

Ljubisa Bojić¹

Senior Research Fellow

Institute for Philosophy and Social Theory, Digital
Society Lab

HUMAN AUTONOMY AND WELLBEING IN THE AGE OF AI: TOWARDS THE NEW SOCIAL CONTRACT?

Abstract: *This paper challenges the pervasive myth of technological neutrality by unmasking the intrinsic ways in which digital devices, algorithms, and infrastructures encapsulate social, political, and economic imperatives from their inception to deployment. Drawing on seminal works by Winner, O'Neil, Heidegger, and Zuboff, among others, the study explores how every design decision—from smartphone interfaces to recommendation algorithms—carries latent biases and agendas that shape human behavior and societal power structures. The analysis traverses the evolution of technology as a cultural force, beginning with the transformative impact of the iPhone and extending to the increasingly immersive realms of the metaverse. By invoking Plato's allegory of the cave, the paper illustrates how algorithmic mediation creates a curated reality that narrows public discourse while eroding individual autonomy. Contemporary discussions on generative AI reveal further complexities, including opaque decision-making processes and emergent risks to cognitive agency and democratic participation. Concluding with a proposal for a new social contract, the*

¹ ljubisa.bojic@ifdt.bg.ac.rs; ORCID: 0000-0002-5371-7975

paper calls for interdisciplinary regulatory reforms, digital literacy programs, and global cooperation to ensure that technological progress shifts towards human empowerment and participatory democracy.

Key words: *Technological Neutrality, Algorithmic Mediation, Digital Culture, Artificial Intelligence, Social Contract.*

Introduction

In recent decades, the rapid expansion of digital technologies has transformed interactions, information flows, and social structures in ways that challenge traditional assumptions regarding the neutrality of technological artifacts. The prevailing notion of technological neutrality—that devices, software algorithms, and digital infrastructures serve merely as impartial tools enabling human intentions—is increasingly questioned by scholars across multiple disciplines. Research by Winner (1986), O’Neil (2016), and Heidegger (1954) has underscored that every design decision encapsulates particular political, economic, and cultural values. This paper examines the extent to which the claim of neutrality constitutes an illusion, analyzing how the embedded choices in technology—from user interface design to algorithmic curatorship—reflect and reinforce underlying socio-political agendas.

A key dimension of this inquiry involves understanding how technological artifacts not only facilitate communication but also actively construct new cultural norms (Bojic, 2022). The emergence of the smartphone, epitomized by the groundbreaking introduction of the iPhone, illustrates a watershed moment where telecommunications, computing, and sophisticated software converged within a single device.

This convergence has redefined patterns of social behavior, cognitive processing, and even the very perception of reality. As Carr (2008) and Turkle (2011) have noted, the technological transformation initiated by the smartphone era has led to both an acceleration of information dissemination and a remapping of cognitive boundaries, ultimately blurring distinctions between work, leisure, and personal identity. The historical evolution from a tool for communication to a pervasive cultural force invites a reexamination of how digital innovations reshape societal expectations and day-to-day behavior.

The conceptual framework of Plato's allegory of the cave offers a metaphor for the phenomenon of algorithmic mediation in the digital age. Just as the prisoners in the cave mistake shadows for the fullness of reality, contemporary users are frequently exposed to a curated digital experience in which algorithms determine the visibility and salience of information. Modern recommendation systems, social media feeds, and personalized content curation serve not only to streamline user engagement but also to construct a mediated reality that may obscure diverse perspectives. As Zuboff (2019) and Noble (2018) have argued, the filtering inherent in these systems can subtly constrain public discourse by privileging content that aligns with preexisting biases and commercial imperatives, thereby reinforcing structural inequities and limiting the scope of social debate.

Concurrently, the evolution of artificial intelligence has introduced a spectrum of new paradigms that further complicate the relationship between technology and human agency. Advances in deep learning, generative models, and adaptive algorithms have enabled systems

capable of producing novel content and decision-making outcomes that blur the boundaries between human creativity and machine computation. The opaque nature of many AI systems, often characterized as “black boxes,” raises complex questions regarding transparency, accountability, and the ethical implications of delegating decision-making to automated agents (Bostrom, 2014). As these systems are increasingly integrated into critical areas such as healthcare, finance, and governance, an understanding of their underlying design principles becomes essential for evaluating their broader social and political impacts.

Another significant aspect of the contemporary digital landscape is the tension between convenience and autonomy (Bojic, 2024). Digital platforms are engineered to optimize user engagement through designs that encourage repetitive interaction and immediate gratification. Features such as push notifications, algorithmically generated news feeds, and immersive virtual environments are calibrated to capture attention and motivate habitual use. Such design strategies can have profound implications for both individual cognitive engagement and collective social dynamics. As suggested by Turkle (2011) and Habermas (1981), the prioritization of efficiency and convenience may come at the expense of deeper, reflective modes of thought and spontaneous interpersonal interaction. The interplay between mediated convenience and the gradual erosion of autonomy raises important questions about the cumulative effects of digital technology on human agency.

The following analysis begins with a detailed exploration of the illusion of technological neutrality, addressing the inherent biases embedded in technology from its conception to deployment. Subsequent sections

trace the cultural shift instigated by innovations such as the iPhone and extend the inquiry into the realm of algorithmic mediation, drawing on classical metaphors such as Plato's cave to underscore the filtered nature of digital realities. Further sections investigate the emergent arenas of artificial intelligence, considering both the potentials and limitations of recent advances, and assess the implications of digital convenience for individual and collective autonomy. Finally, the discussion synthesizes these perspectives within a framework that, rather than prescribing specific normative actions, seeks to advance a comprehensive understanding of the complex interplay between technology and society.

The Illusion of Technological Neutrality

The notion of technological neutrality holds that devices, algorithms, and digital infrastructures serve as impartial tools—mere extensions of human intent that operate without built-in values or biases. However, a closer examination reveals that the claim of neutrality is not only misleading, but it also obscures the myriad ways in which technology is imbued with the social, political, and economic agendas of its creators. As Winner (1986) famously contends, technological artifacts have politics; they are designed choices that often serve to reinforce existing power structures, rather than function as benign intermediaries between human intent and action.

At the heart of this discussion is the recognition that every design decision, from the layout of a smartphone interface to the parameters of an algorithm, embodies specific assumptions about how users should behave and

what is valued. For instance, the push notifications and engagement metrics built into modern smartphones are not neutral features; rather, they are deliberately engineered to capture attention, promote habitual use, and thereby generate data and profit. These design choices subtly steer users towards specific patterns of behavior, turning what might be thought of as innocuous technology into a mechanism for behavioral control (O’Neil, 2016). In this sense, neutrality becomes an illusion—a narrative that masks the intrinsic biases built into every phase of a technology’s lifecycle, from conceptualization to deployment.

This embedded subjectivity is perhaps most evident in the realm of algorithmic decision making. As digital platforms increasingly rely on sophisticated algorithms to curate information and mediate human interactions, the hidden values encoded within these systems become crucial determinants of what users see, read, and believe. Algorithms are frequently presented as objective arbiters, distilling vast amounts of data to deliver personalized content with surgical precision. However, as Zuboff (2019) argues, these algorithms are not neutral conduits of information; they are reflective of the economic imperatives and ideological biases of their developers. The so- called “neutral” recommendation systems, reminiscent of the shadows in Plato’s allegory of the cave, present a filtered version of reality—one that is designed to maintain user engagement while often marginalizing alternative perspectives.

The claim of neutrality frequently serves as a rhetorical tool used by technology companies to deflect criticism and regulatory scrutiny. By insisting that their innovations are mere tools, devoid of social or political

intent, corporations can obscure the fact that their products often contribute to systemic issues such as surveillance capitalism, economic inequality, and the erosion of democratic discourse. This deflection conceals the reality that design choices—for instance, in the prioritization of engagement over well-being—are inherently political acts. Heidegger (1954) similarly warned that modern technology does not merely offer a range of options but instead ‘enframes’ the world in a manner that limits how reality is disclosed to us. In effect, the neutrality of technology becomes a convenient fiction that prevents a deeper examination of its impact on human agency and social structure.

It is also essential to consider the cultural dimensions of this illusion. Digital technologies are often marketed as universally applicable tools that transcend national, cultural, or socio-economic boundaries. Yet, the values encoded within these systems are deeply contextual. The mechanisms that drive user engagement in Silicon Valley, for example, might not be appropriate—or even replicable—in societies with different communicative norms or political traditions. As Noble (2018) and other critical scholars have demonstrated, algorithmic biases can amplify existing societal inequities, reinforcing stereotypes and limiting access to alternative viewpoints. Thus, the presumed universality and neutrality of technology often serve to naturalize and legitimize disparities that are, in reality, products of specific cultural and historical conditions.

The persistence of the neutrality myth has significant implications for policy and governance. When technology is erroneously assumed to be value-neutral, it becomes more challenging to hold developers and corporations accountable for the social and political consequences of their products. This *de facto* abdication

of responsibility can lead to regulatory gaps where the adverse side effects of digital innovations—ranging from privacy intrusions to the manipulation of public opinion—are left unchecked. A critical reexamination of the neutrality claim is therefore indispensable for efforts aimed at establishing a more equitable and transparent technological landscape.

Technology as Cultural Transformation: From the iPhone to the Metaverse

Over the past two decades, technology has increasingly reshaped human civilization by reconfiguring the way people interact, work, and perceive reality. The technological trajectory that began with the introduction of the iPhone in 2007 illustrates a profound cultural transformation—a metamorphosis that now extends into immersive digital realms such as the metaverse. This section examines the evolution of technology from its role as a communication tool to becoming a transformative cultural force that redefines the very terms of human existence.

The iPhone, heralded as a watershed moment in mobile technology, did more than introduce a touch-screen interface; it catalyzed a rethinking of personal communication and media consumption. By integrating telecommunications, computing, and sophisticated software applications into a single handheld device, the iPhone reconfigured everyday life. As Carr (2008) points out, the Internet and connected devices have radically restructured cognitive processes and social behavior. With every tap and swipe, users began to engage with an ecosystem that not only disseminated information but also actively curated personal experience. This shift has had far-reaching consequences, altering the pace of daily

life and infusing it with a digital immediacy that blurs the distinction between work, leisure, and social interaction.

This transformation was swiftly followed by the rise of social media platforms that capitalized on the capabilities of the smartphone. The integration of recommender algorithms into these platforms further accelerated a cultural evolution reminiscent of Plato's allegory of the cave. Like the prisoners of the cave who mistake shadows for reality, individuals increasingly found themselves exposed to mediated representations that, while engaging, often truncated deeper engagement with the world (Plato, ca. 380 BC). Turkle (2011) has noted that while these digital interactions create an illusion of connection, they can also precipitate feelings of isolation and superficiality. In this digital ecosystem, individuals are not simply passive consumers of information but are being actively shaped by interfaces that determine which voices are heard and which perspectives are sidelined.

Beyond social media, the cultural implications of the smartphone era extend to the very fabric of everyday experience. The iPhone transcended mere communication by establishing a personal device that continuously records, mediates, and transforms everyday life. This reconfiguration of existence is evident in the widespread adoption of mobile banking, digital health monitoring, and location-based services, all of which intrude into aspects of personal autonomy that were once considered sacrosanct. The transformation is not solely technological—it signals a paradigmatic shift in which the boundaries between digital and analog, public and private, blur and overlap, forcing a redefinition of both self and community.

As these changes matured, the technological landscape evolved once more with the advent of immersive environments like the metaverse. While the smartphone brought a digital overlay to human experience, the metaverse promises a synthetically constructed world that mirrors and, in some aspects, competes with physical reality. In this new era, digital experiences are no longer confined to the two-dimensional screens of smartphones and computers but are expanding into fully immersive spaces that simulate sensory experiences with uncanny verisimilitude. This evolution raises critical questions about the nature of reality itself. As Zuboff (2019) cautions, the continuous aggregation of personal data—and its subsequent use in crafting hyper-personalized digital experiences—may further fragment public discourse and erode the authenticity of human experiences.

The metaverse represents the next phase in a long journey of cultural transformation driven by technology. By transcending traditional media formats, it introduces a world where virtual interactions may increasingly substitute for physical ones. Although this promises enhanced creative expression and new forms of social engagement, it also harbors significant risks. For instance, as people gravitate towards synthetic digital experiences, there is a possibility for a profound shift in how individuals value and experience reality—potentially leading to an existential disconnect. This is not merely a technical evolution but a deep cultural reorientation that challenges established norms around work, leisure, and interpersonal relationships.

The transition from the smartphone to the metaverse can be seen as part of a broader continuum wherein

technology progressively shapes the parameters of human thought and social organization. Castells (1996) observed that the networked society is characterized by a new form of organizational logic—a logic that rearranges power structures and alters the social contract that binds individuals to their communities. Today's immersive technologies further intensify this dynamic by creating environments that are not only interactive but also algorithmically curated, raising fresh concerns about autonomy, privacy, and democratic participation. In this digitally mediated reality, the decision-making process is increasingly outsourced to algorithms that determine which narratives are amplified and which are suppressed, potentially leading to a diminution of free will and self-determination.

Plato's Cave and the Age of Algorithmic Mediation

Plato's allegory of the cave, as described in his *Republic* (ca. 380 BC), serves as a time-honored metaphor for the limitations of human perception, where prisoners confined to a cave mistake mere shadows for reality. In the digital era, this allegory finds an unexpected and illuminating parallel in the mechanisms of algorithmic mediation. Just as the prisoners are deceived by superficial images, modern individuals are often presented with a curated, algorithmically generated version of reality that masks the richness and complexity of the world beyond the screen.

At the heart of this comparison is the way in which technological systems—specifically recommender algorithms, social media feeds, and personalized content curation—act like the projections on the wall of the cave. These systems filter and prioritize information based on opaque criteria, often designed to maximize user

engagement, profit, or political influence, rather than to provide a holistic picture of reality (Zuboff, 2019). The content displayed becomes a series of “shadows” that are accepted as truth. Over time, users may come to regard this algorithmically mediated experience as equivalent to, or even preferable to, the unmediated, lived world.

The metaphor deepens when we recognize that these digital shadows are not accidental but are the result of deliberate design choices. Every algorithm carries the imprint of its programmers’ assumptions, economic imperatives, and ideological biases. As Noble (2018) has argued, the idea that technology is inherently neutral obscures the structured and value-laden nature of algorithmic curation. In practice, when users consume news feeds, social media posts, or even entertainment recommendations, they are engaging with a filtered reality that reinforces existing preferences and, often inadvertently, societal divisions. This echo chamber effect bears a striking resemblance to the prisoners in Plato’s cave, who are confined to a narrow spectrum of perceptions, never exposed to the broader truths of the world.

Algorithmic mediation also poses a significant challenge to the process of enlightenment that Plato described. In the cave allegory, the journey out of the cave symbolizes the arduous climb toward knowledge and understanding—a process of confronting uncomfortable truths and questioning established perceptions. Parallel to this, the current digital landscape often encourages a passive acceptance of the provided content. The ease of access to algorithmically filtered information can discourage users from seeking alternative viewpoints or engaging in critical reflection.

As Carr (2008) has observed, this pattern of consumption can lead to shallow cognitive processing, where the depth of understanding is sacrificed in favor of the immediacy and convenience of mediated experiences.

The consequences of this algorithmic mediation extend beyond individual cognition to the level of societal discourse. When a limited set of algorithmic criteria dictates which voices are amplified and which are marginalized, public debates risk becoming homogenized, with dissenting opinions or complex narratives being systematically sidelined. This dynamic not only constrains the marketplace of ideas but also compromises the democratic process by limiting the exposure of citizens to the full spectrum of public opinion and debate. In effect, the digital cave constructed by algorithmic systems can lead to an erosion of agency, as individuals become more susceptible to manipulation through targeted content (O'Neil, 2016).

Another dimension of this metaphor is the seductive allure of digital media that can lead to a form of self-imposed captivity. Just as prisoners in the cave might grow accustomed to their confined reality, contemporary users often develop a preference for the tailored, dopamine-inducing experiences provided by social media and immersive virtual environments. The metaverse and other advanced digital domains intensify this effect by replicating the stimuli of the physical world so compellingly that they become irresistibly attractive substitutes. In this scenario, the boundary between mediated experience and genuine interaction blurs, raising concerns about the long-term impact on personal autonomy and the quality of human relationships.

To counteract the effects of this modern-day cave, it is essential to cultivate a form of digital literacy that goes beyond mere technical competence. The challenge is to empower users to recognize and critically assess the ways in which algorithms shape their experience of reality. This involves not only understanding the mechanics behind these systems but also reflecting on the values they propagate and the potential biases they reinforce. Educative initiatives that emphasize “digital hygiene” can help individuals develop the cognitive tools required to step outside the algorithmic shadow play and engage with a broader, more authentic spectrum of experiences.

Emergence of New AI Paradigms and Their Implications

In the past decade, the field of artificial intelligence has undergone seismic shifts that have transformed it from a collection of specialized, narrowly defined applications into a realm of versatile, generative, and adaptive systems. This transformation has been catalyzed by breakthroughs in deep learning, reinforcement learning, and natural language processing, culminating in the emergence of new AI paradigms such as generative models (e.g., ChatGPT), computer vision systems capable of creating art from scratch, and autonomous agents that learn complex behaviors without explicit programming. These advances have not only broadened the functional scope of AI but also reconfigured the very expectations we hold about machine intelligence and its role in society (Bostrom, 2014).

One of the most visible manifestations of these new paradigms is the rise of generative AI systems. Unlike earlier rule-based or narrowly supervised approaches, generative models are designed to create novel content—

from coherent text and realistic images to dynamic decision- making processes. ChatGPT, for instance, leverages vast linguistic datasets to produce human- like responses and engage in contextually rich conversations. This capacity to generate and synthesize information has redefined the boundaries between machine assistance and human creativity. However, this very power comes with significant limitations and risks. The opacity inherent in deep neural networks, often described as a “black box” phenomenon, means that the decision-making process of these systems is not easily interpretable. This lack of transparency raises profound questions about accountability in cases where AI-driven decisions impact critical sectors such as healthcare, finance, or the judiciary (O’Neil, 2016).

Another salient feature of modern AI paradigms is their ability to adapt and generalize across multiple domains. Previously, AI systems were largely designed to perform discrete tasks using limited datasets and fixed models. Today, deep learning architectures can process enormous quantities of data and learn intricate patterns that allow them to function across a variety of contexts. This transferability has enabled innovations in autonomous vehicles, robotics, and even in areas like climate modeling and personalized medicine. With each successive iteration, AI models are becoming more capable of tackling problems that require understanding of context and a flexibility that approximates general intelligence, albeit still bounded by the limitations of their training data and design parameters (Zuboff, 2019).

The implications of these developments extend well beyond technical performance and technological capability. On a socio-economic level, the integration of

advanced AI systems into industry and public administration is reshaping labor markets and challenging traditional economic models. Automation powered by generative AI and related technologies threatens to displace not only routine manual labor but also cognitive jobs that were once considered secure by virtue of their complexity. While advocates point to the potential for AI to free human workers from mundane tasks—thus enabling more creative or strategic roles—the reality is that the displacement could exacerbate existing social inequalities if the benefits of technological advancements remain concentrated among a select group of corporations and capital owners. The looming question is how to distribute the rewards of these innovations equitably, perhaps by considering policy instruments such as universal basic income or extensive job retraining programs (Carr, 2008).

From a political and democratic standpoint, the centralization of AI research and development in the hands of a few industry giants poses risks for political autonomy and public discourse. When a handful of corporations control the data, platforms, and algorithmic processes that shape public information flows, the democratic ideal of open, unmediated debate is put at risk. AI systems can be, and in many instances are, used to micro-target political messages, entrench confirmatory biases, and manipulate public sentiment through algorithmically driven news feeds and social media platforms. This concentration of power, coupled with the opacity of the AI models in use, challenges regulatory frameworks that are based on the assumption of informational transparency and accountable governance. The resulting imbalance mirrors broader concerns about surveillance capitalism, where economic and social

power become increasingly intertwined with technological control (Zuboff, 2019).

The rapid evolution of AI paradigms introduces critical ethical dilemmas regarding the nature of cognition and the concept of creativity. As machines begin to exhibit behaviors that once were considered distinctly human—such as the composition of music, the creation of visual art, or the generation of philosophical treatises—the very definitions of these concepts come under scrutiny. Philosophers and ethicists are now forced to reconsider the boundaries between human and machine artistry, calling into question notions of originality, authenticity, and intellectual property. This transition challenges traditional humanistic perspectives on creativity, prompting debates about whether algorithmically generated content can, or should, be regarded as having the same cultural and intellectual value as that produced by human hands (Noble, 2018).

At the level of individual cognition, the widespread deployment of generative AI and responsive digital agents raises concerns about human agency and decision-making. As people increasingly rely on AI to guide choices—whether in professional, personal, or consumer contexts—the risk emerges that such systems might subtly manipulate perceptions and behaviors, nudging individuals toward outcomes that serve the interests of the technology providers rather than those of the users. There is an emerging risk that human decision-making will be gradually outsourced to or influenced by opaque algorithmic processes, leading to an erosion of critical thinking skills and personal autonomy. This dynamic is reminiscent of the broader concerns about the “digital anesthesia” of the human mind, where convenience and

efficiency are prioritized at the expense of reflective, deliberate action (Turkle, 2011).

The potential emergence of superintelligent systems represents another frontier that amplifies the stakes of these new AI paradigms. While true artificial general intelligence remains a subject of theoretical debate, the prospect of AI systems that surpass human cognitive abilities in a wide range of domains is increasingly being taken seriously by both technologists and philosophers. The implications of such an eventuality are profound: a superintelligent AI could, theoretically, redesign its own cognition, propagate self-improvement, and perhaps even reshape human society in ways that are currently impossible to predict. The dual-use nature of such technology demands rigorous ethical scrutiny and robust regulatory oversight to ensure that any transition toward higher-order AI is managed in a manner that safeguards human values and global stability (Bostrom, 2014).

Autonomy under AI Threat

In an era dominated by rapid technological innovation, the very conveniences that these advances promise are increasingly recognized as double-edged. On one side, digital technologies—from smartphones to immersive virtual environments—offer unparalleled ease and efficiency, liberating individuals from routine tasks and providing instantaneous access to information and social interaction. On the other, these same technologies encroach upon the most private spaces of the human mind and soul. The paradox is stark: the quest for convenience can quietly corrode genuine autonomy by substituting authentic, self-determined engagement with automated, algorithmically curated experiences. As we

delve into this issue, it becomes clear that what is often hailed as progress may, in fact, be a subtle mechanism of coercion, encouraging individuals to exchange their capacity for reflective, independent thought for the seductive immediacy of digital gratification.

Defining autonomy, in philosophical terms, entails an appreciation for the ability of individuals to make decisions based on their own rational deliberation and values. Autonomy is not merely the absence of external compulsion; it is the presence of agency—an ongoing process of self-governance and critical reflection. Yet, in the modern digital economy, the persistent lure of convenience exerts a powerful influence over everyday behavior. Technological systems are meticulously engineered to capture attention, employing design features that encourage habitual engagement, reward fleeting interactions, and ultimately condition users to prefer mediated experiences over direct, unfiltered reality (Turkle, 2011). When convenience becomes the primary criterion for human activity, the space for critical reflection shrinks, leaving individuals increasingly vulnerable to subtle forms of behavioral control and manipulation.

The architecture of convenience is built on interfaces and algorithms that are intentionally optimized for engagement. Consider the ubiquitous smartphone: its design prioritizes ease of use, rapid accessibility, and an endless stream of notifications and alerts. These elements are not accidental; they are derived from rigorous data analytics designed to maximize user retention. The promise of an always-on, always-available digital assistant comes with a trade-off: as users become habituated to short bursts of dopamine-driven interaction,

the foundations of deep, sustained focus and meditative thought begin to erode (Carr, 2008). What emerges is a form of “digital anesthesia,” where the constant pursuit of convenience numbs the faculties necessary for authentic autonomy and self-determination.

Such technological convenience invariably paves the way for a broader phenomenon we might call “corruption”—a gradual, insidious shift in human values and behaviors. By corruption, we refer not solely to moral decay in the traditional sense but to the erosion of robust, critically examined engagements with the world. As individuals grow accustomed to easy substitutes for genuine experience, there is a growing tendency to accept lower-quality, less fulfilling alternatives. In our digital age, these substitutes manifest in various forms: algorithmically generated news feeds, virtual social interactions that lack of the face-to-face conversation, and even entire digital realms such as the metaverse that offer sanitized reflections of reality. Over time, the cumulative effect is a society that privileges simulated convenience over the inherent messiness—and potential richness—of lived experience (O’Neil, 2016).

The allure of convenience, then, is in its capacity to offer a quick fix—a way to bypass the laborious process of independent thought and engagement. At first glance, this might seem beneficial: after all, freeing up time from mundane tasks should enable individuals to redirect their energies toward creative or intellectual pursuits. However, the reality is far more complex. When convenience becomes the default mode of interaction, the very skills that underpin autonomy—such as critical thinking, resilience in the face of complexity, and the capacity for sustained concentration—are gradually

atrophied. This degradation is not merely an abstract loss; it has tangible implications for societal well-being. For instance, when citizens are conditioned to favor readily available mediated content over diverse and challenging perspectives, the democratic process itself may be undermined as public debate becomes increasingly shallow and polarized (Habermas, 1981).

This dual threat of convenience and corruption is vividly illustrated by the evolution of digital media, particularly social networks and immersive virtual environments. The transformation from basic online interaction to the rich, yet synthetic, experiences of the metaverse represents a culminating example of how technological mediation can reshape human engagement. Virtual reality platforms offer a near-perfect simulation of physical spaces and experiences, enticing users with the promise of more controlled, curated worlds that seem to solve the inconveniences of real-life unpredictability. However, what appears as technological mastery over one's environment can, paradoxically, translate into a surrender of control. When individuals begin to prefer these mediated experiences, they risk becoming detached not only from the original sources of satisfaction but also from the organic, spontaneous interactions that constitute the essence of human connection. In many ways, the metaverse can be seen as the ultimate epitome of convenience—an arena where every interaction is sanitized, predictable, and optimized for immediate gratification, at the expense of authentic spontaneity and personal growth.

This shift toward mediated convenience, while offering immediate benefits, portends deeper, long-term consequences for both individual and collective autonomy. The case is perhaps most starkly articulated in the phenomenon of what might be termed “digital

complacency”—a state in which individuals gradually relinquish their critical faculties and become increasingly passive receivers of pre-digested information. This passivity is compounded by the fact that sophisticated algorithms, which drive such passive interaction, are rarely subject to thorough scrutiny. Their criteria for curating content often remain opaque, and their influence on public opinion and individual decision-making processes goes largely unchecked. Such conditions foster an environment where manipulation becomes not only possible but probable; users unwittingly trade their ability to independently evaluate the world for an over-reliance on algorithmically generated proxies of reality (Zuboff, 2019).

The implications of this dynamic are profound. In the political realm, for instance, the prioritization of convenience in information consumption can lead to the rapid dissemination of biased or unchallenged narratives. This is not accidental: as algorithms learn and adapt to user preferences, they often reinforce existing beliefs and, in doing so, create echo chambers that stifle critical discourse and debate. The resulting environment is one in which the democratic process is compromised, as informed, reflective decision-making is supplanted by the comfort of convenient and familiar content. In such a scenario, the autonomy of the individual is not simply a matter of personal freedom but a critical component of the health of the entire democratic system.

The erosion of autonomy through the paradox of convenience is not limited to the realm of cognitive or political life—it also penetrates the personal and emotional domains. The routine consumption of digital content, especially that which is designed to be addictive,

gradually reshapes neural pathways, reinforcing habits that favor short-term rewards over long-term satisfaction. In vulnerable individuals, this process can culminate in a form of digital dependency that mirrors substance addiction, wherein the individual's ability to derive pleasure from natural, unmediated experiences becomes severely compromised. The result is a society in which individuals might physically have more free time due to automation and digital efficiencies, yet find themselves trapped in a cycle of self-imposed distraction and shallow engagement, reminiscent of a "Homer Simpson-like" state of existence—an existence characterized by a lack of purpose, diminished creativity, and a pervasive sense of detachment from meaningful reality (Turkle, 2011).

Toward a New Social Contract in the Digital Age: Conclusion

The analysis presented throughout this article has highlighted a series of profound challenges arising from the rapid evolution of digital technologies—from the false promise of technological neutrality and the cultural transformation initiated by devices such as the iPhone, to the pervasive influence of algorithmic mediation, the emergence of new AI paradigms, and the dual threats posed by convenience and corruption. In response, this concluding section articulates a comprehensive and scientifically rigorous blueprint for a new social contract that aims to rebalance the relationship between technology and human society in ways that preserve, enhance, and ultimately return human autonomy and democratic participation to their central places in the digital age.

A first and fundamental tenet of the proposed social contract is the explicit recognition that technology is not an impartial tool but a mediator that encapsulates particular values, biases, and interests along its entire development trajectory. As Winner (1986) and Heidegger (1954) have argued, every technological artifact is imbued with political dimensions that affect social structures and human behavior. Consequently, public policy and regulatory frameworks must be reoriented to demand transparency and accountability at all stages—ranging from conceptual design to real-world deployment. This entails, for instance, the implementation of rigorous auditing measures for algorithmic systems designed to ensure that the criteria underpinning decision-making processes are comprehensible, contestable, and aligned with broader ethical norms (O’Neil, 2016; Zuboff, 2019). Establishing interdisciplinary oversight committees comprising experts in computer science, ethics, social sciences, and law is one practical step toward democratizing technological governance and mitigating the risk of algorithmically reinforced inequities.

Central to protecting autonomy in a technology-saturated environment is the cultivation of digital literacy programs that go beyond mere technical fluency. The new social contract must emphasize “critical digital hygiene”—a holistic educational strategy that equips citizens with the cognitive and epistemological tools necessary to interrogate the mediating processes of digital interfaces. Academic research (Carr, 2008; Noble, 2018) has demonstrated that exposure to algorithmically curated digital content may induce a state of passivity and superficial engagement. Therefore, educational initiatives at all levels—from primary education to

lifelong learning programs—should focus on teaching individuals how to recognize, question, and engage critically with digital content. Curricula should be designed to enhance reflective thinking, foster media skepticism, and illuminate the socio-political forces underlying digital information flows. Empowered citizens are better positioned to resist manipulation and to reclaim the capacity for independent judgment, mitigating the influence of algorithm-driven echo chambers that threaten democratic discourse (Turkle, 2011; Habermas, 1981).

Technological advancements, particularly in artificial intelligence and automation, have disrupted traditional labor markets and reconfigured the relationship between work, human dignity, and economic participation. The evolving nature of work—characterized by widespread displacement and the shifting boundaries between tasks that are human-centric and those that can be automated—demands bold reforms in economic policy. Drawing upon the insights of van Parijs (1987), the social contract must integrate economic mechanisms such as universal basic income (UBI) and a restructuring of work hours. These measures have the dual purpose of cushioning the disruptive impacts of technological unemployment and liberating individuals from the confines of repetitive, unfulfilling tasks. A reimagined economic framework would recognize that technological progress is not an end in itself but a means to enhance human potential and quality of life. It would, therefore, prioritize policies aimed not solely at maximizing productivity but at ensuring that the benefits of innovation are distributed equitably, thereby preventing further socioeconomic stratification and augmenting collective agency.

A recurring concern in an age of algorithmic mediation is the erosion of public discourse due to the

narrowing effects of digital echo chambers. The scientific literature has documented how algorithmically mediated information flows diminish exposure to diverse perspectives, undermining the public debate essential for a functioning democracy (Zuboff, 2019; Turkle, 2011). As part of the new social contract, regulatory bodies and independent watchdogs should be mandated to oversee digital communication platforms, ensuring that these systems promote pluralism, mitigate polarization, and uphold the principles of democratic deliberation. Mechanisms for algorithmic transparency and accountability must be established so that citizens are informed about the processes behind information curation and can contest biased or manipulative practices. Additionally, support for independent journalism and community-based discourse initiatives should be increased to cultivate spaces where unmediated, in-person dialogue can flourish in tandem with digital engagement. This dual approach is vital in preserving the integrity of public opinion and ensuring that democratic decision-making is responsive to a broad spectrum of voices.

The global nature of digital networks and the transnational character of modern AI systems call for coordinated international regulation. In an interconnected world where digital infrastructures transcend national boundaries, isolated domestic efforts are insufficient to address the profound risks associated with AI, surveillance capitalism, and digital authoritarianism. The new social contract must thus be underpinned by a robust, internationally negotiated framework that sets common standards for data protection, algorithmic accountability, and the ethical deployment of AI technologies. Drawing on the work of scholars such as Bostrom (2014), policy-makers should engage in multilateral dialogues aimed at establishing

binding norms and treaties that regulate the development and use of AI. This approach would enable a harmonization of policies to prevent a regulatory race to the bottom and to ensure that technological innovations contribute to global stability and human flourishing.

While technical, economic, and political dimensions are critical, a comprehensive social contract must also address the subtler, yet equally vital, realm of human subjectivity and emotional well-being. The proliferation of digital mediations has not only impacted cognitive autonomy but has also contributed to a reorientation of personal identity and intersubjective relations. Empirical studies have associated heavy reliance on superficial digital interactions with diminished capacity for empathy, reduced creative engagement, and altered self-perception (Turkle, 2011). The new social contract should promote cultural policies that foster creative, reflective, and communal practices. Initiatives that encourage artistic expression, mindfulness, and meaningful real-world interactions play a crucial role in recalibrating the human experience. Public funding for cultural projects, community centers, and programs encouraging digital detoxes could serve as counterbalances to the pervasive allure of digital convenience, enabling individuals to reclaim a fuller, more authentic sense of self and societal connection.

The challenges highlighted throughout this article are inherently complex and multifaceted—they span the domains of technology, economics, politics, and culture. Addressing these issues requires an integrative policy approach and a sustained interdisciplinary dialogue that brings together experts from diverse fields including computer science, ethics, sociology, political theory, and economics. The new social contract, therefore, must be

dynamic and adaptive, reflecting not only current technological trends but also evolving understandings of human well-being and societal organization. Forum structures that facilitate ongoing conversation between policymakers, scholars, industry leaders, and the public will be essential. Such platforms should aim to continually reassess the ethical implications of new technologies and adjust regulatory frameworks accordingly. By institutionalizing these dialogues, society can ensure that technological progress remains in service of human values and that emerging digital paradigms are harnessed to promote collective flourishing.

The formulation of a new social contract for the digital age emerges as both an urgent ethical imperative and a pragmatic necessity. This contract calls for a radical rethinking of how technology is integrated into all spheres of human life—from ensuring that technological devices and AI systems operate with transparency and fairness, to overhauling educational and economic systems that currently inadequately prepare citizens for a digitally mediated future. It demands the establishment of international regulatory mechanisms to oversee AI progression and the fostering of a human-centric digital culture that revives critical thinking, empathy, and authentic engagement. As we transition toward an era defined by the convergence of ubiquitous digital media, advanced AI, and immersive virtual environments, safeguarding and enhancing human autonomy must become the cornerstone of our collective endeavor.

The envisioned social contract is not merely a defensive bulwark against the encroachments of technological determinism; rather, it is an aspirational

framework that repositions technology as a tool for human empowerment. It posits that true progress is measured not by economic efficiency or computational prowess alone but by the degree to which technological innovation enhances human creativity, deliberation, dignity, and participatory democracy. This comprehensive proposal serves as a roadmap for policymakers, scholars, and citizens alike—a call to reimagine and reconstruct our social arrangements in ways that embrace technology’s promise while steadfastly preserving the fundamental qualities that define our humanity.

References

1. Bojic, Ljubisa (2024). AI alignment: Assessing the global impact of recommender systems. *Futures*, 160, 103383.
<https://doi.org/10.1016/j.futures.2024.103383>
2. Bojic, Ljubisa (2022). Metaverse through the prism of power and addiction: What will happen when the virtual world becomes more attractive than reality? *European Journal of Futures Research*, 10(1), 22.
<https://doi.org/10.1186/s40309-022-00208-4>
3. Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press).
4. Carr, Nicholas (2008). *The Shallows: What the Internet Is Doing to Our Brains* (New York: W. W. Norton & Company).
5. Castells, Manuel (1996). *The Rise of the Network Society* (Oxford: Blackwell Publishers).
6. Habermas, Jürgen (1981). *The Theory of Communicative Action* (Boston: Beacon Press).

7. Heidegger, Martin (1954). *The Question Concerning Technology and Other Essays* (New York: Harper & Row).
8. Noble, Safiya Umoja (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press).
9. O'Neil, Cathy (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown).
10. Plato (2002). *Republic* (For the Allegory of the Cave, see Book VII) (New York: Dover Publications).
11. Turkle, Sherry (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other* (New York: Basic Books).
12. van Parijs, Philippe (1987). *Real Freedom for All: What (If Anything) Can Justify Capitalism?* (Oxford: Oxford University Press).
13. Winner, Langdon (1986). *The Whale and the Reactor: A Search for Limits in an Age of High Technology* (Chicago: University of Chicago Press).
14. Zuboff, Shoshana (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs).