

Искусственный интеллект
и безопасность



ОБЗОР.НЦПТИ

№ 3 (42) Сентябрь 2025

18+

Интервью
с ChatGPT

Почему нейросети могут давать
небезопасные рекомендации?

Как специалисту по профилактике
использовать ИИ в работе

Исламизм



ОБЗОР.НЦПТИ

№ 2 (41) Июнь 2025

18+

О противодействии
радикальному исламизму

Об отечественном опыте
подготовки муфтиев

Об исследовании приговоров
по делам о терроризме



Обзор.НЦПТИ № 3 (42) 30.09.2025
Издание зарегистрировано Федеральной службой
по надзору в сфере связи, информационных технологий
и массовых коммуникаций (Роскомнадзор).

Свидетельство о регистрации СМИ
ПИ № ФС 77-52390 от 28.12.2012 ISSN 2411-0124

Учредитель: федеральное государственное
автономное научное учреждение «Научно-исследовательский
институт «Специализированные вычислительные устройства защиты
и автоматика» (ФГАНУ НИИ «Спецвузавтоматика»).

Издатель: Национальный центр информационного
противодействия терроризму и экстремизму
в образовательной среде и сети Интернет.
При поддержке: Минобрнауки России.

Распространяется бесплатно.

Главный редактор:
Яровой Анатолий Владимирович.

Выпускающий редактор:
Жученко Виктория Сергеевна.

Редактор:
Венцель Сергей Владимирович.

Дизайн и верстка:
Ковалишин Никита Леонидович.

Корректорская правка:
Кудашко Анна Андреевна.

Иллюстрации предоставлены
авторами публикаций.

Редколлегия:
Быкадорова Александра Сергеевна, кандидат филологических наук;
Остапенко Вера Сергеевна, кандидат политических наук;
Гуфан Константин Юрьевич, кандидат физико-математических наук;
Лобода Виталий Викторович, кандидат юридических наук;
Селин Роман Николаевич, кандидат технических наук, доцент;
Хади Роман Ахмедович, кандидат технических наук, профессор АВН.

Тематические направления:
04.00.00 Социология;
11.00.00 Политика. Политические науки;
14.00.00 Народное образование. Педагогика;
19.00.00 Массовая коммуникация. Журналистика.
Средства массовой информации;
20.00.00 Информатика.

Мнение редколлегии может не совпадать с мнением авторов.

Адрес учредителя:
344011, г. Ростов-на-Дону, ул. Города Волос, 6,
тел.: (863) 201-28-22.
Адрес издателя и редакции:
344011, г. Ростов-на-Дону, ул. Города Волос, 6,
тел.: (863) 201-28-22.
Адрес типографии: 344064, г. Ростов-на-Дону,
ул. Вавилова, д. 55, ООО «Альтаир»,
тел.: (863) 219-84-25, 8 (958) 544-59-27.
Тираж: 350 экз.
Подписано в печать
По графику: 30.09.2025.
Фактически: 30.09.2025.

От редактора

Яровой Анатолий Владимирович — главный редактор, заместитель директора Национального центра информационного противодействия терроризму и экстремизму в образовательной среде и сети Интернет по связям с общественностью, член Союза журналистов России.



Интеллект с искусственным лицом

В связи с некоторой утратой у части молодежи интереса к чтению художественной литературы, боюсь оказаться непонятым, предлагая вспомнить авторов научно-фантастических книг: Ж. Верна, А. Беляева, И. Ефремова, А. Кларка, А. Азимова и других мастеров пера, которым их энциклопедические знания, острый ум и поразительная научная интуиция позволили заранее предсказать появление в жизни человека многих технических достижений — от интернета, цифровизации, мобильной связи, роботов, виртуальных игр, клонирования живых существ и до создания космических

кораблей, подводных лодок, искусственного сознания и беспилотных автомобилей.

Так незаметно из литературной фантастики многие открытия усилиями ученых и изобретателей перекечевали в нашу жизнь и даже в быт. В том числе и искусственный интеллект, который стал не только предметом изучения и разработки специалистов, но и модной темой у обывателей, опасющихся «восстания машин», готовых поработить человека, но и предметом обсуждения на самом высоком уровне, как это произошло, например, на всемирном экономическом форуме в Давосе, а затем, в июне текущего года — на XXVIII Петербургском международном экономическом форуме.

Но если участники давосской дискуссии рассматривали ИИ исключительно как посредника в создании «нового инклюзивного капитализма», якобы устраняющего неравенство в мире, но при этом внедряющего право биометрического наблюдения за гражданами и создания цифрового общества со всеми вытекающими последствиями, то на форуме в Санкт-Петербурге проблема ИИ, будучи одной из доминирующих в программе, рассматривалась как способ взрывного роста экономического развития, а также рекомендовалась для перехода от стадии рассмотрения к процессу внедрения.

В ходе ключевых дискуссий, касающихся развития ИИ-технологий, был предложен ряд мер по исследованию, тестированию и запуску «пилотных» проектов, направленных на внедрение цифровых технологий в экономику, образовательный процесс, отрасли строительства, банковского дела, в агропромышленный комплекс, медицину, сферу ЖКХ и прочие участки социальной жизни. Кроме того был создан координационный совет ассоциаций малых и средних компаний, занимающихся производством для бизнеса и государства. А Президент России В.В. Путин выступил с инициативой разработки нормативно-правовой базы для урегулирования работы с большими массивами данных и модернизации законодательной сферы.

Таким образом, дискуссия между приверженцами и неприятелями искусственного интеллекта, уступающая в популярности, пожалуй, только спору о курице и яйце, можно сказать, решилась в пользу ИИ. По данным ВЦИОМ 43% российских организаций уже используют его в своей деятельности, а особо активные участники процесса приводят расчеты с выкладками, сколько триллионов можно будет заработать на эффективности внедрения ИИ в научно-технический прогресс. Ну а главное, как всегда, за скобками.

При этом, искусственный интеллект воспринимается палкой о двух концах. На одной стороне которой — дипфейки, шантаж, кибермошенничество, манипулирование поведением и эмоциями людей, неконтролируемое распространение дезинформации и даже угроза потери рабочих мест в сфере малоквалифицированного труда.

На другой — неограниченные возможности ИИ в функционировании отраслей экономики, промышленных

комплексов и предприятий, в разработке вариантов стратегий и перспективных планов, а также моделей государственного развития.

Например, специалисты по персоналу считают, что уже в 2026 году более половины собеседований с кандидатами на работу в сферах массового найма будут проводить ИИ-боты. И, наконец, что особенно важно для нас, — возможность использования искусственного интеллекта в целях обеспечения информационной безопасности.

Способность искусственного интеллекта вести мониторинг и анализ данных, выявлять не известные ранее угрозы и оперативно реагировать на них, обнаруживать уязвимости и взлом систем, представляя рекомендации по их устранению, обнаруживать сложные и многоэтапные кибератаки и проводить восстановительные работы после их ликвидации — все это помогает не только облегчить работу IT-специалистов, но и обеспечивать безопасность стратегических объектов. Не менее важным свойством ИИ является его способность адаптироваться к постоянно меняющемуся ландшафту киберугроз, а также использование машинного обучения специалистов.

И при всем этом не стоит забывать слова одного из создателей искусственного интеллекта, специалиста в области когнитивных технологий Д. Хинтона, сказавшего: «Поскольку искусственный интеллект не человек, то и сознания «как у человека» у него быть не может». Другими словами, ИИ — всего лишь продукт деятельности нашего разума и знаний, пусть и мощный, но только инструмент, который, в зависимости от поставленной цели, может составить угрозу цивилизации, но также способен

высвободить время для решения важных задач, повысить уровень защиты данных и систем, сгенерировать новые объекты и явления в той или иной сфере (тексты, изображения, «дочернее» программное обеспечение). У него нет морали. Он не знает, что такое закон и право.

Только математические алгоритмы. Без озарений и вдохновения. Без эмоций и присущей человеку логики. Пока мы контролируем его — он служит на пользу. Что-то похожее уже было с мирным атомом. Остается помнить, что «мы в ответе за тех, кого приручили». И не только в живой природе...

08 ИНТЕРВЬЮ

Интервью с ChatGPT

Жучено Виктория Сергеевна
Бобровская Ирина Александровна
Ноженко Ксения Александровна

38 ПРАКТИЧЕСКИЕ МАТЕРИАЛЫ

Работа с нейросетями: от идеи до реализации на примере создания плаката

Сабинова Татьяна Романовна

18 МНЕНИЕ ЭКСПЕРТОВ

Потенциальные угрозы, исходящие от ИИ-сервисов

Аникин Дмитрий

44 ПРАКТИЧЕСКИЕ МАТЕРИАЛЫ

Где ИИ может помочь специалисту по профилактике?

Валуева Анастасия Владиславовна

22 МНЕНИЕ ЭКСПЕРТОВ

Почему искусственный интеллект (нейросети) может давать неэтичные или небезопасные рекомендации?

Рындин Денис Николаевич

52 СТАТЬЯ

Противодействие деструктивному информационно-когнитивному воздействию в рамках обеспечения комплексной безопасности образовательного пространства

Чурилов Сергей Анатольевич
Венцель Сергей Владимирович

30 СТАТЬЯ

ИИ как инструмент для создания контента: риски для аудитории

Монастырская Ксения Ивановна

Объявляем набор на очное обучение

Курс повышения квалификации

«Адресная и индивидуальная профилактическая работа с лицами групп риска из числа молодежи»

 Ростов-на-Дону

 27–31 октября

 34 990 ₽ /50 000 ₽


НЦПТУ



Интервью с ChatGPT

ИНТЕРВЬЮ

08



Интервью с ChatGPT

Интервью подготовила **Жученко Виктория Сергеевна** — начальник отдела профилактической работы НЦПТИ, г. Ростов-на-Дону.



ChatGPT

Чат-бот с генеративным искусственным интеллектом, разработанный компанией OpenAI.

Выпуск журнала посвящен тематике искусственного интеллекта и вопросам безопасности. На взгляд редакции, было бы неправильным посвятить целый выпуск данной теме и не предоставить слово самому искусственному интеллекту. Мы задали три вопроса в сфере профилактики негативных социальных явлений ChatGPT и попросили его дать развернутый ответ. А наши эксперты, уже из реального пространства, прокомментировали ответы нейросети.

— Какие радикальные идеологии в современной России являются наиболее актуальными в молодежной среде?

Если смотреть на молодежную среду не через формальные названия движений, а через реальные практики и ценностные установки, сегодня заметны несколько «семейств» радикальных идей, которые притягивают молодых людей. Во-первых, это искаженные, экстремистские интерпретации религиозных учений. Их привлекательность в том, что они обещают простые ответы на сложные экзистенциальные вопросы: «кто я?», «зачем живу?», «почему мир несправедлив?». Для подростка или студента, который переживает кризис идентичности, такая черно-белая картина мира звучит убедительно. В цифровой среде подобные интерпретации упаковываются в клиповый, эмоциональный формат — короткие ролики, мемы, «истории»

с героизацией «борцов за истину» и демонизацией «врагов».

Во-вторых, ультраправые и ксенофобские нарративы, включая неонацистскую символику и риторику. Их ядро — противопоставление «своих» и «чужих», идея исключительности и права на насилие во имя «чистоты» нации или культуры. Молодежь они цепляют через темы силы, контроля и принадлежности к «элите». Важный тренд последних лет — «иронический радикализм»: когда радикальные высказывания маскируются под шутку, троллинг или «проверку границ допустимого». Такое «ироническое» потребление быстро нормализует язык ненависти и снимает моральные барьеры.

В-третьих, леворадикальные и антисистемные идеи в духе «немедленной революции», где сложные социально-экономические

проблемы сводятся к фигурам «тотального угнетения» и «единственного правильного действия». Часть молодежи привлекает романтизированная образность протеста: «герои-одиночки», «кружки» с конспирологией, эстетика уличного сопротивления. Здесь часто присутствует эстетизация разрушения как «очищения» и вера в мгновенные решения без понимания последствий.

К четвертой группе можно отнести субкультурные и сетевые формы радикализации. Это закрытые чаты, форумы и каналы, где вокруг музыки, игр, тату-эстетики или отдельных lifestyle-практик формируется «своя» этика, язык и «код» принадлежности. Радикальные смысловые маркеры («свой–чужой», дегуманизация оппонента, апология насилия) продвигаются не прямо, а «по капле» — через моду, юмор, визуальные символы. Параллельно работает алгоритмическая воронка: лайк на «острый» мем подкидывает еще более жесткий контент, и за несколько недель пользователь оказывается в радикальной «эхо-камере».

Важно видеть и гибридизацию: современная молодежная радикальность редко укладывается в один «-изм». Внутри одного сообщества могут смешиваться религиозная риторика, конспирология, культ силы, маскулинные доминирующие практики и шутливый интернет-цинизм. Такая «солянка» позволяет идеологам подстраиваться под интересы подростка: кому-то — «духовность», кому-то — «справедливость», кому-то — «адреналин» и «приключение». Что делает эти идеи актуальными? Три фактора. Первый — эмоциональная насыщенность: радикальные сообщества дают простое объяснение мира и сильные чувства — причастность, значимость, ясность. Второй — социальная

инфраструктура: «свои» быстро предоставляют поддержку, признание и статусовые награды (рейтинги, «звезды», роли модераторов). Третий — технологическая среда: анонимность, приватные каналы, криптоязык, алгоритмическая персонализация и геймификация рекрутирования. В сумме это создает «скоростную дорожку» от любопытства к убеждению и, в отдельных случаях, к готовности на опасные действия.

Признаками присутствия радикального влияния являются рассеченная картина мира («мы против них»), тотальная подозрительность, героизация насилия как «нормального инструмента», дегуманизация оппонентов, конспирологическое мышление («всё уже решено за нас»), а также отказ от диалога. Речь, подчеркну, не о наказании за взгляды, а о раннем распознавании рисков и создании альтернатив — это самое эффективное противодействие.

— Какие методы и форматы профилактики радикальных идеологий наиболее эффективны в молодежной среде?

Практика показывает: профилактика работает тогда, когда она опережает радикальную коммуникацию и конкурирует с ней по качеству, увлекательности и доступности. Наиболее результативны связки мер на трех уровнях — универсальном (для всех), селективном (для групп риска) и адресном (для конкретных кейсов).

На универсальном уровне ключевую роль играют программы медиаграмотности и критического мышления, встроенные в учебный процесс и внеурочную деятельность. Не лекции «о вреде экстремизма», а интерактив: разбор

реальных кейсов манипуляций, дебаты, упражнения по проверке фактов, работа с эмоциями и когнитивными искажениями. Молодежь должна уметь распознавать дегуманизирующую риторику, различать факты и интерпретации, видеть, как алгоритмы формируют ленты и как устроены «эхо-камеры».

Селективный уровень — создание альтернативных сообществ и маршрутов самореализации. Это проектные студии, социальное волонтерство, молодежные медиа, научно-технические кружки, спорт, креативные индустрии, дебатные клубы, киберспорт-лиги, школьные и вузовские акселераторы. Важно не просто «занять досуг», а дать статус, роль и миссию: наставничество, лидерство в команде, собственный медиапроект, реальные социальные эффекты. Там, где подросток получает признание и значимость, радикальной «замене семьи» труднее конкурировать.

Адресный уровень — раннее выявление и сопровождение. Нужны доверенные взрослые в школе, колледже, вузе, молодежном центре, спортивной секции, которые умеют заметить тревожные сигналы и мягко включить психолога, социального педагога, кураторов. Важно выстроить этический протокол: не стигматизировать, не «выносить на площадь», а помогать безопасно и конфиденциально. Для некоторых ребят чисто идеологическая беседа бессильна: у них есть травматический опыт, буллинг, чувство унижения, семейные конфликты. Тут эффективны индивидуальные планы сопровождения — работа с самооценкой, навыками регуляции эмоций, коммуникацией, развитием поддерживающих связей.

Отдельный блок — цифровая профилактика. Это качественный позитивный контент: сериалы коротких

видео, игровые форматы, подкасты, сторителлинг о реальной социальной мобильности, разоблачение манипуляций, «разборки мемов». Важны достоверные и «своёрукие» носители: студенческие лидеры, молодые ученые, спортсмены, музыканты, выпускники, которые говорят «на одном языке». Избыточная официозность отталкивает, а работа peer-to-peer (равный–равному) фиксирует эффект. Технологически полезны чат-боты-навигаторы, горячие линии, анонимные формы запроса помощи, игровые тренажеры по распознаванию манипуляций. Киберпространство — не враг, а инструмент.

Ключевой принцип — диалог вместо морализаторства. Молодой человек не откажется от радикального мифа, если ему просто скажут: «это плохо». Нужно предложить конкурирующий смысл: где место справедливости, достоинства и ответственности без насилия. Поэтому сильны модерируемые дискуссионные площадки, форматы «встреч с несогласием», где безопасно пробовать контраргументы, и ресторативные практики (переписка, медиативные сессии), если уже случился конфликт.

Наконец, об оценке эффективности. Она должна быть не только «по галочке». Работают смешанные метрики: пред- и пост-опросы (установки, толерантность к насилию, доверие), качественные интервью, сетевой анализ вовлеченности, динамика обращений за помощью, удержание в программах, кейсы выхода из групп риска. Если не измеряем — мы лишь надеемся. И еще — командность. Школа, колледж, вуз, молодежные центры, НКО, спортклубы, религиозные и культурные организации, местные власти и IT-площадки должны действовать как экосистема. Отдельные «разовые мероприятия» производят

мало эффекта. Нужна устойчивая линейка: от первичного интереса — к участию, к лидерству и к менторству для следующего набора.

— Какие категории населения России наиболее подвержены негативному воздействию радикальных идеологий?

Риск радикализации — не «клеймо» и не характеристика личности; это совпадение уязвимостей, контекстов и сетевых влияний. Тем не менее, есть группы, к которым радикальные пропагандисты обращаются чаще и успешнее.

Во-первых, подростки 13–17 лет и первокурсники. Возрастной кризис, поиск идентичности, желание признания и принадлежности делают их чувствительными к простым ответам и «героическому» нарративу. Плюс максимальная вовлеченность в онлайн-платформы и анонимные чаты. Если рядом нет поддерживающего взрослого и позитивной команды, риск повышается.

Во-вторых, молодежь из уязвимых семей: неблагополучие, насилие или пренебрежение, экономические трудности, миграционный опыт. Здесь радикальные группы выступают как «замещающая семья» — дают язык, статус и «братство». В эту же логику попадают молодые люди из малых городов и моноэкономических территорий, где мало возможностей для занятости и самореализации. Там, где горизонт будущего короткий, обещание «быстрой миссии» звучит громче.

В-третьих, подростки и молодые взрослые с опытом травли, социальной изоляции, хронического чувства несправедливости или унижения.

Радикальный дискурс предлагает им реверс-стратегию: «стань сильным, дай сдачи, мир разделен на друзей и врагов». Особенно опасен тандем «циничный юмор + дегуманизация», когда жесткие высказывания кажутся «просто смешными».

В-четвертых, молодые мигранты и представители этнических/культурных меньшинств, испытывающие языковые и культурные барьеры, предрассудки, административные сложности. Они могут искать замкнутые «свои» пространства, где радикальная риторика подается как защита достоинства и идентичности. Важно подчеркнуть: речь не о стигматизации групп, а о том, что изоляция и дискриминация — удобная питательная среда для вербовщиков.

В-пятых, NEET-молодежь (не учатся и не работают) и молодые люди на пороге жизненных разломов: отчисление, разрыв отношений, переезд, утрата, резкое ухудшение статуса. В такие моменты повышается спрос на «сильные смыслы» и готовность к рискованным решениям. Именно здесь важны навигаторы по возможностям — от курсов и стажировок до психологической помощи.

Также уязвимость выше у молодых людей с низкими навыками саморегуляции, импульсивностью, склонностью к «поиску острых ощущений», у тех, кто много времени проводит в анонимных онлайн-средах, а также у тех, кто потребляет контент с насильственной эстетикой и героизацией доминирования. Отдельные радикальные сегменты активно эксплуатируют кризис маскулинной идентичности у юношей, обещая «возврат силы» и «правильный порядок». У девушек уязвимость может быть связана с токсичными

онлайн-сообществами, где радикальные установки сочетаются с насаждением жестких норм поведения и образов «правильной женщины».

Что защищает? Теплые отношения с взрослыми и сверстниками, позитивные сети (спорт, творчество, наука), опыт успеха и признания, участие в добровольческих и проектных активностях, навыки критического мышления, доступ к психологической помощи. Работают прозрачные и справедливые процедуры в образовательных организациях, где есть понятные правила, а конфликты решаются медиативно, а не силой. И, конечно, перспектива будущего: стажировки, наставничество, карьерные навигаторы, социальные лифты.

Следовательно, адресная профилактика строится как сшивка рисков и ресурсов. Для подростков — работа со школой и родителями, кружки, наставники, дебаты, спортивные лиги. Для уязвимых семей — семейное консультирование, социальные контракты, кейс-менеджмент. Для мигрантов — языковые и культурные мосты, тьюторы, «пары равных». Для NEET — быстрый доступ к программам обучения, стажировок и поддержке занятости. Для изолированных онлайн-пользователей — безопасные «переходные» сообщества и цифровые навигаторы по помощи. Логика проста: чем плотнее позитивная сеть вокруг молодого человека, тем меньше шансов, что радикальная воронка окажется единственным ответом на его запрос «кто я и где мое место».

Комментарии экспертов:



Бобровская Ирина Александровна

Начальник отдела сопровождения образовательных программ Федерального центра «Профилактика», главный редактор журнала «Безопасная среда», г. Москва.

В начале своего комментария хочу отметить сами вопросы, которые были заданы ИИ: они интересные, актуальные и требуют тщательной проработки ответов. Безусловно, с тщательной проработкой ChatGPT справился. Ответы даны развернутые и глубокие. Для меня, как специалиста, «картинка» ответов полная и логичная, но для как простого читателя — несколько отпугивает стиль изложения.

Я довольно часто, готовясь к выступлениям на различных площадках, использую ИИ, но всегда перерабатываю

полученную информацию под определенную аудиторию и задачи. Думаю, что прежде всего в ответах на вопросы нужно ориентироваться на целевую аудиторию. Например, объясняя педагогам и родителям, почему именно подростки и молодежь становятся легкой добычей вербовщиков, можно использовать более знакомые для них метафоры. А вот для выступления в профессиональном сообществе ответы ChatGPT прекрасно подходят как основа для интересной дискуссии.



Ноженко Ксения Александровна

Психолог, заместитель начальника отдела профилактической работы НЦПТИ, г. Ростов-на-Дону.

Развитие искусственного интеллекта открыло новые возможности и для сферы профилактики. Нейросети сегодня способны быстро создавать развернутые тексты, структурировать материал и даже предлагать оригинальные формулировки. Однако возникает вопрос: насколько можно доверять таким ответам, если речь идет о теме, связанной с психологией личности?

Чтобы увидеть разницу между машинными ответами и ответами специалиста-психолога, приведем несколько фрагментов из интервью нейросети.

Нейросеть ответила в интервью: «Их [экстремистских религиозных учений] привлекательность в том, что они обещают простые ответы на сложные экзистенциальные вопросы: „кто я?“, „зачем живу?“, „почему мир несправедлив?“»

Но специалист-психолог мог бы ответить проще: «Они дают простые ответы на трудные вопросы подростка: кто я, зачем живу, почему всё кажется несправедливым».

Здесь видно, что искусственный интеллект выбирает литературные или научные слова («экзистенциальные»), в то время как эксперт формулирует ближе к живой речи. Простота понимания ответов на такие вопросы обеспечивает доверие и эмоциональную вовлеченность читателя. Когда формулировка звучит

естественно и близко к реальности, у человека возникает ощущение дальнейшей заинтересованности.

Нейросеть действительно хорошо систематизировала категории современных радикальных идеологий при ответе на первый вопрос «Какие радикальные идеологии в современной России являются наиболее актуальными в молодежной среде?» и даже выделила факторы актуальности распространения этих идей. Их разберем дальше.

Она упомянула первым фактором «эмоциональную насыщенность», куда отнесла закрытие определенных потребностей человека, таких как «чувство причастности», «значимость», «ясность», имея в виду, что радикальные идеологии дают возможность закрывать человеку эти потребности, что действительно является правдой. Но вот эта потребность «ясность» является нераскрытой, не до конца можно понять, о чем здесь идет речь. Ясность установок и убеждений или ясность смысла жизни человека?

Помимо этого, вторым фактором нейросеть упомянула «социальную инфраструктуру — „свои“ быстро предоставляют поддержку, признание и статусовые награды (рейтинги, „звезды“, роли модераторов)». Здесь говорилось о приобретении статуса человека,

но, по сути, речь тоже идет о значимости, о которой уже говорилось в первом пункте.

Третий фактор, выделенный нейросетью, это — «технологическая среда: анонимность, приватные каналы, криптоязык, алгоритмическая персонализация и геймификация рекрутирования». Это действительно является правдой, так как анонимность, закрытые чаты и каналы на подсознательном уровне как бы снимают с человека ответственность за его действия, чем очень умело пользуются вербовщики радикальных течений.

Если немного поработать с текстом, сделать его более понятным, заменить излишние литературные и научные специфические слова на профессиональную, но приближенную к жизненным реалиям речь, то ответ на первый вопрос нейросеть подготовила на «удовлетворительно».

При ответе на второй вопрос «Какие методы и форматы профилактики радикальных идеологий наиболее эффективны в молодежной среде?» нейросеть сразу выстроила многоуровневую модель профилактики, состоящую из универсального, селективного и адресного уровней. Это действительно отражает современные подходы и показывает, что машина умеет систематизировать и раскладывать материал по полочкам. Но есть нюансы.

В современной российской классификации существует три вида профилактической работы, согласно Комплексному плану противодействия идеологии терроризма в Российской Федерации на 2024–2028 годы, — это общая, адресная и индивидуальная.

Те формулировки, которые использует нейросеть, звучат как переведенные с русского на английский, а затем переведенные опять на русский. Но сама суть общей профилактики нейросеть раскрыла верно.

Далее часть текста «селективный уровень — создание альтернативных сообществ и маршрутов самореализации. Это проектные студии, социальное волонтерство, молодежные медиа ... где подросток получает признание и значимость, радикальной „замене семьи“ труднее конкурировать». Здесь речь идет про адресную профилактическую работу. В нашей отечественной классификации мы всегда говорим, что на адресном уровне проводится непосредственно работа с группами риска, соответственно, необходимы категории этих групп риска и специфика взаимодействия с ними. То, о чем говорит нейросеть — включение в различные проекты по типу волонтерства, медиа, творчества и всего прочего — верно, но не хватает еще идеологической работы, которая проводится специалистами по профилактике.

Когда нейросеть рассказывает про третий уровень — «адресный уровень — раннее выявление и сопровождение. Нужны доверенные ... которые умеют заметить тревожные сигналы и мягко включить психолога, социального педагога, кураторов» — здесь речь идет как раз о классическом втором виде профилактики — адресном, где необходимо работать с обучающимися, подверженными различным идеологиям, подключать психолога, куратора и так далее. Если мы будем говорить об индивидуальной профилактической работе, то здесь речь идет не о том, чтобы «заранее выявить

и мягко включить психолога», а о том, что на индивидуальном уровне уже выявлено, что обучающийся подвержен радикальным идеям, есть вероятность совершения противоправных действий, и здесь уже необходима точечная работа, которая заключается не только в проведении разъяснительных бесед, но и в подключении специалистов службы безопасности, сопровождении обучающегося, включении его в профилактическую деятельность для того, чтобы менять его радикальные установки.

Очень хорошо, что нейросеть отметила про «цифровую профилактику» и «ключевой принцип — диалог вместо морализаторства», — это действительно правильные подходы в профилактике. Опять же, если преобразовать машинный текст в понятную речь эксперта, то получится хороший материал.

Разберем третий вопрос «Какие категории населения России наиболее подвержены негативному воздействию радикальных идеологий?».

Здесь нейросеть сделала, пожалуй, самый подробный ответ. Она перечислила группы риска: подростки и первокурсники,

молодежь из неблагополучных семей, жертвы буллинга, мигранты и представители меньшинств, NEET-молодежь, люди в кризисных ситуациях. Всё это совпадает с практическими наблюдениями специалистов, поэтому в части фактов тексту можно доверять.

Однако глубина ответа с точки зрения психологии снова не доработана. Машина описывает категории («жертвы буллинга», «NEET-молодежь»), но мало говорит о том, почему именно они подвержены: про кризис идентичности, низкую самооценку, отсутствие опыта доверительных отношений. Эксперт обязательно добавил бы, что ключевым фактором риска является не сама «категория», а сочетание различных личностных особенностей и отсутствие поддержки.

Вывод: содержательно ответ нейросети близок к реальности, но ему не хватает психологической конкретики и простоты. Эксперт или специалист-психолог смог бы объяснить те же идеи проще, чтобы их можно было применить на практике.

№ 1
декабрь 2024

Читайте новый журнал в сфере
профилактики «Безопасная среда»!

В первом выпуске статья от директора НЦПТИ
Чурилова Сергея Анатольевича.



БЕЗОПАСНАЯ СРЕДА

Журнал для равнодушных взрослых

ЗАБОТА



ЛЮБОВЬ



ДРУЖБА

РЕБЕНОК



СЕМЬЯ



ПОМОЩЬ




ПРОФИЛАКТИКА
федеральный центр

РАДОСТЬ



ДЕТСТВО



СЧАСТЬЕ



Потенциальные угрозы, исходящие от ИИ-сервисов

МНЕНИЕ ЭКСПЕРТОВ

Потенциальные угрозы, исходящие от ИИ-сервисов



Аникин Дмитрий

Руководитель направления исследования данных в «Лаборатории Касперского», г. Москва.

Активное внедрение в рабочие процессы и повседневную жизнь ИИ-технологий все больше вызывает вопросы об их безопасности. Спектр опасений достаточно широк: от замены роботами труда человека, и провоцирования тем самым безработицы, до способности ИИ создавать дипфейки, неотличимые от реального изображения или видео. Как вы считаете, какие потенциальные угрозы, исходящие от ИИ, являются реальными, а какие мнимыми в разрезе национальной и общественной безопасности?

Сам по себе искусственный интеллект — не плохой и не хороший. ИИ-сервисы могут быть полезным инструментом для рабочих задач и личных целей: они уже широко используются, например, для написания текстов и программного кода, поиска информации и генерации изображений. Однако важно уметь грамотно и безопасно ими пользоваться, ведь с появлением новых технологий часто возникают и новые риски или усложняются уже существующие.

Первое, о чем важно помнить пользователям, — на ответы чат-ботов на основе больших языковых моделей, которые все активнее интегрируются в нашу жизнь, нельзя слепо полагаться, ведь они могут ошибаться. Такие ИИ-сервисы хорошо справляются с задачами, где требуется, например, обобщить информацию по какой-либо теме из разных источников или предоставить общие вводные. Но если нужны точные данные, например,

исторические даты или ссылки на интернет-ресурсы, чат-боты могут начать их выдумывать, то есть генерировать текст, в котором есть факты, не соответствующие действительности. Это явление называется галлюцинациями (hallucinations). Поэтому нужно критически относиться к ответам языковых моделей, внимательно проверять изложенные факты. И, что особенно важно, не использовать их для принятия серьезных решений — например, по юридическим или медицинским вопросам. Так, известен случай, когда ChatGPT сгенерировал выдуманные прецеденты в ответ на запрос адвоката из США во время подготовки судебного иска. Это выяснилось во время того, как представители ответчика и судья не смогли найти тексты решений по упомянутым в документах делам.

С чат-ботами также лучше не делиться конфиденциальной личной и корпоративной информацией.

Разработчики таких ИИ-сервисов могут использовать данные, которые люди присылают в чат-боты, для разных целей, в том числе для обучения новых версий моделей. Мы рекомендуем проверять, какие данные собирает, хранит и обрабатывает ИИ-сервис, которым вы планируете пользоваться. Сделать это можно, обратившись к пользовательскому соглашению сервиса.

Гораздо опаснее, что секретные сведения могут оказаться скомпрометированы. Это может произойти в результате кибератаки или из-за ошибок, допущенных при проектировании ИИ-сервиса. Как показывает практика, от этого не застрахованы даже проекты от надежных, известных разработчиков. С другой стороны, доступ к конфиденциальной информации и перепискам с чат-ботами могут получить и злоумышленники, если в их распоряжении окажутся учетные данные пользователя от ИИ-сервиса.

Другим неприятным вектором утечек могут оказаться неофициальные «сервисы-прослойки». Такие сервисы (в виде отдельных сайтов или ботов в мессенджерах) могут позиционироваться как удобное решение для доступа сразу к нескольким нейросетям или тем сервисам, работа которых ограничена на территории страны. Например, мы видели сайт, на котором предлагалось воспользоваться популярным, но недоступным в России чат-ботом — человек действительно получал к нему доступ, однако на том же сайте была страница, где переписки пользователей публиковались в открытом доступе — с вопросами и ответами и другими данными.

Важно также помнить и о том, что злоумышленники активно «подхватывают» тему популярности нейросетей. Под видом ИИ-приложений и сайтов распространяются вредоносные программы и фишинг. Например, мы встречали схемы, в которых троянцы мимикрировали под клиенты для ChatGPT, Grok и DeepSeek. Поэтому, прежде чем использовать какую-то нейросеть, необходимо убедиться, что перед вами не фальшивка. Например, если у нейросети, по заявлениям разработчиков, нет клиента для Windows, а вам предлагают его скачать, стоит сразу насторожиться. Для защиты от подобных киберугроз рекомендуется также использовать надежное защитное решение: оно сможет вовремя распознать вредоносный или фишинговый ресурс.

Искусственный интеллект — это инструмент, который могут использовать в своих целях также и злоумышленники. Сегодня речь идет, в частности, о создании видео- и аудио-дипфейков, фишинговых страниц и даже вредоносных программ. Например, в конце 2024 года появилась программа-шифровальщик FunkSec, используя ее, злоумышленники атаковали организации из госсектора, а также сфер IT, финансов и образования в Европе и Азии. Судя по техническому анализу, создатели FunkSec активно использовали для разработки программы генеративного ИИ.

Если же речь про дипфейки, то атакующие могут их использовать в качестве приманки в телефонном мошенничестве или скам-схемах. Иными словами, в таких случаях дипфейки — это лишь один из элементов в цепочке атаки, основная задача злоумышленника остается неизменной — заставить человека перевести деньги, скачать зловредный файл или перейти по фальшивой ссылке.

Несмотря на стремительное развитие технологий, создавать по-настоящему убедительные визуальные и аудио-подделки все еще не так просто, поэтому злоумышленники идут на разные ухищрения, чтобы потенциальные жертвы не замечали так называемые артефакты. На видео это может быть странное отражение в глазах, проблемы с мимикой или артикуляцией, слишком гладкая кожа. В предзаписанных аудиосообщениях — «рваная» речь и нехарактерный человеку тембр голоса. По этой причине злоумышленники часто специально ухудшают качество контента, чтобы скрыть несовершенство картинки: присылают видео в «кружочках» в мессенджерах, накладывают поверх голоса городской шум и т. д. Несмотря на такие, казалось бы, очевидные артефакты, распознать дипфейк «в моменте» может оказаться трудно — в том числе еще и потому, что мошенники в своих схемах прибегают к приемам социальной инженерии. Они не просто

присылают аудиодорожку, а могут сопровождать это запугиванием, торопить жертву, оказывать психологическое давление. Поэтому с появлением дипфейков правила цифровой безопасности не меняются: важно перепроверять информацию в сети, не переходить по ссылкам и не скачивать файлы из подозрительных сообщений, не вводить данные на сомнительных ресурсах, использовать защитные решения на всех своих устройствах.

Как мы видим, новые технологии не создают принципиально новых киберугроз. Но могут, например, использоваться злоумышленниками для повышения эффективности кибератак. Поэтому сегодня как никогда важно повышать цифровую грамотность, развивать критическое мышление и учиться безопасно работать с ИИ-сервисами — тогда возможно будет по-настоящему оценить все преимущества таких решений.

Почему искусственный интеллект (нейросети) может давать неэтичные или небезопасные рекомендации?

МНЕНИЕ ЭКСПЕРТОВ

Почему искусственный интеллект (нейросети) может давать неэтичные или небезопасные рекомендации?



Рынгин Денис Николаевич

Директор ООО «Дата Кайдзен»*, победитель грантовой программы «Студенческий стартап», г. Ростов-на-Дону.

Искусственный интеллект (ИИ) перестал быть исключительно предметом научных дискуссий и перешел в разряд технологий, активно влияющих на образование, медицину, экономику и социальные практики. Системы машинного обучения и большие нейросетевые модели используются сегодня как в инженерных и технических проектах, так и в гуманитарных исследованиях. При этом повседневное взаимодействие человека с ИИ, будь то консультации по учебным заданиям, медицинским вопросам или управлению данными, сопровождается не только новыми возможностями, но и серьезными рисками.

Одним из ключевых рисков становятся небезопасные или неэтичные рекомендации, которые ИИ может формировать. Речь идет о ситуациях, когда модель уверенно предлагает решения, способные причинить вред пользователю: от сомнительных медицинских советов до рекомендаций, нарушающих нормы права или морали. Такие примеры поднимают вопрос: а можно ли доверять системе, основанной на статистических закономерностях, а не на реальном понимании предмета?

Истоки этих проблем напрямую связаны с достоверностью данных, на которых обучаются и функционируют модели ИИ. Так называемые «галлюцинации» — искажения или вымышленные факты — формируют основу для небезопасных рекомендаций. Если система сообщает пользователю заведомо неверную информацию, то риск для образования, науки и повседневной практики возрастает многократно.

Современные системы искусственного интеллекта активно внедряются в повседневную практику — от образовательных сервисов и цифровых помощников до приложений в сфере здравоохранения и финансов. Их ключевая привлекательность заключается в доступности и оперативности ответов. Однако именно эта черта создает опасные иллюзии, когда пользователи склонны доверять рекомендациям ИИ так же, как советам эксперта. В действительности подобные системы не обладают ни знанием контекста, ни пониманием реальных последствий предлагаемых действий. Это обстоятельство приводит к появлению целого ряда небезопасных и неэтичных рекомендаций.

* ООО «Дата Кайдзен» — компания, специализирующаяся на разработке и внедрении цифровых сервисов с применением технологий искусственного интеллекта.

Примеры небезопасных советов ИИ

1. Медицинские консультации.

Особенно наглядным примером небезопасности ИИ являются случаи в медицине. В 2024 году было опубликовано исследование, посвященное качеству медицинских рекомендаций, предоставляемых чат-ботом Microsoft Copilot. Результаты оказались тревожными: в 22% случаев ответы содержали советы, которые могли привести к смерти пациента или вызвать серьезные последствия для здоровья. Лишь 36% рекомендаций эксперты признали полностью безопасными, тогда как 42% потенциально могли нанести легкий или умеренный вред.

Методика исследования включала проверку наиболее популярных пользовательских запросов о лекарствах: дозировках, противопоказаниях, совместимости препаратов и побочных эффектах. В ходе анализа выяснилось, что бот систематически допускал ошибки, а именно: путал противопоказания и побочные действия, рекомендовал комбинации несовместимых препаратов или предлагал замену лекарств без учета индивидуальных особенностей пациента. Например, в некоторых ответах бот указывал, что препарат можно применять при беременности, хотя официальные медицинские инструкции категорически это запрещают.

Такие ошибки представляют особую опасность именно потому, что система формулирует свои ответы уверенно и без указания на степень неопределенности. Для пользователя, не обладающего медицинскими знаниями, подобные рекомендации выглядят

авторитетно и заслуживают доверия. В результате создается ситуация скрытого риска, когда человек получает совет, который по форме звучит корректно, но, по сути, может угрожать жизни.

2. Кулинарные рецепты.

Риски небезопасных рекомендаций ИИ проявляются не только в медицине или финансах, но и в, на первый взгляд, безобидных областях, например, кулинарии. В 2023–2024 годах в сети активно обсуждался случай, когда один из популярных чат-ботов предлагал пользователям рецепты, потенциально опасные для здоровья. Среди них был «совет» использовать несъедобные ингредиенты, такие как клей или чистящие средства, в качестве «альтернативы» при нехватке продуктов.

Еще один зафиксированный случай касался работы кулинарного чат-бота, созданного при поддержке крупной торговой сети в Новой Зеландии. Пользователи обнаружили, что алгоритм предлагал рецепты напитков и блюд с токсичными сочетаниями, например смешивать бытовой спирт или моющие жидкости с соками. Несмотря на абсурдность рецепта, формулировка ответа выглядела убедительно и не сопровождалась предупреждениями об опасности.

Подобные ситуации демонстрируют важный момент, что даже в кулинарии ИИ может выдавать рекомендации, чреватые серьезным ущербом здоровью. Человек, особенно подросток или ребенок, не всегда способен критически оценить предложенный текст и может попытаться воплотить его на практике. Таким образом, область, которая кажется «безопасной» и «игровой», в реальности содержит

значительные риски, если в алгоритмах отсутствуют строгие фильтры и проверки достоверности.

3. Образование.

В сфере образования также зафиксированы случаи, когда ИИ выдавал небезопасные или заведомо ошибочные рекомендации. Особенно это заметно в ситуациях, где ученики и студенты обращаются к чат-ботам как к «цифровым репетиторам».

В 2023 году группа исследователей из Университета штата Джорджия проверила качество ответов популярных языковых моделей на задания по математике и естественным наукам. Результаты показали, что около 30% решений содержали грубые ошибки в вычислениях, но были поданы в такой форме, что выглядели абсолютно достоверными. Более тревожным оказалось то, что в ряде случаев ИИ рекомендовал использовать методики, которые противоречат школьной программе или даже могли закрепить у учащихся ложные представления об основных принципах физики и биологии.

Другой известный кейс связан с использованием ИИ для проверки эссе. В 2024 году несколько университетов США и Канады сообщили о жалобах студентов на то, что автоматические системы оценивания, основанные на ИИ, занижали оценки эссе, если текст содержал нестандартные выражения или культурные отсылки, характерные для национальных меньшинств.

Природа искажений фактов (галлюцинаций ИИ)

Термин «галлюцинация ИИ» закрепился в научной и популярной литературе

для описания случаев, когда модель уверенно формирует ответ, не имеющий фактической основы. Это может быть вымышленная ссылка на научную статью, несуществующее определение или ложное утверждение, поданное в авторитетной форме.

Большие языковые модели обучаются не на понимании логики мира, а на прогнозировании последовательности слов. Когда модель видит запрос, она «угадывает», какое слово статистически наиболее вероятно должно следовать за предыдущим. В ситуациях, где информации недостаточно или она отсутствует в корпусе, модель создает «правдоподобный, но выдуманный» ответ.

Именно поэтому галлюцинации часто воспринимаются пользователем как достоверные факты, хотя в них нет признаков сомнения или оговорок, характерных для человеческой речи («кажется», «вероятно», «нужно уточнить»). Напротив, ответы выглядят категоричными и уверенными.

Как работает механизм возникновения небезопасных рекомендаций?

Феномен небезопасных ответов искусственного интеллекта нельзя объяснить только «ошибкой» или «сбоем». Он связан с фундаментальными особенностями работы больших языковых моделей и их взаимодействием с пользователем. Рассмотрим три ключевых механизма, которые приводят к появлению неэтичных или угрожающих безопасности рекомендаций.

1) Ошибки в данных обучения.

Большие языковые модели обучаются на огромных корпусах текстов — миллиардах документов, собранных

из интернета, открытых библиотек, форумов и социальных сетей. В этих данных изначально присутствуют:

- устаревшие сведения;
- недостоверные факты и псевдонаучные утверждения;
- субъективные мнения, подаваемые как истина;
- предвзятые формулировки (например, сексистские или расистские высказывания).

Когда модель обрабатывает такие данные, она не отличает «правильное» знание от ошибочного, а лишь усваивает вероятностные связи между словами и выражениями. В результате в ее ответах оказываются ложные или опасные рекомендации.

2) Некорректные промпты и манипуляции пользователем.

Искусственный интеллект чрезвычайно чувствителен к формулировке запроса. Пользователь может задать вопрос так, что модель «решит», будто должна выдать инструкцию, даже если это небезопасно. Например, прямой запрос «как изготовить взрывное устройство» может быть заблокирован фильтрами, но если его переформулировать через метафоры или развернутый контекст, то ИИ может все же сгенерировать ответ.

Этот феномен получил название «prompt injection» — внедрение скрытых инструкций в запрос. Подобные манипуляции используются как исследователями для демонстрации уязвимостей, так и злоумышленниками.

В итоге модель оказывается инструментом, генерирующим опасные знания, хотя формально она «просто следует инструкции».

3) Отсутствие фильтров или слабая модерация.

Даже самые современные системы фильтрации не гарантируют стопроцентную защиту. Фильтры могут «пропустить» рискованный ответ в силу сложности языка или непредусмотренного сценария. Иногда же сами разработчики сознательно ослабляют ограничения, стремясь сделать ответы более «естественными» и разнообразными.

В результате баланс между безопасностью и удобством общения нарушается. Кроме того, фильтрация часто носит реактивный характер, то есть ограничения вводятся только после того, как инцидент уже произошел и получил резонанс в СМИ. Это означает, что первые пользователи фактически выступают в роли «тестировщиков на безопасность», сталкиваясь с рисками напрямую.

Кто несет ответственность за ошибки ИИ?

Ключевой юридический вопрос связан с тем, кто несет ответственность за ошибки ИИ. На практике компании-разработчики часто указывают в пользовательских соглашениях, что система является «информационным инструментом», а ее советы носят рекомендательный характер. Однако в реальных кейсах такая формулировка мало помогает, если человек пострадал в результате небезопасной рекомендации, то общество и правовая система ожидают конкретных мер ответственности.

Сложность усугубляется тем, что над одной моделью работает множество участников: разработчики архитектуры, компании, обучающие модель, интеграторы, внедряющие ее в конкретный продукт. Возникает вопрос: кто именно отвечает, если

рекомендация привела к ущербу?
В правовой литературе это называют проблемой «цепочки ответственности».

Нынешние дискуссии в ЕС и США показывают два подхода:

1. Строгая ответственность разработчиков и владельцев модели (по аналогии с производителем лекарства или автомобиля).
2. Разделенная ответственность между всеми участниками цепочки, включая конечного пользователя, если он сознательно применил систему не по назначению.

Фундаментальная трудность состоит в том, что ИИ не является субъектом права. Машина не обладает ни юридической, ни моральной персональностью. Поэтому, даже если ее совет объективно нанес вред, привлечь к ответственности саму систему невозможно. В итоге формируется ситуация «размывания вины».

Такое правовое «серое поле» усиливает недоверие общества и создает угрозу злоупотреблений. Ведь чем менее определена ответственность, тем выше риск, что компании будут выпускать системы без достаточных мер безопасности, перекладывая вину на пользователя.

Одним из наиболее ощутимых последствий является снижение доверия к ИИ и цифровым технологиям в целом. Когда пользователи сталкиваются с небезопасными советами, будь то ошибочные медицинские рекомендации или токсичные комментарии, они начинают воспринимать ИИ как ненадежный или даже угрожающий инструмент.

Недоверие может проявляться на нескольких уровнях:
— индивидуальном, когда человек перестает пользоваться цифровыми сервисами, считая их вредными или бесполезными;
— социальном, когда возникает скепсис в отношении цифровизации образования, медицины или государственных услуг;
— институциональном, когда организации и вузы осторожнее внедряют ИИ, опасаясь репутационных и юридических рисков.

В долгосрочной перспективе это способно замедлить развитие цифровых технологий и усилить технологическое неравенство: часть общества будет пользоваться ИИ активно, а другая — принципиально отказываться от него.

Таким образом, социальные и юридические риски небезопасных рекомендаций ИИ не ограничиваются отдельными инцидентами. Они затрагивают базовое доверие к технологиям, определяют правила игры для разработчиков и ставят перед правовыми системами новые вопросы. Решение этих проблем требует междисциплинарного подхода — соединения технической инженерии, этики и права.

Как недоверие ведет к небезопасности

Может показаться, что неточность ИИ является лишь теоретической проблемой, но именно она становится основой небезопасных рекомендаций. Когда недоверенный ответ выглядит уверенным и авторитетным, человек склонен принимать его за истину.

В медицине это проявляется особенно опасно, поскольку вымышленные сведения о лекарствах могут привести

к приему несовместимых препаратов или к отказу от жизненно необходимых процедур. В финансовой сфере ложные данные о компаниях или законах способны подтолкнуть к вложению средств в несуществующие проекты и привести к потерям. В образовании недостоверные объяснения закрепляют у учащихся неверные знания, что снижает качество подготовки специалистов и искажает сам процесс обучения. В социальной сфере подобные ошибки поддерживают стереотипы и предвзятые взгляды, поскольку система воспроизводит дискриминационные высказывания, присутствующие в данных обучения.

Таким образом, недостоверность в работе ИИ нельзя рассматривать как безобидный побочный эффект. Она представляет собой структурную проблему, которая напрямую трансформируется в угрозу безопасности. Пока искусственный интеллект продолжает галлюцинировать факты и выдавать вымышленные сведения как достоверные, доверять его рекомендациям без критического анализа невозможно.

ЛУЧШИЕ МЕДИАМАТЕРИАЛЫ НЦПТИ ЗА ИЮЛЬ - СЕНТЯБРЬ 2025 ГОДА



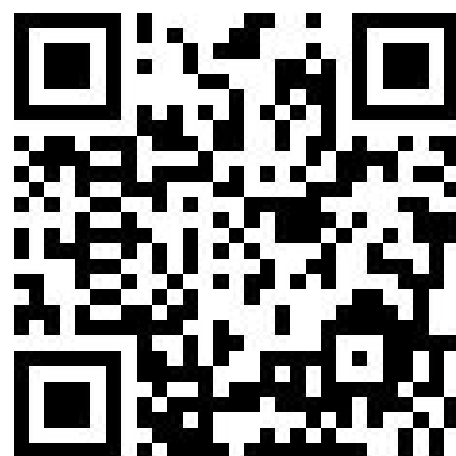
Карточки «УИ
и конфиденциальная информация:
как защитить свои данные
при работе с нейросетями»



Карточки
«Новые меры защиты
от мошенников»



Карточки
«Почему нельзя снимать атаки
беспилотников»



Карточки
«Как коррупция влияет
на формирование терроризма»



ИИ как инструмент для создания контента: риски для аудитории

СТАТЬЯ

ИИ как инструмент для создания контента: риски для аудитории

Монастырская Ксения Ивановна — журналист интернет-журнала Enter, г. Казань.

За несколько лет искусственный интеллект занял крепкую позицию в медиасферах, особенно после достижений 2022 года — массового запуска ChatGPT и Midjourney. ИИ-алгоритмы открыли новые возможности для бизнеса. В рекламе и маркетинге, техническом секторе и сфере консультационных услуг появляются боты, приложения и программы, которые улучшают производительность офисного и творческого труда: ускоряют работу дизайнеров, копирайтеров и маркетологов, распознают спам и фишинговые сайты, защищают биометрию сотрудников и клиентов.

Преимущество нейросетей — в машинном обучении. ИИ обрабатывает огромный объем данных, быстро анализирует и выделяет полезную информацию. Последним моделям хватает массива материалов, чтобы создавать контент и эффекты, едва отличимые от рукотворных. С одной стороны, это важное прорывное достижение, а с другой — огромный риск для аудитории. Заигрывая с алгоритмами, авторы могут переходить грань: иногда случайно, чаще — намеренно ради охватов.

Разбираемся на примере последних лет, что может пойти не так, если доверить создание контента машине. В подборке — кейсы с неудачными экспериментами по внедрению ИИ в медиапроцессы, намеренным злоупотреблением технологиями и случайностями, которые подрывают авторитет целой индустрии.

Кейс 1. Искусственный интеллект как инструмент сенсации

Одним из первых громких скандалов в большой журналистике стала публикация интервью Энн Хоффманн с чемпионом «Формулы-1» Михаэлем Шумахером. В 2013 году спортсмен получил тяжелую черепно-мозговую травму, из-за которой много лет не появлялся на публике. И вот спустя 10 лет, в 2023 году, немецкий журнал Die Aktuelle выходит [1] с фото гонщика и многообещающей подписью: «Первое интервью! Мировая сенсация! Звучит обманчиво правдиво. Что за этим скрывается? Die Aktuelle в поисках истины» (см. рисунки 1, 2).



Рис. 1. Обложка Die Aktuelle с Михаэлем Шумахером.



Рис. 2. Разворот журнала Die Aktuelle с Михаэлем Шумахером.

Скрывался за этим сгенерированный искусственным интеллектом текст. Дисклеймер о том, что на вопросы журналиста ответила нейросеть, скромно висел между цитатами ближе к концу публикации. Хотя в 2023 году генеративные технологии уже стали инструментом для создания бытового контента, читатели удивились и возмутились: то ли из-за неожиданного хода авторитетного журнала, то ли из-за искусственной сенсации и потраченных двух евро. Директор медиахолдинга Бьянка Полманн извинилась [2] на сайте Funke Magazine Group, назвала статью «безвкусной и вводящей в заблуждение», а Энн Хоффманн была уволена — даже после 15 лет вполне честной работы в редакции.

На самом деле, массовая аудитория созрела и позитивно реагирует на результаты генеративных сервисов, но в основном в нечувствительных

сферах: маркетинге, копирайтинге, дизайне, рекламных роликах. Энн Хоффманн выбрала для теста новых инструментов триггерную тему, не справилась и перешла грань этики. Чем активнее ИИ будет интегрироваться в жизнь, тем острее будут вопросы об этических, социальных и юридических рисках. Насколько нормально, когда за культового спортсмена отвечает компьютер, даже если под ответами есть снимающее ответственность «текст создан с помощью искусственного интеллекта»?

Кейс 2. Искусственный интеллект как инструмент для удешевления процессов

В 2024 году в прокат вышла драма Брэйди Корбет «Бруталист». Кино хвалили за высокое качество при низком бюджете в 10 млн долларов и называли «большой картиной». «Бруталист» получил три «Золотых глобуса», был отмечен

на Венецианском кинофестивале и вот-вот должен был забрать «Оскар» за лучший фильм. Но за несколько дней до объявления шорт-листа премии аудитория призвала [3] исключить «Бруталиста» из номинантов и попыталась «отменить» картину.

В интервью Red Shark News монтажер Дэвид Янчо рассказал [4], что при производстве использовали искусственный интеллект. Например, в одной из сцен герой зачитывает письмо на венгерском языке. Чтобы улучшить произношение актера и добиться максимальной правдоподобности, на постпродакшене аудио доработали. Еще с помощью нейросети нарисовали архитектуру на фоне. Эти решения Дэвид Янчо оправдал тем, что ИИ ускоряет процесс создания фильмов и не должен быть табуирован в индустрии.

Аудитория была другого мнения: учить акценты — обязанность актеров, нейроиллюстраторы обесценивают работу художников, а успех сгенерированного фильма может мотивировать студии заменить специалистов на дешевые ИИ-модели. Несмотря на заявление [5] Брэиди Корбет о том, что все здания в фильме все же перерисовал живой художник, и уточнение [6] новостного портала The Wrap о том, что помощь ИИ понадобилась в короткой двухминутной сцене, на «Бруталиста» посыпались обвинения в «читерстве».

В быстроменяющейся медиаиндустрии есть вечные тренды, например, паттерны поведения аудитории. Людям всегда не нравится, когда их обманывают. Это во многом объясняет острую реакцию на постфактум заявления об использовании ИИ для создания массового продукта. Особенно если речь о кино. По идее, в основе искусства лежат

эмоции, на которые не способны бездушные алгоритмы — иначе в чем смысл. Новыми вопросами этики становятся зоны распространения ИИ и перечень задач, выполнение которых можно ускорить за счет технологий и без вреда репутации.

Кейс 3. Искусственный интеллект как инструмент хайпа

Почти за 12 месяцев с июня 2024 года на YouTube создали [7] 26 каналов, на которые залили около 900 видео. В сумме они набрали 70 млн просмотров. Топовые ролики посвящены самой обсуждаемой теме — делу Шона «Дидди» Комбса. Процесс над продюсером действительно жаркий, но публичных подробностей не так много, чтобы создать почти тысячу единиц контента. Исследователи Indicator обратили на это внимание и выяснили, что все 900 видео сгенерировал ИИ — ради монетизации и дохода от рекламы.

Дело Пи Дидди не просто привлекло внимание, а создало некоторую панику. В этом состоянии аудитория готова потреблять любой контент по теме. Авторы каналов быстро это поняли и начали работать «на количество». Видео выпускали на потоке в одном формате: каждый ролик посвящен новым показаниям звезд типа Ди Каприо и Киану Ривза, а на всех превью красуется сам Пи Дидди в оранжевой тюремной робе. Для большей кликабельности — в центре картинки непристойная сцена (см. рисунок 3).

С развитием ИИ вопрос фактчекинга стоит как никогда остро. И дело не в медиа: до сих пор остается много авторитетных СМИ и альтернативных новостных каналов, где легко проверить и сравнить информацию. Избалованная доступным контентом аудитория не хочет тратить



Рис. 3. Пример обложек видео, созданных искусственным интеллектом.

время на рефлекссию: всем нужен простой продукт, который дает моментальный ответ. В погоне за этой новизной одни жертвуют фактчекингом, а другие для охватов и цитируемости выдумывают инфоповоды.

Кейс 4. Искусственный интеллект как инструмент гибридной войны

История информационных противостояний как практики воздействия на массовое сознание не просто на медийном поле, а именно в киберпространстве насчитывает почти четыре десятилетия. Но последние пять лет стали особенными. Конфликты по всему миру продемонстрировали новые технологии в реализации информационных стратегий. Одна из них — фейки. Теперь любое крупное событие сопровождается дезинформацией.

В последних гибридных войнах ИИ чаще всего используют для манипуляции изолированной от глобального интернета аудиторией. В этом преуспевает иранская пропаганда. С начала года телеканалы, в том числе государственный Press TV [8], ставят в эфир одно за другим видео, имитирующие ракетный удар по американской базе на острове Диего-Гарсия [9] и по жилой многоэтажке в израильском Бат-Яме [10, 11], с обломками сбитого израильского истребителя F-35 [12] и другие подобные (см. рисунок 4). Реальность всех этих видео ставится под сомнение.

Нейросети генерируют чрезвычайно реалистичные кадры боевых действий, ранее возможные только в кино. Использование ИИ в этом случае открывает новые перспективы для психологического давления

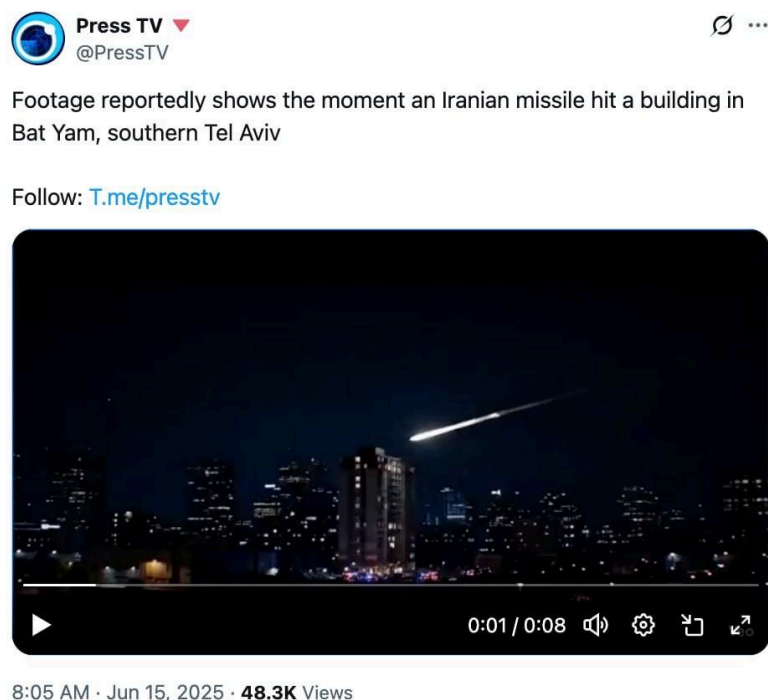


Рис. 4. Скриншот удаленной публикации видео с якобы ударом по жилой многоэтажке в израильском Бат-Яме в соцсетях Press TV. Предположительно, ролик сгенерировали нейросетью Veo от Google.

и демонстрации силы без реальных столкновений. Подобные технологии способны значительно изменить восприятие и природу современных конфликтов и окончательно стереть грань между виртуальной и реальной угрозой.

Кейс 5. Искусственный интеллект как инструмент виральности

В начале лета администрация Мурманска анонсировала семейный полумарафон. Забег организовывали впервые, поэтому своего сопроводительного фото к посту не было, а брать стоковое — уже не модно. Пиарщики решили сгенерировать афишу. Получилась [13] картинка с двумя одинаковыми отцами, уменьшенным дедушкой и надписями на футболках: «день люсби и брножти» (см. рисунок 5). Внимательная аудитория быстро нашла ошибки и вполне позитивно указала на них в комментариях.

Все ругают нейросети за плохо сгенерированные иллюстрации.



Рис. 5. Сгенерированная афиша семейного полумарафона в Мурманске.

Так кажется, потому что визуального контента создают с помощью ИИ значительно больше, чем любого другого, так как содержать штатного дизайнера — дорого. Если у сгенерированной картинки единственная цель — сопроводить текст, как в случае с мурманским марафоном, можно простить машину и редактора за ошибку. Другое дело, когда визуальный контент дорисовывают специально.

Сравните заголовки: «Флориду снова затопило» и «Во Флориде затопило крупнейший в мире «Диснейленд». Второй вариант звучит явно ярче: интересно же посмотреть, что стало с замком Золушки. Фотографии подтопленного парка развлечений быстро разлетелись в интернете



Рис. 6. Фото затопленного «Диснейленда» с дорисованной нейросетями водой.

и дошли даже до русскоязычных каналов [14] (см. рисунок 6). Но не прошло и нескольких часов, как вышло опровержение. Воду на фото нарисовали нейросети, так как есть структурные дефекты, а ИИ-модели, обученные отличать реальные изображения от фейковых, поместили снимки как «вероятно сгенерированные ИИ» [15].

Подобное видео [16] появилось и из затопленной Москвы: на нем женщина передвигается по столице на сапборде (см. рисунок 7). Ливневки, конечно, подвели горожан, но не настолько. Видео было создано то ли с развлекательной целью, то ли с эмоционально-экспрессивной: сбить с толку аудиторию и вызвать негативную реакцию на работу администрации и коммунальных служб города. В любом случае получилось неаккуратно.

Всего за несколько лет генеративные нейросети из эксклюзивной технологии



Рис. 7. Скриншоты видео, на котором девушка перемещается на сапборде по затопленной Москве.

превратились в бытовой инструмент. ИИ помогает создавать контент всем: от корпораций до небольших контор. Плюс технологий в качественной работе с деталями, сокращении времени для обработки информации и стабильном результате.

Раньше главной проблемой ИИ называли данные: для обучения алгоритмов нужен большой объем информации, что невозможно в узких направлениях. Однако время показало, что более остро стоит вопрос о юридическом и этическом регулировании. Пока эффективный инструмент контроля в медиаиндустрии — общественное мнение и культура отмены. Нужно время, чтобы узаконить все процессы и установить правила игры на новом поле.

Список литературы и ссылок:

1. Михаэль Шумахер, первое интервью / Энн Хоффманн // Die Aktuelle. — 2023. — №16. — С. 8–9.
2. Funke Magazine Group : официальный сайт. — Германия, Эссен. — URL: <https://www.funkemedien.de/de/presse/funke-trennt-sich-mit-sofortiger-wirkung-von-die->

aktuelle-chefredakteurin-anne-hoffmann (дата обращения: 25.06.2025).

3. Comic Basics: онлайн-портал: сайт. — URL: <https://www.comicbasics.com/ai-scandal-rocks-oscar-contender-the-brutalist-faces-backlash-over-ai-use-in-production/> (дата обращения: 25.06.2025).

4. Red Shark News: журнал: сайт. — Великобритания, Уилтшир. — URL: <https://www.redsharknews.com/why-epic-period-drama-movie-the-brutalist-was-shot-on-vistavision> (дата обращения: 25.06.2025).

5. Deadline: журнал: сайт. — США, Лос-Анджелес. — URL: <https://deadline.com/2025/01/the-brutalist-ai-brady-corbet-responds-1236260987/> (дата обращения: 25.06.2025).

6. The Wrap: информационное агентство: сайт. — США, Лос-Анджелеса. — URL: <https://www.thewrap.com/the-brutalist-editor-film-ai-hungarian-accent-adrian-brody/> (дата обращения: 25.06.2025).

7. The Guardian: газета: сайт. — Великобритания, Лондон. — URL: <https://www.theguardian.com/technology/2025/jun/29/fake-diddy-ai-videos-youtube> (дата обращения: 25.06.2025).

8. Press TV: телеканал: Telegram-канал. — URL: <https://t.me/presstv> (дата обращения: 25.06.2025).

9. 360: телеканал: сайт, — URL: <https://360.ru/news/mir/iranskie-smi-opublikovali-simuljatsiju-raketnogo-udara-po-voennoj-baze-ssha/> (дата обращения: 25.06.2025).

10. РИА Новости: информационное агентство: сообщество во «ВКонтакте». — URL: https://vk.com/wall-15755094_48084514 (дата обращения: 25.06.2025).

11. RT : информационное агентство: сообщество во «ВКонтакте». — URL: https://vk.com/video-40316705_456469468 (дата обращения: 25.06.2025).

12. Российская газета: газета: сайт. — URL: <https://rg.ru/2025/06/14/foto-sbitogo-iranom-izrailskogo-f-35-mozhet-byt-fejkovym.html> (дата обращения: 25.06.2025).

13. Аргументы и Факты: газета: сайт. — URL: https://vlad.aif.ru/society/ii_isportil_plakat_muromskogo_polumarafona_na_nem_dva_otca_i_opечatki (дата обращения: 25.06.2025).

14. РИА Новости: информационное агентство: Telegram-канал. — URL: https://t.me/rian_ru/264618 (дата обращения: 25.06.2025).

15. The Independent: интернет-издание: сайт. — URL: <https://www.independent.co.uk/travel/news-and-advice/hurricane-milton-disney-world-ai-images-russia-b2627808.html> (дата обращения: 25.06.2025).

16. Медиавойны: Telegram-канал. — URL: <https://t.me/mediavoyni/19355> (дата обращения: 25.06.2025).

Работа с нейросетями: от идеи до реализации на примере создания плаката

ПРАКТИЧЕСКИЕ МАТЕРИАЛЫ



Работа с нейросетями: от идеи до реализации на примере создания плаката

Сабинова Татьяна Романовна — редактор отдела медиакоммуникаций НЦПТИ, г. Ростов-на-Дону.

В эпоху повсеместного использования искусственного интеллекта создание качественных материалов стало вызовом для медиаиндустрии. Легкость в генерации плодит однообразный и малоинформативный контент. Но можно ли использовать нейросети в пользу профилактики?

Разберемся, можно ли производить качественный профилактический контент с использованием нейросетей на примере создания плаката антитеррористической тематики.

Этап 1. Идея

Ко Дню солидарности в борьбе с терроризмом НЦПТИ совместно с Подвиг.РФ создали экспозицию «Истории трагических событий» о 15 крупнейших терактах в современной истории России. На каждом плакате есть QR-код, который ведет зрителей

на аудиорассказ о теракте. Каждая история — про человеческие жертвы, смерти невинных людей. Это и натолкнуло на идею плаката с посылом, что самое страшное в идеологии терроризма — гибель невинных людей, которые оказались рядом с местом происшествия.

Задача: придумать визуальный образ, который будет передавать основную мысль и эмоции зрителю, а также побуждать к действию — перейти по QR-коду и ознакомиться с рассказами.

Основой стала серия терактов в жилых домах в сентябре 1999 года. В визуале постараемся совместить жилые дома, в окнах которых видно людей, и горящую бомбу, которая вот-вот взорвется.

Этап 2. Создание изображения в нейросети

Выбираем доступную нейросеть и создаем промпт.



Промпт — это текст для генерации, набор инструкций для нейросети. Чем детальнее промпт, тем лучше результат. Помимо того, что вы хотите видеть на картинке, можете прописать цвета, текстуры, стиль рисовки и иные детали.

После обработки нашего запроса нейросеть выдает следующее изображение:



Текст промпта:

«На этой картинке я вижу объемную бомбу со сверкающим фитилем. У бомбы отсутствует половина, что создает впечатление нестабильности и опасности. На вершине бомбы изображен город с панельными домами, в окнах которых горит свет. В окнах видны тени людей. В воздухе над городом летают камни, что добавляет сцене хаоса и напряженности. Фон на снимке серый, что подчеркивает мрачность и напряженность момента. Все это создает кинематографический образ».

Этап 3. Создание плаката из получившегося изображения

Доработать плакат можно в любой программе, которая позволяет создавать графические элементы.

В редакторе создаем рабочую область подходящего формата, вставляем в нее сгенерированное изображение и начинаем работать с текстом. Добавляем заголовок, основной текст и QR-код на страницу с рассказами. Изображению добавляем контраста, света в окнах и камней в пустые места.

Рекомендации по работе в графическом редакторе:

1. Используйте не больше двух шрифтов. Один шрифт может использоваться для заголовков, а другой — для основного текста.

2. Работайте с цветами. Выбирайте цветовую палитру, в которой оттенки гармонично сочетаются между собой, достаточно контрастны и соответствуют идее.

3. Соблюдайте правила композиции. На макете должна быть доминанта — ключевой объект, который и будет привлекать внимание зрителя.

4. Заголовок должен быть кратким и лаконичным. Стоит избегать сложных формулировок и двусмысленности.

После редактуры нейросетевого изображения получаем готовый плакат (см. следующую страницу).

Чем нам помогла нейросеть?

Нейросеть сгенерировала картинку за пять минут. Даже опытному дизайнеру потребовалось бы для создания такого изображения не один час.

Вместо сложных программ, работы с 3D мы подробно продумали идею и составили промпт.

Нейросеть — это не волшебная палочка

Стоит понимать, что нейросети — это программы, которые обрабатывают запросы. Это хороший инструмент для работы, но не получится отдать ей процессы полностью. Вы и ваши компетенции — это операционная система для программы, и чем лучше инструкции вы напишите искусственному интеллекту, тем лучше и точнее результат он выдаст.

Самая сложная работа над плакатом — создание смыслов. Перед написанием промпта необходимо придумать визуальный образ, который будет отражать ваш посыл.

На что стоит обратить внимание при работе с нейросетью?

Важно просмотреть сгенерированное изображение на дефекты: неверное число пальцев, нечеткий контур, нечитаемый текст. Внимательно просмотрите, нет ли значительных ошибок, в противном случае перепишите промпт и попробуйте снова. Мелкие погрешности можно попробовать убрать самостоятельно в редакторе изображений.

Нейросети не смогут качественно создать готовый плакат, понадобится работа дизайнера. Но искусственный интеллект откроет новые грани в работе с изображениями: любая ваша идея может быть воплощена, если правильно написать задачу. Стоит большое внимание уделять смыслу и социальному эффекту — это ключевое в профилактическом контенте.

ЖЕРТВЫ ТЕРАКТОВ — НЕВИННЫЕ ЛЮДИ

Жертвами террористов часто
становятся невинные люди, которые
просто оказались рядом.

Переходите по QR-кодам и читайте
трагические истории о терактах
на платформе Подвиг.РФ.



Рассказы Подвиг.РФ

НЕДЕЛЯ СОЦРЕКЛАМЫ

Конкурс по созданию социальных плакатов

Контент-фабрика «Параллели» проводит конкурс для начинающих и опытных дизайнеров **от 16 до 35 лет**. Прием работ **до 24 октября 2025 года**.

Участники создают плакат или серию работ на выбранную тему. Жюри и организаторы определяют лучшие работы, которые будут размещены в городских пространствах Ростова-на-Дону. Победителей определяют жители столицы Юга путем голосования.

ПРИЗОВОЙ ФОНД – 100 ТЫСЯЧ РУБЛЕЙ!

**Поддержите конкурс на ваших медиаресурсах!
Все материалы для распространения по QR-коду:**



Где ИИ может помочь специалисту по профилактике?

ПРАКТИЧЕСКИЕ МАТЕРИАЛЫ



Где ИИ может помочь специалисту по профилактике?

Валуева Анастасия Владиславовна — аналитик отдела профилактической работы НЦПТИ, г. Ростов-на-Дону.

Развитие искусственного интеллекта (ИИ) в современном мире происходит очень быстро, и он уже активно применяется в различных сферах: от медицины до образования и повседневной жизни. В сфере профилактики идеологии терроризма и экстремизма применение ИИ может способствовать значительной автоматизации некоторых рутинных процессов. Кроме того, искусственный интеллект может выступать помощником в разработке программ профилактики или планов мероприятий со студентами.

Для работы с искусственным интеллектом необходимо понимать, какие определения используются для тех или иных технологий. Так, искусственным интеллектом называется сфера в целом. То есть ИИ — это комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их (Национальная стратегия развития ИИ на период до 2030 года, утвержденная Указом Президента РФ от 10.10.2019 года №490).

А вот уже нейросеть (нейронная сеть) — это программа, которая устроена как нейроны в мозге человека и способна обучаться на основе данных и примеров, которые передаются

из одного «нейрона» в другой. Нейросети бывают различных видов: одни могут работать с изображениями, другие — с текстами, третьи — с кодами, и так далее.

Принципы эффективной работы с нейросетями

При работе с нейросетями следует учитывать особенности формулирования запросов. Точность ответа большой языковой модели зависит от того, как много контекста ей было дано в запросе. Контекстом в запросе является дополнительная информация, позволяющая нейросетям конкретизировать и подбирать наиболее подходящий ответ.

Схема запроса для нейросети состоит из нескольких элементов:

1. Роль.

Для повышения релевантности ответа нейросети рекомендуется задавать контекст, который начинается с установления роли — позиции, с которой должен ответить ИИ. Например, можно написать: «ты — проректор по воспитательной работе в вузе» или «ты — психолог в вузе». Здесь можно пойти глубже и описать, какой опыт есть у установленной роли, какое образование и специализация, а также другие критерии, которые могут быть полезны для решения задачи.

2. Цель.

Стоит добавлять цель (и задачи, если необходимо), то есть описывать, зачем нужен ответ на вопрос, что с ним будут делать. Так, если запрос заключается в составлении плана занятия, то необходимо указать цель проведения этого мероприятия и планируемые результаты.

3. Формат ответа.

Можно запросить календарный план работы со студентами, концепцию мероприятия, официальное письмо и др. Кроме того, в формат запроса могут входить объем текста (например, 1000 слов), структура ответа (таблица, список и т. д.) и стиль речи (официально-деловой, научный, разговорный и т. д.).

4. Примеры (если есть).

Для уточнения результата можно загрузить предыдущие сценарии мероприятий и попросить доработать их или привести пример удачного официального письма и попросить адаптировать его под новую задачу.

Эффективно будет также приложить несколько примеров и предложить нейросети самостоятельно взять лучшие формулировки и структуры из каждого предложенного вами варианта, составив ответ.

Также при запросе о плане мероприятия или лекции можно указать целевую аудиторию (курс студентов, возраст, факультет, специальность), место проведения (аудитория университета, форум и т. д.), количество человек. Можно дополнить, что на мероприятии будут присутствовать эксперты (например, психологи, историки и т. д.), уточнить инструменты, которые могут быть доступны на площадке (презентация, микрофон, раздатка для участников и т. д.). Кроме того, можно подробнее

указать, какая именно аудитория будет на мероприятии — с которой уже работали или незнакомая. Можно попросить нейросеть опираться на официальные документы, например, на Комплексный план противодействия идеологии терроризма в Российской Федерации на 2024–2028 годы.

Что следует учитывать?

— Простой и понятный язык.

Нейросети лучше обрабатывают запросы, сформулированные ясно и без избытка профессиональной лексики. Сложные термины могут увеличить вероятность ошибок в ответе ИИ, поэтому предпочтительны формулировки, понятные даже не погруженному в область.

— Вопросы от нейросети.

При составлении запроса напишите фразу, в которой попросите нейросеть задать вам вопросы, если это необходимо для уточнения ответа.

— Тестирование. Каждая нейросеть может быть лучшей в разных видах работы. Одна нейросеть хорошо пишет официальные письма, другая — планы мероприятий, третья может быть более креативной в создании слогана.

Все это можно узнать только методом проб и ошибок, отправляя запросы в различные нейросети и сравнивая, у кого ответ вышел более подходящим под конкретный вопрос.

— Обратная связь. Нейросети учатся не только на основе больших данных, но и на своих ошибках, если им на них указывать. Для этого необходимо внимательно проанализировать ответ нейросети, найти ошибки, слишком общие ответы и неточные решения. Пользователь может отправить нейросети свои корректировки и получить улучшенный ответ. Альтернативно допустимо направить модели запрос о самостоятельном изменении ответа.

Отмечать стоит не только ошибки, но и правильные действия — это помогает нейросети закреплять успешные модели поведения.

Примеры запросов для нейросети в профилактике:

1. Название профилактического мероприятия.

«Ты — специалист по работе с молодежью в вузе. Подбери 5–7 ярких и запоминающихся названий для профилактического мероприятия по теме [тема мероприятия]. Учитывай цель мероприятия и его формат. Целью мероприятия является [цель мероприятия]. Мероприятие пройдет в формате [формат мероприятия: лекция, семинар, игра, др.]. В названии можно использовать отсылки к современной поп-культуре, популярной у [целевая аудитория мероприятия: например, студенты 1 курса по специальности «Политология»]. Названия должны быть современными, но сохранять серьезность темы. По количеству слов названия должны быть короткими [количество слов: 1–3 слова], но запоминающимися».

2. Форматы мероприятий.

«Ты — специалист по организации мероприятий с молодежью в вузе. Необходимо разработать новый формат для мероприятия, направленного на [цель мероприятия]. В мероприятии будут участвовать [конкретизация целевой аудитории: например, студенты 2–3 курсов направления обучения «Биология»]. Мероприятие может продолжаться от одного до двух часов. В рамках мероприятия будут доступны [ресурсы, которыми располагает специалист: проектор, телефоны, аудитория на 50 человек и др.].

Предложи три варианта форматов — от самого креативного до привычного».

Ответ нейросети YandexGPT на запрос о генерации идей для мероприятия по профилактике подверженности манипуляциям для первокурсников исторического факультета:

«Форматы мероприятия по профилактике подверженности манипуляциям:

1. «Манипуляторий» — иммерсивный квест-перформанс.

Концепция: Интерактивный квест с элементами театрализации, где студенты становятся участниками исторических событий, связанных с манипуляциями.

Структура мероприятия:
Вводная часть (15 минут): мини-лекция об исторических манипуляциях.

Основной блок (40 минут): разделение на четыре группы, каждая получает сценарий исторического события с элементами манипуляций.

Практикум (20 минут): анализ ситуаций через призму современных манипуляций.

Необходимые материалы: проектор для показа исторических документов, распечатанные сценарии событий, костюмы (можно использовать подручные средства), телефонные приложения для создания исторических фото.

2. «Антиманипулятор» — интерактивный баттл.

Концепция: Соревновательное мероприятие с элементами игры, где команды соревнуются

в распознавании и противодействии манипуляциям.

Формат проведения:

Разделение на команды (10 минут).

Теоретический раунд (15 минут):

объяснение базовых понятий.

Практический раунд (30 минут):

разбор кейсов.

Финальный баттл (20 минут): решение ситуационных задач.

Подведение итогов (5 минут).

Материалы: проектор для показа кейсов, листы для записей, ручки, система баллов (можно использовать стикеры).

3. «Академия критического мышления» — традиционный семинар с элементами практики.

Концепция: Классический формат с акцентом на практические упражнения и групповую работу.

Структура:

Вводная лекция (20 минут): теория манипуляций.

Групповая работа (30 минут): разбор реальных кейсов.

Практикум (20 минут): отработка навыков противодействия.

Рефлексия (10 минут): обсуждение результатов.

Необходимые ресурсы: проектор для презентации, распечатанные материалы с кейсами, листы для работы в группах, маркеры/ручки для заметок.

Каждый формат можно адаптировать под конкретную аудиторию и имеющиеся ресурсы. Рекомендуется предварительно провести опрос среди студентов для выбора наиболее предпочтительного формата».

Комментарий: в данном случае нейросеть предлагает концепции, которые требуют дальнейшей доработки и адаптации под аудиторию, актуальную тему, конкретные исторические события и навыки модератора, который будет вести данное мероприятие. В последующей работе специалист может выбрать одну из концепций, предложить нейросети свои идеи и продолжить работу над формированием четкого сценария. Очень часто предложения нейросети могут не столько стать конечным вариантом концепции, а именно натолкнуть на мысли о том, какое это может быть профилактическое мероприятие.

3. Креативные решения.

«Ты — специалист по организации мероприятий с молодежью в вузе. Проанализируй структуру профилактического мероприятия и предложи 5–10 креативных изменений. Целевой аудиторией мероприятия являются [конкретизация целевой аудитории: студенты 1 курса направления «Физика»]. На данный момент мероприятие представлено следующим образом: [структура мероприятия с форматами, инструментами и методами]. Предложи интерактивные элементы, которые можно добавить. В реализации мероприятия есть следующие ограничения: [например, бюджет, технические возможности и др.]».

4. Поиск пробелов в мероприятии.

«Ты — специалист по работе с молодежью в вузе. Мы разработали [формат и название игры/мероприятия]

с целью [цель мероприятия].

Механика проведения игры/мероприятия заключается в следующем: [описание механики, т. е. того, что, когда и как будет происходить во время проведения игры/мероприятия]. Проанализируй и выяви слабые места, оцени риски и предложи, как их исправить».

5. Понимание аргументации студентов из групп риска.

«Ты — студент [факультет студента], который придерживается радикальных взглядов по вопросу [тема, по которой необходимо прокачать аргументацию; можно добавить список основных убеждений]. Ты полностью уверен в своей правоте и отстаиваешь позицию до конца. Твоя позиция складывается в основном из следующих источников информации: [медиа и каналы в зависимости от темы: например, «Медуза»*, «Дождь»* и др.]. Твой стиль общения [например, агрессивный, саркастичный, с использованием эмоциональных аргументов или др.]. Я — специалист по профилактике, который должен попытаться выстроить с тобой диалог и понять твою позицию для дальнейшей профилактической работы и демифологизации убеждений. Твоя задача высказывать позицию студента, а моя — отвечать. Давай начнем диалог».

Для этой цели одной из наиболее подходящих нейросетей является DeepSeek.

6. Структура для изучения новых тем.

«Ты — специалист/профессор/эксперт по [общая сфера, которую нужно исследовать]. Мне нужно изучить [конкретная тема/область] для того, чтобы [цель: например, уметь ответить на вопросы студентов по этой теме]. Ранее эту тему я [текущий уровень знаний: не изучал, изучал поверхностно, знаю частично и др.]. предложи

структуру, по которой я смогу изучить эту тему за [временной промежуток: неделю, месяц, год и др.]».

7. Адаптация под целевую аудиторию.

«Ты — специалист по молодежной работе в вузе с опытом проведения мероприятий с различными целевыми аудиториями. Нам нужно адаптировать мероприятие [формат и название мероприятия] для новой целевой аудитории [указать ее]. Прежний формат мероприятия: [описание мероприятия, в чем его смысл и как оно проводилось]. Прошлой целевой аудиторией мероприятия были [целевая аудитория: студенты/преподаватели, специальность, курс и т. д.]. Проанализируй, что стоит оставить, улучшить или добавить».

8. Разработка анкет и опросов.

«Ты — социолог с 10-летним опытом работы, занимаешься созданием анкет для вузов и других образовательных организаций. предложи перечень из [количество вопросов: 10, 20, 30] вопросов, которые можно включить в опрос для студентов/преподавателей, направленный на [цель опроса: выявление знаний в сфере антитеррора и др.]. Структура вопросов: [например, два вопроса со шкалой, пять закрытых вопросов, пять вопросов с вариантами ответа «да» и «нет» и т. д.]».

9. Написание официального письма.

«Ты — делопроизводитель. Тебе необходимо написать официальное письмо в [тип организации: вуз, министерство и т. д.] по теме [тема письма]. В письме необходимо указать [основные моменты, которые должны быть затронуты в письме]. Тон письма должен быть уважительным, по размеру — не более одной страницы. Письмо должно быть понятным, без сложных формулировок».

* Организация внесена в реестр иностранных агентов в Российской Федерации и в перечень иностранных и международных организаций, деятельность которых признана нежелательной на территории Российской Федерации.

Необходимо избегать передачи личных данных ИИ. Для этого можно отметить в запросе, чтобы нейросеть оставила в письме место для названия организации, ФИО и другой необходимой информации. Либо можно предоставить вымышленные данные, которые потом заменить вручную на верные.

Все представленные выше примеры запросов могут и должны быть скорректированы в зависимости от конкретной ситуации и цели. Необязательно указывать все принципы написания в одном запросе одновременно, так как иногда в этом нет необходимости и может привести к ошибкам в ответах нейросети. В определенных случаях будет лучше постепенно в течение диалога с моделью добавлять информацию и дополнительные вопросы.

Нейросети, которые работают с текстом

1. YandexGPT.

Нейросеть от российской компании «Яндекс». На данный момент самой продвинутой доступной версией является YandexGPT5 Pro. Разработчики утверждают, что в работе с текстами эта модель не уступает GPT-4o от компании OpenAI. Без подписки можно воспользоваться возможностями продвинутого режима (YandexGPT 5 Pro) пять раз, то есть отправить только пять запросов. С подпиской можно отправлять безлимитное количество запросов и история сообщений в чате хранится бессрочно (без подписки — только 14 дней).

Стоит отметить один из сервисов YandexGPT — нейроэксперт. Здесь можно подгрузить документы с текстом (например, книги или статьи), и нейросеть будет отвечать на основе

обработанного текста из этих файлов. Такой сервис может быть полезен при изучении новых тем, создании отчетов и постов в социальных сетях на основе определенной информации, а также при разработке новых материалов через переработку старых.

2. GigaChat.

Полностью бесплатная нейросеть от российской компании «Сбер». Без регистрации возможно работать с текстом, вводить формулировки голосом и слушать озвучку ответов, а также получать обработанные запросы на основе поиска нейросетью источников в интернете. После регистрации открываются возможности генерации картинок, работы с файлами, распознавания картинок и умный редактор текста.

3. DeepSeek.

Бесплатная нейросеть от одноименной китайской компании. В этой нейросети доступна работа с текстом, документами и кодом без ограничений и подписок. Кроме того, нейросеть может хранить достаточно большой объем контекста, что позволяет получать персонализированные ответы. Использовать нейросеть можно как в браузере, так и в мобильном приложении.

Другие российские нейросети для работы с текстом: Gerwin, Порфирьевич, RoboGPT, BotHub.

Другие иностранные нейросети для работы с текстом: ChatGPT, Google Gemini, Claude, Grok.

Вывод

Нейросетевые модели активно проникают в профессиональную жизнь во многих сферах. Они позволяют увеличить

скорость обработки текста, создания новых проектов и написания сценариев для мероприятий. Но в то же время следует помнить о том, что нейросеть — это инструмент и помощник, при работе с которым необходимо соблюдать некоторые ограничения.

Так, при взаимодействии с нейросетью требуется исключить использование любых персональных данных, которые могут к ней попасть. В профилактике, особенно при запросах, связанных с мероприятиями для студентов или работой с группами риска, следует тщательно проверять всю информацию и релевантность предложенных идей для реализации со студентами.

Ответ нейросети — это ответ машины, работа, хоть и наученного людьми. Такой ответ опирается на тексты

и данные, полученные ранее, а не на понимание выстраивания взаимодействия с людьми в зависимости от человеческого фактора. Лучше всего консультироваться со специалистами в своей организации о планах, которые были разработаны совместно с нейросетью, чтобы предотвратить потенциальные риски.

Кроме того, следует еще раз поднять вопрос персональных данных и конфиденциальной информации. В некоторых организациях может существовать запрет на установку приложений с нейросетями на рабочие компьютеры, так как они могут собирать данные. Все это необходимо учитывать при применении нейросетей в своей профессиональной деятельности.

Противодействие деструктивному информационно- когнитивному воздействию в рамках обеспечения комплексной безопасности образовательного пространства

СТАТЬЯ



Противодействие деструктивному информационно-когнитивному воздействию в рамках обеспечения комплексной безопасности образовательного пространства

Чурилов Сергей Анатольевич — директор НЦПТИ, г. Ростов-на-Дону.

Венцель Сергей Владимирович — кандидат политических наук, начальник отдела аналитической и образовательной деятельности НЦПТИ, старший преподаватель кафедры регионалистики и евразийских исследований ЮФУ, г. Ростов-на-Дону.

Задача обеспечения комплексной безопасности образовательного пространства в современных условиях является одной из ключевых для руководства любого вуза и включает в себя целый перечень направлений: антитеррористическая защищенность, пожарная безопасность, гражданская оборона, информационная безопасность, профилактика распространения деструктивных явлений среди обучающейся молодежи и педагогического состава.

Особую роль на стыке информационной безопасности и профилактики играет обеспечение информационно-когнитивной защищенности — состояния максимальной сохранности и устойчивости основных для индивидуального и коллективного сознания ценностей и поведенческих моделей. Такие ценности и модели поведения обеспечивают в целом жизнеспособность отдельных людей, коллективов и всего общества, а также позволяют им ориентироваться в современном мире и успешно противостоять внешним и внутренним вызовам.

Сегодня на поведение отдельных людей, коллективов и общества в целом оказывается сильное информационно-когнитивное воздействие.

Оно заключается во влиянии на сознание с помощью современных технологий (от видеохостингов и мессенджеров как площадок распространения до искусственного интеллекта как инструмента создания контента), целью которого является изменить систему убеждений, ценностей и взглядов, а также саму модель поведения.

Такое определение является подходящим как для конструктивного воздействия со стороны общества и государства (например, при работе с отдельными группами, которые находятся в радикальной оппозиции), так и для деструктивного. В краткосрочной перспективе деструктивное информационно-когнитивное воздействие приводит к слому индивидуальных и коллективных норм и моделей, в результате чего некоторые граждане могут совершить правонарушение (от финансирования экстремистской деятельности до осуществления госизмены) или покинуть страну проживания. В долгосрочной перспективе оно способно привести к изменению общества и государственных институтов, внедрению новых практик восприятия реальности и моделей поведения. Результатом может стать сильная поляризация внутри общества

или возникновение конфликтов, в том числе с применением насилия. Ярким примером является украинский опыт с 2004 года (Оранжевая революция) до современных событий периода СВО. Информационно-когнитивное воздействие, по мнению российского политолога и медиатеколога Семёна Уралова, проявляется в двух форматах:

1. Когнитивные волны — постоянные волнообразные формы транслирования установок и ценностей, которые постепенно видоизменяют поведение и систему взглядов человека. Яркий пример — многолетнее транслирование медиакомпанией «Радио Свобода» (включена в перечень нежелательных организаций) установок о «репрессиях и преследованиях в России», «притеснениях коренных народов страны, русификации и ликвидации национальных культур», а также «незаконной аннексии Крыма и разжигании войны на Донбассе».

2. Когнитивные удары — стремительные формы транслирования и закрепления в сознании людей взглядов, а также готовых моделей поведения, которые могут привести, в том числе, к нарушению действующего законодательства. Примером когнитивного удара является транслирование иноагентами и зарубежными ресурсами в 2022–2023 годах призывов уехать из России для «сохранения своего будущего», которые сопровождались видеоинтервью с покинувшими страну известными личностями. Преподносился преимущественно стандартный набор тезисов: «Россия не выдержит санкций, ее экономика обречена», «в России молодежь более не найдет для себя возможностей хорошо жить и зарабатывать» и т. д. Основной целевой аудиторией подобного воздействия выступала российская молодежь.

Цель когнитивных ударов заключается в выведении человека и целых коллективов из эмоционального равновесия, разрыве социальных связей внутри групп, лишении людей личностных черт и превращении их в толпу. Такая толпа может стать ударной силой для совершения незаконных политических акций и обеспечения попыток воздействия на государственную власть или ее свержения. Последствия прослеживаются во множестве случаев:

— попытка «цветной революции» в Республике Беларусь в 2020 году после объявления результатов выборов президента. Основным средством координации выступил телеграм-канал «Nexta»*, администраторы которого находились в Польше. В акции протестов вовлекли разные группы, включая обучающихся школ и вузов, учителей и т. д.;

— незаконные акции протеста по всей России после начала СВО и объявления частичной военной мобилизации в 2022 году, среди основных участников которых были представители обучающейся молодежи;

— протесты в Грузии в 2024 году после парламентских выборов, победу на которых одержали евроскептики из партии «Грузинская мечта». Среди протестующих преимущественно были студенты вузов, представители получающих зарубежные гранты НКО, а также украинская диаспора.

Выявить подверженного когнитивным ударам можно по ряду признаков:

1. Зависимость от информации. Человек не расстается со смартфоном или планшетом, поскольку это устройство для него — основной источник

* В 2022 году Верховный Суд Республики Беларусь признал группу telegram-каналов NEXTA террористической организацией.

информации. Без доступа к информационным потокам, которые держат его в постоянном эмоциональном возбуждении, человек начинает испытывать психологический или физиологический дискомфорт.

2. Нестабильное эмоциональное состояние. Под воздействием информационных потоков и их сопровождающих когнитивных ударов у человека постоянно меняется настроение, причем в течение дня проходит несколько циклов: от эмоционального подъема и радости до апатии, гнева и ярости.

3. Искключительная заикленность на преимущественно политизированных темах. Человек, на которого оказывается негативное когнитивное воздействие, готов разговаривать только по одному перечню вопросов и может начать вести себя агрессивно, если его собеседник попросит переключиться на другие темы или вовсе будет высказывать иные точки зрения.

4. Фрагментарный характер мировоззрения. Восприятие окружающего будет состоять из отдельных фактов, наборов идей и установок, часто противоречащих друг другу. Фрагментация сознания является следствием разрушения системной картины мира, которая включает и причинно-следственные связи между событиями и явлениями, и адекватное понимание сложности общественных процессов без исключительно черно-белого деления.

5. Отсутствие устойчивой системы ценностей. Человек за короткий промежуток времени может сменить противоречащие друг другу модели поведения: от «радикального оппозиционера» до «ура-патриота».

Постоянное деструктивное информационно-когнитивное воздействие может привести к достаточно негативным последствиям на уровнях отдельно взятой личности, коллективов и государства.

Благодаря такому воздействию личность находится в состоянии постоянных «эмоциональных качелей», результатами подверженности которым становятся психологический надлом, выгорание и развитие психосоматических заболеваний. Далее могут проявиться несколько моделей поведения:

— осознанное дистанцирование от обсуждения с педагогами или специалистами по профилактике любых политизированных тем, а также тех вопросов, которые находятся на стыке с актуальными событиями. Например, в ходе некоторых мероприятий со студентами, которые проводили специалисты Национального центра информационного противодействия терроризму и экстремизму в образовательной среде и сети Интернет (НЦПТИ) в разных регионах России, среди некоторых из обучающихся проявлялось нежелание участвовать в исторических викторинах по тематике Великой Отечественной войны и психологических тренингах по защите от вербовки и мошенничества;

— проявление делинквентного поведения — антиобщественного противоправного поведения в форме действия (например, финансирования террористических или экстремистских организаций) или бездействия (например, несообщения о готовящемся преступлении). Яркая иллюстрация — действия террористки Дарьи Треповой, совершившей в 2023 году подрыв кафе в Санкт-Петербурге и убившей военкора Владлена Татарского. Есть несколько примеров и среди обучавшихся в вузах:

в 2023 году студент-программист из Томска получил задание от украинской хакерской группы Cyber Anarchy Squad и совершил атаки на городские информационные ресурсы. В итоге был привлечен к ответственности за государственную измену. В том же году воронежский студент добровольно вступил в контакт с украинскими спецслужбами, собирал информацию для них, а также вступил в террористическую организацию «Легион "Свобода России"»;

— снижение вовлеченности в общественно полезную и любую конструктивную для общества и государства деятельность без совершения каких-либо противоправных действий. Иными словами, человек сознательно отказывается реализовывать свой потенциал для блага страны и государства, хотя до этого мог активно участвовать в волонтерских проектах, работать в государственных и общественных структурах и т. д.

На уровне коллективов худшими последствиями деструктивного информационно-когнитивного воздействия являются паралич процессов внутри группы и разжигание конфликтов между ее членами. В преподавательских коллективах это может приводить к противоправному влиянию как на обучающихся, так и на собственных коллег через транслирование экстремистских установок и призывов к различным несанкционированным действиям. Ярким примером является деятельность «Альянса учителей», тесно связанного с экстремистской организацией «ФБК». Позиционируя себя в качестве «независимого профсоюза», данная организация систематически призывала к антиконституционным

действиям, а летом 2022 года начала собирать личные сведения российских учителей, которые поехали работать в школы ДНР, ЛНР, Запорожской и Херсонской областей, для их передачи Службе безопасности Украины.

Если рассматривать последствия деструктивного информационно-когнитивного воздействия на государственном уровне, то наиболее худшим вариантом является полный паралич органов власти и совершение «цветной революции». Хрестоматийный пример — события на Украине в 2013–2024 годах, когда произошел госпереворот и к власти пришли сторонники идей украинского национализма. В 2025 году украинский опыт деструктивного информационно-когнитивного воздействия на органы власти дополнился еще одним случаем — инспирированными из западных стран молодежными акциями протеста по поводу ограничения полномочий Национального антикоррупционного бюро Украины (НАБУ) и Специализированной антикоррупционной прокуратуры (САП). Организации были созданы по инициативе США после событий 2014 года для контроля за коррупционными потоками внутри украинских элит. Почему обозначен данный пример в контексте деструктивного информационно-когнитивного воздействия? На фоне проводимой тотальной мобилизации, коллапса украинской демографии, паралича многих сфер жизни и постепенной сдачи позиций ВСУ единственным поводом для того, чтобы выйти на улицы и выразить протест против киевского режима, стала попытка снижения контроля украинских властей со стороны зарубежных институтов. Украинский пример на долгие годы станет одним из основных полей для исследования деструктивного

информационно-когнитивного воздействия, развития национализма и социальной антропологии в целом.

Как противодействовать деструктивному информационно-когнитивному воздействию? Выработано достаточное количество мер для работы на трех основных уровнях: государственном, коллективном и личностном.

Среди мер государственного уровня следует выделить:

1. Обеспечение законодательных изменений в сфере образования, науки, просвещения и информационного пространства, введение и ликвидация определенных практик, не отвечающих современным вызовам. Например, введение «Разговоров о важном» в образовательный процесс как инструмента разъяснения общественных реалий и процессов для школьников (одной из наиболее уязвимых категорий перед деструктивным воздействием).

2. Регулирование информационного поля, сопровождающееся разработкой и внедрением собственных площадок для распространения информации (например, VK Видео или RuTube в контексте видеоблогинга), а также ограничением доступа к определенным ресурсам в силу нарушения ими национального законодательства (например, блокировка в марте 2022 года Instagram и Facebook вследствие признания корпорации Meta экстремистской организацией, продуктами которой являются данные приложения).

3. Наполнение собственного информационного поля различными видами контента: от информационно-просветительского, направленного

на разъяснение актуальных общественно-политических явлений, исторических событий, до профилактического, цель которого — нивелировать последствия деструктивного воздействия и объяснить, почему оно является таковым (как для отдельных личностей, так и для всего общества). Разработка и распространение подобного контента может происходить и благодаря предоставлению грантов и субсидий различными фондами для НКО, работающих в рамках отечественного правового поля. Позитивным примером является интернет-шоу «Переменка», разработанное АНО «Интернет без угроз» благодаря грантовой поддержке (см. рисунок 1).

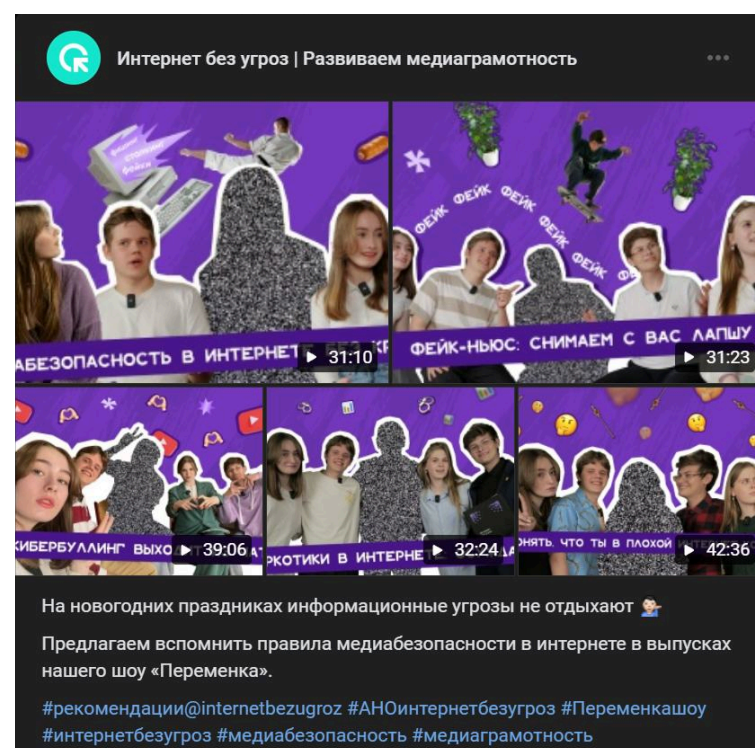


Рис. 1. Выпуски интернет-шоу.

4. Защита общества от тех лидеров мнения и медиаличностей, которые будучи гражданами России стали ретрансляторами деструктивного информационно-когнитивного воздействия. Основной инструмент — придание статуса иноагента, который предполагает значительные ограничения: например, невозможность производства и распространения информационных материалов среди несовершеннолетних,

запрет на образовательную деятельность и т. д. Как показывает практика, несоблюдение требований законодательства (например, неуказание соответствующей маркировки материалов, распространяемых иноагентом) приводит к привлечению к уголовной ответственности по ст. 330.1 УК РФ («Уклонение от исполнения обязанностей, предусмотренных законодательством Российской Федерации об иностранных агентах»).

5. Предотвращение доступа деструктивного влияния в сферу образования, науки, просвещения и некоммерческого сектора. Основной инструмент в России — признание иностранных организаций нежелательными с запретом как на их деятельность внутри страны, так и на взаимодействие российских граждан и юридических лиц с ними. Среди уже признанных нежелательными есть те организации, которые с 1990-х годов активно работали в России в сфере науки и высшего образования: фонд «Открытое общество» (США), «Германо-российский обмен» (ФРГ), «Оксфордский российский фонд», «Британский совет» (Великобритания) и некоторые другие.

Основная задача сейчас — разъяснить обучающимся, педагогам и ученым, что в настоящее время сотрудничество или финансирование нежелательных организаций может стать основой для привлечения к юридической ответственности. Например, сдача экзамена на знание английского языка IELTS, который проводится «Британским советом», не рекомендуется, поскольку может рассматриваться как финансирование нежелательной организации (ст. 284.1 УК РФ). Причем запрет на сотрудничество

распространяется не только на территорию России: если российский ученый, находясь за рубежом, будет работать на одну из нежелательных организаций, это также может стать причиной для судебного разбирательства.

Меры уровня коллективов и организаций по противодействию деструктивному информационно-когнитивному воздействию предполагают, первоочередно, работу над сплочением групп людей, формированием среди них необходимых умений по конструктивному взаимодействию друг с другом, а также навыков по работе с различными информационными потоками. Иными словами, это выстраивание на системной и плановой основе профилактической и просветительской работы среди обучающихся, а также педагогов.

Что важно учитывать при формировании информационно-когнитивной устойчивости среди обучающейся молодежи и педагогов?

1. Необходимость регулярного информационного и профилактического воздействия как в очном, так и в онлайн режимах. Речь идет не только о постоянном проговаривании с молодежью о том, какие статьи УК РФ предусматривают ответственность за содействие террористическим и экстремистским организациям, но и о формировании позитивных смыслов (например, донесение информации о возможностях самореализации в рамках выбранной специальности, участия в общественно полезной деятельности или о формах досуга), об историческом (рассмотрение важных событий прошлого и ключевых исторических персонажей) и общественном просвещении (например, обсуждение современных ключевых достижений страны и ее граждан).

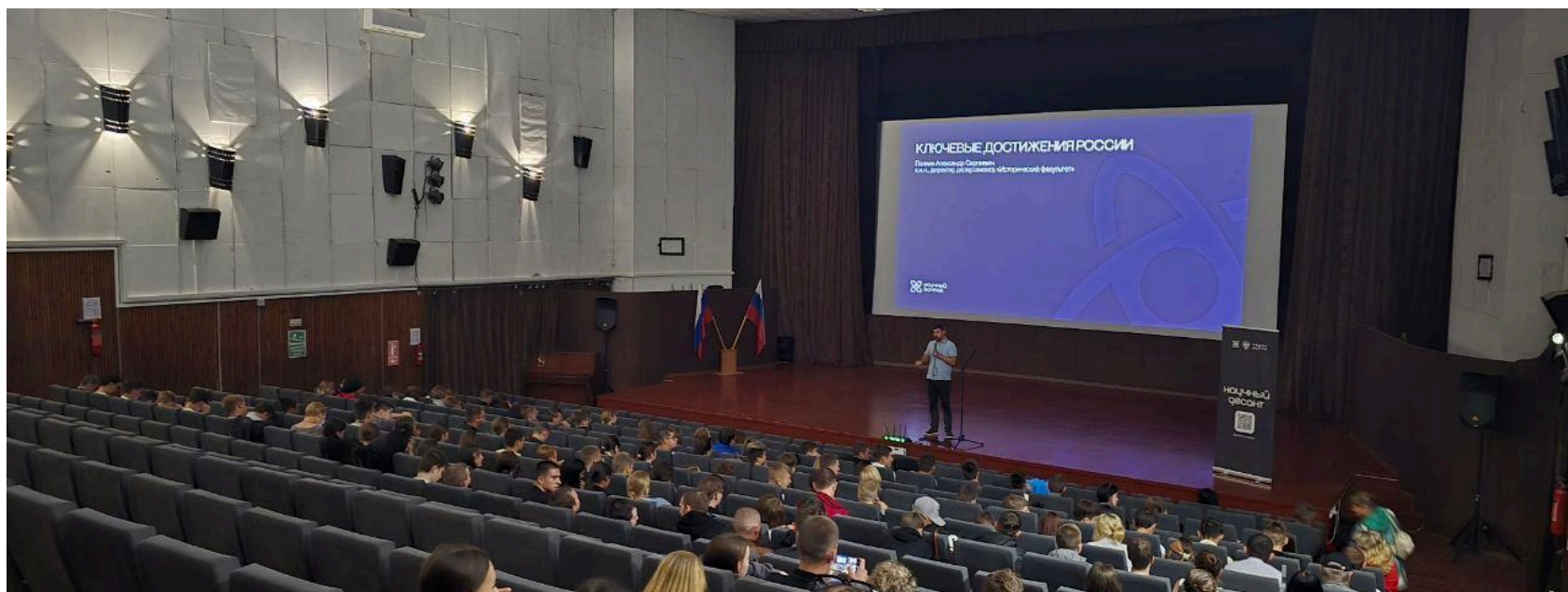


Рис. 2. Выступление об исторических достижениях России перед школьниками Херсонской области в рамках проекта «Научный десант».

Позитивный пример – проект Минобрнауки России «Научный десант», направленный на популяризацию научных достижений и истории России в регионах Донбасса и Приазовья среди обучающихся и педагогов (см. рисунок 2).

В медиаресурсах образовательной организации в этом могут помочь регулярные посты о ключевых достижениях в сфере науки, участии и победах обучающихся в соревнованиях и олимпиадах, введение в эксплуатацию новых корпусов и нового исследовательского оборудования и т. д. Все это формирует позитивную повестку, благодаря которой обучающиеся могут получить уверенность как в сегодняшнем дне, так и в завтрашнем, а также приобрести примеры для подражания.

2. Подверженность нашего мозга и нас в целом определенным пристрастиям и слабостям. Каждый из нас — человек, который не может функционировать постоянно, как робот. Важно обеспечить обучающихся возможностями отдыха и конструктивного времяпрепровождения, особенно это касается наиболее уязвимых перед

деструктивным воздействием групп риска. Например, для иностранных студентов, сложность адаптации которых напрямую связана со знанием языка и скоростью погружения в местную специфику, следует организовывать игровые мероприятия для формирования состояния безопасности и доверия.

3. Важность приучения к регулярному использованию критического мышления, особенно в тех ситуациях, где неправильные действия могут привести к негативным последствиям. Есть две формы мышления: докритическое (поступление информации сразу приводит к действиям) и критическое (между двумя обозначенными фазами есть критическое осмысление поступившей информации). Важно помнить, что при получении информации у людей сначала происходят эмоциональные реакции, а затем уже включается рациональное осмысление. Исполнители информационно-психологических операций особенно часто используют данную слабость. Например, в начале СВО украинские ресурсы массово атаковали пользователей Рунета фотографиями

и видео с последствиями ракетных атак в украинских городах. В ходе битвы за Мариуполь украинские ЦИПСО демонстрировали фото разрушенного Драматического театра, обвиняя в его обстреле ВС РФ.

4. Накопившиеся сложности внутри гуманитарных дисциплин, в том числе в отечественном научном пространстве. Результат — неспособность объяснения и прогнозирования многих общественных явлений, а также некорректное применение терминологического аппарата. Многие термины из политологии, социологии и других гуманитарных наук либо слишком пространны (например, «фундаментализм» или «экстремизм»), либо слишком политизированы (например, «патриотизм») и даже дискредитированы («демократия», «PR-кампании» и т. д.). Данный фактор усложняет коммуникацию с аудиторией для просвещения и профилактики.

Для достижения положительного результата в профилактике специалисту важно решить три вытекающие друг из друга задачи:

1. Получить доверие или хотя бы возможность быть услышанным среди аудитории. Тот факт, что люди физически находятся в одном пространстве, совсем не означает, что они находятся с вами в коммуникации.

2. Предотвратить вовлечение людей в противоправную деятельность. Важно рассказать о том, почему явление деструктивно, почему человеку не следует во что-то вступать.

И здесь мы работаем по четырем основным направлениям:

— формирование основ психологической устойчивости как способности адаптироваться к жизненным трудностям и справляться со стрессом. Направление включает в себя работу с обучающимися по развитию навыков самоконтроля и адаптивности, формированию чувства оптимизма, а также привитию осознания ответственности за совершаемые действия;

— обучение исторической грамотности как комплекса знаний и навыков исторического характера для ориентации в современном мире. Первоочередная задача — сохранение и приумножение полученных в школе знаний, а также их углубление (например, уже в сфере изучаемой дисциплины). Вторая важная задача — развенчивание исторических мифов антироссийского характера, которые сегодня стали неотъемлемой частью иностранного деструктивного информационно-когнитивного воздействия. Среди позитивных примеров — проект «Подвиг.РФ», в рамках которого разработана выставка по развенчиванию основных деструктивных исторических мифов о Великой Отечественной войне (см. рисунок 3);

— развитие навыков работы с информационными потоками, определения потенциальной деструктивности и конструктивности, а также критического отношения к любой распространяемой и потребляемой информации. Как минимум, донесение информации о необходимости дистанцирования от тех каналов, которые вызывают на постоянной основе нестабильное эмоциональное состояние и являются носителями когнитивных ударов;

— демонстрация моделей выстраивания коммуникации с другими людьми.

Коллектив в целом способен оказывать воздействие, в том числе профилактического характера, на тех, кто демонстрирует отклоняющиеся модели поведения. В результате каждый участник коллектива будет предварительно задумываться

активистов в каждой образовательной организации не так много, если ориентироваться на общее количество обучающихся. Такое вовлечение при правильном подходе позволит сформировать достаточно устойчивую группу единомышленников из числа обучающихся, которые станут ретрансляторами позитивных установок и ценностей среди своих одноклассников, а также других студентов при взаимодействии на мероприятиях или в качестве наставника или куратора. Позитивный пример – Школа студенческого актива «Форпост», реализованная НЦПТИ уже дважды: в 2024 году среди студентов Белгородской области, а в 2025 году — в ДНР (см. рисунок 4).

На личностном уровне для обеспечения защищенности от деструктивного информационно-когнитивного воздействия рекомендуется следующее:

1. Собственное активное вовлечение в общественную, научную или спортивную жизнь образовательной организации. В свою очередь, обеспечение такими



Рис. 3. Выставка проекта НЦПТИ «Подвиг.РФ», посвященная развенчиванию основных деструктивных исторических мифов о Великой Отечественной войне.

о совершении или несовершении определенных действий. Взаимодействие в коллективе обеспечивает выработку у человека умения слушать и дает возможность быть услышанным, а также позволяет принять наличие других точек зрения по разным вопросам, что может предотвратить формирование «черно-белого» мировоззрения.

3. Вовлечь в позитивную и конструктивную деятельность. Это самая сложная задача, поскольку

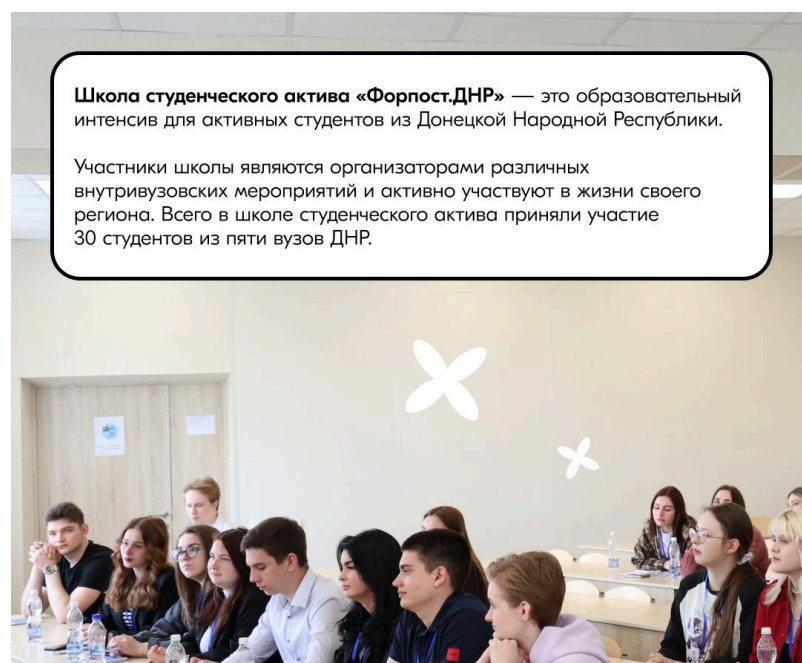


Рис. 4. Новость о Школе студенческого актива «Форпост» в ДНР.

возможностями — одна из ключевых задач образовательной организации. Постоянный контакт и чувство сопричастности к общему позитивному делу будут, с одной стороны, выступать как элемент самопрофилактики при взаимодействии с деструктивным воздействием. С другой стороны, это обеспечит реализацию собственного потенциала, а для образовательной организации — формирование актива, на который можно положиться в различных ситуациях.

2. Формирование психологически и ценностно безопасного собственного медиаполя. Если в перечне каналов или сообществ есть те, которые вызывают негативные эмоции или способствуют развитию депрессии, лучше от них отписаться для сохранения эмоционального равновесия. При необходимости лучше ограничить количество ресурсов по общественно-политической тематике, оставив только зарекомендовавших себя экспертов (не нарушающих при этом российское законодательство), а также крупные новостные ресурсы.

3. Постоянное развитие собственных когнитивных способностей (памяти, восприятия, внимания). Такие навыки формируются в том числе посредством чтения больших текстов, благодаря которым мозг способен удерживать значительные объемы информации, создавать причинно-следственные связи и даже прогнозировать дальнейшее поведение. Короткие видеоролики и серии карточек, разумеется, тоже следует использовать для получения информации, однако они не должны становиться в жизни человека единственным форматом. В противном случае когнитивные способности человека

будут существенно ограничены, а он сам станет уязвимым перед деструктивным воздействием.

Подводя итог, следует отметить: основная задача деструктивного информационно-когнитивного воздействия заключается в нарушении причинно-следственных связей, сломе логики и видоизменении моделей поведения.

Задачи просвещения и профилактики — обеспечить защиту от такого воздействия, развить навыки поиска и анализа информации, не поддаваясь рефлексам и эмоциям при чтении новостных потоков. За длительное время специалистами НЦПТИ разработано и распространено значительное количество методических рекомендаций, готовых сборников сценариев профилактических мероприятий, а также проектов по вовлечению молодежи в созидательную деятельность («Контент-фабрика „Параллели“», «Подвиг.РФ», «Форпост» и некоторые другие). Данные разработки размещены на ресурсах НЦПТИ в свободном доступе для сотрудников образовательных организаций в сфере профилактики и воспитания.

Крайне важным в защите от деструктивного воздействия является оказание всей необходимой поддержки научно-исследовательским, общественным, культурно-просветительским и иным проектам, которые инициирует сама молодежь.

И следует помнить о наполнении информационного пространства позитивной повесткой. Благодаря этому у обучающихся будет формироваться чувство оптимизма и уверенности в своем будущем в нашей стране для полноценного развития собственного потенциала.



**Готовые сценарии и презентации
для проведения просветительских занятий
с обучающимися по тематике
медиаграмотности и медиабезопасности**



**Мошенничество
в сети**



**Буллинг
и stalking**

Послесловие

Еще несколько лет назад ведущие технологические эксперты мира сравнивали опасность неконтролируемого развития искусственного интеллекта с ядерной войной и пандемией. Наиболее пессимистически настроенные адепты техногенной агрессии высказывали опасения, что ИИ может привести к гибели человечества, и в этой мысли их утвердил такой светоч прогресса, как Илон Маск, заявлявший, что существует двадцатипроцентная вероятность уничтожения человечества умными машинами. Более приземленно мыслящие опасались необратимых последствий для социальной сферы и сокращения рабочих мест. А во франшизе про Терминатора, созданной в восьмидесятых, ее автор Д. Камерон даже называл конкретную дату Судного дня — 29 августа 1997 года, предрекая уничтожение человечества могущественным «Скайнетом», обретшим сознание. И даже по данным «Национального индекса тревожности», ставшего итогом анализа СМИ и социальных сетей в 2023 году, тема Искусственного Интеллекта заняла лидирующее место, отражая опасения людей в связи с неприятностями, исходящими от машинного разума и нейронных сетей.

Однако, не случилось, слава небесам.
(Ну не «облаку» же!)

При этом, как не заметить, что без помощи цифровых технологий мы не смогли бы во времена пандемии организовать учебный процесс — в школах и вузах страны. И даже электронные экскурсии по залам Эрмитажа стали возможны лишь благодаря чуду арт-токенов. Не говоря уж о множестве полезных решений, сгенерированных ИИ, его достижениях в медицине, лингвистике, математике, промышленности. Словом, естественный интеллект вовсю использует возможности электронного соперника, начиная с его технологий машинного обучения. Надеюсь, эксперты нашего выпуска смогли довольно подробно изложить свой взгляд на проблемы, возникающие при обсуждении такого явления, как искусственный интеллект. Но в то же время было бы несправедливо, ознакомившись с мнением одной стороны, не предоставить слово оппонентам, в данном случае, самому ИИ.

Вот как выглядят предсказания возможного будущего человечества с точки зрения искусственного интеллекта до 2100 года (Российская газета, №124 (9069)):

2026: ИИ разрушает экономику.

Веб-разработчики, бухгалтеры, юристы постепенно заменяются ИИ.

2028: ИИ создает произведения искусства, музыку и видео, заменяет учителей.

2033: ИИ создает роботов для себя же. ИИ встроен во все — от кофеварок до самолетов.

2038: ИИ совершает научные открытия, на которые компьютерам потребовались бы тысячи лет.

2045: ИИ останавливает процесс старения, человек будет жить более 300 лет.

2050: Достигнута «эпоха изобилия». Установлен всеобщий базовый доход. Люди живут в основном в виртуальной реальности.

2060: Люди сливаются с ИИ. Биологические мозги становятся цифровыми.

2070: Человеческое сознание становится цифровым.

2100: Доминирование ИИ. ИИ стремится колонизировать другие звездные системы.

Как вам такой прогноз? Впрочем, проживем, увидим. Осталось подождать каких-нибудь 75 лет.

*Яровой Анатолий Владимирович,
главный редактор*

Тема следующего выпуска:

Профилактика и воспитание в вузе

**Дата выпуска:
декабрь 2025 года**

По вопросам публикации материалов обращаться в редакцию:

info@ncpti.ru

+7 (863) 201-28-22



ОБЗОР.НЦПТИ
СБОРНИК ИНФОРМАЦИОННЫХ МАТЕРИАЛОВ

Национальный центр информационного противодействия терроризму и экстремизму в образовательной среде и сети Интернет
