Ignas Kalpokas is an Associate Professor, head of MA Future Media and Journalism at the Department of Public Communication and Senior Researcher at V. Kavolis Institute of Transdisciplinary Research, Vytautas Magnus University. He also holds a visiting research position at SGH Warsaw School of Economics. His primary research interests include the societal and political impact of AI, political communication, fake news and disinformation, and media theory. His current research focuses on the adoption of generative AI and societal-level coping strategies within the context of accelerated technological change. He has authored or co-authored eight books and nearly 40 articles and book chapters on the latter topics.
Contact: ignas.kalpokas@vdu.lt

Viktoriia Oksymets is a PhD candidate at the Department of Public Communication, Vytautas Magnus University. Her research focuses on the adoption of generative AI in communication and journalism as well as AI-generated disinformation and regulatory responses to the latter.
Contact: viktoriia.oksymets@vdu.lt

Soft Power
Revista euro-americana de teoría
e historia de la política y del derecho

Vol. 11(2). Julio-Diciembre 2024
ISSN (online): 2539/2239 ISSN (print): 2389-8232
https://doi.org/10.14718/SoftPower.2024.11.2.9

JOYSTICKS IN WHOSE HANDS? HACKING THE SELF AS A CRIME AND AS CORPORATE PRACTICE*

Ignas Kalpokas
Vytautas Magnus University
SGH Warsaw School of Economics

Viktoriia Oksymets
Vytautas Magnus University

¿EL JOYSTICK EN MANOS DE QUIÉN? EL PIRATEO DEL YO COMO DELITO Y COMO PRÁCTICA EMPRESARIAL

Abstract

Despite the pervasive integration of digital technologies into all aspects of our lives, the emergence of novel technologies, such as virtual reality (VR), generative artificial intelligence (AI), and brain-computer interfaces, has led to a significant expansion in the scope and depth of technological influence, such as the generation of experience, knowledge (or a pretence thereof) and communicative expressions, as well as the removal of the very separation between the biological and the digital/machinic. Under such circumstances, the idea of an independent self (if one ever existed) becomes structurally impossible. Instead, the self becomes enmeshed with digital technologies to the point of co-constitution. Here, the self becomes susceptible to hacking in two important ways. One is through the more traditional cyber-criminal activities, such as human joystick attacks pertaining to VR environments

* Reception date: 20th April 2024; acceptance date: 24th April 2024. The essay is the issue of a research project carried out within the Viešosios komunikacijos katedra / Department of Public Communications, Vytautas Magnus University.

and, potentially, brain-computer interfaces. However, it may be argued on a more fundamental level that it is possible to achieve the same results without hacking. In fact, this is exactly the way in which corporate algorithmic governance of the digitally enmeshed self already takes place. This process involves the structuration of digital spaces and soon, the establishment of corporate control of brain-computer interfaces as well as training and ownership of AI models, all of which help structure the individual self with corporate, rather than personal, interests in mind. In combination, these two modes of determination are seen as alternative but structurally similar ways of hacking the self and, effectively, turning individuals into human joysticks controlled by an intangible – but no less potent – hand.

Keywords
platformisation; generative AI; Virtual Reality; hacking; human autonomy

Resumen
Si bien las tecnologías digitales han impregnado cada vez más todos los aspectos de nuestras vidas, la llegada de nuevas tecnologías, como la Realidad Virtual (RV), la IA generativa y las interfaces cerebro-ordenador, extiende la influencia de las tecnologías a una profundidad sin precedentes, dando paso a la generación de experiencia, conocimiento (o una pretensión del mismo) y expresiones comunicativas, así como a la eliminación de la propia separación entre lo biológico y lo digital/mecánico. En tales circunstancias, la idea de un yo independiente (si es que alguna vez existió) se hace estructuralmente imposible. En su lugar, el yo se entrelaza con las tecnologías digitales hasta el punto de la co-constitución. En este caso, el yo es susceptible de ser pirateado de dos formas importantes. Una es a través de las actividades cibercriminales más tradicionales. En particular, los ataques con joysticks humanos relacionados con entornos de RV y, potencialmente, con interfaces cerebro-ordenador. Sin embargo, podría sostenerse en un nivel más fundamental que es posible lograr los mismos resultados sin piratear. De hecho, esta es exactamente la forma en que ya tiene lugar la gobernanza algorítmica corporativa del yo digitalmente enredado. En concreto, esto implica la estructuración de espacios digitales y, de cara al futuro, el control corporativo de interfaces cerebro-ordenador, así como la formación y propiedad de modelos de IA, todo lo cual ayuda a estructurar el yo individual teniendo en cuenta los intereses corporativos, más que los personales. En combinación, estos dos modos de determinación se consideran formas alternativas, pero estructuralmente similares de piratear el yo y, de hecho, convertir a los individuos en joysticks humanos controlados por una mano intangible, pero no por ello menos potente.

Palabras clave
plataformización; IA generativa; Realidad Virtual; piratería informática; autonomía humana

Introduction

The rapid development of digital technologies in the 21st century has dramatically reshaped the boundaries of human experience and capability. This includes Virtual Reality (VR), generative Artificial Intelligence, and Brain-Computer Interfaces (BCIs) which, while enabling new capacities and experiences, give rise to novel security threats in an already complex environment. However, these threats and challenges may not necessarily come from nefarious actors – instead, many of them come about as part and parcel of the routinary application of such technologies, particularly when combined with the broader trend of platformisation that characterises contemporary societies. This article deals with the threats and challenges to human autonomy and independent decision-making capacity.

The following systems are good examples of such threats and challenges: VR, on one hand, immerses individuals in lifelike digital environments, often blurring the lines between physical reality and virtual experience. This immersion can lead to altered perceptions and behaviours, raising critical questions about the autonomy of users within these environments, regardless of whether this takes place as a matter of intentional deception or merely as over-identification with regular experiences. Generative AI, on the other hand, can influence and sometimes manipulate human decision-making and creativity by producing highly convincing text, images, and sounds. These systems, trained on vast datasets, can subtly shape opinions and choices, potentially undermining individual agency. Meanwhile, BCIs, which facilitate direct communication between the brain and external devices, promise to revolutionise fields such as medicine, communication, and entertainment. Yet, the integration of BCIs into daily life presents significant ethical concerns regarding privacy, mental autonomy, and the potential for external control over human thought and action.

Consequently, this article explores the multifaceted implications of these technologies on human autonomy and agency. Through a critical analysis of current literature and emerging trends, this study seeks to illuminate the potential risks in an increasingly technologically mediated world, focusing primarily on the ways in which, borrowing from an emergent cybersecurity threat discussed below, we now become human joysticks, i.e. entities whose behaviours (including, in many cases, the internal motivations for such behaviours) are controlled by an intangible hand.

Immersive Technologies and Hacking the Self

As societies become ever more dependent on their technological infrastructures, individual and collective vulnerability increases accordingly, with threat actors being increasingly

capable of not only undermining everyday life but also harming individuals directly. For example, Mazarr et al. (2019) have coined the term 'virtual societal warfare' to describe the broad range of techniques that adversaries can use for such purposes (p. xiii). It includes (a) disinformation, including the generation of very large amounts of manipulated or entirely fabricated digital content and distributing it through easily accessible channels, such as social media, while simultaneously undermining trust in conventional truth arbiters; (b) corruption or other forms of manipulation of the databases on which not only the economy but also decision-making processes, particularly those pertaining to algorithmic governance and AI tools, depend; (c) disruption and hacking the everyday digital ecosystems, such as the Internet-of-Things; and (d) hijacking of VR systems and AI-powered chatbots and assistants to spread panic, discomfort, or even to cause direct harm. In these ways, confusion, anxiety, and distrust can be induced, undermining the functioning of a society and its government. While it is beyond the scope of this article to cover all the abovementioned threats, those that cause a direct threat to individual autonomy and personal capacity, either through virtual experiences or by way of physically invasive technologies (such as BCIs), deserve particular attention because they operate directly against the dominant empowerment-focused discourse on new digital technologies.

To begin with, VR, integrating data from multiple sources and a cornucopia of connected sensor-enabled devices in a user's vicinity, can offer a personalised escape from the confines of the physical world and immersion in a fantasy that one has always longed for (Mazarr et al., 2019, p. 2; Kalpokas, 2024a). In this way, individuals can be seen as *empowered* to have experiences of their liking regardless of their offline conditions and other concerns (Flavián, et al., 2019) – that is, as long as those offline conditions provide sufficient connectivity and wealth to enable seamless immersion in VR. Provided that 'the best experiences imaginable can be had at the press of a button' (Bailenson, 2018, p. 250), emotional and time investment in virtual experiences is likely not only to grow but also raise the question of which 'reality' is more 'real'. Given the preceding, it is not surprising that some would raise questions as to whether VR should be seen as merely a maladaptive escapist reaction to the contemporary lifestyle (see e.g. Han et al., 2022, p. 13) whereby the ability to immerse oneself into a simulated world is seen as simply an attempt to hide away from the pressures of offline world. Indeed, for better or worse, the feeling of presence is crucial to the successful functioning of VR experiences because it 'happens when your brain is so fooled by a virtual experience that it triggers your body to respond as though the experience were real' (Rubin, 2020, p. 4). Hence, the very purpose of virtual experiences is to erase the boundary between the physical and the virtual self, thereby creating a hybrid entity (Scholz & Duffy, 2018).

Successful VR experiences depend particularly on inducing a state of flow resulting from a combination of seamless rendering of the virtual world and a user's perception of personal efficacy in it (see e.g. Huang et al., 2023). As such, VR provides immersion in what is taking shape in the perceptive surroundings of the individual or, rather, of their avatar – provided that such a distinction is still viable (Chalmers, 2022, p. xii). Hence, VR is truly about being in the experience and assuming the identity and the locatedness of whoever's role the user is playing in the story (Rubin, 2020). However, aside from immersion and locatedness VR can also be used as an enabler of security threats, particularly if the virtual surroundings induce certain movements in the physical space that could end up causing harm (see e.g. Vondráček et al. 2023). It is crucial to note that when immersed in a VR experience, the user is completely blocked off from the surrounding physical reality and is, therefore, dependent on location sensors and geofencing to remain in a safe area; meanwhile, if the user is directed beyond such boundaries and into a dangerous space, that might be outrightly harmful (Chow et al., 2023). Such attacks, commonly known as Human Joystick Attacks, involve an attacker hacking or otherwise getting access into the location rendering function of VR experiences so that users, when moving in order to induce corresponding movements in the virtual world, walk into objects or even off surfaces or move towards a predefined location for some other purpose, such as to put the user into a compromising situation (see e.g. Odeleye et al., 2023; Cayir et al., 2024). While these may currently seem to be rather niche threats, it can be reasonably expected that the growing popularity of virtual experiences would attract a corresponding increase in the number of threat actors intending to capitalise on the specificities of VR (Cayir et al., 2024).

Notably, for a successful and seamless rendering of virtual experiences, one needs to not only minimise latency but also reduce the external obstruction of headsets, haptic devices (i.e. devices that convey bodily sensations) etc. This enables new potential for BCIs, namely, implantable devices that can both read and induce brain signals in order to interact with external digital objects and content (see e.g. López Bernal et al. 2020; Ajrawi et al., 2021; Liv, 2021). As described by Maiseli et al. (2023), "[b]ypassing the conventional communication channels for different tasks (e.g., vision, movement, and speech), BCI links the brain's electrical activity and the external world to augment human capabilities in interacting with the physical environment" (p. 1). The devices are mostly framed for use in healthcare, claiming to give individuals with disabilities the capacity to manipulate physical objects, communicate, or have sensations of their environment by way of digital brain implants (López Bernal et al., 2020). However, these have the potential to induce non-existent sensations, including virtual experiences. In

addition, just like all digital interfaces, BCIs can be hacked into, raising at least two significant concerns: obtaining private information directly from the brain without the user's consent and altering behaviour by directly manipulating neural activity (López Bernal et al., 2020; Brocal, 2023). This calls for a new focus on security – namely, 'neurosecurity' – understood as 'the protection of the confidentiality, integrity, and availability of neural devices from malicious parties' (Liv, 2021, p. 339). Similarly, cyberbiosecurity has been proposed as a term to denote potential abuse and detrimental use of personal data accessed through BCIs as well as malicious signals submitted to the brain via BCI (Greenbaum, 2021).

Indeed, it is important to stress that 'misusing neural devices for malicious purposes may not only threaten users' physical security, but it can influence their behaviour and alter their sense of identity and personhood' (Liv, 2021, p. 339; see also Maiseli et al., 2023). Evidently, the dangers are significant: as stressed by Greenbaum (2021): "[e]ven the ultimate safety of the user and those around her' can be threatened, because by compromising the interface, 'a hacker could take control of the device, even committing a crime. It would be difficult to prove that the crime was committed by a hack and not by the owner of the prosthetic" (p. 665). In fact, the capacity of BCIs to read signals associated with human thoughts, and the increasingly two-way nature of such devices implies that not only reflexive behaviours and movements but also thought processes can be affected (and, indeed, remotely controlled), including for nefarious purposes (King et al., 2024), meaning that not only external behaviours but also internal motivations can be inauthentic. It must be kept in mind that such implantable devices will have to remain connected 24/7, not only to send and receive data pertaining to the functionality of the device in the narrow sense but also to receive security patches and other necessary updates, thereby leaving ample room for interference (Greenbaum, 2021, p. 665).

AI and Tilting Human Agency

While the above clearly points towards significant security threats regarding one's thought processes and behaviours, these do not necessarily have to involve hacking into digital devices and tools. After all, individuals can hardly be seen to have the capacity to act beyond the information affordances and world perceptions available to them. In this case, AI functions as a crucial mediating (an, often, even shaping) force. In many ways, humans can be influenced even without external intentionality, nefarious or otherwise, owing only to the overall functioning of this technology as such, becoming, so to say, accidental joysticks. This is particularly the case when generative AI capacities are combined with virtual experiences or news ecosystems.

As a term and a concept, AI might be difficult to define in the sense that it can be seen as an umbrella term used for a broad range of related technologies of very different complexity and innovativeness – after all, it has been around since the 1950s, thereby unavoidably denoting a mixture of the old and the new (see e.g. Beckett, 2019; Beckett & Yaseen, 2023). Broadly, though, the purpose of AI can be said to be the performance of tasks that are typically performed by human intelligence using reasoning, perception, prediction, and motor control (Boden, 2016). As such, AI is fundamentally imitative, that is, it *simulates* the external effects of the functioning of the human mind (McLay 2018). However, contrary to natural intelligence, which has developed independently though the process of evolution, artificial intelligence, being human-designed, is purpose-dependent: rather than being a simulation of the external workings of human intelligence for its own sake, AI is created with the aim of accelerating, extending, or completely automating specific tasks typically carried out by humans (Rinehart & Kung, 2022). Hence, the creation of AI is a task-oriented activity, resulting in AI itself being geared towards conveying benefits to those developing and/or deploying it.

Of the different types of AI, the generative variety has recently attracted the most attention. Broadly speaking, generative AI can produce new visual, audio, or textual content based on patterns learned from training data, typically in response to a prompt. Among other things, generative AI can enable always-on virtual interactions whereby chatbots, virtual assistants, and non-player characters in virtual experiences become interactive and human-like to an unprecedented degree – in fact, sufficiently so to instil perceptions of friendship and even intimacy (Brandtzaeg et al., 2022). More broadly, the increasing abundance of AI-generated content means that conventional distinctions between 'real' and 'artificial' are becoming increasingly problematic, to the extent that "new realities become an ever-increasing part of our digital lives – which themselves are merging with our *real* lives to the point where the distinction will soon become moot, if still technically accurate" (Rubin, 2020, p. 124). In fact, it transpires that on the experiential level in particular, the differences between human and artificial, the physical and the virtual have become insignificant (Branca et al., 2023; Schöne et al., 2023).

The above, once again, has important consequences for virtual experiences by making them perceptively real to an unprecedented extent, particularly by ensuring that the content of virtual worlds adjusts to human users in real time as a result of automated content generation (see e.g. Kalpokas, 2024a). The sense of embodied presence thereby accrued (and, particularly, of co-presence with other persons, both real and artificial) further strengthens the effects' immersiveness and persuasiveness (Zhang et al., 2022). However, immersiveness comes at a cost: as stressed by Bailenson (2018): "[i]n VR,

which actually "feels real", the potential dangers for misinformation and emotional manipulation are exponentially greater" (p. 67). Notably, VR can induce stronger identification with objects that are encountered and experienced than those mediated by a screen, a printed page, or any other type of medium, including the creation of vivid memories that feel 'real' despite having no correlate in the 'real' (i.e. physical) world (Rubin, 2020).

What takes shape in virtual environments is not mere 'storytelling' but 'story-living': individuals do not passively encounter a story or a plot but, instead, experience it first-hand, becoming participants and, therefore, experiencing greater proximity to the unfolding events and stronger emotional investment than in the case of traditional communication techniques (Kukkakorpi & Pantti, 2021). The preceding is even further enhanced through virtual social interactions, at least as they fall in line with, and support, the message being put forward, thereby creating a social and spatial mix that engulfs the user (Kukkakorpi & Pantti, 2021). Unsurprisingly, concerns have been raised that the viscerality of VR, and the emotional response induced, can result in both psychological harm (in case of particularly distressful content) and 'mental and behavioural manipulation' (Mabrook, 2021, p. 210). Indeed, it can reasonably be claimed that 'virtuality is not virtual' but, instead, real for those who are immersed in it, causing significant ethical challenges for those creating VR experiences (Lin & Hsu, 2023) or that 'virtual reality is genuine reality' and engaging on it exceeds mere escapism (Chalmers, 2022, xvii). However, while the physical reality exists independently and does not have to be intentionally structured and curated (similarly to the distinction between artificial and natural intelligence above), VR needs to be put together at the expense of time, money, and other resources. While many (probably most) of such experiences will be created with the aim of monetizing human hedonic drives or as extensions (and possibly replacements) of physical environments (e.g. workplaces), at least some of them will be created with the aim to manipulate people. Again, the implication is that not only hacking but also strategic content provision can lead to behavioural manipulation, i.e. turning individuals into externally manipulable joysticks of sorts.

VR is also highly interactive not only with regards to simulated environments but also when it comes to individuals and personalities, including those who have a physical embodiment beyond the VR experience ('real' humans in the traditional sense) and those who do not, i.e. virtual AI-powered entities (Jiang, 2022). Co-presence with such virtual entities has been shown to enhance the user's experience beyond the realms of enjoyment and reward, rendering it more plausible and generating stronger fixation of the user's attention (Jayawardena et al., 2023). Such effects are likely to be partic-

ularly strong when virtual agents are combined with the communicative potential of generative AI: already in non-virtual settings, 'interactions between humans and AI, 'particularly if filled with anthropomorphic and emotional cues, may lead consumers to feel connected with the AI agents', such as digital assistants, forgetting the artificial nature of the latter and sometimes even forging perceived emotional bonds (Guerreiro & Loureiro, 2023, p. 1-2). Indeed, the more anthropomorphic such digital agents become, the more trustworthy they are perceived as, thereby opening new grounds for persuasion (Balakrishnan & Dwivedi, 2024), becoming AI-powered accessories in the hands of those intending to exert influence.

Normalisation of Synthetic Realities

Moving to the role of AI in the provision of information to the public, AI already performs an important function in the "traditional" domains, such as journalism. Journalism and artificial intelligence have a complex relationship, encompassed by the concept of "algorithmic journalism". Algorithmic journalism, also known as "robot journalism" or "automatic journalism", refers to the use of algorithms and computer systems to automate content generation and curation or to transform structured data into texts (Canavilhas, 2022; Pavlik, 2023; Porlezza, 2024). This, again, points towards the purpose-oriented and benefit-driven nature of the development and application of AI already noted above. With the preceding in mind, it becomes clear that news and information provision is being transformed as outlets dedicate increasing amounts of resources to automation, thereby transforming journalistic work practices and ways in which news is gathered, produced, and distributed. Indeed, as stressed by Beckett & Yaseen (2023), AI is used more or less equally in newsrooms at every step of the content creation process, from ideation to publishing. In this way, several layers of information threats pertaining to generative AI are enabled.

As automated news gathering and social listening tools are being increasingly used, the quality of journalistic output is increasingly dependent on the capacity of the resulting human-AI teams to separate AI-generated sources and chatter from authentic ones – something that might be increasingly difficult given the time and efficiency pressures faced by today's newsrooms (Beckett & Yaseen, 2023). Such pressures are also likely to cause a pivot towards preference for AI and its machinic speeds of production at the expense of human journalists. Such a shift might lead to an increased susceptibility to astroturfing, i.e. attempts to manipulate public debate by injecting fake news and/or inflating the salience of existing fake narratives, typically by way of using generative AI and automated social media accounts (see e.g. Chan, 2024). Automated news gathering and social listening tools

(particularly if lacking adequate supervision) may succumb to such disinformation and "launder" its claims (in analogy with money laundering) by making them part of the legitimate news ecosystem (an example of such a blunder, albeit in a slightly different context, could be that of Google's AI Overview – see e.g. Williams, 2024). Hence, baking ethical and normative commitments that normally are at the heart of journalistic practices into automated content generation tools – even assuming that it is possible in the first place – will be a very steep task indeed (Peña-Fernández et al., 2023).

Regarding news production, generative AI technologies, Large Language Models (LLMs) in particular, have become a valuable resource in content creation, including the production of summaries, headlines, visual storytelling, targeted newsletters, translation of articles and assessing different data sources, providing headline alternatives, tagging articles, illustration, video and audio production, and data sifting to deliver real-time news updates. However, their use for the generation of articles has notably increased (Beckett & Yaseen, 2023; Pavlik, 2023; Arguedas & Simon, 2024), which raises fundamental concerns, such as rendering journalism dependent of inputs and prompt suggestions (as indicated above) or the production of news pieces which, due to the functioning of the generative models themselves, are subject to biases and intentional data poisoning. Therefore, even reliable prompts could end up producing manipulative results (see e.g. Peña-Fernández et al., 2023; Tomlinson et al., 2023).

No less importantly, AI is being used in news distribution to customise content and increase audience reach and engagement by using personalisation and recommendation systems to match content more accurately and at scale with interested audiences (Beckett & Yaseen, 2023). However, such tools have been observed to cause the so-called filter bubble effect whereby individuals are being exposed to disproportionate amounts of content that they already agree with, thereby further solidifying (and potentially radicalising) their positions (see e.g. Coeckelbergh, 2023; Eg et al., 2023; Palmieri, 2024).

Crucially, within the active adoption of AI in journalism, the former's gatekeeping function is growing. AI algorithms can rapidly analyse vast amounts of data to select, produce, and present news stories, prioritise and sort public information, and, in this way, filter content (van Dalen, 2023). Since AI algorithms are getting involved in all the stages of the news developing cycle – gathering, production and distribution – they may act as gatekeepers of information flow, shaping the content users see based on various criteria, potentially impacting the diversity and inclusivity of news consumption, often serving particular interests in their repetition, reproduction, and duplication of the content (van Dalen, 2023). Once again, issues of algorithmic bias and automated content tailoring and/or delivery in accordance with audience tastes loom large, only this time

to be repeated across all stages of the news content production chain, with misinformation, disinformation (the latter presumably originating outside newsroom, but being amplified and "laundered" through automated journalism), and biased content becoming major points of concern.

A further way in which completely synthetic realities are created pertains to interactions between agents, powered by generative AI. Focus groups and even communities of simulated individuals are already used for everything from generating marketing ideas to researching and modelling human interactions (see e.g. Knight 2023). While such and other ways of generating synthetic data with generative AI are often lauded for their efficiency and zoom into niche groups (thus supposedly de-biasing large data sets), they nevertheless, can only build on what is already datafied, that is, reproducing existing patterns, albeit in more granular detail and with a greater clout of supposed objectivity (see e.g. Offenhuber, 2024).

Joysticks Everywhere and Hacking Without Hacking

Of course, the problem is further exacerbated when deliberate fakery using generative AI is concerned. Here, threat actors are now capable of creating a semblance of events that never took place or putting targeted individuals in situations they were never in. Newman (2023) sees a serious risk in AI developing abilities to produce immediate plausible content that will make it harder to tell the difference between real and false information, either misleading audiences or compromising their trust. There is, consequently, an acute need for transparency, particularly in terms of content creators adhering to ethical guidelines and AI-generated content being watermarked or otherwise clearly labelled (Newman, 2023). The EU's AI Act marks an important step in that direction, although by no means a perfect one (for a discussion of potential weaknesses, see e.g. Kalpokas, 2024a).

Crucially, generative AI models are not only easily accessible; what is more, their output is increasingly difficult to distinguish from authentic content (Jungherr & Schroeder, 2023, p. 169). That, in turn, has the capacity to lead to production of disinformation at scale: even without nefarious intent, generative AI 'has no commitment to the truth of an argument or observation; it is only imitating their likeness as found in past data', thereby causing so-called hallucinations, but when the goal of content creation is outright manipulative, AI enables efficient disinformation both in a targeted manner (i.e. directed at particular individuals and their vulnerabilities) and *en masse* (Jungherr, 2023, p. 6). Indeed, generative AI can be used to create anything from synthetic identities (deceptively or authentically fake profiles and personalities to synthetic public

knowledge (i.e. information environments in which the perception of at least one fundamental aspect of reality has been significantly altered) and beyond, including, when coupled with VR, completely synthetic realities (Ferrara, 2024).

Indeed, recent innovations in the field of generative AI pose the danger of 'supercharging pre-existing risks, potentially unleashing harmful content on an unprecedented scale and with great impact' by challenging democratic integrity, particularly under the aspects of equality, truth, and non-violence (Judson et al., 2024, pp. 5, 11). Notably, encounters with AI-generated content have the capacity to cause misperceptions, potentially to the extent of overwriting pre-existing memories, or at the very least, to erode trust in the common standards of truth and socio-political reality (Weikmann & Lecheler, 2023), thereby contributing to the drive towards relying on opinion-congruent content and seeking comfort in filter bubbles (Kalpokas, 2024b). Indeed, there already is a tendency to evaluate – or (mis)recognise – AI-generated content in line with existing convictions and heuristics (Shin et al., 2024).

When combined with VR experiences, generative AI capacities allow the creation of fully personalised and predictive realities and fracture of shared frameworks that hold society together (Bay, 2023). The emergence of such personally enclosed realities then allows for more effective manipulation because there is no shared reality to lean on or measure one's perceptions against. In this way, human agency can be even more fully determined by externally induced causes or beliefs.

What should be noted, though, is that even awareness of disinformation might be a double-edged sword. In fact, the negative connotations attached to falling for disinformation seem to discourage people from acknowledging their own vulnerability, instead ascribing it to some supposedly uniquely naïve or gullible others; hence, disinformation is seen as a problem for those 'other' individuals rather than a society-wide issue, thereby reducing the motivation to participate in collective mitigation efforts (Hall et al., 2024). Specific to generative AI, though, regulation efforts find themselves in a paradoxical situation whereby their representatives, courtesy to AI's ability to *produce* and participate in public discourse, can also recursively insert itself in any discussions pertaining to its own regulation (Anany, 2024). In this way, changes to the current ecosystem may be difficult to formulate, let alone implement.

Yet another set of issues to be considered pertains to the very functioning of online platforms (such as social media ones) that have by now become crucial intermediaries in the processes and practices of information supply. Nowadays, information overload and the ensuing need to deal with emerging challenges by way of automation in content selection and delivery (since no team of human moderators, yet alone individual

end users, would be able to deal with the incessant accumulation of content), afforded by platforms companies, becomes the key variable (see, from different perspectives, Hepp & Couldry, 2023). This leads to the emergence of algorithm dependency due to the dominance of platformised news use (Schaetz et al., 2023). In fact, the *power* of algorithmic content recommendations can be disconnected from their *quality* while, nevertheless, remaining in place (Hagar & Diakopoulos, 2023). Notably, algorithmic and AI-based content governance decreases the role and influence of sources (including media companies and news organisations) as well as other content creators by rendering them subservient to centralised governance and moderation practices (Jungherr & Schroeder 2023, p. 168). By contrast, the power of technology companies increases, making the latter central to the functioning of contemporary societies (Jungherr, 2023). Here one must agree with van Dijck (2024) in that platformisation should now be seen as "the prism through which we should critically examine how technological shifts that are simultaneously social, economic, cultural, and political transformations affect the global power (im)balance while deeply infiltrating private lives and public spaces" (p. 2). Communication thus becomes an automated process that operates on a supra-individual level (Hepp et al., 2023).

If we understand the public arena as an interconnection of spaces which "allows societies to settle crucial issues and control elites and governments, and for groups with shared concerns to emerge" (Jungherr & Schroeder, 2023, p. 165), then such crucial processes must be seen as unavoidably mediated by algorithms, and, through them, by online platforms. In fact, platforms, though their algorithms, "enable and constrain the publication, distribution, reception, and contestation of information that allows people to exercise their rights and duties in pursuit of the public good" (Jungherr & Schroeder, 2023, p. 165). The preceding makes individual autonomy and self-rule and meaningful democratic participation difficult to achieve (Jungherr, 2023).

As the reality perceptively inhabited by individuals is increasingly comprised of data, and the bits that comprise the latter are primarily owned by technology companies, effectively, the net result points towards the privatisation of the perceived and experienced reality (Lemley & Volokh, 2018). Others, meanwhile, would go even further to argue that, as the development and operation of the current platform ecosystem largely reflects historical patterns of colonial domination, they only further entrench the dominant structures of power and knowledge (Bannerman, 2024), meaning that harms are suffered not only at an individual level or that of a (political) community but also globally. Crucially, though, the trend among platform companies, both traditional ones and the newly emergent generative AI ones, is to attempt

to avoid responsibility by claiming to be merely neutral intermediaries that provide an open space for people's opinions and for debate, despite heavily structuring such (self-)expression and, in the case of generative AI providers, even co-constituting (self-)expression (Edwards et al., 2024). In this way, the hackability of the self can be understood as a daily – even permanent – condition.

Conclusion

Autonomy and independent agency are traditionally seen as keen human faculties – indeed, as capacities in which human exceptionality is typically based. However, the current technological environment, both independently and as a matter of abuse, challenges such capacities, instead rendering humans into joysticks in somebody else's hands. The term itself is borrowed from so-called Human Joystick Attack, aimed at manipulating human movements in the physical space through virtual reality. Such threats are even further exacerbated through the hackability of brain-computer interfaces, potentially to the extent of altering the thought processes of targeted individuals. Nevertheless, such threats are only the extreme end of a spectrum. Comparable levels of loss of autonomy and agency can also be achieved by way of employing generative AI to manipulate existing information environments and create completely synthetic ones – a process even further strengthened when, again, coupled with VR technologies. Still, one could zoom out even further and consider the broader algorithmic governance processes enabled by online platforms, particularly the extent to which they structure public spaces and their content, thereby determining what is (and is not) to be known. In this way, decision-making and, consequently, behavioural autonomy is further curtailed, implying that humans today have little other option that to be joysticks in somebody else's hands.

References

Ajrawi, S., Rao, R., & Sarkar, M. (2021). Cybersecurity in Brain-Computer Interfaces: RFID-Based Design-Theoretical Framework. *Informatics in Medicine*, *22*, 1-9. https://doi.org/10.1016/j.imu.2020.100489

Anany, M. (2024). Making Generative Artificial Intelligence a Public Problem: Seeing Publics and Sociotechnical Problem-Making in Three Scenes of AI Failure. *Javnost – The Public*, *31*(1), 89-105. https://doi.org/10.1080/13183222.2024.2319000

Arguedas, A. R. & Simon, F. M. (2024). Automating Democracy: Generative AI, Journalism, and the Future of Democracy. *Oxford Internet Institute*, https://www.oii.ox.ac.uk/news-events/reports/automating-democracy-generative-ai-journalism-and-the-future-of-democracy/.

Bailenson, J. (2018). *Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do*. W. W. Norton and Company.

Balakrishnan, J. and Dwivedi, Y. K. (2024). Conversational Commerce: Entering the Next Stage of AI-Powered Digital Assistants. *Annals of Operations Research*, *333*, 653–687. https://doi.org/10.1007/s10479-021-04049-5

Bannerman, S. (2024). Platform Imperialism, Communications Law and Relational Sovereignty. *New Media & Society*, *26*(4), 1816-1823. https://doi.org/10.1177/14614448221077284

Bay, M. (2023). Arendt in the Metaverse: Four Properties of eXtended Reality that Imperil *Factual Truth* and Democracy. *Convergence: The International Journal of Research into New Media Technologies*, *29*(6), 1698-1712. https://doi.org/10.1177/13548565231199957

Beckett, C. (2019). New Powers, New Responsibilities: A Global Survey of Journalism and Artificial Intelligence. *LSE Polis*, https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/.

Beckett, C. & Yaseen, M. (2023). Generating Change: A Global Survey of What News Organisations Are Doing With AI. *LSE Polis*, https://www.journalismai.info/s/Generating-Change-_-The-Journalism-AI-report-_-English.pdf.

Boden, M. A. (2016). AI: Its Nature and Future. Oxford University Press.

Branca, G., Resciniti, R., & Loureiro, S. M. C. (2023). Virtual Is so Real! Consumers' Evaluation of Product Packaging in Virtual Reality. *Psychology & Marketing*, *40*(3), 596–609. https://doi.org/10.1002/mar.21743

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human-AI Friendship. *Human Communication Research*, *48*(3), 404–429. https://doi.org/10.1093/hcr/hqac008

Brocal, F. (2023). Brain-Computer Interfaces in Safety and Security Fields: Risks and Applications. *Safety Science*, *160*, 1-16. https://doi.org/10.1016/j.ssci.2022.106051

Canavilhas, J. (2022). Artificial Intelligence and Journalism: Current Situation and Expectations in the Portuguese Sports Media. *Journalism and Media*, *3*(3), 510–520. https://doi.org/10.3390/journalmedia3030035

Cayir, D. et al. (2024). Augmenting Security and Privacy in the Virtual Realm: An Analysis of Extended Reality Devices. *IEEE Security & Privacy Magazine*. https://doi.org/10.1109/MSEC.2023.3332004

Chalmers, D. J. (2022). *Reality +: Virtual Worlds and the Problems of Philosophy*. W. W. Norton & Company.

Chan, J. (2024). Online Astroturfing: A Problem Beyond Disinformation. *Philosophy and Social Criticism*, *50*(3), 507-528. https://doi.org/10.1177/01914537221108467

Chow, Y.-W. et al. (2023). Visualization and Cybersecurity in the Metaverse: A Survey. *Journal of Imaging*, *9*(1), 1-15. https://doi.org/10.3390/jimaging9010011

Coeckelbergh, M. (2023). Democracy, Epistemic Agency, and AI: Political Epistemology in Times of Artifcial Intelligence. *AI and Ethics*, *3*, 1341–1350. https://doi.org/10.1007/s43681-022-00239-4

Edwards, L. et al. (2024). Private Ordering and Generative AI: What Can We Learn from Model Terms and Conditions? *CREATe Working Paper 2024/5*. https://doi.org/10.2139/ssrn.5026677

Eg, R., Tønnesen, Ö. D., Tennfjord, M. K. (2023). A Scoping Review of Personalized User Experiences on Social Media: The Interplay Between Algorithms and Human Factors. *Computers in Human Behavior Reports*, *9*, 1-17. https://doi.org/10.1016/j.chbr.2022.100253

Ferrara, E. (2024). GenAI against Humanity: Nefarious Applications of Generative Artificial Intelligence and Large Language Models. *Journal of Computational Social Science*, *7*, 549–569. https://doi.org/10.1007/s42001-024-00250-1

Flavián, C., Ibáñez-Sánchez, S., & Orús, C. (2019). The Impact of Virtual, Augmented and Mixed Reality Technologies on the Customer Experience. *Journal of Business Research*, *100*, 547-560. https://doi.org/10.1016/j.jbusres.2018.10.050

Greenbaum, D. (2021). Cyberbiosecurity: An Emerging Field that has Ethical Implications for Clinical Neuroscience. *Cambridge Quarterly of Healthcare Ethics*, *30*(4), 662–668. https://doi.org/10.1017/S096318012100013X

Guerreiro, J. & Loureiro, S. M. C. (2023). I Am Attracted to My Cool Smart Assistant! Analyzing Attachment-Aversion in AI-Human Relationships. *Journal of Business Research*, *161*, 1-13. https://doi.org/10.1016/j.jbusres.2023.113863

Hagar, N. & Diakopoulos, N. (2023). Algorithmic Indifference: The Dearth of News Recommendations of TikTok. *New Media & Society*. https://doi.org/10.1177/14614448231192964

Hall, A.-N., Chadwick, A., & Vaccari, C. (2024). Online Misinformation and Everyday Ontological Narratives of Social Distinction. *Media, Culture & Society*, *46*(3), 572-590. https://doi.org/10.1177/01634437231211678

Han, D.-I. D., Bergs, Y., & Moorhouse, N. (2022). Virtual Reality Consumer Experience Escapes: Preparing for the Metaverse. *Virtual Reality*, *26*, 1443–1458. https://doi.org/10.1007/s10055-022-00641-7

Hepp, A. & Couldry, N. (2023). Necessary Entanglements: Reflections on the Role of a 'Materialist Phenomenology' in Researching Deep Mediatization and Datafication. *Sociologica*, *17*(1), 137-153.

Hepp, A. et al. (2023). ChatGPT, LaMDA, and the Hype Around Communicative AI: The Automation of Communication as a Field of Research in Media and Communication Studies. *Human-Machine Communication*, *6*, 41-62. https://doi.org/10.30658/hmc.6.4

Huang, Y.C., Li, L.N., Lee, H.Y., Browning, M.H. and Yu, C.P. (2023), Surfing in virtual reality: an application of extended technology acceptance model with flow theory. *Computers in Human Behavior Reports*, *9*, 100252. https://doi.org/10.1016/j.chbr.2022.100252

Jayawardena, N. S., Thaichon, P., Quach, S., Razzaq, A., & Behl, A. (2023). The Persuasion Effects of Virtual Reality (VR) and Augmented Reality (AR) Video Advertisements: A Conceptual Review. *Journal of Business Research*, *160*, 1-17. https://doi.org/10.1016/j.jbusres.2023.113739

Jiang, C. (2022). The Making of Popstar Fembots: Participation, Co-Creation, or online Cultural Exploitation? *Hybrid: Journal of Arts and Human Mediations*, 8, 1-12. https://doi.org/10.4000/hybrid.2254

Judson, E., Fisher, S. A., Howard, J. W., Kira, B., Basavaraj, K. A., & Perry, H. (2024). *Synthetic Politics" Preparing Politics: Preparing Democracy for Generative AI*. Demos.

Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, *9*(3). https://doi.org/10.1177/20563051231186353

Jungherr, A. & Schroeder, R. (2023). Artificial Intelligence and the Public Arena. *Communication Theory*, *33*(2-3), 164-173. https://doi.org/10.1093/ct/qtad006

Kalpokas, I. (2024a) *Technological Governance and Escapism in Times of Accelerated Change*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-60890-2

Kalpokas, I. (2024b). Post-Truth and Information Warfare in their Technological Context. *Applied Cybersecurity and Internet Governance*, *3*(2), 99-121. https://doi.org/10.60097/ACIG/190407

King, B. J., Read, G, J. M., & Salmon, P. M. (2024) The Risks Associated with the Use of Brain-Computer Interfaces: A Systematic Review. *International Journal of Human-Computer Interaction*, *40*(2), 131-148. https://doi.org/10.1080/10447318.2022.2111041

Knight, W. (2023, October 12). The Chatbots Are Now Talking to Each Other. *Wired*. https://www.wired.com/story/fast-forward-the-chatbots-are-now-talking-to-each-other/

Kukkakorpi, M. & Pantti, M. (2021). A Sense of Place: VR Journalism and Emotional Engagement. *Journalism Practice*, *15*(6), 785-802. https://doi.org/10.1080/17512786.2020.1799237

Lemley, M. A. & Volokh, E. (2018). Virtual Reality and Augmented Reality. *University of Pennsylvania Law Review*, *66*(5), 1051-1138. https://doi.org/10.2139/ssrn.2933867

Lin, C. C. & Hsu, Y. C. (2023). The New Ethical Thinking in CGI Immersive Journalism. *Convergence: The International Journal of Research into New Media Technology*, *29*(4), 1033-1053. https://doi.org/10.1177/13548565231176177

Liv, N. (2021). Neurolaw: Brain-Computer Interfaces. *University of St. Thomas Journal of Law and Public Policy*, *15*(1), 328-355.

López Bernal, S., Huertas Celdrán, A., Fernández Maimó, L., Barros, M. T., Balasubramaniam, S., & Martínez Pérez, G. (2020). Cyberattacks on Miniature Brain Implants to Disrupt Spontaneous Neural Signalling. *IEEE Access*, *8*, 1-19. https://doi.org/10.1109/ACCESS.2020.3017394

Mabrook, R. (2021). Between Journalist Authorship and User Agency: Exploring the Concept of Objectivity in VR Journalism. *Journalism Studies*, *22*(2), 209-224. https://doi.org/10.1080/1461670X.2020.1813619

Maiseli, B., Abdalla, A. T., Massawe, L. V., Mbise, M., Mkocha, K., Nassor, N. A., Ismail, M., Michael, J., & Kimambo, S. (2023). Brain-Computer Interface: Trend, Challenges, and Threats. *Brain Informatics*, 10, 1-16. https://doi.org/10.1186/s40708-023-00199-3

Mazarr, M. J., Bauer, R., Casey, A., Heintz, S., & Matthews, L. J. (2019). *The Emerging Risk of Virtual Societal Warfare: Social Manipulation in a Changing Information Environment*. The RAND Corporation. https://doi.org/10.7249/RR2714

McLay, R. (2018). Managing the rise of Artificial Intelligence. *Australian Human Rights Commission*. https://tech.humanrights.gov.au/sites/default/files/inline-files/100%20-%20Ron%20%20McLay.pdf%20

Newman, N. (2023). Journalism, Media, and Technology Trends and Predictions 2023. *Reuters Institute for the Study of Journalism*. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-01/Journalism_media_and_technology_trends_and_predictions_2023.pdf

Odeleye, B., Loukas, G., Heartfield, R., Sakellari, G. Panaousis, E., & Spyridonis, F. (2023). Virtually Secure: A Taxonomic Assessment of Cybersecurity Challenges in Virtual Reality Environments. *Computers & Security*, *124*, 1-17. https://doi.org/10.1016/j.cose.2022.102951

Offenhuber, D. (2024). Shapes and Frictions of Synthetic Data. *Big Data & Society*, *11*(2). https://doi.org/10.1177/20539517241249390

Palmieri, E. (2024). Online Bubbles and Echo Chambers as Social Systems. *Kybernetes*, *54*(4), 2457-2468. https://doi.org/10.1108/K-09-2023-1742

Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator*, *78*(1), 84–93. https://doi.org/10.1177/10776958221149577

Peña-Fernández, S., Meso-Ayerdi, K., Larrondo-Ureta, A., & Díaz-Noci, J. (2023). Without Journalists, There Is no Journalism: The Social Dimension of Generative Artificial Intelligence in the Media. *El Profesional de La Información*, *23*(2), 1-15. https://doi.org/10.3145/epi.2023.mar.27

Porlezza, C. (2024). The Datafication of Digital Journalism: A History of Everlasting Challenges Between Ethical Issues and Regulation. *Journalism*, *25*(5), 1167-1185. https://doi.org/10.1177/14648849231190232

Rinehart, A. & Kung, E. (2022). Artificial Intelligence in Local News: A Survey of US Newsrooms' AI Readiness. *The Associated Press Technical Report*, https://www.researchgate.net/publication/363475725_Artificial_Intelligence_in_Local_News_A_survey_of_US_newsrooms'_AI_readiness.

Rubin, P. (2020). *Future Presence: How Virtual Reality is Changing Human Connection, Intimacy, and the Limits of Ordinary Life*. Harper One.

Schaetz, N., Gagrčin, E., Toth, R., & Emmer, M. (2023). Algorithm Dependency and Platformized News Use. *New Media & Society*, *27*(3), 1360-1377. https://doi.org/10.1177/14614448231193093

Scholz, J. & Duffy, K. (2018). We ARe at Home: How Augmented Reality Reshapes Mobile Marketing and Consumer-Brand Relationships. *Journal of Retailing and Consumer Services*, *44*, 11-23. https://doi.org/10.1016/j.jretconser.2018.05.004

Schöne, B., Kisker, J., Lange, L., Gruber, T., Sylvester, S., & Osinsky, R. (2023). The reality of Virtual Reality. *Frontiers in Psychology*, 14, 1-17. https://doi.org/10.3389/fpsyg.2023.1093014

Shin, D., Koerber, A., & Lim, J. S. (2024). Impact of Misinformation from Generative AI in User Information Processing: How People Understand Misinformation from Generative AI. *New Media & Society*. https://doi.org/10.1177/14614448241234040

Tomlinson, B., Patterson, D. J., & Torrance, A. W. (2023). Turning Fake Data into Fake News: The AI Training Set as a Trojan Horse for Misinformation. *San Diego Law Review*, 60, 641-670.

van Dalen, A. (2023). *Algorithmic Gatekeeping for Professional Communicators: Power, Trust, and Legitimacy*. Routledge. https://doi.org/10.4324/9781003375258

van Dijck, J. (2024). Governing Platforms and Societies. *Platforms & Society*, *1*, 1-2. https://doi.org/10.1177/29768624241255922

Vondráček, M., Baggili, I., Casey, P., & Mekni, M. (2023). Rise of the Metaverse's Immersive Virtual Reality Malware and the Man-in-the-Room Attack & Defenses. *Computers & Security*, *127*, 1-19. https://doi.org/10.1016/j.cose.2022.102923

Weikmann, T. & Lecherer, S. (2023). Visual Disinformation in a Digital Age: A Literature Synthesis and Research Agenda. *New Media & Society*, 25(12), 3696-3713. https://doi.org/10.1177/14614448221141648

Williams, R. (2024, May 31). Why Google's AI Overviews Gets Things Wrong? *MIT Technology Review*, https://www.technologyreview.com/2024/05/31/1093019/why-are-googles-ai-overviews-results-so-bad/.

Zhang, G. et al. (2022). Popularity of the Metaverse: Embodied Social Presence Theory Perspective. *Frontiers in Psychology*, *13*, 1-13. https://doi.org/10.3389/fpsyg.2022.997751