**УДК 004.89:316.4**

**К ВОПРОСУ О БЕЗОПАСНОСТИ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СОЦИАЛЬНЫХ СЕТЯХ**

**Васин Владимир Владимирович**

АО ЦКБА, г.Тула

***Научный руководитель: Зайцева Татьяна Валентиновна****к.т.н., доцент*

*ФГАОУ ВО «Белгородский государственный национальный*

*исследовательский университет» (НИУ «БелГУ»)*

**Аннотация:** В работе рассматривается роль искусственного интеллекта (ИИ) в развитии социальных сетей, а также потенциальные риски, связанные с его использованием. ИИ обеспечивает персонализацию контента, автоматическую модерацию, анализ пользовательского поведения и выявление трендов. Однако его применение сопряжено с угрозами конфиденциальности, распространением дезинформации, манипуляцией общественным мнением и усилением зависимости от цифровых платформ. Особое внимание уделено вопросам этичности, прозрачности алгоритмов и необходимости регулирования.

**Ключевые слова:** искусственный интеллект, социальные сети, конфиденциальность, дезинформация, алгоритмы, цифровая безопасность, этика ИИ, цифровое регулирование.

**ON THE ISSUE OF THE SAFETY OF USING ARTIFICIAL INTELLIGENCE IN SOCIAL NETWORKS**

**Vasin Vladimir Vladimirovich**

***Scientific adviser: Zaitseva Tatiana Valentinovna***

**Abstract:** The article examines the role of artificial intelligence (AI) in the development of social networks, as well as the potential risks associated with its use. AI provides content personalization, automatic moderation, user behavior analysis, and trend detection. However, its use is fraught with threats to privacy, the spread of disinformation, manipulation of public opinion, and increased dependence on digital platforms. Special attention is paid to the issues of ethics, transparency of algorithms and the need for regulation.

**Key words:** artificial intelligence, social networks, privacy, disinformation, algorithms, digital security, ethics of AI, digital regulation.

Искусственный интеллект (ИИ) сегодня играет ключевую роль в развитии социальных сетей, он стал их неотъемлемой частью. ИИ помогает персонализировать контент, автоматизировать модерацию, улучшать пользовательский опыт и анализировать тренды [1, 3, 4]. Однако вместе с этими возможностями приходят и серьезные вызовы — от угроз конфиденциальности до распространения дезинформации и манипуляции общественным мнением.

ИИ приносит значительную пользу как пользователям, так и платформам. Один из самых заметных эффектов — персонализация. Алгоритмы анализируют поведение пользователей и подбирают контент, который соответствует их интересам, повышая вовлеченность и удовлетворенность.

Еще одно преимущество — автоматическая модерация. ИИ способен быстро выявлять и удалять нежелательный контент, включая спам, разжигание ненависти и порнографию, делая платформы безопаснее. Благодаря ИИ также улучшается поиск информации и коммуникация: интеллектуальные ассистенты и чат-боты могут эффективно обрабатывать запросы и помогать пользователям в режиме 24/7.

Кроме того, ИИ активно используется для анализа трендов. Платформы отслеживают обсуждаемые темы и общественные настроения, что позволяет лучше понимать поведение аудитории и адаптировать стратегию контента.

Несмотря на многочисленные плюсы, ИИ в социальных сетях порождает и существенные риски. Один из главных — угроза конфиденциальности. Алгоритмы собирают и анализируют огромные объемы персональных данных: от демографии до эмоционального состояния пользователей. Это создает риски утечек, несанкционированного доступа и использования информации без согласия — например, в политических целях или для навязчивой рекламы.

ИИ также способствует распространению дезинформации [2]. Алгоритмы могут усиливать влияние ботов и фейковых аккаунтов, распространяющих ложные новости. Использование дипфейков позволяет создавать реалистичные, но фальшивые аудио- и видеозаписи, что подрывает доверие к источникам информации. Эффект эхо-камер, когда пользователи видят лишь то, что подтверждает их взгляды, усугубляет поляризацию общества.

Опасность представляют и манипулятивные механизмы. Таргетированная реклама может эксплуатировать психологические уязвимости пользователей. Алгоритмы способны влиять на эмоциональное состояние и даже формировать зависимость от социальных сетей, втягивая людей в бесконечное потребление контента.

Еще один вызов — непрозрачность алгоритмов [5]. Большинство систем ИИ функционируют как "черные ящики", логика работы которых остается неясной даже разработчикам. Это затрудняет контроль, снижает уровень доверия и ставит под сомнение возможность привлечения к ответственности за негативные последствия их использования.

Для минимизации рисков необходимо комплексное взаимодействие всех сторон: разработчиков, пользователей и регулирующих органов [1-5].

Разработчикам стоит сосредоточиться на:

1. Создании прозрачных и объяснимых алгоритмов, которые позволяют понять, как принимаются решения.
2. Минимизации предвзятости с помощью разнообразных обучающих данных и регулярного тестирования.
3. Надежной защите данных и ограничении их сбора до необходимого минимума.
4. Комбинировании автоматической и ручной модерации, особенно в спорных случаях.
5. Информировании пользователей о работе ИИ и возможностях настройки конфиденциальности.
6. Постоянном мониторинге и аудите алгоритмов для выявления ошибок и нежелательных эффектов.

Пользователи могут повысить свою безопасность, если будут:

1. Проверять информацию из независимых источников и критически относиться к увиденному.
2. Понимать основы работы алгоритмов и быть осведомленными о рисках манипуляций.
3. Настраивать конфиденциальность, избегать публикации чувствительной информации и не переходить по подозрительным ссылкам.
4. Контролировать время, проводимое в сети, и сознательно выбирать источники контента.
5. Участвовать в онлайн-дискуссиях уважительно, избегая распространения слухов и токсичности.

Регуляторы и государственные органы играют важную роль и должны:

1. Разрабатывать этичные стандарты использования ИИ, учитывая принципы справедливости, прозрачности и ответственности.
2. Принимать законодательные меры по защите данных, борьбе с дезинформацией и контролю за алгоритмами.
3. Проводить мониторинг и вводить санкции за нарушения.
4. Содействовать международному сотрудничеству, обмену опытом и согласованным действиям по глобальным угрозам ИИ.

Таким образом, ИИ в социальных сетях — мощный инструмент, который может приносить значительную пользу, но при неправильном использовании становится источником рисков и угроз. Обеспечение безопасного, этичного и прозрачного использования ИИ требует скоординированных усилий всех участников цифровой экосистемы. Только так можно сохранить доверие пользователей, защитить их права и создать здоровую информационную среду, где технологии служат интересам общества.

**Список источников**

1. Горемыко М.С. Эвристический потенциал искусственного интеллекта в анализе социальных сетей // Изв. Сарат. ун-та Нов. сер. Сер. Социология. Политология. - 2024. - №2. – С. 172-177.

2. Искаков А.М., Корнилов А.А. К вопросу поиска людей с помощью социальных сетей и искусственного интеллекта [Электронный ресурс] // Альманах Пермского военного института войск национальной гвардии. - 2025. - №1 (17). – Режим доступа: URL: https://cyberleninka.ru/article/n/k-voprosu-poiska-lyudey-s-pomoschyu-sotsialnyh-setey-i-iskusstvennogo-intellekta (дата обращения: 15.07.2025).

3. Соколов А.В., Шуткин А.С., Епифанова Е.М., Попкова А.А., Бекларян А.Л., Барулина М.А. Применение искусственного интеллекта для привлечения клиентов в социальных сетях: обзор российского и зарубежного опыта // ВЕСТНИК ПЕРМСКОГО УНИВЕРСИТЕТА. МАТЕМАТИКА. МЕХАНИКА. ИНФОРМАТИКА. - 2025. - 1 (68). – С. 118–144.

4. Соцсети и ИИ: вызовы и возможности в 2025 году [Электронный ресурс]. – Режим доступа: URL: https://www.sostav.ru/publication/sotsseti-i-ii-vyzovy-i-vozmozhnosti-v-2025-godu-76688.html (дата обращения: 14.07.2025).

5. Харитонова Ю.С., Савина В.С., Паньини Ф. Предвзятость алгоритмов искусственного интеллекта: вопросы этики и права // Вестник Пермского университета. Юридические науки. - 2021. - №53. – С. 488-515.

© В.В. Васин, 2025