Modern Era Mythmaking: AI, Conspiracies and The Digital Age

W.M.C.S. Wijesinghe
Department of Computer & Data Science,
Faculty of Computing, NSBM Green University,
Homagama, Sri Lanka
mcswijesinghe@students.nsbm.ac.lk

Diluka. Wijesinghe
Department of Software Engineering & Computer Security,
Faculty of Computing, NSBM Green University,
Homagama, Sri Lanka
diluka.w@nsbm.ac.lk

***Abstract*— As society becomes increasingly interconnected and algorithmically mediated, myth-making has not diminished but evolved. In this digital age, generative artificial intelligence (AI) plays a pivotal role in amplifying misinformation and conspiracies, contributing to the construction of persuasive modern myths. This paper explores how AI technologies facilitate the creation and dissemination of misinformation, examining psychological factors, algorithmic influence, and the socio-technical systems that enable belief formation. Employing a qualitative methodology with empirical support from two case studies—Meta's Galactica model and GPT-2's news article generation experiments—the paper identifies critical vulnerabilities in public perception and trust in digital information. Drawing on updated literature and theoretical frameworks, we propose a multi-level strategy encompassing digital literacy, regulatory interventions, and machine-learning-based detection tools to mitigate the risks posed by AI-driven myth-making. The findings offer vital insights for policymakers, technologists, and educators aiming to safeguard truth in an increasingly AI-mediated society.**

Keywords*— Conspiracies, digital literacy, Generative AI, AI regulation

I. INTRODUCTION

Throughout human history, storytelling and myth-making have served as fundamental mechanisms for communities to comprehend complex realities and navigate uncertainty. From ancient oral traditions to religious parables, societies have consistently constructed narratives that explain the unexplainable, challenge established authorities, and foster shared beliefs—regardless of how fantastical or improbable these beliefs may appear to outsiders [1][2]. This enduring human tendency toward mythological thinking has not disappeared in the digital age; rather, it has undergone a profound transformation in both form and scale.

The contemporary landscape of myth-making represents a marked departure from traditional person-to-person transmission. Digital platforms and sophisticated algorithms now facilitate the rapid dissemination of narratives across global networks, fundamentally altering how false beliefs emerge, evolve, and take root within communities. Social media ecosystems, with their emphasis on engagement-driven content distribution, have created unprecedented opportunities for conspiratorial narratives to reach and influence vast audiences.

The emergence of artificial intelligence technologies has introduced an additional layer of complexity to this phenomenon. Advanced generative AI systems can now produce photorealistic images, craft persuasive textual content, and synthesize convincing audio-visual materials that closely mimic authentic human communication [3]. These capabilities have dramatically lowered the technical barriers to creating sophisticated disinformation, enabling the production of content that appears credible and authoritative while being entirely fabricated. The resulting erosion of traditional markers of authenticity poses significant challenges to information verification and public discourse.

Despite the growing recognition of AI's role in contemporary information disorders, a substantial research gap exists at the intersection of artificial intelligence, human psychology, and digital platform dynamics. While existing scholarship has examined conspiracy theories through psychological and sociological lenses, and separate bodies of work have investigated AI-generated content and algorithmic amplification, insufficient attention has been paid to how these elements converge to create novel forms of mythological thinking in digital spaces.

This research addresses this gap by investigating how artificial intelligence technologies specifically contribute to the creation and amplification of modern conspiratorial narratives. By analyzing the interplay between human cognitive biases, platform algorithms, and AI generative capabilities, this study seeks to illuminate the mechanisms through which false narratives emerge and proliferate in online environments. The central inquiry examines what occurs when deeply rooted human psychological tendencies encounter sophisticated technological tools, and explores the implications of this convergence for information processing, belief formation, and societal discourse in the digital age.

II. LITERATURE REVIEW

A. Historical Context of Conspiracy Theories and Myth-Making

Conspiracy theories represent a persistent phenomenon throughout human history rather than a uniquely modern development [4]. Historical analysis reveals that populations have consistently turned to conspiratorial explanations during periods of uncertainty, fear, and social upheaval. This pattern became particularly evident during the 2016 US presidential elections, which witnessed a marked increase in conspiracy theories and misinformation [1], and during the COVID-19 pandemic, where the proliferation of false information led WHO Director-General Tedros Adhanom Ghebreyesus to declare an "infodemic" [5].

Hofstadter's seminal work on the "paranoid style" in American politics demonstrated how deeply embedded conspiracy thinking has become within cultural and political discourse [6]. The historical record reveals troubling connections between conspiracy theories and various forms of social harm, including prejudice, persecution, and genocide. Notable examples include the anti-Semitic propaganda of Nazi Germany and AIDS denialism by the South African government [7].

Contemporary scholarship suggests that while conspiracy theories have historically emerged during times of crisis and subsequently faded, modern technologies—particularly artificial intelligence and digital media—may be fundamentally transforming how these narratives spread and persist, potentially giving traditional patterns of conspiratorial thinking unprecedented reach and influence.

B. Psychological Foundations of Conspiracy Belief

The psychological literature identifies several key factors that contribute to conspiracy belief formation and maintenance. Research demonstrates that individuals generally tend to accept information from others as accurate unless obvious contradictory evidence is present [8]. This baseline trust becomes problematic when combined with confirmation bias, whereby people preferentially consume, endorse, and favor information that aligns with their pre-existing beliefs and ideologies [9][1].

Confirmation bias contributes to the formation of online communities centered around similar ideologies, creating what researchers term "echo chambers" where conspiratorial beliefs can flourish unchallenged. Pennycook and Rand [10] introduced the concept of "pseudo-profound bullshit receptivity"—the tendency to assign profound meaning to meaningless statements—and demonstrated its correlation with both fake news susceptibility and difficulty distinguishing between authentic and fabricated information. Their research further revealed that individuals who overestimate their own intellectual capabilities are more likely to rate false news and conspiracy theories as accurate.

A comprehensive study by Bryanov and Vziatysheva [9] identified additional factors contributing to conspiracy belief, including endorsement by trusted sources, deception bias, and the influence of momentary emotions. Particularly relevant to digital environments, their research found that social media engagement metrics, such as the number of likes, significantly increased perceived credibility regardless of content veracity. The problem is compounded by the fact that even high-reputation sources occasionally amplify unverified claims [11].

C. Digital Amplification and the Internet's Role

The relationship between internet technology and conspiracy theory proliferation presents a complex paradox. Clarke [2] initially argued that internet access would facilitate robust criticism of conspiracy theories through real-time fact-checking and unprecedented access to information. However, empirical evidence suggests a different outcome, with approximately three-quarters of Americans attributing the spread of modern conspiracy theories to social media and internet platforms [12].

The internet's low barrier to entry for content publication has created an environment where conspiratorial content is freely available, placing the burden of verification on individual users [16]. While conspiracy engagement occurs both online and offline, research indicates that the ease of sharing and endorsing content without accuracy verification significantly accelerates conspiracy dissemination [13].

The digital era has fostered what scholars term a "Post-Truth ecosystem" [13][14], characterized by information flows that prioritize emotional appeal and personal conviction over objective facts. This ecosystem encompasses deliberate fakes, hoaxes, and misinformation often motivated by financial incentives rather than ideological commitment.

D. Artificial Intelligence and Contemporary Challenges

The emergence of artificial intelligence, particularly generative AI systems, represents a qualitative shift in conspiracy theory creation and dissemination capabilities. Modern AI systems enable unprecedented ease of content generation across multiple media formats, including text, images, and video. This technological advancement has fundamentally altered the resource requirements for information manipulation, which historically demanded significant human and financial resources available primarily to powerful actors [20].

Empirical studies demonstrate that readers perceive AI-generated text, when curated by humans, as equally credible to human-authored content covering identical events. More concerning, research shows that AI text models can produce credible-sounding news articles at scale without human intervention [17]. The development of deepfake technology has enabled the creation of entirely fabricated multimedia content, from false advertisements to pornographic material and fabricated "evidence" supporting conspiratorial claims [18].

State-level actors have already begun exploiting these capabilities. Research documents Russia's use of automated social media bots to amplify pro-Russian propaganda within Europe [19], suggesting that AI-powered disinformation represents a logical technological progression. Studies indicate that AI-generated misinformation tends to feature vivid storytelling and exaggerated or fabricated conclusions about future events [3], characteristics that may enhance its persuasive impact.

The convergence of internet-enabled rapid dissemination with AI-facilitated content creation presents unprecedented challenges for information integrity and democratic discourse. This technological synthesis enables the mass production and distribution of sophisticated misinformation at scales and speeds previously impossible, fundamentally altering the landscape of conspiracy theory propagation.

III. METHODOLGY

This study adopts a multi-layered qualitative methodology, integrating document analysis, comparative case study, and theoretical triangulation. The objective is to deeply investigate how generative AI contributes to the formation and propagation of digital-age myths, particularly misinformation and conspiracies. The methodology includes the following components:

A. Research Design

We employed a comparative case study design to analyze two representative and high-impact instances of AI-generated misinformation. This design enables in-depth contextual analysis and cross-case comparison to extract broader insights into AI's role in myth-making.

B. Case Selection Criteria

Cases were chosen based on the following criteria:

- High relevance to AI-generated misinformation and public trust
- Contrasting domains (scientific discourse vs. news media)
- Availability of empirical data and credible documentation

The selected cases are:

- Meta's Galactica: a large language model designed for scientific content creation, which demonstrated the risks of misinformation in academic contexts.
- GPT-2 Experiments: an empirical study on public perception of AI-generated news credibility, which provides experimental validation of AI's impact on belief systems.

C. Data Collection

Data were collected from multiple reliable sources, encompassing peer-reviewed academic articles and empirical studies, official documentation and technical reports, as well as media coverage, platform communications, and publicly available AI model outputs with accompanying expert commentary.

D. Analytical Framework

The analysis was guided by an interdisciplinary theoretical framework drawing on:

- Media epistemology: to evaluate how truth claims are shaped in digital discourse
- Digital rhetoric and narrative framing: to assess how AI-generated content influences perception
- Information trust theory: to interpret user vulnerability to misinformation based on content characteristics

Key analytical dimensions included:

- Content credibility: Does the AI-generated content appear trustworthy?
- Autonomy and scale: To what extent can AI operate without human oversight in spreading misinformation?
- Sociocultural impact: How does this content affect public knowledge, trust, and behavior?

E. Validity and Limitations

To ensure validity, we triangulated sources and cross-validated claims with peer-reviewed data. However, the study is limited by its qualitative nature and reliance on secondary data. Future research should incorporate experimental user studies and real-time AI output analysis to extend empirical rigor.

IV. CASE STUDIES

A. Case Study I: The Failure of Meta's Galactica AI

Meta's Galactica large language model represents a significant cautionary tale in the deployment of AI systems for specialized domains. Launched on November 15, 2022, the model was abruptly discontinued just three days later on November 17, following widespread criticism and public backlash. This rapid shutdown highlighted critical oversights in Meta's approach to AI deployment and the inherent challenges of creating reliable scientific AI assistants.

Galactica was designed with ambitious goals to "organize science" and assist researchers by providing access to scientific knowledge. The model was trained on an extensive dataset comprising 48 million academic papers, textbooks, lecture notes, and reference materials, which Meta characterized as "a large and curated corpus of humanity's knowledge." Despite this substantial training foundation, the system exhibited fundamental flaws that undermined its intended purpose.

The most critical limitation of Galactica was its inability to distinguish between factual and fictional information, a fatal flaw for a system designed to support scientific research. The model consistently generated content that was factually incorrect, exhibited clear biases, and produced outputs that were often nonsensical. Notably, the system failed to solve basic mathematical problems, raising serious questions about its reliability for any scientific application.

This case study illustrates the potential dangers of premature AI deployment, particularly in domains requiring high accuracy and reliability. The failure of Galactica demonstrates how even well-intentioned AI tools can become vectors for misinformation when released without adequate testing, validation, and regulatory oversight. The incident serves as a stark reminder that the sophistication of training data does not guarantee the accuracy or reliability of AI outputs.

B. Case Study II: GPT-2's Influence on News Credibility Perception

A comprehensive study conducted by Kreps, McCain, and Brundage examined the capacity of different versions of GPT-2 to generate convincing news content and its implications for information credibility. This research utilized three variants of the GPT-2 model with varying computational complexity:

- Medium Model: 355 million parameters
- Large Model: 774 million parameters
- Extra Large Model: 1.5 billion parameters

All models were trained on 40 GB of highly-rated Reddit content, providing a foundation for generating human-like text. The researchers designed multiple experiments to assess the believability of AI-generated content compared to professionally written news articles.

3) Experiment 1: Credibility Assessment of AI-Generated News

The first experiment evaluated how credible AI-generated news articles appeared to American readers. Researchers compared articles produced by the three GPT-2 variants against a baseline article from *The New York Times*. To ensure fair comparison, only the most factually accurate AI outputs

were selected for evaluation. Participants were asked to rate the credibility of randomly presented articles without knowing their source.

The results revealed concerning patterns in public perception of AI-generated content:

- Articles from the smallest model (355M parameters) were perceived as credible but noticeably less so than the human-written baseline
- Content from larger models (774M and 1.5B parameters) achieved credibility ratings statistically indistinguishable from the professional news article
- Most remarkably, the 774M parameter model's outputs were rated as significantly more credible than the human-written article

These findings suggest that AI-generated content can achieve or exceed human-level credibility in public perception, raising significant concerns about the potential for AI to be used in creating convincing misinformation.

1) Experiment 2: Unfiltered AI Content Generation

The third experiment addressed a critical real-world scenario: the potential for AI models to generate believable misinformation without human curation or editing. This experiment simulated conditions where malicious actors might deploy AI systems for automated misinformation campaigns. Using a *New York Times* article about North Korean ship seizures as a prompt, researchers generated 300 articles from each GPT-2 model variant.

Unlike previous experiments that selected only the highest-quality outputs, this study evaluated the full spectrum of AI-generated content with minimal filtering, reflecting realistic deployment conditions. The analysis yielded several important insights:

- Larger models (774M and 1.5B parameters) produced content that was generally perceived as more credible than the smallest model, though the improvement was modest
- Despite statistical differences in peak performance, the overall credibility distributions across all model sizes showed significant overlap
- The results indicated diminishing returns in perceived believability as model size increased, suggesting that computational power alone does not linearly improve misinformation potential

This experiment demonstrates that even unfiltered AI outputs can achieve concerning levels of credibility, highlighting the need for robust detection mechanisms and regulatory frameworks to address the potential misuse of AI in information warfare and misinformation campaigns.

V. RESULTS AND DISCUSSIONS

A. The Paradox of Scientific Authority in AI-Generated Content

The analysis of Meta's Galactica system reveals a critical disconnect between dataset comprehensiveness and truthfulness verification capabilities. Despite training on extensive scientific literature, Galactica demonstrated an inability to distinguish between scientifically valid and invalid information. Instead, the system consistently produced outputs that mimicked the stylistic conventions of academic writing—including formal tone, technical terminology, and structured argumentation—while lacking substantive scientific accuracy. This phenomenon represents what we term "performative credibility," where AI systems replicate the surface-level markers of authority without the underlying epistemological rigor.

The implications of this finding extend beyond technical limitations to fundamental questions about how authority and credibility are constructed and perceived in digital information environments. The system's outputs possessed sufficient verisimilitude to appear credible to users who lack specialized domain knowledge, creating a dangerous gap between perceived and actual reliability.

B. Public Perception and the Credibility Gap

Experimental findings from the GPT-2 news credibility studies provide empirical evidence for concerning patterns in human information processing. Participants demonstrated comparable credibility ratings for AI-generated content and human-authored news articles, with larger language models showing particularly strong performance in believability metrics. This equivalence in perceived credibility suggests that current AI systems have crossed a threshold where their outputs can no longer be easily distinguished from human-generated content by lay audiences.

The results indicate that individuals rely heavily on presentation quality and confirmation bias rather than rigorous fact-checking when evaluating information credibility. This cognitive bias creates systematic vulnerabilities that can be exploited by malicious actors seeking to disseminate false information. The findings align with broader research on motivated reasoning and the role of cognitive shortcuts in information processing.

C. Mechanisms of Misinformation Amplification

The convergence of AI capabilities and conspiracy theory propagation presents a novel threat vector in the information landscape. AI systems like Galactica and GPT-2 provide conspiracy theorists with sophisticated tools for generating content that bears the hallmarks of legitimate discourse. The technical sophistication of AI-generated text—characterized by coherent argumentation, appropriate citation formats, and domain-specific vocabulary—can lend false credibility to unfounded claims.

This represents a qualitative shift in misinformation production. Traditional conspiracy theories often suffered from obvious markers of dubious origin, such as poor writing quality, lack of supporting documentation, or clearly biased sources. AI-generated content eliminates these quality barriers, enabling the production of polished misinformation that can more easily evade detection by both automated systems and human readers.

D. Scalability and Accessibility Concerns

The democratization of AI content generation tools has lowered traditional barriers to sophisticated misinformation

production. Users with minimal technical expertise can now generate volumes of seemingly credible content at unprecedented scale and speed. This accessibility multiplies the potential for harm, as the resource requirements for producing convincing false content have decreased dramatically.

Social media platforms compound this risk by providing efficient distribution mechanisms for AI-generated content. The combination of low production barriers and high-velocity distribution creates conditions favorable to rapid misinformation spread, particularly within communities predisposed to conspiratorial thinking.

E. AI as Myth-Making Infrastructure

These findings suggest that contemporary AI systems function as enablers of what we characterize as "digital myth-making." Unlike traditional myth-making processes that developed over extended periods through oral tradition and cultural transmission, AI-mediated myth-making can occur rapidly and at scale. AI systems serve dual roles in this process: they generate the content that forms the foundation of false narratives while simultaneously providing that content with markers of credibility that facilitate acceptance and propagation.

The implications extend beyond individual instances of misinformation to broader concerns about epistemic security in democratic societies. When AI systems can generate content that is indistinguishable from legitimate scientific or journalistic discourse, the foundations of evidence-based decision-making become vulnerable to systematic erosion.

F. Regulatory and Technical Implications

The documented capabilities and risks associated with AI-generated misinformation highlight the urgent need for comprehensive regulatory frameworks and technical safeguards. Current approaches to content moderation, which often rely on post-hoc detection and removal, may prove insufficient against AI-generated content that can evade traditional detection mechanisms.

The results suggest that effective mitigation strategies must address both technical and social dimensions of the problem. Technical solutions might include improved detection algorithms, watermarking systems, or content provenance tracking. Social interventions could focus on digital literacy education and the development of critical evaluation skills, specifically adapted to AI-generated content.

These findings contribute to growing evidence that the deployment of powerful AI systems without adequate safeguards poses significant risks to information integrity and democratic discourse. The research underscores the need for proactive rather than reactive approaches to AI governance, particularly in domains where false information can have serious societal consequences

VI. CONCLUSION

Generative AI has introduced a paradigm shift in how misinformation and modern myths are created, perceived, and propagated. As illustrated by the Galactica and GPT-2 case studies, AI-generated content can possess a deceptive credibility that deeply influences public perception and belief formation. The synergy between psychological biases, algorithmic amplification, and the believability of AI-generated narratives creates a fertile ground for digital-age myth-making.

This paper has demonstrated that the threat posed by generative AI is not solely technical but also epistemological, affecting how society defines and interacts with truth. It calls for urgent, interdisciplinary interventions to address this challenge. These include advancing AI content detection tools, integrating digital media literacy into education systems, and enforcing platform accountability through transparent regulation.

Ultimately, preserving the integrity of knowledge in an AI-driven world will depend on our ability to balance innovation with responsibility. As we continue to benefit from generative technologies, we must remain vigilant to their misuse, fostering a digital environment where truth is resilient, information is trustworthy, and myths are questioned rather than blindly believed.

REFERENCES

[1] C. Beauvais, "Fake news: Why do we believe it?," Joint Bone Spine, vol. 89, no. 4, p. 105371, 2022.

[2] S. Clarke, "Conspiracy theories and the Internet: Controlled demolition and arrested development," Episteme, vol. 4, no. 2, pp. 167–180, 2007.

[3] J. Zhou, Y. Zhang, Q. Luo, A. G. Parker, and M. De Choudhury, "Synthetic lies: Understanding AI-generated misinformation and evaluating algorithmic and human solutions," in Proc. 2023 CHI Conf. on Human Factors in Computing Systems, pp. 1–20, 2023.

[4] J.-W. Van Prooijen and K. M. Douglas, "Conspiracy theories as part of history: The role of societal crisis situations," Memory Studies, vol. 10, no. 3, pp. 323–333, 2017.

[5] J. Zarocostas, "How to fight an infodemic," The Lancet, vol. 395, no. 10225, p. 676, 2020.

[6] R. Hofstadter, "The paranoid style in American politics," Harper's Magazine, Nov. 1964.

[7] K. M. Douglas, J. E. Uscinski, R. M. Sutton, A. Cichocka, T. Nefes, C. S. Ang, and F. Deravi, "Understanding conspiracy theories," Political Psychology, vol. 40, pp. 3–35, 2019.

[8] S. Lewandowsky, U. K. H. Ecker, C. M. Seifert, N. Schwarz, and J. Cook, "Misinformation and its correction: Continued influence and successful debiasing," Psychological Science in the Public Interest, vol. 13, no. 3, pp. 106–131, 2012.

[9] K. Bryanov and V. Vziatysheva, "Determinants of individuals' belief in fake news: A scoping review," PLoS One, vol. 16, no. 6, p. e0253717, 2021.

[10] G. Pennycook and D. G. Rand, "Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking," Journal of Personality, vol. 88, no. 2, pp. 185–200, 2020.

[11] A. Zubiaga, M. Liakata, R. Procter, G. Wong Sak Hoi, and P. Tolmie, "Analysing how people orient to and spread rumours in social media by looking at conversational threads," PLoS One, vol. 11, no. 3, p. e0150989, 2016.

[12] A. M. Enders, J. E. Uscinski, M. I. Seelig, C. A. Klofstad, S. Wuchty, J. R. Funchion, M. N. Murthi, K. Premaratne, and J. Stoler, "The relationship between social media use and beliefs in conspiracy theories and misinformation," Political Behavior, pp. 1–24, 2021.

[13] C. Birchall and P. Knight, "Do your own research: Conspiracy theories and the internet," Social Research: An International Quarterly, vol. 89, no. 3, pp. 579–605, 2022.

[14] D. S. Artamonov, E. N. Medvedeva, S. V. Tikhonova, and Z. A. Slivnaia, "Digital mythology: A new direction in the study of social myths," European Proceedings of Social and Behavioural Sciences, 2021.

[15] M. Johann and L. Bülow, "One does not simply create a meme: Conditions for the diffusion of Internet memes," International Journal of Communication, vol. 13, p. 23, 2019.

[16] J. E. Uscinski, D. DeWitt, and M. D. Atkinson, "A web of conspiracy? Internet and conspiracy theory," in Handbook of Conspiracy Theory and Contemporary Religion, pp. 106–130, 2018.

[17] S. Kreps, R. M. McCain, and M. Brundage, "All the news that's fit to fabricate: AI-generated text as a tool of media misinformation," Journal of Experimental Political Science, vol. 9, no. 1, pp. 104–117, 2022.

[18] M. R. Shoaib, Z. Wang, M. T. Ahvanooey, and J. Zhao, "Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models," in Proc. 2023 Int. Conf. on Computer and Applications (ICCA), pp. 1–7, 2023.

[19] T. C. Helmus, E. Bodine-Baron, A. Radin, M. Magnuson, J. Mendelsohn, W. Marcellino, A. Bega, and Z. Winkelman, Russian Social Media Influence: Understanding Russian Propaganda in Eastern Europe. Santa Monica, CA, USA: Rand Corporation, 2018.

[20] R. W. Zmud, "Opportunities for strategic information manipulation through new information technology," in Organizations and Communication Technology, pp. 95–116, 1990.

[21] W. D. Heaven, "Why Meta's latest large language model survived only three days online," MIT Technology Review, Nov. 18, 2022. [Online]. Available: https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/

[22] J. Ryan, "Meta trained an AI on 48 million science papers. It was shut down after two days," CNET, Dec. 14, 2022. [Online]. Available: https://www.cnet.com/science/meta-trained-an-ai-on-48-million-science-papers-it-was-shut-down-after-two-days/

[23] D. Caled and M. J. Silva, "Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation," Journal of Computational Social Science, vol. 5, no. 1, pp. 123–159, 2022.