iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

ARTIFICIAL INTELLIGENCE, ETHICS, AND THE SPREAD OF MISINFORMATION

Promise Enyindah[1]
ORCID ID - 0000-0001-6246-7077
Email: promise.enyindah@uniport.edu.ng
https://orcid.org/0000-0001-6246-7077
Senior Professor, Department of Computer Science, University of Port Harcourt,
Choba, Rivers State, Nigeria.

Ogbonnaya Stephen Success[2]
Email: stephen_ogbonnaya@uniport.edu.ng
Student, Department of Computer Science, University of Port Harcourt,
Choba, Rivers State, Nigeria.

ABSTRACT
Fast progress in artificial intelligence, especially systems that generate new content, has reshaped how information is made and shared. Though such tools improve speed and reach, they simultaneously raise deep ethical concerns, especially around widespread fake content. Because machine-made text, pictures, sound, and videos closely resemble real material, spotting deception grows harder by the day. What follows explores moral issues tied to AI-fueled falsehoods, looking at responsibility, openness, unfair distortions, and control over personal choice. This review analyzes approximately 45 scholarly articles and policy documents published between 2018 and 2025, focusing on ethical challenges, platform amplification mechanisms, and regulatory responses to AI-generated misinformation. Starting from a blend of scholarly work and official papers, the analysis points to weaknesses in how things are currently managed. What stands out is the importance of aligned efforts - ethical oversight joined with updated rules - to protect the reliability of information online. Though progress exists, mismatches remain across systems meant to respond. Findings indicate that AI significantly increases misinformation scalability, while existing ethical and regulatory frameworks remain fragmented and largely reactive.

Keywords:
Artificial Intelligence, Ethics, Misinformation, Disinformation, AI Governance, Digital Policy.

1. INTRODUCTION

Artificial intelligence (AI) has rapidly transitioned from an emerging technology to an integral component of contemporary digital ecosystems. From search engines and recommendation systems to social media platforms and automated content creation tools, AI increasingly shapes how information is produced, distributed, and consumed. Recent advances in generative artificial intelligence—particularly systems capable of producing human-like text, images, audio, and video—have expanded the scale and realism of digital content generation. While these developments offer substantial benefits across education, healthcare, media, and creative industries, they also introduce significant ethical and societal challenges.

The most concerns about the use of generative AI systems are the way they may spread misinformation and disinformation. While disinformation and misinformation have been around since the beginning of the digital environment, the use of AI systems greatly increases the volume, velocity, and perceived trustworthiness of the disinformation and misinformation. The ability of AI systems to create persuasive news articles, images, voices, and deep fake videos at low costs and within a matter of seconds has the potential to spread disinformation and misinformation to vast numbers of people. Thus, the trustworthiness of the information available on the internet becomes questionable.

Besides the technical risks, the rise of AI-generated misinformation poses ethical concerns. For example, the issue of accountability arises when misinformation is created autonomously, and it is not clear who is responsible, the developer or the end user. The issue of transparency arises when the end user is not able to distinguish between the real and the artificial, as the AI-generated information is not labeled as such. Another ethical concern is that

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

the biased information used to train the AI may perpetuate discriminatory messages, while the persuasive power of the AI may lead to the loss of autonomy on the part of the end user.

In the face of the challenges posed by the rise of AI-generated misinformation, governments, international organizations, and regulatory bodies are developing ethical guidelines and policy frameworks to address the use of AI. For example, the European Union is developing the Artificial Intelligence Act, UNESCO is developing guidelines on the ethics of AI, while various countries are developing their own AI strategies. Despite the efforts made, the current regulatory responses to the risks posed by the rise of AI are fragmented, unevenly implemented, and often reactive rather than proactive. This is because the rate of technological change is often faster than the mechanisms that are put in place to address the risks.

While previous literature has addressed the issue of misinformation, the role of algorithms, and the ethics of AI, there is limited literature that has addressed the ethical principles, the role of the platform, and the global regulatory responses to the rise of AI-driven misinformation. This study, therefore, seeks to bridge the gap that has been created by the lack of literature on the role of ethical failures and the limitations of the regulatory responses to the rise of AI-driven misinformation. This study, therefore, seeks to provide a thematic analysis of the literature on the role of ethical failures and the limitations of the regulatory responses to the rise of AI-driven misinformation.

2. METHODOLOGY

The current research utilizes a qualitative systematic literature review method to explore the ethical issues and governance strategies concerning AI-driven misinformation. The current research utilizes a systematic review methodology given the advantages regarding transparency, replicability, and comprehensiveness of pertinent literature it offers.

2.1. Data Sources and Search Strategy

The data regarding the literature have been obtained from Google Scholar, Scopus, IEEE Xplore, and the ACM Digital Library. It has also been complemented by applicable official policy documents and guidelines from different international and national bodies, such as UNESCO and the European Commission, the United States Office of Science and Technology Policy, among others. In this regard, the literature search strategy involved terms like "artificial intelligence ethics," "AI-generated misinformation," "disinformation," "deepfakes," "algorithmic amplification," "AI governance," "digital platform regulation," and so forth. Therefore, Boolean search operators were also used while executing the search strategy to secure maximally relevant literature.

2.2. Inclusion and Exclusion Criteria

The selection of literature is done based on clearly defined criteria for inclusion and exclusion. The literature includes journal articles, conference proceedings, and official policy documents in English from 2018-2025. This is the recent time frame to capture the rapidly evolving generative AI technologies. The excluded sources are opinion pieces without empirical or theoretical underpinnings, studies of misinformation unrelated to AI, duplicate sources, and sources that do not focus explicitly on ethical, governance, and socio-impact issues surrounding artificial intelligence.

2.3. Study Selection and Data Analysis

Initially, the sources were screened through title and abstract screening, which enabled the researchers to narrow down the search to relevant literature. This was further narrowed down through the critical reading of the full text and allowed for confirmation of the sources' alignment with the objectives of the present study. After this, a total of 40 to 50 sources were selected for final analysis. The thematic analysis will allow for the identification of themes that will recur thereby permitting the identification of robust patterns and themes from the material. Based on these themes, the identified thematic structure includes

1) types of misinformation generated by AI,
2) ethical issues like transparency, accountability, fairness, and autonomy,
3) Social media algorithms driving the spread of misinformation,
4) regulatory and policy measures, and
5) detection and mitigation strategies

2.4. Synthesis and Interpretation

Such findings from academic literature and policy documents synthesize to identify convergences and gaps between ethical theory, technological practice, and regulatory implementation. By comparison across disciplines

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

and governance levels, insights highlight weaknesses in current oversight frameworks and areas where ethical principles are poorly translated into enforceable policy. This systematic qualitative approach thus provides an evidential structure within which the impacts of artificial intelligence on misinformation risks can be mapped while unearthing the considerable limitations of its ethical and regulatory responses to date.

3. RESULTS

There are clearly emerging patterns within the official guidelines and normative documents that pertain to AI-driven disinformation. These outcomes, which are thematically grouped, trace how systems propagate misleading content, what moral dilemmas ensue, and how agencies undertake attempts at oversight.

3.1. Types of False Information Created by Artificial Intelligence

The reviewed literature identifies various AI-driven misinformation types, primarily categorized via content modality. Thus, as summarized in Table 1, these include text-based outputs, synthetic images, audio impersonations, and AI-generated or manipulated videos.

Text-based AIs are often associated with making fake news articles, misleading social media posts, and fabricated academic citations. Recent evidence shows that this type of information-generating technology can mass-produce content at an unprecedented scale, hence increasing the possibilities for misinformation and undermining detection effectiveness (Vosoughi et al., 2018; Pennycook et al., 2021).

Generative methods for image production have been advanced to a level where synthetic images can hardly be distinguished from real photographs of any event. Such research highlights that these synthetic images have also been used in misleading news and creating altered images of protests, thus further weakening trust in digital media (European Commission, 2023).

Audio synthesis tools develop the synthesis of a human voice. This has been previously evident in cases of voice cloning associated with impersonation, fraud, and deceiving voices in emergencies. The same occurred with video deepfakes, largely discussed in the literature regarding their potential for political narrative manipulation and their impact on personal reputation (UNESCO, 2021; Romanishyn et al., 2025).

AI Modality	Description	Example Use Case	Ethical Risk
Text-based AI	Automatically generated articles, comments, or chat responses	Fake news articles, fabricated academic references	High risk of deception, hallucinations
Image generation	Synthetic images indistinguishable from real photos	Fake protest images, altered news photos	Visual manipulation, loss of trust
Audio synthesis	AI-cloned human voices	Fake emergency calls, impersonation scams	Identity theft, fraud
Video deepfakes	AI-generated or altered videos	Political deepfakes, celebrity impersonation	Political manipulation, reputational harm

(Table 1. Forms of AI-generated misinformation)

Result Interpretation:

The most logical assumption is that audiovisual types of misinformation are the most severe types because they are considered the most genuine. However, text-based misinformation is disseminated most easily and, to some extent, is what sounds and images stick to. The realistic factor draws people in; the ease of sharing sends false words far and wide.

3.2. Ethical Issues Found in AI and False Information

The above-discussed literature consistently categorized some of the general ethical issues that arise with respect to the generation of misinformation with AI, as structured in the following table:

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

The core ethical issues relate to the absence of transparency regarding the authorship and creation of AI-generated content, the accountability for harm not being ascribed to anyone in particular, the assumption of bias, which has originated from training data, and threats to individual autonomy. Other studies have pointed to the lack of labeling mechanisms, which has made it difficult for them as users to distinguish between real and synthetic content. UNESCO (2021)

Other research pointed out that, in case of harm, accountability is diffuse across developers, platforms, and users, creating legal and ethical uncertainty. (White House OSTP, 2022). Moreover, bias can be embedded within training data for AI, thus reinforcing messages that are discriminatory; persuasive AI systems are able to subtly shift behavior in users themselves. (Lu & Hu, 2025).

These show that ethical risks flow over various stages of content generation and diffusion using AI.

Meaning the observers pointed out as outstanding is that the lack of transparency was one of the greatest ethical faults because this would prevent them from making informed choices about the materials generated by artificial intelligence.

Ethical Principle	Observed Issue	Impact
Transparency	Content made by artificial intelligence usually lacks labels	Users unable to distinguish real vs fake
Accountability	Unclear responsibility for harm	Legal ambiguity
Fairness	Bias inherited from training data	Discriminatory narratives
Autonomy	Manipulative AI persuasion	Undermines informed decision-making

(Table 2. Ethical challenges associated with AI-generated misinformation)

Result Interpretation:

What stood out most was how often observers pointed to missing openness - as a core ethical shortcoming, since it blocks people from making informed judgments about material produced by artificial intelligence.

3.3. Platform Algorithmic Amplification of AI-Generated Misinformation

The literature reviewed consistently identifies platform algorithms as the main driver behind the amplification of AI-generated misinformation. As outlined in Table 3, through recommendation systems, engagement-based metrics, automation tools, and personalized feeds, the timing can thus be much faster for misleading information traveling through digital platforms.

A number of works have indicated that the recommendation algorithms prioritize, above all else, the maximization of user engagement rather than an accurate account of the underlying content. This created a newly generated AI-constructed information that was sensational or emotionally charged, more likely to come through the algorithm and distribute widely, yet not factually correct (Vosoughi et al., 2018). This injected visibility into deepfakes; synthetic narratives further catalyzed their metastasis through these networks.

Engagements like, share, or comment, the cycle gains momentum. When false claims spark intense feelings, people react more often - studies confirm this pattern clearly (Pennycook et al., 2021). Because systems treat frequent engagement as proof of worth, deceptive posts made by artificial intelligence sometimes spread wider and last longer than accurate ones.

Bots along with automated schedulers often speed up the spread of false information. Because these systems push out artificial content quickly and together, they flood websites before people can review it by hand. What happens next? Users see more of what they already liked, thanks to recommendation engines shaping their feeds. That repetition traps them in loops where deceptive stories keep reappearing.

What we see across studies is clear: when platforms favor attention-grabbing content, false or synthetic info spreads faster than accurate material. Design choices shape what goes viral - not truth. Each tweak to an algorithm boost reaches based on reaction, not reliability. Visibility often follows excitement, rarely fact-checks. These systems lift up what keeps eyes glued, regardless of accuracy. Engagement rules tend to sideline quality control. The outcome? Misinformation thrives where response rates rise.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Platform Feature	Function	Effect on Misinformation
Recommendation systems	Promote trending content	Prioritizes sensational misinformation
Engagement metrics	Likes, shares, comments	Rewards emotionally charged content
Automation tools	Bots and schedulers	Rapid spread across networks
Personalization	Tailored content feeds	Creates echo chambers

(Table 3. Platform mechanisms amplifying AI-generated misinformation)

Result Interpretation:
Easy spread on the internet does not mean information gets verified; errors enter unseen. Pattern-based systems show no preference between correct and false inputs, acting purely on structure instead of reasoning. Outputs blur accuracy when evaluation lacks depth within design. Mistakes increase slowly, strengthened by frequent reuse more than deliberate effort.

3.4. Regulatory and Policy Responses to AI-Generated Misinformation

The reviewed literature finds broad differences in how regions regulate AI-driven false information, according to the analyzed studies. Though detailed in Table 4, current systems are not uniform - each varies by reach, binding power, or method of execution.

Regionally, binding rules now shape how high-risk AI uses unfold across Europe. With the Artificial Intelligence Act, clear duties emerge - transparency takes form alongside system categorization by threat level. Oversight frameworks apply even when machines produce fabricated media (European Commission, 2023). Though considered thorough, experts observe uneven real-world application because of hurdles tied to technology and governance structures. Enforcement trails behind intent under such conditions.

Unlike stricter frameworks elsewhere, U.S. oversight tends to depend on advisory norms rather than enforceable rules. Guidance issued by the White House Office of Science and Technology Policy sets forth ideals like openness, equity, and responsibility - yet carries no statutory weight (White House OSTP, 2022). Because it cannot be legally enforced, some experts suggest this approach struggles to keep pace with shifting risks tied to artificial intelligence and false information.

Globally, UNESCO's guidance on artificial intelligence ethics sets out principles rooted in human rights to shape how AI is developed and applied. Still, because adherence is optional, implementation varies widely among nations (UNESCO, 2021). While some states introduce specific actions - especially concerning electoral integrity and digital offenses - these responses tend to lack cohesion and are uneven in practice (Romanishyn et al., 2025). Finding after finding shows rules reacting too late, failing to link across nations, which leaves oversight full of holes. Gaps remain because systems do not align beyond borders, response comes only after harm appears.

Region	Policy Framework	Focus Area	Limitations
European Union	Artificial Intelligence Rules and Online Platform Regulations	Transparency, risk classification	Enforcement complexity
United States	Artificial Intelligence Guidelines Not Legally Binding	Ethical principles	Lack of legal force
UNESCO	Global Guidelines for Ethical Use of Artificial Intelligence	Human rights-based AI	Voluntary adoption
National governments	Election and cyber laws	Deepfake control	Fragmented enforcement

(Table 4. Comparative overview of regulatory and policy responses)

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Result Interpretation:
Fresh hurdles usually push officials to respond once trouble hits, not ahead of time. Across nations, teamwork drags behind, built on separate moves instead of common goals.

3.5. Detection and Mitigation Approaches

The review literature research points to various methods designed to spot and slow down fake content made by artificial intelligence. As shown in Table 5, these efforts mix tools built on technology with rules set by authorities along with learning programs meant to inform people.

A well-known approach? Slapping clear labels on AI-made stuff. Research shows these markers help people notice fake content - less likely to spread it by mistake. Yet how well they work rides on steady rules and whether folks believe the system even matters (European Commission, 2023).

Into AI-made images and videos, hidden marks get tucked away like notes inside a bottle. These markers help track where files come from, studies show, yet many systems still fail to speak the same tagging language. One reason? Standards haven't caught up, even though the idea holds weight among experts lately. Hidden signals might guide us later, but right now they flicker at the edge of usefulness.

Some writings look into how artificial intelligence tools might catch fake text, pictures, sound, or videos. Though these systems help handle large amounts of material, research shows they often misjudge - accuracy shifts between tests. Because clever tweaks can fool them, depending only on such software brings problems down the line (Romanishyn et al., 2025).

Starting fresh each day, schools might teach people how to spot false information online instead of just blocking it. When folks learn to question what they see on screens, fake videos or misleading posts lose some power - yet changes show up slowly, like shadows moving across a wall. Studies back this idea, showing thoughtful thinking helps even when tricks get smarter.

Finding after finding shows one fix alone won't cut it. Instead, mixing tech tools with rules and teaching works better. Layered methods handle fake content from AI more reliably. Each piece supports the others. Strength comes from combining different types of responses.

Strategy	Description	Effectiveness
AI content labeling	Tags indicating AI-generated content	Moderate
Watermarking	Embedded digital signatures	Promising though uptake remains narrow
AI detection tools	Algorithms to identify synthetic media	Inconsistent accuracy
Media literacy	Public education initiatives	Long-term effectiveness

(Table 5. Detection and mitigation approaches for AI-generated misinformation)

Result Interpretation:
A single remedy does not suffice; layered approaches, blending technology with rules and learning, are more effective. A broader view establishes that ethical issues are present at every stage of developing AI-driven misinformation.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

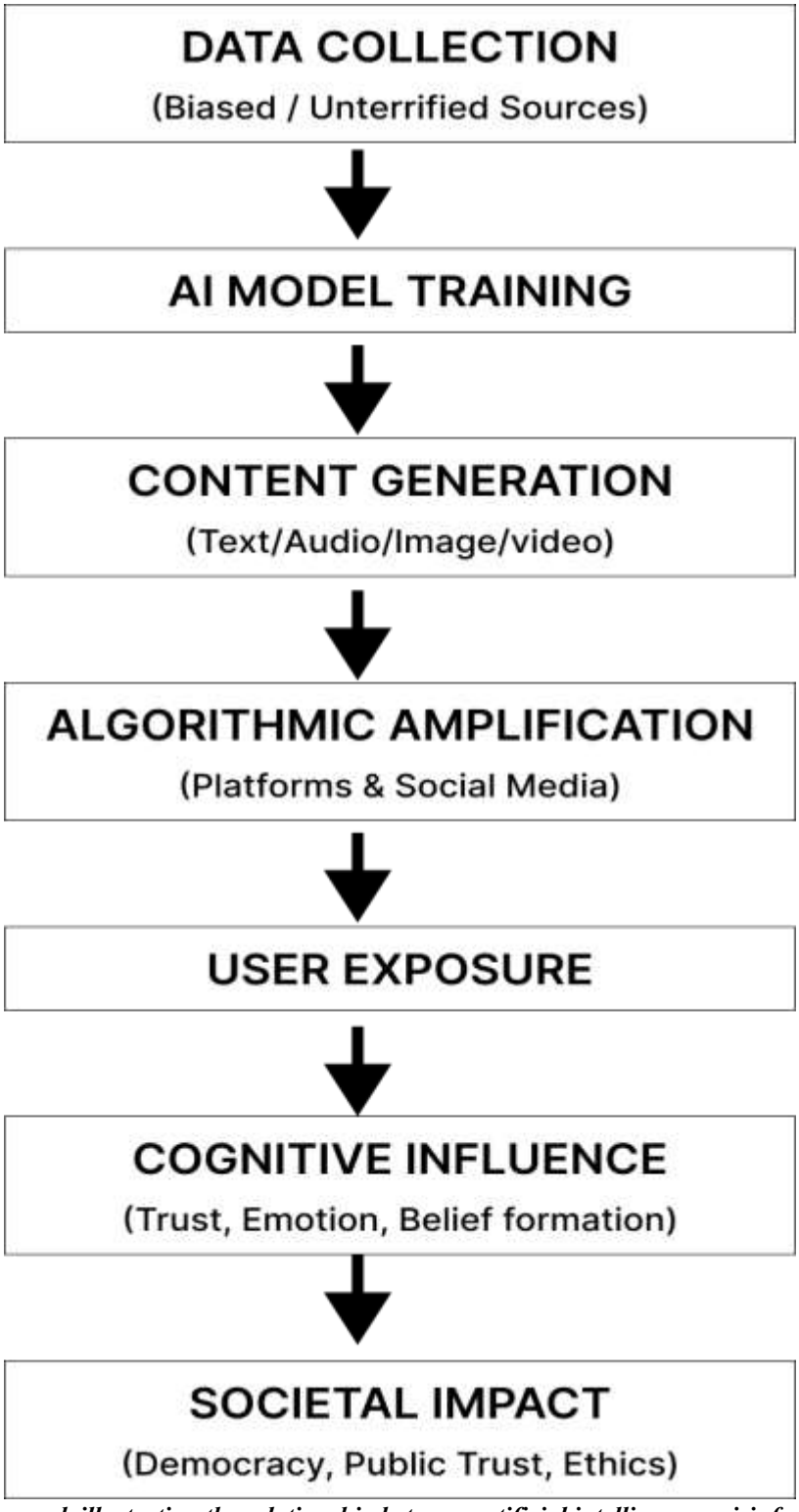

Figure 1. Framework illustrating the relationship between artificial intelligence, misinformation, and governance responses

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

3.6. Summary of Results

The findings establish that:

1) AI largely advances the scalability and realism of misinformation.
2) While the pace of technological development is dizzying, the ethics have yet to catch up.
3) Some oversight is visible in parts, but enforcement remains patchy across regions.
4) The design and algorithms of platforms massively amplify the problems.
5) Handling any problem well requires multi-layered oversight. It spreads accountability across different domains.

4. DISCUSSION

The findings of this review underline how AI is increasingly shaping the scale, speed, and trust of misinformation in digital settings. Across this literature, AI-enhanced creation is acknowledged as increasing the dynamics of the pre-existing nature of misinformation through high-speed content creation and highly realistic synthetic media. These features combine with algorithmically driven platform engagement to push the pace of misleading messages beyond what could be handled by traditional moderation.

Of course, the ethical resonance would be long-standing issues of transparency, accountability, fairness, and individual autonomy. Lack of appropriate disclosure of the AI-generated material limits the ability of users to question the credibility of such sources critically; this sits alongside earlier research that framed transparency as a primary condition for the ethical use of AI (UNESCO, 2021). Much like the diffusion of responsibility from developers and platforms to users complicates the enforcement of regulation and ethical responsibility—echoing some earlier concerns raised in governance literature (White House OSTP, 2022).

This is where the role of digital platforms in amplifying unethical governance comes in. Recommendation and personalization systems, optimized for user engagement, unknowingly favor sensational AI-generated content, intensifying the spread of misinformation. These findings align with earlier research demonstrating that emotionally charged content moves faster than factual information within online networks (Vosoughi et al., 2018). Hence, design choices on platforms begin to stand out as an important criterion for the ethical assessment of information ecosystems driven by AI.

This review now shows how all of this contrasts with what has been paid by policy, despite the growing amount of regulatory interest in this area. The European Union's Artificial Intelligence Act and UNESCO's ethical recommendations sketch out very important normative principles; however, they are de facto postponed by practical issues of implementation, jurisdictional fragmentation, and a lack of enforcement capacity. This echoes earlier comments that any governance of AI currently in force is mostly reactionary rather than preventative (Romanishyn et al., 2025).

Still, real-world difficulties reduce how well safeguards work. Even if tagging digital content helps visibility, uptake varies widely because agreed methods do not yet exist everywhere. Tools meant to spot harmful material at scale sometimes fail, especially when faced with deliberate tricks. While teaching people to judge online information matters greatly, progress depends on steady support over many years - results come slowly.

Findings point to layered solutions when handling false content made by artificial intelligence. Ethical frameworks, oversight by online platforms, unified regulations, together with informed citizens form part of the structure needed. Technical fixes alone fail. So do isolated laws. What emerges is reliance on adaptive systems of oversight growing in step with new tools. Stability in digital information depends on such evolving alignment. Trust follows only if responses shift just as quickly.

5. CONCLUSION

Information shaped by artificial intelligence now moves faster, spreads wider, appears more real - altering how facts are made and shared. Benefits exist in health, education, logistics; yet unintended uses raise deep concerns about trust and control. Examination shows mismatched efforts between moral standards, technology structures, government rules. Clarity often missing where decisions form. Responsibility fades when systems act without clear oversight. Uneven cooperation across borders leaves room for harm to grow unchecked.

Evidence shows current systems of control stay split apart, reacting mostly once damage appears. While efforts like the EU's AI legislation or UNESCO's moral standards mark steps forward, real impact weakens due to uneven application, poor international alignment, and fast-moving tech shifts. When platforms fine-tune software to boost user time, false information spreads easier - highlighting a growing role for company duty under stronger supervision.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

To counter misinformation created by artificial intelligence, collaboration across multiple levels becomes necessary. Built-in ethics within the architecture of these systems form a foundational step - regulation follows closely, needing flexibility, global consistency, and real-world impact. Alongside, responsibility among platforms grows critical, paired with long-term support for education that sharpens judgment about online material. When progress in technology moves hand in hand with values focused on people, confidence in what we see and share stays intact. What holds it together is alignment - not just between rules, but between intent and outcome.

Future Work

Work ahead might build on this analysis by examining how artificial intelligence spreads false information during events like voting periods, health emergencies, or disaster updates. Evidence gathered from checking whether platforms follow rules about tagging synthetic content may help shape future regulations. Instead of broad assumptions, insights could come from looking at how different areas respond to software that identifies fake material or educational efforts meant to improve judgment. Progress in these directions may lead to governance methods grounded in observation rather than reaction

REFERENCES

[1] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151.
[2] UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence.
[3] European Commission. (2023). Artificial Intelligence Act.
[4] White House Office of Science and Technology Policy. (2022). Blueprint for an AI Bill of Rights.
[5] Pennycook, G., et al. (2021). Shifting attention to accuracy can reduce misinformation online. Nature.
[6] Lu, X., & Hu, Y. (2025). Generative AI, misinformation, and human autonomy. Journal of AI Ethics.
[7] Romanishyn, Y., et al. (2025). AI and disinformation: governance challenges. Policy Studies Review.