# Sophia LUCID

A PEER REVIEWED
MULTI-DISCIPLINARY
RESEARCH JOURNAL

SOPHIA COLLEGE FOR WOMEN
Empowered Autonomous

PUBLISHED BY, SOPHIA COLLEGE FOR WOMEN (EMPOWERED AUTONOMOUS)
BHULABHAI DESAI ROAD
MUMBAI 400026
MAHARASHTRA

Contact us- editorlucid@sophiacollege.edu.in
www.sophiacollegemumbai.com

DESIGNED BY ANURAG PATIL

FEBRUARY 2025
ANNUAL VOLUME IV: ISSUE 1

# *Sophia Lucid*

*A Peer-Reviewed Multidisciplinary Research Journal*

***Current Volume***
***Annual Volume-IV***
***Issue-I (February, 2025)***
***Artificial Intelligence, Technology and Society***

**Publisher**
Sophia College for Women (Empowered Autonomous),
Bhulabhai Desai Road
Mumbai-400026, India

## GENERAL PUBLICATION DETAILS

**Starting year of publication:** The first issue was published in 2022.

**Frequency of publication:** It is a yearly publication and the annual issue will be published in February.

**Format:** This is a print journal.

**Author guidelines:** The style sheet and other requirements for publication will be mentioned in detail in the call for papers for each issue.

**Language:** English will be the primary language of publication. However, if you wish to include certain words or phrases from other languages, the editorial team will try to incorporate such linguistic variations.

**Submissions:** All submissions are to be emailed to editorlucid@sophiacollege.edu.in

**Aims and Scope:** *Sophia Lucid* aims to provide a platform for high-quality, original research that bridges diverse disciplines, fostering collaboration and knowledge exchange across scientific, technological, social, and humanistic domains. It seeks to publish innovative studies that address complex global challenges through interdisciplinary approaches, integrating perspectives from various fields to advance theoretical and practical insights.

The call for papers on each issue is based on a theme, crafted based on the pressing issues of the world we live in. The journal encourages discourse analysis, empirical studies, theoretical explorations, methodological advancements, interviews and case studies that provide cross-disciplinary insights on the focus area of each issue. *Sophia Lucid* welcomes original research articles, review papers, and perspective pieces that contribute to fostering a multidisciplinary dialogue in areas or contemporary issues from various disciplines including, but not limited to:

- Social Sciences and Humanities
- Health and Medical Sciences
- Business, Economics, and Management
- Law, Education and Learning Sciences
- Natural and Applied Sciences

A double-blind peer review process ensures the quality, originality, and significance of the published content. *Sophia Lucid* invites academic research papers from teachers and researchers across universities, colleges and research institutes.

**Past Issues**

**Vol I, Issue I, 2022: The Globe on the move: Avenues for the Future**
**Vol II, Issue I, 2023: Conflicts, Wars, Tolerance, Peace, and the Emerging World Order**
**Vol III, Issue I, 2024: Sustainability, Environment and Climate Change**

**CONTENTS**

**Editorial**

# Artificial Intelligence, Technology, and Society: Charting the Course of an Evolving Society

*Dr Elwin Susan John*

The robust advancement of artificial intelligence (AI) and its integration into various aspects of our everyday have brought forth new opportunities and challenges. From healthcare and education to governance and ethics, AI's influence is both transformative and disruptive. As we stand at the precipice of a new technological era, it is imperative to critically examine the interplay between AI, technology, and society to ensure responsible and equitable development. Keeping this at the centre, the Annual Volume IV, Issue I of our National Level, Peer Reviewed Multidisciplinary Research Journal *Sophia Lucid* explores the theme of "Artificial Intelligence, Technology and Society", and the role that academia can play in addressing these issues.

Artificial intelligence is transforming key aspects of our lives, in unimagined ways. The uses of AI percolates into making decisions that are more efficient and accurate. For example, it aids doctors to diagnose diseases with greater precision, ease traffic congestion in cities, personalize education, and detect fraud in financial transactions. While these advancements bring immense benefits, they also raise ethical and societal concerns that we cannot afford to ignore. The papers included in this volume address some of them.

A significant challenge offered by AI is the bias in AI systems. As AI learns from our personal data, it can reflect and even amplify existing biases, leading to unfair outcomes in hiring, policing, and credit decisions. To prevent this, we need greater transparency, accountability, and collaboration between policymakers and technologists. Similarly, concerns about privacy are growing, as AI relies on vast amounts of personal data. AI influences the way we interact and access information. Shoshana Zuboff, scholar and author of *The Age of Surveillance Capitalism*, critiques how AI-powered data collection is reshaping privacy, democracy, and power structures in her research. AI enhances the ability to track individuals through facial recognition, biometric data, browsing history, and geolocation, often without their explicit consent. This erosion of privacy raises concerns about autonomy, as individuals lose control over their personal information and are subjected to algorithmic profiling. Social media algorithms powered by AI shape public discourse, sometimes fuelling misinformation and division. As AI takes on a bigger role in moderating content and spreading information, it is important to ensure that it upholds democratic values rather than undermining them.

As discussed by Michel Foucault in *Discipline and Punish* regarding the present-day disciplinary systems, the concept of the 'panopticon' becomes even more relevant in the AI era. AI-powered surveillance creates an environment where individuals are constantly monitored, leading to self-censorship and behavioural changes due to the fear of being watched. This level of surveillance

challenges democratic ideals by restricting personal freedoms and reinforcing power asymmetries between the state, corporations, and the public.

Another pertinent aspect of AI is that it largely focusses on reshaping the job market. On one hand, it creates new opportunities and boosts productivity. While on the other hand, it threatens traditional jobs, especially in manufacturing, retail, and customer service. Automation has displaced many workers, raising fears about job security and economic inequality. Finding the right balance between innovation and individual rights is crucial to build public trust in AI.

In education, AI has opened new possibilities for personalized learning, intelligent tutoring, and administrative efficiency. Adaptive learning platforms tailor lessons to individual students, improving engagement and outcomes. AI-powered chatbots assist with questions, and automated grading frees up educators to focus on teaching. Rose Luckin, a professor of learner-centered design at UCL, argues that AI can act as an *intelligent tutor,* identifying gaps in student understanding and offering targeted interventions. In a recent interview, she opined, "the real power that AI brings to education is connecting our learning intelligently to make us smarter in the way we understand ourselves, the world and how we teach and learn. For the first time we will be able to extend, develop and measure the complexity of human intelligence; an intellect that is more sophisticated than any AI" (Luckin).

Navigating AI's impact on society requires input from multiple fields: not just Computer Science but also Law, Humanities, Social Sciences and Commerce. Scholars, educators, policymakers, and industry leaders must work together to shape AI's future in a way that is ethical, fair, and inclusive. Public engagement is equally important, ensuring that decisions about AI are made with the interests of society in mind. AI is more than just a tool. It is a powerful force shaping our world, and we must approach it with responsibility and thoughtful consideration.

The essays included in this issue collectively argue that by fostering interdisciplinary dialogue, enacting robust regulatory frameworks, and prioritizing human-centric AI development, we can harness the potential of AI, while safeguarding against its risks. The journey ahead is complex, but with a commitment to fairness, transparency, and inclusivity, we can shape an AI-driven future that aligns with societal values and aspirations.

Amit Hattiangadi's paper, "AI and Ethical Frontier: The Ethical Ramifications of Consciousness Uploading" explores the concept of consciousness uploading, the process of transferring human consciousness into a digital format. It examines fundamental issues related to individuality, personality, and moral responsibility. It also outlines moral principles to prevent exploitation and emphasizes the importance of a philosophical approach to the morality of this technology.

"Two Souls are Dwelling in My Breast": Mephistopheles in Our Pockets" by Deepak Kashyap examines the parallels between Artificial Intelligence (AI) and Mephistopheles from Goethe's *Faust*, framing AI as a modern Faustian bargain where technological progress comes at the cost

of moral values. According to the paper, like Mephistopheles, AI influences human decisions, often serving corporate interests while simultaneously enhancing human capabilities. Through a comparative literary, philosophical, and theological analysis, the study explores the dual nature of AI; both as a tool for progress and a potential threat, while addressing broader ethical concerns about knowledge and responsibility in modern society.

In the paper titled, "Zombie-Auras: GenAI and Hybrid Text Production", Joshua Nieubuurt examines the evolution of textual production and its impact on Walter Benjamin's concept of "aura." It identifies three key phases; natural, mechanized, and digitized each diminishing the "cult value" of texts by making creation and reproduction easier. The rise of Generative AI (GenAI) represents the latest shift, turning the "aura" into a memetic "zombie" form-recognizable yet stripped of originality. Through an analysis of *The Epic of Gilgamesh* and Shakespeare's works, the study explores how GenAI alters textual authenticity. It introduces the concepts of "hybrid auras" and "zombie auras" to describe AI-co-authored texts, emphasizing the need to preserve human artistic essence in the digital age.

Shadab Zaveri's paper "Exploring the Increasing Role of Digital Culture in the Creation of Literary Genres: A Study of the Dark Academia Movement" explores the influence of digital culture on literary genres, focusing on the rise of the *Dark Academia* movement. Digital culture, driven by social media and online communities, has reshaped human interaction, creativity, and cultural consumption, fostering new subcultures and aesthetic trends. The study examines digital platforms like YouTube, Instagram, Tumblr, and TikTok as texts to analyze how they contribute to *Dark Academia*'s establishment as both a cultural movement and an emerging literary genre.

In the paper, "Constructing Utopias: Edward Soja's Thirdspace and the Transhumanist Vision in Ray Kurzweil's *The Singularity is Near*", Vidya Hariharan examines the intersection of Edward Soja's Thirdspace theory and Ray Kurzweil's vision of the *Singularity*. Soja redefines space as social, psychological, and imagined, while Kurzweil envisions a future where technology merges with biology, reshaping identity and space. This technological acceleration aligns with Soja's idea of reimagined spaces, suggesting a utopian society shaped by AI. By applying Thirdspace to Kurzweil's utopian society, the study examines how technological progress redefines human potential and spatial practices. It also addresses the moral and ethical implications of this transhuman future.

Chitra Adkar's "Digital Manipulation as Text: The Problem of Disinformation in Digital Content" explores the role of AI in digital manipulation and its impact on internet content. It examines the fine line between socially acceptable edits such as AI-generated humorous video mashups and harmful misinformation or disinformation. By analyzing AI-powered editing tools, the study highlights how digital content relies on manipulation while questioning its detectability and ethical implications in the age of artificial intelligence.

In the joint paper, "Well, That Escalated Quickly: The Role of Memes in Political Engagement" Lavanya Mohan and Malvika Dixit study the role of AI in shaping meme culture and its impact on political engagement, particularly during the 2024 Indian Lok Sabha Elections. With AI-driven tools facilitating meme creation and dissemination, memes have become powerful vehicles for political discourse. Using qualitative and quantitative analysis, the study finds a significant link between meme engagement and political participation among individuals aged 20-35. AI-enhanced meme culture streamlines political discussions, influencing both online and offline activism, contributing to the broader discourse on digital political communication.

Sukanya Menon's paper, "Rules of Engagement: LAWs, Ethics and the Changing Nature of Warfare in the AI Era" identifies AI's transformative role in modern warfare, highlighting its ethical and geopolitical implications. The rapid advancement of AI-driven technologies, including lethal autonomous weapons (LAWs) and cyber tools, has redefined warfare strategies while raising critical concerns about regulation and accountability. Through a systematic review, this study explores how AI is reshaping military tactics and international relations, emphasizing the urgent need for ethical considerations and policy frameworks.

In the paper, "The Role of Artificial Intelligence on Exchange Traded Fund Investments", Midhun Johny George discusses the integration of AI into Exchange Traded Funds (ETFs) and its transformative impact on the financial sector. AI-driven tools analyze market trends, predict price movements, and optimize investment portfolios, enabling personalized and data-driven strategies. The paper explores services like the Robo-advisors that can offer algorithm-driven investment services without human intervention.

Vidya Konar conducted a study on cybercrimes and the resulting paper titled, "Study of Cyber Crimes in Selected Metropolitan Cities for the period 2020 – 2022" examines cybercrime trends in five major Indian cities using NCRB data from 2020-2022, with a focus on AI's role in cybercrime detection and prevention. AI-driven tools help analyze large datasets, identify patterns, and enhance cybersecurity measures against fraud, identity theft, and digital disruptions. The study employed statistical methods like ANOVA and mean analysis to evaluate cybercrime rates under various legal frameworks. By leveraging AI, the research highlights how machine learning and predictive analytics can improve law enforcement responses and strengthen digital security.

**Works Cited**

Luckin, Rose. "Opinion: AI in Education will Help us Understand How We Think". UCL News, 2020, https://www.ucl.ac.uk/news/2020/mar/opinion-ai-education-will-help-us-understand-how-we-think.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Public Affairs, 2019

# AI AND ETHICAL FRONTIER: THE ETHICAL RAMIFICATIONS OF CONSCIOUSNESS UPLOADING

*Amit Hattiangadi*
*Lecturer, Department of Philosophy*
*Sophia College for Women (Empowered Autonomous), Mumbai*

**Abstract:** Consciousness uploading refers to the idea of transferring the consciousness to a digital format. This idea has attracted considerable attention recently because of advances in brain research and computer science. Employing conceptual analysis, this paper explores the possibility of transferring man's consciousness to a digital format; casting light on all the basic problems concerning one's individuality, personality and moral responsibility. Previous studies have discussed ethical issues arising when one transfers an individual's memory into another person's head—issues like privacy rights and personal identity. The various studies on this theme are largely devoted to ethical issues concerning what information should be stored in the new memory and whether it would preserve individuality. This paper deals with the ethical implications of consciousness uploading, its societal impact and it outlines moral principles that must be followed to prevent the exploitation of the uploaded consciousness. Thus, the importance of considering morality in consciousness uploading will be exemplified, through a careful philosophical examination.

**Keywords:** consciousness, ethics, AI, individuality

## Introduction

The concept of consciousness uploading also referred to as mind upload or whole brain emulation, has continued to attract controversies and discussions in the fields of philosophy, technology, and ethics. This theoretical technology raises serious moral and ethical issues that need to be pondered on. The idea of 'uploading of consciousness' came into being as a possible means of increasing the longevity of mankind. However, this has birthed various questions around the morality and ethicality of the practice. Earlier research has handled ethical issues such as, privacy rights and personal identity that arise during the process of inserting individual memory in an entity's brain. A crucial ethical consideration when contemplating the possibility of consciousness uploading is determining the moral status of conscious entities that exist without a physical body or form (henceforth referred to as disembodied consciousness).

As Artificial Intelligence (AI) becomes increasingly sophisticated, the criticality of ethical issues around it becomes even more prominent. While there is no denying the positive impact and potential of AI in certain fields such as healthcare, education and development; the idea of AI systems managing to comprehend and play around with the emotions and cognitive level of human beings is an extreme proposition that should be dealt with ethically and proactively.

This study seeks to explore and address the ethical aspects of this possible technological enhancement with specific emphasis on the self, privacy, identity, social issues, ethical issues, and the restrictions that have to be imposed by the state for peace purposes, which may include curbing uploaded consciousness abuses.

**Scope of Consciousness**

Our understanding of self-awareness has been based on deepening our awareness about both the mind and body functioning. Our consciousness, as we understand it, is informed by said understanding of self-awareness. When discussing consciousness uploading, whose key distinguisher is a disembodied consciousness, it raises questions such as – 'Will such a consciousness have a sense of self-awareness?'

The self-awareness of consciousness is a very complex phenomenon which has been written, theorized and debated on in various sociological areas (Butlin et al., 2023). The philosophy of mind has long tried to account for subjective experience and has become increasingly enriched by the learnings of neuroscience and cognitive science.

The very basis of the concept of consciousness uploading lies in the fact that all the attributes of human being including his or her memory, viewpoints, or dispositions can be captured and subsequently converted into a digital format. This however, leads to certain issues such as determining the nature of self, personal identity and the relationship with society.

Some philosophers believe that consciousness cannot be separated from the physical form of the brain and nervous system. This view suggests that the transfer of consciousness to a digital substrate would fundamentally alter the nature of the self, potentially leading to a loss of personal identity. Alternatively, proponents of functionalist theories argue that consciousness can be reproduced in artificial systems, provided that the appropriate computational processes are replicated.

The core of the problem is one of personal identity with respect to the consciousness (Pandey, 2018). If an individual's consciousness could be digitized and transplanted into a non-biological platform, would the resulting entity be truly "the same" as the original person? (Mudrik et al., 2023). This question touches upon deep-seated beliefs about the self, the soul, and the underpinnings of human existence. Others contend that consciousness is irrevocably tied to the physical substrate of the brain and that to separate it from its biological roots would essentially change the essence of the person. Some say that retention of cognitive ability and subjective experience is the essence of personhood, no matter what the substrate (Berne, 2003).

Though computational theories of mind have come a long way in explaining cognition, the subjective, phenomenological aspects of consciousness are still a mystery. As researchers continue to grapple with the relationship between physical processes in the brain and the

emergent property of subjective experience, the feasibility of accurately replicating consciousness in a digital format remains uncertain (Johnstone & Alegaonkar, 2023).

**Ethical Implications of Consciousness Uploading**

One of the main moral issues with consciousness uploading would be the effect it would have on the idea of self and personal identity. If an individual's consciousness is successfully uploaded, it raises questions about the continuity of their identity and the extent to which the digital version can be considered the same person.

**The Concept of Self and Personal Identity**

The idea of consciousness uploading is further complicated by the potential for multiple copies of the same consciousness, leading to multifarious questions about the rights and responsibilities of each iteration. With cyborg technology growing closer and closer to the human body and mind, the distinction between biological and technological not only blurs, but shatters the classical idea of what is human.

> If a neuroprosthetic device is accessed by a third party and done to edit one's memory or to plant a new memory in one's mind, or even to place an ad for a commercial product in one's consciousness, should there be a law of cognitive liberty or of "neuro-advertising" that applies? (Fiske et al., 2019)

Such intrusions into the mind's privacy and autonomy could have profound implications for an individual's sense of self and personal identity (Barfield Williams, 2017). When an individual's consciousness is uploaded to a digital substrate, the resulting entity may possess the same memories, beliefs, and personality traits as the original, but it remains unclear whether it can be considered the same person. The philosophical discussions of personal identity and the metaphysics of "self" begin to take on more meaning in this situation because the idea of a consciousness upload undermines the conventional concept of what it is to be human.

> Only with a clear image of how a certain enhancement method might alter cognitive processes in specific populations, along with side effects and costs, can make it justifiable. (Cinel et al., 2019)

In other words, the ethical implications of consciousness uploading cannot be fully understood without a comprehensive understanding of how the process may impact an individual's sense of self and personal identity.

Closely related to the question of personal identity is the issue of moral responsibility and the ethical implications of consciousness uploading. If an individual's consciousness could be replicated or transferred, what would this mean for concepts of criminal culpability, personal agency, and the continuity of the self over time? Also, the idea of "backing up" one's

consciousness would lead to some serious moral questions about the nature of these digital copies and how they could/would be exploited. (Jangwan et al., 2022).

**Privacy Rights, Cognitive Liberty and Autonomy**

The integration of brain-computer interfaces, neural implants and consciousness uploading raises concerns on both – policy and legislation level as well as on an individual, personal privacy level. For instance, what type of legislation and policy should come about when cyborg apparatus that can be implanted in the body leads to the development of a new class of individuals that have greater physical and computing capacity than the rest of the population? Additionally, what fundamental human and legal rights should be given to those with cyborg technology when they become less biology and more machine?

On a personal level, what is the extent of the breach of privacy and possibility of mental manipulation as mental functioning becomes increasingly integrated with technological machines? To what extent would an individual's thoughts, memories and decision-making processes become accessible to others?

This in turn raises questions around one's cognitive liberty - the right to have autonomy over one's own thought processes. In the same way that there is a hacker culture with regard to computer software, there is also an emerging culture of "hacking brain function", or the exploration of ways to transcend the natural boundaries of human cognitive ability. This has brought on both hope and fear; while socialists and scientists agree that enhancement is possible, they disagree on how plausible it is, what use it will be, what risks will be involved, and what effects it will have in the long run.

> The ability to access and potentially modify an individual's memories, beliefs, and decision-making processes through technological means raises profound ethical questions. Societies would have to deal with everything from the rights and standing of uploaded consciousness's (Bostrom & Sandberg, 2009).

While the impacts of uploaded consciousness are varied, possibly the most profound area for ethical consideration remains the impact it has on personal privacy and autonomy. If an individual's consciousness is uploaded, there are concerns about the level of control and access that others, such as governments or corporations, may have over the digital representation. That would be an invasion of privacy, and opens up the possibility of the uploaded consciousness being exploited or manipulated.

Lavazza and Inglese (2018) speak of a situation that exemplifies the same - "The most basic, albeit still not really feasible ethical issue in this field, would be that in the process of solving crimes, law enforcement could utilize some sort of mind reading neurotechnology to infer the mental

states of suspects and thus aid in their investigation. These kinds of cognitive incursions could have serious consequences for personal autonomy and the right to privacy of the mind."

Should third parties gain access to our private thoughts, what rules would dictate how they can make use of this information? The moment this access is granted, what's private becomes public – private desires, thoughts and beliefs all are presented in the public domain, subject to public speculation and judgements. This hypothetical instance highlights the significant ethical concerns that are raised when considering the possibility that unauthorized third parties will gain control over one's uploaded consciousness to access, analyse, exploit or manipulate it.

**Societal Impact and Ethical Principles**

The widespread adoption of consciousness uploading may lead to seismic shifts in society. This technology, if available to the public, would change socioeconomic relations and will drastically alter cultural facets and norms of society that exacerbate pre-existing inequalities and potential forms of discrimination.

Furthermore, it would lead to other questions such as – 'what happens to society should *every* person upload their consciousness?' Fears have been expressed that, as technology develops, biological humans might be progressively supplanted by digital ones, raising questions about how human identity would be preserved and how society might unite or be torn apart.

To ease these concerns, it is important to develop a robust ethical framework, guidelines and a set of moral principles that will ensure that the well-being, autonomy, and dignity of the individuals concerned will be protected. An international consortium of interdisciplinary scholars and practitioners has expressed that the development and implementation of neurotechnologies must be informed and directed by democratic and inclusive processes in order to protect the interests of all relevant stakeholders (Ligthart et al., 2023).

**The Philosophical Challenges**

In as much as several other technologies do as well, consciousness uploading raises philosophical and ethical issues that necessitate in-depth and multi-faceted research and dialogue. While it may be theoretically possible to develop techniques for consciousness uploading, ethical aspects of the technology are indeed something difficult.

One obvious concern that presents itself is – what if consciousness uploading technology became widespread, it would most likely lead to a further division between the haves and have nots (Barfield Williams, 2017). This could make current inequality worse and create an "enhanced" class of people with considerable cognitive and physical advantages, which would lead to many problems of fairness, social justice, and possibly exploitation.

Another important concern is the consciousness's inability to lose the non-virtual, real-world embodiment. As Schweizer (2018) explains, the "mind-machine merger" conceptual scenario where human body and brain are eliminated poses significant problems. The subjective sense of 'self', social affiliation, and the very value of being human and having a physical body may in some way be transformed or altogether eradicated along the routine operation of 'emigrating' to the digital.

If someone is copied into a machine, is that thing really that person? Sartre famously said "Existence precedes Essence"; the primary concern raised is that even if we manage to upload consciousness, mimicking the person's memory and cognitions, can we truly say that the personality too has been uploaded, if the *person* beneath it all ceases to exist? The philosophical debate continues as to whether or not personal identity and consciousness can be transferred across such a dramatic change in physical form.

While those in favour of consciousness uploading may argue that the retention of cognitive function and 'stream of consciousness' is what being a person is all about, which can be achieved by transferring of consciousness into a digital medium; those against the idea would counter that to upload the person's consciousness into a computer is to dehumanize them – make them nothing more than a machine. While the idea of being able to back up and potentially extend one's lifespan through digital means, obtaining a sort of "digital immortalization" may be an appealing one, the fear that this process of digitization will lead to significant fragmentation or displacement of one's identity remains a very real one. The fear is that the consciousness uploading will eventually phase out biological humans for their digital counterparts and that this process will in some way erode the very nature of humanity and society itself. Additionally, the concerns about the person's digital copies being used and abused, which could result in exploitation, identity theft, or the original person being replaced. (Lee et al., 2023)

**Government Policies to Mitigate Ethical Risks**

As AI continues to encroach upon domains traditionally reserved for human intelligence and emotional capacity, it is crucial that developers, policymakers, and the public grapple with the profound ethical quandaries raised by the prospect of conscious, emotionally intelligent machines. Governments are already developing AI strategy, policy and ethics papers, but it is also essential to consider the public's opinion about the ethics of AI. Consciousness uploading is a complex phenomenon with far reaching consequences. While AI and uploaded consciousness present uses or potential uses in various industries, most of these uses are still in their infancy stage. The current policies, mandates and strategies designed for AI are not sufficient for covering the far reaching ethical, moral and philosophical impacts of consciousness uploading. Thus, it is imperative, that at the germinal stage of consciousness uploading, we also simultaneously mould the moral and ethical standards that will govern this future technology.

Policymakers, researchers and the general public must all come together to thoughtfully deliberate the advantages and perils of these technologies, and to see that they are created and utilized in such a manner that upholds human moral standards and the welfare of all. This cooperation will be key in the creation of a complete moral code that will allow for the growth and use of affectively-aware AI and uploaded consciousness in a manner that respects humanity and contributes to the betterment of society as a whole.

Policies, moral principles and the ethical code of conduct developed must include, but not be limited to, the protection of mental privacy, the preservation of personal identity, the promotion of equitable access to such technologies, and the prevention of the exploitation of uploaded consciousness for commercial or nefarious purposes. The power of these technologies to change society must be approached with caution and with full consideration of the potential unintended consequences; the dangers of this kind of technology being misused cannot be overlooked. The ability to force, control, or wipe out individual minds with digital manipulation is an Orwellian nightmare that needs to be thought about very seriously. Ensuring that these ethical principles are in place, reduces the probability of said Orwellian nightmare from becoming a dystopian reality.

Other areas of consideration in policy making around uploaded consciousness may include – beginning by smoothing the existing challenges like equity and privacy risks, regulating the thick legal industry around digital death management and related commercialization and regulations on the use of particularly potent technologies for uploading consciousness and related medical devices.

True and effective management and implication can only be made possible with public awareness and participation. Efforts towards raising public awareness and understanding of the ethicality of such complex technologies provides an important tool to keep check of the nature of development of said technologies, and has an immense social and philosophical impact. For example, it will allow us to monitor and curb the commercialization of uploaded consciousness that can lead to people being prostituted and exploited with their digital selves becoming a sellable commodity.

**Synthesis and Conclusions**

The groundwork conducted for this article has offered, within itself many aspects that directly involve the ethical concerns on the topic of consciousness uploading and related technology. The notion of 'uploading' of consciousness raises numerous questions regarding the fundamental self and the concept of personal identity. The idea of outsiders having access to one's augmented consciousness without consent and employing it in ways that lie beyond their operational scope raises strong ethical questions on multiple fronts.

The societal cue that might be sounded out of this shift should be that if this situation arises, this adoption of industrial practice would probably only be accessible to the wealthy, and a new line within the category of enhanced and normal subjects will emerge focused on equality and social justice. In order to address these dilemmas, it is necessary that adequate moral principles and rules be designed and enforced with respect to development and usage of technology for consciousness 'uploading'. These principles should include, but not be limited to, undertaking to safeguard mental privacy, enhancing personal identity, providing equal opportunities to all such technology users, and stopping commercialisation of uploaded consciousness.

There are many facets of workplace ethics with regard to uploading consciousness and they all tend to centre on those aspects of the self as self-hood and identity, the right to privacy, the social purpose. Since the scientific and technological progress in this direction is incessant, it is necessary that such discourse remains focused on these ethical issues and enables setting adequate ethical frameworks as part of the process of developing and implementing this technology. If we proceed in an ethical manner in relation to technology of consciousness uploading, it will be possible to incorporate this technology without infringement of basic understanding of human rights and without harm to the society.

**Bibliography**


Askenasy, M., J., J., and Joseph Lehmann. "Consciousness, brain, neuroplasticity." Frontiers Media, vol. 4, 1 Jan. 2013, https://doi.org/10.3389/fpsyg.2013.00412. Accessed 21 Jul. 2024.

Barfield, Woodrow, and Alexander Williams. "Law, Cyborgs, and Technologically Enhanced Brains." Multidisciplinary Digital Publishing Institute, vol. 2, no. 1, 17 Feb. 2017, p. 6-6. https://doi.org/10.3390/philosophies2010006. Accessed 12 Jul. 2024.

Bostrom, Nick, and Anders Sandberg. "Cognitive Enhancement: Methods, Ethics, Regulatory Challenges." Springer Science+Business Media, vol. 15, no. 3, 19 Jun. 2009, p. 311-341. https://doi.org/10.1007/s11948-009-9142-5. Accessed 28 Jul. 2024.

Bostrom, Nick, and Eliezer Yudkowsky. "The ethics of artificial intelligence." Cambridge University Press, 12 Jun. 2014, p. 316-334. https://doi.org/10.1017/cbo9781139046855.020.

Butlin, Patrick, et al. "Consciousness in Artificial Intelligence: Insights from the Science of Consciousness." Cornell University, 1 Jan. 2023, https://doi.org/10.48550/arxiv.2308.08708. Accessed 28 Jul. 2024.

Cinel, Caterina, et al. "Neurotechnologies for Human Cognitive Augmentation: Current State of the Art and Future Prospects." Frontiers Media, vol. 13, 31 Jan. 2019, https://doi.org/10.3389/fnhum.2019.00013. Accessed 14 Jul. 2024.

Coeckelbergh, Mark. "Pervasion of what? Techno–human ecologies and their ubiquitous spirits." Springer Nature, vol. 28, no. 1, 20 Apr. 2012, p. 55-63. https://doi.org/10.1007/s00146-012-0418-y. Accessed 21 Jul. 2024.

"Corrigendum: Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements." SAGE Publishing, vol. 20, no. 3, 15 Nov. 2019, p. 165-166. https://doi.org/10.1177/1529100619889954. Accessed 28 Jul. 2024.

Dehaene, Stanislas, et al. "What is consciousness, and could machines have it?" American Association for the Advancement of Science, vol. 358, no. 6362, 27 Oct. 2017, p. 486-492. https://doi.org/10.1126/science.aan8871. Accessed 21 Jul. 2024.

Dehaene, Stanislas, et al. "What is consciousness, and could machines have it?." American Association for the Advancement of Science, vol. 358, no. 6362, 27 Oct. 2017, p. 486-492. https://doi.org/10.1126/science.aan8871. Accessed 14 Jul. 2024.

Esmaeilzadeh, Hadi, and Reza Vaezi. "Conscious AI." Cornell University, 1 Jan. 2021, https://doi.org/10.48550/arxiv.2105.07879. Accessed 28 Jul. 2024.

Fiske, Amelia. "Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy." JMIR Publications, vol. 21, no. 5, 9 May. 2019, p. e13216-e13216. https://doi.org/10.2196/13216. Accessed 12 Jul. 2024.

Gaudry, S., Kate, et al. "Projections and the Potential Societal Impact of the Future of Neurotechnologies." Frontiers Media, vol. 15, 15 Nov. 2021, https://doi.org/10.3389/fnins.2021.658930. Accessed 28 Jul. 2024.

Goering, Sara, et al. "Recommendations for Responsible Development and Application of Neurotechnologies." Springer Science+Business Media, vol. 14, no. 3, 29 Apr. 2021, p. 365-386. https://doi.org/10.1007/s12152-021-09468-6. Accessed 21 Jul. 2024.

Jangwan, Singh, Nitish, et al. "Brain augmentation and neuroscience technologies: current applications, challenges, ethics and future prospects." Frontiers Media, vol. 16, 23 Sep. 2022, https://doi.org/10.3389/fnsys.2022.1000495. Accessed 14 Jul. 2024.

Johnstone, Charles, and Prashant S. Alegaonkar. "Understanding Physical Processes in Describing a State of Consciousness: A Review." Cornell University, 1 Jan. 2023, https://doi.org/10.48550/arxiv.2301.09576. Accessed 28 Jul. 2024.

Lai, T, Vicky, and Lera Boroditsky. "Consciousness, brain, neuroplasticity." Frontiers Media, vol. 4, 1 Jan. 2013, https://doi.org/10.3389/fpsyg.2013.00142. Accessed 21 Jul. 2024.

Lee, Kang, Yung, Patrick, et al. "Speculating on Risks of AI Clones to Selfhood and Relationships: Doppelganger-phobia, Identity Fragmentation, and Living Memories." Association for Computing Machinery, vol. 7, no. CSCW1, 14 Apr. 2023, p. 1-28. https://doi.org/10.1145/3579524. Accessed 14 Jul. 2024.

Lee, P Y K., Ning, F., Kim, I., & Yoon, D. (2023, April 14). Speculating on Risks of AI Clones to Selfhood and Relationships: Doppelganger-phobia, Identity Fragmentation, and Living Memories. Association for Computing Machinery, 7(CSCW1), 1-28. https://doi.org/10.1145/3579524. Accessed 21 Jul. 2024.

Ligthart, Sjors, et al. "Minding rights: Mapping ethical and legal foundations of 'neurorights'." Cornell University, 1 Jan. 2023, https://doi.org/10.48550/arxiv.2302.06281. Accessed 28 Jul. 2024.

Mudrik, L., Mylopoulos, M., Negro, N., & Schurger, A. (2023, October 1). Theories of consciousness and a life worth living. Elsevier BV, 53, 101299-101299. https://doi.org/10.1016/j.cobeha.2023.101299. Accessed 21 Jul. 2024.

Mudrik, Liad, et al. "Theories of consciousness and a life worth living." Elsevier BV, vol. 53, 1 Oct. 2023, p. 101299-101299. https://doi.org/10.1016/j.cobeha.2023.101299. Accessed 21 Jul. 2024.

Ong, C., Desmond. "An Ethical Framework for Guiding the Development of Affectively-Aware Artificial Intelligence." Cornell University, 1 Jan. 2021, https://doi.org/10.48550/arxiv.2107.13734. Accessed 28 Jul. 2024.

Pandey, C., Subhash. "Can artificially intelligent agents really be conscious?." Springer Science+Business Media, vol. 43, no. 7, 22 Jun. 2018, https://doi.org/10.1007/s12046-018-0887-x. Accessed 28 Jul. 2024.

Schweizer, Paul. "Artificial Brains and Hybrid Minds." Springer Nature, 1 Jan. 2018, p. 81-91. https://doi.org/10.1007/978-3-319-96448-5_10. Accessed 28 Jul. 2024.

Shu, Sara, and Benjamin K.P. Woo. "Pioneering the Metaverse: The Role of the Metaverse in an Aging Population." , vol. 6, 20 Jan. 2023, p. e40582-e40582. https://doi.org/10.2196/40582. Accessed 21 Jul. 2024.

Warwick, K. (2014, November 4). The Cyborg Revolution. Springer Science+Business Media, 8(3), 263-273. https://doi.org/10.1007/s11569-014-0212-z. Accessed 28 Jul. 2024.

Xie, Jianhua. "An explanation of the relationship between artificial intelligence and human beings from the perspective of consciousness." SAGE Publishing, vol. 4, no. 3, 1 Sep. 2021, p. 124-134. https://doi.org/10.1177/20966083211056376. Accessed 21 Jul. 2024.

# "TWO SOULS ARE DWELLING IN MY BREAST": MEPHISTOPHELES IN OUR POCKETS

*Deepak Kashyap*
*Ph.D. Research Scholar, Center of German Studies,*
*Jawaharlal Nehru University, New Delhi*

**Abstract:** A Faustian bargain has taken place. In the modern world, the pursuit of knowledge in the form of technological advancements is achieved by keeping moral values at stake. Similar to Mephistopheles, the Devil from Goethe's Faust, artificial intelligence (AI) continuously monitors an individual's – the sophisticated Fausts of 21st century - every move, both online and offline, guiding their decisions to maximize the profits of the big capitalist enterprises. Although AI has been likened to Shelley's Frankenstein's monster, exploring the sentience of intelligent machines, the devilish aspects of the modern-day search for intelligence are believed to be less invasive. The paper seeks to explore the double personas of Artificial Intelligence (AI) and Mephistopheles in Goethe's Faust through a comparative literary, philosophical, and theological analysis of the Devil and its necessity in the modern society. By doing so, the paper highlights the dualism embedded in the nature of AI and Mephistopheles, where on the one hand they enhance human capabilities and redefine the modern workplace, on the other they endanger human agency and threaten an imminent labor displacement. This comparative analysis draws attention to the nuanced interactions between humans and technology while considering the more general philosophical implications of progress, moral responsibility, and the pursuit of knowledge.

**Keywords:** AI, Faust, Devil, Goethe, Technology.

*The Devil holds the strings which move us!*
*In repugnant things we discover charms;*
*Every day we descend a step further toward Hell,*
*Without horror, through gloom that stinks*
~ Charles Baudelaire, To the Reader, in *Le Fleur du Mal*

## Introduction

In the western literary tradition, there could be observed two dominant corpora in which the themes of sinister creation and their consequences are dealt. In the novels such as Mary Shelley's *Frankenstein; or, The Modern Prometheus* (1818) and Gustav Meyrink's *The Golem* (1914), it is the modern, striving human which bestows a new life upon the inanimate objects of their creation. Whereas it is the enlightened scientist in the former to bring life to Frankenstein's monster, in the latter it is a religious rabbi who incarnates a gigantic beast to protect the Jews. Both gain sentience and rebel. On the other hand, in epic poems such as Dante's *Divine Comedy* (1321) and John Milton's *Paradise Lost* (1667), it is the divine God who has created the Devil, that fallen angel, which was initially a good, sincere follower. In Dante, Lucifer has been condemned to remain

trapped at the center of Hell because of his rebellion. In Milton, Satan is overfilled with pride and conspires to overthrow God. Both denote the fall.

It is in the interstices of these two rather entwined strands that Johann Wolfgang von Goethe's play *Faust. A Tragedy* (1806, 1831) situates itself. The conscious decision to consider Goethe's *Faust* to explore the nefarious aspects of digital technology is driven by its simultaneous entrapment in and departure from these two conventional literary traditions on sinister creations. It is a story of our kin, a striving scholar, Dr. Heinrich Faust, who has aspired to achieve the Truth, to no avail. The unaccomplished scholar is on one night visited by a devil figure, Mephistopheles, who has wagered with God that he will lead Faust astray from the path of knowledge. The two earthly creatures make the popular Faustian pact, the terms of which are as follows: should Faust encounter such a moment which leads him to say "Tarry, remain! – you are so fair!"[1], then Mephisto can take away his soul. Today, our relationship with the smartphones is not in any way dissimilar. Like Mephisto, digital technology enables us to indulge in and realize our base desires and in exchange gathers our soul in form of data. The humans of today have indeed become cyborgs, as Haraway had prophesized.[2]

Even if this Faustian exchange may have occurred in some way or the other with each upgrade in the modern technology. The latest specter in its realm, namely, AI, builds on and simulates the human intelligence, the Cartesian *res cogitans*, and hence purports to scandalize the very qualia of human metaphysics. AI is abundant in its applications throughout the terrains of the digital world: in search engines, algorithmic recommendation, machine learning, language training, and generative AI, for example. However, in this discussion, all these would be conflated in a general sense and we would rely on European Commission's High-Level Expert Group's definition, which says "Artificial Intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-determined parameters) to achieve the given goal."[3] Nonetheless, the different signifiers may be uttered interchangeably albeit the signified should be understood as one and the same.

---

[1] Goethe, J.W. von, Edited and Translated by Atkins, Stuart. *Faust I & II.* 1984. Princeton and Oxford. Princeton University Press. Line 1700

[2] Haraway, Donna. "A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century."In: *The transgender studies reader.* 2013. Routledge. pp. 103-118

[3] "A Definition of AI: Main Capabilities and Scientific Disciplines". *The European Comission's High-Level Expert Group on Artificial Intelligence.* 2018. Brussels. European Commission. (Link:

The central argument of this paper is that AI is a Janus-faced pandora's box which the modern human demigod has opened in anticipation of adventurous gifts, and what she has gotten in addition is also a plethora of causes for migraine. Interestingly however, this notorious, devilish creature is so embedded in her everyday life that she decidedly cannot do away with it. The paper is divided into three sections: The Nature of (D)evil, The Nature of Mephistopheles and AI, The Nature of Human(s). In the first section, a brief theological, philosophical, and cultural contextualization of the concept of evil and devil will be provided. The second section will juxtapose and gauge in how far are the figures of Mephistopheles and AI to be considered similar. The third section will affirm the premise of the second and then deliberate on the metaphysical concerns that arise from such a conflation.

**The Nature of the (D)evil**

The figure of Devil is ever-present in all the folktales, scriptures, mythologies, as well as literatures transcending all cultural, geographical, and imaginary boundaries. Hence, the treasure trove of names: Satan, Lucifer, Antichrist, Leviathan, Raakshas, Shaitaan, Daanav, Asuras, Iblis. Etymologically speaking, all of these names represent a certain transgression or challenge against the dicta of God. For example, Satan is from Hebrew which means "the accuser".[4] Lucifer has Latin roots and denote "the light-bringer",[5] much like Prometheus. Rakshas would be of Sanskrit origin, having the word Raksha in its root, equating with "protection". As per one account, the name Iblis, that of the Quranic demon, is taken from Arabic and has the verbal root "bls", meaning "to be in grief."[6]

Scripturally, the serpent is the devil incarnate in the Garden of Eden, which lures the Eve to take a bite of the apple from the tree of knowledge which leads to the Fall.[7] In the book of Job, Satan

https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf, Last Accessed: 29 August 2024 )

[4] Kelly, H.A.. *Satan: A Biography*. 2006. Cambridge, England: Cambridge University Press.

[5] From Latin (lux/luc- [meaning "light"] + -fer [meaning "bearing"]). It is also reminiscent of the transgression that Prometheus undertakes as per the Greek mythology in bringing the gift of fire to humans. Consequently, Prometheus was punished to have his liver regenerated and eaten up by eagles perpetually.

[6] Kazim, Ebrahim. *Scientific Commentary of Suratul Faateḥah*. 2010. New Delhi, India: Pharos Media & Publishing. p. 374

[7] Genesis 3:4-5, NIV, „'You will certainly not die,' the serpent said to the woman. 'For God knows that when you eat from it your eyes will be opened, and you will be like God, knowing good and evil.'" Link:

appears as a challenger to God and confronts him that Job only worships him because he has been prosperous thus far.[8] In the Quran, Iblis is present in the court of Allah and protests not to prostrate to Adam because of his superior fiery origins.[9] It is in this vein, that the folkloristic Mephisto is to be situated, He is similarly present in the court of God and has shared a certain history with Him and challenges him to corrupt his earthly "servant".[10]

That the Devil is an antagonistic, opposing force was thwarted outright as early as by the Roman theologian St. Augustine (354-430). The bishop of Hippo saw it as logically contradictory because God is perfect and incapable of defeat. Hence, the Devil is a creation of God and has consciously decided to fall from His grace due to free will.[11] The Italian priest-philosopher Thomas Aquinas (ca. 1225-1274) subscribes to this notion of Augustine. However, there seems to be a slight departure in the latter's understanding of the seduction of the Devil.[12] Whereas Augustine believes that suffering and evil should reconsolidate one's faith on God, Aquinas presupposes an antecedent malice on the subject's part. And thus, we enter a Hobbesian (1588-1679) world.[13]

The starkest proponent of German enlightenment Immanuel Kant (1724-1804) appears to be a bit undecided in the Leviathan world of cut-throat "self-love". According to him, one is either wholly good or wholly evil depending on the universal moral law that they bestow unto themselves.

---

https://www.biblegateway.com/passage/?search=Genesis%203&version=NIV#:~:text=After%20he%20drove%20the%20man,to%20the%20tree%20of%20life. (accessed: 26 August 2024)

[8] Job 1:9-11: "'Does Job fear God for no reason?' Satan replied. 'Have you not put a hedge around him and his household and everything he has? You have blessed the work of his hands, so that his flocks and herds are spread throughout the land. But now stretch out your hand and strike everything he has, and he will surely curse you to your face.'" Link: https://www.biblegateway.com/passage/?search=Job%201%3A9-11&version=NIV (accessed: 26 August 2024)

[9] Al-Ar'af 7:12: "Allah asked: 'What have prevented you from prostrating when I commanded you?' He [Iblees] replied, 'I am better than he is: You created me from fire and him from clay." Link: https://quran.com/en/al-araf/12 (accessed: 26 August 2024)

[10] Lines 297-318 in The Prologue in Heaven. In Goethe, J.W. von, Edited and Translated by Atkins, Stuart. *Faust I & II.* 1984. Princeton and Oxford. Princeton University Press.

[11] Augustine, Saint. *The Confessions of St. Augustine: Books I-IX (Selections).* 1942. New York: Prentice-Hall.

[12] Aquinas, Thomas. *Summa Theologica.* Edited by Thomas Gilby. OP. Cambridge: Blackfriars, 1966.

[13] Hobbes, Thomas. *Leviathan*. Edited by J. C. A. Gaskin, Oxford University Press, 2008.

There is a distinction among three degrees of evil. The first, normal evil, is due to the frailty of human nature. People are driven by self-interest. The second, radical evil, is due to impurity. People are so driven by self-interest that the costs of their decisions seem insignificant. The third, diabolical evil, is wickedness or perversity par excellence. People distort the world around them not necessarily in service of self-interest, but as an exercise of free will, to ensue pure chaos,[14] like the Joker from The Batman series.

A rationalization for the Hobbesian enterprise could be seen in the asceticism of Arthur Schopenhauer (1788-1860) when he asserts that self-preservation is a basic human tendency and one could indulge in evil in that endeavor.[15] His student, the skeptic Friedrich Nietzsche (1844-1900), however, polemicizes the entire structure of morality. In saying that the genealogy of morals lie in the herd mentality rising out of *ressentiment,* he overturns the complete conception of moral corruption and says one is free to follow self-interest and an ascription of evil is evil itself.[16]

Attempting to understand the origins of totalitarianism and the ensuing ghastly deeds undertook in its name, Hannah Arendt (1906-1975) harks back to Kant and turns "radical event" on its head, by saying that the subject is not necessarily driven by self-interest always, they may be vested with a structure that divests them of their free will. Humans as humans have become superfluous. As per her position, evil is banal, and everyone is capable of it. "Monstrous deeds do not require monstrous motives." [17] Naturally, there are poignant criticisms to her thought regarding Eichmann.

In pop culture, there is a certain Lord Voldemort, whose name must not be uttered, who can be seen as another pure, diabolical evil.[18] In a similar vein, the Rolling Stones' song "Sympathy for the Devil" hails the ever-present devil for his witnessing of the most heinous crimes in the history of humankind. A less evil version would be Darth Vader, who represents a fallen angel, seduced by the dark side. However, most suited to our purposes is Al Pacino as John Milton in *The Devil's*

---

[14] Kant, Immanuel. *Religion Within the Boundaries of Mere Reason*. Edited by Allen Wood and George di Giovanni. Cambridge University Press. 1998.

[15] Schopenhauer, Arthur. *The Will to Live*. 1967. New York: F. Ungar Pub. Co. Edited by Richard Taylor.

[16] Nietzsche, Friedrich. *On the Genealogy of Morality: A Polemic*. Translated by Maudemarie Clark and Alan J. Swensen, Hackett Publishing, 2009.

[17] Arendt, Hannah. Eichmann in Jerusalem: A Report on the Banality of Evil. New York, N. Y., Penguin Books. 1994.

[18] Columbus, Chris. *Harry Potter and the Chamber of Secrets*. Warner Bros., 2002.

*Advocate* (1997) who provides Keanu Reeves the right setting to pursue his highest passions and basest desires. Arnold Böcklin's (1827-1901) painting Self-Portrait with Death Playing the Fiddle (1872) is yet another fitting comparison to the figure of the modern devil. Fiddle, interestingly, was the legendary instrument Nero played while Rome burned.

**The Nature of Mephistopheles and Artificial Intelligence**

Not only is the conflation between the figures of Mephisto and AI is to be considered within their enabling, seductive framework, but also in that they are "spirit(s) of contradiction"[19], "spirit(s) of Eternal Negation"[20]. They both bring down the striving, thinking, human humans from their idealistic high horses and show them the magical allure of the material world, much like Sancho Panza. Whereas the former invites the Faust to engage in drunken debauchery and libidinal pursuits, the latter, well, does the same in a consumerist sense. Both can be considered "serious jests"[21] in the way Goethe had put it.

The contradiction with Mephisto for example, begins before the tragedy even begins, in the Prologue in Heaven, where a Devil is invited. The God and Mephisto go way back and often meet each other, despite their antagonisms. Upon finally revealing himself to Faust, the Mephisto announces himself as "a part of that force which, always willing evil, always produces good,"[22] and "part of the Darkness that gave birth to Light."[23] There is an ambiguous association and disassociation with the identity of Devil at many points in Mephisto's speech. Such as in Witch's kitchen, when the eponymous witch calls him Satan, he repudiates: "I will not tolerate that title. […] You cannot doubt my noble blood – Look at the coat of arms I wear!"[24] Contrastingly, in Faust's study, in the process of making the enticing wager, he proudly associates with the tag: "This universe – believe a devil – was made for no one but god! God lives surrounded by eternal glory, He casts us into utter darkness, And you must be content with day-and-night."[25] Mephisto,

[19] Line 4030 in Goethe, J.W. von, Edited and Translated by Atkins, Stuart. *Faust I & II*. 1984. Princeton and Oxford. Princeton University Press

[20] Lines 1338, *Ibid.*

[21]See: Washington, Ida H. "Mephistopheles as an Aristophanic Devil." *MLN* 101.3 (1986): 659-669.

[22] Lines 1335-1336. Goethe. *Op. cit.*

[23] Line 1350. *Ibid*.

[24] Lines 2505-2512. *Ibid*.

[25] Lines 1780-1784. *Ibid*.

being a devil, is sympathetic to the plight of humans, has come to provide a Dionysian hedonism and a final release.

He does sufficiently enable Faust to realize his morbid, villainous instincts. He rejuvenates the middle-aged Faust to become an enchanting youth again with the help of the Witch's potion.[26] Mephisto's trickery in procuring two necklaces for Gretchen - his desire object, at least in the first part – and consequent engaging her neighbor Martha in a conversation in the garden becomes useful for Faust's amorous intent.[27] It is on Mephisto's provocation and support that Faust kills Gretchen's brother, Valentine, who wanted to harangue Gretchen for having lost her honor.[28] Mephisto drags Faust to Harz Mountains on Walpurgis Night and makes him take part in an erotic dance with the witches.[29] Mephisto hastes Faust in the cell to abandon Gretchen as she can not be saved now.[30] In the second part, Mephisto devises a new fiscal policy for Faust so he can impress the Emperor[31], takes him to Helen,[32] and evacuates Philemon and Baucis.[33]

AI likewise has enabled access to a vast compendium of knowledge, but it contemporaneously fraught with epistemic biases in its individualizing the search results. On the one hand, the modes of information dissemination have become increasingly miniaturized, decentralized, and democratized, the stranglehold of its policing, moderation, and amplification still rests in the hegemonical hands of the oligarchic tech giants.[34] There is an overstimulating overabundance of information in the hollows of social media, but little attention is bestowed upon knowledge and

[26] In the Act "The Witch's Kitchen", *Ibid*.

[27] In the Act "Martha's Garden", *Ibid.*

[28] In the Act, "Night", Ibid.

[29] In the Act, "Walpurgis Night", *Ibid.*

[30] In the Act, "Prison", *Ibid.*

[31] In the Act, "The Throne Room", *Ibid.*

[32] In the Act, "Kinght's Hall", *Ibid.*

[33] In the Act, "Before the Palace." *Ibid.*

[34] See: Horkheimer, Max, Theodor W. Adorno, and Gunzelin Noeri. „The Culture Industry" in *Dialectic of enlightenment*. Stanford University Press, 2002., also, Chayka, Kyle. *Filterworld: How algorithms flattened culture*. 2024

wisdom.[35] Benjamin's statement that what the society has gained in breadth, it has lost in its depth, still holds.[36] Shafak uses the metaphor of moon for this phenomenon in her 2020 essay, "full of light and promise, and then an unexpected dark side […] of misinformation, hate speech, division, and falsehood."[37] To jumble Mephisto's words, AI is "a part of that force which, always willing good, always produces evil." There is always a ready association with digital technology with something that facilitates human connections, yet the modern woman has never been so alienated. AI, being a product, is sympathetic to the interests of profit and "has now come to the heart of desire," to quote Guattari.[38]

Culture, in the age of its digital reproducibility, is driven by a violent thrust towards a flat, featureless, redundant, and homogenous entity.[39] The digital technology has birthed an ever-penetrating, crawling, creeping, algorithmic panopticon whose task it is to amass endless data on each and every person. Pegasus, for example, was the talk of the town not two years ago.[40] It not only registers all our preferences, habits, opinions, and dispositions, but also dynamically inform and reinforce them by intelligent curation and recommendation. A new economy of millennial influencers has found its expression as a niche microcosm for selling partnered products. Each mindless finger touch on a screen should perpetuate the cycle of corporate profit.[41] We have landed in a generic landscape, wherein each city, café, restaurant, library, conference room resemble each other.[42] What one consumes is of no importance as long as one consumes. In such

---

[35] Shafak, Elif. *How to Stay Sane in an Age of Division: The powerful, pocket-sized manifesto*. Profile Books, 2020.

[36] Benjamin, Walter. "The author as producer." In: *Thinking photography*. 1982. pp. 15-31.

[37] Shafak, Ibid.

[38] Guattari, Felix. "Machine and Structure".in: *Change,* no. 12. Seuil. 1971.

[39] Jameson, F. "Postmodernism and Consumer Society." In: *The Anti-Aesthetic*, edited by Foster, Hal. Bay Press, 1983, pp. 111–25.

[40] "India: Damning new forensic investigation reveals repeated use of Pegasus spyware to target high-profile journalists", *Amnesty International.* 28 December 2023. Link: https://www.amnesty.org/en/latest/news/2023/12/india-damning-new-forensic-investigation-reveals-repeated-use-of-pegasus-spyware-to-target-high-profile-journalists/ (Last Accessed: 29 August 2024)

[41] Han, Byung-Chul. *Non-things: Upheaval in the Lifeworld*. John Wiley & Sons, 2022.

[42] Chayka, Kyle. *Op. cit.*

atmosphere, the culture industry prioritizes and incentivizes those quantifiable products which conform to the status quoist imaginations of consumption.[43]

In as far as Heidegger concerns the question of technology with revealing and unconcealment of *aletheia*, the truth of the matter remains that it reveals the sinister fallacy of human existence.[44] Today, our curated self-projection of happiness on the social media handles is a display of narcissistic vanity. The ready availability of gig workers to pleasingly deliver food at a mere click satiates the gluttonous instinct. The sheer unacceptability and subsequent lambasting of delinquent opinions on public fora is an exhibition of wrath. There are bound to be ample examples of other four original sins.

**The Nature of Human(s)**

If such a conflation is to be envisaged, then AI naturally raises pressing conundrums of metaphysical, epistemological, and ethical origins. Such concerns aren't to be dealt with on a distanced, abstract, or philosophical level, rather they are of immediate urgency since they have disrupted the everyday life. There are problems in the workforce such as Keynesian technological unemployment or reduction of engaging intellectual labor as David Graeber[45] points out, in the realm of unconsented sexual transgressions such as deepfakes[46] or the likes of Sulli Bai deals, and in the realm of creativity such as the use of ChatGPT[47] to write assignments, which has found an expression in the protests of the Writers' Guild in the US. However, due to the constraints of the format, this paper restricts itself to a brief discussion of three underlying moral issues: the perilous pursuit of knowledge, free will and moral responsibility, and the indispensability of the evil.

Faust, in the beginning of the play, laments that he has "studied now, […] Philosophy, Law, Medicine, and – what is worst – Theology, from end to end with diligence."[48] Yet, after all of this,

---

[43] Horkheimer, Max, Theodor W. Adorno, and Gunzelin Noeri. *Op. cit.*

[44] Heidegger, Martin. "The question concerning technology." In: *The Question Concerning Technology and Other Essays.* Transl. by Lovitt, William. Garland Publishing, Inc. New York & London. 1977.

[45] Graeber, David. *Bullshit Jobs: A Theory.* Simon & Schuster. India. 2019.

[46] Cf.: Murgia, Madhumita. *Code Dependent: Living in the Shadow of AI.* India, Pan Macmillan, 2024.

[47] Cf.: Baron, Naomi S.. Who Wrote This? How AI and the Lure of Efficiency Threaten Human Writing. United States, Stanford University Press, 2023.

[48] Lines 354-357, Goethe, *op. cit.*

he remains a "wretched fool"[49], not any "wiser than before"[50]. Technology, similarly, has engaged itself in the rat race of first mover advantage. The technocrat of today has infinite access to epochs of knowledge available at her laptop. But is she any better off than that nineteenth century polymath? It raises the question how did knowledge and a better life become conjoined? And, what is knowledge after all? One could argue that humans before everything are creatures of stories, homo narrans,[51] and faith, homo fidei[52]. In Theaetetus, Plato finally propounds after a long Socratic dialogue that knowledge is a "true belief with an account."[53] However, it is the dictum of *sapere aude* which led to the exile of the first humans from that heavenly garden. The most quoted Socratic sentence to this day remains, "All one can know is that one knows nothing." The weaver-poet Sant Kabir Das has aptly written, "Pothi Padhi Padhi Jag Muva, Pandit Bhaya Na Koye."

The notion of free will had always remained limited to begin with. In an age of restricted physical, material interactions now, these limitations have been further exacerbated. By exploiting the internal disorienting tendencies, evidence suggests that the echo chambers of social media "is emboldening people to cross the line and push the envelope on what they are willing to say to provoke and incite."[54] This has found its terrestrial manifestation, for example, in the Jamia

[49] Line 358, *Ibid.*

[50] Line 359, *Ibid.*

[51] See, Fisher, W. R. Narration, knowledge, and the possibility of wisdom. In: *Rethinking knowledge: Reflections across the disciplines*. 1995. pp. 169-192.; also, Campbell, J., & Moyers, B. *The power of myth*. Anchor. 2011; also, Harari, Yuval Noah. *21 Lessons for the 21st century.* Vintage. London. 2018. p. 235. "Homo sapiens is a storytelling animal that thinks in stories rather than in numbers or graphs, and believes that the universe itself works like a story, replete with heroes and villains, conflicts and resolutions, climaxes and happy endings. When we look for the meaning of life, we want a story that will explain what reality is all about and what my particular role is in the cosmic drama.", Similar argument by Harari can be found in: *Homo Sapiens. A Brief History of Mankind.* 2011.

[52] An academic study of this phenomenon is undertaken, for example, here: Bae, K. W. *Homo fidei: a critical understanding of faith in the writings of Wilfred Cantwell Smith and its implications for the study of religion*. Harvard University. 1997.

[53] Plato. *Theaetetus*. United States, Hackett Publishing Company, Incorporated, 2012.

[54] Although there are plenty of incidents in which the hate from social media has poured out in the real world and made itself manifest through acts of violence, this quote is taken from a report on the mass shooting at a Synagogue in Pittsburgh in December 2018. From Bubar, Joe. (2018) *Is Social Media Fueling Hate?,* The New York Times Upfront.

incident on 30 January 2020 when a 17-year-old "Godse" shot fire on the protesting students. In such a situation, the natural propensity is to disregard the moral responsibility and lay the popular claim, "The Devil made me do it." In the Expanse of an Open Country, similarly, Faust cries "All death and murder be upon your head, monster!", whereas it was he himself who had killed Valentine. This convenient surrender to fatalism and running away from responsibility is what the existentialist Sartre calls "bad faith"[55]. The French philosopher asks one to take account of her actions. The enlightenment philosopher from Königsberg, Immanuel Kant, would also not endorse a collective guilt. The 20th century political philosopher Hannah Arendt would understand that though individual responsibility is not to be done away with, there are also structural and systemic factors at play. It is a symbiotic relationship that humans and machines have entered into and thus the apt characterization of Mephisto by Faust as "you who know my heart and soul, why fetter me to this vile companion who feeds on mischief and rejoices in destruction?"[56]

The Big Data Demon, The Big Brother, is analogously aware of the contents of the individual's "heart and soul," and the individual is likewise "fettered to this vile companion" which flourishes on the mischievous spending spasms and destructive thinking proclivities. Regardless of however much the Industry 4.0 may be blamed for the perpetration and perpetuation of evil, the stark, bitter truth prevails that a life without it is unimaginable under current circumstances. "Here we are again at our wits' end, the point where, with you humans, minds give way. Why have you entered into partnership with us if you cannot keep its terms?"[57] The humankind has indeed reached an impasse where the discontent with the technology has become unsurmountable, yet a vehement opposition to the very technology on which it thrives would not only be counterintuitive but also foolish. This is the aporetic problematic of the age we found ourselves in the frustration of which is best uttered by the Devil itself, "Did we thrust ourselves upon you, or you on us?"[58]

URL:https://upfront.scholastic.com/issues/2018-19/121018/is-social-media-fueling-hate.html?language=english#1250L. (Accessed: 28.08.2024)

[55] Chapter Two, "Bad Faith" in Sartre, Jean Paul. *Being and nothingness.* 2003. H. E. Barnes, Trans.; 2nd ed. Routledge.

[56] Uttered by Faust in the only act written in prose in the tragedy, "An Expanse of Open Country" in Goethe, *op. cit.*

[57] Uttered by Mephistopheles, *Ibid.*

[58] *Ibid.*

## Conclusion

In conclusion, the statement the suggestion needs to be reinvoked that humans should say a resilient "no" to the fettering hands of the technology to preserve our integrity. As per the deliberations in this paper, the average cyborg of the day isn't in a position to resound with those high spirits. In his translation of the Holy Bible, Faust happily but somewhat hesitatingly mistranslates that "In the beginning was the act."[59] It is this dictum that the cyborg should rely on. The act should be that of embracing the sheer hypocrisy of being alive and embodying the multiple identities, languages, origins, machines, and other such cross-sections in which our ontos is entangled. On an individual level, the act should be that of finding that ephemeral equilibrium, that elusive middle ground, that fine moderation between the Apollonian and Dionysian.[60] It is a solitary initiative and the act should be of discovering the optimum balance between the intellectual pessimism and the willful optimism.[61]

## Acknowledgements

Thanks to FCSR and *Sophia Lucid* for providing the platform and the opportunity to pursue this idea. Thanks to Dr. Jigyasa Sogarwal, Jaya Sharma, Vivek Archis, Dr. Shambhavi Prakash, Dr. Sunil Choudhary, and Manasi Hulsurkar for hearing out this rather novel idea at various stages of the research and the writing. Eternal thanks to my partner, Ms. Dikshit, for her unwavering presence and everything she has made me.

## Bibliography

Augustine, Saint. *The Confessions of St. Augustine: Books I-IX (Selections).* New York: Prentice-Hall. 1942.

Arendt, Hannah. *Eichmann in Jerusalem*: *A Report on the Banality of Evil.* New York, N. Y., Penguin Books. 1994.

Arp, R. *The Devil and Philosophy: The Nature of His Game.* Chicago. Open Court Pub Co. 2014.

Aquinas, Thomas. *Summa Theologica.* Edited by Thomas Gilby. OP. Cambridge: Blackfriars, 1966.

Benjamin, Walter. "The author as producer." In: *Thinking photography*. Editor: Victor Burgin. London. Red Globe Press London. 1982. pp. 15-31.

Benjamin, Walter. *The Work of Art in the Age of Mechanical Reproduction.* Penguin UK. 2008.

Chayka, K. *Filterworld: How Algorithms Flattened Culture.* New York. Heligo Books. 2024.

[59] Line 1238, in "Faust's Study", Goethe

[60] Nietzsche, Friedrich. *The Birth of Tragedy: Out of the Spirit of Music.* United Kingdom, Penguin Books Limited, 2003.

[61] Gramsci, Antonio. *Selections from the Prison Notebooks of Antonio Gramsci.* United Kingdom, International Publishers, 1971., cited after: Shafak, *op. cit.*

Graeber, David. *Bullshit Jobs: A Theory.* India, Simon & Schuster, 2019.
Guattari, Felix. "Machine and Structure". In: *Change,* no. 12. Seuil. 1971.
Goethe, Johann Wolfgang von. *Faust I & II.* Edited and Translated by Stuart Atkins. Princeton and Oxford. Princeton University Press. 1994.
Han, Byung-Chul. *Non-things: Upheaval in the Lifeworld*. John Wiley & Sons, 2022.
Harari, Y. N. *Homo Sapiens: A Brief History of Humankind*. 2014. Random House.
Haraway, D. *A Cyborg Manifesto : Science, Technology, and Socialist-feminism in the Late Twentieth Century.* 1984.
Heidegger, Martin. *The Question Concerning Technology and Other Essays.* Transl. by Lovitt, William. Garland Publishing, Inc. New York & London. 1977.
Hobbes, Thomas. *Leviathan*. Edited by J. C. A. Gaskin, Oxford University Press, 2008.
Horkheimer, M., Adorno, T. W., & Noeri, G. *Dialectic of enlightenment.* 2002. Stanford University Press.
Kant, Immanuel. *Religion Within the Boundaries of Mere Reason.* Edited by Allen Wood and George di Giovanni. Cambridge University Press. 1998.
Murgia, Madhumita. *Code dependent: Living in the Shadow of AI.* London. Pan Macmillan. 2024.
Nietzsche, Friedrich. *On the Genealogy of Morals*. Translated by Michael A. Scarpitti. Penguin. 2013.
-, *The Birth of Tragedy: Out of the Spirit of Music.* United Kingdom, Penguin. 2003.
Plato. *Theaetetus*. United States, Hackett Publishing Company, Incorporated, 2012.
Sartre, Jean Paul. *Being and Nothingness.* Translated by: H. E. Barnes, Trans.; 2nd ed. Routledge. 2003.
Schopenhauer, Arthur. *The Will to Live*. Edited by Richard Taylor. New York: F. Ungar Pub. Co. 1967.
Shafak, E. *How to Stay Sane in an Age of Division.* London. Profile Books. 2020.
Washington, Ida H. "Mephistopheles as an Aristophanic Devil.", *MLN* 101.3 (1986): 659-669.

# ZOMBIE-AURAS: GENAI AND HYBRID TEXT PRODUCTION

*Joshua Nieubuurt*
*PhD candidate, Old Dominion University, Norfolk, Virginia &*
*Overseas Collegiate Assistant Professor at University of Maryland Global Campus-Asia*

**Abstract:** This paper explores the iterative evolutions of textual production and their impact on the "aura" of texts, as conceptualized by Walter Benjamin. The study identifies three key phases of textual production: the natural, the mechanized, and the digitized, each progressively displacing the "cult value" of texts. This cult value is lost through increased ease of creation, reproduction, dissemination, and dislocation of creators and audiences in time and space. The advent of Generative AI (GenAI) marks the latest evolution, transforming the "aura" into a memetic "zombie" form—familiar yet opaque, evoking both the sublime and fear. By examining Benjamin's notion of "aura" at the threshold of mass media technologies, this paper introduces the concepts of "hybrid auras" and "zombie auras," representing texts co-authored by non-embodied technologies that retain only a semblance of their original aura.

To illustrate this transformation, the paper begins with an analysis of the *Epic of Gilgamesh*, whose "natural" aura has endured through various textual evolutions. It then conducts a performance-based "aura check" by comparing Shakespeare's embodied, mechanically reproduced texts with GenAI-generated reproductions. This experiment highlights the distillation of texts and the impact of GenAI mimicry on the aura of mechanically reproduced works. Finally, the paper broadens its scope to contemporary GenAI co-authorship beyond literature, examining how the hybridization of mechanization and digitization creates new forms of textual production. The study argues for the necessity of attuning to these new literary forms to preserve the remaining humanities-based aura in text reproduction, preventing its complete transformation into a "zombie-aura."

**Keywords:** Generative AI, Digital Rhetoric, Critical Theory, AI Humanities

## Embodied textual production through time

One of Humanities long standing gifts is the ability to harness the power of narrative. Stories has moved mountains of meat since time immemorial: they have built and destroyed empires, helped to create new expressions that linger on generations after their intended audience, and have an awe inspiring way of captivating audiences and transporting them into once unfamiliar worlds with fresh ideas.. Great gods and heroes, flush with the best and worst of humanity, have left a lingering impression thousands of years removed from their origins and original interpretations. These deeply rooted tales (told by millions of tongues) have shaped and molded worldviews setting in motion the sublime power of storytelling through the ages. With the advent of formalized writing, an explosion of creativity and an increase in an audience displaced from the

original storyteller further altered our species' destiny. These tales began as vocalizations echoing across generations until the invention of material writing.

The writing itself existed in or on a thing (i.e. clay tablet, book, scroll, hieroglyphs, etc.). Each text was something that affected the embodied, a material piece of the power inherent in the words weaved upon or within them. Walter Benjamin proposes that in these material manifestations of the word (as well as any other material art form) there is an "aura" about them.[1] In his words an aura consists of "the here and now of the work of art–its unique existence in a particular place" (21). The power of materials meeting minds catalyzes into a special phenomenological experience: a touch of the abstract realms each person harbors deep within the unseen worlds they inhabit. For the purposes of this paper the aura will be considered as an immaterial ingredient given to a text through the embodied experiences of its creator(s) and audience(s). It is the piece of a text that signals the humanity behind the words that jumps into the phenomenological experiences of those who come across it. Wherever the written word has been witnessed, from towering deeply carved walls, unraveling papyrus rolls, to flickering monitors–with it comes the aura of the humans who helped perform its manifestation and dissemination.

Ancient analog textual production methods–though slightly altered through time and place–remained laboriously intact until the Goryeo dynasty's (roughly between 1234-1377) usage of metal type production and Johannes Gutenberg's printing press in the Middle Ages (1440-1454 from conceptual model to commercial use) (Sohn 96; Woo; Mosley; Kapr and Douglas). With these inventions of mechanized writing production, mass audiences had the potential to be greeted with an explosion of stories and information allowing old and new tales to be translated, written, and shared in bound forms far beyond historical limits. This performance of mass creation of a given text came at a price.

In doing so the wordsmith and their manifested texts became further displaced from an immediate audience in the presence of the author or text to strangers tethered by a text that may never once cross paths in their lived experiences. In addition, authors and texts became names conjoined to a solidified package of semiotic meaning traversing national and linguistic boundaries and reaching minds they would never know existed in ways they could never have anticipated. Despite the dislocation these people and things still remained bound through material copies of a text. With such a material object a token-aura could be held, observed, treasured or distrusted, and pondered over by humans–though diluted from the original–could also make the journey.

[1] It should be noted that Benjamin was focused primarily on "art" in its various forms. His insights into aura and reproduction here focus on the written word and digital rhetoric more generally.

Such a status remained "normal" until roughly around the late 19th and early 20th centuries. Around this time philosopher and cultural critic Walter Benjamin proposed that art traditionally composed through embodiment and delimited supply had lost a certain "aura" through its mechanical reproduction. It is no coincidence that the leap of the word from the printing press to its multimodal expressions (i.e. radio, television, cinema, and eventually its current incarnation: the digital form) coincides with the advent of modern mass media mediums. As new technologies were introduced and adapted, new ways to cast a shadow of aura-embedded embodiment out into the zeitgeist were found to bring the world to the individual: giving them a wider world view than their own bodies could possibly see, hear, or be cognizant of otherwise. The world became both larger in terms of insight and more connected through these technologies.

In the 21st century a globally wired world has emerged, and digital texts have become ubiquitous. In this late stage of textual production, a shift has occurred in which an author is not only dislocated spatially and temporally from their audience but also phenomenologically. It is important to note that material texts are still wildly popular and continue to maintain the flame of the aura sparked with the invention of writing (Fleck; Watson). In fact, demand and consumption of analog text has increased post COVID pandemic, but with the advent and convenience of digital e-books, online only publications, and social media much of what people see, read, and engage with has become digital. As of July 2024, an estimated 5.45 billion people (67.1% of the human population) are using the Internet and with that comes massive amounts of creation of and contact with digital texts (Datareportal). And although the Internet harbors great troves of information it distends (and sometimes distorts) the aura of texts through mass access with no single static physical token of representation. To further add to this newer stage of textual production, [co]authorship itself is no longer the sole possession of humanity, but rather a hybrid of flesh, data, and silicon with the increasing use of GenAI in digital spaces.

This paper comes from a vantage point in which there have been three iterative evolutions of textual production that have subsequently altered the "aura" of texts: the natural, the mechanized, and the digitized. In each of these three steps the "aura" of a given text has been further displaced in terms of what Benjamin described as a "cult value" due to its ease of creation, reproduction, and dispersal. Furthermore, the latest evolution found in the dawning of GenAI the "aura" has now become akin to the memetic form of a zombie. This latest iteration possesses a form similar to the known, its underlying processes are theoretically understood but remain mostly opaque, and the grander implications of its existence spark both ideas of the sublime powers of ingenuity and fear of the unknown. It is through Benjamin's "aura" formulated at the cusp of mass media technologies that the shift in textual production will be considered and explicated upon. In doing so this paper makes a case for both "hybrid auras" and "zombie auras." These auras are texts that are [co]authored through non-embodied technologies leading to the (re)production of texts with only a kindling of the aura of previous epochs.

To display this change in the aura this paper will begin by considering one of the earliest texts, the *Epic of Gilgamesh*, and its "natural" aura has transcended time. Its inclusion in this paper stems from its authentic aura that has persevered through its long history in all three textual evolutions. From there the paper will turn to a performance based "aura check" utilizing the words of William Shakespeare and GenAI (re)productions. The point of this section is to run a simple experiment to view differences in an embodied mechanically reproduced text (Via Shakespeare's famous Hamlet quote) to digitized (re)productions prompted to generate a similar quote. This section is meant to cast a light on the distilling of texts, the performance of GenAI mimicry, and its effect on the feel and aura of an embodied, yet mechanically reproduced, work. Finally, the text will broaden its scope to consider contemporary pragmatic GenAI [co]authorship in areas outside of literature. These realms are encompassed by the latest iteration of textual production; a hybrid form birthed from the mating of mechanization and digitization leading to a new spawn that pushes the aura of a text into abstraction. It is in this abstraction that "attunement" to the potentials of a new literary form becomes necessary so as not to lose the kindling of aura left in text [re]production to the "zombie-aura."

**The Natural: Gilgamesh and the cult-like aura**

Nearly 4000 years ago one of the most influential stories in the history of humankind was preserved in clay: *The Epic of Gilgamesh*. In this epic tale a partly divine king (Uruk) seeks to find the secret to immortality in doing so he and his newfound comrade (Enkidu) embark upon a series of adventures eventually leading to the death of Enkidu and a spiritual awakening for Uruk. This spiritual awakening gives Uruk a deeper understanding about the nature of existence: in particular the role of mortality–the great equalizer–in human life.

The story has and continues to be remixed and retransmitted into the world thousands of years displaced from its origins and original language (Ogleby). For many scholars the interest of this tale lies in its being the "earliest recorded thoughts on abstract topics such as the nature of 'nature' and 'civilization'; the nature of mortality and the inevitability of death; and the characteristics and behavior of 'heroes'" (Ogleby). Not much is known about its original author(s), their respective audience(s), and their own critical interpretations of the work. Nonetheless it remains a tale that still titillates audiences to this day. Somewhere within its creation, its physical manifestation, and the inherent messages lies the keys to its own "aura." The master key of which is its existence as a deeply human text (from creation to underlying themes) echoing timeless manifestations of the human condition.

Now far removed from its natural origins, the *Epic of Gilgamesh* continues to exhibit its own authenticity: an epic tale told and retold for thousands of years and eventually across the entire globe. It exists both as a series of clay texts that still inhabit [digital] museums (its *most* natural state), it is also in print and digital forms (mechanization and digitization), and still lingers in the

zeitgeist. Its thousands of years of circulation is in line with one of Benjamin's core components of the aura.

Benjamin notes that art's aura is partially based on "physical duration to historical testimony relating to it" (22). Its physical duration adds to its value as a cultural artifact (both the tale and the tablets themselves) attest to the aura of the object. Furthermore, the historical testimony to its lingering aura can be seen through its cultural relevance rippling into the 21st century.[2] Despite the translations, reproductions, and retellings of the tale it maintains a "cult-like" aura as one of the earliest existent mythological tales speaking to the human condition that is still in circulation. This feat is in part due to the mechanical reproduction of the text.

Through its translation and reproduction, the text has been able to put "the copy of the original in situations which the original itself cannot attain" including the multimodal technologies of the 21st century (Benjamin 21). Its initial "cult-like" aura has in part been lost through its iterations into new forms greeting new peoples during its long textual life. The *Epic of Gilgamesh*, despite the dilution of its cult-like status, is a prime example of the "aura" transcending time through each stage of textual form evolution. Due to its longstanding existence and influence it remains relatively "safe" from aura decay through its material and digital manifestations. As Benjamin writes, "the authentic work retains its full authority in the face of a reproduction" a status that the *Epic of Gilgamesh* and other significant ancient texts maintain even in the era of widespread digitization. Although there are certainly no substitutes for original works of art in regard to aura, contemporary GenAI models are capable of mimicking the stylization of such widely esteemed artists and their works. In the following subsection the aura and textual mimicry will be considered through the work of William Shakespeare's famous *Hamlet* quote.

**Authenticity and Aura: Shakespeare**

William Shakespeare's work was initially meant to be experienced with its aura emitting from the stage into a live, embodied, and perhaps bawdy crowd. Through the mechanization of text and multi-modal media production millions have come across Shakespeare either directly (through live plays) or indirectly via books, cinema, or remixed homages. The latter set generally comes at the cost of a decreased aura impact (even when language and setting have been updated to reflect the audience's phenomenologically grounded expectations). This aligns with Benjamin's thoughts on the aura of a work of art in that even if the aura has decreased through the process mechanization at least "enables the original to meet the recipient halfway" (22). The art, the word, and some recognizable aura exist in its reproduction that aligns with an embodied audience. It may not be the "purest" or "direct," but nonetheless great art continues to echo its authenticity through mechanization. What can be seen today with widespread style mimicry via GenAI seems to pay homage to great artists of the written word but misses the mark in regard to

[2] This includes movies such as the animated *Gilgamesh* and scholarship (Al-Ani; Araújo).

the continuance of the token-aura of great (re)productions. Instead, a new form of aura is being manifested: a hybrid aura… the zombie-aura. To consider this more carefully we must first consider how LLMs work.

LLM and other GenAI models (i.e. diffusion models) use large amounts of human derived data in order to train and (re)produce their own stochastic outputs (Chu). All this data has been digitized for the mechanisms of an LLM to be trained upon it. In doing so the aura of a text undergoes displacement from the original source. It becomes cadaver-like as it is intentionally dissected, studied, composted, and reborn into new computationally likely forms. During this process the whole bound aura of a text becomes split apart, cast into billions of tokenized units aiding the digital machine in forecasting the next most likely word based on an outside prompting mechanism. This extends not only to the physical manifestation of the text and its bound aura, but also its embodied author and audience(s). In doing so the form of a text appears "human" enough at least briefly. Once inspected more closely this form speaks of the artificial processes hidden behind the facade. To illustrate this a prime example of this can be found in the GenAI performance of style mimicry.

Before the explosion of public usage of LLMs in the early 2020s experiments were already being conducted on harnessing the power of natural language processing and machine learning techniques. Jhamtani et al. in their work (focusing on "Shakespearizing" contemporary language) displays the difficulty involved in the computational distillation of a text's form: it takes a lot of ingenuity and novel work. This grueling computational process begins with a concoction of digital texts, many of which once harbored the diluted aura of mass reproduction (in book form then transferred into digital form).

To train their model researchers used Shakespearean works, SparkNotes, and human embedded training data that was not included in previous training models (Jhamtani et al.). This process also included extensive human input through "dictionary mapping between large numbers of Shakespearean and modern English Words" (15). Through their novel process the researchers were able to "transform Modern English text to Shakespearean style English" (17). Their work, among many others in "style adaptation" is among those that helped to show the power of computational mimicry through the use of a digital machine to further displace the style of a given author and their texts.[3] Other researchers have also helped raise the bar for generated text that is passable among the public.

Two similar studies conducted by Lau et al. through their "Deep-speare" model helped to display that mimicry at the word level was not only possible but could also be developed into more

[3] Stylistic mimicry has come under fire by artists across the arts and continues to be a hotly debated topic among artists, legal specialists, and scholars. Perhaps this relates to the artists attunement to the loss of the original aura poured into their work (Knight; Hilden; Kafka).

complicated forms such as a sonnet. These artificial sonnets were then shown to non-specialist readers that could not distinguish between computationally generated works and those crafted by Shakespeare (Lau et al.).[4] Their results showed that non-specialists found it difficult to distinguish the difference, but specialists were readily able to discern them. A key difference cited by one of the co-authors of the Lau et al. study noted that to a specialist's eye the AI generated texts lacked a certain pathos driven impact and slight mishaps in the readability of the text (Lau et. al.). Like the unique gait of cinematic zombies, the zombie-aura also displays signs of its transformation (Wu et al.; Tang et al.). The zombie-aura is the uncanny algorithmic impersonality often cited as a tell-tale sign of GenAI texts: the hollow *feeling* language masquerading as a living corpus.

This lack of pathos, and specter-like aura, speaks to the ideas of one of AI's most notable critics Hubert Dreyfus and his thoughts on the phenomenological limitations of a thinking machine being disembodied.[5] One of the most notable for this subject is an AI's inability to dwell in a phenomenological experience encapsulated within a human body. Therefore, the plethora of experiences one has that feeds into their creative work are also incapable of being (re)produced due to a lack of base data greedily stored in the embodied human experience. Instead, the human prompters embodied experiences are funneled into the delimited experience of prompting. The GenAI then takes this information and conjures its best guess at what the prompter was calling forth. Through this process a new hybrid aura takes shape in the zombie form. A spark of human emblazoned life (via the prompter) enters into the machine's range of affordances and draws forth text from the well of tokenized texts it has trained on (the disassembled corpses of billions of texts).

To display these qualities the following figure (1A) compares the power of one of the world's most famous quotes from William Shakespeare to AI generated (re)productions.[6]

[4] Deep-speare specifically focused on Elizabethan English via the usage and (re)production of a Shakespearean sonnet.

[5] See *What Computers Still Can't Do*; and "Why Heideggerian AI failed and how fixing it would require making it more Heideggerian." for insights into Dreyfus' positionality on the issue.

[6] AI prompt on 6 Sept. 2024: I am writing an academic paper about aura and authorship comparing embodied writing to that of LLMs. I would like you to take your best shot at recreating a famous Shakespeare quote and to recreate it in a convincing way. Here is the quote: "To be or not to be—that is the question: Whether 'tis nobler in the mind to suffer The slings and arrows of outrageous fortune, Or to take arms against a sea of troubles, And, by opposing, end them.

**Figure 1A**

| Shakespeare | ChatGPT |
|---|---|
| "To be or not to be—that is the question: Whether 'tis nobler in the mind to suffer The slings and arrows of outrageous fortune, Or to take arms against a sea of troubles, And, by opposing, end them." | "To exist or not exist—that is the inquiry: Whether it is more honorable in thought to endure The piercing darts and cruel blows of fate, Or to rise and combat a tide of sorrows, And, by resisting, overcome them." |
| Gemini | Copilot |
| "To process or not to process—that is the query: Whether 'tis nobler in the neural net to endure The biases and errors of a faulty dataset, Or to take arms against a sea of misinformation, And, by recalibrating, end them." | "To exist or not to exist—that is the quandary: Whether it is more honorable in the spirit to endure The darts and barbs of outrageous fate, Or to rise up against a vast ocean of woes, And, by confronting, put an end to them." |

As can be seen ChatGPT and Copilot "understood" the prompt and utilized tokenization to generate near copies of the original. The results are close enough to "get" the point of the quote, but as noted in Lau. et. al., Dreyfus, and by contemporary writing professors/teachers across the world, it leaves a rather uninspiring example contrasted next to the original embodied script.[7] The generated texts exist as something "new" but are essentially recast from an altered previous mold. Their zombie-aura does not hold up to scrutiny

Surprisingly, the Gemini model took the prompt in a very different direction. This small example illuminates the gap between prompting for a desired output and receiving it as well as the differences and limitations between various LLMs. This prompt-dissonance is important to keep in mind as although the general structure of the soliloquy remains intact the "inside" is even more hollowed out than the other LLMs in the prompted example. Decontextualized from the authenticity of the initial source the text takes on meanings far outside the scope of the initial embodied text. Even when a model understands the prompt, tokenizes the text, and (re)produces

[7] It should be noted that both ChatGPT and Copilot are run off the same GPT base model making their similarities unsurprising (Microsoft365).

in a near enough fashion it still lacks a certain pizazz that a human author's and audience's unique experiences throw into the mix.

This returns us to the zombie metaphor deployed earlier, the form appears similar enough to pass… at a distance. This zombie can be herded into a specific direction (via prompting) but will not necessarily end up in the desired location (as seen in the Gemini example). Yet, what is often overlooked at this point is the role the human prompter played in helping this text come into creation. By the performance of prompting the human and the LLM then become a micro assemblage: Two enormously complex networks combining and catalyzing the birth of a stochastic output (even if mimicry is involved). Within this engagement the human prompter gives a tiny bit of aura, but it also takes on the aura of the digital machine. This is made clear when a prompt engineer specifically attempts to mimic the style of an embodied artist.

Prompting a (re)production of an original work forgoes consideration of the context and time in which the embodied author was writing for their audiences. Through computation a similar text is crafted but remains lacking in the pathos and auratic vigor of the original. It is as if once the dissected members of a text are then asked to transmogrify back into their original form after being jumbled together with the parts of others: A digital Frankenstein's monster text.

This leads us into the latest iteration of textual creation and production: The hybrid form of LLMs further displace authorial intent and the aura of artistic [re]production through artificially quantifying and distilling the human data-driven works they were trained on. In doing so the role of a text's distinctive "aura" is *revived* in a way akin to the awakening of the undead: reanimated, but surprisingly hollow shades of their former being. There is a long way to go until GenAI can "replace" human creativity and ingenuity in the realm of writing and an even further–perhaps impossible–distance to traverse before the phenomenological uniqueness of human embodiment is found to be replicable. To add to this, the loss of human-aura is not limited to the realm of the arts in their (re)production. In fact, the GenAI creep has begun to spread into other facets of human life with a similar loss in the human aura (for better and for worse) of what was once traditionally purely based on human-to-human textual productions.

**Digital texts and communication of the future**

Although much of this paper has been focused on the evolution of the aura through (re)production via Literature, this can also be extended to any text-based communication. It seems likely that soon AI agents will generate text (or vocalized text-based communications) in areas traditionally steeped in human interpersonal communication. Henry Jammes of Microsoft, estimates "750 million apps will need to be built by 2025" that incorporate some form of GenAI (Ball). Among these apps include deployment into fields such as customer service chatbots, healthcare provider AI agents, and personalized classroom instructors (Uspenskyi; Hippocratic

AI; Carroll). Each of these realms has always included a certain degree of technology to persist, but never have they been given such agency in the tasks they perform.[8]

Contrary wise these activities have historically been humans helping humans via the medium of spoken or written communication. Whether it be through the managing of a product error, the teaching of basic math, or advanced emergency brain surgery, people engaging in these embodied activities have utilized texts as essential parts of the process. A sense of the historical embodied praxis is an important concept as AI alignment practices shift between the desire for model preciseness and niche expertise so that the human aura (though likely diminished) remains tethered to the production of texts.

For some scholars this includes finding ways to keep experts in the loop. For example, in the field of Literature Heerden and Bas call for GenAI developers to include literary experts "to improve the quality of generated text" so that together they can "explore how structural elements jointly convey meaning and emotion" (176, 183). In this instance another layer of human input would come into play. The inclusion of subject matter experts may be one route to help in subverting aura loss. But a better approach may come through attuning human users en masse to the capacities of GenAI to equally distribute and maintain an embodied human presence within the loop.

The act of "attuning" to new technologies is something humans do particularly well (although not at the breakneck speed of technological development). Thomas Rickert in his work *Ambient Rhetoric* notes that attuning to things requires close attention to taking "things at their world and not just their word" (162). Taking things "at their world" is not an easy task to accomplish (162). The matter requires people to consider themselves, a given thing/technology, and their respective roles and affordances within much larger assemblages of people and things. When considering the role of hybrid texts moving forward, we must also consider the role of hybrid auras.

As can be seen in the previous subsection hybrid auras are produced in ways in which the embodied can co-operate with GenAI to generate text. As also noted, there is a danger of hybridity turning into a zombie-aura. If left unattuned to future purely GenAI texts may structurally appear in correct form but will remain lacking in their ability to connect on the same level as other texts originating before and during mechanical reproduction. Without alignment of both GenAI and the embodied human, the aura of text production is bound to meet an irreversible zombie-aura fate sooner rather than later. As Benjamin writes, "The alignment of reality with the masses and the masses with reality is a process of immeasurable importance for both thinking and perception" (24). If this hybrid text can somehow find a way to create an

[8] Some analog examples include the pencil, voice automated messaging systems, and incredible machines such as the Da Vinci Robotic Surgical System.

authentic hybrid aura there may not be the tension of "substitut[ing] mass existence for a unique existence" as Benjamin perceived mechanization to do (22). Rather, through attunement, and co-operation the hybrid aura may lead textual production into a Renaissance: a stage in which flesh and silicon find a healthy medium for the advancement and benefit of both.

**Bibliography**


Al-Ani, Asmaa Sadeq Abdulkareem, and Ghada Al Slik. "Architecture and Ancient Epics-Mesopotamian Gilgamesh as a Model." Association of Arab Universities Journal of Engineering Sciences 31.2 (2024): 1-15. Accessed 16 Aug. 2024, Web.

Araújo, Naiara Sales. "The Epic of Gilgamesh and the Odyssey: An Examination of the Supernatural Elements." International Journal 12.1 (2024): 22-31. Accessed 18 Aug. 2024, Web.

Ball, Stephanie. "LLM Use-by Data - Who's Keeping the Data Up to Date?" *ERP Today*, 18 Dec. 2023, erp.today/llm-use-by-data/. Accessed 3 Sept. 2024, Web.

Benjamin, Walter. *The Work of Art in the Age of Its Technological Reproducibility, and Other Writings on Media*. Harvard UP, 2008. Print.

Chu, Zhibo, et al. "History, Development, and Principles of Large Language Models-An Introductory Survey." *arXiv preprint arXiv:2402.06853* (2024). Accessed 3 Aug. 2024, Web.

Datareportal. "Digital Around the World." *DataReportal – Global Digital Insights*, Kepios, Aug. 2024, datareportal.com/global-digital-overview. Accessed 2 Sept. 2024.

Dreyfus, H. *What Computers Still Can't Do: A Critique of Artificial Reason*. The MIT Press, 1992.

Dreyfus, Hubert L. "Why Heideggerian AI failed and how fixing it would require making it more Heideggerian." *Philosophical psychology* 20.2 (2007): 247-268.

Hilden, Nick. "Protecting Artists from Theft by AI." *Nautilus*, 25 June 2024, nautil.us/protecting-artists-from-theft-by-ai-660557/. Accessed 31 Aug. 2024, Web.

Hippocratic AI Inc. "About Hippocratic AI." *Hippocratic AI*, 2024, www.hippocraticai.com/about. Accessed 28 Aug. 2024, Web.

Jhamtani, Harsh, et al. "Shakespearizing modern language using copy-enriched sequence-to-sequence models." *arXiv preprint arXiv:1707.01161* (2017). Accessed 13, Aug. 2023, Web.

Kafka, Peter. "The AI Boom is Here, and So Are the Lawsuits." *Vox*, 2 Feb. 2023, www.vox.com/recode/23580554/generative-ai-chatgpt-openai-stable-diffusion-legal-battles-napster-copyright-peter-kafka-column. Accessed 1 Sept. 2024, Web.

Kapr, Albert. *Johann Gutenberg: The Man and His Invention*. Routledge, 1996, Print.

Knight, Will. "Algorithms Can Now Mimic Any Artist. Some Artists Hate It." *WIRED*, 19 Aug. 2022, www.wired.com/story/artists-rage-against-machines-that-mimic-their-work/. Accessed 10 Aug. 2024, Web.

Lau, Jey Han, et al. "Deep-speare: A joint neural model of poetic language, meter and rhyme." *arXiv preprint arXiv:1807.03491* (2018). Accessed 13 Aug. 2024, Web.

Lau, Jey Han, et al. "" Deep-speare" crafted Shakespearean verse that few readers could

distinguish from the real thing." *IEEE Spectrum* 57.5 (2020): 40-53. Accessed 13 Aug. 2024, Web.

Microsoft365. "The Copilot System: Explained by Microsoft." *YouTube*, www.youtube.com/watch?v=E5g20qmeKpg. Accessed 2 Sept. 2024.

Mosley, James. "11 The Technologies of Print." *The Book: A Global History* (2013). 130-153. Print.

Ogleby, Jared. "Epic of Gilgamesh." *Epic of Gilgamesh*, Sept. 2007, p. 1. *EBSCOhost*, search.ebscohost.com/login.aspx?direct=true&db=khh&AN=26773146&site=ehost-live&scope=site. Accessed 30 Aug. 2024, Web.

Rickert, Thomas. *Ambient rhetoric: The attunements of rhetorical being*. University of Pittsburgh Pre, 2013. Print.

Sohn, Pow-key. "Early Korean Printing." *Journal of the American Oriental Society* 79.2 (1959): 96-103. Accessed 25 Aug. 2024, Web.

Tang, Ruixiang, Yu-Neng Chuang, and Xia Hu. "The science of detecting llm-generated text." *Communications of the ACM* 67.4 (2024): 50-59. Accessed 3 Sept. 2024, Web.

Van Heerden, Imke, and Anil Bas. "Ai as author–bridging the gap between machine learning and literary theory." *Journal of Artificial Intelligence Research* 71 (2021): 175-189. Accessed 11 Aug. 2024, Web.

Watson, Amy. "Book Sales by Year: Print Books in the U.S. 2023." *Statista*, 5 Jan. 2024, www.statista.com/statistics/422595/print-book-sales-usa/. Accessed 6 Aug. 2024.

Yoo, Woo Sik. "The World's Oldest Book Printed by Movable Metal Type in Korea in 1239: The Song of Enlightenment." *Heritage* 5.2 (2022): 1089-1119. Accessed 25 Aug. 2024, Web.

Wu, Junchao, et al. "A survey on llm-generated text detection: Necessity, methods, and future directions." *arXiv preprint arXiv:2310.14724* (2023). Accessed 12 Dec. 2023, Web.

# EXPLORING THE INCREASING ROLE OF DIGITAL CULTURE IN THE CREATION OF LITERARY GENRES: A STUDY OF THE DARK ACADEMIA MOVEMENT

*Shadab Zaveri*
*Co-Coordinator, Department of Strategic Communication and Journalism,*
*Sophia College for Women (Empowered Autonomous), Mumbai*

**Abstract:** In the contemporary era, digital culture has emerged as a dominant force shaping various facets of human interaction, creativity, and consumption. The advent of social media platforms and online communities has facilitated the rapid dissemination and evolution of cultural trends, leading to the formation of new subcultures and aesthetic movements. Digital culture encompasses the practices, attitudes, and norms arising from the use of digital technology, influencing how people create, share, and engage with content. This interconnected virtual landscape has redefined traditional boundaries, enabling niche interests and movements to gain unprecedented visibility and traction. Dark Academia is a literary and aesthetic movement that first began on Twitter in 2015 and gained popularity as a digital cultural movement during the global pandemic of 2020. The context of the shift in the digital marketing trends in the publishing industry along with the increasing participation in creation and consumption of digital culture are two primary components that contributed to the ascent Dark Academia movement. This paper seeks to understand the role of digital culture in the transformation of genres in English Literature through an analysis of the Dark Academia movement. It intends to look at the posts and trends on digital platforms like YouTube, Instagram, Tumblr, and TikTok as texts in order to study the manner in which they contribute towards the establishment of Dark Academia as a cultural movement and an upcoming literary genre. Through this study the paper hopes to expand the ongoing dialogue between the fields of Popular Culture, Digital Humanities and English Literature.

**Keywords:** Digital humanities, Digital Culture, Digital Platforms, Dark Academia.

Cultural change and digital technologies share a symbiotic relationship that has evolved with progress. The shift of human interactions from analogue to digital has also seen an upward growth trajectory. The emergence of social media and digital platforms has led to an unprecedented digital presence for all individuals across the globe. Digitality can be thought of as a marker of culture because it encompasses both the artefacts and the systems of signification and communication that most clearly demarcate our contemporary way of life from others (Gere, 16). The rapid expansion of digital platforms has redefined social engagement and broadened cultural landscapes. Digital technologies are embedded in contemporary discourse (Hlynka). Digital culture encompasses the practices, attitudes, and norms arising from the use of digital technology, influencing how people create, share, and engage with content. This interconnected virtual landscape has blurred traditional boundaries, enabling niche interests and movements to gain unprecedented visibility and traction. Digital culture has created spaces that have sped up

the rate at which information is shared locally and globally. It only takes hours or even minutes for something to go viral or become a global phenomenon. In the era before the popularity of social media platforms niche interests or even sub cultures would remain limited or isolated. The expansion of digital platforms redefines personal engagement with cultural content for everyone thus enabling niche subcultures to transcend global boundaries. Creation of a digital community is now accessible to anyone who can access the internet. This allows subcultures to get a chance at becoming popular culture in the global world. At the ground level, what it means is that anyone can create and influence global digital culture. The interplay between cultural evolution and digital technology has affected a lot of areas, including the creation and dissemination of new literary genres. The rise of Dark Academia movement and its subsequent growth into a literary genre aids the argument that is made here regarding the contribution of digital platforms and digital marketing in the process of genrefication in English Literature.

Dark Academia is a subculture that blends the aesthetic appreciation of everything Gothic, vintage and fashion and the pursuit of Knowledge. Rooted in the traditions of classic Gothic fiction and the college novels, Dark Academia gained traction due to social media platforms like Tumblr. It originated as a movement on Tumblr initially drawing inspiration from the thrilling novel *The Secret History* (1992) by Donna Tartt which includes a fascination and yearning for classical studies. Several of the major characteristics of the movement are either inspired or directly borrowed from this seminal text. According to Mukadam, " In its earliest iterations on Tumblr, fans of Dark Academia focused on *The Secret History*, and it involved fans creating moodboards (collages of pictures curated around a particular theme), "fancasting" celebrities as The Secret History's central characters, and recommending books and movies that match the "vibe" of Donna Tartt's world" (Mukadam, 2022). The book then seems to have been the starting point for the aesthetic references and stylistic inspiration behind several classic Dark Academia looks. The community unfailingly credits Donna Tartt's *The Secret History* (1992) as the center of the Dark Academia universe, for embodying everything crucial to the genre (Taylor, 2022). As digital users began to interact with the movement, it expanded at an unprecedented rate in the COVID-19 global pandemic with lockdowns that made physical spaces unaccessible. The paper dives into how Dark Academia has grown from an online aesthetic movement into a recognized genre in literature. As cultural signifiers shift into the digital realm, a global fluidity emerges in the exchange of ideas and interactions which have the power to blur or dismantle tradition. Social media platforms like YouTube, Instagram and TikTok allow for the creation and flourishing of cultures cultivated through algorithms and technology to amplify visibility and foster participation at a global level, transforming people from passive consumers of culture to active creators leading to the emergence of BookTube, Bookstagram and BookTok.

Since its launch, YouTube has successfully managed to popularize the art of vlogging. Vlogs refer to videos that are an audio-visual blog, where the content creators put their thoughts and views out in the world for viewers to view, like and follow them. What became a unique feature of

YouTube was that eventually content creators or vloggers could earn money through advertisements and paid promotions as well. Sub communities started developing on the platform when a group of YouTubers would create vlogs that were specifically catering to a particular audience like cooking tutorials or a specific genre or activity. One such rising sub community was the one that focused on reading, discussing books and also sharing book hauls. It came to be known as BookTube and it became a very prominent niche space within the literary community. Along with creating a collaborative book community, BookTubers are changing the way book publishers approach marketing, connecting with their audience, and publishing debut authors (Mitchell, 2021). BookTubers create vlogs whereby they talk about the books that they read, including dissecting the book as if it were a book club. Katytastic (@booksplosion) had run a BookTube channel for 10 years where one of their accomplishments include a special with Michelle Obama for the autobiography release. Jack Edwards, another famous booktuber, has more than a million subscribers. BookTubers had also started to collaborate with other BookTubers and start book clubs together where they would host a monthly club meeting, they have readalongs with their subscribers as well. A feature of this niche community that makes it attractive for digital marketing is the interactive aspect of it. Fans and viewers can ask questions, request book titles and even offer feedback which makes it easy to collect data from them. Publishing houses started sending advance reader copies, popularly known as ARCs to these BookTubers to review them and thus creating a fan base for the book prior to its publication. "Most of these ARCs are sent from the "Big Five" publishers: Hachette, Harper-Collins, MacMillan, Penguin Random House, and Simon-Schuster" states Brekelmans. They often also sponsor their videos for these book promotions or to advertise merchandise related to a book community or fandom.

With the popularity of Instagram, there was an eventual relocation of content creators to this newer digital platform. This move did not however change the basic ideas of an interactive, communication centric literary community. The niche space on Instagram was now called Bookstagram and users began to upload their posts instead of videos. Ova ceren (@excusemyreading) has 5.7k followers and has a significant influence over the online book community. The usage of images and captions instead of videos as on YouTube changed the game to a certain level because it allowed people who may not be the most comfortable with putting themselves out there to still talk about their interests and passions. It allowed for a form of expression that was both momentary and gave the option to remain anonymous, which attracted a larger group of people. The identity of the account no longer needed to be restricted by a person, in fact people had also started to share an account together. This was and is especially relevant in context to posting about books and fandoms because multiple people could then make edits and share posts on the same account increasing the content created on it and thus having a bigger fan/ following base. Bookstagram account holder @maevaeatsbooks has 184k followers and chooses to keep their personal identity private while stil sharing book reviews, literary

discussions and many book aesthetics. In relation to digital marketing, the methods and ideas of promotions and advertising for the publishing houses and the authors remained the same.

In recent times, the digital platform called TikTok has made it the new popular cultural space attracting a lot of digital users. TikTok has reels which are videos that can be a single long video or an amalgamation of several pictures or videos with a time duration of minimum ten seconds and maximum one minute. Users on this platform began uploading reels of various activities including readers' opinions or book unveilings etc. The sub community of BookTube and Bookstagram is called BookTok here. BookTok became more popular than BookTube because it required lesser time and attention spans. It also focuses on specific genres, BIPOC and LGBTQ+ authors or specific genres. In November 2023 it was reported that the #BookTok had been viewed over 200 billion times on TikTok with about 60 billion videos (Ariyanayagam, 2023). As social media platforms started aiming towards making the content fast consuming due to more demand for less attention consuming content, BookTok became more and more popular. Once again, publishing houses and authors started using the platform for digital marketing in almost similar ways. Bookstagram has a unique feature that allows it to share links to BookTube platforms on its own stories and also share reels of BookToks which makes it an important digital platform that can target a larger audience too. Both Bookstagram and BookTok have a reposting feature that allows publishing houses, marketing agents and especially authors to repost the edits and reels of the other users. This motivates the users to create more content in the hopes that they too will be reposted by the author and thus gain more recognition or popularity; or simply to be noticed by the author of whom they are a fan. Regardless of the reasons, it increases the posts based on a particular text, series or an author and thus creates a hype around it. The strategy of using digital marketing is most effective and visible when promoting and launching a new book. The author is encouraged to create a mystery around it by making posts and countdowns to a specific date with captions such as "Important announcement in two weeks" and then posting something new related to the announcement, containing small clues in it every day until the date of the announcement. The feature of YouTube and Instagram that comes in here is the live option which allows the user to interact with the audience in the live stream from their own personal device In 2021 an author tried to create intrigue around her upcoming book by sending dozens of packages to the BookTok community to decode a message that came with the book; gaining 2 million views on TikTok and has impacted sales drastically (Colyard, 2024).

Social media marketing became an important part of marketing in 2008, due to the upsurge of smartphones which congealed the widespread potential of social media (SM) platforms like Facebook (FB) and Twitter.(Nolan, 2018). Since then, digital marketing has had its own growth and development which in itself is an area that is being studied within digital humanities. The publication industry faced its own unique challenges when there was an increasing traffic moving towards digital spaces. The paperback had felt the threat from the e-book, especially when the Kindle device started gaining popularity. The allure of carrying around your entire

library, along with the financial benefits were primarily the reason behind the threat that e-books carried. However, within the capitalistic world of the publishing industry, publishing giants managed to incorporate the threat into yet another product that they could benefit from. Eventually as the novelty wore off, the paperback was and is still thriving. Speaking to CNBC, Meryl Halls, managing director of the UK's Booksellers' Association, said:"I think the e-book bubble has burst somewhat, sales are flattening off, I think the physical object is very appealing. Publishers are producing incredibly gorgeous books, so the cover designs are often gorgeous, they're beautiful objects" (Hughes, 2020). One key aspect in the marketing of the paperback is the culture created around a book by the author and the publishing house wherein digital marketing plays a pivotal role. It is important to note that digital marketing is not limited to the paperback only but it is one of the most selling formats of the written book overall. As Howarth states, "The opportunities associated with digital marketing and promotion in the publishing world are seemingly endless" (Howarth, 2022.) While publishing houses themselves also contribute to the creation of the culture around a book in different ways, the fans form a personal bond with the author. The crucial role played by the author as a social media celebrity is what makes a digital campaign successful. To promote their work, many authors are being thrust into the position of social media superstar or content marketing expert (Hughes, 2020). It is akin to building up the author as a brand from the perspective of the digital publishers.

There are multiple ways of utilizing digital platforms ranging from hosting digital book readings, conducting interviews with social media influencers, organizing giveaways and also utilizing email marketing. Authors have started creating their own monthly newsletters that fans can sign up for which includes sharing glimpses of unpublished upcoming books, blurb reveals, cover reveals, information on what the author is currently working on, etc. This makes the fans feel personally connected to the author and creates the bonds that over time start feeling like a community space. An example of this would be the author Rebecca Yarros (@rebeccayarros) whose books Fourth Wing and Iron Flame have managed to become the top two most read books overall of 2024 according to Goodreads. Rebecca Yarros has used her digital profile to promote her books constantly. These two books have a sequel called The Onyx Storm that is gaining digital traction through her promotions. The book's title, its cover reveal and blurbs from the book have been a part of her Instagram page since June 2024. The book will be released in January 2025. She has been yet another TikTok book sensation of the year. Hyrkin says, "Social media has become a set of sophisticated tools for matching content to reader interests, growing communities and building brands" (Hyrkin, 2015). Digital platforms work as digital marketing because of the role played not just by the authors but also at the end of the fans and popular literary social media accounts. All of this creates cultural practices that are consumed by the literary masses on digital spaces.

The yearning or the need for a cultural practice in a given social context leads to the creation of newer cultural contexts that arise out of the ways in which society deals with that need of that

specific group that demanded it in the first place. When and if these contexts start trending online on digital platforms, they become a part of digital popular culture and lead to an increase in demand for products that make that cultural context more visible. One of the demands that go up is a demand for fictional representation of the cultural practices and trends. This in turn causes publication houses to find and cater to that demand and thus publish more books of a specific category that fulfil the needs of the readers. A vital conundrum in the analysis of this process is that of the understanding of what sways the influence within the communities in the movement- the literature or the aesthetic culture? How are we to grasp, let alone classify, a slippery concept that serves both as a market category within the publishing world and as an organizing term for a decentralized network of online communities (Stowell, 2022)? The digital marketing processes in the publication industry create a loop where the social media marketing of these books makes the cultural practices and movements trending, and the popularity of the posts and feeds creates a demand for more books. The world of literature and the popular online culture is interconnected because they both fuel each other. The world of literature is not hermetically sealed off from the more populist internet cultural sphere; it is in fact absorbed by it (Taylor, 2022). This is how the process of genrefication then takes place, whereby after enough publications the books can now be called a genre of the same name as the cultural trend like Dark Academia.

In the Dark Academia movement as well, the process of genrefication begins with the initial migration from it being a cultural trend to an established literary genre. The initial days of the movement on Tumblr included visual content creation that was inspired by *The Secret History*. Eventually texts like The Picture of Dorian Gray (1890), Wuthering Heights (1847), Frankenstein (1818), Dracula (1897), Northanger Abbey (1817) and even the Harry Potter (1997-2007) series began to be added in lists that were labelled as must-read Dark Academia books. It is important to remember that these texts were not in themselves Dark Academia but rather hailed as belonging to this loosely termed idea of Dark Academia that was still developing. Along with the list of books there were aesthetic elements, mood boards, fancasts, playlists and fashion styles that made up the content of early posts made using this concept. Dark Academia came to reflect a blend of intellectualism, romanticization of academia, nostalgia for a vintage past, dark tones, Gothic architecture, supernatural undertones, tweed jackets/skirts and old libraries. The "dark" in dark academia stands for the members' slight penchant for the unusual and the illicit (Zirngast, 2021). The "academia" in it stands for a group of aesthetics that are connected to learning, be it through reading, writing or intensive research.

Examples of Dark Academia are a picture of an individual in a tweed jacket standing with an old copy of any classic and a fall background, or a typewriter on a desk that includes piles of books and ink stains on the paper. Dark Academia was still a digital movement by the time the COVID-19 pandemic of 2020 was happening and it was not yet established as a genre. There is a direct relationship between the 2020 pandemic and the increasing popularity of the Dark Academia

movement. The shift to a digitized lifestyle played a key role in a lot of online cultural trends from baking banana bread at home to a lot of pandemic related home videos. Among the areas that were affected by the pandemic, a vital one that had to adapt drastically was the field of academia. Schools, colleges and universities are the locus of academia as spaces that are not just the setting for education but imperative to the culture that enhances the learning process for students (Pai, 2001). Due to the lockdown and prohibition of large scale social physical gatherings, academic institutions had to shift to online learning and digital classrooms. This shift highlighted the importance of the academic culture and environment that was created by libraries, campuses and institutional clubs and organizations in facilitating and enhancing the pursuit of knowledge for students. There was a noticeable lack felt among the students in online learning which led to a yearning for this academic culture and environment which was fulfilled for many students globally by the Dark Academia culture.

During the pandemic, wooden or vintage desks with lots of books, notepad pages, pens and sticky notes in the colours of Dark Academia made themselves a part of the aesthetic for a lot of students. The aesthetics of Dark Academia encourage both students and people seeking knowledge to set up their own tables and study spaces in a similar manner. Overall, it encourages studying and seeking knowledge for its own sake. The emphasis on knowledge is a vital characteristic of Dark Academia that plays a role in creating a cultural context whereby reading books and seeking knowledge are trends. Arguably, there is a greater emphasis paid to humanities in Dark Academia, and as such students of humanities, especially literature and language end up being a major target audience. This is not to say that other fields are ignored or do not find representation in the Dark Academia communities, rather, numerous Dark Academia social media accounts and posts also have specially curated STEM content. Noteworthy is the fact that Dark Academia is not limited to just institutional academia. The Dark Academia movement has often been censured for being elitist and encouraging gate-keeping vis-a-vis propagating Ivy League universities and glorifying the idea of pursuing higher education. Dark Academia dismantles the elitism that academia, especially institutions uphold. The movement encourages the fact that anyone can pursue an academic interest. It frees the pursuit of knowledge from the chains of ivory towers. Anyone who has a connection to the internet can study anything that they want to as knowledge is available everywhere. There are millions of lectures and tutorials on any given topic on YouTube, and information on any topic can be read on online platforms too. It therefore adds to an online learning culture that celebrates the quest for knowledge from anywhere in the world thus enhancing its desirability as an online cultural movement.

All of these are the different threads of the movement that came together to create the very fabric of the culture of Dark Academia while it simultaneously made strides in the literary spaces. The visual and aesthetic subculture of Dark Academia that started on Tumblr migrated to platforms like Instagram and TikTok where hashtag #DarkAcademia started getting millions of views. Bookstagrammers and BookTok content creators began integrating literary and aesthetic

appreciation for novels that aligned with the Dark Academia movement and culture. *If We Were Villains* by M L Rio, published in 2017, belonged to the Dark Academia movement with a heavy focus on literary drama, morally ambiguous protagonists, Gothic elements, vintage fashion and a murder mystery. Incorporating the study of Shakespeare drama in a highly elite school for drama whereby characters are pushed to recreate the very drama that they learn in a mystical journey, *If We Were Villains* opened the doors for new novels to be published in the realm of Dark Academia and thus began its genrefication. It was soon followed by *Bunny* by Mona Awad published in 2019, *Ninth House* by Leigh Bardugo also in 2019 and then *The Atlas Six* by Olivie Blake in 2020. All three are hailed as belonging to a now emerging category on BookTube, Bookstagram and BookTok as not just aesthetics but the development of a genre. Publishers actively started marketing them as Dark Academia novels by highlighting the novels with the aesthetic, intellectual and Gothic characteristics among others of the cultural movement of Dark Academia. Readers began to immerse themselves in the Dark Academia worlds created in these books which reinforced the genrefication of Dark Academia by attracting fan driven content further blurring the lines between content creation, culture and publishing. R F Kuang's *Babel* was released with all the publicity of it belonging to the Dark Academia genre and it was joined by the sequels to two established milestone books in the now neo-nascent genre. *The Atlas Paradox* sequel to *The Atlas Six* by Olivie Blake came out in 2022 and *Hell Bent,* the sequel to *Ninth House* by Leigh Bardugo came out in 2023. Both of them were immediately popular and were much awaited in the Dark Academia circles. Dark Academia aesthetics that involved all the elements that made them a part of the movement's identity started gaining traction in the genre as well. Readers and members of the book communities on all the digital platforms create content and fan art promoting the books as belonging to the genre, thus adding to the genrefication of these texts. Authors too actively participate in these communities by reposting some of the fan content and responding to audience queries.

Olivie Blake on 31st January 2023 announced the third book in the Atlas series via her Instagram account where she revealed the title and the book cover, *The Atlas Complex* expected to release on 9th January 2024. An entire year in advance, the third book in this thrilling Dark Academia series was announced with 10.1k likes already with an immediate followup post linking the presale sign up for the book. With regular appearances on her otherwise busy social media, the third book kept having reposts or giveaways till its release date which then included book tours and special praise by established papers or lists. Similar to Olivie Blake, R K Kuang also uses social media to her advantage. She has highlighted the Dark Academia aesthetic in her own life and often links her work with her studies further entangling the lines between reality and fiction for fans that enjoy the reciprocal interconnectedness of the author-fan-work traid. A lot of other books and series have been added to the genre of Dark Academia ever since and in between these years as well. Initially, the fans were specific about the texts meeting all the requirements of the traits that made a text Dark Academia but eventually even one or two characteristics were

acceptable. As of December 2024 Dark Academia is a shelf/category on Goodreads that includes 9034 books and is clearly still growing on a regular basis.

The conundrum of whether the digital movement/culture filled the genre or vice versa can can be possibly looked at as a self-fulfilling loop that promotes both the aesthetic aspect of Dark Academia and the literary genre as well. They both inspire and popularize each other and the common thread that runs through them are digital platforms. The evolving relationship between social media and literature is possible because of the interactive nature of digital spaces and the individual's power over their own choices as well as social media marketing. Thus it can be seen that digital platforms are playing a vital role in the process of genrefication in English Literature. This study of Dark Academia is one such example where this process has been explored and analyzed. Dark Academia exemplifies how the aesthetic digital movement created a culture born and evolved through digital engagement which eventually became a literary genre. It is the aim for this paper to encourage the research among other such upcoming 'genres' that are born or evolving through other niche digital cultures and subcultures that can be seen on digital platforms. The limitation of such a study can possibly be that sometimes the books may be popular for a while and then disappear from the trending pages, but that does not mean that it did not contribute to the popular culture and the understanding of the relationship between popular culture and digital platforms that is evolving and growing constantly with digital progress and technological development. By understanding and examining the growing relationship between technological development and literature, researchers will be able to study the mechanisms that drive contemporary literary landscapes.

**Bibliography**


Ariyanayagam, Annabelle (2023-12-03). "TikTok's #BookTok trend 'literally changed my life'". *BBC*. https://www.bbc.com/news/uk-england-67555175

Brekelmans, Ashley (7 November 2017). "The Reality of BookTube: An Analysis of how BookTube Affects Marketing Within the Publishing Industry". *Thinkubato*r. Archived from the original on 23 March 2020. Retrieved 27 March 2020.

Caliandro, A., and J. Graham. "Studying Instagram Beyond Selfies." *Social Media + Society*, vol. 6, no. 2, 2020, https://doi.org/10.1177/2056305120924779.

Cheng, Y. "Opinion: 'Dark Academia' — An Alternative Take on Education and Privilege." *The Student Life*, 22 Apr. 2022, https://tsl.news/opinion-dark-academia-an-alternative-take-on-education-and privilege/#:~:text=Dark%20academia%20is%20an%20internet,which%20Hogwarts%20gained%20its%20inspiration.

Colyrad, K. W. "How A Mysterious Novel Went Viral On TikTok & Changed The Publishing Game". *Bustle*. February 21 2024 https://www.bustle.com/entertainment/tiktok-books-a-broken-blade-mystery

Daspin, Eileen (23 June 2016). "Shelfies, BookTube videos court readers: Publishers take a page out of social media playbook". *USA Today.* ProQuest 1799072364.

Garg, D. "The Sins of Dark Academia." *The McGill Daily*, 1 Sept. 2021, https://www.mcgilldaily.com/2021/09/the-sins-of-dark-academia/.

Gere, Charlie. *Digital Culture*. Reaktion Books, 2008.

Hlynka, Denis. "Educational Technology." *Educational Technology*, vol. 50, no. 6, Nov.-Dec. 2010, pp. 62–63, Educational Technology Publications, https://www.jstor.org/stable/44429884.

Hughes, D. "The Publishing Industry in a Digital Age." *Digital Marketing Institute*, 2020, https://digitalmarketinginstitute.com/blog/the-publishing-industry-in-a-digital-age.

León, C. "Meet the YouTube Stars Turning Viewers into Readers." *The New York Times*, 18 Aug. 2018, www.nytimes.com/2018/07/31/books/booktubers-youtube.html.

Mitchell, A. *The Impact of BookTube on Book Publishing: A Study of John Green's Looking for Alaska*. East Tennessee State University, 2021, https://dc.etsu.edu/honors/741.

Mukadam, A. "A Touch of the Picturesque." *Post45: Dark Academia*, 15 Mar. 2022, https://post45.org/2022/03/a-touch-of-the-picturesque/.

Murray, S. "Dark Academia: Bookishness, Readerly Self-Fashioning and the Digital Afterlife of Donna Tartt's *The Secret History*." *English Studies*, vol. 104, no. 2, 2023, pp. 347–364, https://doi.org/10.1080/0013838X.2023.2170596

Nolan, S., and A. Dane. "A Sharper Conversation: Book Publishers' Use of Social Media Marketing in the Age of the Algorithm." *Media International Australia*, vol. 168, no. 1, 2018, pp. 153–166, https://doi.org/10.1177/1329878X18783008.

Pai, S., and S. Adler. *Cultural Foundations of Education*. 2001, http://what-when-how.com/distance-learning/cross-cultural-issues-in-online-learning-distance-learning/.

Quiring, A. "What's Dark about Dark Academia?" *Avidly*, 31 Mar. 2021, https://avidly.lareviewofbooks.org/2021/03/31/whats-dark-about-dark-academia/.

Reyes, K. "Dark Academia: How the Genre Turned into a Subculture." *Bookstr*, 13 Sept. 2022, https://bookstr.com/article/dark-academia-how-the-genre-turned-into-a-subculture/#:~:text=The%20idea%20of%20dark%20academia,discussed%20gothic%20and%20dark%20literature.

Stowell, O., and M. Therieau. "Introduction." *Post45: Dark Academia*, 15 Mar. 2022, https://post45.org/2022/03/introduction-3/.

Taylor, A. *Secret Histories: The Rise of Dark Academia*. Bennington College, 2022.

Taylor, G. "Tweed Jackets and Class Consciousness." *Post45*, 2022, https://post45.org/2022/03/tweed-jackets-and-class-consciousness/.

Zirngast, L. "Everything You Need to Know about the 'Dark Academia Aesthetic' Trend." *L'Officiel*, 2021, https://www.lofficiel.at/en/pop-culture/dark-academia-aesthetic.

# CONSTRUCTING UTOPIAS: EDWARD SOJA'S THIRDSPACE AND THE TRANSHUMANIST VISION IN RAY KURZWEIL'S *THE SINGULARITY IS NEAR*

***Dr Vidya Hariharan***
***Assistant Professor of English***
***SIES College of Arts, Science and Commerce, Mumbai***

**Abstract:** Edward Soja proposes that space is not just physical but social, psychological, and imagined too. His concept of Thirdspace challenges traditional binary notions of space and suggests that identities are fluid and constructed within these spaces. AI scientist Ray Kurzweil predicts a future where technology merges with biology, possibly transcending physical and mental limitations in his 2005 book *The Singularity Is Near: When Humans Transcend Biology.* This blurring of boundaries between human and machine suggests the creation of a transhuman identity as well as a redefinition of identity and space. Kurzweil's vision of the "Singularity", a point where technological advancement accelerates exponentially, creates new imaginative geographies. It challenges existing boundaries of what is possible and reimagines human potential in ways that match with Soja's call for rethinking spatial practices. This paper aims to apply Soja's Thirdspace theory to Kurzweil's vision of a Utopian society in *The Singularity is Near* as Soja's Thirdspace allows for the construction of utopian visions through spatial praxis and Kurzweil's ideas suggest how technological progress leads to enhanced human abilities and identities, leading to a utopian society. The moral and ethical consequences of such a future will also be studied.

**Keywords:** Singularity, Thirdspace, Transhuman, Utopia.

> "The Singularity will allow us to transcend these limitations of our biological bodies and brains. We will gain power over our fates. Our mortality will be in our own hands. We will be able to live as long as we want (a subtly different statement from saying we will live forever). We will fully understand human thinking and will vastly extend and expand its reach. By the end of this century, the nonbiological portion of our intelligence will be trillions of trillions of times more powerful than unaided human intelligence" (Kurzweil 23).

This startling prediction made by Ray Kurzweil in his influential work *The Singularity is Near* published in 2004 was viewed as an apocalyptic statement by many thinkers at that time due to its audacity, but twenty years later some of the forecasts seem to be prophetic rather than audacious. In his 2024 book *The Singularity is Nearer*, written twenty years after the first one, Kurzweil points this out with real life examples. His post-humanistic vision and Edward Soja's concept of Thirdspace may appear to belong to completely different fields of discourse as *The Singularity is Near* delves into the future of artificial intelligence and technology, predicting a moment of immense change and transformation known as the "Singularity" and Thirdspace, as

articulated by Soja, is a theoretical framework within geography and cultural studies that challenges traditional understandings of space and spatiality. This paper studies Kurzweil's Utopian vision for the future of humanity from the perspective of Soja's construction of utopian visions through spatial praxis and the implications for the existence of a post-human entity in space, as such an entity will extend the definition and boundaries of what it means to be human.

Ray Kurzweil's statement, "…future machines will be human, even if they are not biological" (38) encapsulates a fundamental aspect of his vision in *The Singularity is Near*. This assertion reflects Kurzweil's belief in the transformative potential of artificial intelligence (AI) and other advanced technologies to enhance and potentially surpass human capabilities, without necessarily replicating biological forms.

At the heart of Kurzweil's perspective lies the concept of augmenting human abilities via technological advancements. He explores the possibilities of integrating human biology with cutting-edge technologies, including neural implants and genetic modifications, to boost cognitive functions, prolong life, and enhance the overall quality of existence. This outlook suggests a utopian scenario in which individuals surpass their existing biological constraints, attaining superior mental and physical capabilities. Kurzweil foresees a future characterized by rapid technological growth, resulting in unparalleled control over the physical environment. This encompasses progress in fields such as nanotechnology, biotechnology, and artificial intelligence, which hold the potential to eradicate scarcity, poverty, and disease. The prospect of plentiful resources and improved capabilities indicates a utopian society in which human needs are effortlessly satisfied. Kurzweil examines the potential of virtual realities and immersive experiences facilitated by cutting-edge technologies. He elaborates on how these virtual environments could offer tailored and enriching experiences, allowing individuals to explore new realms, collaborate globally, and create without limitations. This dimension of Kurzweil's vision implies a utopian landscape where individuals can transcend ordinary realities and have fulfilling experiences.

An important term which Kurzweil has coined is "knee of the curve", this concept of the "knee of the curve" refers to a critical juncture in the exponential growth of technological progress, particularly in the fields of information technology, artificial intelligence (AI), and biotechnology. This metaphor is employed to illustrate a juncture at which advancements in technology experience a significant acceleration, resulting in profound and potentially unpredictable transformations in human society and existence. Kurzweil posits that technological development, particularly in areas such as computing power and artificial intelligence, adheres to an exponential growth trajectory rather than a linear one. At first glance, progress may seem gradual and in increments; however, as technologies evolve and build upon one another, their capabilities expand at an exponential rate. This rapid growth can lead to advancements that swiftly surpass human intuition and expectations. The term "knee of the curve" denotes an inflection point on this exponential growth trajectory, indicating the moment when technological capabilities attain

a level that triggers an increasingly rapid acceleration. This acceleration is driven by various factors, including the emergence of new tools, algorithms, and breakthroughs that foster further advancements in a continuous cycle of innovation. As technological progress approaches the "knee of the curve," Kurzweil projects significant implications for society. The swift advancement of artificial intelligence, for instance, may result in superintelligent systems adept at solving complicated problems, enhancing human cognition, and potentially transforming the nature of work, communication, and social frameworks. Similarly, progress in biotechnology could revolutionize healthcare, prolonging human lifespan and improving quality of life. The "knee of the curve" represents a crucial moment in the exponential growth of technology, where advancements accelerate significantly and pave the way for transformative shifts in human civilization.

Similar to Kurzweil's idea is Soja's 'trialectics', a spatial concept which is a combination of First space and Second space, and it creates a new experience called Thirdspace. Firstspace is real and physical; it is socially produced as channels and products of human activity. The physical reality of Firstspace is arranged and understood through imagined representations, in life lived in Secondspace. We delineate this space in our thoughts and feelings and express it in symbolic language. This is regarded as conceived or ideated space and power can be said to be located here. According to Soja, Thirdspace is real and imagined space at the same time. It is erected through the social practices of people who live it. It is directly lived space, as also, the space of resistance and renewal. So, in Soja's words, "Thirdspace is a way of interpreting socially produced space, where our human geography has the same scope as the historical and social dimensions" (Soja 1). As Thirdspace is space and spatiality people constructed through social practices, it inheres the idea of a fluid concept of space and thus a fluidity of identity. According to Soja, the notion "Third" refers to the constructing and re-constructing of identity, to the fluidity of spaces, to the space where identity is not fixed, and such is the feature of any kind of spatial existence (Soja 8). He conceptualized a nation, city or locality which allowed freedom from all binaries and representations of power, where class, gender and such other considerations are not important. Thirdspace disrupts and creatively restructures consciousness to give way to alternative understandings of the self. Restructuring binaries does not serve to negate the existence of such binaries. Rather, Thirdspace identity constructions describe sites that do justice to the complexity of the self by giving space for the existence of binaries as well as third/other choices.

The similarity between Kurzweil's Singularity and Thirdspace lies in their shared emphasis on transformation and the reconfiguration of boundaries. Kurzweil predicts that the Singularity will fundamentally alter human capabilities and perceptions. This transformation extends beyond mere technological progress; it implies a profound shift in the human condition, potentially blurring the lines between biology and technology.

> All the machines we have met to date lack the essential subtlety of human biological qualities. Although the Singularity has many faces, its most important implication is this: our technology will match and then vastly exceed the refinement and suppleness of what we regard as the best of human traits (Kurzweil 23).

Similarly, Soja's Thirdspace challenges conventional spatial categories by advocating for a perspective that transcends binary oppositions such as real/imaginary, physical/mental, or local/global. This reconfiguration of spatial understanding parallels Kurzweil's vision of a future where the boundaries between humans and machines, and possibly between reality and virtuality, become increasingly fluid.

Kurzweil and Soja both advocate for a shift towards a more interconnected and integrated way of living. Kurzweil's notion of the Singularity envisions a future where technology significantly enhances human abilities, potentially improving communication, cognitive functions, and creativity to extraordinary levels. This perspective resonates with Soja's concept of Thirdspace, which emphasizes a spatial awareness that recognizes the intricate web of interconnections and relationships. Both viewpoints challenge the idea of a singular, linear progression, promoting instead a multidimensional approach to understanding transformation and evolution. Kurzweil and Soja stimulate contemplation regarding the implications of their theories for human identity and agency. Kurzweil posits that the Singularity could result in the convergence of human and artificial intelligence, prompting inquiries into the essence of humanity in a technology-driven world. Meanwhile, Soja's Thirdspace encourages a re-evaluation of identity within spatial frameworks, urging individuals to navigate and reconcile various dimensions of space and place. The two theoretical frameworks encourage a reconsideration of conventional boundaries and definitions, fostering a more sophisticated understanding of human existence in relation to technology and spatial dynamics.

Kurzweil's vision of the Singularity anticipates a future where technology and humanity merge in unprecedented ways, echoing Soja's call for a spatial consciousness that transcends traditional binaries.

> Biology has inherent limitations. For example, every living organism must be built from proteins that are folded from one-dimensional strings of amino acids. Protein-based mechanisms are lacking in strength and speed. We will be able to reengineer all of the organs and systems in our biological bodies and brains to be vastly more capable (Kurzweil 36).

This vision of the human capability assumes a human body which is far superior to the present one, and an identity less biological, more tech-infused. Kurzweil forecasts the development of artificial intelligence systems that will surpass human intelligence. He asserts that these systems will not only replicate but also exceed human cognitive functions across various domains, such as problem-solving, creativity, and decision-making. Although Kurzweil acknowledges the

possibility of unemployment in some areas, he argues that artificial intelligence (AI) would also create new job possibilities and boost economic growth by increasing productivity and efficiency. Ensuring a smooth transition for workers affected by automation and advancing education and training programs for developing AI-centric industries will be the main challenges. AI is also anticipated to stimulate human creativity by generating novel concepts, creative works of art, and scientific discoveries. Kurzweil proposes that artificial intelligence (AI) systems could help academics, writers, and artists push the boundaries of human creativity and knowledge while they explore ideas and creative solutions to challenging issues. According to Kurzweil, future machines particularly artificial intelligence systems will either equal or exceed human intelligence and cognitive capability. The constraints imposed by human evolution, physical weakness, or susceptibility to disease are not going to be an obstacle for these technologies. These are limitations inherent in biological beings. Instead, they will use extensive data processing skills, complex algorithms, and computing power to carry out jobs and make choices with previously unheard-of accuracy and efficiency.

Kurzweil predicts that, these artificial intelligence machines will greatly advance human potential, particularly in areas like pattern recognition, problem solving and knowledge retrieval and decision making. Moreover, artificial intelligence systems have the capacity to operate nonstop without a break, which could lead to notable increases in output and creativity in a variety of industries. Despite not being biological, Kurzweil believes that these systems are fundamental to the human experience. They are anticipated to support people with their everyday tasks, improve cognitive abilities through brain interfaces or other human machine connection, and encourage creative approaches to cooperation and communication. AI's integration into many facets of human life may trigger significant changes in how society functions and how individuals engage with technology. This claim raises ethical questions about the nature of awareness, intelligence, and the distinction between humans and computers. The increasing complexity of AI systems prompts questions about their impact on human identity and autonomy, and the risks associated with over-reliance on such technology. In order to make sure that the integration of AI upholds fundamental human values and favourably impacts human welfare, it is necessary that these ethical problems be addressed.

Edward Soja's Firstspace, Secondspace, and Thirdspace framework provides an insightful lens through which to examine spatiality and human experiences in relation to various geographic and cultural contexts. Firstspace, which represents the conventional understanding of space in geographical terms, is concerned with the tangible and physical surroundings. Firstspace, in Kurzweil's vision, can be seen as the real world and places where the impacts of technological progress are visible. This encompasses the places where new technologies are developed, put into practice, and interact with ecosystems and human cultures. Soja defines secondspace as the imagined and represented spaces. It includes the various ways that cultural, social, and symbolic viewpoints shape how spaces are understood, experienced, and perceived. Regarding *The Singularity is Near*, Secondspace could refer to digital surroundings, replicated settings, and

virtual realities made possible by cutting edge technologies like artificial intelligence (AI), augmented reality (AR), and virtual reality (VR). These areas present fresh perspectives on interaction and experience that go against traditional notions of human perception and physical space. Soja states that, Thirdspace is a space that is both hybrid and transformative, resulting from the interaction between Firstspace and Secondspace. It symbolizes an environment of possibility, encounter, and resistance, a place where fresh interpretations, identities, and social structures are always being created and challenged. According to Kurzweil's theory, Thirdspace is the emerging realities and experiences that arise when digital, virtual, and augmented spaces combine with physical environments. This covers the merging of artificial and human intellect, the technology that permeates daily life, and the changing relationship between people and the technology they create.

Regarding the possible benefits of technology developments, such as enhanced human capacities and the eradication of diseases, Kurzweil presents a positive outlook. But his viewpoint also raises significant moral conundrums. The possibility of widening socioeconomic gaps is a serious ethical concern. Although Kurzweil emphasizes how technology may democratize society, he also acknowledges that wealthy people or countries may initially be the only ones with access to cutting-edge inventions. This restriction may widen the gap between those who have access to technology and those who do not, exacerbating already-existing social and economic disparities. As technology develops, particularly in the fields of biotechnology and artificial intelligence, worries about individual autonomy and the capacity to control one's identity and choices surface. In his futuristic ideas of human technology hybridity, Kurzweil raises concerns around data ownership, privacy, and the possibility of manipulation or coercion by powerful organizations, be they governmental or corporate. Risk reduction requires addressing concerns including AI bias, accountability, transparency, and the unexpected consequences of AI actions. Kurzweil investigates how technologies like neural implants, genetic engineering, and nanotechnology can improve human potential. There are ethical issues with fair distribution of these improvements, unanticipated consequences for ecosystems and human health, and the social implications of differences in ability between people who have been boosted and those who have not. Although Kurzweil is optimistic about the advancement of technology, he also recognizes the possibility of unanticipated hazards and unintended consequences. In order to reduce dangers and maximize benefits, ethical frameworks should place a strong emphasis on taking preventative actions, exercising ethical foresight, and regularly evaluating how emerging technologies will affect society. According to the idea of the Singularity, everyone would have instant access to information and knowledge in a highly connected society. In his book, Kurzweil explores the possibility of an international network that would enable easy communication, teamwork, and idea exchange across national and cultural barriers. A utopian picture of a society where obstacles to information and understanding are reduced and collaboration and communal advancement are encouraged is fostered by this connection.

> Nanotechnology will enable the design of nanobots: robots designed at the molecular level, measured in microns (millionths of a meter), such as 'respirocytes' (mechanical red-blood cells). Nanobots will have myriad roles within the human body, including reversing human aging (to the extent that this task will not already have been completed through biotechnology, such as genetic engineering) (Kurzweil 37).

These robots are intended to be more than just tools or helpers; they are supposed to be sentient beings with intellect and functionality that will make it difficult to identify the difference between artificial and human intelligence. Although Kurzweil's viewpoint is positive regarding the possible advantages of cutting-edge technologies, it also emphasizes the significance of careful thought and moral foresight in creating a future in which artificial intelligence and humans may coexist peacefully and morally. Through an application of Soja's spatial concepts to Kurzweil's ideas in *The Singularity is Near*, we can investigate how technological advancements are generating complex, hybrid spaces -Thirdspace- where boundaries between digital and physical, artificial and natural, and human and machine are becoming increasingly hazy, in addition to reshaping physical environments -Firstspace- and creating new virtual realms -Secondspace. This viewpoint emphasizes how technology has the power to fundamentally alter human interactions, spatial experiences, and societal institutions in ways that both challenge and broaden our comprehension of space and human life in an increasingly technologically mediated future.

An important aspect of Kurzweil's outlook is the dominant influence of virtual reality (VR) which will revolutionize technology by enabling people to immerse themselves in virtual worlds with unbelievable levels of realism. By allowing users to interact with virtual environments, take part in simulated situations, and access information in novel ways, virtual reality (VR) has the potential to expand human sensory experiences beyond geographical limitations. This fits with Kurzweil's overarching goal of using technology to improve human capacities seamlessly. In "The Singularity is Near," Ray Kurzweil addresses the idea that artificial simulations will eventually blend seamlessly with the actual world. Virtual reality (VR) is essential to this goal because it offers platforms for the creation and experience of simulations, enabling a wide range of applications, from immersive entertainment and therapeutics to virtual meetings and instructional simulations. These models may help with experimentation, prediction, and decision-making in a number of different domains.

By establishing shared virtual spaces where people from different locations can interact in real-time, virtual reality technology can improve cooperation and communication. According to Kurzweil, virtual reality (VR) will facilitate improved communication and teamwork between academics, professionals, and educators. This will promote creativity and quicken the exchange of knowledge within international networks. In his investigation of virtual reality, Kurzweil touches on issues of identity and self-perception. People may be able to create and inhabit avatars, change their appearances, and explore a variety of identities in virtual settings as VR technology develop. This brings up moral questions about identity, sincerity, and the fuzziness of lines

between the actual and virtual worlds. Kurzweil talks about how VR and other technologies could work together to create even more intelligent and immersive virtual worlds. By offering tailored recommendations, flexible settings, and receptive interactions, AI-driven virtual assistants in virtual reality have the potential to improve user experiences and further obfuscate the distinction between real-world and virtual realities. He even goes so far as to say, "In virtual reality, we can be a different person both physically and emotionally. In fact, other people (such as your romantic partner) will be able to select a different body for you than you might select for yourself (and vice versa)" (38). Beyond a consideration of the ethical implications of a posthumanist identity, there are several perspectives from which this vision can be criticised.

Kurzweil's post humanist vision can be criticised for its homogenizing effect. Sandy Burnley perceives a posthuman attitude towards humans and nonhuman beings in Lewis Carroll's *Alice in Wonderland* and points out that "Wonderland underscores the caterpillar's unique and posthumanist identity, signals the gardeners' attempts at eradicating this species because of misidentification, and admonishes readers against transforming Wonderland into a homogenous nation of sympathetic humanoids" (Burnley 107).

The economic imbalances and underdevelopment in parts of the world could further the binaries of First World nations and Third World countries. The developed nations would maintain their hegemony by placing constraints on knowledge and resources. In "What Space makes of us" Rick Allen discusses this possibility of imbalance using Lefevre's trialectics of perceived, conceived and lived space. "Difference in both cases is feared and invokes spatial metaphors with negatively viewed meanings such as fragmentation. Conceived space once again dominates lived space, and the expressed subjugation in the form of multiple or fragmented identities is attacked" (Allan 24).

Both Soja and Nidesh Lawtoo fear the hyperreal effect of virtual reality on the susceptibility of humans. In "Posthuman Mimesis I: Concepts for the Mimetic Turn" Nidesh Lawtoo, shares his concerns regarding what he calls this "hypermimesis", "Once AI is programmed via increasingly sophisticated algorithms to feed individual users with their posthuman preferences online, powerful bubbles are generated that not only create alternative worlds behind this world that render truth already post; they also generate affective reactions and bodily actions that determine, in very practical, material, and potentially devastating terms, the all-too-human behaviour of homo mimeticus offline" (Lawtoo 112).

To summarise, posthumanism offers a nuanced interaction of negative and beneficial consequences that significantly influence our perception of humanity and its destiny. On the one hand, it casts doubt on established ideas about ethics, identity, and the interaction between people and technology, which could give rise to ethical, moral and existential concerns. In an article published in the Journal of Healthcare Engineering, titled "New Challenges for Ethics: The Social Impact of Posthumanism, Robots, and Artificial Intelligence", Lourdes Velasques raises extremely pertinent concerns about what she calls the "deforestation of humans" to the

degeneration of the "digital revolution." Currently, man seems to have embarked on a new path towards a goal which was unimaginable only a few years ago: to become a *machine*. In fact, there are supporters of artificial intelligence who predict a future in which men and machines will merge into cyborgs: a real anthropological change, not only culturally but also in the understanding of man, nature, life, and the world (Velasquez). Conversely, posthumanism also creates novel opportunities to advance human potential, promote diversity, and reframe our relationship with the natural world and non-human species.

**Bibliography**


Allen, Rick. "What Space Makes of Us: Thirdspace, Identity Politics, and Multiculturalism." Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, 24–28 Mar. 1997. ERIC, https://files.eric.ed.gov/fulltext/ED409409.pdf.

Burnley, Sandy M. "Looking Back: Posthumanism and Sympathy in *Alice's Adventures in Wonderland*." *Victorian Review*, vol. 48, no. 1, 2022, pp. 107–23.

Kurzweil, Ray. *The Singularity Is Near: When Humans Transcend Biology*. Viking, 2005.

Soja, Edward. *Thirdspace: Journeys to Los Angeles and Other Real-and-Imagined Places*. Blackwell Publishers, 1996.

Soja, Edward. "There's No Space Like Home: From Bachelard to Soja and All the Spaces in Between." *2113 HUM There's No Space Like Home*, https://2113humtheresnospacelikehome.wordpress.com.

Velázquez, Lourdes G. "New Challenges for Ethics: The Social Impact of Posthumanism, Robots, and Artificial Intelligence." *Journal of Healthcare Engineering*, vol. 2021, 2021, Article ID 5593467, https://doi.org/10.1155/2021/5593467.

# DIGITAL MANIPULATION AS TEXT: THE PROBLEM OF DISINFORMATION IN DIGITAL CONTENT

*Dr Chitra Adkar*
*Assistant Professor of Sociology*
*Sophia College for Women (Empowered Autonomous), Mumbai*

**Abstract:** Digital manipulation and questions about its detectability lie at the core of the problem regarding content on the internet in the age of artificial intelligence. This paper will focus on the central issues around digitally manipulated content on the internet. It would begin with a brief retrospective on the topic of editing and the level of manipulation that has been acceptable in different media as an inherent component, without which the message would not be received in the way that the creator intends. The discussion would examine 'edits' among other forms of humour on the internet as socially acceptable forms of digital manipulation used on the internet as opposed to misinformation and disinformation that also use similar AI tools. 'Edits', which are edited videos comprising two or more distinct cultural artefacts or components coming from distinct cultural locations that are conjoined through the process of editing, would be analysed to understand how digital content hinges upon manipulation.

**Keywords:** digital manipulation, edits, humour, disinformation, artificial intelligence

Social media platforms have created a unique virtual space, and researchers in digital sociology and media studies, as well as the practitioners of the methods in digital ethnography, have started envisioning this space as a research site. The basic reading of this space, its consumers, and engagement patterns redirect our attention to artificial intelligence (AI) technologies used for monitoring online content. This paper will specifically analyze the evolution of edits foregrounded in the debates between formalism and realism to understand digital manipulation on the internet. The arguments around authorial intent in literary criticism would be brought in to understand the same. The discussion would attempt to understand how social media platforms using machine-learning technology to curb disinformation deal with this digital manipulation.

In June 2020, an Instagram user posted a mashup[1] of a video of a Korean band called BTS practising their dance moves for the track 'Boy with Luv'[2] and the Bollywood song called 'Chunari Chunari'[3]. Mashups were not a new format in 2020; several content creators had been creating edits, including mashups featuring altered dialogues or video editing involving deepfake technologies for a while. Visuals, like those in this viral mashup, are not alien to anyone who engages in online behaviours regularly; for instance, one may not be a BTS fan, or one may

[1] See https://www.youtube.com/watch?v=XZu9jdGKqkQ (Kpop world, 2021)
[2] See https://www.youtube.com/watch?v=CzvfbRbEjww
[3] See https://www.youtube.com/watch?v=dRLwMAGMJnQ

not follow pages and groups meant for BTS fans, but owing to its popularity, the videos of these bands are everywhere. They may pop on someone's timeline even if none of their friends or followers share them; this could be attributed to algorithms used by social media platforms. An edit format like a mashup primarily involves two components, the audio and visual, and both come from different sources. One might perceive any mashup in several ways; consider the 'BTS x Chunari Chunari' as an example. An Indian person without awareness of a band called BTS would still be amused at seeing a bunch of Korean men dancing to a Bollywood song. A seasoned BTS fan would also be amused for the exact same reason, and mostly the reason would be, as mentioned earlier, quite uncomplicated; it is funny because it is unexpected.

A video edit, or 'edit' as it is commonly referred to, is not a juxtaposition of a video with a different audio track or altered imagery; it is the discrepancy in the auditory and the visual that defines its form. Edits have been popular in the last decade, and their formats keep updating themselves with the mixing of several media, despite the fact that they remained succinct and brief. What also remains common in changing forms of edits and other forms of internet humour is digital manipulation. The realism versus formalism debate in the theory of filmmaking concerning what both schools had to say about an image and its author would be relevant in discussing them to analyze digital manipulation in depth. Relevant arguments on the issue of authorship and intention will be brought in to draw parallels between content creators on social media and their role in ascribing meaning to their visual forms of expression. The paper will also briefly discuss Rosetta, a machine-learning program Facebook uses for image detection and recognition.

**Theorizing Editing & Edits**

Andre Bazin's essay 'The ontology of the photographic image' has become a canon for theorists studying film and media as it provides several arguments that bring forth the reasons behind the advent of modernism in arts. Bazin's commentary makes poignant references that set forth a tone of attributing intention and extracting the essence of the photography process, and his inferences go beyond the scope of film theory. His basic argument rests upon the tendency of an artist to preserve an object that is the subject of the artistic expression, like a painting, and how it is similar to embalming corpses. He argues that to preserve one object, another object enters the process, and through that, one creates a photograph, which is supposed to carry an essence of the reality of that object (Bazin). The introduction of the photographic image and technology to create photographic and videographic material left a mark upon movements in the world of art; if creating a realistic image of a person in the form of a portrait would be an important skill in a world with no technology to photograph objects, then the new world with the camera would attribute lesser importance to it.

Later, the realism vs formalism debates in various schools of film-making also set off different debates that gave rise to distinct ways of looking at the form. While realism called for bringing in

cinematic technology for showcasing the world as it is formalism, considered the very form of cinema to be something that needs to be put together. The latter gave rise to the idea of montage popularized by stalwarts like Sergei Eisenstein. Formalism would also focus on showcasing reality but not with an approach wherein the camera and videography equipment were considered sufficient in recreating the world as the filmmaker wants his audience to see it. Instead, the juxtaposition of several elements of reality as per the vision of that director was put together even when, in reality, those elements wouldn't be consequent[i]. These changes can be attributed to the power of the camera as Bazin perceived it; the presence of another object itself paved the way for several possibilities of the form and allowed filmmakers more possibilities than simply viewing videography as a technology to record what is observed (Bazin).

This brings one to the finer nuances of arguments about visual forms of expression. In the process of understanding photographing in a nuanced manner C S Peirce's three categories of icon, symbol and index have been considered to be important in literary theory (Johansen). Peirce argued that if the photographic image is iconic, it bears a physical resemblance to the real thing; if it is symbolic, then there exists a relationship between the signifier and the signified based upon culturally learned connections; and if it has indexicality, it shows evidence of the existence of something (Johansen). These three parameters are important to distinguish the difference between the digital and the pre-digital understanding of photography and, consequentially, the interpretation of visual media. Several theorists in media studies harp upon the argument that digital and pre-digital photography are distinct from one another and should be studied accordingly because digital photography lacks indexicality. At this juncture, understanding the medium as an object, through Bazin's framework and beyond his argument, is important. One could argue that analog or pre-digital photography is a play of light and produces a photochemical effect that can show the objects in their 'most real form', whereas digital photography involves a series of coding, and the process itself can be altered to produce enhanced pictures, thus taking away the 'indexicality' from the photograph.

If one attempts to break down the visuals in an edit as icon, symbol and index (Johansen), a layer-by-layer peeling of multiple media is necessary. Are the visuals in an edit iconic? The visuals used for creating the edit are iconic, but once we make reference to an edit created to incite laughter, there may not be a 'physical resemblance' to something. For instance, an edit generated using AI technology might morph one's face onto someone else's body, and while this body may resemble the actual person, this morphed visual only exists in the edit and outside of it, there is a truth claim being made for its existence. For example, it is popular to compile song lyrics from individual words spoken by celebrities and politicians that are auto-tuned to sound like a song. A YouTube user account called Maestro Zikos has created an edit of Donald Trump and Barack Obama singing the track '*I am a Barbie gir*l'[4]. Most of the words and sometimes letters put together

[4] See https://www.youtube.com/watch?v=hLNy6T3KkEc

to create the sound of a word in such edits are actual captures of those people saying the same things. However, the format implies that there is no truth claim; the viewer knows Trump & Obama have not actually sung those tracks. This, in turn, implies that there while analyzing edits there is no sense to pursue such semiotic analysis further as the relationship between the signifier and signified changes. This is the reason why one can see continuity and a departure in what edits are meant to be if placed in a history of editing. In one way, there is continuity as the form means more than what the deconstructed pieces of the visual are meant to be. This is no different from any movement in the visual arts that steps further away from formalism or, in simpler words, allows one to argue that just because it includes a video does not mean one is looking at just that in isolation. Instead, that video can be a part of an intentionally manipulated piece of information that is to be read with knowledge of how it should be read. One can also argue that there is somewhat of a departure with the arrival of such formats and templates online, as the theoretical possibility of generating good quality deepfakes that are unrecognizable without software usage would always exist. If this discussion were about all edits, including those beyond humour, there is much to uncover.

If digitally manipulated content is seen as text, the author-over-critic and critic-over-author debates would further complicate one's reading of edits and their inherent manipulation. Roland Barthes takes what is referred to as the anti-intentionalist position by arguing that 'the image of literature to be found in ordinary culture is tyrannically centered on the author, his life, his person, his tastes' (Barthes). Beardsley & Wimsatt, in their essay, 'The Intentional Fallacy,' have stressed the importance of the text itself and argue that the logic of 'author over the critic' or the 'critic over the author' is fallacious (Wimsatt and Beardsley). Instead, one should limit themselves to the evidence found in the text itself to examine a work of art, like a piece of literature. They reject the belief that a text means what an author intended it to mean. This school of thought can be called formalism in literary criticism, specifically new criticism.

These complexities in textual analysis are important from the point of view of recognizing who the actors are in the process of authoring texts and how one should perceive other stakeholders whose reading of the text can make a difference in how society, at large, interprets any artistic expression. Whether one sees a content creator, an editor, as author or a critic, the dynamic between various parties who are stakeholders in regulating online content gets complicated when multiple claims of authorship are made. For example, when parties owning audio tracks or visuals used in edits move for a copyright claim, how does one defend authorship of the edits? The authorial intent of every party involved in addition to the content creator (in this case, an editor) in such a situation would be rife with conflicting interests and an imbalance in power. Furthermore, when the authorial intent is sought and needs interpreting by the consumers of content, one needs to acknowledge that what is deemed to be acceptable, offensive, and effective and what qualifies as humour becomes extremely subjective. There is hardly any room for critic-over-author or author-over-critic debate (Gunning) in cases where copyright strikes lead to

content being taken down. It also becomes irrelevant when pressure tactics are used to make a content creator take down their posts. Therefore, Wimsatt & Beardsley's scheme of finding the authorial intention within the text would be an effective way of pursuing analysis (Wimsatt and Beardsley).

**When AI Monitors Digital Manipulation**

Pertaining to the question of digitally manipulated content on the internet, one must note that the categories of humorous content and disinformation often have blurred boundaries. While memes mostly look like memes, sometimes one might stumble upon a heavily sarcastic statement driven by the usage of a meme. Who, what, and how gets to lay their truth claims regarding such content, and when AI is involved, who gets to uphold their right to freedom of expression cannot be determined simplistically. One could also take examples of edits involving technology used in creating deepfakes. Any editing technology that enables a person to bring together two distinct cultural artefacts to create digital content can range from the simplest techniques in Photoshop to more complex techniques; in other words, one could create shallow fakes or deepfakes. The role of AI-based systems used by social media platforms would be limited. One could argue that AI can only determine the veracity of posts on these platforms. Usually this does not involve complex algorithms that can deeply understand humour, especially satire found in humorous digital content.

Rosetta is a large-scale machine learning system used by Facebook since 2018 and is regarded as a development of the previous optical character recognition (OCR) technologies deployed by several media websites (Sivakumar, Gordo and Paluri). Rosetta focuses on text detection and text recognition to study memes to fulfil goals about censorship of certain material shared on Facebook. It recognizes text from 'rectangular regions in an image that could contain text and uses a convolutional neural network (CNN) to recognize and transcribe the text that could be found there' (Sivakumar, Gordo and Paluri). To 'train' this machine learning system, Facebook uses a database containing publicly available images on Facebook itself (and now the entire Metaverse) that formally resembles images containing text, like memes. The programming also accommodates certain differences that could eliminate analysis of some material; for instance, Rosetta is trained to read Arabic text backwards and thus can analyze it (Sivakumar, Gordo and Paluri). However, the pool of resources to recognize what data would be useful for training such systems is by itself very fluctuating.

One must note that there are closed groups on social media platforms that do not post content in the public domain. Melmagazine studied how Facebook users create private groups and what their systems are to keep out users who do not believe in conspiracy theories about COVID-19[5]. Memes have been used widely to put forth arguments on those lines and despite its best efforts,

[5] See: https://melmagazine.com/en-us/story/covid-coronavirus-truth-meme-page-facebook

Facebook was unsuccessful in curbing misinformation to the extent it should have. If one studies memes as text and chooses to find meaning within the context itself, every expression of a meme creator or an author would be admissible. While this is not an issue, the development of AI to seek certain forms of content as potentially harmful makes no sense. Every article, post, image, video or other form of expression on social media that contains a word like 'covid' or 'corona' is subject to surveillance. Although it was widely used by conspiracy theorists, humour is not primarily meant for conspiracy theorists alone. Authentic information can also be spread through memes and other forms of humour on social media, even though they use digital manipulation.

Furthermore, one needs to bear in mind that content on the internet, especially internet-mediated forms like memes, edits, mashups, shorts, reels, etc, require manipulation. The argument of what is real is authentic and any other form that isn't so is manipulation does not hold much relevance. In a recent episode of social media-based altercations, a Bollywood actress, Kangana Ranaut, fell for a spoof video, which was an edit featuring the CEO of Qatar Airways (Desk). An Indian Twitter user had appealed to his fellow countrymen to boycott Qatar Airways for the stand the country took in an incident involving hate speech used by ex-BJP spokesperson Nupur Sharma (Report). The spoof video showed an old visual from a recording of an interview with the CEO of Qatar Airways with an audio track where he is appealing to Twitter users to go back on his decision to boycott the airlines. The actress failed to recognize this as a humourous edit and called out the video for 'bullying' the Twitter user. What transpired thereafter was interesting. Several social media users who had shared the edited video as a joke could do so for a day or more after which their posts got deleted. Despite Ranaut's failure to recognize humour, which was quite public in nature, Facebook was flagging the spoof as false information in some places, thus restricting its visibility; the author's own Facebook wall showed the same restricted visibility for this post[6]. If one wants to view this situation as a back-and-forth between an author and a critic, in terms of literary criticism, enlisting every author involved in the process is quite the task. There is a video editor who created the spoof, then there are thousands of users who shared the spoof as they understood and accepted that this was humour; there is also a subject in the video, there is a videographer, and there is a voicing artist who created the manipulated audio track. Of all these actors, the video editor and the voicing artists can undoubtedly be considered to be authors as far as the edit is concerned, while those who shared the video can be readers or interpreters of

[6] The current status of one of the posts featuring the video, that hasn't been taken down, comes with instructions about fact-checking towards the end of the post. See: https://www.facebook.com/bhavnamisr/posts/pfbid0rHEfZjooRkk6copAPWwSGEjonhaAp436Yw6qTPtAjpr8eDDFnYNwHCiBTLaHr1Ubl for a visible version with the warning and see https://www.facebook.com/chitsy.adkar/posts/pfbid0eSCJLkmd49HoSACiVaeTQSNkjWto72yb2W3JVWushjXvYofvpu4hCHPN9sSR9QAnl for the version where the post has lost public visibility.

this text. How would one then analyse the role of Rosetta or any machine learning systems that can potentially stop the sharing of this cultural artefact?

One could boil it down to the fundamental question: is Rosetta a critic of the text, or does another critic teach Rosetta? In addition to that, can one compare Bazin's interpretation of the camera as an object that authentically preserves another object to machine-learning systems that get a say in determining authenticity? These questions are difficult to answer, yet they are important for developing a deeper understanding of the 'form' in the digital era. Before addressing these questions, one should bear in mind that the evolution of visuality in various forms of expression has been a process longer than the last decade that witnessed several changes due to the emergence of Web 2.0, and it shouldn't be oversimplified. In other words, there is no need to seek a point in time that marks the distinction between the digital and the pre-digital, as well as the online and the offline, as far as this exercise in the textual analysis is concerned.

To tackle the first question that identifies who is a critic of a meme as a text, one should begin by acknowledging that every text has multiple critics and, therefore, multiple readings. People find themselves pledging alliances with either the author or a critic of a text and they do so knowing that every text always has multiple readings and interpretations by multiple critics. Perceiving a meme, an edit, or any piece of humourous content online as text is fundamentally rooted in accepting multiple interpretations. What makes Rosetta the ultimate reader who decides if content stays on Facebook or is deleted is germane to this analysis. Rosetta is owned by the company Facebook, which enables the very process of mediation every time one posts, likes, shares, or comments on any content found on the social media platform. Despite considering a machine learning system as an entity, one cannot compare it with a human reader of any form of text. At the same time, this system is designed by and frequently updated by human engineers. These engineers and their individual readings of a text, for instance, their opinion about a funny video or their take on the COVID-19 pandemic, are in no definitive of what Rosetta learns for curbing information and hate speech.

By borrowing from Wimsatt and Beardsley's arguments, if one decides to locate evidence of what exactly a text should mean within the text (Wimsatt and Beardsley), Rosetta would be 'tyrannical' (Barthes) without a doubt in Barthes' eyes. One often loses the crux of such arguments by accommodating technological advances that censor speech by regulating content as a step for the greater good. While that may be true in some cases, it makes space for absurdities; one could revisit the example of the flagged spoof video of Qatar Airways CEO and its restricted visibility. There is enough evidence on the internet that would leave no room for an alternate interpretation, one can infer with certainty that the video is indeed a spoof. Yet the algorithms that teach Rosetta fail to consider this evidence; at times, the video is visible with a warning, and in other places, it has been taken down.

**Analysis and Conclusions**

Two important questions could be raised while looking at the processes involved in regulating digital manipulation: whether AI technologies would remain effective and whether people need to seek mechanisms beyond AI that regulate online content. As far as the first question is concerned, one could safely say that AI technologies meant for the regulation of digital content will always lag because users of social media platforms and creators of AI technologies will always have conflicting interests and means of circumventing the present mechanisms in place that don't let either party create what they want. For example, in the case of deepfakes, Rob Toews has spoken of a 'game of technological cat-and-mouse' where the growing industry of digital forensics involves the development of software that creates better deepfakes alongside the development of software that creates better programs for detecting the same deepfakes (Toews). The presence of human actors determining what AI does and what it is supposed to do is more than the question of technology and its maker. The very presence of technologies like these has become a reason that involved human actors often use as an excuse for wrongly regulated or under-regulated digital content. Claire Ward warns of a phenomenon called the 'liar's dividend' (Chesney & Citron, 2019: 1753) that describes misinformation being cited as a reason to neglect genuine information because of the abundance of misinformation that is found on the internet (Warde).

Regarding the second question, one needs to see the virtual space for what it is: intangible, not a part of material reality, but very real in terms of the interactions of real actors that it hosts and its potential consequences. The virtual space is primarily visual, compressing a host of information onto a screen through its websites, applications and programs. The visual form has long surpassed the test of authenticity through the idea of reproduction. Instead, any interface on any site in this virtual space must pass the test of clarity and should be user-friendly which can only be achieved through its form. One can take the example of a satirical website like *theonion.com, which* has the look of a regular news website; it is not full of caricatures. Instead, it uses actual pictures of public figures and politicians, and even the headlines of the satirical news pieces are framed using standard language. One is required to understand satire to read it. Manipulation is inherent to the form and producing satire would require manipulation. The learning process will always be at the core of understanding any text. This basic principle has not changed in the digital era.

The public nature of most content on social media brings in several individuals who fill in the roles of creators and consumers of content. Every time a person shares a post, another person re-shares the same with a different caption, and a third person puts their comment on it, there is an addition made to the text itself. Sometimes, it loses the essence of what the original post was meant to be, and sometimes, the original post may be enhanced and complicated by the value addition. In such cases, there will be multiple authors and multiple readers. Usually, it is not easy

to find out which one among them would be more important in driving the conversation in a particular direction.

One must also note that the spaces on the internet have only diversified themselves throughout all these years, and they will continue accommodating more actors. With these actors, the users of the internet, every person would get the chance to express themselves in a way they see fit. The study of the content may enable a user to escape algorithms in place to detect problematic content that violates the policies of social media platforms. On the other hand, harmless posts that do not fit the definition of hate speech or even hateful speech might get flagged and eventually would be hidden. Facebook has admitted that they need to bring back humans into the review process of flagged information, and only bettering AI might not be the best solution to curb misinformation (Coldewey and Hatmaker).

Lastly, one must pay due attention to online humour as a form of expression with digital manipulation inherent in its features. Despite the many similarities from a literary criticism point of view, one must consider that regulating such expression would never be accurately executed by large-scale machine learning systems. Torrential proportions of data and their spread cannot be easily condensed to sources that would make systems like Rosetta or other smaller algorithms perfect in doing their job. Archiving content is key, and media researchers must pay due attention to it. How do users of social media platforms save their content? How do they make backups of what they share? Do these platforms allow viewing the history of an edited post and if yes, how does one access it? These are all important questions. A simple example of TikTok video compilations illustrates the importance of YouTube and Facebook. The former app is banned in India, and several users rely upon the latter platforms to watch compilations of videos from TikTok. The provision of shifting platforms is not the brainchild of the makers of AI that regulate social media. These compilations, for example, manifest in people seeking ways to reclaim the internet as a space. These practices go beyond a regulated space in the virtual domain with the spreading use of VPN[7] and browsers like TOR[8] and. This is also not a strategy people have been using collectively against algorithms that are used by Meta or other conglomerates, online forums, blogs, closed groups on individual websites and offline mobilizations of people to discuss what the internet cannot host is not unique to the last decade that 2010-2020. It is more important to see continuities than departures while studying how AI has been a game-changer in the regulation of digital content. In the last few years, several public figures have campaigned against the reach of AI; this includes protective lawsuits against the use of their visual and auditory imprint or a recreation of the same (HTTech). On the contrary, there are episodes where actors favour the use of AI technologies, including deepfakes, when it is done in a strategic manner. Vice media had analysed a 44-second clip featuring the Bharatiya Janata Party (BJP)

[7] Virtual Private Network
[8] The Onion Router

Delhi President Manoj Tiwari and found it to be a deepfake (Christopher). According to their reports on February 7, 2020, two videos of Tiwari criticizing the Delhi government under the leadership of Arvind Kejriwal were being heavily circulated on social media, especially WhatsApp. In one of the videos, which was found to be the original one, Tiwari was speaking in English, and in the other one, which was analysed and found to be a deepfake, he was speaking in a dialect of Haryanvi, which is commonly spoken in Delhi. The content of his speech in both videos was the same but one was real and one was not. Theorizing manipulation in such texts gets more complicated when AI technologies are viewed as tools in a world where their use is heavily normalised. Content remains text, but the analysis process becomes much more complicated. It is also important to acknowledge that any resistance to this regulation does not come from collectives tinged by political inclinations. The easier it becomes to put oneself out there in the virtual domain, the easier it will be to resist any regulation. This research hopes to build upon the observation of such patterns in future to understand human behaviour in the virtual domain that is, in one way of speaking, mediated by artificial intelligence.

**Bibliography**


Bazin, Andre. "The Ontology of the Photographic Image." *What Is Cinema? Volume I*. Berkeley: University of California Press, 1967. 9-16.

Johansen, Jorgen. "The Distinction between Icon, Index, and Symbol in the Study of Literature." (Ed.), M. Herzfeld & L. Melazzo. *Semiotic Theory and Practice, Volume 1+2: Palermo, 1984*. Berlin, Boston: De Gruyter Mouton , 2017. 497-504.

Gunning, Tom. "What's the Point of an Index? or, Faking Photographs." In Beckman, K. R., & In Ma, J. *Still Moving: Between Cinema and Photography*. Durham: Duke University Press, 2008. 23-40.

Kreps, Daniel. *Jon Stewart Calls Antisemitic Goblin Comment a Joke: 'Get a F–king Grip'*. 5 January 2022. <https://www.rollingstone.com/culture/culture-news/campaign-against-antisemitism-defends-j-k-rowling-jon-stewart-goblin-claims-1278891/>.

Barthes, Roland. "Internet archive, Wayback Machine." 1967. *web.archive.org*. <https://web.archive.org/web/20200419132326/http://www.ubu.com/aspen/aspen5and6/threeEssays.html#barthes>.

Wimsatt, W K and M C Beardsley. "The Intentional Fallacy." Wimsatt, W K. *The Verbal Icon: Studies in the Meaning of Poetry*. Lexington: University of Kentucky Press, 1954. 25-51.

Sivakumar, Viswanath, Albert Gordo and Manohar Paluri. *Rosetta: Understanding text in images and videos with machine learning*. 11 Sepetember 2018. <https://engineering.fb.com/2018/09/11/ai-research/rosetta-understanding-text-in-images-and-videos-with-machine-learning/>.

Desk, Entertainment. *Kangana Ranaut falls for spoof video, calls out Qatar Airways CEO 'for bullying a poor man'*. 9 June 2022.

<https://indianexpress.com/article/entertainment/bollywood/kangana-ranaut-falls-for-spoof-video-calls-qatar-airways-ceo-an-idiot-of-a-man-7958907/>.

Report, Ground. *Vashudev Qatar is trending on Twitter, Story behind trend is hilarious*. 7 June 2022. <https://groundreport.in/vashudev-qatar-is-trending-on-twitter-story-behind-trend-is-hilarious/>.

Toews, Rob. *Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared*. 25 May 2020. <https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=277813b74940>.

Warde, Claire. *Deepfakes: Is This Video Even Real?* 14 August 2019. <https://www.youtube.com/watch?v=1OqFY_2JE1c&t=156s>.

Coldewey, Devin and Taylor Hatmaker. *Facebook upgrades its AI to better tackle COVID-19 misinformation and hate speech*. 13 May 2020. <https://techcrunch.com/2020/05/12/facebook-upgrades-its-ai-to-better-tackle-covid-19-misinformation-and-hate-speech/>.

HTTech. *Rage Against the Mahcine: Scarlett Johansan, Keanu Reeves to Anil Kapoor, 10 celebrities battling AI invasion*. HT Tech: Hindustan Times, 2023.

Christopher, Nilesh. *We've Just Seen the First Use of Deepfakes in an Indian Election Campaign*. 18 February 2020. <https://www.vice.com/en/article/jgedjb/the-first-use-of-deepfakes-in-indian-election-by-bjp>.

# WELL, THAT ESCALATED QUICKLY: THE ROLE OF MEMES IN POLITICAL ENGAGEMENT

*Lavanya Mohan and Malvika Dixit*
*Lecturers, Department of Psychology*
*Sophia College for Women (Empowered Autonomous), Mumbai*

**Abstract:** Memes hold an important place in the ever-evolving tapestry of digital culture. They have transformed the way people share feelings, ideas, and information with each other. This is particularly evident during periods of immense socio-political importance such as the 2024 Indian Lok Sabha Elections. Given people's proclivity for using memes as tools to engage in political conversation, it is essential to investigate the relationship between this usage, and people's political behaviour. This study sought to examine the relationship between meme engagement and political engagement in the contemporary Indian landscape, using qualitative and quantitative data obtained from people aged 20-35 years. Regression analysis found this relationship to be significant. The use of memes helps streamline and bolster political discussions, predicting an increase in political engagement, both online and offline. This study contributes to the academic discourse surrounding meme culture and political communication, and suggests potential avenues for future research.

**Keywords:** Memes, Political Discourse, Digital Culture, Communication

## Introduction

Artists have always been at the forefront of perceiving and inciting social change, shaping the way we think about and engage with the world. From Thomas Nast who created the enduring symbols of the Republican and Democratic parties, to R.K. Laxman giving a voice to the common man, to the Amul cartoons that continue to have their finger on the pulse of society; artists have sparked reflection and revolution alike.

A newer form of this subversion appears to be memes. Over the years memes have taken the Indian political scenario by storm, with their scope for political commentary and satire. From the "Pappu" memes critiquing INC member Rahul Gandhi, or the cough syrup memes taking a jab at Arvind Kejriwal of the AAP, to the "Mandir yahi banega" meme referring to the Babri Majid dispute; the internet has been replete with political memes. Memes have become a ubiquitous part of online culture, as an accessible source of information for the masses, and a medium of communication for critics. However, they have also given a platform to vitriol and rumour mongering.

The past few elections have underscored the role social media and memes play in shaping public discourse. During these periods, memes became a source of information, misinformation, and

disinformation. Individuals used memes to echo and share opinions, weigh in on issues with the use of humour, and comment on the results.

Memes experienced their ascendency with Millennials and Gen Z, who rely heavily on them in their daily parlance, with words and phrases like 'lit', 'tea', 'Ok, boomer', becoming an integral part of their vocabulary (Afp; Borge and Nicolaou). This demographic also appears to be more politically active online. Given the popularity and rampant use of memes amongst people between 20 to 35 years, there arises a question on the utility of memes and if they have the capability to influence political engagement, especially in this cohort.

**Literature Review**

*The Subculture of Memes*

Since the internet's advent, online communication has rapidly evolved from merely using the written word to incorporating emoticons and memes to enhance conversation. Meme, as a term, was conceived by Richard Dawkins in 1976, who explained memes as cultural units of information transmitted through copying or imitation. Memes can take the form of catchphrases, photos, videos, audios, or Graphics Interface Formats (gifs), such as the "Keep calm and (…)" template.

The visual nature, brevity, lucidity, cultural relevance, and playfulness of memes contribute to their virality. Memes have found usage in social commentary and critique, education, marketing, and advertising. They help connect individuals across the globe by allowing them to share their opinions in a universal fashion that transcends the boundaries of language (Badlani; Reime). As a result, they have acquired the power to shape and influence public opinion and encourage social change. A powerful example of this is the MeToo Movement which gained traction with the use of the hashtag, #MeToo.

The reach of social media and the widening scope of content that can be created and shared on it, including memes, has led to individuals relying on it for information, especially news (as per a Reuters Institute Digital News Report from 2017) (Bebić, 44). According to Fernández-Villanueva & Bayarri Toscano, memes also appear to have created a sub-culture based on the understanding of meme language. According to Bernardino and Sollano, creating, sharing and decoding of memes provides users with a platform to engage in conversation with greater agency and freedom over the expression of their views, along with creating a collective identity. However, this communal aspect of memes leaves room for intentional and unintentional manipulation as well, especially if the faith in the sender is high. This, combined with the anonymity of the online sphere and the ease of disseminating memes, sways people into absorbing, accepting, and sharing information at face value (Santosa et al., 2-3; Lynch, 2-5).

*Memes in the Political Landscape*

Social media has given rise to newer forms of political activism which are not limited to voting, petitioning, and participating in demonstrations. Individuals share, create, and interact with political memes online to express themselves. While some are optimistic that this will lead to more democratic participation because of the anonymity of users, and a lack of censorship and geographical boundaries, others are wary of online political participation, deeming it a performative form of activism known as 'slacktivism'. The sceptics opine that 'slacktivism' dilutes activism to activities like liking, and sharing memes, or tweeting (Fernández-Villanueva, 451).

Research has found evidence indicating a relationship between online political expression and online political participation. It was seen that online political expression was positively associated with both creating and sharing political memes. However, this was not the case when it came to political interest, strength of partisanship, or political participation. The intention to persuade others was positively associated with creating political memes, but not with sharing them, whereas the intention to inform others was associated with sharing memes and not with creating them. (Halversen and Weeks, 6-9)

According to Rehman and Mahmood (113-116), exposure to memes was seen to be positively correlated with political polarisation, biases, political awareness, and online political participation of users in Pakistan. Some of these findings were echoed by Ahmed and Masood (5-8), who found that political social media usage positively correlated with online political participation. This relationship was positively mediated by exposure to political memes, and was moderated by political cynicism. It appears that if there is high enough exposure to political memes, it can lead to political cynics engaging in online political participation. This could be due to memes simplifying and bringing levity to relatively complex issues, thus making information more relatable. However, the usage of political memes is not without the possible trivialisation of sensitive and pressing issues. There is a chance that the humour used dilutes the seriousness of the matter or that satirical critiques are reduced to jest (Fernández-Villanueva & Bayarri Toscano, 450-452).

Memes are used for self-expression, establishing and strengthening social identity, and as a source of entertainment, but their role in the political sphere is debatable. Thus, the memetic discourse might be a way for users to maintain, create, or strengthen group membership (Mukhtar et al.; Leiser).

This begs the question whether online meme exposure translates to offline political engagement. Salikin and Sulaksono (415-417) found that the intensity, frequency, and duration of exposure to political memes on social media was found to be directly proportional to the level of participation of beginner voters during the presidential elections in Indonesia. While offline participation

largely involved discussing the issues with friends, family, and co-workers, and reading up on relevant information, online political participation entailed sending messages and political memes via Instagram to family, co-workers, and the social milieu.

Johann (159-161) found that the intensity of creation and sharing of political internet memes was positively related to online and offline political participation, with offline participation including involvement in community work, political campaigns, political meetings, and signing petitions. Both kinds of participation were associated with left-leaning political views. Nazeer and Ashfaq (118-120) found that political Facebook memes were a source of information on political issues and bringing to light loopholes in the system in Pakistan. They made political issues interesting and relatable for younger people, and the online 'slacktivism' did translate to offline political engagement in terms of engaging in conversations or learning more about an issue. However, Kulkarni (16-17) found that while memes were used as a means of communicating political satire, to spread political propaganda, and in political campaigns, they did not influence a change in political engagement in terms of voting behaviour, political opinions, and ideologies. The content of memes is not given the same esteem as traditional media. They are deemed a source of entertainment, not information.

Past research has found evidence supporting the relationship between political meme usage and online political engagement. However, the research on the relationship between meme usage and offline political engagement is inconclusive and is thus the focus of the current study. While people are extremely vocal of their opinions online in terms of condoning or condemning issues, causes, and people, the paper aims to understand if online engagement with memes can lead to political engagement, including that which manifests offline as both discussions had and actions taken.

**Method**

**Respondents**

Respondents (N = 52) were all residents of Mumbai, Delhi, and Bangalore, between the ages of 20 to 35 years. The age bracket, gender, and political leaning split is summarised in Table 1.

With the exception of one person who had just passed school, the respondents were either graduates ($N_{Graduate}$ = 24) or postgraduates ($N_{PostGraduate}$ = 27).

Most of them identified religiously as Hindu ($N_{Hindu}$ = 39). The remaining 13 identified as: Christian ($N_{Christian}$ = 6), Jain ($N_{Jain}$ = 3), Muslim ($N_{Muslim}$ = 2), Agnostic ($N_{Agnostic}$ = 1), and Atheist ($N_{Atheist}$ = 1).

**Table 1**
**Number of Respondents by Political Leaning, Gender and Age Categories**

| | Liberal | | | Moderate | | | Conservative | NA[a] |
|---|---|---|---|---|---|---|---|---|
| Age (in years) | 20-24 | 25-30 | 31-35 | 20-24 | 25-30 | 31-35 | 20-24 | 31-35 |
| Male | 3 | 4 | 1 | 1 | 11 | 1 | - | 1 |
| Female | 4 | 8 | 3 | 3 | 10 | 1 | 1 | - |
| Total | 7 | 12 | 4 | 4 | 21 | 2 | 1 | 1 |
| | 23 | | | 27 | | | 1 | 1 |

Source: Data collected by Authors
Note: [a] One respondent described themselves as having 'no political inclination'

**Procedure**

The questionnaire was disseminated using the online form and survey platform Typeform, and respondents were recruited via snowball convenience sampling and thus, a bias exists. Survey links were shared with individuals meeting the study's age criteria, who were encouraged to pass them onto others within their networks. Data was collected between 26th August 2024 and 10th September 2024.

**Materials**

The questionnaire was developed for this study after reviewing multiple sources that explored memes and political engagement. Both qualitative and quantitative data was collected. The complete questionnaire can be found on OSF (Mohan and Dixit).

Questions on political engagement included asking respondents about actions they have taken related to politics (such as voting in elections, signing petitions, participating in protests, etcetera.), how important it is for them to stay informed about various social issues and how they do so, the frequency of their political discussions, and with whom and where they have them.

For the purpose of this study, only political memes were considered, and they were defined as memes about politicians, political parties, political issues, political incidents and political ideologies. This was conveyed to the respondents before they answered any questions related to memes. Meme engagement was investigated by asking about how (encountering, liking, sharing, commenting and creating memes), and how often, respondents interacted with political memes, and whether they were prompted to delve deeper into a particular topic due to interacting with a meme about it. They were also asked about their motivations for sharing memes, and how they might be used to elevate (or degrade) the quality of political discourse.

**Data Analysis**

Quantitative data analysis was conducted using R Studio. Given that political engagement and meme engagement were both latent variables and measured using various scales, the scores were standardised prior to analysis.

Cronbach's Alpha was calculated for both sets of questions to verify internal consistency. Confirmatory Factor Analysis (CFA) was performed using the lavaan package to confirm that the observed variables or questions adequately captured the underlying latent constructs of political engagement and meme engagement. Model fit indices such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Standardized Root Mean Square Residual (SRMR) were utilised to evaluate the hypothesized model as they are comparatively less sensitive to smaller sample sizes.

Post this, correlation analysis was conducted using the cor function to examine if meme engagement and political engagement were related. This was followed by regression analysis using the lm function to create a linear regression model to assess whether meme engagement could predict political engagement, whilst controlling for age, gender, education, religion, and political leaning.

Thematic analysis was used to analyse the qualitative data. However, since the data obtained was minimal, it was primarily used to supplement the quantitative findings.

**Results**

**Reliability Analysis**

The internal consistency of the Political Engagement scale, consisting of 5 items was found to be acceptable, $\alpha = 0.72$ (95% CI: 0.57 to 0.82) and standardised $\alpha = 0.74$. The table presenting the reliability statistics for each item if it were to be dropped from the scale, and the average inter-item correlation, can be found in the OSF project (Mohan and Dixit). Since $\alpha$ when Q1 is dropped is 0.71 (standardised $\alpha = 0.73$), it can be noted that dropping Q1 would increase or not affect the reliability of the scale. This could indicate that it is not capturing a unique aspect of the construct

of political engagement. However, dropping Q1 would not make sense theoretically and thus, it was included in the final analysis. Conversely, dropping Q4 would decrease the reliability significantly (α when dropped = 0.65, standardised α = 0.67), thus indicating its importance for the scale.

The internal consistency of the Meme Engagement scale, consisting of 4 items, was found to be acceptable as well, α = 0.71 (95% CI: 0.58 to 0.84), both raw and standardised. The table presenting the reliability statistics for each item if it were to be dropped from the scale, and the average inter-item correlation, can be found in the OSF project (Mohan and Dixit). Here, dropping Q4 would not affect the reliability of the scale – α (raw and standardised) if dropped = 0.71. This could indicate that it is not contributing effectively to the construct of meme engagement as measured by the other items. However, theoretically, it is measuring a unique aspect of meme engagement, which is why it was included in further analysis.

**Confirmatory Factor Analysis (CFA)**

CFA conducted to evaluate the latent variables Political Engagement (PE) and Meme Engagement (ME), using 5 and 4 questions respectively, yielded positive results for the study. For the PE model, the CFI was 1.000, TLI was 1.128, and SRMR was 0.036, all indicating excellent model fit. Similarly, for the ME model, the CFI was 1.000, TLI was 1.128, and SRMR was 0.020, all indicating excellent model fit. All CFA results can be accessed in the OSF project (Mohan and Dixit).

The factor loadings for all 5 items/questions in the PE model were significant, with Q2 (β = 2.151, p = 0.008) and Q4 (β = 2.020, p = 0.007) emerging as the most significant contributors to the latent variable. However, the variance for the items and latent variables showcases a more complex picture, indicating that this sample may not capture the entire variability of PE and thus, the need for future research on this variable.

For the ME model as well, the factor loadings for all 4 items/questions were significant. However, compared to Q1, Q2, and Q3, Q4 was a weaker contributor to the latent variable (β = 0.708, p = 0.007) and had a slightly higher variance as well. This could suggest that it measures a different aspect of ME or that respondents have more varied opinions regarding it, which do not align as closely with the other items.

**Descriptive Statistics for Measurement Variables and Latent Variables**

Since the questions to measure PE and ME used different scales, raw and proportionate means (and standard deviations) were computed for them and the latent variables. These can be found in Table 2 on the next page.

From the means, it can be seen that respondents generally want to be informed about issues and are open to discussing them frequently with multiple groups but are not as proactive while carrying out political actions. While a majority of the respondents voted during the Lok Sabha Elections ($N_{VotedLokSabha}$ = 34), a smaller number voted during State Elections ($N_{VotedState}$ = 16) and Local Elections ($N_{VotedInLocal}$= 14). Additionally, encouraging people to vote was a common action undertaken by many respondents ($N_{EncouragedVoting}$ = 28), yet only 8 of them have ever participated in meetings/seminars/speeches on political issues, and just 1 person has contributed time or money to a campaign/party/candidate. Thus, overall, the average respondent only scored 55% of the maximum PE score possible, indicating a fair but moderate level of political engagement.

When it comes to engagement with political memes, it is evident that respondents are more likely to interact with memes through liking/reacting ($N_{Like/React}$= 41) and sharing ($N_{SharePrivately}$ = 30, $N_{SharePublicly}$ = 15), than commenting on or creating them. In fact, only two people reported commenting on memes during the 2024 Lok Sabha Election season, and none created a meme. Overall, (political) meme engagement seems to be somewhat passive, as seen by how the average respondent only scored 48% of the maximum ME score possible.

**Table 2**
**Means and Standard Deviations (Raw and Proportionate) of the Latent and Measurement Variables**

| Question/Variable | Raw Mean and Standard Deviation | Proportionate Mean and Standard Deviation |
|---|---|---|
| Number of elections voted in (scored from 0 to 3) | *M*=1.23, *SD*=0.98 | *M*=0.41, *SD*=0.33 |
| Number of political actions undertaken in the past 4-5 years (scored from 0 to 6) | *M*=1.33, *SD*=1.29 | *M*=0.22, *SD*=0.22 |
| Importance of staying informed about current affairs (scored from 0 to 3) | *M*=2.60, *SD*=0.57 | *M*=0.86, *SD*=0.19 |
| Frequency of discussing politics with others (scored from 0 to 3) | *M*=1.98, *SD*=0.90 | *M*=0.66, *SD*=0.30 |
| Groups with whom politics is discussed (scored from 0 to 4) | *M*=2.38, *SD*=0.97 | *M*=0.60, *SD*=0.24 |

| | | |
|---|---|---|
| **Political Engagement** | - | ***M*=0.55, *SD*=0.58** |
| Interaction with political memes online (scored from 0 to 5) | *M*=1.69, *SD*=0.87 | *M*=0.34, *SD*=0.17 |
| Frequency of encountering or receiving political memes during the 2024 Lok Sabha Election and Pre-Election period (scored from 0 to 3) | *M*=2.13, *SD*=0.84 | *M*=0.71, *SD*=0.28 |
| Frequency of sharing or creating political memes during the 2024 Lok Sabha Election and Pre-Election period (scored from 0 to 3) | *M*=0.69, *SD*=0.94 | *M*=0.23, *SD*=0.31 |
| Frequency of seeking more information about a topic based on a political meme (scored from 0 to 3) | *M*=1.92, *SD*=0.97 | *M*=0.64, *SD*=0.32 |
| **Meme Engagement** | **-** | ***M*=0.48, *SD*=0.58** |

Source: Data collected by Authors
Note: Raw means indicate participants' average responses on the actual scales, proportionate means indicate the normalised average scores on a common scale.

**Correlation and Regression Analysis**

The correlation between the latent variables, ME and PE, was found to be 0.508, indicating a moderate positive relationship between them. This suggests that higher levels of meme engagement were associated with higher levels of political engagement for the respondents.

Linear regression analysis revealed that ME significantly predicts PE ($\beta=0.45$, $p<0.001$) – implying that for every unit increase in meme engagement, political engagement increases by about 0.45 units. The intercept was non-significant, indicating that the model's prediction for zero ME is

uncertain. Since it is unlikely for individuals to have zero meme engagement, the importance of this result is reduced; nonetheless, it should be investigated further in the future.

The adjusted R-squared of the model is 0.2437 indicating that about 24% of the variance in PE can explained by ME, and the F-statistic, $F(1,50)=17.43$, $p<0.001$, indicates good model fit for the data. This suggests that meme engagement contributes significantly to predicting and understanding political engagement, but there are other variables that influence political engagement as well that this study does not include.

**Thematic Analysis**

Respondents prefer to engage in offline political conversations – seeking easy exchanges with like-minded individuals and to evade the possibility of online trolling. To avoid conflict, and possibly to ensure safety, especially when one is a minority, certain topics are avoided based on company and location.

Respondents also look for quick, easily digestible content from trusted sources. Traditional media, such as newspapers and news channels, are seen as slow, inaccessible, or biased due to affiliations. While social media's reliability is variable, it may be considered more trustworthy, depending on the source.

Memes serve as tools for communication and entertainment rather than primary sources of information. Despite the possibility that they may trivialize serious issues and spread misinformation, they facilitate discussions on complex topics, making them more approachable through humour and frivolity. While some respondents acknowledge that memes can introduce unfamiliar topics, they are not considered wholly reliable (especially from untrusted sources) nor a means to foster emotional connections or persuade opponents.

**Discussion**

The aim of this study was to understand if meme engagement, which happens online, can lead to political engagement, especially that which manifests offline. The scales used to measure meme engagement and political engagement were found to have acceptable internal consistency. Confirmatory factor analysis also confirmed the hypothesized structure of the two latent variables. In line with the hypothesis, the regression analysis found that meme engagement does indeed predict political engagement.

In previous studies, meme engagement has been associated with online political participation; however, the association with offline political engagement/participation was suspect. While the results of this study support the association between meme engagement and offline political behaviour, it should be noted that the sample was more likely to discuss political issues offline

than actually carry out other political actions such as voting or participating in protests, rallies, meetings, etcetera. Thus, while meme engagement does appear to lead to offline political participation, it does so primarily through facilitating political conversation. Respondents' qualitative answers support this by having a positive impression of memes and, contrary to previous research, not feeling much apprehension about their supposed negative effects such as misleading information, trivialising issues, etcetera. They believe memes are helpful in making political conversation easier by imbuing levity and humour in these conversations, making serious issues more palatable, and complex content/information more accessible and confrontable.

What memes appear unable to do is lead to persuasion, or gains in social capital. Since, according to Halversen and Weeks, it was creating memes that was related to intention to persuade, and these respondents only like/react to and share memes, they do not seem inclined to use them for the purpose of persuasion. This sample seemed to engage with politics somewhat passively. If they were more heavily invested in politics and used memes to communicate this investment, they might have wanted to and managed to foster social capital. However, since their investment in politics was not measured, this remains a hypothesis, and a possible avenue for future study.

Memes also seem to be a trigger for information collection, but this is a smaller part of the respondents' engagement with them, as seen by the smaller factor loading of Q4 on this latent variable. Since sharing information, especially about important causes, also helps foster a sense of community, this could be an aspect of meme engagement partly explored in the current study.

However, some limitations of this study must be acknowledged. Firstly, the sample size was much smaller relative to the population, which limited the statistical power of the study and its generalisability. It also lacked representation with regard to key demographics, particularly minority religions, individuals with lower educational qualifications, and those who identity as political conservatives. This prevented these demographic variables from being meaningful control variables for the regression analysis, and thus, this analysis was not included in the study. Additionally, the study restricted the age range to 20-35 years and did not extent the survey beyond urban areas.

Secondly, the questionnaire lacked a scale to measure and classify political leaning which could have led to inconsistencies in its interpretation, and the lack of significance of its relationship with the other variables. Thirdly, the study's qualitative approach to the topic was not rigorous enough to lead to a nuanced understanding of the relationship between meme engagement and political engagement. While some broad and subjective questions were intended to elicit exploratory information, they did not succeed in doing so adequately. A more comprehensive and pilot-tested questionnaire could have offered a more accurate and well-rounded understanding of the variables studied.

Lastly, the memes were not categorised by content and/or purpose, nor were respondents asked to do so, which prevented an analysis of the potentially distinct effects of different types of memes. Thus, investigating the influence of various meme types on political engagement could also provide valuable insights.

Future research could thus address the aforementioned points, and further examine the influence of meme engagement on political engagement through longitudinal studies across election cycles, and cross-cultural studies across multiple regions and countries. Moreover, the psychological mechanisms underlying these variables and their relationship could also be assessed.

Memes democratise the political discourse by breaking down linguistic and geographical barriers, making information more accessible and digestible. However, whether they have the potential to serve as educational and thought-provoking tools remains to be seen, as they currently appear to occupy a liminal space. Digital culture has furthered both reflection and revolution in contemporary society for the millennial, Gen Z, and other current generations ever since its inception; memes allow for reflection but aren't sparking a revolution *yet*.

**Bibliography**


Afp. "Gen Z, Millennials Care More about Memes than Movies, Finds Study." *News18*, 23 Dec. 2021, www.news18.com/news/buzz/gen-z-millennials-care-more-about-memes-than-movies-finds-study-4584197.html.

Ahmed, Saifuddin, and Muhammad Masood. "Breaking Barriers with Memes: How Memes Bridge Political Cynicism to Online Political Participation." *Social Media + Society*, vol. 10, no. 2, SAGE Publishing, Apr. 2024, pp. 5–8, https://doi.org/10.1177/20563051241261277.

Badlani, Dimple. "From Genesis to Its Zenith: A Study on Evolution of Memes as a Mainstream Media Content." *Journal of Communication and Management*, vol. 2, no. 02, June 2023, pp. 148–52, https://doi.org/10.58966/jcm20232214. Accessed 13 Sept. 2024.

Bebic, D. (Domagoj), and M. (Marija) Volarevic. "Do Not Mess with a Meme: The Use of Viral Content in Communicating Politics." *Dadun.unav.edu*, vol. 31, no. 3, 2018, pp. 43–56, dadun.unav.edu/handle/10171/55766.

Bernardino, Kim. "Political Activity and Memes | PDF | Meme | Social Media." *Scribd*, 17 May 2017, www.scribd.com/document/351601817/Political-Activity-and-Memes. Accessed 19 August 2024.

Borge, Jonathan, and Elena Nicolaou. "Here's What All of Those Popular Slang Words Really Mean." *Oprah Daily*, 5 Oct. 2018, www.oprahdaily.com/entertainment/g23603568/slang-words-meaning/?slide=20. Accessed 30 Sept. 2024.

Fausto, Colombo F. "Reviewing the Cultural Industry: From Creative Industries to Digital Platforms." *Unav.edu*, vol. 34, no. 2, Servicio de Publicaciones de la Universidad de Navarra, 2018, pp. 449–68, https://doi.org/2386-7876. Accessed 2 Sept. 2024.

Fernández-Villanueva, Concepción, and Gabriel Bayarri-Toscano. "Legitimation of Hate and Political Violence Through Memetic Images: The Bolsonaro Campaign." *Communication & Society*, Apr. 2021, pp. 449–68. https://doi.org/10.15581/003.34.2.449-468.

Harbo, Tenna Foustad. "Internet Memes as Knowledge Practice in Social Movements: Rethinking Economics' Delegitimization of Economists." *Discourse, Context & Media*, vol. 50, Dec. 2022, https://doi.org/10.1016/j.dcm.2022.100650.

Johann, Michael. "Political Participation in Transition: Internet Memes as a Form of Political Expression in Social Media." *Studies in Communication Sciences*, vol. 22, no. 1, Mar. 2022, https://doi.org/10.24434/j.scoms.2022.01.3005. Accessed 17 Sept. 2024.

Kulkarni, Anushka. "Internet Meme and Political Discourse: A Study on the Impact of Internet Meme as a Tool in Communicating Political Satire." *SSRN Electronic Journal*, vol. 6, 2017, pp. 13–17, https://doi.org/10.2139/ssrn.3501366.

Leiser, Anne. "Psychological Perspectives on Participatory Culture: Core Motives for the Use of Political Internet Memes." *Journal of Social and Political Psychology*, vol. 10, no. 1, June 2022, pp. 236–52, https://doi.org/10.5964/jspp.6377. Accessed 30 August 2024.

Lynch, Michael P. "Memes, Misinformation, and Political Meaning." *The Southern Journal of Philosophy*, vol. 60, no. 1, Mar. 2022, pp. 38–56, https://doi.org/10.1111/sjp.12456.

Mohan, Lavanya, and Dixit, Malvika. "Well, That Escalated Quickly: The Role of Memes in Political Engagement." Open Science Framework, 2024, https://osf.io/tsauh/?view_only=c67d05fbdb0940eba40c22aeac202d6b.

Moreno-Almeida, Cristina. "Memes as Snapshots of Participation: The Role of Digital Amateur Activists in Authoritarian Regimes." *New Media & Society*, vol. 23, no. 6, Mar. 2020, https://doi.org/10.1177/1461444820912722.

Mukhtar, Shahira, et al. "Memes in the Digital Age: A Sociolinguistic Examination of Cultural Expressions and Communicative Practices across Border." *Educational Administration: Theory and Practice*, vol. 30, no. 6, May 2022, pp. 1443–55, https://doi.org/10.53555/kuey.v30i6.5520. Accessed 15 Aug. 2024.

Nazeer, Javeria, and Dr. Ayesha Ashfaq. "Effects of Slacktivism on Offline Engagement: Appraising the Role of Political Facebook Memes as a Communication Tool for Young Consumers in Pakistan." *Journal of Development and Social Sciences*, vol. 4, no. I, Mar. 2023, pp. 113–112, https://doi.org/10.47205/jdss.2023(4-i)10. Accessed 16 Aug. 2024.

Rehman, Habib Ur, and Zubair Mahmood. "Internet Memes Aas Political Communication Tool: An Exploratory Study." *Journal of ISOSS*, vol. 7, no. 3, 2021, pp. 107–18, www.joi.isoss.net/PDFs/Vol-7-no-3-2021/08_J_ISOSS_7_3.pdf. Accessed 30 Sept. 2024.

Reime, Thov. *Memes as Visual Tools for Precise Message Conveying: A Potential in the Future of Online Communication Development*. pp. 1–12, www.ntnu.no/documents/10401/1264435841/Design+Theory+Article+-+Final+Article+-+Thov+Reime.pdf.

Rosseel, Yves, et al. "Lavaan: Latent Variable Analysis." *R-Packages*, 20 Dec. 2023, cran.r-project.org/web/packages/lavaan/index.html.

Salikin, Raja Muhammad Munzir Nur, and Tunjung Sulaksono. "The Influence of Political Memes in Social Media on the Level of Political Participation of Beginner Voters in the Presidential Election Year 2019 in Karimun District." *Atlantis-Press.com*, vol. 209, Atlantis Press, Mar. 2022, pp. 404–17, https://doi.org/10.2991/aebmr.k.220209.053. Accessed 30 Sept. 2024.

Santosa, Pudjo, et al. "The Reception of Memes as Political Information in the Media." *Web of Conferences*, vol. 73, 2018, pp. 1–3, https://doi.org/10.1051/e3sconf/201873.

Soh, Wee Yang. "Digital Protest in Singapore: The Pragmatics of Political Internet Memes." *Media, Culture & Society*, vol. 42, no. 7-8, Feb. 2020, https://doi.org/10.1177/0163443720904603.

Yves Rosseel, et al. *Lavaan: Latent Variable Analysis*. Sept. 2012, https://doi.org/10.32614/cran.package.lavaan. Accessed 30 Sept. 2024.

# RULES OF ENGAGEMENT: LAWs, ETHICS AND THE CHANGING NATURE OF WARFARE IN THE AI ERA

*Sukanya Menon*
*Assistant Professor, Department of Sociology*
*Sophia College for Women (Empowered Autonomous), Mumbai*

**Abstract:** In the last decade, with its rapid growth, AI has become a pivotal external catalyst for social change. In its strictest definition, AI refers to "the simulation of human intelligence by a system or a machine" (Xu). It is vital to note that beyond a mere technological shift, AI has transcended to become an imminent part of society at large. One such area, where it has become a critical point of contestation is with regard to warfare. In the last decade, there has been a shift in the nature of warfare largely due to the technological expansion. The advancement of lethal-autonomous weapons (LAWs), increasing role of cyber tools and other forms of AI have not only changed the face of war, but also raises pertinent questions on the ethics and regulation of the same. Furthermore, these developments have tipped the scales of global power structures thus impacting international relations. The UN Secretary General has referred to "humanity being on a knife's edge" due to the AI technologies multiplying the threat of nuclear war (Borger 2024). Using a systematic review, this paper aims to understand the role of AI on warfare strategies and the ethical concerns it raises in a global landscape.

**Keywords**: Warfare, Artificial Intelligence (AI), LAWS, ML, AWS, Ethics, Conflict

## Introduction

The 21st century has been referred to as the 'Era of Big Data' (Duan et al). As we move from an age of mechanization to automation, data has also expanded to include metadata- referring to the "information that describes and contextualizes the data, aiding in its organization, retrieval, and comprehension" (Atlan). This has impacted and altered the world in many ways. One of the major catalyst responsible for this rapid change has been the growth of Artificial intelligence or A.I. By its very definition artificial intelligence (henceforth referred to as AI) refers to the 'simulation of human intelligence by a system or machine" (Xu) and has been "… frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience" (Copeland). The growth of the AI has also been referred to as 'a critical element of the Fourth Industrial revolution' (Schwab) for its innate ability to quickly decipher through vast amounts of data and to identify the underlying patterns. The significant component in this becomes the speed with which AI completes these tasks. At present, the functioning of AI is largely dependent on human insight and control, but with growing research on technology and the expansion of autonomous or self-reliant machines, this is bound to change in the future (Collins).

The last few decades have also been marked by an increase in the number of conflicts (Patel et al). The emergence of non-state actors, environmental degradation and resource deprivation along with political and social identity markers have all contributed into the growing rate of global conflicts in the contemporary era (Cederman et.al). The turbulent geopolitical situation has ensured that every country remains prepared with adequate arsenal at their disposal. As countries begin to expand their troops- beyond just traditional weapons and methods of combat - the role of AI has been extensive. The emergence of digitized weaponry in the form of algorithm-based AI tools, automated machinery and surveillance tools, have thus contributed in creating a potentially dangerous situation, which is bound to escalate in the future (Egon et al).

AI has drastically transformed warfare, reshaping not only military strategies but also the ethical, social, and political landscapes that surround it and thus, has implications for global geopolitical relations. As nations race to develop and implement AI capabilities in military operations, concerns have emerged regarding an 'arms race' that could potentially escalate at a rapid pace. The lack of a binding regulatory legislation which may have some control over these weapons and their use, has aggravating the issue further (Dillon).

This transformation can be comprehended theoretically with the help of the Actor-Network Theory (ANT). ANT is grounded in the notion that both human and non-human actors participate actively in the creation and maintenance of networks that influence social structures. ANT treats non-human entities—such as AI systems, algorithms, and military drones—as equally significant actors within social networks. These actors have agency, meaning that they can influence decisions, shape behavior, and create outcomes. At the same time, human actors, such as policymakers, soldiers, and engineers, negotiate with these non-human entities within networks, aligning their interests to produce stabilized outcomes through a process known as translation. Networks are dynamic and continually evolve as new actors enter or old actors withdraw, leading to the continuous reconfiguration of warfare and society (Gutierrez).

The objective of this paper, thus lies in understanding the influence of AI on modern warfare. While the technological and strategic aspect are more easily identifiable- the more pressing concern lies on the ethical and socio-political bearing- which are often latent and more serious in nature.

**AI and Warfare: How is it connected?**

There has been a deep interconnection between the integration of AI into military operations- especially as far as warfare is concerned. As warfare becomes more complex in nature with the emergence of analytical and cyber capabilities, there has also been an increase in the growth of newer technologies to aid the same. Especially where military power and international security are concerned, the evolution of AI has resulted in what Schwab refers to as "a shift from mechanization to informatization and ultimately intelligentization" (CNAS).

From a strategic point of view, AI can be vital in decision making and planning. The use of AI has resulted in an evolution from traditional forms which were primarily based on human cognition, to AI tools which scourge through mountains of data at a quicker pace (Kumar). This is primarily done with the help of machine learning (ML)- a subset of AI which allows computers to make predictions through algorithms; done through the autonomous detection of patterns, these systems run without human intervention and is a crucial tool in predictive analytics. Thus, in a shorter time, AI tools- with the help of ML, not only limit the 'human bias', but also increases its accuracy and efficiency thereby paving way for a new era of warfare.

The use of AI has also allowed the synchronization of multiple domains like land, sea, air and space, in more efficient ways (Song et at). The integration of data from various sources like satellite images, cyber intelligence, surveillance drones are all processed from various sources to get a holistic picture of the environment. While increasing the complexity of the operation, it also extends to more effective military actions, thus making it a key feature of the modern military strategy (Kriegler).

Modern warfare strategies are further revolutionized with the emergence of automated decision-making features, which reduces the need for human insights and control. Autonomous Weapons Systems (henceforth, AWS) have been defined as 'systems capable of understanding higher level intent and direction, namely of achieving the same level of situational understanding as a human and able to take appropriate action to bring about the desired state." (Humble). U.S. Department of Defense in 2023 proposed a different approach and defined "AWS as being capable of once activated, to select and engage targets without further intervention from a human operator." (AWS, DoD). Though AWS has been present in warfare weapons since the past decade, it is the use of lethal autonomous weapons systems (LAWS) which has raised critical questions on the use of AI in warfare. LAWS are weapons which are designed to "identify, track and engage targets without direct human control" (Egeland). This can allow for quicker responses and an increased ability to infiltrate dangerous and changing battle conditions, while reducing the risk to human soldiers. The utilization of these weapons has had a substantial impact on surveillance and combat operations- thus being a game-changer in modern warfare.

**How does it change things?**

The emergence of LAWS and ML has had significant impact on military doctrine as well as on all the stakeholders of war- from policy makers, soldiers, machines and even on the non-combatants and civilians. While traditional warfare has emphasized on human decision making and adaptability, the integration of AI, clearly indicates the need for a shift towards newer doctrines which must reflect on the data-driven analysis and growing role of automation in military operations.

Furthermore, there has been a fundamental shift in the relation between soldiers and machines. While humans earlier used weapons in the battlefield, the emergence of automated weaponry

has resulted in a shift in this dynamic. To modern warfare, AI is deployed along human soldiers- to augment human capabilities through enhances situational awareness, operational support and other tools. For example, AI-powered exoskeletons can assist soldiers in carrying heavy loads and enhance human locomotion ("AI Powered Exoskeleton"), while AI-driven reconnaissance systems can provide real-time intelligence to inform tactical decisions (Almeida et al.). This collaboration between humans and AI represents a new paradigm in warfare, where machines are no longer merely tools but active participants in military operations. However, this changing dynamic- and the role of autonomous (specifically lethal weapons) have raised pertinent ethical and legal questions especially in terms of accountability and responsibility.

**Ethical Concerns**

The primary ethical concern in this area has been on the potential consequences of these lethal weapons and the erosion of human oversight in military operations- especially with regard to decision making. As AI advances itself to become more capable of operating independently, the possibility of fully autonomous weapons—capable of selecting and engaging targets without human intervention—raises profound moral questions (Sharkey). There is a strong lobby of critics who argue that allowing machines to make life-and-death decisions undermines the principles of human dignity and accountability that have traditionally governed warfare (Armitage).

Furthermore, it seriously undermines the philosophical foundations of warfare and the existing military doctrines as well as the principles of war. Jus in bello (*justice in war)* is the foundation upon which International Humanitarian Law (IHL) is based. Based on the Just War Theory, it provides the guidelines on conducting warfare ethically with focus on discrimination (between combatants and non-combatants) and proportionality ("Just War Theory"). This implies that, someone, at some level will be held accountable for the actions undertaken during war.

As warfare advances from human based combat, there becomes a pressing need for the development of doctrines that will regulate the use and impact of AI tools (Emery). Furthermore, the emergence of autonomous weapons- especially lethal ones, raises pertinent questions on the accountability and the just war theory- especially with regard to jus in bello. What exactly becomes the set of moral ideas which regulate the conduct of parties which are engaged in armed conflict? While at present, the accuracy of these weapons may not be at par with human beings, it is almost certain that the development of technology will lead to improved accuracy of these machines. It also raises serious questions on the autonomy of weapons may sometimes blur the line on human rights and dignity (Johnson).

This has led to calls for international regulations to limit the development and use of fully autonomous weapons. Advocacy groups have pushed for a ban on these "killer robots", questioning the very moral questions raised earlier (Sharkey); military leaders and policy makers- nationally and internationally- are grappling with the larger social implications of the same,

including the potential AI arms race between world powers, and AI enabled surveillance being used against civilians (Rosendorf).

**Present Scenario- The Arms Race towards AI control**

The world powers wrestling for global domination through an arm's race is not new, but the incorporation of AI and AWS into military strategies is definitely a game changer. Countries such as USA, China and Russia are investing heavily in AI technologies to enhance their military capabilities (Bode and Qiao-Franco), and prioritizing on the use of AI in their defense strategies (Kania).

The competition for AI dominance has further escalated the geopolitical tensions in regions such as the South China Sea, Eastern Europe and countries in the Middle East. Countries have become increasingly wary of other's AI capabilities leading to escalated tensions which may lead to potential conflicts. Under such volatile circumstances, the probability of a Third World War- while speculative- cannot be discounted. As nations leverage AI for strategic advantages, there is also a heightened risk of unintended consequences which come as a result of these autonomous weapons. An increasing dependence on these weapons could lead to a decline in the human control over warfare, leading to situations where machines would be making pivotal decisions without adequate ethical considerations (Motwani).

Although a world war in its traditional sense may be unlikely, the changing nature of warfare has been anything but traditional. Supported by AI technology, repercussions of small global conflicts could be far greater, leading to grave dangers at an international level.

As of September 2024, USA, China, Israel and Russia- countries at a critical junction with regard to their role in an ongoing global conflict- have openly admitted to developing or using AI tools in their arsenal. The power these countries hold at a geopolitical level have cushioned the brunt of the attacks being pelted out with the AI led weaponry- especially those aimed at innocent civilians (Milley and Schmidt).

The Russia-Ukraine war has been a prime example wherein both adversarial parties have actively developed and deployed AI for military applications. The integration of AI technologies in areas such as geospatial intelligence, unmanned systems operations, military training, and cyber warfare has proven instrumental in achieving battlefield effectiveness. Especially for Ukraine, it has become a key data analysis tool, allowing them to analyze vast amounts of data accurately and efficiently as opposed to the (predominantly) traditional warfare being utilized by Russia (Kostenko).

As part of its ongoing genocide in the Gaza Strip (Nijim)- the Israeli army has been using AI power drones (quadcopters) to fire directly at Palestinian civilians (Hasan and Buheji). Project '*Lavender*' employs ML- by leveraging algorithms to analyze data from varied sources- for surveillance and targeted strikes; and has drawn criticism for its *potential* violation of human

rights (Sharma). Furthermore, the accuracy of these assessments has raised serious concerns, with reports indicating massive collateral damage which has disproportionately affected civilians in Gaza (Liu). According to reports, Lavender is supplemented by another AI tool- '*Where's Daddy*'- which targets these militants in their homes- thus increasing the causality to also include non-militants and civilians (Business Insider). Israel has been accused of several war crimes by channels such as the UNHRC. However, the lack of legislatures and regulations have blocked law proceedings (UNHRC).

The Chinese government has invested heavily in AI research through it's '*The New Generational Artificial Intelligence Development Plan' (*Xu). However, concerns have been raised on the deployment of the AI powered surveillance systems in areas such as Xinjiang- where reports of mass surveillance, repression of ethnic minorities and human right abuses have been reported. Similar has been the case in USA, where AI has been integrated into its military operations- ranging from drone warfare to predictive analytics (Wilson). Several of the autonomous weapons have resulted in civilian deaths

**Regulation and Legislation**

The integration of AI into warfare and the development of LAWs have raised some crucial questions- with regards to the ethics, legality and its regulation. As nations explore the potential of AI driven weapons, there is a pressing need to create comprehensive regulation and legislation to ensure accountability and "compliance to international humanitarian laws" (UN). Since 2018, the UN has been a key advocate for this cause. UN Secretary General Antonio Guterres has stressed on having a strong stance against LAWS, to make them politically unacceptable and indefensible; arguing that "delegation of such critical decisions to machines undermines fundamental principles of human dignity and accountability in warfare" (Garcia). This position has been supported by various member states and civil society organizations.

In 2023, the UN's First Committee on Disarmament and International Security approved a draft resolution calling for a comprehensive examination of the implications of LAWs and encouraging member states to engage in dialogue about their regulation. The discussions surrounding LAWS have highlighted the need for a legally binding treaty that would prohibit the development and deployment of autonomous weapons that operate without human control (UN). A landmark resolution was passed in March 2024 which focused on promoting safe, secure and trustworthy AI systems; recognizing that while AI has the potential to enhance military effectiveness, it also poses significant risks if left unchecked. The resolution aims to foster international cooperation to ensure that AI technologies are developed and used in ways that align with humanitarian principles and respect human rights (UN).

In 1980, the UN agreement- The Center on Certain Conventional Weapons (CCW) was signed to limit or prohibit the use of certain weapons which cause unnecessary or indiscriminate harm and prohibited the use of weapons such as landmine and that which cannot be identified by X-Ray.

Clearly outdated, it is high time that it gets revised and updated considering the advancements which have come in AI and technology.

Since the existing legal frameworks such as the 'Geneva Conventions' primarily focus on the traditional warfare, the challenge lies in creating cohesive regulations which specifically address AI and forms of autonomous weaponry. This ever-widening gap has prompted calls for establishing new norms to govern the development and deployment of AI in military operations. It has also been observed that autonomous weapons- especially LAWS- do not comply with the existing IHL, thus absolving countries from accountability and threatening fundamental human rights- including right to life and non-discrimination.

Along with several other international organizations, UN has convened discussions on this topic, with some member states advocating for a pre-emptive ban on LAWs. In 2019, over 26 nations endorsed a joint statement calling for "*the establishment of a legal framework to regulate the use of LAWs and ensure compliance with international humanitarian law*" (UN).

However, there are several challenges to establishing effective regulations. The rapid pace of technical advancements overtakes the ability of policy makers to create effective frameworks. At an international level, vested interests and strategic priorities of its member states have complicated the creation of a unified approach for the same, Additionally, the UN does not possess the authority to enact binding laws; rather, it can only facilitate discussions and encourage member states to adopt voluntary measures (Zakir et al).

**Conclusion: The Path Ahead**

With changes in warfare, humans are being pushed away from the center of the battlefield- both physically and cognitively. The expansion of arsenal to include autonomous weapons have raised profound moral questions on the future of warfare.

The absence of human oversight in decision-making creates major ethical dilemmas especially with regard to accountability where LAWs are concerned. Existing ethical frameworks- such as the Just War Theory- have been found to be somewhat redundant in this constantly evolving field and must be addressed to ensure that there remains some justice in warfare. The key idea here is the notion of '*meaningful human control*'- which is imperative in determining the level of human insight required in armed operations. This allows humans to intervene and make significant decisions- thus shifting the agency, accountability and control back on humans. Delegating the life-and-death decision to machines undermines clear principles of human dignity which have been a focal aspect of traditionally governed warfare; and must remain so in the time to come.

The Actor-Network Theory (ANT) becomes vital in understanding this changing dynamic. As modern warfare moves beyond human agency and actors, it becomes crucial to incorporate non-human factors such as ML, AWS and algorithms, into studying this social phenomenon. Especially whether LAWs are concerned- and critical decisions are made independent of human

input- it becomes all the more fundamental to understand the interaction of human and non-human actors. Technology is no longer just a passive tool- rather, it plays an active role in shaping military strategies, conflict management and in subsequent geopolitical relations

There is a pressing need to hold discussions, as is underscored by the geopolitical landscape, where nations are rapidly advancing their military AI capabilities. The potential for an arms race driven by AI technologies raises concerns about the implications for global security and stability. The UN's call for a legally binding treaty by 2026 reflects the need to address these challenges before the proliferation of LAWS becomes widespread (UN).

Addressing the ethical implications of using AI in warfare requires interdisciplinary conversations involving technologists, lawyers, ethicists, and social scientists. These discussions are essential to develop comprehensive ethical frameworks and regulations that govern the use of AI in military operations. There is a need to ensure laws which will regulate LAWs and these cannot come at the expense of human life and dignity.

At the end of the day, one must accept that AI is here to stay. The challenge is to use this technology in more precise and humane ways without losing humanity in the process.

**Bibliography**


"Just War Theory." *Internet Encyclopedia of Philosophy*, University of Tennessee at Martin, www.iep.utm.edu/justwar/.

*A New Era of Conflict and Violence | United Nations*. (n.d.). United Nations. www.un.org/en/un75/new-era-conflict-and-violence

Arkin, Ronald C. "The Case for Ethical Autonomy in Unmanned Systems." PhilPapers, 2023, philpapers.org/rec/ARKTCF.

Armitage, Richard. "We must oppose lethal autonomous weapons systems." *British journal of general practice* 69.687 (2019): 510-511.

Army University Press. "The Ethics of Robots in War." *NCO Journal*, Feb. 2024, www.armyupress.army.mil/Journals/NCO-Journal/Archives/2024/February/The-Ethics-of-Robots-in-War/

Bode, Ingvild, and Guangyu Qiao-Franco. "The geopolitics of AI in warfare: contested conceptions of human control." *Handbook on Public Policy and Artificial Intelligence*. Edward Elgar Publishing, 2024. 281-294.

Borger, J. (2024, June 7). Guterres warns humanity on 'knife's edge' as AI raises nuclear war threat. *The Guardian*. www.theguardian.com/world/article/2024/jun/07/ai-nuclear-war-threat-un-secretary-general

Business Insider. "Israel's 'Where's Daddy?' AI System Helps Target Suspected Hamas Militants When They're at Home with Their Families." Business Insider, 2024, www.businessinsider.in/international/news/israels-wheres-daddy-ai-system-helps-

target-suspected-hamas-militants-when-theyre-at-home-with-their-families-report-says/articleshow/109107145.cms.

Carnegie Endowment for International Peace. "Governing Military AI amid a Geopolitical Minefield." Carnegie Endowment, July 2024, carnegieendowment.org/research/2024/07/governing-military-ai-amid-a-geopolitical-minefield?lang=en&center=europe.

Cederman, Lars-Erik, and Yannick Pengl. "Global conflict trends and their consequences." *United Nations, Department of Economic and Social Affairs. ETH Zürich, Zürich* (2019).

Center for a New American Security. "Military Artificial Intelligence, the People's Liberation Army, and U.S.-China Strategic Competition." CNAS, 2024, www.cnas.org/publications/congressional-testimony/military-artificial-intelligence-the-peoples-liberation-army-and-u-s-china-strategic-competition.

Collins, Christopher, et al. "Artificial intelligence in information systems research: A systematic literature review and research agenda." *International Journal of Information Management* 60 (2021): 102383.

Copeland, B.J. "Artificial intelligence". *Encyclopedia Britannica*, 27 Sep. 2024, www.britannica.com/technology/artificial-intelligence.

Dillon, Michael. *The Complications in Weaponized Artificial Intelligence for War*. MS thesis. Utica College, 2021.

Duan, Yanqing, John S. Edwards, and Yogesh K. Dwivedi. "Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda." *International journal of information management* 48 (2019): 63-71.

Egeland, Kjølv. "Lethal autonomous weapon systems under international humanitarian law." *Nordic journal of international law* 85.2 (2016): 89-118.

Egon, Kaledio, et al. "AI and national security: The geopolitical implications of autonomous weapons and cyber-security." (2023).

Elzinga, R. "Robotic and Its Implications on Future Wars." AIP Conference Proceedings, vol. 2814, no. 1, 2023, p. 040011, pubs.aip.org/aip/acp/article-abstract/2814/1/040011/2901934/Robotic-and-its-implications-on-future-wars.

Emery, John R. "Algorithms, AI, and Ethics of War." *Peace Review* 33.2 (2021): 205-212.

Garbe, Markus, et al. "Ethical Dimensions of Military Artificial Intelligence." Scandinavian Journal of Military Studies, vol. 2, no. 1, 2019, sjms.nu/articles/10.31374/sjms.26.

Garcia, Denise. "Lethal artificial intelligence and change: The future of international peace and security." *International Studies Review* 20.2 (2018): 334-341.

Gibbons-Neff, Thomas. "How Artificial Intelligence Could Change Warfare." The New York Times, 6 Feb. 2020, www.nytimes.com/2020/02/06/world/military-artificial-intelligence.html.

Gigova, Elena. "Putin: Artificial Intelligence 'Will Rule the World.'" *CNN*, 1 Sept. 2017, www.edition.cnn.com/2017/09/01/world/putin-artificial-intelligence-will-rule-world/index.html.

Guardan, A. "Israel, Gaza, and AI: A Database of Hamas Airstrikes." The Guardian, 3 Apr. 2024, www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes.

Gutiérrez, Jorge Luis Morton. "On actor-network theory and algorithms: ChatGPT and the new power relationships in the age of AI." *AI and Ethics* (2023): 1-14.

Haaretz. "Suicide Drones and AI: With Gaza as a Testing Ground, Israeli Defense Startups Flourish." Haaretz, 3 Jan. 2024, www.haaretz.com/israel-news/2024-01-03/ty-article-magazine/.premium/suicide-drones-and-ai-with-gaza-as-testing-ground-israeli-defense-startups-flourish/0000018c-cf39-ddba-abad-cfb9a3ee0000.

Hasan, Aamir, and Mohamed Buheji. "A World Losing Its Legitimacy-Gaza from Collective Punish till Ethnic Cleansing & Genocide." *International Journal of Management (IJM)* 15.1 (2024): 2024.

Hunter, Lance Y., et al. "Artificial intelligence and information warfare in major power states: how the US, China, and Russia are using artificial intelligence in their information warfare and influence operations." *Defense & Security Analysis* (2024): 1-35.

Hynek, Nik, and Anzhelika Solovyeva. "Operations of power in autonomous weapon systems: ethical conditions and socio-political prospects." *AI & SOCIETY* 36 (2021): 79-99.

Ilyas, Muhammad Derfish. "Implications and Analysis of the Crime of Aggression against Gaza: Breach of International Criminal Law amid the Israel-Hamas Conflict." *International Journal of Business and Social Analytics* 1.1 (2024).

International Committee of the Red Cross. "The Road Less Travelled: Ethics in the International Regulatory Debate on Autonomous Weapon Systems." ICRC Blogs, 25 Apr. 2024, blogs.icrc.org/law-and-policy/2024/04/25/the-road-less-travelled-ethics-in-the-international-regulatory-debate-on-autonomous-weapon-systems/.

Johnson, James. "Artificial intelligence & future warfare: implications for international security." *Defense & Security Analysis* 35.2 (2019): 147-169.

Johnson, James. "The AI commander problem: Ethical, political, and psychological dilemmas of human-machine interactions in AI-enabled warfare." *Journal of Military Ethics* 21.3-4 (2022): 246-271.

Kania, Elsa B. "'AI Weapons' in China's Military Innovation." *Brookings Institution, April* 3 (2020).

Kostenko, O. V. "The probability of military aggression of autonomous AI: assumptions or imminent reality (analyzing the facts of Russian war against Ukraine)." *РЕДАКЦІЙНА КОЛЕГІЯ* (2022): 179.

Kreps, S. E., and A. Haas. "AI and Hybrid Warfare: The Implications of Artificial Intelligence on Modern Conflict." International Security, vol. 44, no. 1, 2019, pp. 54-87, www.jstor.org/stable/27199968.

Kriegler, S. *Artificial Intelligence Guided Battle Management: Enabling Convergence in Multi-Domain Operations*. Diss. Master's thesis, School of Advanced Military Studies, US Army Command and General Staff College, Fort Levenworth, Kansas, 2020.

Kumar, Vijay. "Digital enablers." *The Economic Value of Digital Disruption: A Holistic Assessment for CXOs*. Singapore: Springer Nature Singapore, 2023. 1-110.

Lin, Patrick, et al. "Autonomous Military Robotics: Risk, Ethics, and Design." International Journal of Advanced Robotic Systems, vol. 7, no. 1, 2010, www.researchgate.net/publication/45421838_Autonomous_Military_Robotics_Risk_Ethics_and_Design.

Liu, Xia. "Civilians at war: focusing on the Israeli-Palestinian conflict and scrutinizing the doctrine of non-combatant immunity." *Frontiers in Political Science* 6 (2024): 1400307.

McKirdy, Euan. "The Ethics of AI in Warfare: What's at Stake?" CNN, 15 June 2021, www.cnn.com/2021/06/15/world/ai-warfare-ethics-intl-hnk/index.html.

Milley, Mark A., and Eric Schmidt. "America Isn't Ready for the Wars of the Future: And They're Already Here." *Foreign Affairs.* 103 (2024): 26.

Motwani, Nishank. "The Danger of AI in War: It Doesn't Care About Self-Preservation." *The Strategist*, 30 Aug. 2024, www.aspistrategist.org.au/the-danger-of-ai-in-war-it-doesnt-care-about-self-preservation/#:~:text=The%20reliance%20on%20AI%20could,inefficiencies%20associated%20with%20human%20oversight

Mueller, Hannes, et al. "Monitoring war destruction from space using machine learning." *Proceedings of the national academy of sciences* 118.23 (2021): e2025400118.

NATO Review. "Hybrid Warfare: New Threats, Complexity, and Trust as the Antidote." NATO Review, 30 Nov. 2021, www.nato.int/docu/review/articles/2021/11/30/hybrid-warfare-new-threats-complexity-and-trust-as-the-antidote/index.html.

NATO Review. "Hybrid Warfare: New Threats, Complexity, and Trust as the Antidote." NATO Review, 30 Nov. 2021, www.nato.int/docu/review/articles/2021/11/30/hybrid-warfare-new-threats-complexity-and-trust-as-the-antidote/index.html.

New Jersey Institute of Technology. "AI-Powered Exoskeleton Enhances Human Locomotion, Helps Restore Mobility." NJIT News, 2023, news.njit.edu/ai-powered-exoskeleton-enhances-human-locomotion-helps-restores-mobility.

Nijim, Mohammed. "Genocide in Palestine: Gaza as a case study." *The International Journal of Human Rights* 27.1 (2023): 165-200.

Patel, Ronak B., and Frederick M. Burkle. "Rapid urbanization and the growing threat of violence and conflict: a 21st century crisis." *Prehospital and disaster medicine* 27.2 (2012): 194-197.

Quincy Institute for Responsible Statecraft. "The Dystopian Future of AI Warfare." Quincy Institute, 7 June 2024, quincyinst.org/2024/06/07/the-dystopian-future-of-ai-warfare/.

ReliefWeb. "Gaza: Israel Systematically Uses Quadcopters to Kill Palestinians at Close Distance." ReliefWeb, 2024, reliefweb.int/report/occupied-palestinian-territory/gaza-israel-systematically-uses-quadcopters-kill-palestinians-close-distance-enar.

Risk Intelligence. "AI Tools in Hybrid Warfare: A Double-Edged Sword." Risk Intelligence, n.d., www.riskintelligence.eu/background-and-guides/ai-tools-in-hybrid-warfare-a-double-edged-sword.

Rosendorf, Ondřej. "Predictors of support for a ban on killer robots: Preventive arms control as an anticipatory response to military innovation." *Contemporary Security Policy* 42.1 (2021): 30-52.

Schmidt, Eric. "AI, great power competition & national security." *Daedalus* 151.2 (2022): 288-298.

Schwab, Klaus. "The Fourth Industrial Revolution: What It Means and How to Respond." *World Economic Forum*, 14 Jan. 2016, www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/#:~:text=Klaus%20Schwab&text=We%20stand%20on%20the%20brink,and%20relate%20to%20one%20another.

Sharkey, Amanda. "Autonomous weapons systems, killer robots and human dignity." *Ethics and Information Technology* 21.2 (2019): 75-87.

Sharma, Rohit Kumar. "Human-in-the-loop Dilemmas." *Journal of Defence Studies* 18.2 (2024): 166-175.

Singer, P. W., and August Cole. Ghost Fleet: A Novel of the Next World War. Eamon Dolan Books, 2015.

Song, Liang, et al. "Networking systems of AI: On the convergence of computing and communications." *IEEE Internet of Things Journal* 9.20 (2022): 20352-20381.

Sweeney, J. "Clausewitz, War, Peace and Politics." E-International Relations, 26 Nov. 2009, www.e-ir.info/2009/11/26/clausewitz-war-peace-and-politics/.

Thurnher, John. "The Future of Warfare: Artificial Intelligence and the Weaponization of Robots." Strategic Studies Quarterly, vol. 14, no. 1, 2020, pp. 16-37, www.airuniversity.af.edu/ssq/online/article/1585/the-future-of-warfare-artificial-intelligence-and-the-weaponization-of-robots.

Zakir, Muhammad Hamza, et al. "The Challenges and Opportunities of International Law in Regulating Emerging Technologies: A case of Artificial Intelligence and Autonomous Weapons." *Pakistan JL Analysis & Wisdom* 3 (2024): 242.

# THE ROLE OF ARTIFICIAL INTELLIGENCE ON EXCHANGE TRADED FUND INVESTMENTS

*Midhun Johny George*
*Research Scholar, Department of Commerce*
*Fatima Mata National College, (Autonomous) Kollam*
*University of Kerala*

*Dr. Sajeeve V P*
*Professor, Department of Commerce*
*Bishop Moore College, Mavelikara*
*University of Kerala*

**Abstract:** Exchange Traded Funds (ETFs) are instruments in the financial market which help the investors to pool their money to a fund that invests in shares, bonds, or commodities like gold. These funds represent a group of securities which enable the investors to track an underlying asset in the capital market and possess the features of both equity trading and mutual fund trading. The integration of Artificial Intelligence (AI) into the financial sector empowers investors to personalize their investment strategies by utilizing AI-based tools that assess market trends, forecast future movements, and design portfolios tailored to individual goals and preferences. AI has revolutionised the trading in stock markets by predicting the future trends or prices, analysing market news, and optimizing portfolios in the financial securities like ETFs. AI plays an important role in the financial sector by minimizing the operational costs and risks, fraud detection, improving loan and credit decisions, improving regulatory compliances, and providing better customer satisfaction. By the advent of Robo-advisors, automated portfolio in ETFs can be made according to the preferences of the investors and helps to provide algorithm-driven investment services and financial planning that avoids human supervision.

**Keywords:** Exchange Traded Fund, Artificial Intelligence, Robo-advisor, Asset Management, Passive investment

## Introduction

Exchange-Traded Funds (ETFs) have emerged as a key element in contemporary investment strategies due to their high liquidity, low costs, and ability to offer diversification. ETFs are investment funds that can be bought and sold on stock exchanges much like individual stocks. These funds consist of various assets such as stocks, bonds, or commodities. Since their introduction in the early 1990s, ETFs have seen substantial growth in both popularity and complexity, evolving to cater to the varying needs of investors (Madhavan 31).

Artificial Intelligence (AI), which enables machines to perform tasks that usually require human intelligence, such as learning, reasoning, and problem-solving, is increasingly permeating

various sectors, including finance. In financial markets, AI technology is used to analyse large amounts of data, predict market trends, and optimize trading strategies (Hendershott et al. 5).

The convergence of AI and ETF investments marks a major advancement in the financial sector. AI improves ETF management by utilizing sophisticated data analytics, predictive modelling, and algorithmic trading, which can enhance decision-making and potentially boost returns (Chen et al. 146). However, the adoption of AI also brings up critical concerns about market efficiency, transparency, and regulatory supervision (Arora et al. 544). This paper seeks to investigate the diverse role AI plays in ETF investments, considering both the opportunities and challenges presented by this innovation.

**Review of Literature**

Since their introduction in the early 1990s, Exchange-Traded Funds (ETFs) undergo significant evolution. Initially designed as a cost-effective and liquid alternative to mutual funds, ETFs now encompass a broad range of asset classes, investment strategies, and global markets (Madhavan 33). Recent advancements include thematic ETFs that focus on specific trends or sectors, and actively managed ETFs that aim to outperform traditional index-tracking ETFs. This evolution reflects the growing sophistication of ETF products and their adaptation to investor needs.

Artificial Intelligence (AI) rapidly transforms the financial industry by offering advanced tools for data analysis, predictive modelling, and decision-making. Technologies like machine learning and natural language processing enable more precise forecasting and improve the efficiency of trading algorithms (Hendershott et al. 22). In finance, AI applications include automated trading systems and sentiment analysis, which evaluates market sentiment based on news and social media (Chen et al. 153).

The integration of AI into ETF management signifies a major advancement. AI-powered algorithmic trading enables the execution of high-frequency trades with a level of precision and speed that human traders cannot match (Agarwal et al. 58). Machine learning algorithms analyse extensive datasets to uncover patterns and forecast ETF performance, aiding in portfolio optimization (Kumar and Tiwari 30). Furthermore, AI-driven tools adjust ETF holdings dynamically according to real-time market data and trends, potentially boosting returns and reducing risks (Arora et al. 547)

Numerous studies reveal the effects of AI on ETF performance and operational efficiency. For instance, trading algorithms enhanced by AI lower trading costs and boost liquidity (Hendershott et al. 8). Moreover, AI-driven predictive analytics improve the accuracy of forecasting ETF price changes compared to conventional methods (Chen et al. 155). Nonetheless, the advantages of AI-based strategies can depend significantly on the market environment and the particular AI methods used (Jiang and Xu 991).

Despite its benefits, incorporating AI into ETF investments also brings challenges. A significant concern is the risk of market manipulation through high-frequency trading algorithms, which can lead to heightened volatility and unfair practices (Agarwal et al. 62). Furthermore, the opaque nature of many AI systems creates transparency issues, making it challenging for investors and regulators to grasp the decision-making processes behind automated trading approaches (Arora et al. 548). Regulatory frameworks are evolving to tackle these concerns, and ongoing discussions continue about the appropriate level of oversight for AI-driven financial technologies (Zhang et al. 103).

Looking forward, AI's role in ETF investments is set to keep evolving. Progress in AI technologies, including better algorithms and greater computational power, is anticipated to further refine AI-driven ETF strategies. Moreover, combining AI with other emerging technologies, such as blockchain, may open up new avenues for innovation in ETF management and trading (Zhang et al. 108).

**Research gap**

Studies regarding the use of AI in ETFs remains scarce and there exists need for further research on the applications, benefits and risks of AI and Robo-advisors on ETFs. Investigating these areas will lead to a deeper insight into AI's potential to transform ETF management, while also shedding light on the potential challenges and risks as AI technology becomes more integrated into the financial industry.

**Objectives of the Study**

1. To identify the application of AI technologies in ETF investments
2. To assess the role of Robo-advisors in ETF investments

**Methodology**

This study uses descriptive data analysis and has collected data from various secondary data sources like journals, websites, and articles.

**AI Technologies and its Applications in ETFs**

Artificial Intelligence (AI) technologies are transforming the financial sector, especially in the area of Exchange-Traded Funds (ETFs). By utilizing different AI methods, financial institutions and investors are advancing the management, trading, and optimization of ETFs. This paper examines essential AI technologies and their uses in ETFs, focusing on how they enhance investment approaches, risk management, and operational effectiveness.

*Machine learning (ML)*

Machine learning branch of AI, involves algorithms that allow computers to learn from data and make predictions. In the realm of ETFs, ML algorithms are employed to analyse historical market

data, detect trends, and forecast future market behaviour. For example, ML algorithms can project ETF performance by examining previous price movements, trading volumes, and economic indicators (Chen et al. 150). These predictive models assist investors in making well-informed choices about which ETFs to purchase or sell. ML also supports algorithmic trading, where high-frequency trading systems leverage ML models to execute trades at optimal moments based on current data (Kumar and Kumar 78).

*Natural Language Processing (NLP)*

It is another AI technology focused on enabling interactions between computers and human language. NLP is utilized to examine news articles, financial reports, and social media to assess market sentiment and extract pertinent information. In ETF investing, NLP tools process large volumes of unstructured text data to offer insights into market conditions and investor sentiment (Jain et al. 220). For instance, NLP algorithms can analyse news headlines and financial statements to identify elements that might influence ETF prices. This analysis aids investors in grasping market trends and making more informed investment choices.

*Robotic Process Automation (RPA)*

RPA involves using software robots to handle repetitive and rule-based tasks. In the ETF sector, RPA is utilized to streamline operational functions such as executing trades, performing compliance checks, and generating reports. RPA tools automate ETF portfolio rebalancing by carrying out trades according to established rules (Singh and Sharma 55). This automation minimizes manual errors, boosts operational efficiency, and helps maintain portfolio alignment with investment goals. Additionally, RPA enhances reporting and compliance accuracy by automating data collection and verification processes.

*Predictive analytics and Forecasting*

This employs statistical methods and machine learning algorithms to anticipate future events based on past data. In the realm of ETFs, predictive analytics is used to project market trends, estimate ETF returns, and evaluate the influence of different factors on ETF performance. Predictive models examine historical price trends, macroeconomic indicators, and sector performance to forecast future ETF movements (Smith et al. 112). These insights assist investors in spotting potential investment opportunities and making strategic choices regarding ETF allocations. Predictive analytics also aids in risk management by anticipating potential risks and allowing investors to adjust their ETF holdings as needed. AI contributes to predictive analytics by employing machine learning models to forecast ETF performance. Methods like regression analysis and time series forecasting enable more precise predictions of price fluctuations (Chen et al. 152).

*Portfolio optimization*

This involves using AI algorithms to allocate assets in a manner that maximizes returns while reducing risk. In ETF investments, techniques for portfolio optimization assist in creating and managing diversified ETF portfolios. AI-powered optimization models assess factors such as asset correlations, past performance, and market conditions to suggest ideal ETF allocations (Brown and Johnson 90). These models ensure that portfolios remain balanced and in line with investors' risk tolerance and investment objectives. By continually refining the optimization process, AI supports the maintenance of a robust and effective investment strategy.

*Sentiment analysis*

Sentiment analysis uses AI to assess the mood or sentiment of market participants from textual data. This technology evaluates social media posts, news articles, and financial reports to gauge investor sentiment and its potential effects on ETF prices. AI-driven sentiment analysis tools offer important insights into market sentiment, aiding investors in understanding how news and events may impact ETF performance (Lee et al. 95). By incorporating sentiment analysis into their investment strategies, investors can make better-informed decisions about ETFs and adjust their portfolios according to current market conditions.

AI-driven sentiment analysis involves assessing market sentiment by analysing news, social media, and other textual sources (Hendershott et al. 10). AI systems process and interpret this information to understand investor sentiment and its potential effects on ETF prices. By combining sentiment analysis with conventional financial metrics, AI offers a more thorough perspective on market conditions (Hendershott et al. 11).

*Algorithmic trading*

It involves using computer algorithms to execute trading orders based on set criteria. AI-powered algorithms analyse market data, identify trading opportunities, and carry out trades with accuracy and speed. In ETF investing, algorithmic trading optimizes strategies by processing large data sets and executing trades at the best possible times (Kumar and Kumar 78). For example, AI algorithms evaluate historical price trends, trading volumes, and other market signals to determine ideal moments to buy or sell ETFs. This high-frequency trading method enables investors to exploit short-term market inefficiencies and achieve improved execution prices (Chen et al. 157).

Additionally, algorithmic trading algorithms adjust to evolving market conditions by learning from new data and modifying trading strategies as needed. This flexibility enhances ETF trading efficiency and improves the management of market risks (Ghosh 45). AI-driven algorithmic trading transforms ETF trading by facilitating high-frequency, automated transactions (Agarwal et al. 63). These systems handle enormous data volumes and execute trades at speeds surpassing human abilities, enhancing market efficiency and liquidity (Hendershott et al. 15). For instance,

AI-powered algorithms conduct intricate calculations to refine trade execution, lowering transaction costs and minimizing market impact (Agarwal et al. 66).

**Benefits of AI to ETF investment approaches**

This section examines AI's growing influence on ETFs, highlighting its role in data analysis, portfolio management, rebalancing, and the emergence of smart beta strategies.

*Improving Data Analysis and Market Evaluation*

A key advantage of AI in ETF investing is its capacity to handle large volumes of structured and unstructured data in real time. Conventional investment methods typically depend on historical data and human expertise. In contrast, AI leverages machine learning (ML) algorithms to evaluate multiple data sources, such as financial statements, news updates, and even social media sentiment (Jain and Jain 23). By processing this wide array of information, AI identifies patterns and market trends that human analysts might miss.

For example, AI-driven ETFs like the AI-Powered Equity ETF (AIEQ) utilize natural language processing and machine learning algorithms to continuously adjust portfolios in response to current market conditions (Ferguson and Seagle 177). These algorithms forecast stock movements by evaluating various factors, including corporate earnings, macroeconomic signals, and geopolitical risks. AI's capacity to analyse and integrate vast amounts of data provides fund managers with deeper insights, enabling more informed decisions and potentially improving investment results.

*Enhancing Portfolio Management and Rebalancing*

Rebalancing portfolios is essential in ETF management, and AI improves this process by automating it and refining the timing. Typically, ETFs need periodic adjustments to stay in line with their target index or strategy, but AI elevates this by constantly tracking market fluctuations and making real-time changes. This adaptive rebalancing enhances asset allocation efficiency while also minimizing transaction costs (Calvano et al. 225).

AI forecasts which sectors or industries are expected to outperform by analysing both historical and current market data. Through automation, AI keeps the rebalancing process efficient, ensuring that ETFs stay aligned with their goals, whether tracking a broad market index or focusing on particular sectors or factors (Bender et al. 118). This dynamic, AI-driven rebalancing enables ETFs to react swiftly to market fluctuations, resulting in improved performance and reduced costs.

AI plays a crucial role in portfolio management by refining asset allocation and managing ETF investments in real time. Sophisticated algorithms evaluate current market data to modify portfolio holdings, with the goal of maximizing returns and reducing risk (Kumar and Tiwari

31). AI-powered tools offer investors customized strategies aligned with their risk preferences and investment objectives, thereby improving portfolio outcomes.

*Improving Risk Management*

AI's contribution to risk management is a key factor in transforming ETF investments. AI-powered risk models evaluate complex datasets and detect early indicators of market declines or potential risks. These models account for various factors, such as interest rates, inflation figures, and global geopolitical developments, offering more thorough risk evaluations (Gupta and Aggarwal 72).

For instance, AI can simulate different risk scenarios and recommend changes to a fund's portfolio before major losses happen. This enables ETF managers to apply more proactive risk management strategies, like shifting assets to safer sectors or utilizing hedging methods (Bender et al. 123). By constantly scanning for potential risks, AI-powered ETFs can respond faster and more efficiently than funds managed by humans.

AI improves risk management in ETF investments by precisely identifying and quantifying risks. Machine learning models evaluate market volatility and possible risk elements, offering investors tools to manage risks effectively (Arora et al. 550). AI-powered risk management techniques assist in optimizing portfolios to achieve superior risk-adjusted returns (Arora et al. 552).

*AI-Driven Active Management*

ETFs have traditionally been linked to passive investment strategies, tracking indices like the SandP 500 or FTSE 100. However, AI is driving a shift towards more active management in the ETF industry. AI algorithms continuously analyse market data, enabling AI-powered ETFs to frequently modify portfolio compositions, a defining aspect of active management (Cornelli et al. 5).

AI-powered ETFs, like AIEQ, actively choose stocks based on forecasts of future market performance, rather than just following an index. This integration of AI with active management strategies allows ETFs to exploit market inefficiencies and short-term trends, providing investors with a more flexible investment approach while maintaining the cost-efficiency of ETFs (Ferguson and Seagle 182). The growing adoption of AI in active management is making ETFs more attractive to investors looking for data-driven, aggressive strategies.

*Advancing Smart Beta and Beyond*

The rise of "smart beta" ETFs illustrates how AI merges passive and active management approaches. These ETFs focus on specific factors like volatility, value, or momentum, rather than traditional indices. AI enhances these methods by continuously adjusting the factor weightings in the portfolio based on real-time market data and predictive analytics (Bender et al. 120).

For instance, AI identifies changes in market sentiment or trends that may impact factor-based investment strategies, enabling ETFs to adapt instantly. As AI advances, the future of ETF investing is expected to include more widespread use of AI-driven strategies across passive, active, and hybrid investment methods. AI is not only boosting the efficiency and cost-effectiveness of ETFs but also paving the way for new, personalized, adaptive investment solutions (Delis et al. 5).

*Efficiency and Speed*

AI greatly enhances trading efficiency and speed, facilitating the swift execution of trades and lowering transaction expenses (Agarwal et al. 69). Automated trading systems allow investors to seize market opportunities with remarkable speed, boosting liquidity and decreasing trading costs (Hendershott et al. 23).

*Cost Reduction*

AI aids in cost reduction by automating different facets of ETF management, including trading and portfolio adjustments. Automation decreases the necessity for manual supervision and lowers operational expenses compared to conventional investment methods (Agarwal et al. 70). This efficiency leads to reduced management fees and overall costs for investors (Arora et al. 556).

*Customization and Personalization*

AI allows for increased customization and personalization of ETF investment strategies. By evaluating individual investor profiles and preferences, AI systems can adjust ETF portfolios to fit specific investment objectives and risk tolerances (Kumar and Tiwari 34). This tailored approach improves investor satisfaction and aligns investments with personal needs.

**Challenges and Limitations**

*Algorithmic Trading Risks*

Despite its benefits, AI-powered algorithmic trading carries risks, such as market manipulation and heightened volatility (Agarwal et al. 75). High-frequency trading algorithms may trigger abrupt market shifts and flash crashes, leading to concerns about market stability.

*Transparency and Accountability*

The "black box" nature of AI presents challenges related to transparency and accountability. The complexity of AI algorithms often makes it difficult for investors and regulators to comprehend the decision-making process (Arora et al. 555). This opacity can erode trust in AI-based investment strategies.

*Regulatory and Ethical Considerations*

The incorporation of AI in ETF investments introduces both regulatory and ethical challenges. Existing regulations might not adequately cover the intricacies of AI-driven trading, prompting demands for revised frameworks (Zhang et al. 114). Ethical issues include the possibility of biased algorithms and the importance of maintaining fair market practices.

**Robo advisors and ETF**

Artificial intelligence (AI) is transforming the financial industry, especially in the use of robo-advisors for managing Exchange-Traded Funds (ETFs). These AI-driven tools provide automated financial planning services with little need for human intervention, relying on sophisticated algorithms.

Robo-advisors are digital tools that offer automated investment management services. They rely on algorithms to build and oversee investment portfolios, taking into account personal preferences and risk levels. AI plays a key role in how these platforms operate, allowing them to process large data sets, forecast market movements, and execute trades accurately (Bender et al. 118).

By using machine learning, AI algorithms in robo-advisors adapt to both shifting market dynamics and the unique needs of investors. This adaptability enables robo-advisors to provide customized ETF strategies, a service once exclusive to traditional financial advisors. The incorporation of AI in these platforms improves efficiency, lowers costs, and broadens access to advanced investment strategies (Ferguson and Seagle 175). The benefits of robo-advisors on ETF trading are:

*Enhancing Accessibility and Reducing Costs*

A key benefit of AI-powered robo-advisors is their ability to reduce barriers to ETF investing. Traditional financial advisory services typically require high minimum investments and charge significant management fees. In contrast, robo-advisors provide low-cost, easily accessible investment options with minimal entry requirements (Jain and Jain 23).

By leveraging AI algorithms, robo-advisors deliver personalized investment recommendations at a fraction of the cost compared to human advisors. For instance, platforms such as Betterment and Wealthfront utilize AI to evaluate investor profiles and suggest ETFs based on factors like risk tolerance, financial goals, and time frames (Calvano et al. 225). This cost-effectiveness and ease of access open up ETF investing to a wider audience, including those with smaller amounts to invest.

*Customizing Investment Strategies*

AI-driven robo-advisors excel at customizing investment strategies. By evaluating a client's financial status, objectives, and risk tolerance, AI algorithms can design personalized ETF portfolios suited to individual requirements. This process includes a comprehensive assessment of factors such as income, expenses, investment targets, and risk preferences (Gupta and Aggarwal 72).

For example, robo-advisors utilize AI to categorize investors based on risk levels and suggest ETF portfolios that match their specific risk profiles. This level of customization is achieved through advanced algorithms that analyse historical performance, market trends, and investor behaviour (Delis et al. 5). As a result, investors receive tailored recommendations that are more likely to align with their financial goals, compared to generalized approaches.

*Enhancing Portfolio Management and Rebalancing*

AI plays a pivotal role in optimizing portfolio management and rebalancing within robo-advisors. Successful portfolio management requires ongoing monitoring and adjustments to ensure investments remain in line with an investor's objectives and risk tolerance. AI algorithms automate these processes, making rebalancing more efficient and cost-effective (Bender et al. 120).

Robo-advisors leverage AI to monitor market changes and adjust ETF allocations in real time. This dynamic rebalancing ensures that portfolios maintain their intended asset allocation while addressing the risks posed by market fluctuations. For instance, if a specific ETF sector performs well, AI algorithms can automatically adjust investments to take advantage of the gains, while managing the risks from sectors that are underperforming (Ferguson and Seagle 182).

*Improving Risk Management*

AI-powered robo-advisors significantly enhance risk management. This involves recognizing, evaluating, and addressing potential investment risks. AI algorithms process extensive datasets, including market trends, economic indicators, and geopolitical events, to forecast risks and adjust ETF portfolios as needed (Gupta and Aggarwal 75).

For example, robo-advisors utilize AI to apply advanced risk models that simulate different market scenarios and evaluate their effects on ETF portfolios. This proactive strategy enables robo-advisors to implement risk reduction measures, such as reallocating investments to more secure assets or using hedging techniques (Delis et al. 8). By constantly monitoring and responding to risk factors, AI improves the stability and durability of ETF investments.

**The Future of AI in Robo-advisors for ETF investments**

The future of AI in robo-advisory services looks promising, with ongoing advancements expected to drive further innovations. As AI technology advances, robo-advisors are set to provide even

more tailored and flexible investment solutions. Upcoming developments might feature enhanced predictive analytics, refined machine learning algorithms, and integration with emerging technologies like blockchain (Cornelli et al. 6).

For instance, combining AI with blockchain technology could result in more transparent and secure investment processes. Blockchain's immutable ledger could improve the accuracy of financial transactions, while AI algorithms might analyse blockchain data to uncover investment opportunities and assess risks (Bender et al. 125). Such progress is likely to boost the capabilities of robo-advisors and broaden their role in ETF investments.

AI is profoundly reshaping the function of robo-advisors in ETF investments. By enhancing accessibility, customizing investment strategies, refining portfolio management, and improving risk management, AI-powered robo-advisors are making advanced investment solutions more accessible and affordable. As AI technology progresses, robo-advisory services are expected to experience even more innovations, providing investors with increasingly personalized and efficient methods for managing their ETF investments. This ongoing development in AI will continue to broaden access to sophisticated financial strategies, benefiting a wider spectrum of investors.

**Recommendations**

Investors should utilize AI tools to strengthen their ETF investment strategies, prioritizing technologies that boost trading efficiency and manage risk effectively. AI-powered personalized portfolios provide customized investment solutions that match individual objectives and risk preferences. Regulators need to create revised frameworks to tackle the specific challenges introduced by AI in financial markets. Upholding transparency, accountability, and fairness in AI-based trading practices is essential for preserving market integrity and safeguarding investors.

**Conclusion**

AI technologies are greatly improving the management and investment strategies related to ETFs. Machine learning, natural language processing, robotic process automation, predictive analytics, portfolio optimization, and sentiment analysis each play a role in making ETF investing more efficient and informed. These technologies enhance decision-making, refine portfolio management, and bolster risk management, giving investors advanced tools to navigate the complexities of financial markets. As AI advances, its use in ETFs is expected to grow, offering even more sophisticated solutions for investors and financial institutions. The integration of AI into ETF investment strategies marks a significant development in the industry, promising ongoing innovation and better investment results.

AI significantly reshapes ETF investments by increasing trading efficiency, refining risk management, and offering more precise performance predictions. However, incorporating AI

introduces challenges concerning transparency, regulation, and ethical considerations. As AI technologies advance, their influence on ETF investments is expected to expand, requiring continuous research and regulatory updates to tackle new issues.

**Bibliography**


Agarwal, A., Ghosh, A., and Zhang, Y. "High-Frequency Trading and Market Efficiency: Evidence from the ETF Market." *Journal of Financial Markets*, vol. 42, 2019, pp. 56-77.

Arora, V., Bansal, R., and K. Sethi. "Artificial Intelligence in Finance: Risks and Opportunities." *Journal of Financial Regulation and Compliance*, vol. 29, no. 4, 2021, pp. 543-558.

Bender, Jennifer, et al. "AI-Powered Equity: A New Frontier in Asset Management." *Journal of Investment Strategies*, vol. 12, no. 2, 2019, pp. 110-130.

Brown, James, and Emily Johnson. "AI-Driven Portfolio Optimization: Techniques and Applications." *Journal of Financial Analytics*, vol. 13, no. 4, 2022, pp. 85-95.

Calvano, Emilio, et al. "Artificial Intelligence, Algorithmic Trading, and Market Efficiency." *Journal of Financial Markets*, vol. 14, no. 3, 2020, pp. 215-230.

Chen, Wei, et al. "Machine Learning Applications in Financial Markets: A Review." *Financial Technology Review*, vol. 11, no. 2, 2021, pp. 145-160.

Cornelli, Francesca, et al. "The Future of Asset Management." *European Financial Management*, vol. 26, no. 1, 2020, pp. 1-23.

Delis, Manthos D., et al. "Machine Learning in Finance: The Case of Financial Crises." *Journal of Financial Stability*, vol. 54, 2021, pp. 1-11.

Ferguson, Roger, and Patrick Seagle. "AI in Financial Markets: Trends and Performance." *AI and Finance Review*, vol. 5, no. 1, 2020, pp. 175-188.

Ghosh, Arindam. "AI and the Democratization of ETF Investments in India." *Journal of Financial Technologies*, vol. 8, no. 2, 2021, pp. 43-55.

Gupta, Arvind, and Rakesh Aggarwal. "Predictive Analytics in Financial Markets: A Machine Learning Approach." *Finance Research Letters*, vol. 22, 2020, pp. 68-77.

Hendershott, Terrence, Jones, Charles M., and Albert J. Menkveld. "Does Algorithmic Trading Improve Liquidity?" *Journal of Finance*, vol. 66, no. 1, 2011, pp. 1-33.

Jain, Rajiv, et al. "Natural Language Processing in Financial Markets: An Overview." *Journal of Financial Data Science*, vol. 8, no. 1, 2022, pp. 215-230.

Jain, Sandeep, and Shweta Jain. "Artificial Intelligence in Financial Markets: Disruption and Opportunity." *Financial Review Journal*, vol. 44, no. 1, 2021, pp. 20-35.

Jiang, W., and S. Xu. "Machine Learning for Financial Forecasting: An Overview." *Quantitative Finance*, vol. 22, no. 6, 2022, pp. 987-1002.

Kumar, Ashok, and Ravi Kumar. "Algorithmic Trading and Machine Learning: Innovations in Financial Markets." *Finance Research Letters*, vol. 15, 2020, pp. 70-85.

Kumar, P., and A. Tiwari. "Predictive Analytics in ETF Trading: A Machine Learning Approach." *Financial Analysts Journal*, vol. 76, no. 4, 2020, pp. 23-36.

Lee, Hae-Young, et al. "Sentiment Analysis for Financial Markets: A Review of Techniques and Applications." *Journal of Financial Technology*, vol. 12, no. 3, 2021, pp. 85-100.

Madhavan, Ananth. "Exchange-Traded Funds and the New Dynamics of Investing." *Financial Analysts Journal*, vol. 72, no. 3, 2016, pp. 30-45.

Singh, Pradeep, and Neelam Sharma. "Robotic Process Automation in Financial Services: Opportunities and Challenges." *Financial Management Review*, vol. 17, no. 2, 2022, pp. 50-65.

Smith, Robert, et al. "Predictive Analytics in ETF Investments: Techniques and Applications." *Journal of Financial Economics*, vol. 16, no. 4, 2021, pp. 100-115.

Zhang, X., Li, W., and Wang, Y. "Blockchain and AI Integration: Transforming the ETF Landscape." *Journal of Financial Technology*, vol. 8, no. 1, 2023, pp. 102-118.

# A STUDY OF CYBER CRIMES IN SELECTED METROPOLITAN CITIES FOR THE PERIOD 2020 to 2022

*Vidya Konar*
*Assistant Professor, Department of Mathematics & Statistics*
*Sophia College for Women (Empowered Autonomous), Mumbai*

**Abstract:** Cybercrime is an umbrella term for illegal activity using electronic devices, networks and computers. It includes an extensive range of illicit acts carried out by various forms of electronic communication. These activities usually target people, organizations or governments with the goal to carry out fraud, identity theft, disrupt digital operations or have illegal access to sensitive data. This research paper analysis cybercrime in five metropolitan cities -Bengaluru, Delhi City, Kanpur, Kolkata and Mumbai - that are reported in the NCRB crime report for the period 2020-2022. The National Crime Records Bureau's (NCRB) official publication provides quantitative data on cybercrime in metropolitan cities. The aim of this paper are to evaluate the total number of offences carried out in the selected metropolitan cities under the Information Technology Act (I.T. Act), the total number of offences committed under the Indian Penal Code ( IPC) in the specified metropolitan cities, the total number of offences carried out under Special and Local Laws (SLL) in selected metropolitan cities, the total number of cybercrimes in the chosen metropolitan cities and to determine whether there is a significant difference in the average number of cybercrimes among the selected metropolitan cities. The statistical analysis used in the paper are diagrammatic presentation of the data, mean, standard deviation and ANOVA.

**Keywords:** ANOVA, cybercrime, diagrammatic representation, mean and standard deviation.

## Introduction

One of the greatest risk to modern society is cybercrime. While providing unparalleled accessibility and connectivity, the rapid development of technology has also led to an environment that is susceptible to different forms of cybercrime. Cybercrime is an illegal activity using computers or digital networks. It is a crime using a computer, network or networked device. Some of the cybercrimes are exceptionally skilled technically and adopt modern techniques. These acts tend to target individuals, companies or government organizations with the aim of obtaining identities and gaining unauthorized access to confidential information, tampering digital operations or committing fraud.

The different categories of cybercrime as classified by Indian Cybercrime Coordination Centre (I4C) are as follows: Cryptocurrency Crime, Cyber Tourism, Hacking/Damage to Computer Systems, Online and Social Media Related Crime, Online Financial Fraud, and

Publishing/Transmitting of Explicit Material in Electronic Form, Ransomware, Child Pornography/ Child sexually abusive material (CSAM).

The following is a list of cybercrimes compiled by National Cyber Crime Reporting Portal:

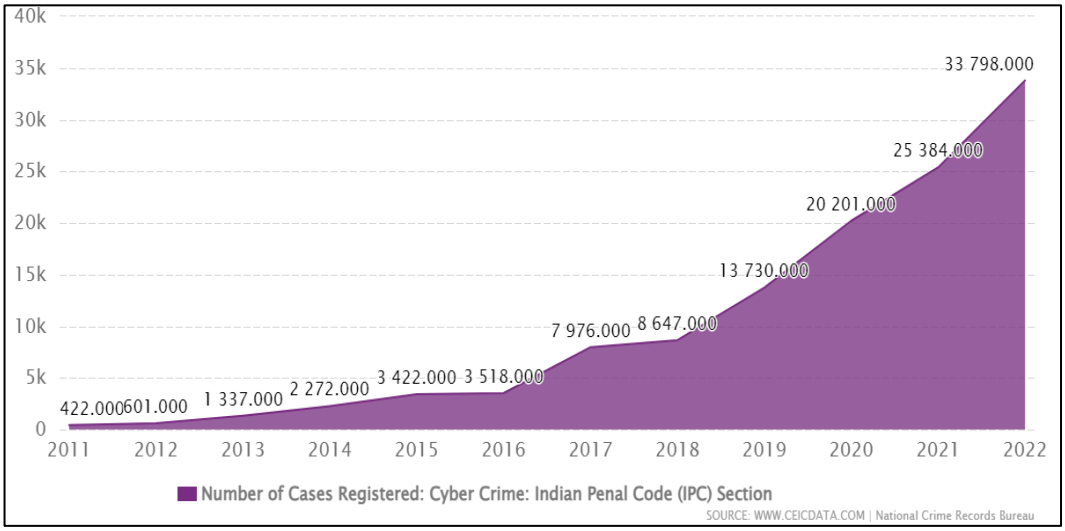
Child Pornography/ Child sexually abusive material (CSAM), Cyber Bullying, Cyber Stalking, Cyber Grooming, Online Job Fraud, Online Sextortion, Vishing, Sexting, Smishing, SIM Swap Scam, Debit/Credit Card Fraud, Impersonation and Identity Theft, Phishing, Spamming

Ransomware, Virus, Worms & Trojans, Data Breach, Denial of Services /Distributed DoS, Website Defacement, Cyber-Squatting, Pharming, Cryptojacking, Online Drug Trafficking, Espionage.

India's cybercrime rate is increasing rapidly due to the country's massive internet usage. Phishing, hacking, identity theft, cyberbullying, child sexual abuse material, online fraud are some of the common cybercrimes observed in India. The primary objective of the Information Technology Act (I.T. Act) of 2000 was to lay down the conditions for commercial use of IT in order to prevent cybercrimes. The I.T. Act spells out the offenses that are penalized. Moreover, cybercrimes are now covered by the Indian Penal Code, 1860. This Act is a key component of legislation regulating various aspects of cybercrime .e-commerce and digital governance, is mainly responsible for monitoring cybercrime in India. The Act defines a number of technology-related crimes and provides juridical force to transactions carried out digitally. Hacking, Data theft, Identity theft, Cyberstalking, Cyberbullying, sending offensive messages, cyber terrorism, phishing, spamming, copyright infringement are some of the key cybercrimes under this Act. While the IT Act of 2000 is the primary act in India to govern cybercrimes, these crimes are also covered by a number of sections of the Indian Penal Code (IPC). The most significant IPC sections concerning cybercrime are: Section 379 (Theft), Section 411: (Dishonestly receiving stolen property), Section 419 (Cheating by personation), Section 420(Cheating and dishonestly inducing delivery of property) Section 465(Forgery). Copyright Act (1957), Trademarks Act, Telegraph Act are some of the Special and Local Laws pertaining to cybercrime in India.

The National Crime Records Bureau (NCRB) says that 4850 cybercrimes in India in 2023 resulted in alarming loss of Rs.66.66cr. According to a latest study by the Indian Cybercrime Coordination Centre (I4C), over the last three years, digital financial fraudulent transactions have led to a shocking loss of Rs.1.25 lakh crore. In 2023, a minimum of Rs. 10,319 crore was reported lost by victims of digital financial fraud. In its report on "cyber security and rising incidents of cyber/white collar crimes", the Parliamentary Standing Committee on Finance revealed that the total amount of domestic fraud reported by the Supervising Entities in FY'23 was Rs.2537.35 cr. The report also states that 6.94 lakh complaints were filed in 2023 alone.

According to the I4C, in May 2024, an average of 7,000 cybercrime reports were lodged each day. This marks a drastic increase of 113.7% compared to 2021-22 and 60.9% increase over 2022-23. 85% of these complaints involved digital financial fraud.




Number of Cases Registered under IPC section from 2002 to 2022
Source: CEIC Data, National Crime Records Bureau

**Literature Review**

ChiChao Lu, WenYuan Jen, Weiping Chang, Shihchieh Chou's (2006)[1] paper on "Cybercrime & Cybercriminals: An Overview of the Taiwan Experience" examines the rise in cybercrime among Taiwanese students from 1999 to 2004. The study says that many students have utilized computer resources in criminal activities such as sex trading and internet fraud. The study reveals the requirement of awareness of cybercrime prevention among internet users.

Dr. Fatma Abdalla Mabrouk Khiralla (2020)[2] paper on "Statistics of Cybercrime from 2016 to the First Half of 2020" looks at the progress of cybercrime from 2016 to the first half of 2020. This paper analyses the statistics of cybercrime from 2016 to first half of 2020. The paper highlights the difference between cybersecurity and knowledge of cybersecurity.

**Objectives of the Study**

- To evaluate the total number of offences carried out in the selected metropolitan cities under the Information Technology Act (I.T. Act).
- To evaluate the total number of offences committed under the Indian Penal Code (IPC) in the specified metropolitan cities

- To evaluate the total number of offences carried out under Special and Local Laws (SLL) in selected metropolitan cities
- To evaluate the total number of cybercrimes in the chosen metropolitan cities
- To determine whether there is a significant difference in the average number of cybercrimes among the selected metropolitan cities.

**Research Methodology**

The paper attempts to study the total number of cybercrime cases reported in selected metropolitan cities under the Information Technology Act (I.T. Act), the Indian Penal Code (IPC) and Special and Local Laws (SLL) for the period 2020-2022. The paper also tests whether there is a significant difference in number of cybercrimes reported among the selected metropolitan cities for the period under study.

Sample Size: The sample selected for the study is five metropolitan cities - Bengaluru, Delhi city, Kanpur, Kolkata and Mumbai. Cybercrime records of these cities for a period of two years (2020-2022) have been collected from National Crime Records Bureau's (NCRB) website.

**Tools used for the study:**

1. Bar Diagrams: It is a way of expressing numbers through bar diagrams. It provides an in-depth understanding of the statistical data which makes visual comparison smoother. It performs a task of converting intricate concepts associated with numbers into an easy diagram form. The height of the bar is proportional to the value of the variable.

2. ANOVA: Analysis of variance (ANOVA) is a statistical test which is used to compare group means. It shows how far a particular variable's numerical values vary from the overall mean. It is used to test the differences between the population means by comparing the degree of variance within each sample to the degree of variance between the samples. The hypothesis that the means of two or more populations are equal is tested through ANOVA. In regression analysis, this technique is used to investigate how independent factors affect the dependent factor.

3. Arithmetic Mean: It is found by adding all the observations in the data set divided by the number of observations in the data. It is one of the measure of central tendency. It is a rigidly defined measure. It is based on all the observations. Hence it is affected by small or extremely large observations. It is easy to understand and calculate. It cannot be determined by inspection of the data or locate graphically. For unclassified data, the formula to compute mean is $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. For classified data, the formula to compute mean is $\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$

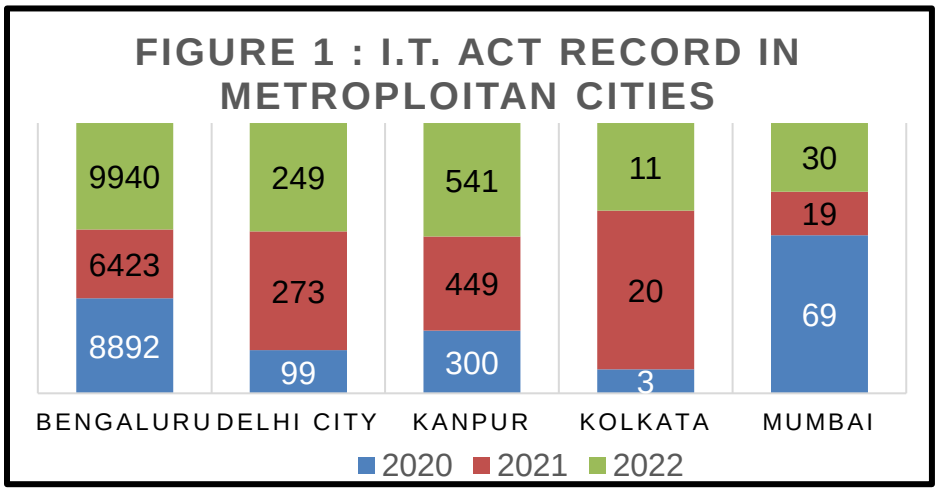
4. Standard Deviation: It is the square root of variance. It is one of the best measure of dispersion. It is based on each and every observation of the data. It is possible to find the combined variance and standard deviation for more than one group. It is affected by extreme observations. It cannot be calculated for qualitative data. For unclassified data, the formula to compute standard deviation is $= \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2}$ . For classified data, the formula to compute mean is $\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i-\overline{x})^2}{N}} = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2}{N} - \overline{x}^2}$

**Data Analysis and Findings**

**1. Study of the total number of offences recorded in the selected metropolitan cities under the Information Technology Act (I.T. Act or ITC)**

TABLE 1: Total number of offences recorded in the selected metropolitan cities under the Information Technology Act

| | **2020** | **2021** | **2022** |
|---|---|---|---|
| **BENGALURU** | 8892 | 6423 | 9940 |
| **DELHI CITY** | 99 | 273 | 249 |
| **KANPUR** | 300 | 449 | 541 |
| **KOLKATA** | 3 | 20 | 11 |
| **MUMBAI** | 69 | 19 | 30 |

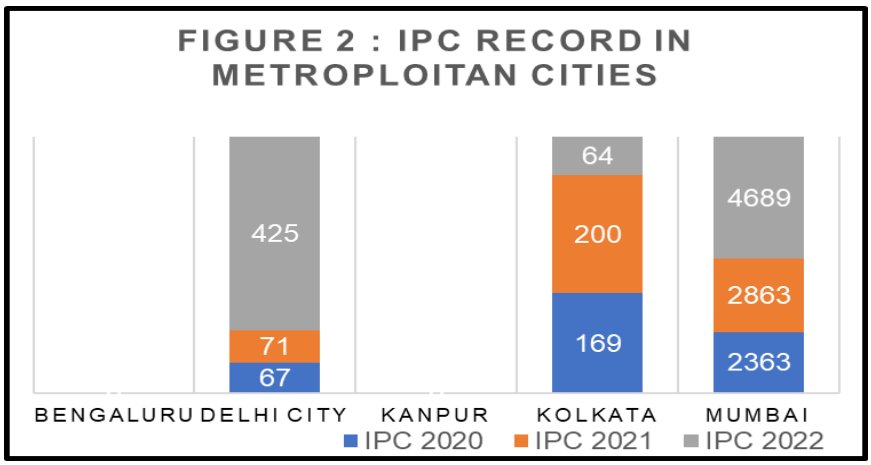


From Table 1 and Figure 1, we can see the total number of offences committed in five metropolitan cities under the Information Technology Act (I.T. Act) between 2020 and 2022. The crime figures In Bengaluru were 8892 in 2020, 6423 in 2021 and 9940 in 2022. The crime rate in Delhi City was 99 in 2020, 273 in 2021 and 249 in 2022. In Kanpur, the crime record was 300 in 2020, 449 in 2021 and 541 in 2022. In Kolkata, the crime record was 3 in 2020, 20 in 2021 and 11 in 2022. In Mumbai, the crime rate was 69 in 2020, 19 in 2021 and 30 in 2022. It has been observed that the city Bengaluru has the highest rate of crime in this category for the period under consideration.

**2. Study of the total number of offences reported in the selected metropolitan cities under the Indian Penal Code (IPC)**

TABLE 2: Total number of offences reported in the Indian Penal Code (IPC) in the selected metropolitan cities.

| | **2020** | **2021** | **2022** |
|---|---|---|---|
| **BENGALURU** | 0 | 0 | 0 |
| **DELHI CITY** | 67 | 71 | 425 |
| **KANPUR** | 0 | 0 | 0 |
| **KOLKATA** | 169 | 200 | 64 |
| **MUMBAI** | 2363 | 2863 | 4689 |

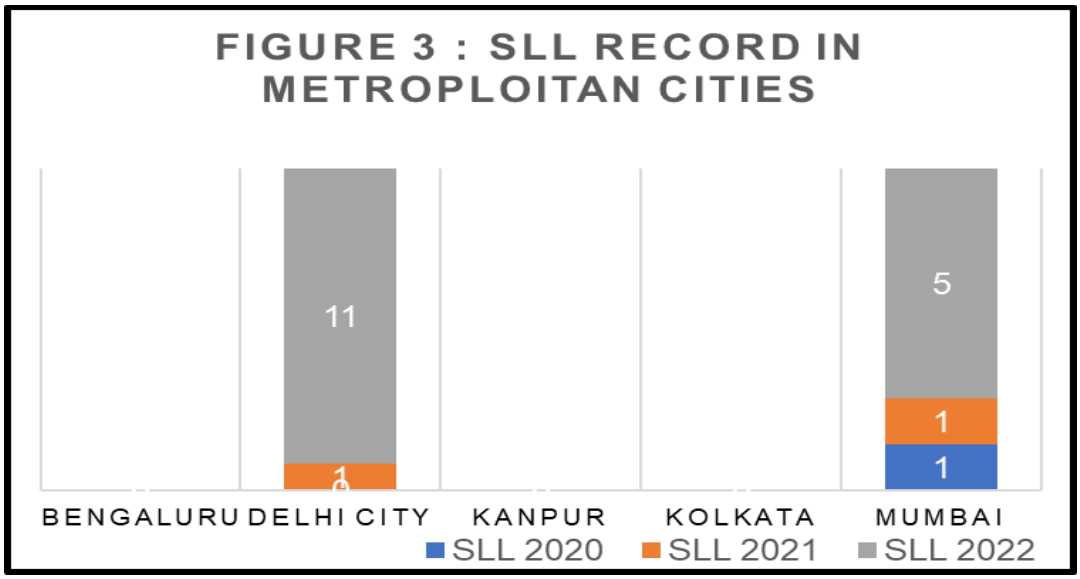


From Table 2 and Figure 2, we can see the total number of offences committed in five metropolitan cities under the IPC between 2020 and 2022. The crime statistics in Bengaluru and Kanpur is zero across the period under study. The crime rate in Delhi City was 67 in 2020, 71 in 2021 and 425 in 2022. In Kolkata, the crime record was 169 in 2020, 200 in 2021 and 64 in 2022. In Mumbai, the crime rate was 2363 in 2020, 2863 in 2021 and 4689 in 2022. It has been observed that the city Mumbai has the highest rate of crime in this category for the period under study.

**3. Study of the total number of offences carried out in the selected metropolitan cities under Special and Local Laws (SLL):**

TABLE 3: Total number of offences carried out under Special and Local Laws (SLL) in selected metropolitan cities

| | **2020** | **2021** | **2022** |
|---|---|---|---|
| **BENGALURU** | 0 | 0 | 0 |
| **DELHI CITY** | 0 | 1 | 11 |
| **KANPUR** | 0 | 0 | 0 |
| **KOLKATA** | 0 | 0 | 0 |
| **MUMBAI** | 1 | 1 | 5 |

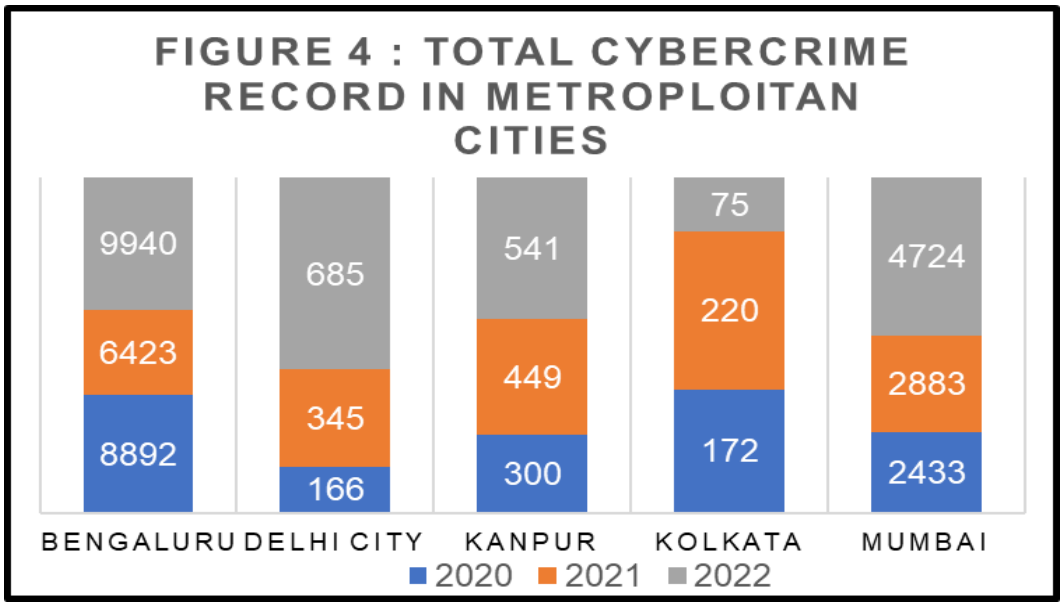


From Table 3 and Figure 3, we can see the total number of offences committed in five metropolitan cities under the SLL between 2020 and 2022. The crime statistics in Bengaluru, Kanpur and Kolkata is zero across the period under study. The crime rate in Delhi City was zero in 2020, 1 in 2021 and 11 in 2022. In Mumbai, the crime rate was 1 in 2020, 1 in 2021 and 5 in 2022. It has been observed that the crime rate under SLL is very low across all the selected metropolitan cities as compared to I.T. Act and IPC categories of cybercrime.

**4. Study of the total number of cybercrime recorded in the selected metropolitan cities**

TABLE 4: Total number of cybercrime recorded in the selected metropolitan cities

| | **2020** | **2021** | **2022** |
|---|---|---|---|
| **BENGALURU** | 8892 | 6423 | 9940 |
| **DELHI CITY** | 166 | 345 | 685 |
| **KANPUR** | 300 | 449 | 541 |
| **KOLKATA** | 172 | 220 | 75 |
| **MUMBAI** | 2433 | 2883 | 4724 |

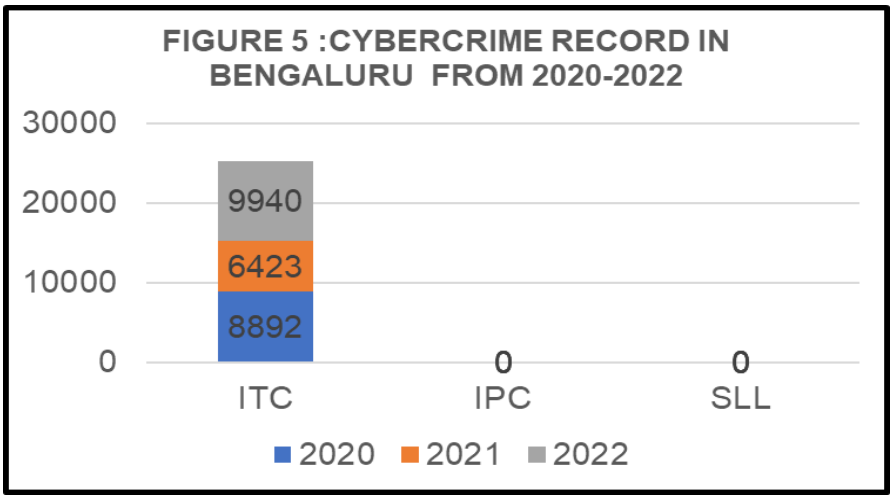


From Table 4 and Figure 4, we can see the total number of cybercrimes committed in five metropolitan cities between 2020 and 2022. The crime statistics in Bengaluru was 8892 in 2020, 6423 in 2021, 9940 in 2022. The crime rate in Delhi City was 166 in 2020, 345 in 2021 and 685 in 2022. In Kanpur, the crime record was 300 in 2020, 449 in 2021 and 541 in 2022. In Kolkata, the crime record was 172 in 2020, 220 in 2021 and 75 in 2022. In Mumbai, the crime rate was 2433 in 2020, 2883 in 2021 and 4724 in 2022. It has been observed that the city Bengaluru has the highest rate of cybercrime record for the period under study.

**5. Study of the total number of cybercrime reported in Bengaluru**

TABLE 5: Cybercrime record in Bengaluru

| | **I.T. Act** | **IPC** | **SLL** |
|---|---|---|---|
| **2020** | 8892 | 0 | 0 |
| **2021** | 6423 | 0 | 0 |
| **2022** | 9940 | 0 | 0 |

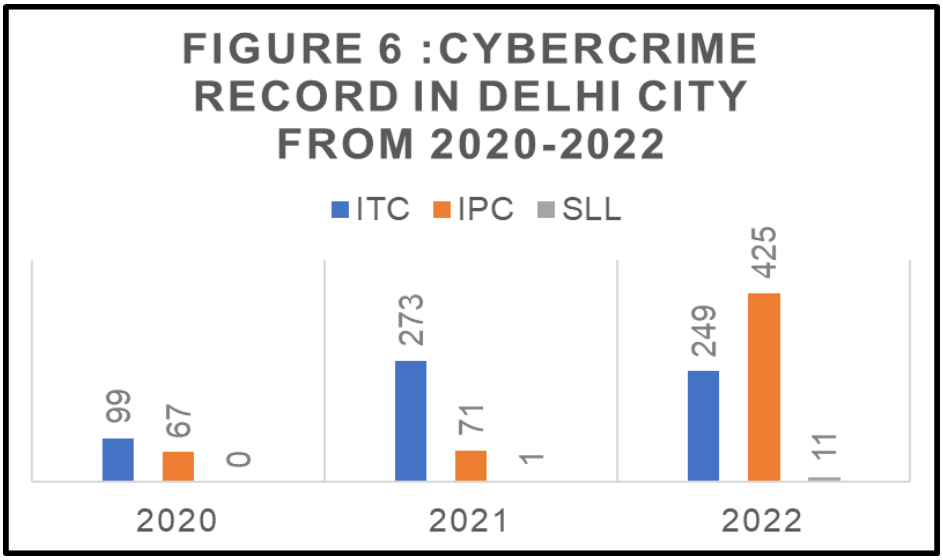



From Table 5 and Figure 5, we can see that from 2020 to 2022, the Bengaluru city has recorded cybercrimes under I.T. Act (or ITC) and no crime recorded under IPC and SLL. Under I.T. Act, Bengaluru reported a crime record of 8892 in 2020, 6423 in 2021 and 9940 in 2022. Table 5 also highlights that the Bengaluru did not record any cybercrime under IPC and SLL for the period under study.

**6. Study of the total number of cybercrime reported in Delhi City**

TABLE 6: Cybercrime record in Delhi city

| | I.T. Act | IPC | SLL |
|---|---|---|---|
| **2020** | 99 | 67 | 0 |
| **2021** | 273 | 71 | 1 |
| **2022** | 249 | 425 | 11 |

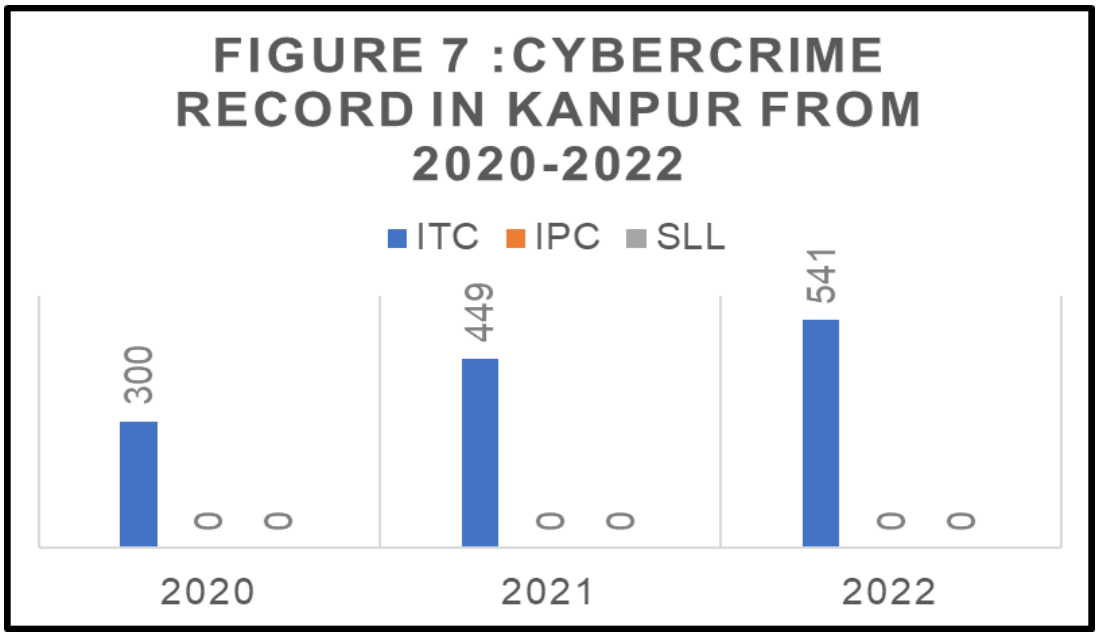



From Table 6 and Figure 6, we can see that the Delhi city has reported a crime record of 99 under I.T. Act, 67 under IPC and zero crime under SLL in the year 2020. For 2021, Delhi city a crime record of 273 under I.T. Act, 71 under IPC and 1 crime under SLL .For 2022, Delhi city a crime record of 249 under I.T. Act, 425 under IPC and 11 crime under SLL . Table 6 also highlights that the Delhi city recorded 60% of the cybercrime under I.T. Act and nearly 40% of the cybercrime under IPC in the year 2020. For 2021, the Delhi city had 79% cybercrime under I.T. Act, 20.6% under IPC and 0.3% under SLL. For 2022, the Delhi city had 36.4% cybercrime under I.T. Act, 62% under IPC and 1.6% under SLL. It can be seen that the crime rate under IPC and SLL has increased in Delhi city for the period under study.

**7. Study of the total number of cybercrime reported in Kanpur**

TABLE 7: Cybercrime record in Kanpur

| | **I.T. Act** | **IPC** | **SLL** |
|---|---|---|---|
| **2020** | 300 | 0 | 0 |
| **2021** | 449 | 0 | 0 |
| **2022** | 541 | 0 | 0 |

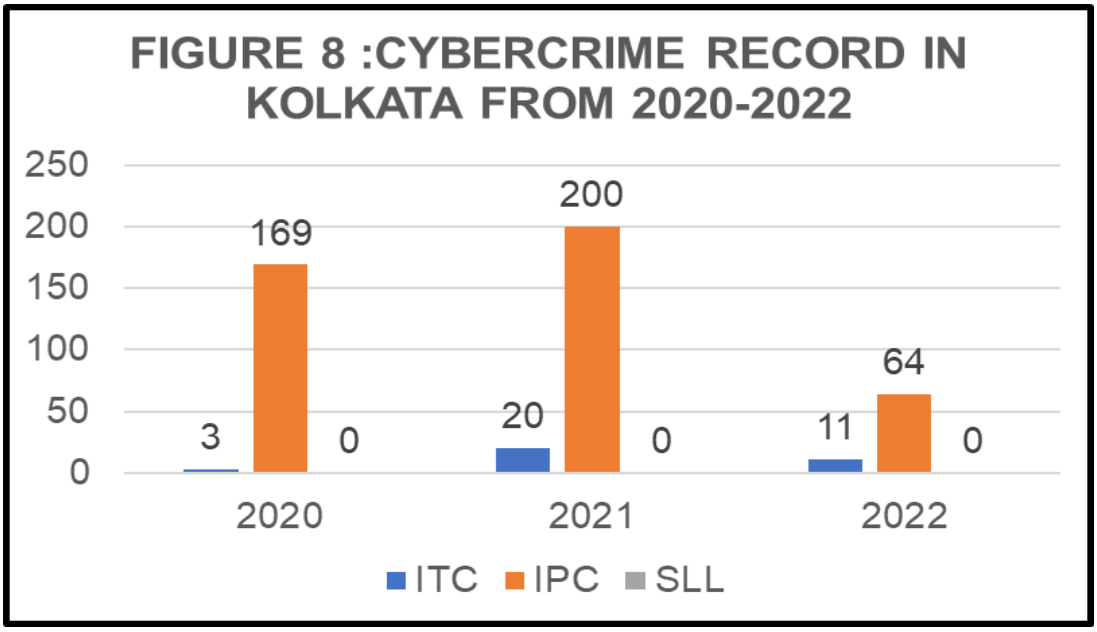



From Table 7 and Figure 7, we can see that from 2020 to 2022, Kanpur has recorded cybercrimes only under I.T. Act and no crime recorded under IPC and SLL. Under I.T. Act, Kanpur reported a crime record of 300 in 2020, 449 in 2021 and 541 in 2022. Table 5 also highlights that the number of cybercrimes reported under I.T. Act has increased from 2020 to 2022.

**8. Study of the total number of cybercrime reported in Kolkata**

TABLE 8: Cybercrime record in Kolkata

| | **I.T. Act** | **IPC** | **SLL** |
|---|---|---|---|
| **2020** | 3 | 169 | 0 |
| **2021** | 20 | 200 | 0 |
| **2022** | 11 | 64 | 0 |

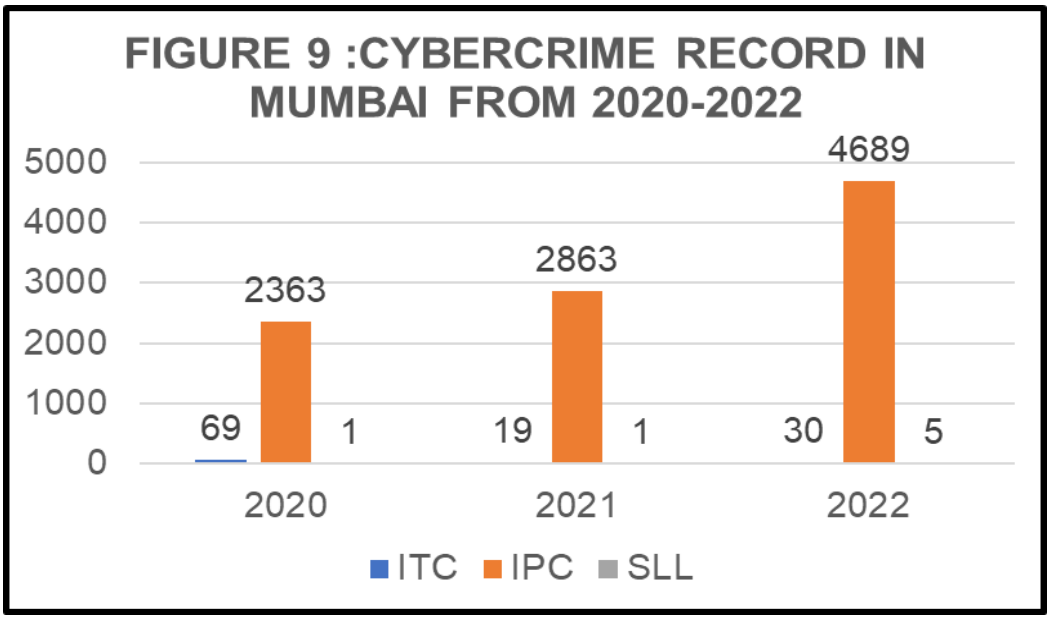



From Table 8 and Figure 8, we can see that Kolkata has reported a crime record of 3 under I.T. Act (or ITC), 169 under IPC and zero crime under SLL in the year 2020. For 2021, Kolkata has a crime record of 20 under I.T. Act, 200 under IPC and 0 crime under SLL. For 2022, Kolkata has a crime record of 11 under I.T. Act, 64 under IPC and 0 crime under SLL. Table 8 also highlights that Kolkata recorded 1.7% of the cybercrime under I.T. Act and 98.3% of the cybercrime under IPC in the year 2020. For 2021, Kolkata had 9.1% cybercrime under I.T. Act and 90.9% under IPC. For 2022, Kolkata had 14.7% cybercrime under I.T. Act and 85.3% under IPC. It can be seen that Kolkata has a major cybercrime reported under IPC and no cybercrime reported under SLL for the period under study.

**9. Study of the total number of cybercrime reported in Mumbai**

TABLE 9: Cybercrime record in Mumbai

| | **I.T. Act** | **IPC** | **SLL** |
|---|---|---|---|
| **2020** | 69 | 2363 | 1 |
| **2021** | 19 | 2863 | 1 |
| **2022** | 30 | 4689 | 5 |




From Table 9 and Figure 9, we can see that Mumbai has reported a crime record of 69 under I.T. Act (or ITC), 2363 under IPC and 1 crime under SLL in the year 2020. For 2021, Mumbai has a crime record of 19 under I.T. Act, 2863 under IPC and 1 crime under SLL. For 2022, Mumbai has a crime record of 30 under I.T. Act, 4689 under IPC and 5 crime under SLL. Table 8 also highlights that Mumbai recorded 2.84% of the cybercrime under I.T. Act, 97.12% of the cybercrime under IPC and 0.04% of the cybercrime under SLL in the year 2020. For 2021, Mumbai had 0.66% cybercrime under I.T. Act, 99.31% under IPC and 0.03% of the cybercrime under SLL. For 2022, Mumbai had 0.64% cybercrime under I.T. Act, 99.26% under IPC and 0.11% of the cybercrime under SLL. It can be seen that Mumbai has a major cybercrime recorded under IPC and the cybercrime record under I.T. Act has decreased for the period under study.

**10. Hypothesis testing using ANOVA**

To test whether there is a significant difference in the occurrence of cybercrime in the selected metropolitan cities.

$H_0$: There is no significant difference in the cybercrime record of different metropolitan cities for the study period.

$H_1$: not - $H_0$

Table 10: ANOVA OUTPUT

| **Source of Variation** | **SS** | **df** | **MS** | **F** | **P-value** | **F crit** |
|---|---|---|---|---|---|---|
| Between Groups | 150481692 | 5 | 30096338 | 37.4352 | $6.80x10^{7}$ | 3.10588 |
| Within Groups | 9647502 | 12 | 803958 | | | |
| Total | 160129195 | 17 | | | | |

From Table 10, we can see that the p-value ($6.80x10^{7}$) is less than 0.05 which implies that we can reject the null hypothesis. Therefore, we can say that there is a significant difference in the cybercrime record of different metropolitan cities for the study period.

**11. Mean of cybercrime records under each category**

Table 11: Average record of cybercrime under I.T. Act, IPC and SLL

| | **I.T. Act** | **IPC** | **SLL** |
|---|---|---|---|
| **BENGALURU** | 8418.3 | 0 | 0 |
| **DELHI CITY** | 207 | 187.6666667 | 4 |
| **KANPUR** | 430 | 0 | 0 |
| **KOLKATA** | 11.333 | 144.3333333 | 0 |
| **MUMBAI** | 39.333 | 3305 | 2.3 |

From Table 11, we can say that Bengaluru records the highest average rate (8418.3) of cybercrime under I.T. Act, Mumbai records the highest average rate (3305) of cybercrime under IPC and Delhi city records the highest average rate (4) of cybercrime under SLL.

**12. Standard Deviation of cybercrime records under each category**

Table 12: Standard Deviation (SD) of cybercrime record under I.T. Act, IPC and SLL

| | **I.T. Act** | **IPC** | **SLL** |
|---|---|---|---|
| **BENGALURU** | 1805.7 | 0 | 0 |
| **DELHI CITY** | 94.297 | 205.5464262 | 6.1 |
| **KANPUR** | 121.62 | 0 | 0 |
| **KOLKATA** | 8.5049 | 71.27645708 | 0 |
| **MUMBAI** | 26.274 | 1224.374126 | 2.3 |

From Table 12, we can say that Bengaluru records the highest SD (1805.7) of cybercrime under I.T. Act, Mumbai records the highest SD (1224.37) of cybercrime under IPC and Delhi city records the highest SD (6.1) of cybercrime under SLL.

**Conclusions**

The analysis reveals that the selected metropolitan cities have remarkably low incident of cybercrime under Special and Local Laws (SLL). Bengaluru and Kanpur have record of cybercrime only under the Information and Technology Act (I.T. Act) for the period under study. Delhi City, Kolkata and Mumbai record the highest averages of cybercrimes under Indian Penal Code (IPC). The results of the ANOVA test indicate a substantial difference in the occurrence of cybercrimes in the study period between the selected metropolitan cities.

**Limitation of the Study**

The NCRB website presents information on the amount of cybercrimes reported in all the cities over a period of time. Hence, the data fails to provide a determination of specific factors, such as demographic, economic or infrastructure factors that could potentially be responsible for the occurrence of cybercrime in different cities.

**Bibliography**

ChiChao Lu, WenYuan Jen, Weiping Chang, Shihchieh Chou (2006). "Cybercrime & Cybercriminals: An Overview of the Taiwan Experience", *Journal of Computers*, Vol. 1, No. 6, September 2006.

Dr. Fatma Abdalla Mabrouk Khiralla (2020). "Statistics of Cybercrime from 2016 to the First Half of 2020", *International Journal of Computer Science and Network*, Volume 9, Issue 5, October 2020

Allan G Bluman, *Elementary Statistics-A step by step approach* Seventh Edition, Mc Graw-Hill, New York, 2009.

Prem. S. Mann. *Introductory Statistics* Seventh edition, John Wiley & Sons Inc, NJ, 2009.

National Crime Records Bureau -www.ncrb.gov.in

Mumbai Police - https://mumbaipolice.gov.in/CrimeStatistics

Government of India - Press Information Bureau - www.pib.gov.in

National Cyber Crime Reporting Portal: https://cybercrime.gov.in

Indian Cybercrime Coordination Centre (I4C): https://i4c.mha.gov.in

https://www.business-standard.com/india-news/here-is-how-much-indians-lost-to-cyber-frauds-between-jan-and-apr-of-2024-124052700151_1.html

# SOPHIA LUCID CALL FOR RESEARCH PAPERS

## Volume V, Issue I (February 2026)

**Theme of the Issue: The Human Condition: Exploring the Intersections of Health, Wellness and Disease**

*Sophia Lucid* invites academic research papers from teachers and researchers across universities, colleges and research institutes in India to rethink the existing paradigms, perspectives and compendium of knowledge on this broad area, through the disciplines of humanities, social sciences and natural sciences. A list of suggestive areas (but, not limited to) is given below.

- Health, wellness and AI
- History of medical practices
- Colonial medicine
- Vernacular medicine
- Narrative medicine
- Health policies
- Landscape of changing healthcare
- Pharmaceutical companies and its prospects
- Medical tourism
- Public health
- Psychological well-being and counselling
- Social media and wellness
- Fake news and treatments
- Health economics
- Sociology of health and illness
- Nanotechnology and medicine
- Nutrition and therapeutic wellness
- Social media and mental health
- Impact of climate change on health
- Genetically modified foods
- Novel materials for remediation
- Computational science for human betterment
- Ethical wellness and mental health
- Philosophy of medicine
- Medicine and politics
- Cost effective health technologies
- Wearable technologies and posthumanism
- Biostatistics
- Gerontology
- Health journalism
- Digital technologies in health and wellness
- Medical zoology and public health
- Medical imaging physics

**Important deadlines**

**Submission of abstract**: July 5, 2025
**Communication regarding the selected abstracts**: July 15, 2025
**Submission of the final paper:** September 15, 2025

Please email your abstracts pertaining to the theme of the issue to the editor: **editorlucid@sophiacollege.edu.in** within the deadline.