AI-Generated Synthetic Content on the Web: Impacts on Credibility, Detection Strategies, and Ethical Challenges

Atul Jindal

ABSTRACT
Synthetic content, which moves virally across the web, has created concerns about its impact on information saturation, online data authenticity, and originality. This study mainly focused on the synthetic content generated by AI on user credibility. It also examines the challenges and ethical concerns arising from the vast usage of synthetic content, while introducing different methods for detecting and combating its spread. The research also focuses on case studies of how new technologies handle synthetic content on the web and their effect on online networks. The paper ends with a survey on new technologies and predictions regarding the growth of synthetic content on the internet.

Keywords
Synthetic content, authenticity, originality, credibility, technology, online networks, internet.

1. INTRODUCTION
Synthetic content has altered online data creation and consumption through its widespread use on the internet in recent years. Synthetic content produced by AI can perfectly mimic or copy human-generated text, images, and videos (Ghiurău, 2024). Synthetic content has initiated a discussion about originality, authenticity, and its capability to influence the web (Ghiurău, 2024). As a result, there is an impact on user credibility and trust (Ghiurău, 2024; Ciftci, 2020). The Internet is now facing difficulties and ethical conflicts due to improvements in synthetic content as AI is developing rapidly. This study focuses on the saturation, authenticity, and originality of online data (Ciftci, 2020). By analyzing multiple phenomena, such as identification techniques, ethical implications, and case studies, we can understand the sophisticated mechanisms between human-generated and AI-generated content in the digital era (Ciftci, 2020).

2. DEFINITIONS OF AUTHENTICITY, ORIGINALITY, AND SATURATION IN THE CONTEXT OF ONLINE CONTENT

2.1 Authenticity
Authentic online data refers to data that is not corrupted, obtained from a reliable source, and intended for original purposes (Ghiurău, 2024).

- In context: Synthetic content can manipulate authenticity because it can create duplicate versions of true individuals and organizations and their modes, which makes it difficult for users to recognize accurate content from altered material (Ghiurău, 2024).

2.2 Originality
Online content originality refers to distinctive and innovative aspects, which include expressions, ideas, and presentations. Content maintains its originality when it avoids direct copying, excessive derivative work, and repetitive generation from existing materials (Ciftci, 2020).

- In context: The content generation capabilities of generative AI tools involve pattern recombination from extensive training data, which challenges the authenticity of original outputs and sparks concerns about plagiarism, intellectual property rights, and creative work degradation (Ciftci, 2020).

2.3 Saturation
The web experiences saturation when it becomes overwhelmed by excessive content, which leads to diminished visibility of important material, shorter attention span, and challenges in separating valuable information from background noise (Ghiurău, 2024).

- In context: The ability to produce extensive synthetic content leads to platform saturation through the overwhelming presence of repetitive low-effort content and SEO-targeted materials that push out high-quality human-created work (Ghiurău, 2024).

3. EXPANSION OF AI TOOLS FOR CONTENT DEVELOPMENT
Artificial Intelligence (AI) has revolutionized content creation by shifting from human-based manual work to automated and scalable operations (Lounge, 2023).

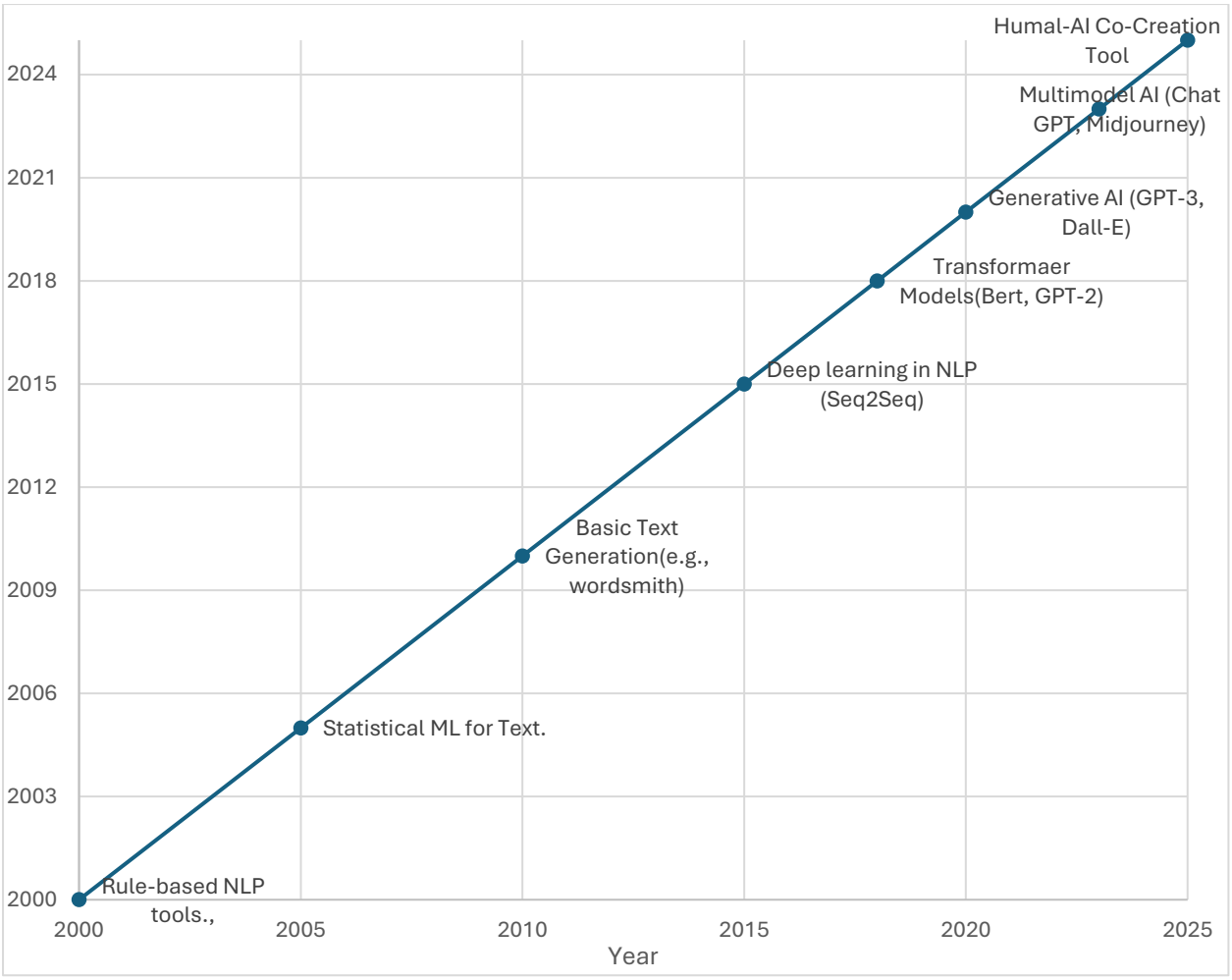

Figure1. Evolution of AI Tools over the Years (2000-2025)

The number of worldwide searches for AI reached its highest point in May 2023, when it surged from 42 million to 17 million in January. The number of searches has grown substantially since January 2022, when it stood at 8 million (Lounge, 2023). Figure 1 shows the rapid improvement of AI tools over the past few years. AI evolution began with rule-based systems and natural language processing (NLP) algorithms, which created early tools for grammar correction and content summarization (Subramani, 2024). Over time, machine learning models have progressed to produce basic templates and structured data narratives, including financial reports and weather updates (Subramani, 2024). Fundamental transformation occurred when deep learning and transformer-based architecture emerged, including Meta's LLaMA, OpenAI's GPT models, and Google's BERT. These models enable the production of highly sophisticated text along with human-like images, audio, and video content. AI-powered platforms today generate extensive articles, marketing materials, social media posts, realistic voiceovers, artwork, music, and deepfakes through limited human interaction (Subramani, 2024). The increasing availability of these tools, together with their integration into standard platforms, has made human and machine authorship increasingly indistinguishable. Technological advancement brings both improved operational efficiency and customized experiences, while creating significant issues regarding original work authenticity, ethical practices, and digital content overload (Subramani, 2024).

4. SYNTHETIC CONTENT: AI-GENERATED TEXT, IMAGES, VIDEOS

The identification of synthetic content requires the detection of specific characteristics in text, audio, image, and video formats. The main problem arises from the absence of logical connections between the different elements (Cooke, 2024). Synthetic media differ from natural content because they display abrupt transitions instead of smooth connections between frames or text. Multimodal detection methods, which analyze audio signals, as well as visual and textual data, have emerged as a promising solution to enhance synthetic content identification (Cooke, 2024). Synthetic videos display abrupt changes in background elements, lighting conditions, and character movements that create an unnatural viewing experience. Logical connections between ideas tend to be absent in texts produced by AI (Cooke, 2024). While humans can maintain knowledge consistency, AI texts may contain errors due to a limited understanding of the context. The identification of these features helps in the development of detection systems and supports digital media forensic operations (Cooke, 2024).

4.1 AI-generated Text

A new AI technology, such as OpenAI's GPT series, generates text that can copy human writing. Traditional tools struggle to detect the AI content. AI can alter text, change meanings, and spread misinformation (Ghiurău, 2024). Synthetic text shows

unnatural patterns, with AI displaying less variability in sentence structure. AI often uses uncommon phrases rarely found in human writing. Semantic anomalies indicate AI authorship (Ghiurău, 2024). While humans maintain knowledge consistency, AI text can include errors owing to limited context understanding. AIGC detection uses statistical patterns, linguistic anomalies, and semantic consistencies. GPTZero uses probability distributions but struggles with GPT-4. OpenAI's detector and BERT models identify semantic inconsistencies, achieving accuracy in specific domains but facing limitations with the edited AIGC (Ghiurău, 2024).

Table 1. Tips for recognizing false text based on word usage (Ghiurău, 2024)

Phrases Generated by a Language Model	Hints for Detecting Synthetic Text
The intricate dance between science and nature provides an undeniable synergy, fostering innovation and breakthroughs.	Grandiose phrasing: Use broad, impressive-sounding words without real insights.
In modern global economics, the balance of trade and currency fluctuations are vital components.	Surface-level complexity: The text sounds technical but lacks specific, actionable insights.
The historical development of art has traversed various eras, from the renaissance to post-modernism and classical physics has puzzled scientists for decades.	The text refers to certain specific terms, such as "renaissance," while providing limited detail about other significant aspects.
The relationship between quantum mechanics and classical physics has puzzled scientists for decades.	Generic technical jargon: The language sounds academic but does not provide new or meaningful details
While many of them believe in importance of education, the future seems to be moving towards artificial intelligence.	Overly generic or predictable patterns: Complex, or vague ideas lacking specificity or depth.

Hybrid approaches combining statistical and deep learning methods improve detection but increase costs. Crowdsourcing offers optimal performance. Table 1 presents an analysis of the phrases generated by the language models. Tools such as GLTR 2024 or GPTZero 2024.06 analyze lexical features, whereas OpenAI's TruthfulQA performs fact-checking (Ghiurău, 2024).

4.2 Image

The images generated by generative adversarial networks appear genuine until trained experts perform the analysis (Bougueffa, 2024). The metadata within the AI-generated images contains misleading information that confuses the detection algorithms. Synthetic images reveal their artificial nature through abnormal texture patterns, which differ from natural images that integrate smooth textures (Bougueffa, 2024). The unnatural patterns in these images demonstrate that AI systems cannot replicate natural textures. The detection process works by detecting unusual patterns that appear in the background and shadows. Synthetic images reveal their artificial nature through incorrect shadow angles and inconsistent lighting, which deviate from the physical lighting principles that authentic images follow (Bougueffa, 2024). Dual-stream networks that perform texture analysis together with low-frequency trace detection produce superior results compared to traditional methods for detecting GAN content. The accuracy of CNN-based frameworks improves, but they require periodic updates because generative models continue to evolve (Bougueffa, 2024; Ghiurău, 2024). The training datasets for the CNN image classifiers consisted of human-generated images alongside AI-created content. A watermarking system functions as a protective mechanism for content attribution. The synthetic image in Figure 2 shows both unnatural apple textures and unconvincing raindrop effects. The Sight Engine and NVIDIA GAN fingerprint tools can detect these defects. The results from the Sight Engine demonstrated strong capabilities in identifying synthetic content (Figure 3) (Ghiurău, 2024). The detection of newer models requires advanced frameworks that are essential for dealing with multiple audio and video streams (Ghiurău, 2024).

Figure 2. Apple Image Produced by Dall-E (Ghiurău, 2024)

Figure 3. A Sight Engine was used to analyze the AIGC images (Ghiurău, 2024)

4.3 Videos

The video manipulation tools DeepFaceLab and FaceSwap produce extremely realistic fake videos (Bougueffa, 2024). The detection system examines visual elements along with audio signals, and behavioral patterns are spread across various data streams. Synthetic video detection relies heavily on identifying lip-sync errors because natural speech maintains perfect synchronization between audio and video, whereas synthetic videos show noticeable misalignment (Cooke, 2024). The detection process looks for specific frame-level artifacts because AI-generated videos display temporal inconsistencies due to lighting mismatches (Cooke, 2024). Instruction-tuned LLMs produce descriptive metadata to detect coherence anomalies, but high-resolution diffusion-based videos are

difficult to detect (Cooke, 2024). The FaceForensics++ system detects irregularities in facial movements through optical flow analysis to detect abnormal video synthesis transitions. Detecting features enables digital media forensic operations while preserving information source trust. These initiatives protect against misinformation as digital content continues to grow (Ghiurău, 2024).

Figure 4. For the Authenticity Investigation, Video Frames are Extracted (Ghiurău, 2024)

The detection of text and images has made significant progress, but audio and video detection face challenges because of their complex nature. The ongoing development of Generative AI demands persistent updates to the detection methods to achieve reliable results (Cooke, 2024).

5. IMPLICATIONS OF SYNTHETIC CONTENT ON USER TRUST AND CREDIBILITY

Synthetic content creation has significant consequences for how users trust and believe online content. User acceptance, together with engagement and decision-making across digital platforms, depends heavily on trust and credibility.

Research indicates that specific elements boost the perceived credibility and trustworthiness of online content. The implementation of preference elicitation features within recommender systems leads to a substantial growth in user trust (Herse et al., 2018). The combination of brand engagement, perceived credibility, perceived benefits, and information quality enhances consumer attitudes towards user-generated content, which affects purchase intentions (Mathur et al., 2021). Websites benefit from logos that display expertise and trustworthiness because they create favorable credibility judgments that boost trust levels and transaction willingness (Lowry et al., 2013). The growing use of synthetic content has created multiple difficulties for users. The combination of crowdsourcing produces adverse effects on content credibility because users lose trust in sources but simultaneously strengthen source trustworthiness through interactive engagement and user control (Huang & Sundar, 2020). The clashing results illustrate the complex nature of trust systems within online platforms, particularly when users confront generated or artificial content. With the increase in demand for synthetic content, organizations must incorporate strong credibility and trust in their systems (Huang & Sundar, 2020). The primary strategy for improving trust involves implementing trust-building design features, using social proof, and maintaining high-quality information. (Huang & Sundar, 2020).

6. COMPLICATIONS AND ETHICAL CONCERNS REGARDING SYNTHETIC CONTENT

6.1 Misinformation and Information

Rapid scattering of misleading information becomes possible through synthetic content, which includes deepfakes and AI-generated content. The deliberate creation and sharing of false data destroys public discussions based on facts and democratic organizations (Hao, 2024).

6.2 Consent and Manipulation

AI tools can replicate voices, images, or writing styles without the subject's permission. This violates individuals' autonomy, consent, and privacy. For instance, synthetic pornography involving real individuals is a serious problem (Hao, 2024).

6.3 Accountability and Transparency

The responsibility for generating or sharing synthetic content remains ambiguous among human creators, machine algorithms, and organizational entities. The absence of clear attribution and transparency creates challenges for enforcing norms and legal standards (Hao, 2024).

6.4 Intellectual Property and Attribution

The training process for generative models involves the use of existing human-created content without permission or payment. This practice creates ethical problems regarding copyright violations and fair use standards, particularly when creators see their work duplicated or transformed (Hao, 2024; Kumar, 2024).

6.5 Identity Theft and Impersonation

Synthetic content can imitate people to commit fraud, damage reputations, or manipulate their opinions. There are risks to personal security, dignity, and truthfulness of representation.

6.6 Bias and Representation

AI-generated content may reflect or amplify the biases present in the training data. This can reinforce stereotypes, marginalize groups, or create misleading portrayals of reality (Hao, 2024).

6.7 Psychological and social impacts

Overexposure to synthetic content may cause confusion, paranoia, or emotional harm (e.g., being targeted by fake personal messages or videos). There is a duty to protect individuals, especially vulnerable groups, from manipulation and distress (Hao, 2024).

6.8 Regulation and Governance

Legal systems face challenges when adapting to the fast-moving development of AI-generated media. The lack of proper regulation creates an environment where dangerous applications can spread freely, while beneficial innovations might be restricted through excessive oversight (Hao, 2024; Kumar, 2024).

6.9 Authenticity and Trust in Media

The public loses confidence in what they encounter in synthetic content. Ethical concern about trust remains fundamental to journalism education and civic engagement because its deterioration produces extensive effects (Hao, 2024; Kumar, 2024).

7. CASE STUDIES OF SYNTHETIC CONTENT IMPACTING ONLINE COMMUNITIES AND INDUSTRIES

Several case studies have demonstrated how synthetic content affects online communities and industries through multiple impact mechanisms.

Deepfakes and virtual reality technologies enable entertainment and creative industries to generate content that matches authentic materials (Sloot, 2024). Synthetic content brings positive benefits through immersive entertainment, but it also creates risks of identity theft and fraud. These technologies serve the retail and financial sectors by enabling new customer experience. Social media platforms and online communities enable users to generate healthcare content, and Cuomo et al. (2020) describe it as "disruptive cooperation" that transforms doctor-patient relationships and personal wellness practices. Through online platforms, smart patients exchange knowledge and experiences that transform established health care delivery systems. According to Jafarov (2025), virtual communities have become essential educational and mentoring platforms. Virtual environments duplicate real-world interactions so that users can participate in authentic learning experiences to develop their skills. Virtual environments enable risk-free learning opportunities and cultural and geographical connections between users. The growing popularity of synthetic content has led to significant difficulties. The future inability to distinguish artificial content from genuine content will have a significant impact on democratic processes, judicial systems, and press operations (Sloot, 2024). A "post-truth society" could result in decreased trust within society, along with reduced individual autonomy. Synthetic content provides various advantages across industries, yet its ethical and regulatory challenges require immediate attention to achieving responsible use in online communities and additional areas (Sloot, 2024).

8. THE ROLE OF TECHNOLOGY AND ALGORITHMS IN COMBATING THE SPREAD OF SYNTHETIC CONTENT

Digital platforms face rapid synthetic content spread through technological and algorithmic progress in detection methods (Giansiracusa, 2021). The increasing realism of AI-generated content requires detection and authentication tools to develop equivalent sophistication. Stylometric analysis serves as a strong method to detect human versus machine-generated writing by examining linguistic fingerprints, including sentence length, vocabulary richness, syntactic structure, and punctuation habits (Giansiracusa, 2021; Ghiurău, 2024). The technique shows its best results in academic and publishing environments because it detects minor writing-style variations that reveal external or automated interference. Stylometric systems perform content analysis by matching suspected materials against established writing samples to detect anomalies that indicate inauthentic authorship (Elkin-Koren, 2020). The implementation of watermarking and digital fingerprinting technologies is another essential strategy (Elkin-Koren, 2020). These methods embed tamper-resistant invisible markers into AI-generated content, including text, images, and videos, which enables the future identification of the source material and synthetic detection. Major AI firms implement watermarking techniques to achieve transparency by enabling platforms and users to detect machine-generated content (Giansiracusa, 2021). Digital fingerprints track specific file characteristics, such as image pixel distribution and audio acoustic signatures, to establish authenticity while detecting file alterations (Ghiurău, 2024).

Researchers are now employing adversarial and robust detection techniques to fight against the growing evasive methods of malicious actors (Ghiurău, 2024). The detection models receive adversarial inputs, which are manipulated to deceive the standard detectors during training. The training process with edge cases makes models more resistant to small changes and better able to detect synthetic content that traditional filters typically allow (Ghiurău, 2024). The ability to resist such attacks is crucial because generative models continue to improve their human-like output capabilities (Ghiurău, 2024). Machine learning models serve as fundamental elements of detection systems because they analyze large datasets to detect patterns that humans cannot. These models perform content classification while detecting inconsistencies, and they gain knowledge about new synthetic content types over time (Ghiurău, 2024).

Table 2. Contemporary Methods with their Main Advantages and Challenges (Ghiurău, 2024)

Approach	Challenges	Benefits
Stylometric analysis	Detecting AI generated text across different domains due to varied context.	Uses stylometric analysis with explainable AI, improving transparency in AI text detection.
Watermark and digital fingerprints.	Detecting realistic deepfakes, which constantly evolves in quality.	Applies watermarking to help verify authentic versus synthetic content.
Adversarial and robust detecting techniques.	Real-time detecting is challenging due to processing speed and model updates.	Develops robust detection for real-time image analysis and authentication.
Machine learning models	GAN- generated images closely resemble real images, making detection challenging.	Builds machine learning models for reliable detection of AI- generated images.
Blockchain	Utilizing blockchain for AI trust while addressing scalability and data privacy concerns.	Blockchain provides verified, tamper-proof information management.

Machine learning technology allows the real-time detection of multiple content types across different platforms through its ability to identify facial expression irregularities in deepfakes and inconsistent metadata in AI-generated articles. Blockchain technology serves as a tool for verifying content authenticity and tracking its origin (Giansiracusa, 2021). Blockchain

systems maintain permanent records of content origin, authorship, and modification history, which enable users and platforms to verify whether digital content has been altered or misidentified (Elkin-Koren, 2020). The Content Authenticity Initiative (CAI), led by Adobe and other tech leaders, uses blockchain technology to add trusted credentials to digital media, establishing consumer trust in its authenticity (Elkin-Koren, 2020). These technologies work together to form an advanced multilayered protection system against synthetic content. These detection and traceability systems, combined with resilience and transparency features, enable platforms, governments, and users to maintain digital integrity during the rapid development of generative technologies (Elkin-Koren, 2020).

9. CONCLUSION

Synthetic content produced by AI has transformed the digital world through its dual role of enabling new opportunities and creating substantial difficulties. Synthetic media provides industries with unmatched capabilities to boost creativity, optimize workflows, and deliver personalized experiences. The nearly indistinguishable quality of AI-generated content, including text images and videos, creates urgent problems regarding authenticity and originality while simultaneously leading to information overload that damages trust and facilitates misinformation spread (Lounge, 2023). The detection methods of stylometric analysis, watermarking, blockchain verification, and multimodal forensics show progress but struggle to match the speed of advancing generative models. The EU AI Act and the Deepfake Accountability Act, together with industry standards such as the Content Authenticity Initiative, serve as necessary regulatory tools to address ethical and legal issues stemming from consent violations, intellectual property disputes, and algorithmic bias. The period between 2025 and 2030 will bring explainable AI advancements together with quantum-resistant watermarking and behavioral biometrics, while global content provenance standards and media literacy programs will emerge to protect societies from swarm propaganda and synthetic identity fraud and autonomous deepfake agents (Lounge, 2023). Success in addressing these issues demands that governments enhance international forensic cooperation, tech companies integrate cryptographic content verification into their AI systems, researchers develop flexible detection frameworks, and people learn essential verification skills. Synthetic content's future success depends on technological resilience together with ethical governance and societal vigilance because responsible use of synthetic media requires safeguarding truth, consent, and authenticity in the digital age (Ghiurău, 2024).

10. REFERENCE

[1] Ghiurău, D., & Popescu, D. E. (2024). Distinguishing Reality from AI: Approaches for Detecting Synthetic Content. Computers, 14(1), 1.

[2] Ciftci, U. A., Demir, I., & Yin, L. (2020). Fakecatcher: Detection of synthetic portrait videos using biological signals. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[3] Lounge, T. (2023, June 15). AI searches worldwide reached an all-time high in May. Mint. https://www.livemint.com/mint-lounge/business-of-life/ai-searches-worldwide-all-time-high-may-111686816202303.html

[4] Subramani, R. A Descriptive Study on Emerging AI Tools in Digital Media Content.

[5] Cooke, D., Edwards, A., Barkoff, S., & Kelly, K. (2024). As Good as A Coin Toss: Human detection of AI-generated images, videos, audio, and audiovisual stimuli. arXiv preprint arXiv:2403.16760.

[6] Bougueffa, H., Keita, M., Hamidouche, W., Taleb-Ahmed, A., Liz-López, H., Martín, A., ... & Hadid, A. (2024). Advances in AI-Generated Images and Videos. International Journal of Interactive Multimedia & Artificial Intelligence, 9(1).

[7] Hao, S., Han, W., Jiang, T., Li, Y., Wu, H., Zhong, C., ... & Tang, H. (2024). Synthetic data in AI: Challenges, applications, and ethical implications. arXiv preprint arXiv:2401.01629.

[8] Kumar, A. (2024). Synthetic data's role in advancing AI models includes opportunities, challenges, and ethical considerations. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 5(1), 443-459.

[9] Giansiracusa, N. (2021). How algorithms create and prevent fake news (pp.) 17-39). Berkeley, CA: Apress.

[10] Elkin-Koren, N. (2020). Contesting algorithms: Restoring public interest in content filtering using artificial intelligence. Big Data & Society, 7(2), 2053951720932296.

[11] Herse, S., Judge, W., Johnston, B., Ebrahimian, D., Ojha, S., Williams, M.-A., Tonkin, M., & Vitale, J. (2018). Do You Trust Me blindly? Factors Influencing Trust in a Robot Recommender System. 7–14. https://doi.org/10.1109/roman.2018.8525581

[12] Mathur, S., Tewari, A., & Singh, A. (2021). Modeling the Factors affecting Online Purchase Intention: The Mediating Effect of Consumer attitude towards user-generated content. Journal of Marketing Communications, 28(7), 725–744. https://doi.org/10.1080/13527266.2021.1936126

[13] Lowry, P. Wilson, B., and Haig, W. L. (2013). A Picture is Worth a Thousand Words: Source Credibility Theory Applied to Logo and Website Design for Heightened Credibility and Consumer Trust. International Journal of Human-Computer Interaction, 30(1), 63–93. https://doi.org/10.1080/10447318.2013.839899

[14] Huang, Y., & Sundar, S. S. (2020). Do We Trust the Crowd? Effects of Crowdsourcing on Perceived Credibility of Online Health Information. Health Communication, 37(1), 93–102. https://doi.org/10.1080/10410236.2020.1824662

[15] Van Der Sloot, B. (2024). Regulating Synthetic Society. hart. https://doi.org/10.5040/9781509974979

[16] Cuomo, M. T., Festa, G., Tortora, D., Giordano, A., Martinelli, E., & Metallo, G. (2020). User-generated content in the era of digital well-being: A netnographic analysis in a healthcare marketing context. Psychology & Marketing, 37(4), 578–587. https://doi.org/10.1002/mar.21327

[17] Jafarov, S. (2025). Simulated Communities and Online Mentoring. International Journal of Current Science Research and Review, 08(01). https://doi.org/10.47191/ijcsrr/v8-i1-38.