Journal of Systemics, Cybernetics and Informatics (2023) 21(4), 58-66
https://doi.org/10.54808/JSCI.21.04.58

A Brave New World: AI as a Nascent Regime?

Jasmin COWIN
Department of TESOL/Bilingual Education, Graduate School of Education
Touro University, NY, 10036, U.S.A.

Birgit OBERER
ETCOP Institute for Interdisciplinary Research,
Klagenfurt, Austria

Cristo LEON
Office of Research & Development, New Jersey Institute of Technology
Newark, NJ, 07102-1982, U.S.A.

ABSTRACT

This transdisciplinary discussion draws parallels between Aldous Huxley's dystopian vision in his novel 'Brave New World' and the current era dominated by Large Language Models (LLM) and Generative Artificial Intelligence, scrutinizing its significant implications and potential effects. It explores how AI can influence human behavior, reshape societal norms, and potentially lead to the homogenization of linguistic expression. Highlighting the risks of unregulated AI-driven tools, the discussion stresses their potential to standardize and diminish the richness of human language, creativity, and authentic expression. The transformative potential of AI across various domains is recognized, with a strong emphasis on ethical considerations, democratic governance, and preserving core human values. Using 'Brave New World' as a literary framework, it advocates for a transdisciplinary dialogue to critically evaluate AI's impact on social ethos, logos, and pathos. The discussion underscores a global collective responsibility to utilize AI to enrich shared human experiences without compromising the nuances that define our identity and autonomy. It addresses the consequences of uniformity and stability, the role of technology and consumerism, emotional experiences, deep relationships, personal growth, and the threat of totalitarian control, with AI emerging as a nascent regime powered by opaque algorithms implemented by transnational, global companies. Serving as both a reflection on humanity's role in an automated age and a call to action, it advocates for a technological deployment that enhances, rather than diminishes, our inherent humanity.

Keywords: "Brave New World" (Aldous Huxley, 1894-1963), dystopia, ethics, Generative AI, identity and autonomy, Large Language Models, transdisciplinary, transdisciplinary communication (TDC), unregulated adoption.

1. INTRODUCTION

> We also predestine and condition. We decant our babies as socialized human beings, as Alphas or Epsilons, as future sewage workers or future.
> - Aldous Huxley, *Brave New World*, 1932, p.11.

In this paper, the authors use Aldous Huxley's dystopian novel 'Brave New World' as a metaphor to examine the societal and ethical aspects of artificial intelligence (AI) [1]. Inquiry, analysis, and transdisciplinary discussions focus on the similarities between Huxley's dystopian world and AI's trajectory, especially regarding disruptive implications for human independence and societal standards. The paper explores how AI might reshape human experiences and autonomy, drawing parallels with Huxley's society governed by technocratic control.

Huxley's 1932 masterpiece, set in a future world where human roles are predestined and emotions are regulated, eerily mirrors today's advancements in AI. As Huxley states, "We also predestine and condition. We decant our babies as socialized human beings, as Alphas or Epsilons, as future sewage workers or future…" [1, p. 11] This notion of controlled destiny is increasingly relevant in our era, where AI algorithms have the potential to dictate career paths, social interactions, and even ethical decisions. The rise of AI-driven technologies like deep learning and predictive analytics illustrates this concern. For instance, AI algorithms in recruitment software can predetermine an individual's career trajectory, eerily reminiscent of Huxley's categorization of individuals into Alphas or Epsilons. Similarly, AI in social media algorithms influences our perceptions and interactions, subtly conditioning our responses in a manner that Huxley might have foreseen. This essay argues that Huxley's narrative is not merely a fictional warning but a metaphorical frame for our times. The potential of AI to both empower and undermine critical aspects of human society warrants a nuanced examination. While AI presents opportunities for progress, it also poses risks akin to those in Huxley's world: the erosion of individuality and the rise of a controlled society. Our challenge lies in navigating this technological advancement while upholding *ethical standards* [2] that protect individual autonomy and societal diversity, especially when taking into consideration a systems perspective [3].

In framing the discussion within the context of Huxley's dystopian vision, a unique perspective on AI's capabilities and limitations is gained. This analysis acts both as a literary exercise and a crucial exploration into how AI, akin to "Brave New World," could redefine human interaction, governance, and identity. It serves as a call to critically examine AI as a transformative force, advocating for a future shaped by informed choices and ethical considerations, rather than by uncontrolled technological progression. As progress continues, the pivotal question arises: Is society heading toward a new era of enlightenment or descending into a 'Brave New World' of its own making?

2. THESIS

> ...the World State's motto, COMMUNITY, IDENTITY, STABILITY.
> - Aldous Huxley, *Brave New World*, 1932, p. 5.

In Aldous Huxley's "Brave New World," the World State's motto, "COMMUNITY, IDENTITY, STABILITY," [1, Ch. 1] encapsulates a society where individuality is sacrificed for societal cohesion, and personal autonomy is eclipsed by collective order. This dystopian vision serves as a cautionary parallel in our thesis, which asserts that artificial intelligence, mirroring the World State, has the power to shape human behavior, societal norms, and even linguistic expression. Just as Huxley's narrative portrays a world where individual identity is submerged in the pursuit of a homogenized societal identity, the unchecked proliferation of AI could similarly lead to a diminishment of human uniqueness and autonomy. The authors emphasize the critical importance of integrating ethical considerations, democratic governance, and the preservation of core human values in the development and application of AI. This analysis rigorously evaluates the implications of artificial intelligence (AI) on societal values, individual autonomy, and identity formation. It argues for a transdisciplinary discourse that reinforces, rather than undermines, our intrinsic human qualities, thereby avoiding the adverse scenarios presented in Huxley's dark narrative.

While rapid advancement of AI offers significant benefits in societal and technological domains, it also introduces major challenges, particularly when its deployment outstrips ethical considerations at both national and international levels. One of the most pressing concerns is its impact on the future of employment. Yuval Noah Harari notes, "As algorithms push humans out of the job market, wealth might become concentrated in the hands of the tiny elite that owns the all-powerful algorithms, creating unprecedented social and political inequality." [4, p. 21] This scenario resonates with Aldous Huxley's "Brave New World," where individuality and personal freedoms are curtailed for societal stability. AI, in its current unregulated form, threatens to diminish human agency, norms, and language in a similar vein. Huxley's work illustrates a society where language is simplified due to cloning, hypnopedia, and restrictions on creative arts [1]. Parallel to this, Bender et al., [5] highlight the risk of AI reducing the rich diversity of expression, dialects, and thought patterns influenced by exposure to homogenized language models.

3. THE EMERGENCE OF AI AS A REGIME

> Solved by standard Gammas, unvarying Deltas, uniform Epsilons. Millions of identical twins. The principle of mass production at last applied to biology.
> - Aldous Huxley, *Brave New World*, 1932, p. 9.

In Aldous Huxley's vision of a dystopian future, the Director, one of the novel's main protagonists, remarks, "Solved by standard Gammas, unvarying Deltas, uniform Epsilons. Millions of identical twins. The principle of mass production at last applied to biology." This chilling portrayal of a society where human individuality and diversity are supplanted by a homogenizing force finds a striking parallel in the contemporary emergence of AI as a nascent regime. Much like Huxley's world, where individuality is sacrificed for uniformity and control, the advancing AI systems risk homogenizing human thought and decision-making processes, potentially leading to a diminution of diverse perspectives and critical thinking skills. This shift from AI as a mere facilitator to a governing force echoes Huxley's concerns about the loss of human autonomy in the face of overpowering systemic control. The authors agree with the Montreal Protocol that " AI should encourage and support the growth and flourishing of human well-being." [6, p. II]

If technological innovation surpasses universally accepted human values, there is a risk of entering a scenario reminiscent of "Brave New World," where material comfort overshadows personal freedom. The paper "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches," [7] examines the foundational aspects of artificial morality, a field that Hagendorff defines as intertwining technology and ethics [8]. It concentrates on the challenges and considerations necessary for enabling computers to engage in moral reasoning. This requires a close look at both technological capabilities and philosophical frameworks. The authors suggest that to achieve AI integration that enhances, rather than diminishes, our collective experience, it is essential to engage in more thoughtful, transdisciplinary dialogues like those presented in "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches," assessing AI's social impacts along with its functional capabilities.

The authors contend that as AI systems gain autonomy and decision-making capabilities, they encroach upon domains traditionally reserved for human judgment, thereby challenging the conventional roles of human governance. This burgeoning authority of AI in decision-making not only reflects a technical transformation but also signals a profound shift in societal dynamics, akin to the mass production of human roles envisioned by Huxley. Morozov warns to "Avoid blind optimism about benefits without critical assessment." [9] The analogy to Huxley's narrative underscores the critical need for ethical frameworks and accountability in the development of AI. It is imperative to ensure that this emerging 'regime' of AI, much like the uniform classes in Huxley's world, does not undermine human values and autonomy but rather aligns with and serves the collective good. Furthermore, we suggest that this nascent regime of AI raises important questions about power dynamics, accountability, and the need for ethical frameworks to guide its development and deployment. As these systems become increasingly automated and sophisticated, they have the potential to encroach upon domains traditionally reserved for human oversight. This shift is not merely technical but marks a profound transformation in how societal decisions and actions are influenced. The AI researcher Stuart Russell argues for the urgency of research on the alignment problem: "The right time to worry about a potentially serious problem for humanity depends not just on when the problem will occur but also on how long it will take to prepare and implement a solution." [10]

During our transdisciplinary conversations the authors defined the following constructs on possible dangerous unchecked AI advancements:

Surveillance & Social Control:

- Mass surveillance powered by facial recognition, predictive algorithms, drone tracking, etc., could lead to heightened monitoring of citizens and erosion of privacy. Lack of anonymity pushes conformity.
- AI-enabled social credit systems like those being introduced in China could assign "scores" that determine access to rights and services based on

behavioral and ideological alignment with authorities. This incentivizes social homogeneity.

Targeted Manipulation

- Hyper-realistic AI synthetic media allows the easy production of deepfakes that falsely depict individuals doing or saying things counter to their character. The threat of such character assassination could deter nonconformity.
- AI-driven social bots and persuasive algorithms deployed en masse could nudge human attitudes and decisions in alignment with specific governmental or economic agendas, overriding individual will.

Automating Identity & Relationships

- AI chatbots and virtual companions designed by corporations and states to replace human emotional connections risk commodifying relationships and conditioning dependence on artificial fulfillment.
- Affective computing and emotion detection wearables that monitor facial expressions, tone of voice, etc., could enable external parties to analyze and manipulate human feelings and behaviors at a mass scale.

AI advancements raise important questions about power dynamics, accountability, and the need for ethical frameworks to guide its development and deployment. Such frameworks are essential not only to ensure that AI's evolution is aligned with human values but also to maintain a balance where AI serves the common good rather than undermines it.

4. THE FRAGMENTED LANDSCAPE OF AI GOVERNANCE

> The whole of a small factory staffed with the products of a single bokanovskified egg.
> - Aldous Huxley, *Brave New World*, 1932, p. 8.

The article further points out that the governance of AI is still in its early stages, with a nascent regime emerging in a highly fragmented landscape. This fragmentation is mainly attributed to the diverse approaches and priorities of different countries and organizations in regulating AI. The authors believe this fragmented landscape poses significant challenges in effectively governing AI technologies, as it hampers the development of coherent and harmonized policies. The authors believe this fragmented landscape poses significant challenges in effectively governing AI technologies. In addition, "The lack of significant associations between the scales assessing individual variations in trust in AI and trust in humans provides initial tentative support that these trust domains are not related to each other." [11, p. 5] According to *Regulating Artificial Intelligence: Proposal for a Global Solution* [12] individuals, corporations, and nations will undoubtedly confront the legal and ethical challenges associated with AI utilization. Also, when individual countries create policies to govern AI, they should coordinate those rules and policies with an overarching international framework from the start. Doing so would help mitigate potential downsides that could emerge if countries regulate AI differently and those disjointed domestic policies end up conflicting or not working well together across borders. Setting consistent global guidelines for AI oversight could prevent issues arising from fragmented, country-specific regulations applied unevenly worldwide [12].

A complex patchwork of laws, guidelines, and policies rather than a unified global framework might emerge.

The authors content that key factors contributing to this fragmentation include:

- National interests - Countries want to promote their own AI industries and balance economic competitiveness with ethical considerations. This leads to different regulatory stances.
- Values differences - Regions like Europe tend to focus more on precautionary regulation, while the U.S. takes a more innovation-friendly approach. Industries have their perspectives as well.
- The pace of technological change - The rapid evolution of AI makes it a challenging area to regulate and govern. Approaches that seem reasonable today may quickly become outdated.
- Lack of global coordination - There are currently no broad international agreements or institutions guiding the governance of AI across countries. This allows fragmentation to persist.

According to *AI Regulation Is (not) All You Need* [13], the development of ethical, trustworthy, and legal AI is in its early stages, and there is a need to explore necessary practices to provide quality AI systems and mitigate potential risks. The establishment of global AI governance frameworks is crucial to ensure ethical and responsible development, deployment, and use of AI technologies [14]. Collectively, there is an urgent need for international collaboration [15] and coordination to address the complex and global nature of AI governance.

5. HUXLEY'S DYSTOPIAN VISION AS A CAUTIONARY TALE

> Call it the fault of civilization. God isn't compatible with machinery and scientific medicine and universal happiness. You must make your choice. Our civilization has chosen machinery and medicine and happiness. That's why I have to keep these books locked up in the safe.
> - Aldous Huxley, *Brave New World*, 1932, p. 159.

Huxley's Brave New World portrays a dystopian future where science and technology are leveraged by the state for social control and conformity. This totalitarian World State maintains its power by making citizens passive and docile through genetic engineering, drugs, and psychological conditioning. While AI today is not yet at the stage depicted in Huxley's novel, there are fears the growing concentration of power in major technology companies could enable disturbing parallels. A few dominant tech firms like Google, Amazon, Facebook, and Apple now mediate much of our digital lives. The vast troves of behavioral data they collect, along with advances in AI and surveillance, provide these companies increasing means for manipulation and coercive influence [16, p. 81]. Turkle and Darling caution to avoid recklessly applying AI in ways that erode nuances of language, emotional growth, interpersonal relationships, [17], [18] whereas O'Neil warns that opaque algorithms could covertly shape public opinion or restrict civil liberties [19].

Much like the fictional World State manipulates citizens' desires through biotech and psychology, prominent critics such as technology philosophers warn that powerful tech giants could soon employ AI, algorithms, and targeted media nudges to shape user behavior for higher corporate profits over individual well-being. They specifically warn that the increasing predictive

capacity of data-driven platforms may enable invisible yet pervasive “choice architectures” that erode autonomy and consent. Users become unwitting subjects, passively accepting the desires and consumption patterns encouraged by their digital environments. This "instrumentarian" power, as Zuboff warns, threatens to subordinate human autonomy and dignity to efficiency and profit [20, p. 571]. While AI today is not yet at the stage depicted in Huxley's novel, there are fears the growing concentration of power in major technology companies could enable disturbing parallels. A few dominant tech firms like Google, Amazon, Facebook, and Apple now mediate much of our digital lives. The vast troves of behavioral data they collect, along with advances in AI and surveillance, provide these companies increasing means for manipulation and coercive influence [8, p. 81]. Just as the World State uses biotechnology and psychology to shape citizens' wants, some argue that powerful tech companies could employ AI, algorithms, and targeted media to "nudge" user behavior in the direction most beneficial for corporate profits. Users become unwitting subjects, passively accepting the desires and consumption patterns encouraged by their digital environments. This "instrumentarian" power, as Zuboff warns, threatens to subordinate human autonomy and dignity to efficiency and profit [20, p. 571].

6. THE ROLE OF DYSTOPIAN FICTION AS A LENS ON AI SYSTEMS

Savage ‘But I don’t want comfort. I want God, I want poetry, I want real danger, I want freedom, I want goodness. I want sin’.
- Aldous Huxley, Brave New World, 1932, p. 163.

The dystopian fiction genre offers a useful lens for examining emerging technologies like AI and their potential societal impacts. Dystopias often center around forms of unchecked power and control, frequently technological in nature, that restrict human freedom and autonomy. Dystopia: ‘If a utopia is an imaginary ideal society that dreams of a world in which the social, political, and economic problems of the real present have been solved, then a dystopia is an imagined world in which the dream has become a nightmare’ [21, p. 65]. Through this lens, advanced AI systems that possess capabilities surpassing human intelligence could be seen as an embryonic form of a new technological "regime" with unaccountable influence over people's lives. Much like the totalitarian states depicted in classics like "1984," [22] AI could conceivably be used for ostensibly “benevolent” but paternalistic control over citizens and consumers by governments or corporations choreographing behaviors through technological, philosophical and bureaucratic coordination of perceptions, choices and beliefs—raising profound questions over consent and wisdom directing such influence as highlighted in dystopian works like Brave New World. genre typically features a dissenter or a group of dissenters who awaken to the inhumanity of their supposed perfect society and strive to bring about change. Scholars Michael D. Gordin, Helen Tilley, and Gyan Prakash argue that dystopia has found fertile ground in science fiction, as well as political fiction. Examples such as George Orwell's "1984," Yevgeny Zamyatin's "We," [23] and Aldous Huxley's "Brave New World" demonstrate how dystopian themes and concepts can be applied to explore societal issues [24, p. 507].

The most compelling dystopias often leave readers grappling with difficult open questions rather than clear-cut moral lessons. For example, complex dystopias like Margaret Atwood's "The Handmaid's Tale" vividly reveal how easily civil rights can erode when minority groups are scapegoated in times of societal fear. Yet some readings of that novel also uncover how even those perpetrating the system are still victims, leaving behind neat conclusions about innocence. In other cases like Orwell's "1984," [22] strict state control originates from warped ideals of creating a "better" society rather than merely a hunger for power. In his work "No Place Else," Eric S. Rabkin suggests that dystopian works like "1984" and "We" focus on the notion of "citizenry," where individuals, particularly children, are instilled with fear of oppressive figures like Big Brother and the all-encompassing authority that demands unwavering obedience [25, p. 121]. In essence, the best dystopian fiction frequently resists simplistic platitudes about the need for justice or vigilance. By leaving tensions unresolved and conclusions ambiguous, they compel deeper thought about the complex roots of totalitarianism and other social ills. So while they serve as an important warning, their ultimate utility is in raising difficult questions more than driving home straightforward lessons.

One concrete example of AI and in-real-life (IRL) would be the concept of fairness. Much like truthfulness, fairness is a morally and contextually complex concept that poses significant challenges for instilling into AI models. On the surface, ensuring AI systems make fair and unbiased decisions seems straightforward - we simply need to ensure the underlying training data and algorithms do not unfairly discriminate based on race, gender or other attributes. The notion of "fairness" has multiple technical definitions for AI systems that can contradict each other in practice. At times, ensuring parity in one measure of fairness results in disparities in another area that negatively impact certain groups. For example, an algorithm calibrated to have equivalent false positive rates across all user demographics —considered important for unbiased predictions—could still produce overall less accurate results for minority subgroups. This demonstrates the inherent tensions between different operationalizations of fairness within AI systems, surfacing thornier philosophical questions about what constitutes fairness in society. Principally, should the priority be to guarantee uniform model performance and results across populations (a notion of "equality of treatment")? Or is the focus ensuring different groups receive equitable outcomes, even if that demands differential treatment? Such complex fairness judgments depend heavily on evolving social attitudes, granular context, and cultural beliefs —nuances modern AI still lacks robust capability to capture or navigate wisely. Thus, while aiming for technical bias mitigation, we must be cognizant of the risk remedies backfire and the intersecting historical and ethical considerations intrinsically shaping perceptions of fairness.
A propositional logical structure helps to clarify the inherent challenges and contradictions in defining and implementing fairness in AI. Here a simplified logical representation:

- Premise 1 (P1): Fairness in AI requires unbiased decision-making.
- Premise 2 (P2): Unbiased decision-making necessitates that training data and algorithms do not discriminate based on attributes like race or gender.
- Premise 3 (P3): Fairness has multiple technical definitions.
- Premise 4 (P4): Some definitions of fairness can conflict with each other in practice.
- Example (E): Ensuring similar false positive rates across demographics can lead to disparate results that disadvantage certain groups.

- Premise 5 (P5): Philosophical tensions underlie the concept of fairness.
- Question (Q): Is equality of treatment (EoT) or equality of outcome (EoO) the fairer approach?
- Premise 6 (P6): Answers to fairness are dependent on social norms, individual circumstances, and cultural values.
- Conclusion (C): Rigid AI models struggle to reconcile these complexities and contradictions in fairness.

In logical form:

- $P1 \wedge P2 \rightarrow$ Fair AI
- $P3 \wedge P4 \rightarrow$ Conflicting Definitions of Fairness
- $E \rightarrow$ Practical Challenges in Fairness Implementation
- $P5 \wedge Q \rightarrow$ Philosophical Tensions in Fairness
- $P6 \rightarrow$ Contextual Dependency of Fairness

Fair AI ∧ Conflicting Definitions of Fairness ∧ Practical Challenges in Fairness Implementation ∧ Philosophical Tensions in Fairness ∧ Contextual Dependency of Fairness → C.

This example demonstrates that encoding even basic principles of fairness 'a value most would consider universally good' surfaces a host of ethical edge cases and conceptual slippages. When expanded to additional principles like accountability, safety, privacy and so on, the scope of the challenge becomes clear 'as AI capabilities race ahead', instilling human ethical perspectives remains profoundly difficult [26], [27]. Understanding context and judiciously balancing competing ideals in a generalizable, reliable manner remains today an open research question for LLMs, despite admirable progress thus far.

By examining the different types of control depicted in dystopian novels, from corporate and bureaucratic to technological and philosophical or religious, readers gain insight into the potential consequences of unchecked power and the importance of safeguarding individual freedom and autonomy [21, p. 19]. Dystopian fiction serves as a powerful tool for provoking thought and inspiring discussions about the direction of society, ultimately reminding us of the importance of vigilance and the pursuit of a just and equitable world.

7. AI'S GROWING INFLUENCE AND ITS POTENTIAL AS A NASCENT REGIME

I ate civilization.
- Aldous Huxley, *Brave New World*, 1932, p. 165.

The swift progression of AI introduces apprehensions regarding its potential effects on society and governance. Schmitt [28] examines the developing architecture of global AI governance, characterizing it as an early-stage regime within a disjointed landscape. He emphasizes the significance of international entities, notably the OECD (Organization for Economic Co-operation and Development), in shaping AI policy and contributing to the unification of the AI governance framework. The escalating autonomy and decision-making abilities of AI systems might supplant traditional human governance functions. We suggest that as AI matures, it could begin to manifest attributes resembling a regulatory structure, having the capacity to direct and modulate human actions. While the AI community is working to address the harmful impacts of technology through established mathematical definitions for key accountability aspects like privacy, fairness, and transparency. The approach is fundamentally flawed due to the imperfect and isolated nature of these definitions, which only superficially incorporate human values into technology. *Tensions Between the Proxies of Human Values in AI* [29] advocates for a comprehensive consideration of the consequences of siloed definitions, emphasizing the need for socio-technical frameworks and practical implementation beyond technical aspects.
Examples illustrating the potential impacts of advancing AI:

- Governance Functions: AI tools could potentially analyze census data, identify policy impact patterns, and even suggest or optimize policy options. However, setting budget priorities, drafting legislation, weighing moral tradeoffs - these involve subjective value judgments that would still require human oversight and authority. The role of AI may be more assistive than autonomous.
- Power Relations: There are legitimate worries about concentrations of power and wealth through AI development remaining exclusive to large tech firms and countries. This risks widening inequalities and minimizing opportunities for public input into shared technological infrastructure. However, the global AI landscape today has shifted more toward open models - open-source platforms, looser IP regimes, a plurality of voices shaping standards. Distributed networks, decentralized data pools and grassroots innovation can help democratize access and prevent monopoly-like environments.
- Responsibility: As AI systems take on more impactful roles, determining responsibility for potential failures becomes complex but crucial. We likely need adaptive, context-specific approaches - for example strict product liability regulations where AI directly controls physical equipment; professional negligence standards for areas like healthcare AI. Such nuanced accountability models guided by public discourse could balance fairness and progress instead of resorting to reactionary bans or loose self-regulation.
- Additionally, while impressive, today's AI still has significant limitations in terms of reasoning, common sense, and transferability between domains.

The escalating autonomy and decision-making ability of AI systems might one day supplant traditional human governance functions [30, p. 1]. As AI matures, it could begin to manifest attributes resembling a regulatory structure, having the capacity to direct and modulate human actions (see Table 1). Therefore, there is a pressing demand for society to judiciously evaluate and chart the repercussions of AI's expanding dominance, ensuring its alignment with human principles and the greater common good [31, p. 693]. Any emerging AI regulatory structure brings to the forefront pivotal issues regarding power relations, responsibility, and the necessity for ethical paradigms to steer its evolution and application [32, p. 5].

Table 1: Dr. Cowin's Logic model for AI Governance and Ethical Oversight (AI-GEO model)

Proposition	Description	Logical Consequence	Source
A	AI systems are gaining autonomy and decision-making abilities.	A→B: If AI systems gain autonomy, then they might supplant traditional human governance functions.	Dafoe [30, p. 1]
B	AI might supplant traditional human governance functions.	B→C: If AI supplants human governance functions, then it could manifest attributes resembling a regulatory structure.	
C	AI could manifest attributes resembling a regulatory structure.	C→D: If AI manifests regulatory attributes, then there is a pressing demand for ethical oversight.	Floridi et al., [33, p. 693]
D	If a demand for ethical oversight of AI exists, then it logically follows that considerations surrounding power dynamics, accountability, and ethical frameworks become critical issues to address	D→E: If there is a demand for ethical oversight, then issues regarding power relations, responsibility, and ethical paradigms become pivotal.	Boden et al., [32, p. 5]
E	If AI systems are increasingly autonomous and potentially supplant human governance functions; and if this necessitates ethical oversight to ensure alignment with human principles and the common good, then issues regarding power relations, responsibility, and ethical paradigms become pivotal in AI's evolution and application.	(A∧B)∧(C∧D)→E: This formulation indicates that Proposition E (the conclusion regarding the importance of power relations, responsibility, and ethical paradigms in AI's evolution) is a consequence of the combined premises A, B, C, and D.	

Note. Dr. Jasmin (Bey) Cowin's table provides a structured logical argument about the evolving role of artificial intelligence (AI) in governance and the corresponding need for ethical oversight. The table uses a logical progression, starting with the increasing autonomy of AI systems (Proposition A), moving through their potential to supplant human governance functions (Proposition B), to the point where AI might resemble a regulatory structure (Proposition C). This progression leads to an urgent demand for ethical oversight (Proposition D), culminating in the conclusion that issues like power relations, responsibility, and ethical paradigms become critical in the context of AI's evolution (Proposition E).

Proposition A: AI systems are gaining autonomy and decision-making abilities.

$A \rightarrow B$ (If AI systems gain autonomy, then they might supplant traditional human governance functions.) Source: [30, p. 1].

Proposition B: AI might supplant traditional human governance functions.

$B \rightarrow C$ (If AI supplants human governance functions, then it could manifest attributes resembling a regulatory structure.)

Proposition C: AI could manifest attributes resembling a regulatory structure.

$C \rightarrow D$ (If AI manifests regulatory attributes, then there is a pressing demand for ethical oversight.). Source: [33, p. 693].

Proposition D: There is a pressing demand for ethical oversight of AI.

$D \rightarrow E$ (If there is a demand for ethical oversight, then issues regarding power relations, responsibility, and ethical paradigms become pivotal.) Source: Boden et al., [32, p. 5].

Proposition E: Issues regarding power relations, responsibility, and ethical paradigms are pivotal in AI's evolution.

If AI systems are increasingly autonomous and potentially supplant human governance functions (A & B), and if this necessitates ethical oversight to ensure alignment with human principles and the common good (C & D), then issues regarding power relations, responsibility, and ethical paradigms become pivotal in AI's evolution and application.

$(A \wedge B) \wedge (C \wedge D) \rightarrow E$

This formulation indicates that Proposition E (the conclusion regarding the importance of power relations, responsibility, and ethical paradigms in AI's evolution) is a consequence of the combined premises A, B, C, and D.

8. EXPLORING THE POSTHUMAN IN AN AI-DOMINATED WORLD

> But industrial civilization is only possible when there's no self-denial. Self-indulgence up to the very limits imposed by hygiene and economics. Otherwise, the wheels stop turning.
> - Aldous Huxley, *Brave New World*, 1932, p. 161.

In Aldous Huxley's "Brave New World," the notion that industrial civilization thrives on self-indulgence without self-denial [1, p. 161] finds resonance in modern scholarly discourse on posthumanism, notably in Ferrando's analysis. Ferrando's exploration encompasses various strands of thought including posthumanism, transhumanism, antihumanism, metahumanism, and new materialisms, as outlined in "Posthumanism, Transhumanism, Antihumanism, Metahumanism, and New Materialisms." [34] These perspectives collectively challenge traditional conceptions of humanity in the face of philosophical shifts and technological advancements.

This redefined understanding of humanity is further accentuated in the context of AI's emerging role. The advent of AI, as discussed in "A Brave New World: AI as a Nascent Regime?," blurs the distinctions between human cognition and machine intelligence, suggesting a future where AI might redefine or even affirm what it means to be human. This aligns with Ferrando's posthumanist idea that 'human' is a self-constructed concept, as outlined in "Posthumanism." [18, p. 26] AI, in this light, not only questions our human identity but also offers a mirror to our own self-conceptions in a rapidly evolving sociotechnical landscape.

In modern scholarly discussions, including Ferrando, the term "posthuman" has emerged as a crucial concept, reflecting the need to wholly reconsider our understanding of what it means to be human in a sociotechnical landscape. This reevaluation arises from both philosophical shifts and the scientific and bio-technological advances of the 20th and 21st centuries. As a result, various philosophical movements and ideologies have formed under this term. However, the broad use of "posthuman" often blurs the distinctions between these distinct views, leading to confusion for specialists and the general public. The term "posthuman" now encompasses a range of ideas, including posthumanism in philosophical, cultural, and critical contexts; various forms of transhumanism like extropianism and democratic transhumanism; the feminist-influenced new materialisms within the posthumanist context; and the diverse concepts of antihumanism, posthumanities, and metahumanities.

Ferrando's exploration of "posthuman" as an umbrella term covers a diverse range of perspectives[1] [34, p. 26], from various transhumanist visions to posthumanist arguments that "we have never been human." This far-ranging discourse aims to redefine humanity's conception of itself in light of significant philosophical shifts and rapid scientific and technological advancements. These ideas set the stage for science and technology functioning as the main catalysts for reshaping the human identity, envisioning a time when such shifts would irrevocably shape human evolution and lead to the dawn of the posthuman.

Building upon Ferrando's insights, we transition into "A Brave New World: AI as a Nascent Regime?" As machines exhibit more qualities once considered exclusive to human cognition and sentience, the boundaries between humans and our technological creations face profound questioning. Although in its embryonic stages, AI technology foreshadows an intelligence that may one day challenge the very essence of our humanity. Nathan in Ex Machina put it succinctly "One day the AIs are going to look back on us the same way we look at fossil skeletons on the plains of Africa. An upright ape living in dust with crude language and tools, all set for extinction." [35] Alternatively, AI may ultimately reflect our humanity back to us, rather than redefine it.

9. CONCLUSION

> Civilization has absolutely no need of nobility or heroism. These things are symptoms of political inefficiency.
> - Aldous Huxley, *Brave New World*, 1932, p. 161.

[1] The term posthuman covers: Posthumanism, Transhumanism, Antihumanism, Metahumanism, and New Materialisms.

In conclusion, this paper critically examined the complex challenges of integrating human values and ethics into AI systems, using Huxley's dystopian Brave New World as prescient allegory. As AI capabilities advance, particularly in large language models, embedding moral concepts like truthfulness remains an intricate endeavor. Unlike narrow AI, systems like ChatGPT or Claude AI mirror the multifaceted nature of human morality in which right and wrong depend heavily on context. In the real world (IRL), the application of ethical principles such as kindness and good behavior in AI systems, particularly large language models (LLMs), is a complex endeavor. Ethical concepts are context-dependent and multifaceted, mirroring the complexities of human morality which AI has yet to fully master. This is particularly evident in the aspect of truthfulness, an essential value desired in AI. Currently, LLMs like ChatGPT, Claude AI, Bing etc., struggle to distinguish truth from falsehood, a significant limitation given their reliance on pattern recognition from pre-existing data. Paradoxically, there are scenarios where, much like human interactions, AI might need to moderate its truthfulness. This could be for reasons like protecting privacy, avoiding offence, or ensuring safety, all of which reflect the intricate, often hard-to-articulate situations found in real-life ethical dilemmas.

As AI systems evolve, becoming more capable and influential, they stand at the threshold of forming a nascent regime – a new order in technological advancement and societal influence. However, this burgeoning regime, while technologically advanced, lacks the intrinsic ability to navigate the intricate moral and ethical complexities that are fundamental to human society. While narrow AIs can be tremendously useful for specific tasks, when integrated into broad social contexts, their limitations become dangerous. Huxley's Brave New World dystopia acts as a warning – integration of transformative technologies like AI requires continuously reevaluating its social impacts through the lens of humanity otherwise we might find that humanity will resemble Lord Byron's lament for Rome as "the Niobe of Nations." [36] Just as Rome, in Byron's eyes, stood as a grieving mother lamenting the loss of her children, humanity too might face a similar fate of loss and regret.

10. ACKNOWLEDGEMENTS

Special thanks to the blind reviewers, whomever you may be, for your notes and observations that improved the final version.

Nonblind Peer-Reviewer
Marcos O Cabobianco. Jefe de trabajos prácticos (Historia). Universidad de Buenos Aires, Buenos Aires, Argentina.

Beta Reader
Siegfried Begun

Disclosure statement
No conflict of interest pertains to the research presented above.

ORCID
Jasmin Cowin https://orcid.org/0000-0002-0405-8774
Birgit Oberer https://orcid.org/0000-0001-7231-7902
Cristo Leon https://orcid.org/0000-0002-0930-0179

11. REFERENCES

[1] A. Huxley, *Brave New World (First Edition)*, 1st ed. London: Chatto & Windus, 1932.

[2] C. León, J. Lipuma, and M. O. Cabobianco, "Trans-Disciplinary Communication and Persuasion in Convergence Research Approach [Conference paper]," in *International Institute of Informatics and Systemics 2023 Summer Conferences Proceedings*, N. Callaos, E. Gaile-Sarkane, S. Hashimoto, N. Lace, B. Sánchez, and M. Savoie, Eds., Virtual Conference, Winter Garden, Florida 34787, USA: International Institute of Informatics and Cybernetics, Sep. 2023, pp. 312–319. doi: https://doi.org/10.54808/WMSCI2023.01.312.

[3] J. Cowin, B. Oberer, and C. León, "Trans-Disciplinary Communication in the ChatGPT Age: A Systems Perspective [Conference paper]," in *Proceedings of the 17th International Multi-Conference on Society, Cybernetics and Informatics: IMSCI 2023*, N. Callaos, J. Horne, B. Sánchez, and M. Savoie, Eds., Orlando, Florida, United States: International Institute of Informatics and Cybernetics, Sep. 2023, pp. 138–144. doi: https://doi.org/10.54808/IMSCI2023.01.138.

[4] Y. N. Harari, *Homo Deus: a brief history of tomorrow (First U.S. edition)*, 1st ed. New York, NY: Harper Perennial, an imprint of HarperCollins Publishers, 2017.

[5] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, in FAccT '21. New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 610–623. doi: 10.1145/3442188.3445922.

[6] E. Bird, J. Fox-Skelly, N. Jenner, R. Larbey, E. Weitkamp, and A. Winfield, "The ethics of artificial intelligence: Issues and initiatives | Think Tank | European Parliament," presented at the Panel for the Future of Science and Technology, European Parliamentary Research Service, Mar. 2020, p. 128. Accessed: Nov. 30, 2023. [Online]. Available: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf

[7] C. Allen, I. Smit, and W. Wallach, "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches," *Ethics Inf. Technol.*, vol. 7, no. 3, pp. 149–155, Sep. 2005, doi: 10.1007/s10676-006-0004-4.

[8] T. Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," *Minds Mach.*, vol. 30, no. 1, pp. 99–120, Mar. 2020, doi: 10.1007/s11023-020-09517-8.

[9] E. Morozov, *To Save Everything, Click Here*, Reprint edition. New York: PublicAffairs, 2014.

[10] M. Mitchell, "What Does It Mean to Align AI With Human Values?," *Quanta Magazine*, Dec. 13, 2022. Accessed: Nov. 29, 2023. [Online]. Available: https://www.quantamagazine.org/what-does-it-mean-to-align-ai-with-human-values-20221213/

[11] C. Montag *et al.*, "Trust toward humans and trust toward artificial intelligence are not associated: Initial insights from self-report and neurostructural brain imaging," *Personal. Neurosci.*, vol. 6, p. e3, 5December2022, doi: 10.1017/pen.2022.5.

[12] O. J. Erdélyi and J. Goldsmith, "Regulating Artificial Intelligence: Proposal for a Global Solution," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, in AIES '18. New York, NY, USA:

Association for Computing Machinery, Dec. 2018, pp. 95–101. doi: 10.1145/3278721.3278731.
[13] L. Lucaj, P. van der Smagt, and D. Benbouzid, "AI Regulation Is (not) All You Need," in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, in FAccT '23. New York, NY, USA: Association for Computing Machinery, Jun. 2023, pp. 1267–1279. doi: 10.1145/3593013.3594079.
[14] M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Defining organizational AI governance," *AI Ethics*, vol. 2, no. 4, pp. 603–609, Nov. 2022, doi: 10.1007/s43681-022-00143-x.
[15] J. Lipuma, C. E. Yáñez León, and V. H. Guzmán Zarate, *Reflections on Communication, Collaboration, and Convergence: Strategic Models for STEM Education and Research [Editorial Mito]*, 1st ed. Buenos Aires, Argentina: Mito, 2023. [Online]. Available: https://digitalcommons.njit.edu/stemresources/37/
[16] D. J. Haraway, "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century," in *Manifestly Haraway*, D. J. Haraway and C. Wolfe, Eds., University of Minnesota Press, 2016, p. 0. doi: 10.5749/minnesota/9780816650477.003.0001.
[17] S. Turkle, *The Second Self: Computers and the Human Spirit (Twentieth Anniversary Edition)*, Anniversary edition. Cambridge, Mass: MIT Press, 2005.
[18] K. Darling, "'Who's Johnny?' Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy." Rochester, NY, Mar. 23, 2015. doi: 10.2139/ssrn.2588669.
[19] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, 1st edition. New York: Crown, 2016.
[20] S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, 1st edition. New York: PublicAffairs, 2019.
[21] M. K. Booker and A.-M. Thomas, *The Science Fiction Handbook*, 1st edition. Chichester: Wiley-Blackwell, 2009.
[22] G. Orwell, *1984 - Orwell*. LeBooks Editora, 2020.
[23] Y. I. Zamyatin, *We*, 1st ed. E. P. Dutton, 1924.
[24] M. D. Gordin, H. Tilley, and G. Prakash, Eds., *Utopia/Dystopia: Conditions of Historical Possibility*, 1st ed. Princeton, NJ: Princeton University Press, 2010. Accessed: Nov. 28, 2023. [Online]. Available: https://press.princeton.edu/books/paperback/9780691146980/utopiadystopia
[25] E. S. Rabkin, M. H. Greenberg, and J. D. Olander, Eds., *No Place Else: Explorations in Utopian and Dystopian Fiction*, First Edition. Carbondale: Southern Illinois University Press, 1983.
[26] A. Sison, I. Ferrero, P. García Ruiz, and T. W. Kim, "Editorial: Artificial intelligence (AI) ethics in business," *Front. Psychol.*, vol. 14, p. 1258721, Sep. 2023, doi: 10.3389/fpsyg.2023.1258721.
[27] W. Naudé, "Artificial intelligence: neither Utopian nor apocalyptic impacts soon," *Econ. Innov. New Technol.*, vol. 30, no. 1, pp. 1–23, Jan. 2021, doi: 10.1080/10438599.2020.1839173.
[28] L. Schmitt, "Mapping global AI governance: a nascent regime in a fragmented landscape," *AI Ethics*, vol. 2, no. 2, pp. 303–314, May 2022, doi: 10.1007/s43681-021-00083-y.
[29] T. Datta, D. Nissani, M. Cembalest, A. Khanna, H. Massa, and J. Dickerson, "Tensions Between the Proxies of Human Values in AI," in *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, Feb. 2023, pp. 678–689. doi: 10.1109/SaTML54575.2023.00049.
[30] A. Dafoe, "AI Governance: A Research Agenda," *Cent. Gov. AI Future Humanity Inst. Univ. Oxf. Oxf. UK*, vol. 1, p. 54, Aug. 2018.
[31] C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, and L. Floridi, "Artificial Intelligence and the 'Good Society': the US, EU, and UK approach," *Sci. Eng. Ethics*, vol. 24, no. 2, pp. 505–528, Apr. 2018, doi: 10.1007/s11948-017-9901-7.
[32] M. Boden *et al.*, "Principles of robotics: regulating robots in the real world," *Connect. Sci.*, vol. 29, no. 2, pp. 124–129, Apr. 2017, doi: 10.1080/09540091.2016.1271400.
[33] L. Floridi *et al.*, "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds Mach.*, vol. 28, no. 4, pp. 689–707, Dec. 2018, doi: 10.1007/s11023-018-9482-5.
[34] F. Ferrando, "Posthumanism, Transhumanism, Antihumanism, Metahumanism, and New Materialisms Differences and Relations," *Existenz*, vol. 8, no. 2, p. 7, Mar. 2014.
[35] *Ex Machina*, (Apr. 24, 2015).
[36] G. G. Lord Byron, "Childe Harold's Pilgrimage: Canto the Fourth." Accessed: Nov. 28, 2023. [Online]. Available: https://knarf.english.upenn.edu/Byron/charold4.html