Шапошник Кристина Дживаовна
педагог-организатор
ГБОУ «Шебекинская гимназия-интернат»
г. Шебекино, Белгородская область

НЕЙРОСЕТИ В АРХИТЕКТУРЕ МЕНТАЛЬНОГО ЗДОРОВЬЯ: ОТ ИНСТРУМЕНТА МАНИПУЛЯЦИИ К РЕСУРСУ ЗАЩИТЫ

Аннотация: *в статье анализируется двойственная роль технологий искусственного интеллекта (ИИ) и нейросетей в контексте психологической безопасности личности и общества. Рассматриваются основные векторы угроз: генерация дезинформации (дипфейки), манипуляция массовым сознанием, кибербуллинг с использованием ботов, формирование цифровой зависимости и когнитивных искажений. Одновременно ИИ позиционируется как ключевой инструмент защиты: системы раннего выявления суицидального риска, персонализированные платформы психологической помощи, алгоритмы модерации деструктивного контента и инструменты для диагностики ментальных состояний. Делается вывод о необходимости разработки этических и правовых норм для минимизации рисков при интеграции ИИ в социальную сферу.*

Ключевые слова: *искусственный интеллект, нейросети, психологическая безопасность, дипфейки, кибербуллинг, ментальное здоровье, цифровая гигиена, этика ИИ.*

Психологическая безопасность традиционно определяется как состояние, при котором отсутствует угроза психотравмирующих факторов и сохраняется целостность личности. В условиях тотальной цифровизации границы этого состояния стремительно размываются. Искусственный интеллект и нейросетевые алгоритмы перестали быть узкоспециализированным инструментом и превратились в инфраструктуру повседневности. Они формируют информационную повестку, определяют круг общения и влияют на эмоциональный фон человека. Данная трансформация ставит перед психологами, социологами и разработчи-

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

ками технологий новую задачу: купирование рисков, которые несет ИИ, и конвертация его вычислительных мощностей в ресурс для укрепления ментального здоровья.

Нейросети обладают беспрецедентными возможностями по генерации и дистрибуции контента, что создает несколько критических угроз для психологической безопасности.

Эпоха постправды и дипфейки. Генеративные нейросети способны создавать фотореалистичные изображения, видео и аудиосообщения, неотличимые от реальных. Использование дипфейков для шантажа, дискредитации публичных лиц или создания ложных новостей о катастрофах вызывает у потребителей информации хроническую тревогу, подрывает базовое доверие к реальности и социальным институтам. Человек теряет надежный критерий истинности, что ведет к росту паранойидальных настроений.

Алгоритмическая манипуляция и поляризация. Рекомендательные системы социальных сетей и новостных агрегаторов работают на максимизацию вовлеченности. Алгоритмы ИИ создают «информационные пузыри» (эхо-камеры), где пользователь видит только подтверждающую его убеждения информацию. Это провоцирует когнитивную ригидность, рост агрессии к носителям иных взглядов и социальную поляризацию, разрушая чувство общности и безопасности.

Кибербуллинг и синтетические атаки. ИИ-боты позволяют масштабировать травлю, создавая иллюзию массовой агрессии против одного человека. Нейросети генерируют оскорбительные тексты и изображения, делая кибербуллинг более изощренным и психологически разрушительным. Особую опасность представляет использование ИИ для создания синтетических интимных материалов (NSFW) с лицами реальных людей.

Когнитивная деградация и цифровая зависимость. Чрезмерная опора на ИИ-ассистентов (от навигаторов до генераторов текстов) ведет к атрофии критического мышления и памяти. Геймифицированные алгоритмы удержания внимания, встроенные в приложения, провоцируют дофаминовую зависимость, синдром дефицита внимания и эмоциональное выгорание.

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

Синтетические отношения и социальная изоляция. Популярность ИИ-компаньонов и чат-ботов, имитирующих эмпатию, несет риск подмены реальных социальных связей суррогатными. Это может приводить к усугублению чувства одиночества и неспособности выстраивать коммуникацию с живыми людьми.

Парадоксально, но те же технологии, которые создают угрозы, являются наиболее эффективным средством защиты от них и инструментом психологической помощи:

Проактивная модерация контента. Нейросети способны в режиме реального времени анализировать тексты, изображения и видео на предмет деструктивного контента (призывов к насилию, пропаганды самоповреждения, травли). Алгоритмы удаляют опасную информацию до того, как она нанесет вред уязвимым группам пользователей, в первую очередь детям и подросткам.

Раннее выявление ментальных расстройств. Анализ цифровых следов (паттернов речи, скорости набора текста, изменений в активности в социальных сетях, интонаций голоса) позволяет нейросетям выявлять маркеры депрессии, биполярного расстройства или суицидальных наклонностей на субклинических стадиях. Системы раннего предупреждения могут рекомендовать пользователю обратиться за помощью или сигнализировать доверенным контактам.

Доступность психологической помощи. Чат-боты на базе больших языковых моделей (LLM) выступают первым рубежом поддержки. Они доступны 24/7, анонимны и снимают барьер первичного обращения к специалисту. Нейросети помогают структурировать запрос пользователя, обучают базовым техникам саморегуляции (например, дыхательным практикам) и снижают остроту панической атаки, выполняя роль «санитарного кордона» перед визитом к психотерапевту.

Персонализация терапии и супервизия. ИИ анализирует сессии когнитивно-поведенческой терапии, помогая специалистам выявлять упущенные паттерны в поведении клиента. Для пациента создаются персонализированные программы реабилитации и виртуальные тренажеры для отработки социальных навыков (в том числе с использованием технологий виртуальной реальности).

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

Снижение профессиональной нагрузки. Автоматизация рутинной документации (составление протоколов сессий, анализ результатов тестирования) высвобождает время психологов для непосредственной работы с клиентом, предотвращая их профессиональное выгорание.

Для трансформации ИИ из угрозы в ресурс необходим комплексный междисциплинарный подход.

Технологический: внедрение систем «цифровых водяных знаков» для сгенерированного контента и разработка нейросетей-детекторов, способных со 100% точностью идентифицировать дипфейки.

Образовательный: формирование навыков цифровой и психологической гигиены. Обучение критической оценке информации и распознаванию манипулятивных техник ИИ должно стать частью школьной программы.

Этико-правовой: разработка стандартов прозрачности алгоритмов. Пользователь должен понимать, почему ИИ предлагает ему тот или иной контент, и иметь право на аудит этих решений. Необходима правовая база, регулирующая ответственность за вредоносное использование генеративных моделей.

Искусственный интеллект не является по своей природе добрым или злым; он является мощнейшим усилителем человеческих интенций. В контексте психологической безопасности это означает, что ИИ масштабирует как деструктивные, так и созидательные практики. Будущее ментального благополучия общества зависит не от ограничения развития технологий, а от способности человека интегрировать их в свою жизнь осознанно. Психологическая безопасность в XXI веке – это не отсутствие информационных угроз, а наличие у личности когнитивных инструментов и социальной поддержки для их нейтрализации, где ИИ должен выступать надежным помощником, а не невидимым манипулятором.

Список литературы

1. Асмолов, А. Г. Преадаптация к неопределенности как стратегия навигации развивающихся систем: маршруты эволюции / А. Г. Асмолов, Е. Д. Шехтер, А. М. Черноризов // Вопросы психологии. – 2017. – № 4. – С. 3–26.

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

2. Войскунский, А. Е. Психология и Интернет / А. Е. Войскунский. – Москва : Акрополь, 2010.

3. Зинченко, Ю. П. Психологическая безопасность в цифровом обществе: риски и ресурсы / Ю. П. Зинченко // Вестник Московского университета. Серия 14. Психология. – 2020. – № 4. – С. 3–18.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)