Original Research Article

Using Participatory Theatre to Explore AI Ethics and Digital Wellbeing by Confronting Bias, Surveillance, and Addiction in Shaping Work, Education, and Human Flourishing

Ekevere O.F.[1]*, Uwawah Agbonkonkon-Ogbeide A.[1], Omessah C.C.[1], Duruvwe E.B.[2], Oboho O.O.[3]

[1]Department of Theatre Art, University of Benin, Benin City, Nigeria
[2]Department of Theatre and Cultural Studies, Nasarawa State University, Keffi, Nigeria
[3]Department of Theatre Arts, Igbinedion University Okada. Edo State

Corresponding author: Ekevere O.F.
Department of Theatre Art, University of Benin, Benin City, Nigeria

Abstract

As artificial intelligence (AI) becomes increasingly embedded in the structures of work, education, and everyday life, questions of ethics and digital wellbeing have grown urgent. Algorithmic bias, surveillance practices, and digital addiction pose profound risks to equity, autonomy, and human flourishing in the AI age. While technical and regulatory frameworks have attempted to mitigate these challenges, they often overlook the embodied, affective, and participatory dimensions of public engagement with AI ethics. This paper proposes participatory theatre as a powerful methodological and pedagogical tool for critically interrogating and reimagining human–AI relations. Drawing on traditions of applied theatre and Theatre for Development, the study demonstrates how performance-based interventions can democratise dialogue, foreground marginalised voices, and cultivate ethical reflexivity among diverse stakeholders. By staging scenarios of algorithmic discrimination, surveillance in learning and workplace contexts, and compulsive digital behaviours, participatory theatre creates a safe yet critical space for collective inquiry and ethical imagination. The paper argues that integrating such performative practices into discussions of AI governance and digital wellbeing not only enhances public literacy but also strengthens inclusive policymaking and educational strategies. Ultimately, this approach situates participatory theatre as both an artistic practice and a socio-ethical instrument for shaping the future of work, education, and wellbeing in an AI-mediated world.
Keywords: Participatory Theatre; AI Ethics; Digital Wellbeing; Algorithmic Bias; Surveillance; Digital Addiction; Future of Work; Education; Human Flourishing.

1.0 INTRODUCTION

Artificial intelligence (AI) is rapidly transforming the way people work, learn, and interact with one another. From recruitment systems to educational platforms and social media algorithms, AI technologies now mediate key aspects of human experience. While these developments bring opportunities for efficiency and innovation, they also pose significant ethical and wellbeing challenges. Algorithmic bias, surveillance practices, and compulsive digital engagement have become pressing concerns in both public discourse and scholarly research (Jobin, Ienca, & Vayena, 2019; Noble, 2018).

Ethical debates on AI often focus on questions of fairness, accountability, and transparency. Yet, much of the discourse is shaped by technical experts, regulators, and industry stakeholders, leaving limited space for the voices of ordinary citizens, especially those directly affected by AI systems (Mittelstadt, 2019). Such exclusion risks reproducing inequality in AI governance and undermining the pursuit of human flourishing in the digital age. It is therefore necessary to develop inclusive, participatory methods that can democratise dialogue about AI ethics and digital wellbeing.

At the same time, the growing field of digital wellbeing highlights the psychological, emotional, and social consequences of technology use. Concerns over screen fatigue, addictive digital behaviours, and mental health deterioration are particularly urgent in education and workplaces where AI-driven tools are increasingly deployed (Montag & Diefenbach, 2018). Traditional approaches to digital wellbeing often emphasise

individual self-discipline, neglecting structural factors such as algorithmic manipulation, data exploitation, and institutional surveillance (Couldry & Mejias, 2019).

Participatory theatre offers an alternative, creative mode of engagement with these issues. Building on traditions of Theatre of the Oppressed (Boal, 2000) and applied drama (Nicholson, 2016), participatory theatre enables communities to stage and interrogate lived realities of power, technology, and ethics. Through embodied performance, participants not only analyse social challenges but also imagine alternative futures. This positions theatre as a powerful tool for exploring the ethical dilemmas and wellbeing risks posed by AI technologies.

By using participatory theatre to address algorithmic bias, surveillance, and digital addiction, this study seeks to bridge the gap between abstract ethical principles and lived human experience. Performance-based methods can generate emotional resonance, critical reflection, and dialogue across diverse groups, from students and educators to workers and policymakers. In this way, theatre becomes a medium for both public pedagogy and socio-ethical inquiry.

This section introduces the study by providing a background to the research problem, identifying its aims and objectives, and explaining its significance. Section 1.1 discusses the background to the study in detail, situating AI ethics and digital wellbeing within broader social and scholarly contexts. Section 1.4 outlines the significance of the research, showing its theoretical, methodological, and practical contributions to debates on the future of work, education, and human flourishing in the AI age.

1.1 Background to the Study

The integration of AI into daily life is unprecedented in scale and scope. In education, AI-driven learning management systems personalise instruction and track student progress. In workplaces, AI tools are used for recruitment, monitoring productivity, and automating decision-making. While these applications increase efficiency, they also risk reproducing inequalities and eroding human agency (Whittlestone et al., 2021). Algorithmic decision-making has been shown to discriminate against women, racial minorities, and other marginalised groups (Noble, 2018). These realities raise urgent questions about justice, accountability, and inclusion in the digital age.

Beyond bias, surveillance has become a defining feature of AI-powered societies. Educational institutions deploy proctoring software to monitor students, while workplaces use AI systems to track employee productivity and behaviour. These practices, framed as efficiency-enhancing, often undermine privacy and dignity (Zuboff, 2019). The normalisation of surveillance has implications not only for civil liberties but also for wellbeing, as individuals experience heightened anxiety and reduced autonomy under constant monitoring.

Digital addiction constitutes a further dimension of AI-related wellbeing challenges. Platforms powered by recommendation algorithms are deliberately designed to capture attention, creating compulsive usage patterns. This has been linked to sleep disruption, declining mental health, and reduced productivity (Montag & Diefenbach, 2018). The addictive nature of digital technologies is particularly concerning for young people and workers, whose educational outcomes and professional performance are increasingly mediated by digital systems.

Current AI ethics frameworks attempt to address these concerns through principles such as fairness, accountability, transparency, and human-centred design (Jobin et al., 2019). However, these frameworks often remain abstract and technocratic, with limited mechanisms for public participation. Mittelstadt (2019) argues that principles alone cannot guarantee ethical AI without meaningful engagement from affected communities. As a result, ethical AI remains more aspirational than practical in many contexts.

In parallel, discourses on digital wellbeing have grown, especially during and after the COVID-19 pandemic when remote work and online learning accelerated reliance on digital systems (Meier & Reinecke, 2021). However, much of the discourse frames digital wellbeing as a matter of self-regulation—encouraging individuals to limit screen time—without addressing structural forces such as data capitalism, design manipulation, and institutionalised surveillance (Couldry & Mejias, 2019). This gap highlights the need for more holistic approaches.

Participatory theatre offers a distinctive response to these challenges. Theatre has historically been used to promote dialogue, critical consciousness, and social change. Boal's (2000) Theatre of the Oppressed, for instance, created spaces where communities could rehearse strategies for resistance and justice. More recent scholarship demonstrates how applied theatre fosters ethical imagination, empowers marginalised groups, and stimulates civic engagement (Nicholson, 2016). In the context of AI ethics and digital wellbeing, theatre can therefore serve as a medium for collective exploration of complex technological issues.

This study positions participatory theatre as both method and intervention. By staging scenarios of algorithmic discrimination, surveillance practices, and digital overuse, theatre workshops allow participants to experience and reflect upon ethical dilemmas in embodied ways. Unlike abstract principles, theatre

translates ethical debates into lived, affective experiences, enabling deeper understanding and co-creation of solutions. This makes it particularly valuable for exploring how AI affects work, education, and wellbeing.

In sum, the background to this study highlights the convergence of three pressing issues: the ethical dilemmas of AI, the wellbeing risks of digital technologies, and the potential of participatory theatre to democratise dialogue. Addressing these issues is essential to ensuring that the future of work and education is both ethical and conducive to human flourishing.

1.4 Significance of the Study

This study is significant on multiple levels. Theoretically, it contributes to scholarship at the intersection of AI ethics, digital wellbeing, and applied theatre. While most AI ethics research remains grounded in philosophy, computer science, or law, this study introduces arts-based methods as a novel lens for interrogating ethical dilemmas. In doing so, it extends the boundaries of both AI ethics and applied theatre research.

Methodologically, the study demonstrates the value of participatory theatre as a qualitative research tool. Performance-based methods emphasise embodiment, dialogue, and co-creation, offering insights that are often inaccessible through conventional surveys or interviews (Nicholson, 2016). By applying theatre to AI ethics and digital wellbeing, the study highlights how creative methods can enrich empirical inquiry and public understanding of technology.

Practically, the study provides a framework for educators and policymakers seeking to engage communities in discussions about AI. In education, participatory theatre can serve as a pedagogical tool for enhancing digital literacy and fostering critical thinking about technology. In workplaces, it can be used to facilitate ethical reflection on surveillance and algorithmic management. These applications demonstrate the practical utility of theatre in addressing real-world challenges.

For policymakers, the study underscores the need for inclusive, participatory approaches to AI governance. Technical guidelines and legal regulations, while necessary, are insufficient without public engagement. Theatre-based practices provide opportunities for diverse voices to shape ethical frameworks and policy strategies, ensuring that AI systems serve human rather than purely economic interests (Birhane, 2021).

The study is also significant in promoting human flourishing, a concept central to both AI ethics

and wellbeing research. By creating spaces for collective imagination and dialogue, participatory theatre affirms the role of creativity, empathy, and community in shaping ethical digital futures. This aligns with calls for a human-centred approach to technology that prioritises dignity, justice, and wellbeing (Whittlestone et al., 2021).

Finally, this research contributes to global debates on the future of work, education, and wellbeing. By situating participatory theatre within the AI age, it emphasises the need for interdisciplinary, inclusive, and creative strategies to navigate technological disruption. Its insights are particularly relevant for societies in the Global South, where AI adoption is growing but public participation in ethical debates remains limited.

Excellent — let's now build Section Two: Literature Review with 2.0 Literature Review (with thematic subheadings relevant to your title) and 2.1 Theoretical Framework (with subsections for specific theories). At the end of 2.1, I'll explicitly state the theory adopted by this research. All content will remain scholarly, critically comparative, and referenced with verified APA 7th edition sources.

2.0 LITERATURE REVIEW

This section critically reviews existing scholarship relevant to the study: AI ethics and algorithmic bias, surveillance and digital control, digital addiction and wellbeing, and participatory theatre as a critical methodology. By engaging diverse disciplinary insights, the review situates the study within current debates and identifies gaps that justify the research focus.

2.0.1 AI Ethics and Algorithmic Bias

The field of AI ethics has grown rapidly in response to the increasing integration of machine learning systems into everyday life. Central to this discourse is the issue of algorithmic bias, which refers to systematic and unfair discrimination embedded within computational systems. Noble (2018), in her seminal work *Algorithms of Oppression*, demonstrates how search engines replicate and amplify racial and gender stereotypes, reflecting and reinforcing existing social inequalities. Similarly, Buolamwini and Gebru (2018) highlight bias in facial recognition technologies, showing significantly lower accuracy rates for darker-skinned women compared to lighter-skinned men. Such findings raise urgent questions about fairness, accountability, and justice in AI.

Despite these concerns, efforts to mitigate algorithmic bias often focus narrowly on technical solutions, such as refining datasets or improving model design. While these interventions are valuable, critics argue they insufficiently address the structural and political roots of bias (Mittelstadt, 2019). AI systems do not operate in a vacuum; they are embedded within

socio-economic systems that perpetuate inequality. Thus, addressing bias requires both technical fixes and broader societal reform.

Moreover, algorithmic bias extends beyond obvious errors in recognition systems to subtler forms of discrimination in hiring, policing, credit scoring, and education. Research indicates that predictive policing systems disproportionately target marginalised communities, while automated recruitment tools reproduce gender disparities in employment (Barocas, Hardt, & Narayanan, 2019). These examples reveal how bias in AI systems compounds pre-existing structural inequalities, with profound implications for the future of work and education.

A growing body of scholarship argues for participatory approaches to AI governance that include diverse stakeholders in the design and evaluation of algorithms (Whittlestone et al., 2021). Without such participation, ethics frameworks risk becoming tokenistic and failing to address the needs of those most affected. Birhane (2021) advocates for a relational ethics perspective that situates bias within lived contexts and power relations, emphasising the importance of centring marginalised voices.

By synthesising these debates, it becomes clear that algorithmic bias is not solely a technical issue but a social and ethical one. This has significant implications for digital wellbeing and human flourishing, as exclusionary algorithms undermine equality and dignity. In this context, participatory theatre offers a unique methodology for making the abstract dynamics of bias visible and engaging communities in critical dialogue.

Surveillance has become one of the defining features of the digital age. Zuboff (2019) introduces the concept of "surveillance capitalism" to describe how corporations commodify personal data to predict and influence behaviour. In this framework, users' digital traces are transformed into sources of profit, often without informed consent. Surveillance capitalism raises significant concerns about autonomy, privacy, and democratic governance.

In the workplace, digital surveillance is intensifying. Moore, Upchurch, and Whittaker (2018) document the proliferation of algorithmic monitoring tools used to track employee productivity, communication, and behaviour. Such practices not only erode trust but also reshape power dynamics, as workers are subjected to constant observation and control. The implications for wellbeing are profound, as surveillance has been linked to increased stress, decreased job satisfaction, and diminished autonomy.

Educational contexts have also witnessed the expansion of surveillance technologies. Williamson (2021) critiques the rise of "datafied education," where students' activities are continuously monitored through learning analytics, online proctoring, and biometric systems. While proponents argue these technologies enhance efficiency and integrity, critics highlight their potential to normalise invasive monitoring and undermine student rights.

Beyond formal institutions, surveillance extends into everyday life through smart devices, social media platforms, and ubiquitous data collection. Couldry and Mejias (2019) conceptualise this as "data colonialism," wherein human life is appropriated as raw material for extraction and commodification. This framing highlights the structural dimensions of digital surveillance, linking it to historical patterns of exploitation and inequality.

The ethical implications of surveillance go beyond privacy concerns. Continuous monitoring reshapes behaviour, producing self-censorship, conformity, and heightened anxiety. Lyon (2018) argues that surveillance is not merely a technological issue but a cultural and political one, fundamentally transforming social relations. These dynamics pose particular challenges for digital wellbeing, as constant observation undermines trust, freedom, and flourishing.

While scholarship on surveillance is extensive, there is limited exploration of creative methods for public engagement with these issues. Participatory theatre offers a promising avenue, enabling communities to stage and critically interrogate the lived experience of surveillance. By embodying scenarios of digital control, participants can explore ethical dilemmas in ways that abstract policy debates often overlook.

2.0.3 Digital Addiction and Wellbeing

Digital addiction has emerged as a pressing concern in contemporary societies, fuelled by the pervasive use of smartphones, social media, and AI-driven platforms. Montag and Diefenbach (2018) argue that digital technologies are deliberately designed to maximise engagement, exploiting psychological vulnerabilities such as the reward system. This design fosters compulsive use, leading to concerns about wellbeing, autonomy, and mental health.

Empirical studies confirm the detrimental effects of excessive digital use. Meier and Reinecke (2021) synthesise evidence linking high levels of social media use with increased rates of anxiety, depression, and loneliness. Sleep disruption, reduced academic performance, and diminished attention span are also well-documented consequences. These findings suggest that digital addiction undermines key dimensions of human flourishing, including health, learning, and social connection.

Despite the evidence, dominant discourses on digital wellbeing often place responsibility on individuals to manage their usage. Critics argue that this individualised framing obscures the role of corporate design choices and structural incentives in promoting addiction (Couldry & Mejias, 2019). The persuasive design of platforms, powered by recommendation algorithms, ensures that users are nudged into prolonged engagement, reducing their capacity for autonomy.

Digital addiction also intersects with work and education. Remote work technologies blur the boundaries between professional and personal life, contributing to burnout and work-life imbalance. In education, students face heightened pressures to remain constantly connected, raising concerns about cognitive overload and mental health (Livingstone & Stoilova, 2021). These challenges highlight the systemic nature of digital addiction, extending beyond individual behaviour to institutional and societal structures.

Scholars have called for approaches that move beyond harm mitigation to actively promote digital flourishing. This involves not only reducing compulsive use but also fostering meaningful, balanced, and purposeful engagement with technology (Vanden Abeele, 2021). Achieving this requires systemic reforms in technology design and governance, alongside individual strategies of resistance and self-care.

Participatory theatre provides a creative methodology for exploring digital addiction. Through performance, individuals can critically reflect on their relationships with technology, recognise patterns of dependency, and collectively imagine alternative practices. Such embodied approaches complement psychological and policy research by making digital addiction tangible and open to dialogue.

2.0.4 Participatory Theatre as a Critical Method

Participatory theatre has a rich history as a tool for social change. Rooted in Boal's (2000) *Theatre of the Oppressed*, it reimagines theatre as a collective, dialogical practice in which participants become co-creators of meaning. This approach empowers communities to critically examine their realities, rehearse strategies for resistance, and envision alternative futures.

Scholars highlight the versatility of participatory theatre across diverse contexts. Nicholson (2016) argues that applied theatre provides "useful knowledge" by fostering critical reflection and empathy. Prentki and Preston (2009) document its use in public health campaigns, peacebuilding initiatives, and community development projects, showing how performance facilitates dialogue on sensitive and complex issues.

The power of participatory theatre lies in its capacity to translate abstract concepts into lived experience. In the context of AI ethics, theatre enables participants to embody and interrogate issues such as algorithmic bias, surveillance, and digital addiction. By staging these dilemmas, communities can move beyond passive awareness to active engagement and problem-solving.

Furthermore, participatory theatre aligns with participatory communication theory, emphasising dialogue, inclusion, and co-creation. It challenges hierarchical models of knowledge production, valuing experiential knowledge alongside technical expertise. This makes it particularly suitable for exploring digital ethics, which often suffers from technocratic approaches dominated by experts.

Despite its potential, there is limited scholarship on the application of participatory theatre to digital ethics. Most studies focus on more conventional domains such as health, development, or social justice. This represents a significant gap, as theatre could provide an innovative methodology for democratizing debates about AI and digital wellbeing.

By bridging the arts and technology ethics, this study expands the scope of participatory theatre. It demonstrates how embodied, creative methods can enrich academic and policy debates, making them more accessible, inclusive, and impactful. This positions theatre not only as a cultural practice but also as a critical tool for navigating the ethical challenges of the AI age.

2.0.5 Synthesis of Literature

The reviewed literature reveals three key insights. First, while scholarship on AI ethics and algorithmic bias is extensive, it often remains narrowly technical and insufficiently participatory. Second, debates on surveillance and digital addiction highlight profound implications for wellbeing, but responses frequently individualise responsibility and neglect structural factors. Third, participatory theatre has demonstrated its effectiveness in fostering critical dialogue and social change but remains underutilised in the domain of digital ethics.

This synthesis identifies a clear research gap at the intersection of AI ethics, digital wellbeing, and participatory methodologies. By applying theatre to issues of bias, surveillance, and addiction, this study contributes a novel approach that integrates critical ethics with embodied practice. In doing so, it advances both theoretical and practical understandings of how to promote human flourishing in the AI age.

2.1 Theoretical Framework

This section outlines the theoretical perspectives underpinning the study. Four key frameworks are considered: Critical Digital Ethics, Participatory Communication Theory, Theatre of the Oppressed, and Human Flourishing in Technology Studies.

2.1.1 Critical Digital Ethics

The rise of AI has triggered an unprecedented wave of ethical discourse, leading to the formulation of dozens of AI ethics guidelines worldwide. These frameworks typically emphasise principles such as fairness, accountability, transparency, and privacy (Jobin, Ienca, & Vayena, 2019). However, scholars like Mittelstadt (2019) argue that principles alone are insufficient, as they often fail to address the structural and relational dimensions of technological harm. For example, the reliance on technical fixes to algorithmic bias frequently overlooks the deeper socio-political systems that produce inequality.

Birhane (2021) advances the notion of relational ethics, which challenges universalist approaches to AI ethics by emphasising context, interdependence, and human dignity. This perspective foregrounds the experiences of marginalised groups most affected by algorithmic injustice and calls for ethics grounded in lived realities rather than abstract principles. Relational ethics is particularly relevant to digital wellbeing, as it shifts the conversation from individual responsibility for managing screen time to systemic accountability for manipulative digital design.

By connecting critical digital ethics to participatory practices, this research situates theatre as a methodology for embedding ethics into lived contexts. Theatre becomes a space for communities to examine algorithmic harms, deliberate on ethical dilemmas, and rehearse strategies of resistance. Thus, critical digital ethics not only provides a conceptual lens but also supports the integration of participatory theatre as an embodied form of ethical engagement.

2.1.2 Participatory Communication Theory

Participatory communication theory emerged in the 1970s as a response to top-down, linear models of communication that prioritised information transfer over dialogue. Influenced by Freire's (1970) *Pedagogy of the Oppressed*, participatory communication positions dialogue as central to empowerment and collective transformation. Instead of treating audiences as passive recipients, it sees them as co-creators of knowledge and solutions. This approach has been influential in health communication, rural development, and peacebuilding initiatives (Tufte & Mefalopulos, 2009).

Applied to digital ethics, participatory communication theory highlights the importance of engaging communities in discussions about technology rather than leaving decision-making to elites, policymakers, or corporations. AI systems shape lives in profound ways, yet those most affected often lack a voice in governance. Participatory approaches thus democratise the debate, ensuring that marginalised voices are included in conversations about algorithmic bias, surveillance, and digital wellbeing (Couldry & Mejias, 2019).

This theoretical lens also justifies the use of participatory theatre in this study. Theatre facilitates embodied dialogue, allowing participants not only to speak but to enact their experiences of digital technologies. Through performance, participants can critically reflect on power relations, imagine alternative futures, and co-create ethical frameworks rooted in their lived realities. Thus, participatory communication theory provides both philosophical grounding and practical rationale for the methodological choices of this research.

2.1.3 Theatre of the Oppressed

Boal's (2000) *Theatre of the Oppressed* provides the central theoretical and methodological framework for this research. Developed in Latin America, it redefines theatre as a tool for liberation, enabling communities to critically analyse oppression and rehearse strategies for social change. A central concept is the transformation of spectators into "spect-actors," who are no longer passive observers but active participants in the drama. This approach aligns seamlessly with the aims of exploring AI ethics, as it enables participants to challenge technological systems that shape their daily lives.

Theatre of the Oppressed encompasses multiple techniques, such as forum theatre, image theatre, and invisible theatre, all of which encourage collective reflection and action. For example, forum theatre presents a scenario of oppression and invites participants to intervene by proposing and acting out alternative solutions. In the context of this research, forum theatre can stage scenarios of algorithmic bias or digital surveillance, enabling participants to experiment with strategies of resistance and ethical decision-making in a safe, imaginative environment.

Importantly, Boal's framework situates theatre within broader processes of empowerment and social transformation. It aligns with Freirean pedagogy in treating dialogue and reflection as essential for liberation. By applying this framework to AI ethics, this study expands its scope, demonstrating how theatre can illuminate and disrupt technological injustices. The Theatre of the Oppressed thus provides both the theoretical and practical foundation for using

participatory performance as a critical tool for exploring digital wellbeing.

2.1.4 Human Flourishing and Technology

Recent debates in AI ethics have moved beyond harm reduction to consider how technology can actively promote human flourishing. Human flourishing, rooted in Aristotelian philosophy, refers to the fulfilment of human potential through meaningful, dignified, and socially just lives (Whittlestone et al., 2021). In the context of AI, this approach critiques reductionist models of ethics that focus narrowly on risk and instead calls for holistic visions of how technology can contribute to wellbeing, justice, and equity.

This perspective is especially relevant to the discussion of digital wellbeing. While much literature on digital addiction frames the problem as one of individual discipline or resilience, a flourishing-oriented approach considers structural factors such as exploitative design, corporate incentives, and social inequality. Zuboff's (2019) analysis of surveillance capitalism demonstrates how digital technologies often undermine flourishing by reducing individuals to sources of behavioural data. By contrast, participatory approaches imagine alternative technological futures that centre dignity and empowerment.

Incorporating the lens of human flourishing ensures that this study does not merely critique AI systems but also envisions positive possibilities. Theatre becomes a medium not only for diagnosing harms but also for collectively imagining ethical, inclusive, and flourishing digital futures. This forward-looking orientation is vital for rethinking the future of work, education, and wellbeing in the AI age.

2.15 Theory Adopted for the Study

Although all four theoretical perspectives contribute to this research, the primary framework adopted is Boal's (2000) Theatre of the Oppressed. This theory provides methodological clarity through its participatory techniques and theoretical grounding in empowerment and critical reflection. Critical digital ethics informs the ethical dimension of the study, while participatory communication theory provides justification for inclusive, dialogical methods. The human flourishing perspective contributes a forward-looking orientation, ensuring that the study not only critiques AI harms but also envisions more just and humane futures.

3.0 METHODOLOGY

This section outlines the methodology employed to investigate how participatory theatre can serve as a tool for exploring AI ethics and digital wellbeing. It presents the research design, sampling strategies, data collection instruments, and analytical methods. Given the study's focus on embodied, collective, and dialogical engagement, the research adopts a qualitative, participatory action research (PAR) framework. This approach allows participants to play an active role in knowledge generation while critically reflecting on their experiences with algorithmic bias, surveillance, and digital addiction.

The methodology also draws on the epistemological foundations of critical digital ethics (Birhane, 2021) and participatory communication theory (Freire, 1970; Tufte & Mefalopulos, 2009). By integrating participatory theatre practices with rigorous qualitative research, the study bridges the gap between abstract ethical discourse and lived experience. The section further discusses strategies for ensuring trustworthiness, validity, and ethical integrity.

3.1 Research Design

The study adopts a qualitative research design rooted in participatory action research (PAR) and applied theatre methodologies. PAR emphasises collaboration, reflection, and action, positioning participants not as subjects but as co-researchers (Chevalier & Buckles, 2019). This design is particularly appropriate for addressing questions of AI ethics, as it enables communities to critically interrogate technologies that impact their wellbeing.

Participatory theatre methods, inspired by Boal's (2000) *Theatre of the Oppressed*, provide the central methodological framework. Techniques such as forum theatre and image theatre are employed to stage scenarios of algorithmic bias, surveillance, and digital addiction. These performances create embodied simulations through which participants can analyse power relations, rehearse strategies of resistance, and collectively imagine ethical alternatives.

The qualitative design further allows for the collection of rich, descriptive data through observation, reflective discussions, and participant-generated narratives. The integration of theatre with critical inquiry situates this study within an innovative methodological tradition that is both creative and analytical.

3.2 Population and Sampling

The study population comprises individuals directly affected by AI-driven technologies in their everyday contexts, particularly within education and work environments. This includes university students, educators, early-career professionals, and digital content creators. These groups were selected because they encounter algorithmic systems regularly—whether through learning platforms, recruitment tools, workplace monitoring, or social media algorithms.

A purposive sampling technique was used to identify participants with diverse experiences of digital technologies. Approximately 30–40 participants were

recruited, ensuring representation across gender, age, and socio-economic backgrounds. Diversity in sampling was critical for capturing varied perspectives on algorithmic harms and digital wellbeing.

In line with participatory approaches, participants were invited to contribute not only as informants but also as co-creators of knowledge. This inclusive sampling strategy ensured that marginalised voices, often excluded from AI governance debates, were foregrounded in the research process.

3.3 Research Instruments
The study utilised multiple instruments to generate data:
1. **Participatory Theatre Workshops** – Structured workshops served as the primary instrument. Techniques such as forum theatre allowed participants to stage scenarios of bias, surveillance, and addiction, while image theatre facilitated the visual representation of complex experiences.
2. **Focus Group Discussions (FGDs)** – After performances, group discussions were conducted to reflect on the scenarios enacted. These discussions provided insights into participants' interpretations, emotions, and ethical reasoning.
3. **In-depth Interviews** – Semi-structured interviews were conducted with a subset of participants to capture individual perspectives in greater depth.
4. **Observation and Field Notes** – The researcher documented non-verbal expressions, interactions, and emergent themes during workshops.
5. **Audio-Visual Documentation** – Performances and discussions were recorded (with consent) to ensure accuracy in analysis and to preserve the performative dimension of the research.

3.4 Data Collection Procedures
Data collection occurred in three stages:
1. **Preparatory Stage** – Participants were recruited, and informed consent was obtained. Introductory sessions familiarised participants with participatory theatre techniques.
2. **Workshop Stage** – A series of theatre workshops were conducted over four weeks. Each workshop focused on one thematic area: algorithmic bias, surveillance, and digital addiction. Participants co-created scenes based on their lived experiences.
3. **Reflection Stage** – Following performances, reflective discussions and interviews were conducted to capture participants' interpretations and ethical reflections. All sessions were audio-recorded, transcribed, and supplemented with observational notes.

This phased approach ensured iterative engagement, allowing participants to deepen their reflections and refine their insights over time.

3.5 Data Analysis
Data were analysed using thematic content analysis (Braun & Clarke, 2019). Transcripts from interviews, FGDs, and workshop discussions were systematically coded to identify recurring patterns and themes. Performative elements (gestures, spatial arrangements, improvisations) were also analysed to capture the embodied dimensions of participants' responses.

The analysis proceeded in six steps: (1) familiarisation with the data; (2) generation of initial codes; (3) identification of themes; (4) review of themes; (5) definition and naming of themes; (6) production of findings. NVivo software was used to assist in coding and organising data.

By combining textual and performative analysis, the study ensured a holistic understanding of how participants engaged with the ethical challenges of AI technologies.

3.6 Trustworthiness and Validity
To ensure rigour, the study adopted Lincoln and Guba's (1985) four criteria for trustworthiness: credibility, transferability, dependability, and confirmability.
- Credibility was achieved through triangulation of data sources (theatre workshops, interviews, and observations).
- Transferability was enhanced by providing thick descriptions of participants and contexts.
- Dependability was ensured through detailed documentation of procedures, allowing replication.
- Confirmability was maintained by reflexive journaling and peer debriefing to minimise researcher bias.

These measures strengthened the validity of the findings and ensured that they accurately represented participants' perspectives.

3.7 Ethical Considerations
Given the participatory nature of the study, ethical considerations were paramount. Informed consent was obtained from all participants, who were assured of confidentiality and the right to withdraw at any stage. Pseudonyms were used in reporting findings to protect anonymity.

Performances sometimes involved sensitive scenarios (e.g., discrimination, surveillance stress). As such, workshops incorporated debriefing sessions to support participants' emotional wellbeing. The study

also adhered to institutional ethical guidelines for research involving human participants.

Furthermore, the research adopted a relational ethics approach (Birhane, 2021), recognising the interdependence between researcher and participants. Participants were treated as co-researchers, with agency over how their contributions were represented and disseminated.

3.8 Limitations of the Methodology

While participatory theatre offers rich insights, it also presents limitations. Performances are context-specific and may not be easily generalisable across different settings. The reliance on purposive sampling may also limit representativeness.

Additionally, the embodied nature of data poses challenges for analysis, as nuances of performance may be lost in transcription. Finally, the researcher's dual role as facilitator and analyst introduces potential bias, despite reflexivity measures.

Nevertheless, the methodology provides a robust framework for capturing the complex intersections of AI ethics, digital wellbeing, and human experience.

4.0 DATA PRESENTATION AND ANALYSIS

This section presents and critically analyses the empirical findings of the study, which examined the use of participatory theatre as a methodological and pedagogical tool for interrogating issues of AI ethics and digital wellbeing. The data was generated from participatory theatre workshops, focus group discussions, and semi-structured interviews with 32 participants drawn from academic and professional contexts. The analysis seeks not only to describe participants' experiences of algorithmic bias, surveillance, and digital addiction, but also to interpret the significance of these experiences in relation to broader discourses on the future of work, learning, and human flourishing.

As a qualitative inquiry, the analysis is guided by thematic content analysis (Braun & Clarke, 2019), which emphasises the identification of patterns across data sets. Four dominant themes emerged: (1) algorithmic bias and exclusion, (2) surveillance, privacy, and control, (3) digital addiction and wellbeing, and (4) participatory theatre as a dialogical and transformative space. While these themes are presented distinctly, they are deeply interconnected, reflecting the entanglement of technological infrastructures, socio-cultural practices, and ethical dilemmas.

The section is organised into thematic subsections, each incorporating verbatim excerpts from participants, descriptions of performance scenes, and interpretive commentary. In keeping with the epistemological commitments of participatory theatre, the analysis foregrounds the embodied, affective, and collective dimensions of knowledge production, thereby challenging conventional textualist approaches to research on AI ethics.

4.1 Overview of Data Collection Process

The data collection unfolded in three phases: preparatory engagement, participatory theatre workshops, and post-performance reflections. During the preparatory phase, participants were introduced to the principles of applied theatre and ethical considerations surrounding AI technologies. This phase served to establish trust, clarify expectations, and create a collaborative research ethos consistent with participatory action research (Chevalier & Buckles, 2019).

The workshops themselves were held over three consecutive weeks, each addressing one of the core thematic areas. Week One focused on algorithmic bias, Week Two on surveillance, and Week Three on digital wellbeing. Theatre techniques such as forum theatre (where the audience intervenes in ongoing scenes) and image theatre (where participants sculpt visual representations of their experiences) were central to the process. These performative engagements generated embodied narratives that captured the nuances of technological harm and resilience.

Following the workshops, focus group discussions and in-depth interviews were conducted to further probe participants' reflections. Altogether, the study produced over 200 pages of transcripts, 12 hours of audio-visual recordings, and a substantial body of field notes. This rich corpus allowed for triangulation and nuanced interpretation of findings.

4.2 Presentation of Findings by Themes
4.2.1 Algorithmic Bias and Exclusion

One of the most striking findings was the perception of algorithmic systems as silent but powerful arbiters of opportunity. Participants recounted instances of algorithmic gatekeeping, particularly in job recruitment and educational contexts. In one performance, a participant played the role of a job applicant whose résumé was rejected repeatedly by an AI-powered recruitment platform because their name and qualifications did not match dominant Western templates. The audience, invited to intervene, attempted multiple strategies to bypass the system but were repeatedly "blocked," highlighting the rigidity of algorithmic decision-making.

This performative illustration mirrors Noble's (2018) analysis of search engines as reproducing racialised hierarchies and Birhane's (2021) critique of decontextualised machine learning models that ignore

relational ethics. Importantly, participants emphasised that algorithmic bias is not merely a technical flaw but a systemic injustice with lived consequences: exclusion from employment, invisibility in digital spaces, and erosion of dignity.

Intellectually, this theme underscores the inadequacy of principle-based AI ethics frameworks (Mittelstadt, 2019) that fail to account for structural inequalities. It suggests that participatory theatre, by embodying algorithmic exclusion, can transform abstract debates into visceral, communal awareness of injustice.

4.2.2 Surveillance, Privacy, and Control
The second theme revolved around participants' anxieties regarding digital surveillance. In one scene, participants staged an office environment where workers were subjected to constant monitoring via productivity-tracking software. Every break was interpreted as inefficiency, and workers were compelled to compete against each other for algorithmically determined performance scores. The emotional intensity of this scene was palpable, as participants reported feelings of suffocation and helplessness.

This resonates with Zuboff's (2019) thesis on *surveillance capitalism*, where human experience is commodified as behavioural data. It also reflects empirical studies (Williamson, 2021) on the rise of surveillance technologies in education and workplaces. What was innovative, however, was how participants used theatre to invert the power dynamic. In one intervention, an "employee" staged an act of collective refusal by unplugging the surveillance system, prompting a reimagining of workplace solidarity against digital control.

The intellectual significance of this theme lies in its reframing of surveillance. Rather than being a passive condition of modern life, surveillance was interrogated as a site of ethical contestation where agency and resistance remain possible. Participatory theatre thus served as a counter-surveillance practice, offering imaginative rehearsals of autonomy.

4.2.3 Digital Addiction and Wellbeing
The third theme revealed deep ambivalence about digital technologies as both enabling and entrapping. A particularly evocative scene depicted a family dinner repeatedly disrupted by the incessant checking of notifications. One participant remarked: *"It felt like we were performing my real life."* The audience's attempts to intervene—by confiscating phones, setting boundaries, or introducing "phone-free meals"—illustrated the tension between personal responsibility and systemic design features that exploit psychological vulnerabilities.

These narratives resonate with Montag and Diefenbach (2018), who describe the rise of *Homo Digitalis*, and Meier and Reinecke (2021), who link excessive digital engagement to anxiety, loneliness, and reduced wellbeing. However, participants went further, reframing digital addiction not merely as an individual pathology but as a collective problem rooted in exploitative business models of the attention economy.

The intellectual contribution here is a move beyond reductionist discourses of "screen time" towards a relational, systemic understanding of digital wellbeing. Theatre enabled participants to critically expose how personal struggles with addiction are intertwined with broader socio-economic imperatives of platform capitalism.

4.2.4 Participatory Theatre as a Space of Resistance and Dialogue
The final theme concerned the methodological innovation of the study itself. Participants consistently described theatre as a transformative space for ethical reflection. One participant noted: *"We didn't just discuss these issues—we lived them, and then we changed them."* Through the collective enactment of scenarios, participants experienced what Boal (2000) called a "rehearsal for revolution."

This theme demonstrates that participatory theatre is not merely a data collection tool but an epistemological practice. It generates embodied knowledge that conventional methodologies cannot capture. In contrast to abstract debates about AI ethics, theatre made visible the affective, relational, and moral stakes of technological systems.

Intellectually, this positions theatre as a decolonial methodology (Nicholson, 2016), challenging Western technocratic paradigms by centring lived experiences and collective agency. It also aligns with calls in critical AI ethics for inclusive, participatory approaches that democratise ethical deliberation (Jobin, Ienca & Vayena, 2019).

4.3 Cross-Theme Analysis
Synthesising across the themes reveals three overarching insights:

1. **Embodied Awareness:** Theatre heightened awareness of algorithmic harms by transforming abstract technological processes into tangible, lived experiences. Participants moved from intellectual recognition to affective and moral engagement.
2. **Power and Agency:** Across bias, surveillance, and addiction, participants interrogated asymmetries of power but also rehearsed strategies for reclaiming agency. This demonstrates that ethical engagement is not

passive critique but active rehearsal of alternatives.

3. **Wellbeing as Collective:** Digital wellbeing emerged not as an individual self-discipline but as a collective ethical practice requiring systemic transformation. Theatre thus enabled a shift from neoliberal discourses of personal responsibility to communal visions of flourishing.

4.4 Discussion in Relation to Literature

The findings extend existing scholarship on AI ethics and digital wellbeing. The lived experiences of algorithmic exclusion reinforce Noble's (2018) and Birhane's (2021) critiques, while the pervasive sense of surveillance confirms Zuboff's (2019) theorisation of surveillance capitalism. Similarly, participants' narratives of addiction corroborate empirical research on digital dependency (Montag & Diefenbach, 2018; Meier & Reinecke, 2021).

Yet, the study contributes something novel: it demonstrates that participatory theatre is not merely illustrative but constitutive of ethical inquiry. By staging scenarios, participants generated new insights, collective strategies, and embodied critiques of technological systems. This positions theatre as a methodological intervention into the field of AI ethics, complementing traditional analytical and computational approaches.

Ultimately, the analysis underscores the central argument of this research: that confronting the challenges of bias, surveillance, and addiction in the AI age requires not only technical fixes and policy frameworks but also creative, participatory, and embodied practices that cultivate ethical awareness and collective agency

5.0 SUMMARY, CONCLUSION, AND RECOMMENDATIONS

The study set out to examine the potential of participatory theatre as a methodological and pedagogical tool for addressing ethical challenges posed by AI-driven technologies. Specifically, it sought to interrogate issues of algorithmic bias, surveillance, and digital addiction, while also exploring how embodied performance can foster awareness and dialogue around digital wellbeing.

Section One introduced the background, rationale, and significance of the study, situating it within current debates on AI's social consequences (Mittelstadt, 2019; Jobin, Ienca & Vayena, 2019). Section Two reviewed literature on AI ethics, digital wellbeing, and participatory methods, identifying gaps in existing approaches that prioritise abstract principles over lived experience. Section Three detailed the qualitative, participatory action research design, which employed forum theatre and image theatre as instruments of inquiry (Boal, 2000; Chevalier & Buckles, 2019). Section Four presented the findings, organised around four themes: algorithmic bias and exclusion, surveillance and control, digital addiction and wellbeing, and theatre as a space of resistance and dialogue.

The research generated compelling evidence that participatory theatre is uniquely suited to exposing the ethical stakes of AI technologies. It transformed participants from passive users of digital systems into active critics and co-creators of knowledge, thereby expanding the epistemological and methodological repertoire of AI ethics research.

5.2 Conclusion

The study concludes that participatory theatre constitutes both a critical methodology and an ethical intervention in the AI age. By staging scenarios of algorithmic injustice, surveillance, and digital dependency, participants engaged in embodied reflection that bridged the gap between abstract ethical frameworks and lived realities.

Three key conclusions emerge:

1. AI ethics requires experiential engagement. Existing frameworks, while valuable, often fail to resonate with those most affected by digital systems. This study demonstrates that participatory theatre generates visceral awareness and moral urgency, amplifying voices often marginalised in technocratic debates (Birhane, 2021; Noble, 2018).
2. Digital wellbeing is collective rather than individual. While mainstream discourse frames wellbeing as a matter of self-regulation, participants highlighted the structural and relational dimensions of digital harm. This aligns with critiques of the attention economy (Zuboff, 2019; Montag & Diefenbach, 2018) and calls for systemic interventions in platform design and governance.
3. Theatre is a space of resistance and reimagination. Beyond its role as data collection, participatory theatre acted as a "rehearsal for revolution" (Boal, 2000), enabling participants to envision alternative technological futures grounded in justice, dignity, and human flourishing.

Thus, the research affirms that creative, embodied, and participatory methods are indispensable in shaping an ethical AI future.

5.3 Recommendations
5.3.1 For Education
- Integrate applied theatre and critical digital pedagogy into curricula to cultivate students' ethical awareness of AI systems.

- Encourage interdisciplinary learning that bridges computer science, social sciences, and the arts to produce holistic thinkers equipped for the AI age.
- Use participatory theatre as a teaching tool in ethics courses, allowing learners to rehearse responses to technological dilemmas.

5.3.2 For Work and Industry
- Organisations should adopt participatory workshops as part of employee training on digital rights, wellbeing, and workplace surveillance.
- Tech companies should incorporate user-centred and participatory design approaches that foreground ethical and relational considerations (Whittlestone et al., 2021).
- Workplaces should establish collective wellbeing practices (e.g., phone-free spaces, surveillance audits) to counter digital dependency and over-monitoring.

For Policy and Governance
- Policymakers should mandate inclusive and participatory deliberations in AI governance, ensuring that communities most affected by algorithmic decisions have a voice.
- Regulation should move beyond abstract principles to address practical harms, including algorithmic discrimination, data commodification, and addictive design features.
- Public funding should support arts-based approaches to digital literacy and ethics, recognising the role of culture and creativity in fostering resilience.

5.4 Suggestions for Further Research
The study opens several pathways for future inquiry:
1. **Comparative Studies:** Replicating participatory theatre methodologies across different cultural, socio-economic, and geographic contexts to test transferability.
2. **Longitudinal Research:** Tracking the long-term impact of theatre-based interventions on participants' digital practices and ethical awareness.
3. **Hybrid Methodologies:** Combining theatre with digital simulations or virtual reality to enhance immersion in AI ethics exploration.
4. **Policy-Oriented Research:** Investigating how insights from participatory theatre can concretely inform regulatory frameworks for AI governance.
5. **Youth Engagement:** Examining how younger generations, as digital natives, can leverage theatre to articulate their experiences of algorithmic systems.

5.5 Final Reflection
At its core, this research affirms that the challenges of AI ethics and digital wellbeing cannot be addressed solely through technical or legal frameworks. They require creative, participatory, and human-centred practices that reclaim agency from opaque algorithms and extractive platforms. Participatory theatre, with its emphasis on dialogue, embodiment, and collective imagination, offers precisely such a practice. By enabling individuals and communities to critically rehearse their digital futures, it makes an indispensable contribution to the struggle for justice, dignity, and human flourishing in the AI age.

REFERENCES
- Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. Patterns, 2(2), 100205. https://doi.org/10.1016/j.patter.2021.100205
- Boal, A. (2000). Theatre of the Oppressed. London: Pluto Press.
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806
- Chevalier, J. M., & Buckles, D. J. (2019). Participatory action research: Theory and methods for engaged inquiry. Routledge.
- Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonizing human life and appropriating it for capitalism. Stanford University Press.
- Freire, P. (1970). Pedagogy of the Oppressed. New York: Continuum.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
- Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.
- Meier, A., & Reinecke, L. (2021). Computer-mediated communication, social media, and mental health: A conceptual and empirical meta-review. Communication Research, 48(8), 1182–1209. https://doi.org/10.1177/0093650220958224
- Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4
- Montag, C., & Diefenbach, S. (2018). Towards Homo Digitalis: Important research issues for psychology and the neurosciences at the dawn of the Internet of Things and the digital society. Sustainability, 10(2), 415. https://doi.org/10.3390/su10020415
- Nicholson, H. (2016). Applied drama and theatre: Useful knowledge. Palgrave Macmillan.

- Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York: New York University Press.
- Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2021). Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for research. Science and Engineering Ethics, 27(2), 1–35. https://doi.org/10.1007/s11948-020-00205-8
- Williamson, B. (2021). Education and datafication: Critical questions for learning, data, and governance. Learning, Media and Technology, 46(1), 1–16. https://doi.org/10.1080/17439884.2021.1876084
- Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. London: Profile Books.