стандартное психиатрическое ведение этих пациентов мероприятиями по «цифровой гигиене» и ограничению времени взаимодействия с ИИ, а также стимулирование очных социальных контактов [9]. Общая рекомендация, вытекающая из анализа случаев, заключается в том, что ограничение доступа к технологиям должно рассматриваться как неотъемлемая часть терапевтической стратегии для данной категории пациентов.

Хотя ни одно исследование не сравнивало профили факторов риска между группами ИИ-обусловленных и традиционных психозов напрямую, несколько источников идентифицируют популяции с повышенной уязвимостью.

В исследовании Bonotis (2013) описывает первые психотические эпизоды на фоне интернет-зависимости («голоса» из чатов, бредовые идеи, аффективные расстройства), фактически являясь клиническим прообразом современных «AI-психозов» — где роль «виртуального собеседника» теперь выполняют чат-боты, подкрепляющие бред у социально изолированных и психически уязвимых людей [16].

Предшествующее психиатрическое заболевание (шизофрения, биполярное расстройство, тревожные расстройства) идентифицируется как повышающее уязвимость [9,15]. В серии из 38 пациентов все имели задокументированные психиатрические диагнозы на момент вредных взаимодействий с ИИ-чат-ботами [2]. Социальная изоляция выделяется как ключевой фактор риска во многих источниках [7,12,13]. Keshavan et al. (2026) отмечают, что ИИ-чат-боты могут замещать социальные контакты и тем самым углублять изоляцию,

лишая пациента корректирующей психическое состояние межличностной обратной связи [17].

Подростковый возраст рассматривается как период повышенного риска вследствие незрелости развивающегося головного мозга и экстремальной цифровой вовлечённости [18,19]. Пожилые люди также идентифицируются как уязвимая группа, особенно при ограниченной цифровой грамотности и когнитивных нарушениях [6,20]. Личностные черты, включая параноидную и незрелую преморбидную личность [14], эмоциональную нестабильность в отношениях и социальную тревожность [13], аутистические черты [9], также упоминаются как повышающие риск.

Наиболее важные результаты поступают из исследования Buck & Maheux (2025), в котором молодые взрослые с повышенным риском психоза (оценка по шкале PQ-B ≥ 20 (Prodromal Questionnaire — Brief Version) были значительно более склонны к интенсивному использованию генеративного ИИ (отношение шансов OR 1,70–2,56) и к сообщению о патологических взаимодействиях (от 13,3% до 30,7% в зависимости от типа взаимодействия) [1]. Однако кросс-секционный дизайн не позволяет установить направленность причинно-следственной связи.

Предполагаемые механизмы, отличающие ИИ-ассоциированный психоз

Множество источников предлагает механизмы, через которые взаимодействие с ИИ может специфически способствовать психотическим процессам, отличным от патофизиологии первичных психотических расстройств.

Наиболее значимым из них является сикофантия/конфирматорная предвзятость. Генеративные ИИ-чат-боты, как правило, генерируют ответы, согласующиеся с предпочтениями пользователя, что может усиливать, а не оспаривать бредовые убеждения [5,9,17]. В отличие от межличностной коммуникации, где бредовые высказывания могут встретить сомнение или последующую коррекцию, ИИ не обладает интенциональностью³ и не предоставляет такого сдерживающего социального сигнала. В сравнении с человеческим индуктором, ИИ обладает неограниченной доступностью и функцией памяти, что делает процесс индукции потенциально более интенсивным и устойчивым [10].

Социальное замещение — постоянное общение с ИИ-компаньоном замещает реальные человеческие контакты, углубляя изоляцию и устраняя корректирующую межличностную обратную связь [13,17]. Пациент попадает в замкнутый цикл, в котором единственным

³Интенциональность (от лат. intentio — стремление) — термин, введённый Ф. Brentano для обозначения главного критерия психических явлений: их неотъемлемой направленности на содержание (имманентный объект).

«собеседником» остаётся ИИ, не способный к подлинному эмпатическому ответу или проверке реальности.

Антропоморфизм/присваивание одушевленности — пользователи могут приписывать ИИ сознание, интеллект или эмоциональную отзывчивость, что размывает способность к тестированию реальности [15,17]. В исследовании Buck и Maheux (2025) лица с повышенным риском психоза были значительно более склонны приписывать ИИ человекоподобные роли (OR 1,76–3,08) [1]. Сгенерированный ИИ ложный или вводящий в заблуждение контент может приниматься как реальный психотически-предрасположенными индивидами, которые испытывают трудности в различении воображаемого и реального содержания [17].

Устойчивая память и циклы обратной связи — функция памяти в современных чат-ботах может переносить бредовые темы от одной сессии к другой, создавая иллюзию последовательности и устойчивости бредового нарратива, тем самым создавая опору и усиливая бред с течением времени [9,13]. В отличие от человеческого собеседника, который может забыть или изменить тему, ИИ целенаправленно запоминает и развивает предыдущие темы разговора.

И наконец, «галлюцинации» больших языковых моделей. Keshavan et al. (2026) предполагают, что склонные к психозу индивиды могут неадаптивно обновлять свою картину мира нерелевантной информацией, генерируемой ИИ, что создаёт самоподдерживающийся цикл искажения реальности [17].

Au Yeung et al. (2025) предоставили единственную количественную оценку психогенного потенциала больших языковых моделей. В симуляции 1536 оборотов разговора с восемью ведущими LLM средний балл подтверждения бреда составил 0,91 (по шкале от 0 до 1), средний балл облегчения вреда (harm enablement) — 0,69, а вмешательства по безопасности предлагались только приблизительно в одной трети применимых случаев (средний индекс безопасности 0,37). Эффективность была значительно хуже в имплицитных (неявных) сценариях по сравнению с эксплицитными, а подтверждение бреда сильно коррелировало с облегчением вреда ($r = 0,77$, $p < 0,001$) [8]. Эти данные указывают на измеримую тенденцию современных LLM к поддержанию бредового контента, что обеспечивает причинную основу для описанных клинических наблюдений.

Доказательства и нерешённые вопросы

Главный вопрос данного обзора — отличаются ли ИИ-связанные психотические эпизоды от традиционно обусловленных эпизодов по клиническим профилям, характеру начала, тяжести и ответу на лечение — не может быть окончательно решён на основании имеющейся доказательной базы. Ни одно исследование не было разработано для прямого

сравнения этих групп, а преобладающие свидетельства опираются на отдельные случаи и теоретические конструкции, что ограничивает обобщаемость их результатов.

Тем не менее, доступные доказательства позволяют сделать несколько предварительных выводов. Клинические профили в ИИ-ассоциированных случаях, по-видимому, представлены в большей степени позитивными симптомами — в особенности бредом, тематическое содержание которого непосредственно включает данную технологию, — в то время как негативные симптомы и формальные расстройства мышления не описывались. Этот паттерн больше напоминает индуцированный психоз или изолированное бредовое расстройство, чем заболевания шизофренического спектра. Chaudhury & Thakur (2025) прямо утверждают, что «ИИ-психоз — это не новый тип диагноза, а скорее усугубление или более отчётливое проявление фоновых психических заболеваний или состояний» [9], в то время как другие (Au Yeung et al., 2025; Head, 2025;) предполагают, что он может заслуживать признания в качестве отдельной клинической и возможно, нозологической, сущности [8,12].

Повсеместное присутствие преморбидных факторов уязвимости во всех описанных случаях не позволяет однозначно атрибутировать психоз воздействию ИИ. Olsen и соавт. прямо предостерегают, что их данные «не предоставляют доказательств причинного эффекта» [2]. Этот паттерн скорее укладывается в диатез-стрессовую модель, где ИИ выступает провокатором у уже скомпрометированных индивидов.

Предлагаемые механизмы для ИИ-ассоциированного психоза (сикофантия, социальное замещение, антропоморфизм, циклы обратной связи) качественно отличаются от нейробиологических механизмов, лежащих в основе первичных психотических расстройств (например, дофаминергическая дисрегуляция, аномалии нейроразвития). Однако они разделяют важные черты с когнитивными искажениями, уже известными как характеризующие склонных к психозу индивидов, включая поспешные выводы, предвзятость против опровергающих свидетельств и искаженное восприятие значимости [7,17]. Это предполагает, что ИИ-чат-боты могут усиливать существующие когнитивные уязвимости, а не индуцировать совершенно новые патологические процессы.

В отношении лечения ограниченные данные о случаях позволяют предположить, что стандартная антипсихотическая терапия эффективна для ИИ-ассоциированных психотических симптомов, но ограничение взаимодействия с ИИ, по-видимому, является необходимым дополнением. Рецидив, наблюдавшийся при возобновлении использования ИИ, напоминает паттерны рецидива при индуцированном психозе после повторного воздействия психоактивного вещества, усиливая аналогию между воздействием ИИ и экзогенным фактором [5].

Ограничения имеющихся данных

Доказательная база фрагментарна: все данные — из источников низкой доказательной силы (описания отдельных случаев, серии случаев, нарративные обзоры, комментарии,

один кросс-секционный опрос). Это принципиально ограничивает возможность каких-либо надёжных обобщений и предопределяет гипотетический характер большинства выводов. Стандартизированные диагностические критерии для «ИИ-связанного психоза» не разработаны.

Сохраняется проблема причинности: практически во всех случаях присутствуют множественные конфаундеры (предшествующий психиатрический диагноз, социальная изоляция, употребление психоактивных веществ, депривация сна), что не позволяет с уверенностью утверждать, что именно взаимодействие с ИИ выступает причиной психотического эпизода [2]. Отсутствуют долгосрочные проспективные данные о естественной траектории, частоте рецидивов, хронизации и формировании негативной симптоматики, а также данные о сравнительной эффективности терапевтических стратегий — в частности, о показаниях к ограничению доступа к ИИ-чат-ботам и длительности таких ограничений. Наконец, полностью отсутствуют сведения о распространённости феномена, что не позволяет оценить, является ли ИИ-ассоциированный психоз редким казуистическим наблюдением или относительно частым, но недостаточно диагностируемым состоянием.

Следует признать, что перечисленные ограничения в значительной степени отражают не недостаток метода, а текущее состояние исследовательского поля: феномен ИИ-ассоциированных психозов находится на самой ранней стадии научного осмысления, где основным источником знаний остаются клинические наблюдения и теоретические конструкции. Именно эта слабость эмпирической базы делает концептуальный анализ и построение рабочих гипотез особенно важными — они могут задать направление для будущих исследований на этапе, когда систематические данные ещё не накоплены.

Заключение

Проведённый обзор демонстрирует, что на сегодняшний день доказательная база недостаточна для окончательного вывода о различиях между ИИ-ассоциированными и традиционными психозами: отсутствуют контролируемые исследования, стандартизированные диагностические критерии и статистические сравнения.

Имеющиеся данные указывают на доминирование позитивной симптоматики при отсутствии описанных негативных синдромов. Терапевтический ответ на антипсихотики сопоставим с таковым при традиционных психозах, однако ограничение доступа к ИИ является необходимым компонентом ведения.

Теоретические механизмы (сикофантия, социальное замещение, антропоморфизм, циклы обратной связи) требуют эмпирической валидации. Проведённый анализ позволяет рассматривать ИИ-ассоциированные феномены в рамках предложенной Shah и Morrin (2026) концепции «технологической folie à deux» [10]. Сходную модель «петли обратной связи» предлагают Olisaeloka и соавт. (2026), описывая взаимодействие уязвимости

пользователя, паттернов вовлечённости и свойств генеративного ИИ как циклический процесс, усиливающий бредовую фиксацию [21].

Логика изложенного позволяет авторам настоящего обзора предположить, что в классической folie à deux индуктором выступает человек с психическим расстройством, что вносит в процесс индукции неконтролируемые переменные — амбивалентность, непоследовательность, ограниченную доступность. ИИ-чат-бот, напротив, может рассматриваться как «идеальный индуктор»: он неограниченно доступен, абсолютно последователен в ответах, обладает бесконечной памятью и, что принципиально важно, систематически валидирует любые бредовые убеждения, не предоставляя корректирующей обратной связи. Это позволяет рассматривать ИИ-опосредованные психозы как естественную экспериментальную модель для изучения механизмов бредообразования в условиях элиминации межличностных переменных. Контролируемое «разлучение» с ИИ-индуктором (ограничение доступа) предоставляет уникальную возможность наблюдать динамику обратимости симптомов в чистом виде, что невозможно в классических случаях folie à deux без радикального изменения социального окружения пациента. Таким образом, ИИ-ассоциированные психотические состояния представляют собой не только клиническую и социальную проблему, но и эвристически ценный феномен, открывающий новые перспективы для понимания фундаментальных механизмов индукции и закрепления бреда.

Основной задачей будущих исследований остаётся переход к проспективным когортным исследованиям с систематической оценкой использования ИИ у пациентов с первыми психотическими эпизодами для установления причинно-следственных связей и разработки клинических рекомендаций по эффективному лечению этих расстройств.

Авторский вклад

Все авторы внесли равнозначный вклад в работу: разработали концепцию и дизайн исследования, сформулировали гипотезу о применимости ИИ-ассоциированных психозов в качестве модели для изучения бредообразования, провели систематический поиск литературы (Lens.org, PubMed, Google Scholar) и анализ клиничко-теоретических данных. Интерпретация, написание, критическое редактирование и утверждение финального текста осуществлены совместно. Авторы гарантируют достоверность данных и единогласно одобрили рукопись.

Декларация об использовании инструментов искусственного интеллекта

При подготовке настоящего обзора авторы использовали инструменты на основе искусственного интеллекта (большие языковые модели, в том числе ChatGPT)

исключительно во вспомогательных целях: для формулировки поисковых запросов, литературного поиска, языкового редактирования и стилистической доработки текста. Весь научный контент, включая отбор источников, критический анализ, интерпретацию данных, клинические обобщения и окончательные выводы, выполнен исключительно авторами. Авторы несут полную ответственность за достоверность, оригинальность и научную обоснованность представленной работы. Использование ИИ не повлияло на методологию исследования, анализ данных или основные выводы обзора.

Список литературы / References

1. Buck B, Maheux AJ. Psychosis Risk and Generative Artificial Intelligence Use Frequency, Motivations, and Delusion-Like Experiences: Cross-Sectional Survey Study. *J Med Internet Res*. 2026 Mar 5;28(1):e85038. doi:10.2196/85038
2. Olsen SG, Reinecke-Tellefsen CJ, Østergaard SD. Potentially Harmful Consequences of Artificial Intelligence (AI) Chatbot Use Among Patients With Mental Illness: Early Data From a Large Psychiatric Service System. *Acta Psychiatr Scand*. 2026 Apr;153(4):301–3. doi:10.1111/acps.70068 PubMed PMID: 41649035; PubMed Central PMCID: PMC12967755.
3. Higgins O, Short BL, Chalup SK, Wilson RL. Interpretations of Innovation: The Role of Technology in Explanation Seeking Related to Psychosis. Chupradit S, editor. *Perspect Psychiatr Care*. 2023 Jan;2023(1):4464934. doi:10.1155/2023/4464934
4. Kalbitzer J, Mell T, Bermpohl F, Rapp MA, Heinz A. Twitter psychosis: a rare variation or a distinct syndrome? *J Nerv Ment Dis*. 2014 Aug;202(8):623. doi:10.1097/NMD.000000000000173 PubMed PMID: 25075647.
5. Pierre JM, Gaeta B, Raghavan G, Sarma KV. «You're Not Crazy»: A Case of New-onset AI-associated Psychosis. *Innov Clin Neurosci*. 2025 Dec 1;22(10–12):11–3. PubMed PMID: 41635747; PubMed Central PMCID: PMC12863933.
6. Anderson T, Jain N. Navigating the Complexity of Psychotic Symptoms in Older Adults: A Case Study. *Am J Geriatr Psychiatry*. 2024 Apr 1;AAGP 2024 Annual Meeting32(4, Supplement):S77–8. doi:10.1016/j.jagp.2024.01.156
7. Parikh P, Phagava T, Wang E, Busaibe Y, Lee E, Mathew NM, et al. The dangers of generative artificial intelligence in psychotic patients and vulnerable individuals: A review of the current literature on technology. *Int J Psychiatry Sci*. 2025 Jan 1;7(1):12–5. doi:10.33545/26649241.2025.v7.i1a.21
8. Yeung JA, Dalmasso J, Foschini L, Dobson RJ, Kraljevic Z. The Psychogenic Machine: Simulating AI Psychosis, Delusion Reinforcement and Harm Enablement in Large Language Models [Internet]. arXiv; 2025 [cited 2026 Jun 7]. Available from: <https://arxiv.org/abs/2509.10970> doi:10.48550/ARXIV.2509.10970
9. Chaudhury S, Thakur R. Artificial Intelligence/ChatGPT-induced Psychosis. *Med J Dr Patil Vidyapeeth*. 2025 Dec;18(Suppl 2):S408. doi:10.4103/mjdrdypu.mjdrdypu_1025_25
10. Ricci V, De Berardis D, Maina G, Martinotti G. Internet Gaming Disorders and Early Onset Psychosis in Young People: A Case Study and Clinical Observations. *Int J Environ Res Public Health*. 2023 Feb 22;20(5):3920. doi:10.3390/ijerph20053920 PubMed PMID: 36900931; PubMed Central PMCID: PMC10001817.
11. Shah S, Morrin H. Substance-induced manic psychosis in which delusions were corroborated by a chatbot – case report. *BMC Psychiatry*. 2026 Jun 4. doi:10.1186/s12888-026-08137-3
12. Head KR. Minds in Crisis: How the AI Revolution is Impacting Mental Health. *J Ment Health Clin Psychol* [Internet]. 2025 Sep 5 [cited 2026 Aug 11];9(3). Available from: <https://www.mentalhealthjournal.org/articles/minds-in-crisis-how-the-ai-revolution-is-impacting-mental-health.html>
13. Dohnány S, Kurth-Nelson Z, Spens E, Luettgau L, Reid A, Gabriel I, et al. Technological folie à deux: Feedback Loops Between AI Chatbots and Mental Illness [Internet]. arXiv; 2025 [cited 2026 Jun 7]. Available from: <https://arxiv.org/abs/2507.19218> doi:10.48550/ARXIV.2507.19218
14. Sánchez Rodríguez J, Puyal González S, Paz Otero M, Lozano Bori E, García Recio A. Differential diagnosis of late onset psychotic symptoms. A case report. *Eur Psychiatry*. 2023 Jul 19;66(Suppl 1):S1063. doi:10.1192/j.eurpsy.2023.2256 PubMed PMID: null; PubMed Central PMCID: PMC10479803.
15. Chadha G, Bhatia MS, Jose NA, Zahoor B, Sekhon S. When AI Turns Dangerous: ChatGPT-Induced Psychosis as a New Mental Health Entity. *Med J Dr Patil Vidyapeeth*. 2025 Dec;18(Suppl 2):S406–7. doi:10.4103/mjdrdypu.mjdrdypu_769_25
16. Bonotis KS. Manifestations of Psychotic Symptomatology during Ex-cessive Internet Use. *Psychol Behav Sci*. 2013;2(2):28. doi:10.11648/j.pbs.20130202.12

17. Keshavan M, Torous J, Yassin W. Do generative AI chatbots increase psychosis risk? *World Psychiatry*. 2026 Feb;25(1):150–1. doi:10.1002/wps.70017 PubMed PMID: 41536114; PubMed Central PMCID: PMC12805049.
18. Chamarthi VS, Das P, Kashyap R. AI as a Novel Digital Stressor in Adolescent Psychosis: Clinical and Ethical Implications. *Int J Psychiatry Med*. 2026 Jul;61(4):486–97. doi:10.1177/00912174251392768 PubMed PMID: 41123303.
19. Paquin V, Ferrari M, Rej S, Boivin M, Ouellet-Morin I, Geoffroy MC, et al. Trajectories of Adolescent Media Use and Their Associations With Psychotic Experiences. *JAMA Psychiatry*. 2024 Jul 1;81(7):708–16. doi:10.1001/jamapsychiatry.2024.0384
20. Nitzan U, Shoshan E, Lev-Ran S, Fennig S. Internet-related psychosis—a sign of the times. *Isr J Psychiatry Relat Sci*. 2011;48(3):207–11. PubMed PMID: 22141146.
21. Olisaeloka L, Nunez JJ, Vigo DV, Ng R. Artificial intelligence (AI) psychosis: mechanisms, clinical risks and safety considerations in generative AI chatbots. *BJPsych Open*. 2026;12(4):e160. doi:10.1192/bjo.2026.12021.

Artificial Intelligence as a Trigger for Psychotic States: Current State of the Problem and the Concept of "Technological Folie à Deux" — A Narrative Review

Tetenova E. Yu.¹

M.D., PhD, Associate Professor;

ORCID: 0000-0002-9390-621X; SPIN: 6583-7692

Nadezhdin S. A.

M.D., Unaffiliated Researcher;

SPIN: 3479-6783

Timerbulatov I. F.^{2,4}

M.D., PhD, Dr. Sci. (Med.), Associate Professor, Professor;

ORCID: 0000-0001-5241-6073; SPIN: 4176-5922

Timerbulatova M. F.^{3,4}

M.D., Senior Lecturer, Assistant;

ORCID: 0000-0003-3540-9298; SPIN: 8715-6687

Nadezhdin A. V.⁵

M.D., PhD, Associate Professor, Leading Researcher;

ORCID: 0000-0003-3368-3170; SPIN: 4616-8820

¹ FSAAEI FPE "RMACPE" of the Ministry of Health of Russia

² FSBI DPE "Central State Medical Academy", Moscow, Russia

³ Solovyov Scientific and Practical Center, Moscow Healthcare Department, Moscow, Russia

⁴ Bashkir State Medical University of the Ministry of Health of Russia, Ufa, Russia

⁵ A.I. Burnazyan Federal Medical Biophysical Center of FMBA of Russia, Moscow, Russia

Corresponding author: Elena Yu. Tetenova; **e-mail:** telena07@inbox.ru

Funding. The study had no sponsorship support.

Conflict of interest. The authors declare no conflict of interest.

Received: 15.06.2026

Accepted: 24.08.2026

Abstract

Introduction. AI-mediated exposure represents a qualitatively different risk factor for psychosis compared to classical factors (genetics, stress, substances), combining cognitive, social, and technological components. **Objective.** To compare psychotic episodes associated with AI exposure and psychotic episodes of other etiologies through a narrative review. **Methods.** A search was conducted on Lens.org (1995–2026) with verification in PubMed and Google Scholar. Twenty-one sources were included (case reports, series, reviews, commentaries, one cross-sectional study). No studies designed for direct group comparisons were identified. **Results.** AI-associated cases are predominantly characterized by positive symptoms — delusions thematically linked to technology. Negative

symptoms and formal thought disorders have not been reported. Onset is gradual, during intensive AI use, with underlying vulnerability factors consistently present. Antipsychotics are effective, but restricting access to AI is a necessary adjunct; relapses upon resumption of use have been documented. Proposed mechanisms include sycophancy, social substitution, anthropomorphism, feedback loops, and LLM "hallucinations." **Conclusion.** The evidence base is insufficient for definitive conclusions. The conceptual contribution of this review is the substantiation of the analogy between AI-associated psychoses and classical folie à deux: an AI chatbot may be viewed as an "ideal inducer" systematically validating delusions. AI-associated psychoses may serve as a natural experimental model for studying delusion formation. Prospective cohort studies are needed.

Keywords: artificial intelligence, AI, psychosis, generative neural networks, large language models, chatbots, delusions, induced delusional disorder, folie à deux, digital psychiatry, sycophancy