РИСКИ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СОЦИАЛЬНЫХ СЕТЯХ: ДЕЗИНФОРМАЦИЯ, МАНИПУЛЯЦИЯ И УТРАТА ПРИВАТНОСТИ

Скежанов Арман Айыпұлы
магистрант 1 курса ЕНУ им. Л.Н. Гумилева
г. Астана, Казахстан
e-mail: a.skezhanov@gmail.com
Научный руководитель: **Шурентаев А. М.**

Аннотация: В статье рассматриваются риски применения искусственного интеллекта в социальных сетях: дезинформация, алгоритмическая манипуляция и утрата приватности. Актуальность темы связана с растущим влиянием ИИ на отбор, распространение и персонализацию цифрового контента. Цель статьи - проанализировать основные угрозы использования ИИ в социальных медиа. Сделан вывод о необходимости прозрачности алгоритмов, развития медиаграмотности и защиты персональных данных.

Ключевые слова: искусственный интеллект, социальные сети, дезинформация, манипуляция, приватность, цифровая безопасность.

1. Введение

В условиях цифровизации социальные сети стали важным каналом получения и распространения информации, а искусственный интеллект - одним из ключевых инструментов отбора и персонализации контента. Он анализирует поведение пользователей, формирует рекомендации и влияет на то, какие публикации чаще появляются в информационной ленте.

С одной стороны, ИИ делает цифровую коммуникацию удобнее, помогает находить нужные материалы и выявлять вредоносный контент. С другой стороны, его применение связано с рисками дезинформации, создания фейковых материалов, алгоритмической манипуляции вниманием пользователей и сбора персональных данных.

Особую опасность представляют генеративные модели, способные создавать тексты, изображения, аудио и видео, похожие на достоверные материалы. Поэтому актуальность статьи определяется необходимостью рассмотреть риски применения ИИ в социальных сетях, связанные с дезинформацией, манипуляцией и утратой приватности.

2. Методология

Данная статья имеет обзорный характер и основана на методах анализа, систематизации, сравнения и обобщения научной литературы. В работе рассматриваются современные исследования. Выбор обзорного подхода обусловлен междисциплинарным характером темы, поскольку она затрагивает вопросы журналистики, социологии, информационной безопасности, права и цифровой этики. Анализ проводится по трём основным направлениям: распространение дезинформации, влияние алгоритмов на поведение пользователей и сбор персональных данных.

3. Основная часть

3.1 ИИ и дезинформация в социальных сетях

Одним из наиболее заметных рисков применения искусственного интеллекта в социальных сетях является усиление дезинформации. Бонтриддер подчёркивает, что цифровые технологии значительно расширили возможности распространения ложной

информации, поскольку контент в онлайн-среде распространяется быстро, массово и часто без предварительной проверки[103]. Искусственный интеллект изменил механизм создания фейкового контента. Генеративные технологии позволяют быстро создавать тексты, изображения, аудио- и видеоматериалы, внешне похожие на реальные. Шоаиб отмечает, что развитие генеративного ИИ усиливает угрозы, связанные с дипфейками и фейковым контентом, так как искусственно созданные материалы становятся всё более убедительными[104].

Особую опасность представляют дипфейки и автоматизированные аккаунты - боты. Бонтчева указывает, что реалистичность ИИ-контента затрудняет его распознавание обычными пользователями[105]. Хайли рассматривает социальных ботов как инструмент искусственного усиления популярности отдельных сообщений и влияния на восприятие аудитории[106]. Таким образом, ИИ ускоряет создание фейкового контента, повышает его убедительность и облегчает массовое распространение через ботов и алгоритмические рекомендации. Главная опасность заключается в том, что ложная информация становится всё более правдоподобной и требует от пользователя критического мышления и медиаграмотности.

3.2 Алгоритмическая манипуляция поведением пользователей

Алгоритмическая манипуляция является одним из наиболее скрытых рисков применения искусственного интеллекта в социальных сетях. Пользователь может считать, что самостоятельно выбирает контент, однако значительная часть публикаций, видео и новостей подбирается рекомендательными системами. Эти системы анализируют лайки, комментарии, подписки, время просмотра и другие цифровые следы, чтобы предлагать материалы, способные дольше удерживать внимание. Родилоссо отмечает, что рекомендательные алгоритмы могут усиливать поляризацию, поскольку чаще предлагают пользователю контент, соответствующий его уже существующим интересам и взглядам[107]. В результате формируется «информационный пузырь», при котором человек всё реже сталкивается с альтернативными точками зрения. Ариб также подчёркивает, что такая персонализация может сужать информационное поле пользователя[108].

Особую роль играет эмоциональный контент. Социальные сети заинтересованы в удержании внимания, поэтому алгоритмы часто продвигают публикации, вызывающие страх, тревогу, возмущение или раздражение. Огнибене указывает, что цифровая зависимость и манипулятивное влияние усиливаются, когда рекомендательные системы

[103] Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, *3*, e32.
[104] Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023, November). Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models. In 2023 international conference on computer and applications (ICCA) (pp. 1-7). IEEE.
[105] Bontcheva, K., Papadopoulous, S., Tsalakanidou, F., Gallotti, R., Dutkiewicz, L., Krack, N. & Verdoliva, L. (2024). Generative AI and disinformation: recent advances, challenges, and opportunities.
[106] Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social bots and the spread of disinformation in social media: the challenges of artificial intelligence. *British Journal of Management*, *33*(3), 1238-1253.
[107] Rodilosso, E. (2024). Filter bubbles and the unfeeling: How AI for social media can foster extremism and polarization. Philosophy & Technology, 37(2), 71.
[108] Areeb, Q. M., Nadeem, M., Sohail, S. S., Imam, R., Doctor, F., Himeur, Y., ... & Amira, A. (2023). Filter bubbles in recommender systems: Fact or fallacy—A systematic review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(6), e1512.

ориентированы прежде всего на удержание пользователя на платформе[109]. Персонализированные рекомендации, короткие видео, уведомления и постоянное обновление ленты могут формировать зависимость от социальных сетей. Коч отмечает, что дальнейший подбор контента зависит от активности пользователя и его взаимодействия с платформой[110]. Таким образом, ИИ постепенно подстраивает информационную среду под реакции человека, влияя на его внимание, эмоции и поведение.

3.3 Утрата приватности и сбор персональных данных

Утрата приватности является одним из ключевых рисков применения искусственного интеллекта в социальных сетях. Пользователь оставляет множество цифровых следов: лайки, комментарии, подписки, поисковые запросы, просмотренные публикации, фотографии, геолокацию и время активности. Алгоритмы ИИ способны объединять эти данные и формировать подробный цифровой профиль человека.

Гилберт отмечает, что пользовательское профилирование связано с анализом поведения, интересов и активности человека в социальных сетях[111]. Это означает, что платформы получают информацию не только из публикаций пользователя, но и из его взаимодействия с контентом: какие материалы он читает, комментирует, просматривает дольше или чаще открывает.

Особую роль играют лайки, комментарии и геолокация. Ивуанянву подчёркивает, что данные, собранные во время просмотра и взаимодействия с контентом, превращают социальные сети в крупный источник персональной информации[112]. Пелле указывает, что сведения о местоположении позволяют анализировать передвижения пользователя, что создаёт риски для его физической приватности[113]. ИИ способен анализировать изображения и видео, размещённые пользователями. Таким образом, ИИ делает сбор персональных данных более глубоким и точным, что усиливает риск их использования в коммерческих, политических или манипулятивных целях.

4. Обсуждение

Проведённый обзор показывает, что искусственный интеллект в социальных сетях имеет двойственный характер. С одной стороны, он помогает персонализировать контент, выявлять вредоносные публикации и быстро обрабатывать большие объёмы информации. С другой стороны, ИИ может усиливать дезинформацию, манипулировать вниманием пользователей и создавать угрозы приватности. Главная проблема заключается не в самой технологии, а в недостатке прозрачности и контроля за её применением. Поэтому снижение рисков ИИ должно включать защиту персональных данных, развитие медиаграмотности.

5. Заключение

[109] Ognibene, D., Wilkens, R., Taibi, D., Hernández-Leo, D., Kruschwitz, U., Donabauer, G., ... & Eimler, S. (2023). Challenging social media threats using collective well-being-aware recommendation algorithms and an educational virtual companion. *Frontiers in Artificial Intelligence*, *5*, 654930.

[110] Koç, B. İ. R. K. A. N. (2023). The role of user interactions in social media on recommendation algorithms: Evaluation of TikTok's personalization practices from user's perspective. *Istanbul University*.

[111] Gilbert, J., Hamid, S., Hashem, I. A. T., Ghani, N. A., & Boluwatife, F. F. (2023). The rise of user profiling in social media: review, challenges and future direction. *Social Network Analysis and Mining*, *13*(1), 137.

[112] Iwuanyanwu C. C. Facebook artificial intelligence algorithm: Users' awareness and response to data privacy issues. – Robert Morris University, 2023.

[113] Pellet, H., Shiaeles, S., & Stavrou, S. (2019). Localising social network users and profiling their movement. *Computers & Security*, *81*, 49-57.

Таким образом, искусственный интеллект в социальных сетях имеет двойственное значение: он повышает скорость обработки информации и удобство цифровой коммуникации, но одновременно создаёт риски дезинформации, алгоритмической манипуляции и утраты приватности. Проведённый обзор показывает, что главная опасность связана не с самой технологией, а с непрозрачностью её использования и недостаточным контролем со стороны платформ. Поэтому для снижения рисков необходимо развивать медиаграмотность пользователей, усиливать защиту персональных данных и обеспечивать ответственность социальных сетей за применение ИИ.

Список литературы

1. Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. Data & Policy.
2. Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models.
3. Bontcheva, K., Papadopoulos, S., Tsalakanidou, F., et al. (2024). Generative AI and disinformation: Recent advances, challenges, and opportunities.
4. Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social bots and the spread of disinformation in social media: The challenges of artificial intelligence. British Journal of Management.
5. Rodilosso, E. (2024). Filter bubbles and the unfeeling: How AI for social media can foster extremism and polarization. Philosophy & Technology.
6. Areeb, Q. M., Nadeem, M., Sohail, S. S., et al. (2023). Filter bubbles in recommender systems: Fact or fallacy - A systematic review. WIREs Data Mining and Knowledge Discovery.
7. Ognibene, D., Wilkens, R., Taibi, D., et al. (2023). Challenging social media threats using collective well-being-aware recommendation algorithms and an educational virtual companion. Frontiers in Artificial Intelligence.
8. Koç, B. (2023). The role of user interactions in social media on recommendation algorithms: Evaluation of TikTok's personalization practices from user's perspective.
9. Gilbert, J., Hamid, S., Hashem, I. A. T., Ghani, N. A., et al. (2023). The rise of user profiling in social media: Review, challenges and future direction. Social Network Analysis and Mining.
10. Iwuanyanwu, C. C. (2023). Facebook artificial intelligence algorithm: Users' awareness and response to data privacy issues.
11. Pellet, H., Shiaeles, S., & Stavrou, S. (2019). Localising social network users and profiling their movement. Computers & Security.