Научная статья
УДК 070:316.77
DOI 10.47475/2949-3390-2025-5-1-140-146

РЕВОЛЮЦИЯ, КОТОРУЮ МЫ НАЧАЛИ, И ЕЁ ПОСЛЕДСТВИЯ, КОТОРЫЕ МЫ НЕ МОЖЕМ КОНТРОЛИРОВАТЬ

Зарина Мирзо
Учебный центр по авиации «Таджикаэронавигация», Душанбе, Таджикистан, z_mirzo@mail.ru

Аннотация. Мы живём в эпоху новых технологий, интернета и искусственного интеллекта, где каждый человек может создавать контент в цифровом пространстве. Любой может выступить рассказчиком, провокатором, влиятельным лицом или многоликой фигурой в цифровой среде, изменяя институты и формируя новые глобальные дискурсы. Контент, созданный пользователями (UGC), стал одновременно силой, раскрывающей правду, и оружием обмана, стирая границы между подлинностью и манипуляцией на таких платформах, как «ВКонтакте», YouTube, RuTube, TikTok и многих других. Однако эта свобода самовыражения несёт риски, включая распространение дезинформации, пропаганды, этические дилеммы и проблемы регулирования. В данной статье исследуется баланс между свободой слова и ответственностью за создание контента, роль искусственного интеллекта в контроле цифровых пространств, а также влияние UGC на бренды, политику и средства массовой информации.

Ключевые слова: контент, созданный пользователями (UGC), модерация с помощью ИИ, дезинформация, манипуляция, предвзятость, дискурс, свобода слова, этические дилеммы, медиаграмотность, проверка фактов, гражданский журнализм, репутация бренда, прозрачность

Для цитирования: Мирзо З. Революция, которую мы начали, и её последствия, которые мы не можем контролировать // Динамика

медиасистем. 2025. Том 5, Вып. 1. С. 140–146. doi: 10.47475/2949-3390-2025-5-1-140-146

Original article

THE REVOLUTION WE STARTED AND ITS CONSEQUENCES, WHICH WE CANNOT CONTROL

Zarrina Mirzo
Aviation Training Center "Tajikairnavigation", Dushanbe, Tajikistan,
z_mirzo@mail.ru

Abstract. We are living in an era of new technology, the Internet, and AI, where each individual can create content in digital communication. One can be a storyteller, troublemaker, influencer and many-faced characters in the digital arena, changing institutions and reshaping global discourses. User-generated content (UGC) has become both a force for truth and a weapon of deception, confusing the lines between authenticity and manipulation across platforms like VKontakte, YouTube, RuTube, TikTok and many others. However, this freedom of expression comes with risks, including misinformation, disinformation, propaganda, ethical dilemmas and regulatory challenges. This paper studies the balance between free expression and responsibility for content creation, the role of AI in controlling digital spaces, and the impact of UGC on brands, politics and media.

Key words: user-generated content (UGC), AI moderation, misinformation, disinformation, propaganda, bias, discourse, freedom of expression, manipulation, ethical dilemmas, media literacy, fact-checking, citizen journalism, brand reputation, transparency

For citation: Mirzo Z. (2025). The revolution we started and its consequences, which we cannot control. *Dynamics of Media Systems*, Vol. 5, Issue 1: 140–146. doi: 10.47475/2949-3390-2025-5-1-140-146 (In Engl).

Introduction

Social media has changed communication, shifting power from traditional media to individual internet user. Today, a single post can ignite public movements, expose injustices, or spread misinformation. This shift raises critical concerns on credibility, regulation, and the influence of digital algorithms.

AI moderation plays a crucial role in managing the high volume of UGC. However, studies suggest AI alone cannot fully address ethical and contextual nuances, requesting human intervention. Moreover, automated moderation risks reinforcing biases, suppressing disagreement, and aligning with corporate or government agendas.

Molina and Sundar indicate that the success of AI moderation depends on human collaboration, as machines alone cannot fully capture the nuances of context or ethical considerations in online discourse (Molina, Sundar 2022).

Similarly, Gorwa, Binns, and Katzenbakh explore political and technical challenges of automated content regulations (Gorwa, Binns, Katzenbakh 2020). They agree that AI moderation is not only a technological issue but also a political, as filtering algorithms are often influenced by the agendas and values of the platforms deploying them. They caution that relying on AI may reinforce power dynamics controlled by governments or corporations.

As digital arenas become battlefields where truth and deception collide, the struggle to control the discourse has never been more intense. The challenge is not only to identify and eliminate misinformation but also ensure that the systems don't suppress legitimate disagreements or infringe upon freedom of expression.

1. AI, content moderation, and the future of digital governance

Artificial intelligence serves as both a safeguard and a risk in moderating online content. While AI can identify misinformation, it can also reinforce biases or suppress legitimate discourse. Floridi warns that AI moderation systems are vulnerable to manipulation, enabling malicious actors to spread disinformation unchecked (Floridi 2020).

Zhou and Zafarani support these concerns, stating that AI driven moderation systems are vulnerable to attacks (Zhou, Zafarani 2020). Malicious users can manipulate algorithms to evade detection, flooding digital platforms with synthetic misinformation or political propaganda.

Bontridder and Poullet examine AI's in dual role in both spreading and mitigating disinformation (Bontridder, Poullet 2021). Their study analyzes how AI contributes to misinformation while also exploring AI – driven countermeasures. They emphasize the need for a balanced approach, which AI does not reinforce biases or suppress legitimate discourse.

Zhou and Zafarani further caution that disinformation networks will likely adapt, exploiting new technologies to spread falsehoods more effectively (Zhou, Zafarani 2020). This highlights the urgent need for collaboration between the regulatory bodies, tech companies, and researchers to develop adaptable solutions that balance free expression with accountability. It will ensure that future digital spaces remain both open and trustworthy requires ongoing vigilance.

2. Successful and challenging uses of UGC

UGC has played vital role in shaping public opinion, both positively and negatively. During significant global events, citizen journalism has exposed human rights violations, yet misinformation has also grown.

Pennycook and Rand explore the "Implied truth effect", demonstrating how misinformation can sometimes have unintended consequences (Pennycook, Rand 2019). When certain content is labeled as false, users may mistakenly assume that everything without a label is true. In such situations the spread of misinformation, particularly during global crises or public health emergencies rapidly accelerates.

As digital landscape continues to expand, society faces a pressing dilemma, harnessing the power of UGC while mitigating its risks. Spreading misinformation threatens the public trust and disrupts democratic institutions. To mitigate these risks, digital platforms must invest in improved fact- checking, algorithmic transparency, and stronger verification systems.

3. UGC and brand reputation

UGC has both positive and negative consequences for brands. While user-generated campaigns can boost engagement, misinformation can severely damage reputations. Pournaki et al. emphasize the importance of proactive trust building strategies, arguing that countering misinformation is crucial to maintaining consumer confidence (Pournaki et al., 2021). They highlight that brand reputation is constantly negotiated within the dynamic and unpredictable ecosystem of social

media.

Companies like Nike and Apple have used UGC effectively, turning customers into brand ambassadors. However, misinformation related incidents, such as the Apple "Bendgate" controversy and Tesla's self-driving vehicle are of safety concerns, illustrate the risk of viral UGC.

Pournaki et al. caution that negative UGC, including viral complains and misleading reviews, can damage a brand's reputation overnight (Pournaki et al. 2021). Platforms like VKontakte and RuTube have been used to expose unethical business, forcing companies to respond publicly.

To mitigate these risks, brands must invest in real-time monitoring, crisis management strategies and AI-driven tools to detect harmful misinformation before it spreads. Pournaki et al. suggest that brands actively monitor digital discourse, engage with consumers, and implement AI-driven content verification systems (Pournaki et al. 2021). Transparency and responsiveness can transform UGC from a liable into an asset.

4. Ethical dilemmas in AI content moderation

The rapid expansion of UGC introduces complex ethical and legal dilemmas. Automated moderation systems, driven by AI, frequently misinterpret cultural and contextual nuances, resulting in over-censorship or implicit biases across various digital platforms.

Udupa et al. discuss the ethical challenges of scaling AI content moderation, particularly in relation to extreme speech (Udupa et al. 2022). Their study provides evidence demonstrating the difficulty of machine learning moderation methods in keeping pace with the evolving nature of hateful language online. This highlights the limitations of AI systems in accurately identifying and moderating content without suppressing legitimate expression.

Gapani advocates for an ethical framework to combat the uncontrollable spread of online misinformation, suggesting that governments, platforms, and civil society must collaborate to mitigate the negative impact of UGC on society (Gapani 2023).

Moreover, Huschens and Briesch highlight that users are more likely to trust AI when its decision-system process is clear, accurate, and competent (Briesch 2023). They emphasize the importance of ensuring that AI moderation systems not only function effectively but

also communicate their decisions transparently to enhance user trust.

A multi-stakeholder approach involving policymakers, platforms, and civil society is necessary to establish ethical frameworks for UGC governance. Transparency in content moderation, media literacy initiatives, and legal interference are crucial to maintaining fair digital spaces.

Conclusion

UGC has undoubtedly transformed digital communication, but its power comes with challenges. AI-driven moderation must be combined with human oversight to ensure fairness and prevent algorithmic bias. To strengthen fact- checking mechanisms, media literacy education, and ethical content policies will be key to navigating the future of digital discourse.

The balance between freedom and control will define the future of online spaces. Without proper oversight, UGC risks become an instrument for misinformation and manipulation. By promoting transparency, accountability, and digital responsibility, we can harness the potential of UGC while protecting its integrity.

References

Molina, M. D., Sundar, S. S. (2022). When AI moderates online content: Effect of human collaboration and interactive transparency on user trust. *Journal of Computer – Mediated Communication*. Vol. 27, Issue 4, zmac010. https://doi.org/10.1093/jcmc/zmac010.

Gorwa, R., Binns, R., Katzanbach, C. (2020). Algorithmic content moderation: Technical and political challenges in automation of platform governance. *Big Data & Society*, 7 (1). https://doi.org/10.1177/2053951719897945.

Floridi, L. (2020). The fight for digital sovereignty: what it is, and why it matters, especially for the EU. *Philosophy and technology*, 33. 369–378. https://doi.org/10.1007/s13347-020-00423-6.

Zhou, X., Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Comput. Surv,* 53, 5, Article 109 (September 2021), 40 pages. https://doi.org/10.1145/3395046

Bontridder, N., Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, 3, e32. https://doi.org/10.1017/

dap.2021.20.

Pennycook, G., Rand, D. G. (2019). The implied truth effect. *Management science*, Vol. 66, 11. 4944–4957. https://doi.org/10.1287/mnsc.2019.3478.

Udupa, S., Maronikolakis, A., Schutze, H., Wisiorek, A. (2022). Ethical Scaling for Content Moderation: Extreme Speech and the (In) Significance of Artificial Intelligence. *Discussion papers, public interest technology, publications*, available at: https://shorensteincenter.org/ethical-scaling-content-moderation-extreme-speech-insignificance-artificial-intelligence/ (accessed: 01.02.2025).

Huschens, M., Briesch, M. (2023). Perceived credibility of human and AI-generated content. *Human-computer interaction*. https://doi.org/10.48550/arXiv.2309.02524.

Информация об авторе

Зарина Мирзо – кандидат филологических наук, преподаватель английского языка

Information about the author

Zarrina Mirzo – Candidate of Philological Sciences, English Language Teacher

Автор заявляет об отсутствии конфликта интересов.

The author declares no conflicts of interests.