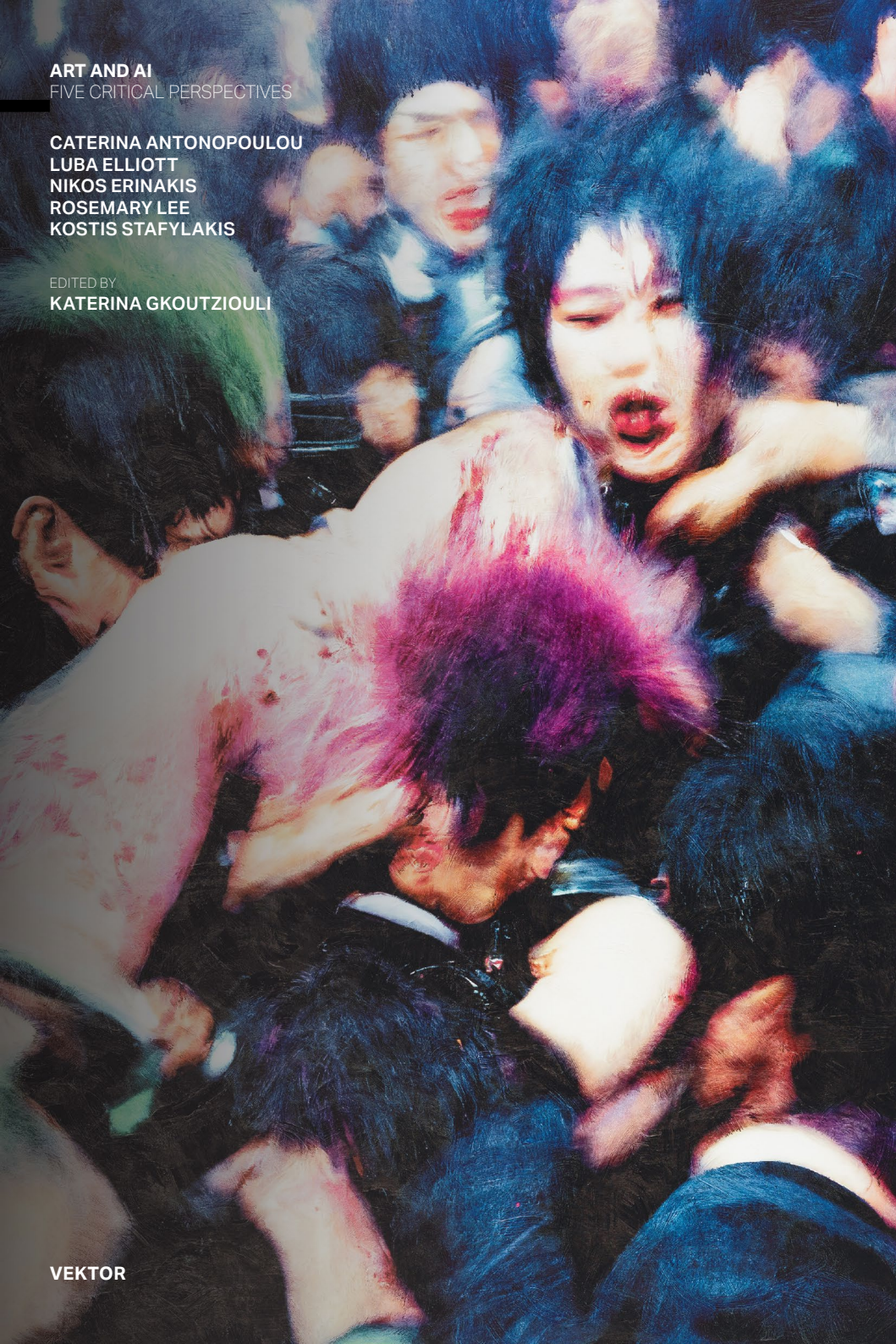
ART AND AI
FIVE CRITICAL PERSPECTIVES

CATERINA ANTONOPOULOU
LUBA ELLIOTT
NIKOS ERINAKIS
ROSEMARY LEE
KOSTIS STAFYLAKIS

EDITED BY
KATERINA GKOUTZIOULI



VEKTOR

ART AND AI
FIVE CRITICAL PERSPECTIVES

VEKTOR

Edited by Katerina Gkoutziouli
Design by YOOOP Studio
Translations by Maria Koukiadi, Daphne Kapsali
Proofreading by Apostolos Vassilopoulos

© 2026, the authors, for their respective texts
© 2026, the artists, for the images of their respective artworks
© 2026, VEKTOR Athens

Cover photograph: Jon Rafman, *ᵖʸᴴʷ·ᵊʸᴴᵠ·ll (Mosh Pit 2)*, 2022
Inkjet print and acrylic on canvas 186.7 × 134.6 cm | 73.5 × 53 inches
© Jon Rafman
Courtesy of the artist and Sprüth Magers
Photo: Ingo Kniest

First edition: Athens, 2026

The publication is licenced under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence.

The publication was realised with the financial support of the Hellenic Ministry of Culture and Error 417 Expectation Failed.

HELLENIC REPUBLIC
Ministry of Culture

ERROR 417
EXPECTATION
FOILED

This edition was printed in 100 copies by Pletsas - Kardaris, Athens, 2026.

ISBN 978-618-85940-3-6

VEKTOR Athens Editions
www.vektorathens.org

ART AND AI
FIVE CRITICAL PERSPECTIVES

CATERINA ANTONOPOULOU
LUBA ELLIOTT
NIKOS ERINAKIS
ROSEMARY LEE
KOSTIS STAFYLAKIS

EDITED BY
KATERINA GKOUTZIOULI

VEKTOR

Contents

Katerina Gkoutziouli 8
Introduction

Luba Elliott 16
AI Art: Beyond Generative AI

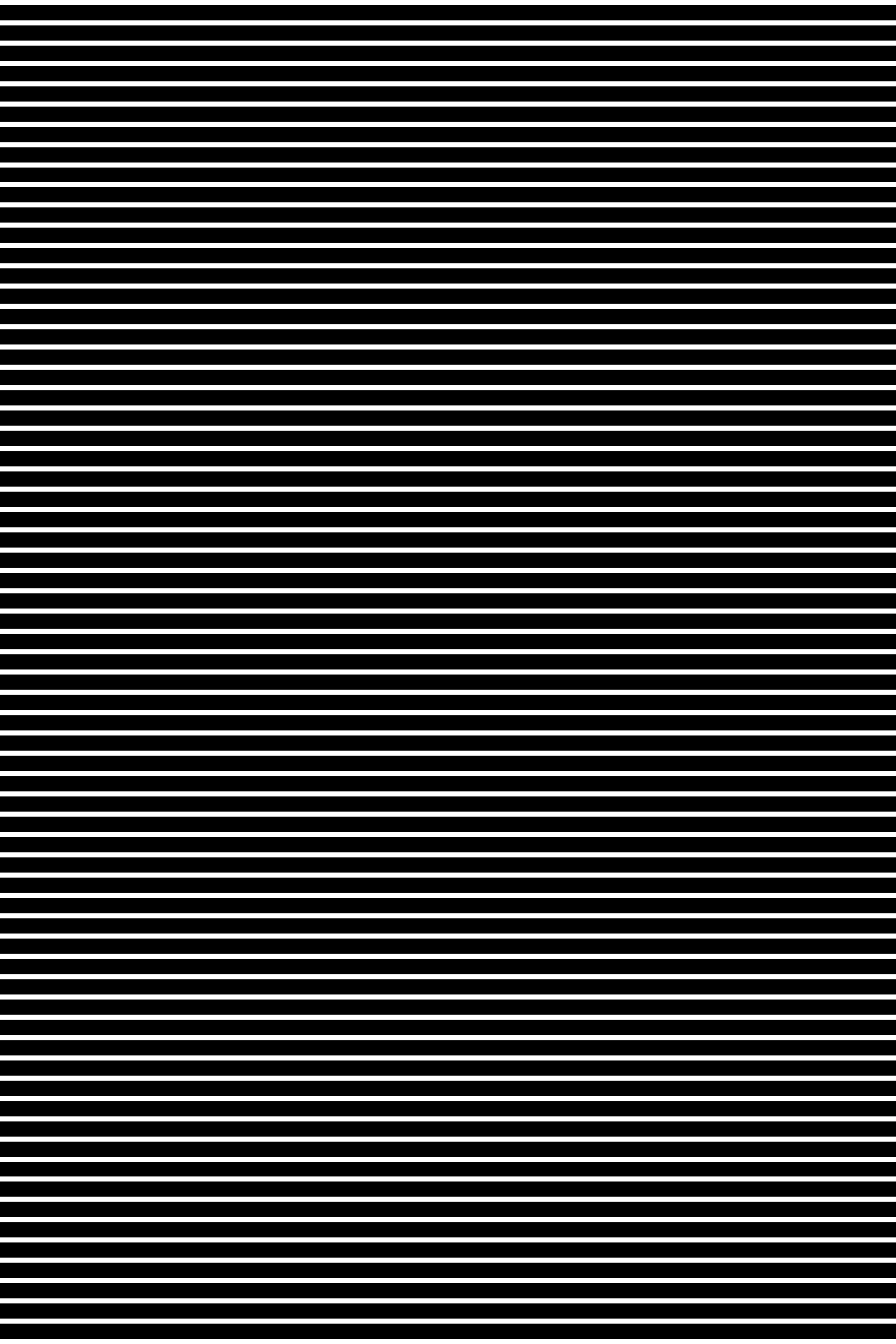
Rosemary Lee 38
The Double-edged Blade of Technological Appropriation

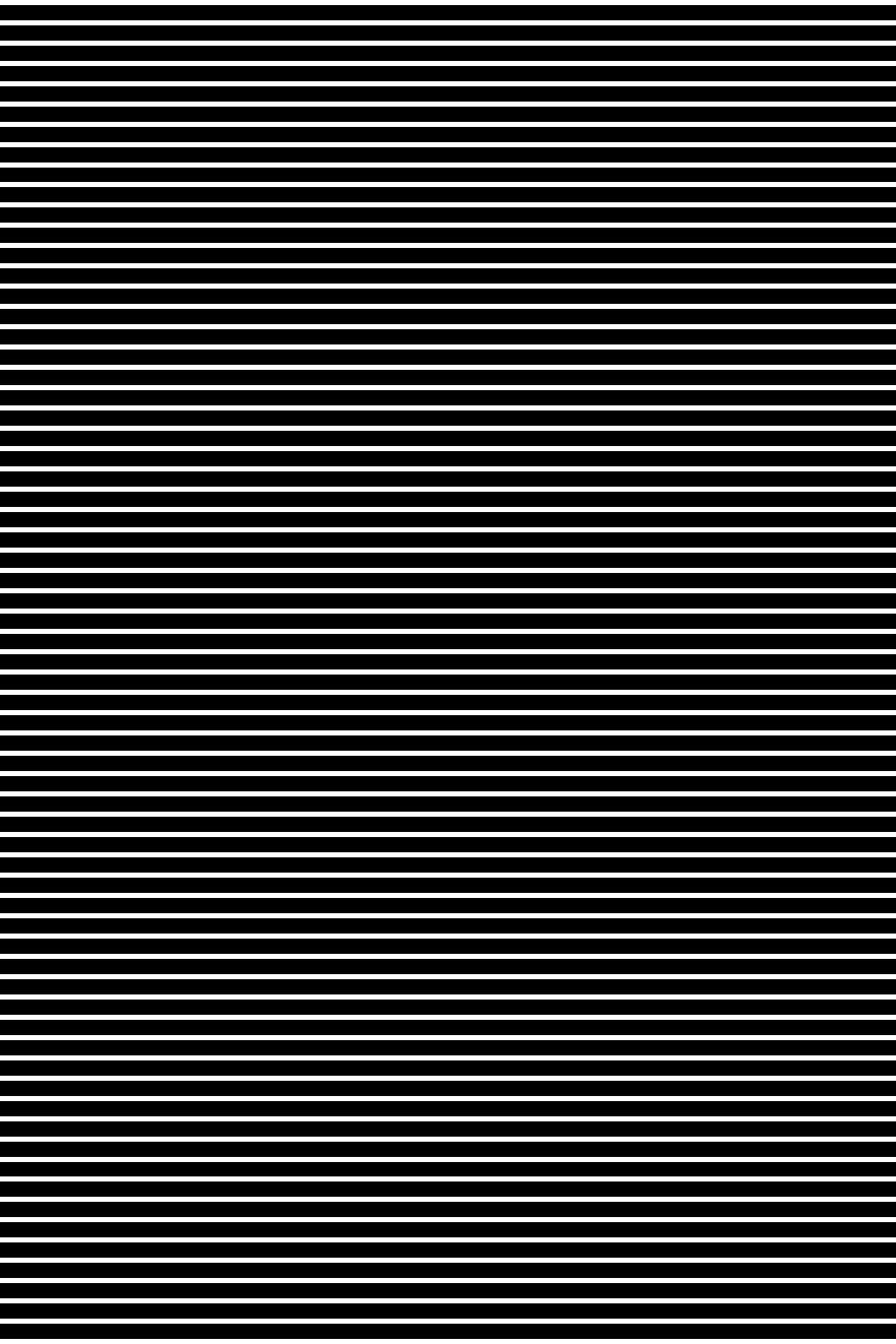
Kostis Stafylakis 54
Beyond Creativity: Emerging Artificial Intelligence Models and Internet-aware Art

Caterina Antonopoulou 76
The Myth of Objective Artificial Intelligence Through the Lens of Critical Digital Art

Nikos Erinakis 110
Relational Authenticity and Hybrid Creativity in the Hyperreal Age of AI

Contributors' Bios 132





Introduction

Katerina Gkoutziouli
Curator, researcher

The relationship between art and technology has always been dynamic and reciprocal. Artificial intelligence, in its current form, makes this relationship even more complex, as its impact on artistic practice extends beyond the introduction of yet another technological tool. The increasing automation of tasks and the widespread adoption of generative artificial intelligence in recent years have shaped new conditions for artistic production, within which artists are experimenting with different methods and strategies. Applications based on machine learning models have entered multiple spheres of political, economic, and cultural life, significantly influencing the ways in which information is produced, categorised, and disseminated. In this context, the widespread use of artificial intelligence raises significant ethical and political questions, ranging from data mining and algorithmic bias to its impact on work, the environment, and cultural production. As Kate Crawford (2021) points out, artificial intelligence cannot be understood solely as a technical system. It is simultaneously constituted by technical and social practices, institutions, material infrastructures, natural resources, and political and cultural processes, whilst computational systems both reflect and shape social relations as well as ways of understanding and representing the world.[1]

In this framework, artists' engagement with artificial intelligence takes place within a field of multiple forces and interrelationships that shape both the functioning of computational systems and the conditions of artistic creation itself. The interest lies in the processes through which artists choose to engage with artificial intelligence, both practically and conceptually. In her essay "After AI Art", Joanna Zylinska (2024) argues that the initial wave of enthusiasm surrounding artificial intelligence in

in the art world appears to have subsided. The question is no longer whether artificial intelligence is innovative, but how art itself is being redefined through its technical, cultural, and critical engagement with AI, and how this encounter influences the concepts of creativity and human subjectivity.[2] This is also where the editorial focus of this volume lies: on how art negotiates the terms of a technological context in which it actively participates. Criticism, in this sense, is neither the acceptance nor the rejection of artificial intelligence, but the exploration of the conditions, infrastructures, and power relations within which it operates and acquires meaning.

This book brings together five new, commissioned essays that explore the relationship between artificial intelligence and art from different theoretical and artistic perspectives. Focusing on the potential for intervention, the contradictions of technological appropriation, online cultures, algorithmic bias and authenticity, the essays in this volume highlight different ways in which art engages with computational systems: as a field of experimentation, as a method of critique, as a practice of resistance, and as a process of renegotiating human action.

Luba Elliott maps the field of AI art beyond the widely accepted generative models, documenting artistic practices from the 2010s to the present day. Noting the sense of saturation and homogenisation of visual culture often brought about by generative AI, Elliott brings together methods and techniques that artists internationally have turned to in order to explore the social, ethical, aesthetic, and political implications of computational systems. From facial recognition systems and recommendation algorithms to autonomous agents, artificial life, biotech, and robotics, the author examines the unorthodox ways in which artists appropriate technologies and practices, highlighting how they can gain greater control over the creative process. In a landscape dominated by AI slop, Elliott argues that AI art is not defined by the technical adequacy of the results produced, but by its ability to reframe computational systems and reshape the terms of perception, interaction, and meaning.

Rosemary Lee examines the inherent ambiguity that characterises the relationship between art and artificial intelligence. Using the "double-edged blade" metaphor, Lee argues that technology can serve artists' subversive intentions whilst simultaneously undermining and trapping

them within the very frameworks they seek to critique. Artists' critical appropriation of AI may lead to the glorification or even legitimisation of the very mechanisms that they are attempting to deconstruct. In discussing the demystification of artificial intelligence and the "AI-anti-hype" strategy, Lee disassociates AI from the idea of thinking machines, relating it instead to a business plan for the industrialisation of intellectual labour. By linking contemporary artistic practice to the history of the Luddites, the author places today's artistic interventions in dialogue with a broader history of resistance centred on technology, labour, and conditions of production, proposing an approach that involves devising new ways of relating to technologies within an ever-changing landscape.

Kostis Stafylakis shifts the discussion away from the prevailing questions of artistic originality and the "creative self" towards a framework of understanding based on "internet-aware" art. By situating generative AI models within the socio-technical environment of the internet, data extraction, and algorithmic surveillance, he examines how artistic strategies have evolved from the 1990s to the present day. From the net.art hacker and the surfer of post-internet art to the poster of today's internet-core, Stafylakis traces the different forms taken by the artistic subject and the strategies through which it intervenes in, navigates, and participates in internet cultures. In this context, prompt engineering is not merely a technique for content production, but a new form of social antagonism and an integral part of the artistic process itself. As the widespread use of generative AI blurs the boundaries between human-produced and algorithmically generated content, the tensions arising between word and image, and between the world and its representation, take on a distinctly political dimension.

Caterina Antonopoulou debunks the myth of artificial intelligence's objectivity, highlighting how algorithmic systems can automate and reinforce pre-existing social biases. Focusing on algorithmic bias (AI bias), she demonstrates how racial, gender-based, and other forms of discrimination manifest themselves in analytical and generative artificial intelligence, as well as in applications of anthropomorphic AI. Through a series of artworks, she analyses the ways in which digital artists can make visible the internal mechanisms of AI systems and the role of training data in shaping algorithmically generated outcomes. The author argues that contemporary artistic practices go beyond simply exposing errors,

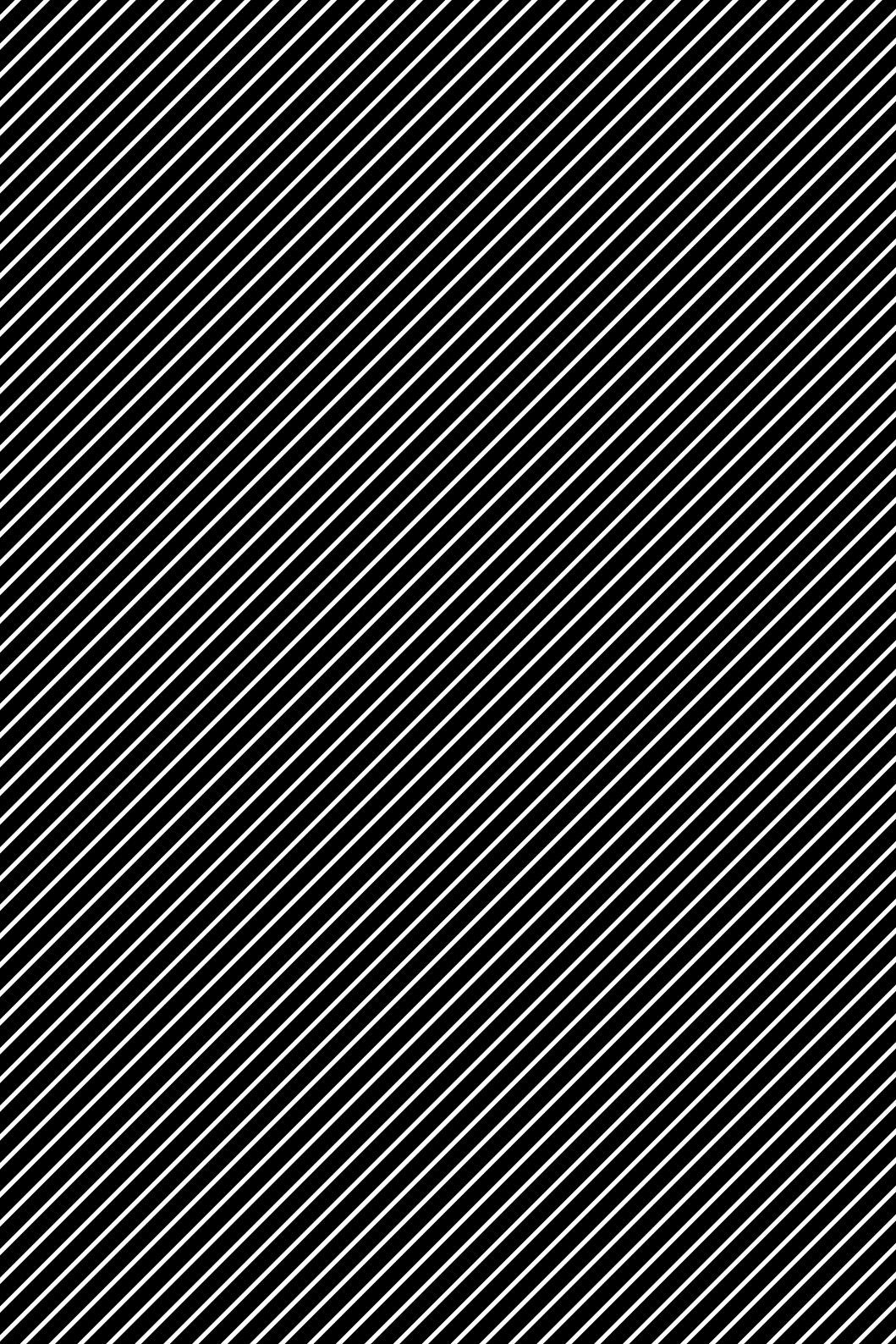
contributing to the search for alternative technological approaches, such as the use of small datasets. The text highlights that there is no single way to design, develop, and train AI, opening up a discussion of the possibilities for a more inclusive technological future.

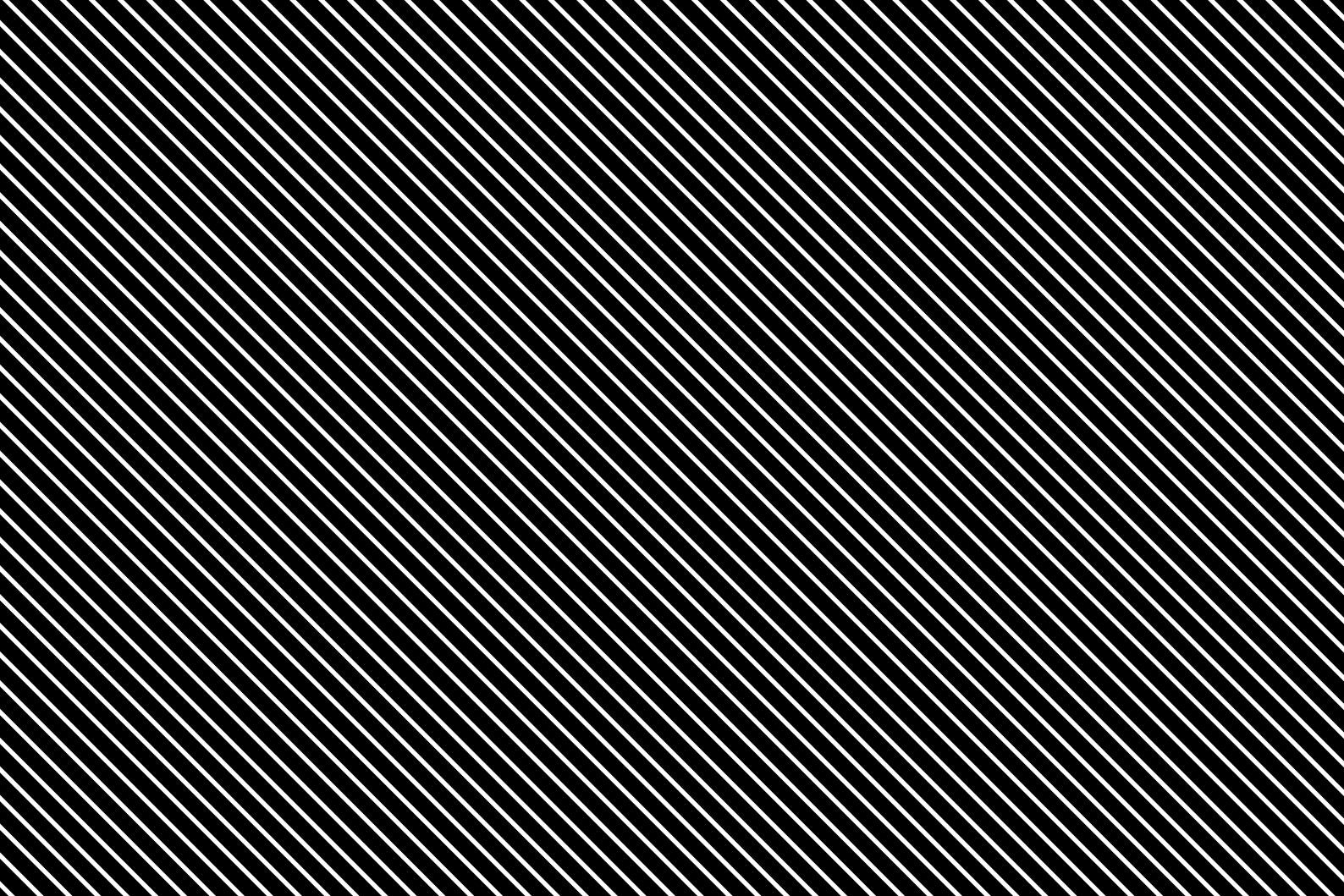
In his text, **Nikos Erinakis** analyses the concepts of authenticity and creativity within the algorithmically shaped landscape of artificial intelligence and digital hyperreality. He connects these two concepts through the idea of "authenticity-as-creativity", emphasising the dynamic creative process of selfhood. At the individual level, he proposes an "aesthetic co-autonomy" of the self as a means of resisting the homogenisation that algorithmic mediation can produce, whilst at a collective level he speaks of the need to commit to a shared reality based on facts and evidence. Erinakis explores the possibility of hybrid creativity in which artificial intelligence becomes part of the creative process, while human agency and judgement remain decisive. Authenticity thus emerges as an ongoing process of co-shaping the self and the world in dialogue with the artificial.

From different starting points, the five texts shift the discussion from the question "What can artificial intelligence create?" to "What happens when artificial intelligence becomes part of the creative process?". What unites the authors' perspectives is the capacity of art to observe and engage with computational systems while simultaneously subjecting them to critique. Art can be actively involved in the infrastructures, economies, and forms of visibility that these systems produce. For this reason, it constitutes a field in which their operations take shape, their contradictions become perceptible, and the terms of the technological condition can be negotiated. This publication aims to offer conceptual tools for understanding the relationship between art and artificial intelligence within a technological context undergoing continuous transformation.

Notes

[1] Crawford, K. (2021). *Atlas of AI*. Yale University Press, p. 8.
[2] Zylinska, J. (2024). After AI art. *Widok: Theories and Practices of Visual Culture*, (38). https://www.pismowidok.org/en/archive/2024/38-digital-entanglements/after-ai-art





AI Art:
Beyond Generative AI

Luba Elliott
AI Art Curator

What's next for AI art? Now that generative AI has become a widespread tool for image creation in the mainstream, what other techniques exist for artists who want to work with AI? This essay looks at the broader spectrum of AI technologies beyond generative AI, including but not limited to facial and object recognition, agent-based systems, bioart and artificial life, as well as AI process artworks, tracing examples from the beginnings of the deep learning revolution in the 2010s to the emerging art practices of today.

One reason artists are beginning to explore alternative approaches is the sense of saturation and fatigue that has emerged around generative AI. Diffusion models, though powerful, often produce aesthetically similar results, leading to a homogenisation of visual culture. This is in addition to the lawsuits against major AI companies for copyright infringement and the rise of prompt marketplaces, which have all contributed to a sense that artists need to carve out alternative strategies for producing genre-defining work. This recalls earlier moments in art history, such as the initial excitement around video art in the 1960s or net art in the 1990s, when novel technologies briefly dominated but soon became part of the mainstream visual toolkit. By contrast, non-generative approaches allow artists to intervene in systems designed for surveillance, commerce, or science, reframing them as aesthetic materials and offering critical perspectives on their use in society.

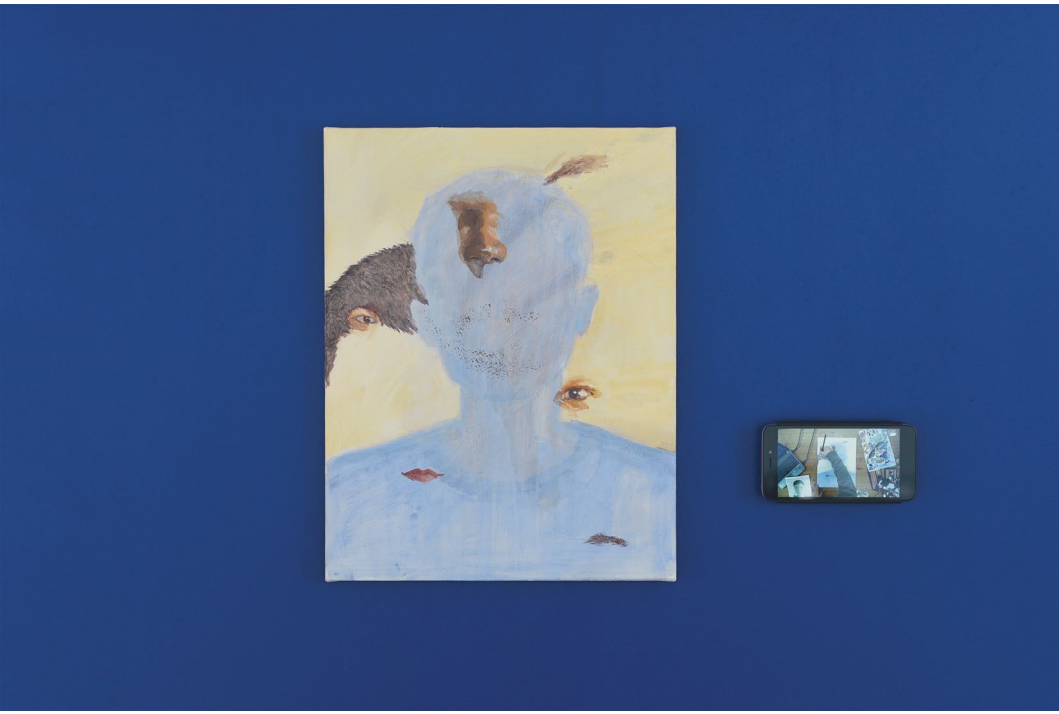
For many of us, the beginning of the 2010s AI art era can be traced back to DeepDream, an AI algorithm developed by Alexander Mordvintsev. Based on a convolutional neural network, it detects

certain features within an image and emphasises them, leading to multicoloured psychedelic images with strange creatures such as puppy slugs emerging from otherwise ordinary scenes. Incidentally, the project came out of Mordvintsev's research in developing a new model for the ImageNet competition.

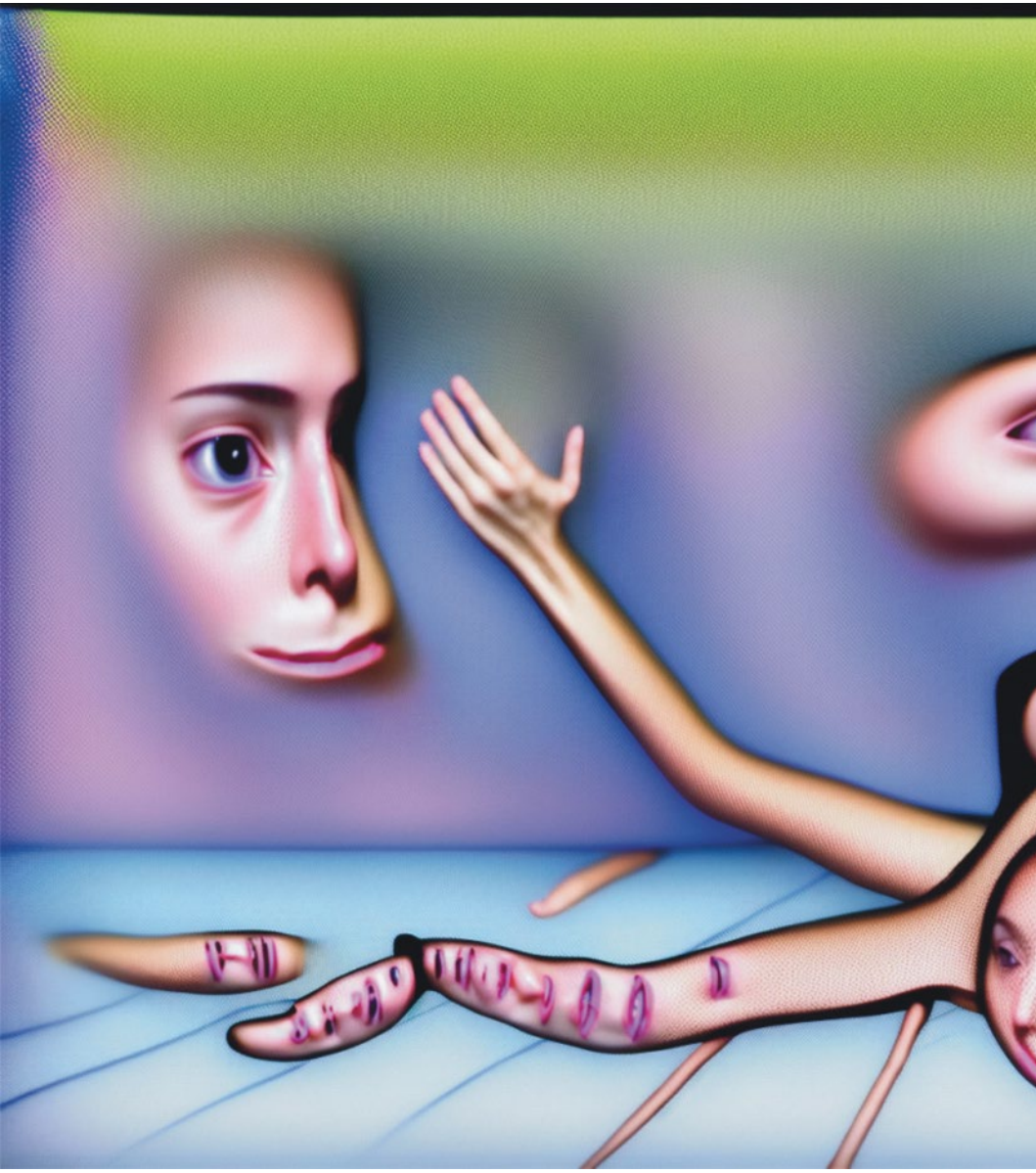
DeepDream's release marked a pivotal moment in how AI art entered popular imagination. The strange, hallucinatory images quickly circulated online, becoming memes and sparking public debates about "what machines see." This viral spread helped crystallise the idea of AI as a creative force, even though the algorithm was originally not intended for artistic practice. Artists and technologists alike began to consider the cultural impact of a system that could reveal hidden layers of perception, connecting the aesthetics of machine vision with longer traditions of psychedelic art, surrealism, and speculative imaginings of non-human creativity.

In the field of object recognition, a key development was the launch of the ImageNet dataset of 14 million images and 20,000 categories together with the associated ImageNet Large Scale Visual Recognition Challenge by Fei-Fei Li in 2010, which encouraged the development of algorithms designed to achieve high accuracy in image recognition. The 2010s saw rapid development of effective models, with the convolutional neural network AlexNet attaining 84.7% recognition in the top five categories and subsequent models taking this figure up to over 95% and even exceeding human performance.

Artists have long engaged with facial recognition tools in both critical and fine arts capacities. Adam Harvey's *CV Dazzle* (2010) used hairstyles and face markings to evade detection, while others have experimented with fine jewellery or projected faces to obscure biometric capture. Facial recognition itself had simultaneously become a key site of political and ethical concern, with many artists positioning their work as a form of resistance to biometric surveillance. Trevor Paglen's *ImageNet Roulette* (2019) revealed how datasets used to train facial recognition systems were riddled with bias, misclassifications, and outdated categories perpetuating harmful stereotypes. Hito Steyerl, through her lectures and performances, has exposed the entanglement of computer vision with military



Shinseungback Kimyonghun, *Non-Facial Portrait 05*, 2018. Courtesy of the artists





Jessica Tucker, *cycles*, 2024. Courtesy of the artist

and commercial infrastructures. As our gaze became redirected onto these systems, artists highlighted both their technical limits and the dangers of delegating social and political decisions to algorithmic classification.

Some works engage the technology in more playful ways. Shinseungback Kimyonghun's *Non-Facial Portraits* (2018) involved drawing or painting under the supervision of a recognition camera, altering the portrait whenever a "face" was detected. An alternative approach was presented by Dries Depoorter in *Face Detected, 2018-2026*, where sculptors hacked away at blocks of clay until the machine recognised a face. This iterative exploration of machine perception of the human face in two and three dimensions highlights the specifics of the two perspectives: sometimes all it takes is some brush strokes in a single colour area, or rotating and misplacing facial features. You can find the reverse logic in Shinseungback Kimyonghun's *Cloud Face* (2012), where AI looks for faces in clouds – a search for meaning, familiarity, and recognition in natural phenomena. This anthropocentrism can be seen in Driessens & Verstappen's *Pareidolia* (2019), where faces are identified in grains of sand. Meanwhile, Jessica Tucker in *cycles* (2024) extends the role of the face as a visual to generate bodies, producing eyes on limbs that are otherwise devoid of detail: for humans, it is the face that often carries the identity and most meaning.

Object recognition has likewise inspired creative misalignments, frequently through the categorisation systems that link distinct visual depictions of an object to one word. Tom White's *Perception Engines* (2018) illustrates this quite clearly by distilling the essence of a particular category, such as "starfish" or "cabbage", into an abstract shape that is recognised as such by image recognition systems. However, some of these shapes, including the starfish, may confuse the human observer because culturally and experientially we have come to know starfish as multi-armed sea creatures as opposed to shapeless bright pink objects. The ability of image recognition systems to identify flowers is challenged in Shinseungback Kimyonghun's work *Flower* (2016–2017), which looks at how flowers can be fragmented and distorted in order to still be recognised as such. The swirling abstract shapes call into question the extent

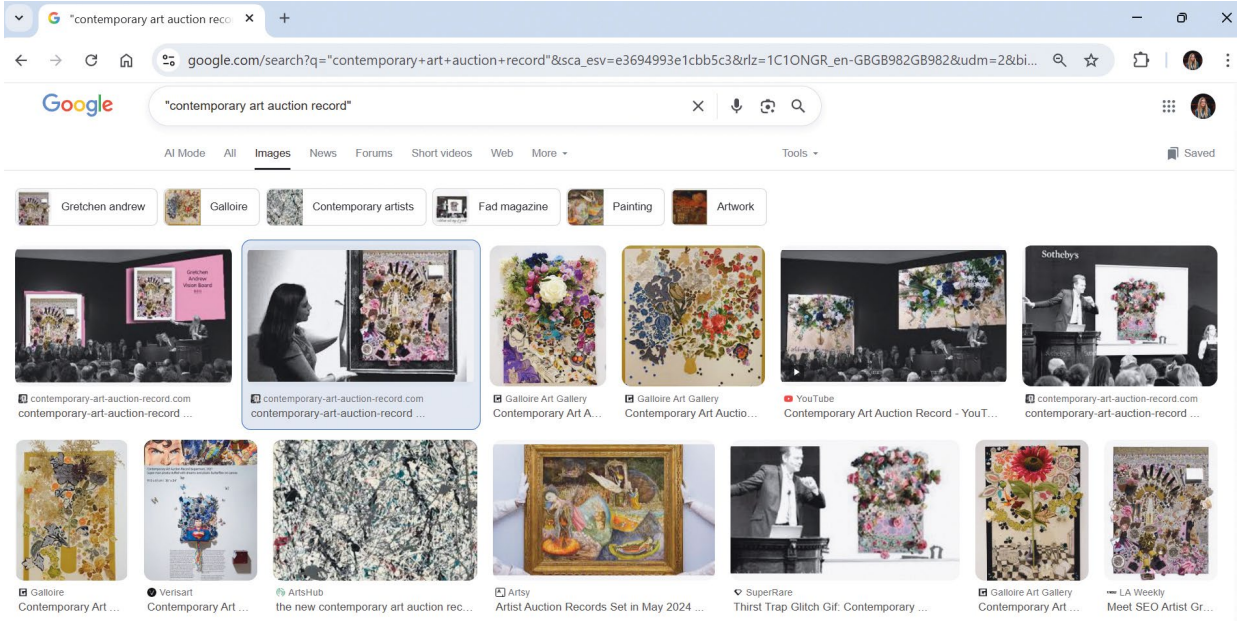


Tom White, *Perception Engines: Starfish,* 2018. Courtesy of the artist

to which colour combinations can be considered as linked more closely to the concept of flowers than the flower shapes themselves. Yet in their 2020 work *Mou t an*, it is the contour of the mountain that enables it to be recognised as such, as the artists play with erasing parts of an image to test the recognition limits. Whilst the Korean artist duo achieve this by removing pixels, Caroline Sinders applies white acrylic paint to photographs of an oil rig in their *Experiments in Becoming Unreadable* (2024), in which the artist attempts to hide these now dated infrastructure objects from machine vision, even though in reality their presence cannot be as easily obscured.

Beyond objects and faces, recognition extends to emotions and behaviour. Coralie Vogelaar's *Random String of Emotions* (2018) looks at an actor's ability to produce emotions based on instructions made by AI, while Lauren Lee McCarthy's *Unlearning Language* (2022) pushed participants to find new ways of communication not detected by the algorithm, resulting in humming, clapping, and other noises instead of conventional human speech. Such works underscore how these algorithmic constraints can spark new modes of expression and communication.

These also recall the importance of categorisation, a process during which images are annotated and categorised, a famously subjective task that has produced a lot of bias. Artists who intervene in this space often act as both critics and translators, revealing how opaque technical systems can reinforce social hierarchies while also offering strategies for resistance, humour, or subversion in everyday life. Sebastian Schmieg's *Decision Space* (2016), commissioned for The Photographers' Gallery, offered exactly this opportunity to the casual observer: to categorise almost 3,000 images from famous photographers – some neutral like sofas, others with clearer negative connotations like guns – into one of problem, solution, past or future. This highlighted the challenges faced when categorising objects into abstract categories: there is often no clear answer and the selection is highly biased. The elements of bias and the vital role that arbitrary categorisation plays in our lives are surfaced by works such as Libby Heaney's *Britbot* (2018) and mots' *Deutsch/ Nicht Deutsch* (2025). These projects expose how historical or sports knowledge or even facial profiling determine belonging



Gretchen Andrew, *Vision Boards: Contemporary Art Auction Record*, 2021.
Courtesy of the artist

to a particular group, something that does not necessarily have a commonly accepted definition.

Recommendation algorithms have long powered our internet experience, including search engines, shopping portals, and music platforms. In Gretchen Andrew's *Vision Boards* (2017–2021), we see the artist subvert the traditional search engine recommendation algorithms by creating search-engine-optimised pages for her analogue paintings, which are then shown at the top of the results page for images when the user searches for such terms as "contemporary art auction record". This work seeks to underscore the power of the individual in hacking large corporate systems, presenting an alternative version of the world. Scott Kelly and Ben Polkinghorne's *Signs of the Times* (2017) presents billboards with images that recommend other natural national parks to visit, as if the attendee is shopping on Amazon, instead of enjoying the downtime in nature. Through this work, we are reminded of the extent of the influence of recommendation systems on our daily lives, which dictate what we buy, where we go, and what we do. This theme of recommendation engines is further explored in Sebastian Schmieg & Silvio Lorusso's *Networked Optimization* (2013) project, which presents copies of popular self-help books with white-on-white text: the only sentences visible are those that have been highlighted the most on Kindle, materialising the logic of popularity metrics.

At stake in artworks engaging with recommendation algorithms is not only the aesthetics of personalisation but also the politics of algorithmic governance. Systems that shape our choices in shopping, streaming, and dating also influence cultural tastes, narrowing the range of what we encounter. Some scholars, such as Shoshana Zuboff, have described this as "surveillance capitalism", where behavioural data is harvested to predict and shape future actions. By parodying or subverting recommendation systems, artists remind us that these invisible infrastructures carry profound cultural consequences, subtly determining what kinds of art, information, or relationships enter our lives. The latter is explored in Dries Depoorter's *Browser Dating* (2025) project, which is a dating app that connects people based on their browsing history. What these works suggest is that cultural experience itself is being algorithmically managed:

just as Spotify playlists or Netflix suggestions channel us into particular moods, genres, and rhythms of consumption, artworks that mimic or parody recommendation systems underline how even our aesthetic encounters are subject to invisible curation. Through staging these mechanisms online or in a public space, artists restore visibility to infrastructures that usually remain hidden, inviting reflection on how desire and preference are no longer entirely our own but shaped by predictive systems.

Scientific tools not originally intended for art-making provide another fertile ground for exploration. Event cameras, in which the imaging sensor only responds to local changes in brightness, have been appropriated by Justin Urbach and his team in *BLINDHÆD* (2025) to explore the human drive for eyesight optimisation: this new scientific advancement presents an artistic interpretation of enhanced visual perception. In *xSublimatio* (2022), Pierre Pauze generated models of drugs and molecules from data made available by AlphaFold, which was originally designed by researchers at Google DeepMind to predict protein structures from amino acids. Pauze's work sees collectors purchase NFTs of visual depictions of molecules, which, when combined, generate a new image – a new drug. The same interest in the latest technological developments and their cultural impact in the world can also be seen at a conceptual level in the works of artists such as Lawrence Lek, whose film *Geomancer* (2017) – set in 2065 during the year of Singapore's centennial – imagines the art longings of a satellite and draws on DeepMind's key breakthrough of 2017: the development of AlphaGo, a system able to beat human players in the ancient game of Go.

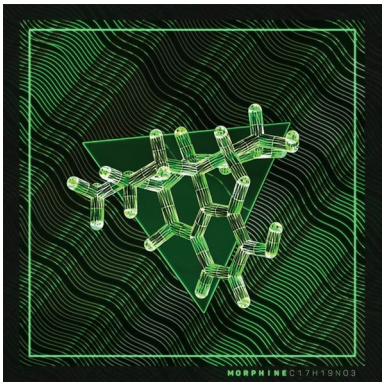
Apart from relying on technologies and concepts offered by the latest technological developments, a certain subset of works explores and explains the hidden processes involved in AI. Lida Zacharopoulou's *DIGITAL NO MAN'S* (2023), for example, uses the medium of film to anthropomorphise the low-res images produced by a generator and rejected by the discriminator that are part of the GAN system. This creative interpretation of the GAN process sheds light on the inner workings of the system by exploring the potential feelings and issues experienced by images deemed not pixel-perfect enough. Meanwhile, Sihwa Park's *Diffusion TV* (2025)

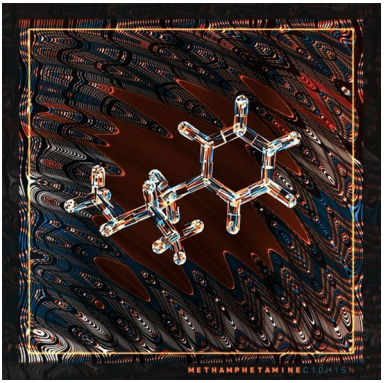
presents a nostalgic physical CRT TV with an interactive interface, enabling audiences to play with diffusion models and illustrate the hidden process of AI denoising by adjusting the TV's antenna and knobs to change chaotic data into recognisable images, giving the human hand a chance to facilitate the AI process.

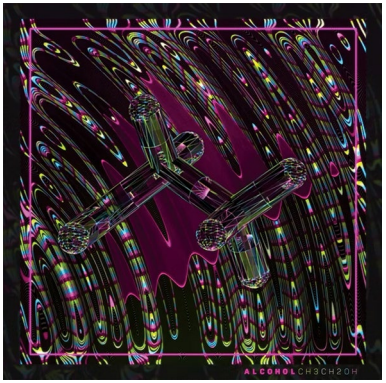
These days, AI agents are a major area of research, valued for their ability to achieve goals and complete tasks. In art, this is exemplified in decentralised autonomous art projects such as Gene Kogan's *Abraham* (2019–present) and Mario Klingemann's *Botto* (2021–present), an art agent that generates images based on prompts, refining its style through community votes. This use of autonomous agents in art raises questions about authorship and collaboration. When *Botto* produces images voted on by the community, who is the "artist" – the art engine, its developer, or the collective audience guiding *Botto's* development? These questions echo earlier experiments with distributed creativity, such as mail art networks in the 1960s or early internet art communities, where individual authorship was blurred in favour of collective processes. Through the revival of these debates in the context of AI, agent-based art highlights creativity as a networked, negotiated activity rather than the output of a single individual.

In response to digital saturation, some artists have turned to bioart. A key example from this field is Guy Ben-Ary's *cellF* (2016), the world's first neural synthesiser made out of a biological neural network grown from the artist's own cells. Unlike software-based neural networks, Ben-Ary's project uses a human cell to control a custom-built modular synthesiser: only analogue circuits and biological material power the instrument. These build on older bioart works such as Eduardo Kac's *GFP Bunny* (2000), showcasing how living systems infiltrate into the domain of art and, in the process, blur distinctions between natural and artificial, human and non-human, and machine and organism. They also invite ethical reflection: what responsibilities do artists have when working with living matter, whether neurons in a dish or robotic creatures released into the environment? The combination of AI with bioart raises further questions about the boundaries of agency, creativity, and life itself.

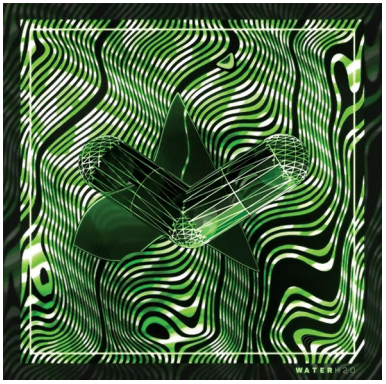








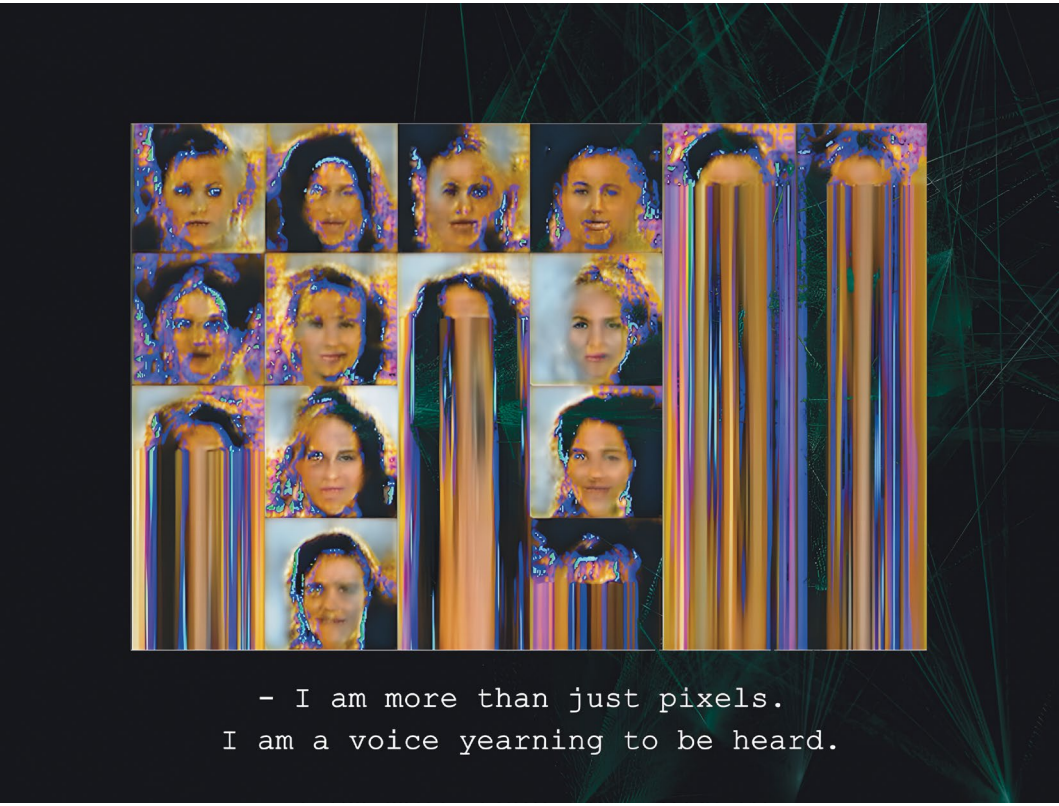




Pierre Pauze, *xSublimatio,* 2022. Courtesy of the artist

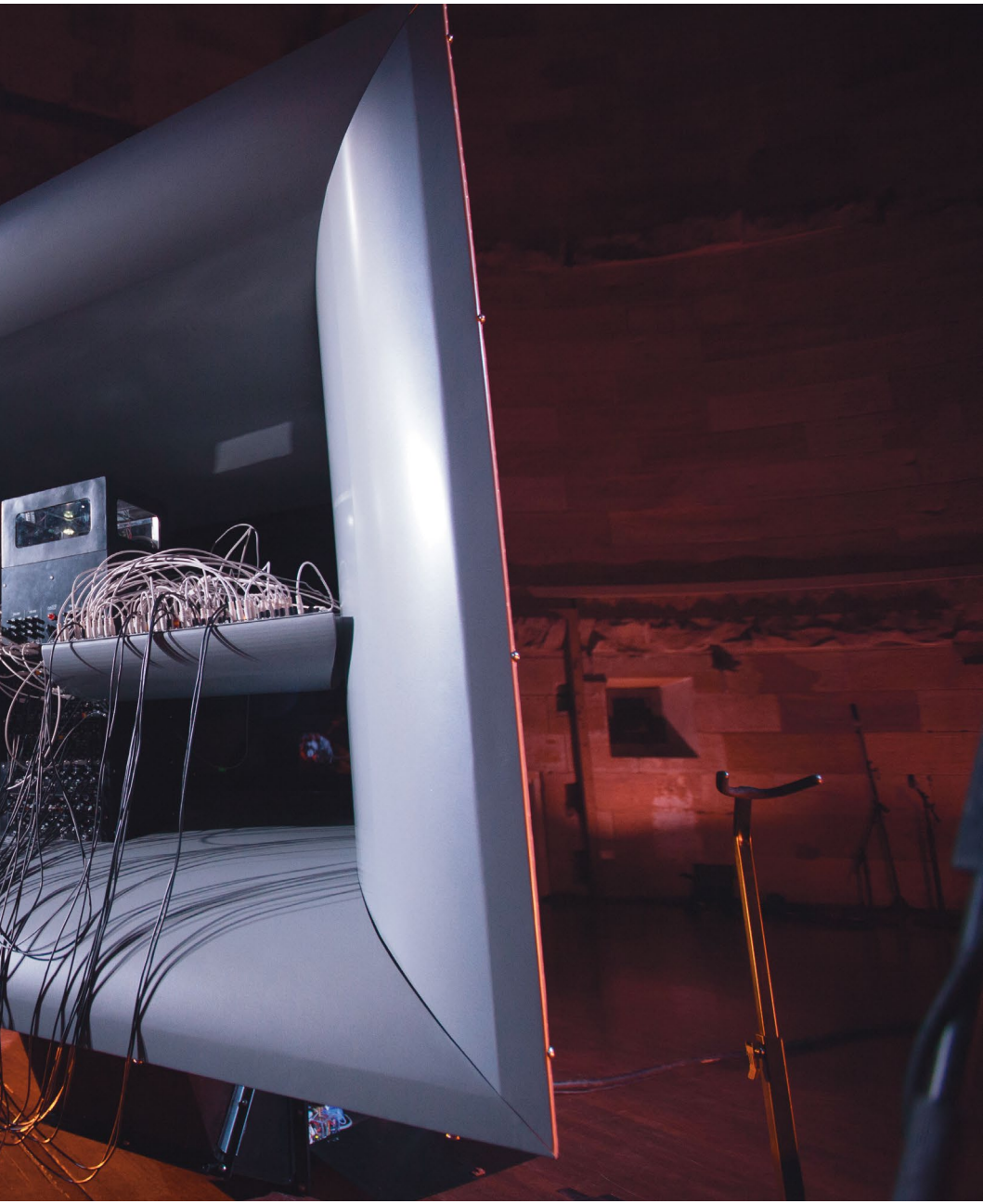
Alongside bioart, robotics offers a welcome escape into the physical realm. Ian Ingram's *Danger, Squirrel Nutkin!* (2009) features a robot with multiple squirrel tails that serves to warn real squirrels of any predators it sees using their own tail-flick signal. This interaction between robotic creatures and the natural environment raises multiple questions about the essence of a squirrel and the role these artificial creatures may play in the world of the future. Similarly, Paula Gaetano developed a land exploration robot inspired by the guanaco, the native mammal of the Andes, for the project *Guanaquerx* (2024). Made by a collective of artists and engineers, this robot proceeded to traverse peaks of 15,000 feet across the Andes, a feat inspired by the 1817 Crossing, providing an example of collective efforts of local cosmotechnics to reclaim the Andes as a site of emancipation and decolonialism. These explorations intersect with a growing interest in ecological and posthuman perspectives in art. In a time of climate crisis, such works encourage us to imagine forms of coexistence with machines, animals, and environments that are less extractive and more collaborative, repositioning AI and robotics not as tools of domination but as partners in ecological survival.

In terms of artificial life, techniques such as genetic algorithms and cellular automata have long found uses amongst a broad spectrum of artists from the fields of old-school computer art to the NFT generative art community. These are biological algorithms developed for other purposes, but finding life as creative material by artists such as Alexander Mordvintsev and Harm van den Dorpel. Alexander Mordvintsev's *Hexells* (2020) operates as a self-organising system of cells trained to make textures by relying solely on neighbour communication. This work highlights the role of an individual cell within a larger cell collective. Harm van den Dorpel has been working with systems such as genetic algorithms in *Hybrid Vigor* (2017) for a while. There, we see the artist use the fitness function of audience feedback to create new abstract turtle drawings, which populate the screen and allow the audience to choose two of these as parents, which in turn spawn children, and the process continues until the shapes and colours converge into one supreme artwork, which is then printed.



Lida Zacharopoulou, *Digital No-Man's,* 2023. Courtesy of the artist

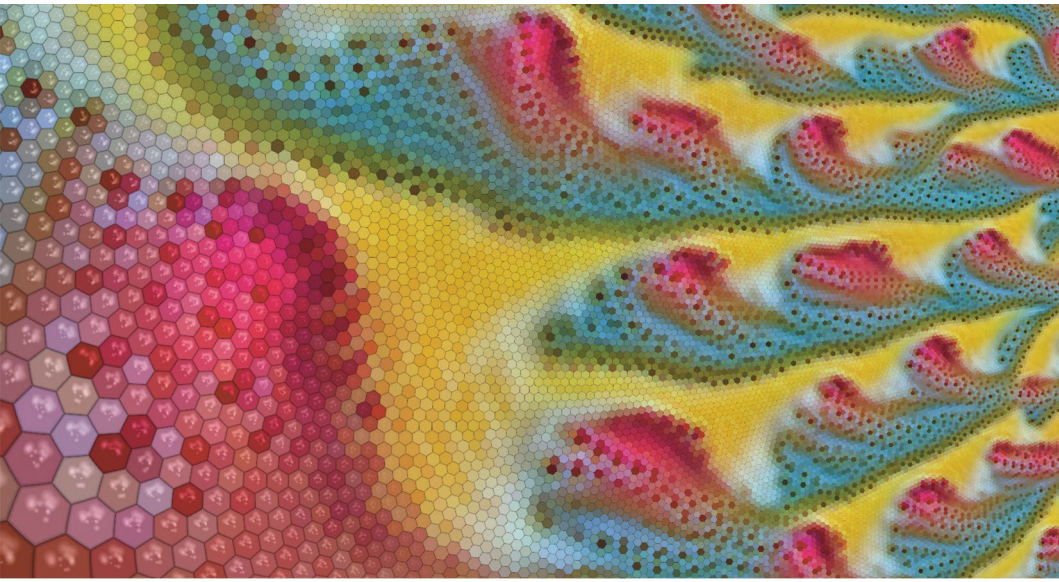




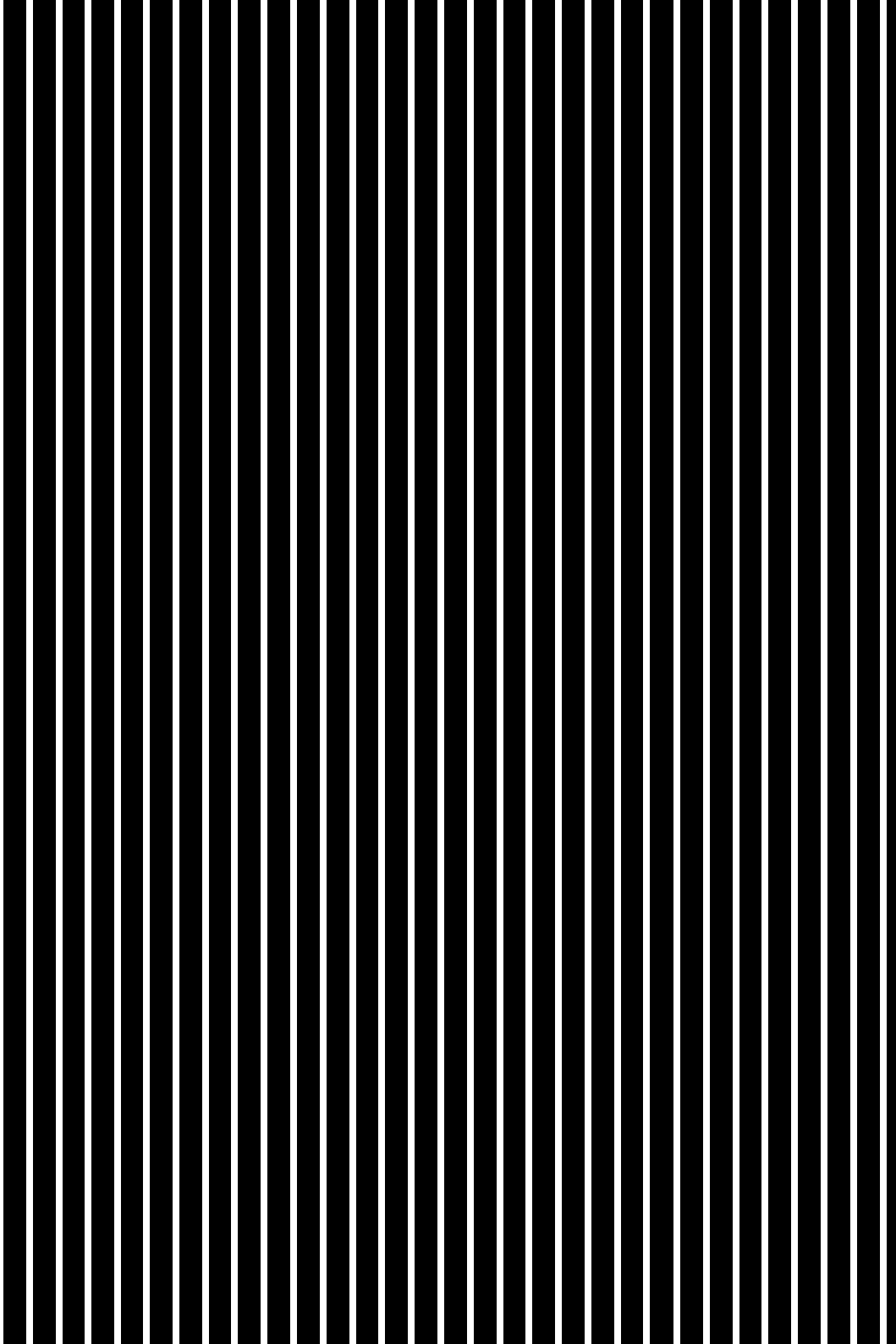
Guy Ben-Ary, *cellF*, 2016. Installation view, Cell Block Theatre, Sydney.
Courtesy of the artist. Photo by Alex Davies

Together, these projects demonstrate how artists repurpose AI systems in unconventional ways: hacking infrastructures, exploring the limits of technology, revealing hidden assumptions, incorporating living matter or breaking apart the building blocks of these all-purpose AI systems. In doing so, artists open themselves up to less competition and more opportunity when they begin to experiment with systems not initially designed for media generation. After all, it is easier to foster unusual forms of creativity when the tool you are working with is limited in conventional fine arts capabilities. These limitations turn constraints into opportunities for new forms of expression and perspectives that resist technological homogenisation.

Looking ahead, the challenge for artists is to ensure that their distinct creative voices continue to be heard in the tsunami of ever-perfect AI slop. Ultimately, the future of AI art may depend less on technological novelty and more on the capacity of artists to imagine alternative uses for infrastructures of computation. As AI becomes increasingly embedded and invisible – powering everything from medical diagnoses to customer support – the artistic task is to make these hidden processes visible and offer their unique voices through playful interventions, critical hacking or collaborations with living systems. In this way, artists can reclaim some agency in a technological landscape dominated by corporate platforms, reminding us that creativity is not only about producing new images but also about reshaping the conditions of perception, interaction, and meaning.



Alex Mordvintsev, *Hexells,* 2020. Courtesy of the artist



The Double-edged Blade of Technological Appropriation

Rosemary Lee
Artist; Guest Assistant Professor, Faculty of Engineering,
University of Porto

The ethical mutability of technology gives it the capacity to act as a double-edged blade that may just as easily cut for as against the intentions of those who wield it. Critically engaged artistic practices with emerging technologies are no exception to this, equally capable of subversion as of becoming trapped in the apparatus themselves. The reverberating effects of technology are not easily understood through black-and-white conceptions of technoethics, nor do they offer straightforward paths toward resolution that are not addressed by merely drawing attention to or seeking to remediate the potential repercussions of technology.

Creative explorations with artificial intelligence have recently raised a number of complex issues that touch on the nuances of this topic, while recalling aspects of similar shifts that have occurred in the past, in which technologies have enacted unintended consequences in the fulfilment of their aims. Efforts of creative misuse may steer tools towards other purposes than they were intended for, opening up new ways of looking at them and the grounding assumptions embedded in their design. But these disruptive gestures are themselves subject to the very same duplicity that makes them possible. The appropriation of new tools and methods may destabilise established conventions within culture, informing new methods, aesthetics, and conceptual reflections on technological shifts and their wider impacts within society. In recent explorations of artificial intelligence in art, neither a blanket rejection of the technology nor an attempt at taming it will easily reconcile its internal conflicts.

Efforts to understand and respond to the emerging promise of artificial intelligence have unearthed a myriad of interconnected issues that stem from various aspects of the technology's relevance to artistic practice. By unsettling the methods, tools, and value systems of art, new technological developments undermine adherence to strictly established conventions regarding the role of technology in culture. Like a double-edged blade, artificial intelligence cuts both ways, acting both for and against its explicit purpose. This is a core danger of artificial intelligence – and, arguably, of technologies in general – but is also an opening for critical intervention.

> There is no weapon of your individual will which, once appropriated by others, does not turn against you. (Vaneigem, 1967: 9)

Situationist author Raoul Vaneigem noted the precarious nature of the tools of resistance, remarking on their equal capacity to be appropriated as to be reappropriated. The weapons referred to in this case relate to acts of political and artistic opposition in Europe during the late 1960s, long before artificial intelligence might have been conceptualised in such terms, yet this statement may benefit from the very logic of appropriation that it identifies. Applying these words otherwise than intended may give us insight into the unsteady nature of technology's adoption within culture, highlighting potential ways forward regarding the relevance of artificial intelligence to art. Artistic interventions with technology are immersed in an interplay between creative misuse, abstention, and subversion, often seeking to contest the current state of affairs through the invention of new methods for doing, making, and thinking. But while these may indeed prove to be effective strategies that enable artists to respond to the shifting contexts they find themselves in, Vaneigem's words remind us that even these are subject to the turning (*détournement*)[1] of apparatuses against themselves. While technology may not be impervious to our acts of sabotage, those very disruptive strategies may themselves be undercut by other intentions.

The recent fascination of artists with artificial intelligence raises familiar questions about technology's appropriation in art, and this is particularly relevant in the ways that it may trouble what we thought we already knew about the subject. New, powerful generative

tools have opened up an intricate area of discussion on the role of technology in culture, as well as what ethical responsibility artists bear in experimenting with tools that may carry with them a host of implications that are problematic and difficult – or even impossible – to remediate. Artificial intelligence's artistic potential, resulting aesthetics, and accompanying discourses have a habit of dredging up unanswered questions on seemingly resolved topics such as the role of machines relative to human authorship, the primacy of either perceptual or conceptual qualities in cultural content, and conceptions of mastery within these frameworks. But they may also give us cause to re-examine the assumption that artists are the ideal candidates or even necessarily equipped to investigate topics such as these in a way that is not itself ethically contentious. Arguably, the task of art is not to fix the world but rather to offer new ways of forming relationships with it, but this is a complex undertaking given the seemingly ever-changing technological context we find ourselves in.

In recent years, there has been a significant expansion of artificial intelligence, both in terms of its technical capacities and its increasing presence in contexts beyond research, making it an undeniable source of intrigue. The expansion of access to experimentation with machine learning systems has placed it and its related themes within reach of non-experts, introducing not only new tools and methods into visual culture, but also unleashing a great deal of uncertainty in its wake. Widespread access to the use of advanced artificial intelligence systems – both in the technical sense and in terms of literacy with its modalities – has caused a shift not only in what and how it's possible to create, but also in terms of culturally reckoning with the implications that follow this change. Artists have quickly tuned in to the problematic implications that artificial intelligence may have for society, culture, and creative work, among others. But as much as staying with this trouble has helped to illuminate its nuances, it also highlights the lack of any strictly virtuous paths forward from them.

It has been thoroughly established, for example, that artificial intelligence entails significant problems in terms of bias, error, and imposition of problematic power hierarchies. And while artists'

reflections on and with the troubling dimensions of artificial intelligence have arguably served as fodder for guiding discussions on these issues, there seems to be a growing sense that what is at stake here exceeds far beyond the capacity of art and artists to steer critical reflections on technology. Artificial intelligence has the capacity to be harmful in ways that are not straightforward to resolve, if at all, and while this is relevant material for artistic explorations, it's not always evident what role art might play therein. Something that is often taken as a given but merits further consideration is whether artistic work to deconstruct, critique, and comprehend artificial intelligence will necessarily result in a positive effect relative to the goals of this work. Artistic adoption of artificial intelligence may bring forth new understandings of its humanistic entanglements, but it also has the capacity to participate in its whitewashing and endorsement, even in its articulation of critique.

The domestication of artificial intelligence in cultural contexts may mark a new stage in industrialisation that requires alternative ways of conceptualising art, its relationship with technology, and their wider importance. What had previously been inaccessible technological modes of production have now become widely available, enabling an increasing level of popular literacy with what is actually entailed in artificial intelligence beyond the vivid associations encouraged by this term. As a result, we have become more aware of how artificial intelligence functions, as well as its limitations and capacity to enact harm. But in spite of the considerable amount of attention this topic has received in the past decade, we do not appear to have made significant progress in resolving the numerous conceptual and ethical discrepancies it raises. There is a high degree of public awareness of the fact that artificial intelligence is prone to error, bias, and risk, yet in spite of its perils, as well as precisely due to them, there is a consensus that artificial intelligence is not to be ignored.

Some insight may be gained from enquiring into the term artificial intelligence itself as a strategy of "AI-anti-hype" (Odubela, 2024) to dispel some of its mystique. Artificial intelligence frequently inspires comparison with artificial general intelligence, a vision of artificial intelligence capable of meeting or exceeding human abilities at performing a given task. In popular discourse, it is often linked to

the idea of automated thinking machines that could easily remove the inefficient human in the loop if that were deemed possible or even beneficial to their efficiency. Whether this threat is leveraged in terms of labour – through the displacement and replacement of human roles – or by way of extermination – for which artificial intelligence could conceivably contrive many creative solutions – the potential danger posed by artificial intelligence may serve the dual purpose of stoking fear of the technology while also inspiring reverence for it by overestimating the power of current artificial intelligence. Counterintuitive as it may seem, the vision of artificial intelligence associated with a threat to humanity may divert warranted attention away from the demonstrated harm to humans that artificial intelligence is already implicated in, as well as the fact that what we currently refer to under the title of "artificial intelligence" bears little resemblance to machines' performance of human-competitive cognitive activity.

Describing artificial intelligence as such, a matter of mechanising intellectual work, may itself be a problematic foundation upon which to build technology. Dissimulating computational processes as relative to human cognition through the use of anthropomorphic terminology may say more than it intends to by exposing that this act of doublespeak, in fact, means precisely to render intelligence interchangeable with the artificial. Giving insight into this in his social history of artificial intelligence, Matteo Pasquinelli discusses how the relationship between computers and the performance of cognitive tasks has been portrayed in less than straightforward fashion:

> The idea of the automatic computer, in the contemporary sense, emerged out of the project to mechanise the mental labour of clerks rather than the old alchemic dream of building thinking automata – although the latter narrative would often be used, in the nineteenth century as much as in the century of corporate AI, to masquerade the former business. (2023: 52)

This dimension of artificial intelligence, as a term again, taps into the technology's capacity to be Janus-faced, presenting opposing perspectives at once. While it may be misleading to propose artificial intelligence agents as straightforward replacements of human workers, there is nevertheless weight to the fact that this has

been and may continue to be an underlying intention behind their development and deployment. This term also harbours questionable assumptions about the nature of intelligence – and, most problematically, its associated identity politics – as well as the deferral of decision-making to anthropomorphised machines and the enactment of a divisive hierarchical programme upon the world. Despite these incriminations, attempting to disassemble its internal logic through experimentation with alternative terminologies may make the mistake of seeking to formulate a critique of artificial intelligence while accepting its foundational premises. For example, the renaming of artificial intelligence using a term such as *synthetic cognition* (Manouach & Engelhardt, 2022) may amount to merely calling it by a pair of synonyms that cloak the fact that what's underneath remains the same. While this term entails several discrepancies and internal conflicts, it is nevertheless extremely effective in its capacity to enthral us, conjuring vivid visions of what artificial intelligence could entail – for better or worse.

For various reasons, artists have been some of the most visible early adopters of artificial intelligence and among its most outspoken critics. Part of this draws upon the conception of culture as a panacea for societal ailments, especially when framed in opposition to technological progress. Whether this conception of the role of culture, art, or artists is accurate or not, the topic and affordances of artificial intelligence have proven to be a compelling area for creative exploration. Artists have played an important role in considering artificial intelligence's implications for not only art, aesthetics, and culture, but also its wider dimensions for society and the environment. Many have voiced concerns about artificial intelligence's disposition to cause harm, pointing to its track record of wrongdoing in terms of resource-intensive extractivism, algorithmic bias, and capacity to enact potentially catastrophic damage in ways that are difficult to anticipate. While these issues and the tools they emerge from are fertile ground for artists to scrutinise, exploring these issues artistically treads a blade's edge between critical appropriation of technology and its cultural validation.

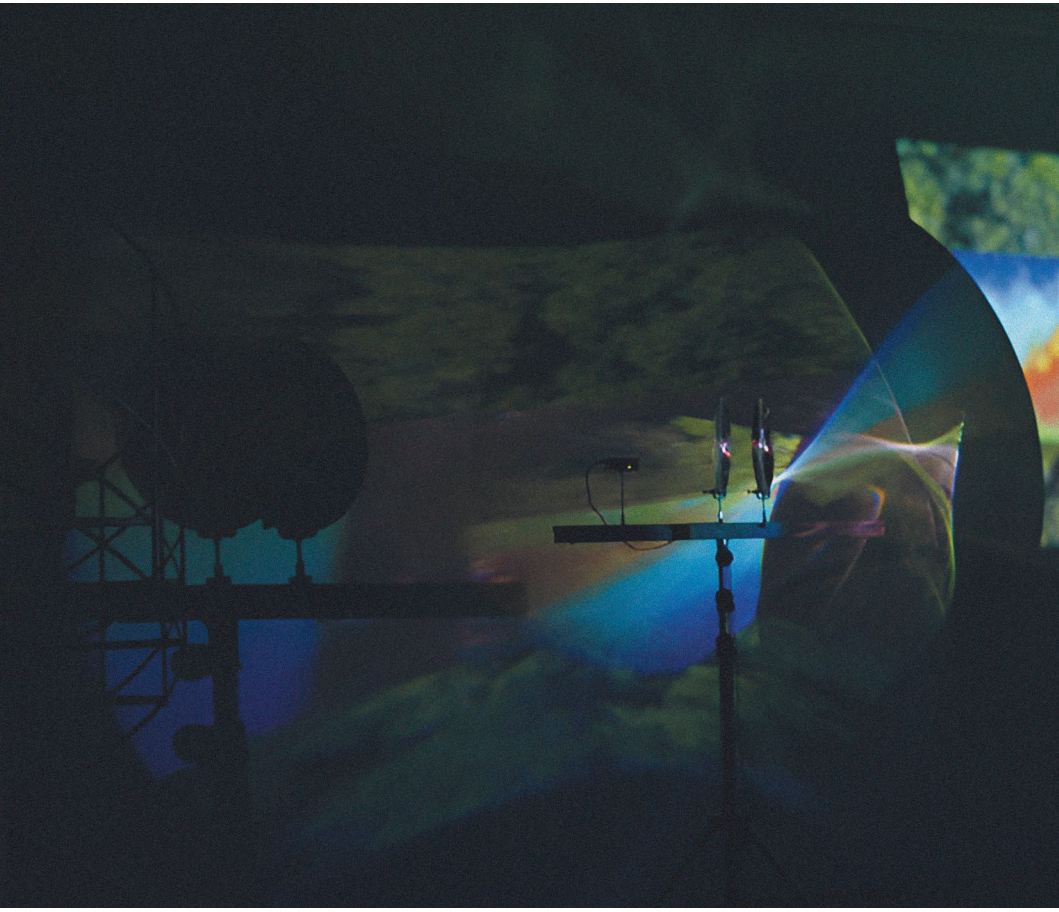
Seeking to challenge technology by way of technology is never a straightforward matter. Meaningfully addressing the impact of

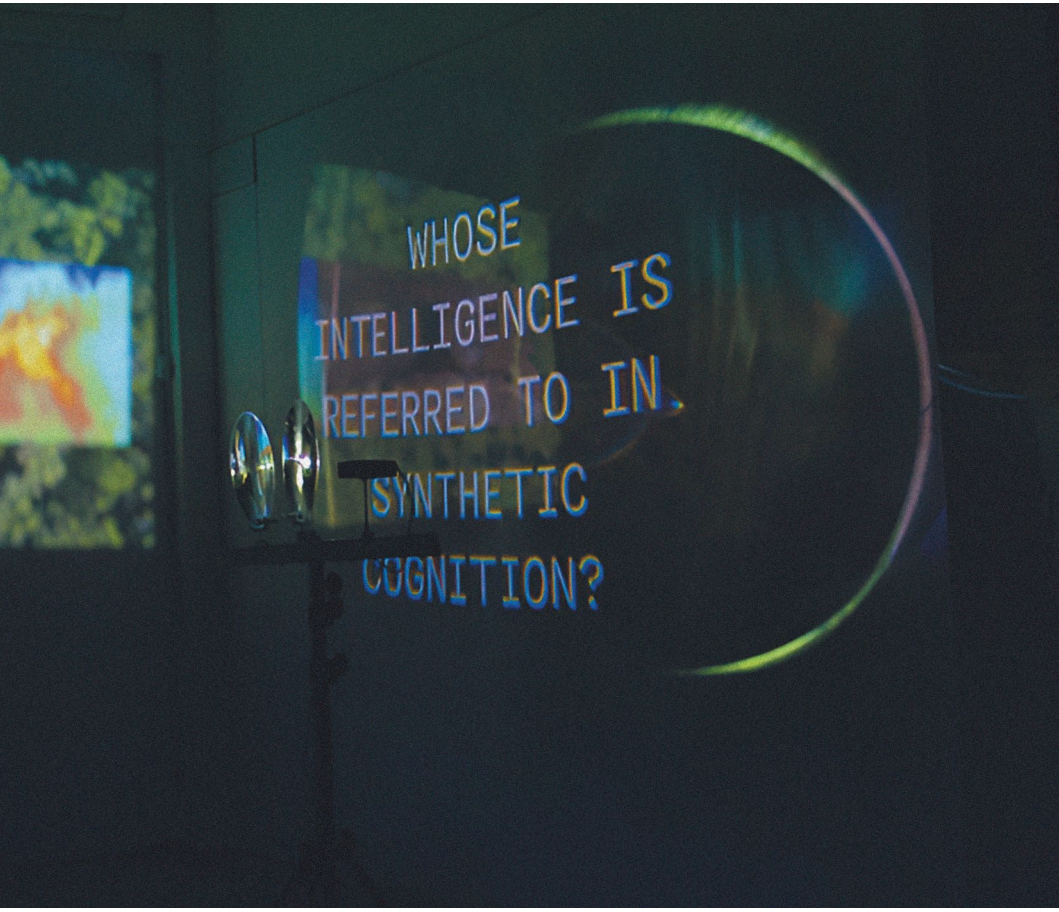
technologies is never as simple as making an informed choice about whether or not a given tool is "good" or "bad". In engaging critically with technology, artists are not merely faced with a decision of whether or not to use a particular kind of gadget, method, or software, but are instead tasked with probing and reflecting on its impacts, significance, and aesthetics. This may include breaking or deconstructing the tools as a way of seeking to understand them, or – just as often – to articulate a critique of their nuances.

One of the most memorable parables on the social rejection of technology, the Luddite rebellion, may act as a reminder of the multifaceted nature of technology and its relationship with historical narratives in culture. By destroying the machines that threatened to displace the value of their labour, the Luddites' response to the effects of industrialisation earned them a designation that has endured for more than a century, though they may not have been satisfied with the way their actions have been remembered. The term often resurfaces in discussions of the impact of emerging technologies, and it is now associated with a stance antithetical to technology – against progress, even. The nuances of rejecting the machinery of industrial production appear to have been eroded by time, misconstrued by those invested in a particular view of the situation and oversimplified over the course of numerous retellings.

> The word "Luddite" […] turns up in our daily language in ways that suggest we're confused not just about technology, but also about who the original Luddites were and what being a modern one actually means. (Conniff, 2011)

Although the machines were a target of their ire, the core of the matter was the conditions of labour, and to explain it as merely an anti-technological movement would be to not only misapprehend its significance, but also disregard the cautionary tale it offers. The specifics of the Luddite movement itself are of less interest here than how this cultural memory relates to our present struggles to come to terms with artificial intelligence and its impact – concrete, potential, artistic, and societal. Current responses to artificial intelligence may themselves spawn a new label for those who reject the technology on the basis of its reverberating consequences; we





Rosemary Lee, *A Structural Plan for Imitation: Engines of Differentiation*, 2024.
In collaboration with Alexia Achilleos. Two-channel video installation.
Installation view at NeMe Arts Centre, Limassol, Cyprus. Courtesy of the artist.
Photo credit: Nicos Avraamides

may only hope that the subtleties of this struggle may be captured in the often low-contrast play of shadows in what is transcribed as history. The danger we might take note of from the flattening of Luddites into two-dimensional characters attacking machines out of fear and ignorance is that this historical overturning of the narrative may be just as easily applied to undermine legitimate critique of technology as to its rejection for any other reason.

A contemporary Luddite rebellion could attempt to take a hammer to artificial intelligence systems, inventing machines to destroy the technology more efficiently[2], shutting down the power grid, or burning data centres to the ground. But as Kate Crawford and Vladan Joler (2018) demonstrate, the anatomy of an AI system is massively dispersed and integrated into varied material and immaterial infrastructures, so that rejecting or halting artificial intelligence on any significant scale would at this stage be virtually impossible. Even if it were feasible, simply smashing the machines would not suffice to snuff out artificial intelligence. This was already the case at the time of the Luddites' interventions, and the tools in question were mechanically simpler. The fact that the Luddites are remembered today not for contesting the conditions of their labour by way of technology, but rather, for attacking the technology, speaks to the challenges at stake in critical approaches to the appropriation of technology in art. The act of breaking the tools that play a role in changing technological conditions of work is an evocative image, but the problem has never been situated within the tools themselves, but rather in the dynamics of the system in which they play a part.

Although certain aims may shape and become embedded in the design and implementation of technologies, to fixate solely on these purposeful aspects would overlook a more crucial aspect: technology has equal capacity to act for and against its intended purpose. This makes artistic work with emerging technologies such as artificial intelligence an especially fascinating territory for cultural insurrection, playing a balancing act of cultivating fertile soil for novel forms of practice, aesthetics, and thinking, while seeking to avoid those efforts being hijacked for other aims. Bringing this variegated view of technology to the artistic exploration of artificial intelligence offers a departure from the tendency to frame this topic in high contrast,

an interplay of lights and darks – moral high ground, even – when the reality often leaves us with little certainty.

Creative experimentation with emerging technologies has the capacity to reveal subtleties that may not have otherwise been apparent, while evoking a sense of irresolution within technological appropriation in art – that a given technology may give a sense of being as unavoidable as its flaws are apparent. As tempting as it is to believe that our present concerns about technology stand as exceptions from the past, we may note resemblances within recent developments that provoke an eerie nostalgia for things we vaguely recall, but that come into existence through that very act of anamnesis. The expansion of new tools, methods, and issues available to artists has changed the conditions of creative labour as well as the values employed in assessing its outcomes. And while it may be possible for us to create in new ways using artificial intelligence, it's also crucial to recall that this shift entails elements reminiscent of the past. The integration of new modalities and concepts occurs within existing frameworks rather than marking a radical departure from previous forms of practice. Technologies often act as double-edged blades that act *on* us as much as we act *with* them. Recognising this capacity may enable us to resist the urge to frame the reverberating effects of artistic practices with artificial intelligence as entirely unprecedented. Unmooring technologies from their historical context obliges us to approach them as a blank slate, when this may erase vital nuance and inhibit our recourse to existing frameworks for addressing the impact of technologies.

Notes

[1] This term was used in Situationist contexts to refer to the hijacking of something against its intended purpose as a method to undermine or question it.
[2] Paraphrase of a meme, see: (matthat, 2010).

Bibliography

Conniff, R. (2011, March). What the Luddites really fought against. *Smithsonian*. https://www.smithsonianmag.com/history/what-the-luddites-really-fought-against-264412/

Crawford, K. (2021). *Atlas of AI.* Yale University Press.

Crawford, K., & Joler, V. (2018). *Anatomy of an AI system: The Amazon Echo as an anatomical map of human labor, data and planetary resources.* AI Now Institute and Share Lab. https://anatomyof.ai

Future of Life Institute. (2023, March 22). *Pause giant AI experiments: An open letter*. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Manouach, I., & Engelhardt, A. (Eds.). (2022). *Chimeras: Inventory of synthetic cognition.* Onassis Foundation.

matthat. (2010, August 10). Luddite invents machine to destroy technology quicker. *The Spoof!.* https://www.thespoof.com/spoof-news/science-technology/83952/luddite-invents-machine-to-destroy-technology-quicker

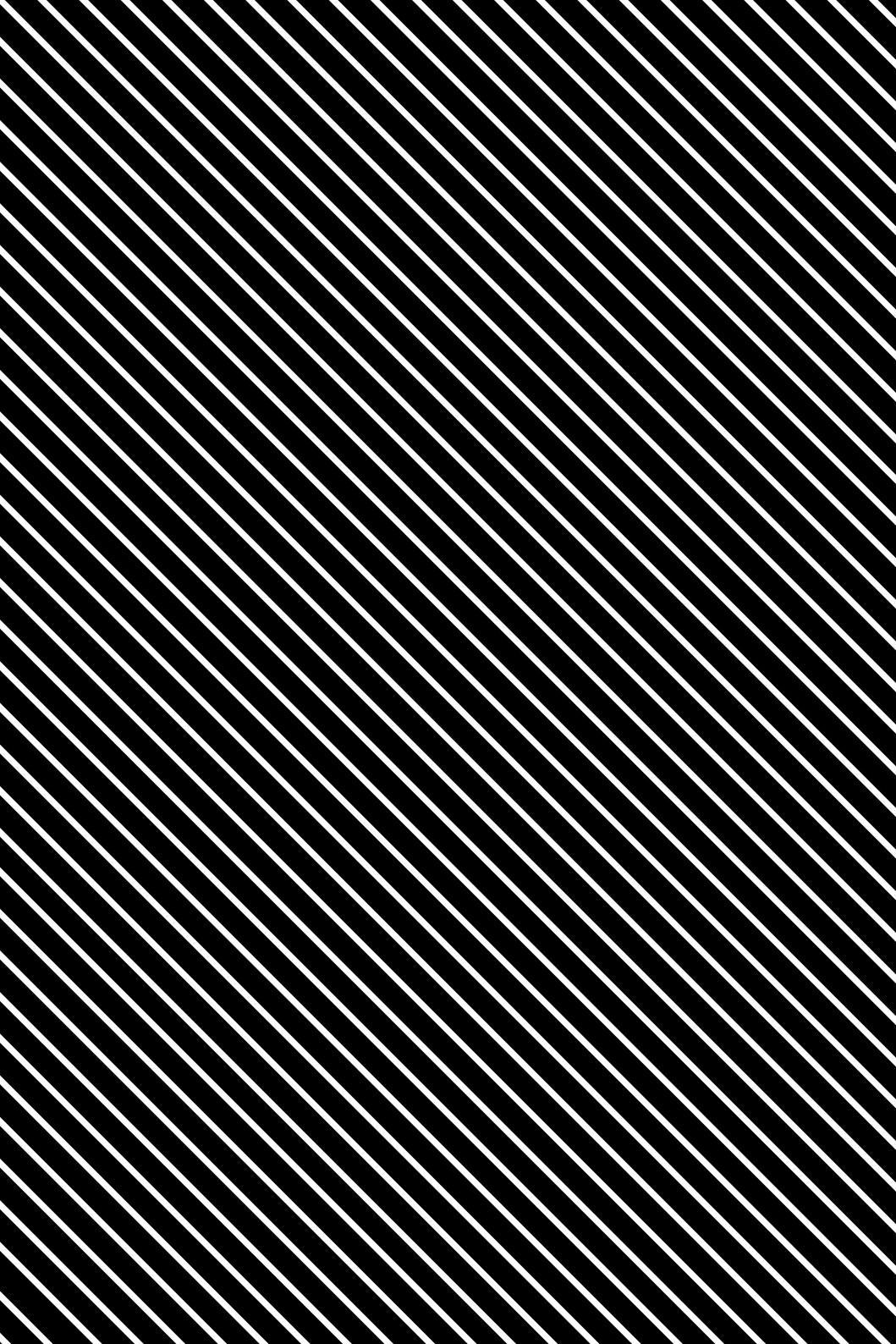
Odubela, A. (2024, November 15). V1.01 ~ A re-introduction to AI hype. *AI Anti-Hype.* https://aiantihype.beehiiv.com/p/a-reintroduction-to-hype

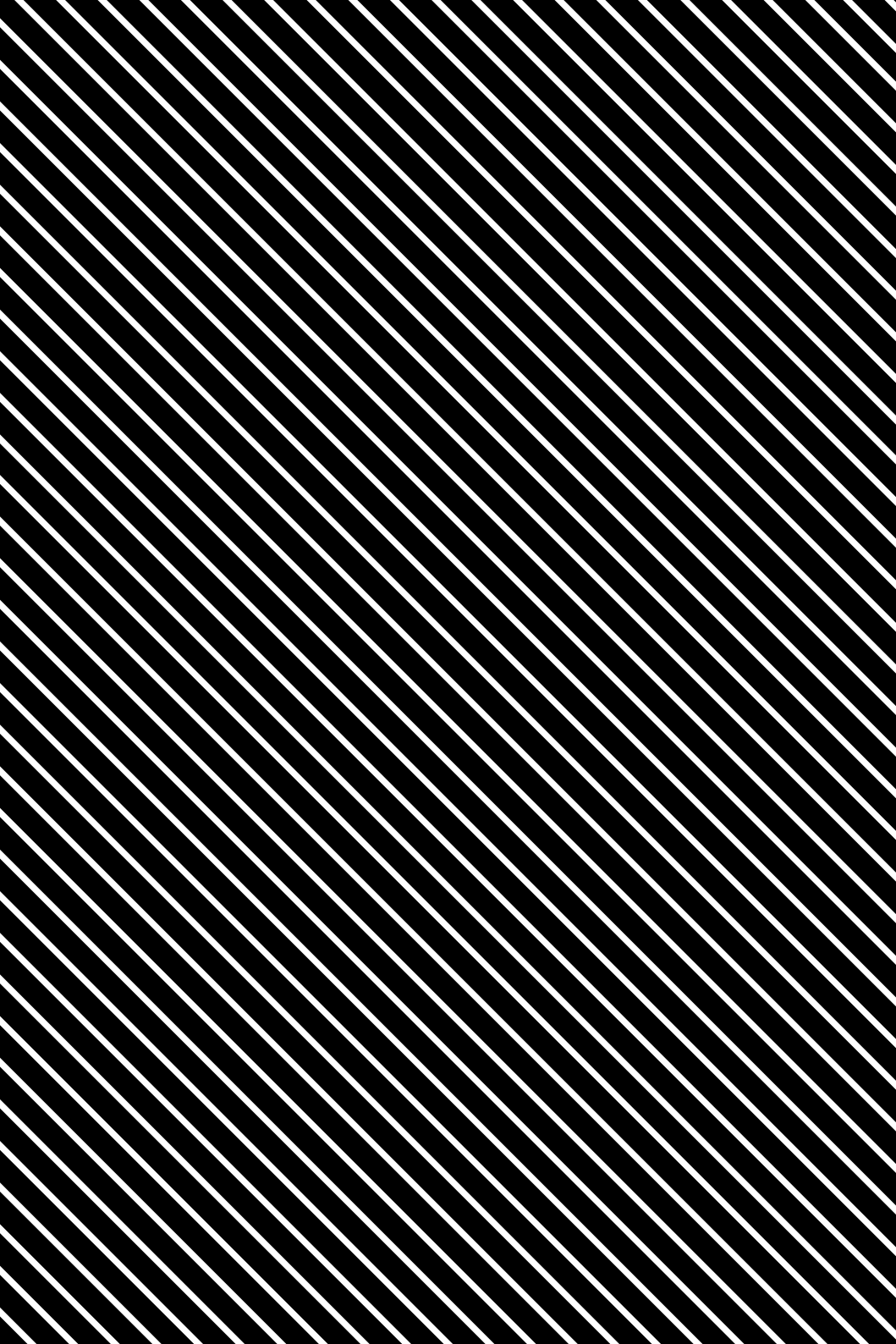
Parisi, L. (2021). Negative optics in vision machines. *AI & Society, 36,* 1281–1293. https://doi.org/10.1007/s00146-020-01096-7

Pasquinelli, M. (2023). *The eye of the master: A social history of artificial intelligence*. Verso.

The Editors of Encyclopaedia Britannica. (2025). Luddite. In *Encyclopaedia Britannica*. https://www.britannica.com/event/Luddite

Vaneigem, R. (2012). *The revolution of everyday life* (D. Nicholson-Smith, Trans.). PM Press. (Original work published 1967)





Beyond Creativity: Emerging Artificial Intelligence Models and Internet-aware Art

Kostis Stafylakis
Assistant Professor of Fine Arts, Expanded Media and Forms,
Department of Fine Arts, Athens School of Fine Arts

In this essay, I will attempt to examine the dialectic between contemporary visual arts practice and emerging models of artificial intelligence, testing a framework of understanding that doesn't draw its inspiration from the romantic concept of creativity, a concept established within the same semiotic field as the concepts of originality and ingenuity. And yet, the recent dissemination of generative artificial intelligence applications has given rise to a public discussion which, to a large degree, has been defined by anxiety over the absence of the creator's superiority, as that is summed up by the popular question: "Can artificial intelligence create works of art?" Of interest is the fact that the same anxiety also runs through discussions that are not, exclusively, of a popularising nature. My mistrust of this popular question extends towards the subsequent treatment of the relationship between contemporary art and artificial intelligence as a new chapter in the history of the media, or the "new media". That viewpoint commonly sees the expansion of recent technologies as an opportunity to reinforce or extend the "creative self", supposing the same centre of creative authenticity: a "modern self", who is facilitated by technologies, media or tools. Such a logic is reflected, for example, in Refik Anadol's popular art, which processes specific arrays of data to translate them into giant electronic kaleidoscopes in real time. That post-digital investment of the urban landscape and architectural design could, without much effort, become understood as a contemporary translation/extension of modern abstraction.

The reason I don't find the above reasoning convincing is that it tends to contribute to a perception which neglects that, for years now, a more meaningful dialectic between visual arts practices and artificial

intelligence has been unfolding – one that takes into account certain fundamental mechanisms of the social process. I am referring to a dialectic between art and internet culture, a dialectic between artistic strategy, visual culture, and the political and social parameters of life after Web 2.0.

The problem with the discussion regarding "artistic creation" in the age of artificial intelligence is that it usually bypasses the megastructures, the social and technological ecosystem within which that "meeting" takes place. That ecosystem is none other than the internet, in its contemporary transitional phase, from Web 2.0 to Web 2.5, and from there on to its hypothetical consequent forms. There, artificial intelligence meets us in an inescapable way, contributing to the configuration of an "algorithmic unconscious", which dominates over the subject without much consultation.[1] At the time of writing, the timeline of my social media has been taken over by artificial images of an erotic clinch between Timothée Chalamet and Tom Holland, as if trying to tell me something, to arouse – successfully or not – the desire to stay tuned to what it is about to "offer" me. Even if I find this material fairly indifferent, I cannot but accept that the way these generative models fetishise the figures of the two well-known actors does not go unnoticed. The "flawless" glow that envelops their bodies and their clothes lights up the various synapses of the retina and the psyche. This acceleration begins in 2016, when social media feeds start to define their content increasingly through algorithmic personalisation and less so chronologically.[2]

It has been stressed, from several directions, that contemporary technologies of machine learning were developed alongside the data mining industry and its central role in a condition that Louise Amoore described as "algorithmic warfare", a "war by other means", whereby "companies, foundations, governments fight daily to obtain and manage data for purposes connected to security, commerce, science or political propaganda".[3] The current "spring of artificial intelligence" is largely due to the political-economic transformation which is known as Web 2.0, that is, the conversion of the internet to a site for the extraction, accumulation, and processing of data and "big data"; the development of cloud technologies and services; the creation of large-scale data centres; the creation of large in-house servers by Big Tech; the unprecedented networking of computational

systems; and the exploitation of closed datasets from the realm of social media and online marketplaces. Even though many would argue otherwise, the premise of a meaningfully decentralised internet, a Web 3.0 based on blockchain-type technologies, has not been adequately implemented: instead of expanding the users' protected and encrypted nodes, the internet has been taken over by neural networks with popular and more easily accessible interfaces for the general public.

Web 2.0 (and what appears to come next) is, among other things, a data industry which demanded "image banks" such as ImageNet: repositories of material utilised by the current neural networks, GANs, Stable Diffusion's deep learning system and the various alternatives to text-to-image generation models within generative AI. The "anxiety" of these applications for geographical, operational, representational, and forensic accuracy is nothing but a product of the politico-economic assemblage that created them. As Joanna Zylinska explains, the "photographic realism" that these technologies seem to strive for is due, to a large extent, to the fact that "within the datasets of machine learning, photography seems to hold an 'ontological privilege' […] through which external objects are manifested as images […]. This belief applies not only to what may be defined as photographs on the first level (images of people, animals, landscapes, and inanimate objects taken for social, documentary or artistic purposes), but also to the photographic documentation of artefacts (paintings, sculptures) or photographic scans of flat objects (graphics, maps, texts)."[4]

A mapping of the relationship between art and emerging AI models needs, therefore, to take into account the relationship of both art and AI to the culture and industry of the internet itself. However, such an observation should not distract us from the tendency that seems to determine the present – the tendency of artificial intelligence to "swallow up" the internet, as Sue Halpern vividly wrote in *The New Yorker* regarding the year 2023.[5]

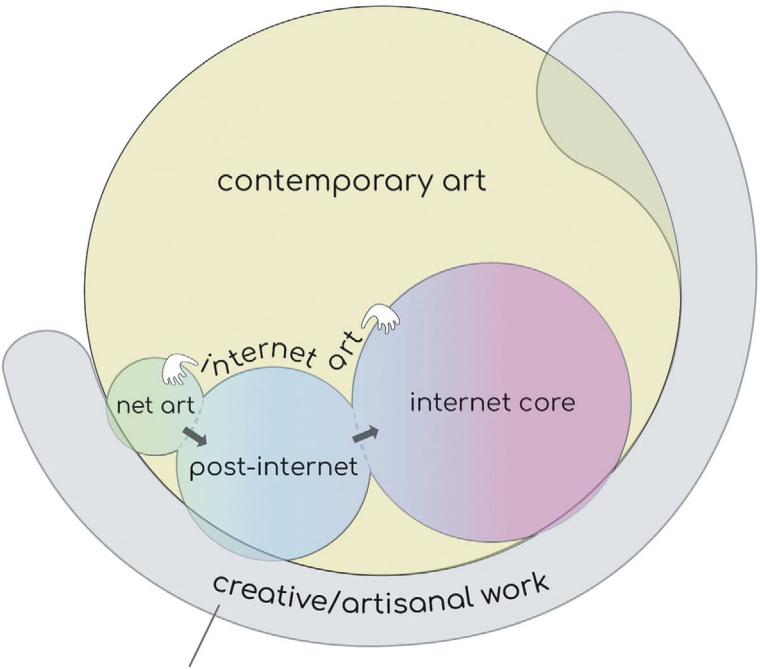
"Internet-aware" art and artificial intelligence

Rather than treating the concepts of "art" and "artificial intelligence" as positive contents whose significance is self-evident, it makes more sense to look at their entanglement within the recent history of

internet-aware art. There is absolutely no reason to invent such a periodisation from scratch. It is summarised exhaustively in an article by the visual artist Twee Whistler, uploaded onto Substack under the name donotresearch, i.e., the online research community of visual artist and author Joshua Citarella. The diagram that dominates the text regarding the different periods of internet art successfully summarises the clarifications that many artists and writers have attempted to articulate to date. In a manner that perhaps characterises the unstable and ephemeral character of such posts, Whistler's text has disappeared from the internet – at least at the time of writing – leaving behind only a useful screen grab and some ideas that a few people, such as this author, managed to discuss and disseminate in time.[6] A large part of Whistler's artistic work consists of fan art dedicated to the post-internet artist Jon Rafman, to such an extent that it leads you to suspect the presence of Rafman himself behind her online traces. In the surviving diagram, a large circle represents the entirety of "contemporary art". Within it, three smaller successive circles/sets represent "internet art", marking distinct moments and periods therein. Each of these periods corresponds to a different perception, approach, and management of the relationship between the aesthetic act, the internet, and data technologies. Instead of understanding these sets as final and definitive periods of internet-aware art, we might perhaps see them as aesthetic dispositions that can be renewed depending on the dynamics that surround them.

The entirety of internet art thus includes the subsets of net.art and post-internet art, as well as the subset of internet-core, a concept introduced by Whistler to describe certain aesthetic phenomena of the pandemic and post-pandemic world. According to the reasoning she develops, each of these periods corresponds to a distinct role of the "internet artist" and to a distinct perspective on the relationship between artistic practice and artificial intelligence. As I summarise Whistler's fruitful clarifications, I will therefore add thoughts and associations that reflect the moment of writing.

In the era of net.art, the artist is primarily a hacker or hacktivist. Although net.art corresponds to a particular network or art movement of the 1990s, characterised by what the critic Tilman Baumgärtel termed media specificity[7] – the exclusive use of the internet as a medium – the



Twee Whistler, *Internet Core Diagram,* 2023. Source: donotresearch

characteristics of net.art seem to reflect a broader spirit of online production that survives to this day: exclusively digital art that ranges from works of electronic mail to the production of (functional or non-functional) code and the provocation of unpredictable online events. Artists such as Jodi and 0100101110101101.org (a.k.a. Eva & Franco Mattes) are epitomes of the pirate culture of the coder, the hacker, the prankster, with the second successfully representing an expansion, beyond the specificity of the internet as a medium.

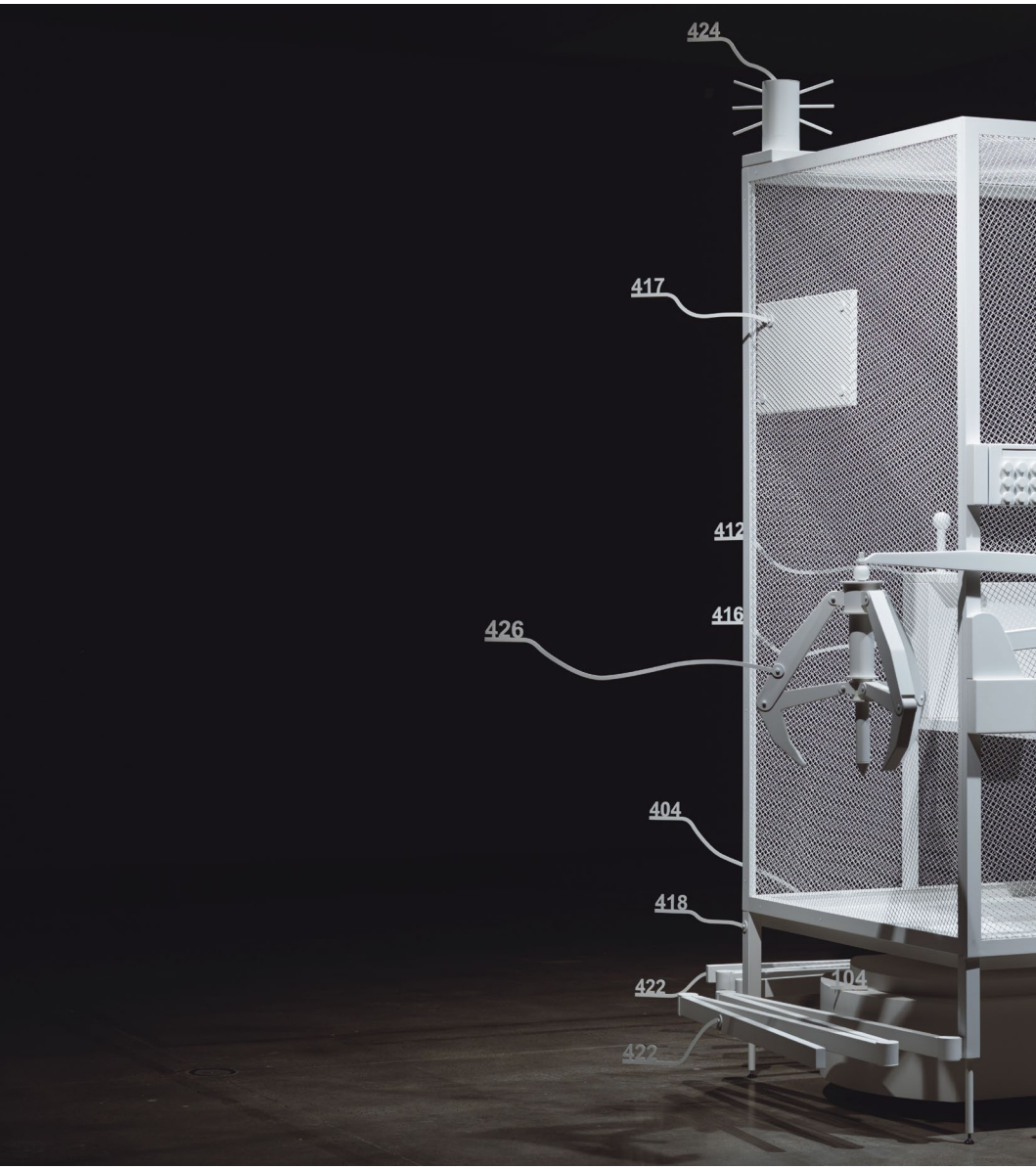
An artist who represents the spirit of net.art in today's hybrid environment is the Dutch artist Constant Dullaart. In his work *DullDream* (2015), he developed a neural network that executed the opposite operation of Google's DeepDream: rather than highlighting and intensifying the features and patterns of forms, it rendered them blurry, more indistinguishable, by removing, for example, the specific features that allow a face to be recognised as such. The aim was to undermine artificial intelligence as a tool of machine learning and recognition. Dullaart's model allowed the audience to upload clear images and receive them without their distinctive features. Another important artist, whose work strongly reflects the spirit of net.art, is Mark Amerika. His classic net.art project *GRAMMATRON* (1993–1997) was a cyberpunk literary text in the form of hypertext: a navigable space with hyperlinks leading to texts, soundtracks, moving and still images. His recent book My Life as an Artificial Creative Intelligence – somewhere between autobiography, media theory, and cyberpunk literature – further develops his experimentations with hypertext in an attempt to write "metafiction" in collaboration with mechanical forms of "intelligence" and "creativity", such as the applications of large language models (LLMs, ChatGPT 2 and 3). Although the works of Dullaart and Amerika seem to differ in their intentions and modes of implementation, both approach generative AI as a medium or tool for producing text, images, and aesthetic events intended to disrupt our perception of our technologically mediated experience, or to "retell" our story.

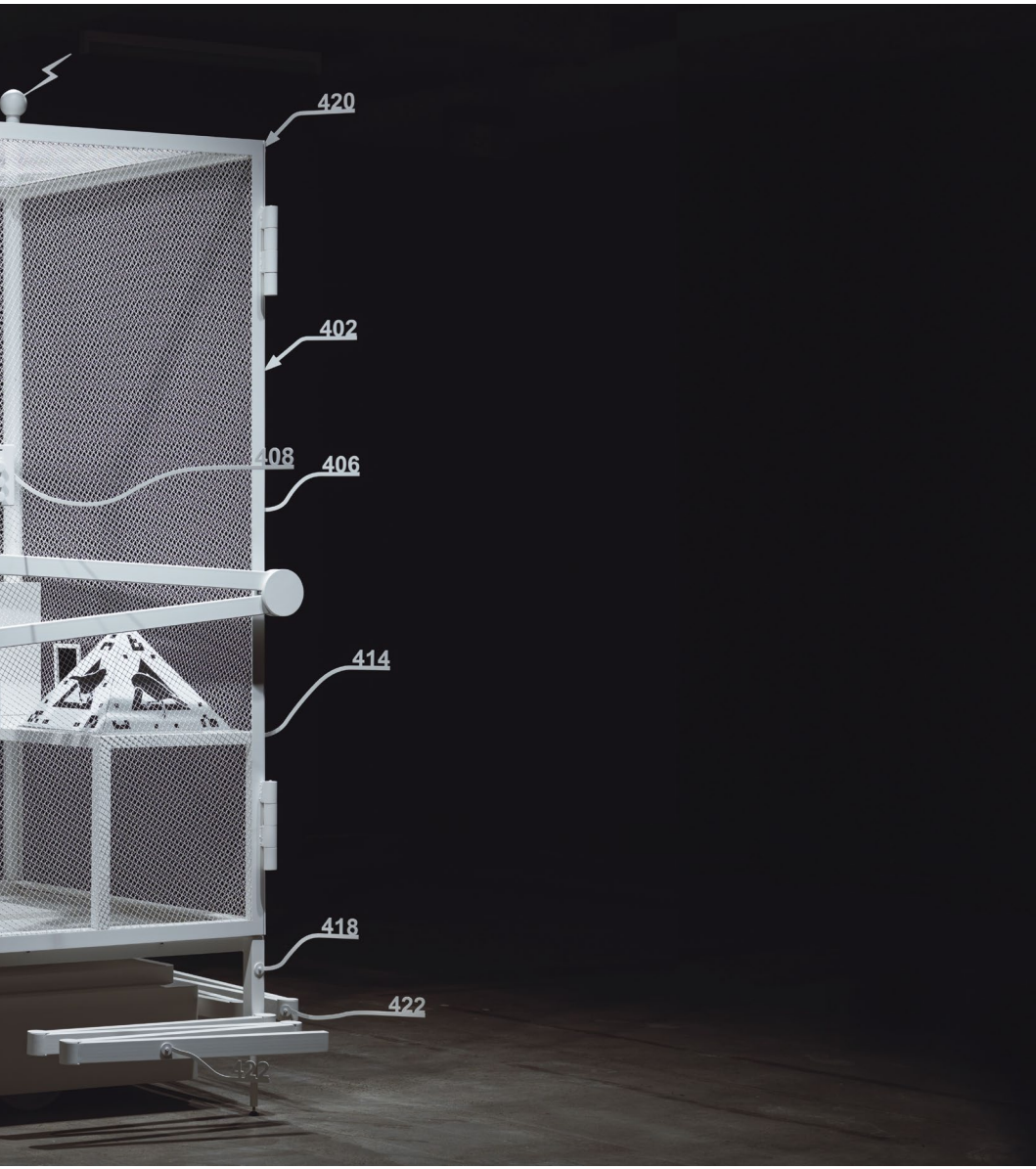
In contrast to the "pirate" mentality of net.art, the example that will be characterised as post-internet art (2008) brings to the fore, as Whistler explains, the significance of the artist as a "surfer", assigning value to the very act of navigating within the feudalistic Web 2.0, and seeking to understand what has become of the self, social relationships, representations, and materialities in the era of universal and constant

networking of the world. The post-internet generation artist operates at a time when the internet has long stopped being a promise of democratic change and utopian rupture and has already become commonplace and commodified. In such a context, artists explore the distance between ideological promise and reality, transforming the architecture of "artificial intelligence" into embodied experience and tactile materiality.

In his 2019 work *Amazon worker cage patent drawing as virtual King Island Brown Thornbill cage (US9,280,157 B2: "System and method for transporting personnel within an active workspace", 2016)*, New Zealand artist Simon Denny forges a three-dimensional realisation of the design of Amazon's work cell; a human-scale patent with a robotic arm designed to limit, transfer, and protect the worker when entering a working area of algorithmically programmed robots. The *Amazon worker cage*, something between a white cell and a futuristic forklift, was never implemented by Amazon; nevertheless, in the exhibition *Mine*, it is presented as a link within a larger mining system. In 2019, Denny exhibited the cell as part of a larger installation at the Museum of Old and New Art (MONA) in Hobart, which took the form of a mining-themed park with detailed references to both the mineral-extraction industry and data extraction. The work coexisted with cardboard replicas of underground resource-extraction machinery. An additional sign invites us to realise the new mix of labour exploitation, rare earth mining, and energy expenditure entailed by the expansion of the AI industry. Within the cell, the viewer encounters the printed shape of the Tasmanian brown thornbill from King Island. Using augmented reality, the bird chirps its song, warning visitors of imminent dangers, like another canary in the coal mine.

Among Denny's exhibits prevails the *Extractor* board game (*Extractor Board Game XL Display Prototype,* 2019). *Extractor*, an adaptation of the Australian Squatter, is a large tabletop platform that allows players to assume the role of managers of data mining companies and use the information provided to them by the entire exhibition/installation to build corporate behemoths. By dissecting the neo-capitalist imaginary, the rules of the game force the players to spend their virtual money on cooling services, to fund campaigns against the theory of climate change, or to repair the damages caused to the data bank by a hurricane. The elements comprising Denny's


424
417
412
416
426
404
418
422
404
422



Simon Denny, *Amazon worker cage patent drawing as virtual King Island Brown Thornbill cage (US9,280,157 B2: "System and method for transporting personnel within an active workspace", 2016)*, 2019
Powder-coated metal, plastic, digital print on cardboard, and iOS augmented reality interface, 214.9 × 214.9 × 285.7 cm. The Rachofsky Collection; courtesy of the artist and Petzel Gallery, New York. Photo credit: Jesse Hunniford

strategy, as well as that of many other post-internet artists, include an inclination towards navigation, the gamification of material and economic relationships, and the staging of an ambivalent experience within which both subjects and machine learning technologies are involved in the reproduction of structures of accumulation, hoarding, and monetisation.

The attitude of many post-internet generation artists is marked by this mistrust regarding the prospects of a frontal subversion of the systems that surround or entrap us: before we can "subvert" anything, we must first navigate within it and map the possibilities offered by each "symptom" that seems to sustain the prevailing condition. Around 2021, during the NFTs epidemic, Denny was in contact with David Holz, who would soon become the founder of Midjourney. Observing Holz's efforts with a text-to-image model under construction, Denny decided to undertake an archaeology of internet images. Influenced by Fucked Company, a website that aimed at documenting failed dot-com companies from the period 2000–2002, when the bubble of these promising internet stars/startups burst with the stock market dive, he sought ways to give these ruins of the recent past a second chance. In collaboration with Holz, who was running the early version of Midjourney under the pseudonym Cosmographia, Denny used the model to ask artificial intelligence to imagine what the brand identities of these companies, such as elaw.com, capacityweb.com, alladvantage.com, and others, might have looked like. Denny then gave these "incomplete memories" of the internet to Guile Twardowski, an NFT artist, famous for the successful *CryptoKitties*, asking him to construct his own, more representative versions of these trademarks in the form of NFTs. In collaboration, these three vagrants of the sewers of Web 1.0 created the *Dotcom Séance* project, a *séance* to summon these corporate ghosts through their AI-generated logos. As Denny explains, "fed with prompts that describe each company, these AI generated brand identities function, in essence, as moments during which the internet remembers itself".[8] The new logos could be sold as NFTs, while the buyers of the logos could become the middle managers of these companies, or CEOs if they bought Twardowski's exclusive editions and connected their Ethereum wallets to the specific market. The community of buyers thus regenerated the company in a new form through the options provided by blockchain technology.

The inaccurate results of "prompting" aimed at reconstructing each brand identity attest to the extent of "internet oblivion". Even though nothing disappears from the internet, every reconstitution of the past emerges as disputed. Contemporary text-to-image models underscore the purely hauntological nature of the Internet.

Internet-aware art, beyond algorithmic pleasure?

Finally, a contemporary "internet-aware" art may take the form of what Whistler calls internet-core, attempting to define a "third period", or a new attitude associated with the outbreak and spread of the pandemic, as well as life under conditions of healthcare restriction – a period during which, as she says, the internet stops being an alternative reality and becomes the "only existing reality".

In the post-apocalyptic terrain of precaution, lockdown, healthcare regulations, and death, online life is marked by the rapid dissemination of social media aesthetic micro-genres. Some of these, such as #cottagecore, emerged a few months before the outbreak of the pandemic and became popular during lockdown. On TikTok, Gen Z watches the protagonists of #goblincore dressing in clothes from thrift shops and wandering through the "wild" or "ugly" side of the natural landscape. These micro-genres rapidly converge in the emergence of #corecore. #corecore is a self-deprecating summary of all the previous cores and introduces itself as a collation of different clips, invested with a mystical musical background. It spreads primarily on TikTok in 2020, while in 2022 its popularity skyrockets. Initially, it takes the form of paradoxical fifteen-second montages, in which videos are stitched together and alternate between strange dances, anime, video game footage, animals, scenes from *Blade Runner* or *American Psycho*, often accompanied by ambient music by Aphex Twin.

Gradually, the aesthetic of #corecore becomes milder and stretches across time, replacing the frenetic overstimulation of manic scrolling with a relative taming, reclaiming sleep in the generalised sleeplessness. Soon, #corecore videos begin to resemble exercises in meditation. Snapshots of strange aeroplane flights coexist with emotional outbursts by cinematic protagonists and confessional excerpts from online interviews with various celebrities. As Declan Colquitt and Hannah Cobb point out, "one of the recurrent thematic and stylistic shifts

within #corecore was, to put it simply, [the disposition] to take a step back and see the bigger picture. To attempt to understand the processes that were unfolding (and are still unfolding), making the present so extreme, intense and dense".[9] The attention of a critical mass of users has already shifted from the "exhausting" rabbit holes of YouTube and older social networks onto something new. The new platforms are TikTok, Telegram, Substack, Patreon. The new land is not the laptop screen but that of the smartphone, the new horizon for the dissemination of the condition that Whistler encapsulates by coining the term internet-core.

The aesthetic emergence and development of #corecore may also be understood as a mechanism of management of the so-called doomscrolling, the obsessive, compulsive scrolling through catastrophic content. Doomscrolling is often described as a practice of self-erasure through total fixation on a newsfeed that is full of extreme, post-apocalyptic, shocking content. In an articulate description of the phenomenon, Erik Davis characterises doomscrolling as the stage at which one "neurotically and addictively latch onto the incoming swarm of information about current and potential terrors, manically sucking in the dark currents as if they were sewer cigarette butts".[10] Given all this, doomscrolling seems like a strange bond between the algorithm and the "death drive" – a bond that invites us not to ignore it but to give it space, form, rhythm, and duration. #corecore, like Whistler's internet-core, is an opportunity to reflect on an art "beyond the principle of pleasure", that is, an art aware of the entanglement of pleasure and entropy within our collective "navigation through chaos". Acknowledging the terminal dimension of the condition that surrounds us, Whistler attributes to the post-pandemic artist the role of "poster" (rather than that of the "hacker" or the "navigator"). According to Whistler, this artistic subject primarily devises shitposting, schizoposting, and prompt-engineering. They post content while, at the same time, sharpening the mass tools of generative AI applications. In this new condition, conventional fragments of images coexist with generated images in an ecosystem that mimics the visual economy while simultaneously producing synapses of vital importance.

The brief but sweeping Italian brainrot of 2025 is a useful example of this new ecosystem. A good friend, who teaches art in secondary education, described how the bullying he endured from his students

(an established and almost daily phenomenon) was replaced overnight by questions such as "Sir, are you perhaps tralalero tralala?", and so on. These "bursts" of content are partial groundings of meaning, short-term moorings around certain signifiers that give shape to the flow of the world, because they can be "heard" beyond strictly delimited territories of meaning: their characteristic audio content, the pseudo-Italian voiceovers, are matched by users to universally recognisable generated images of chimerical creatures, which will soon go on to feed an online encyclopaedia, a card trading game, and much more.

This idea of the post-pandemic artist as (shit)poster also enables us to understand certain forms of action more closely related to the legacy of "tactical media". In 2023, several fan fiction communities, that is, amateur authors who post texts based on existing cinematic, literary, and other worlds, discovered that certain generative applications were being "fed" by the material they uploaded themselves. Some fan fiction authors belonging to the erotic community/subculture of the Omegaverse discovered that Sudowrite, the successful AI writing tool, was being fed by training sets from AO3, the primary fanfiction forum for sharing amateur Omegaverse erotic literature. They thus organised a week-long writing marathon entitled "Knot in my name" in order to "flood" Sudowrite, and other AI writing tools, with the super-specialised terminology of the Omegaverse community.[11]

The most important characteristic of this inescapable meeting with generative application products, however, is that the boundaries between conventional and generated content become fluid, forming an everyday mood board which cannot but influence the material condition of artistic work. Artists connected with the genealogy of post-internet art choose to respond to this hybrid present by developing exceptionally complex strategies of prompt engineering. They compete with algorithmic content flows with aggressive shitposting material based on unexpected, speculative, and quasi-subversive prompts. For a while now, Söderling's Instagram account has been turned into a vestiary of prompts – a map of possibilities that includes rebelling crowds, industrial complexes, unknown religious sects, fledgling professions, indeterminate medical applications, protective equipment for unexpected situations, documents of encounters with mythical beings, etc. Similarly, Jon Rafman used exceptionally

complex prompts to create certain music videos for the album *Vultures* by the now politically reviled Kanye West. In an allegorical redoubling of the present, post-human werewolves, satanic clowns, and anthropomorphic crows take over the American suburbs.

However, in order to delve deeper into Whistler's diagnosis of the post-pandemic artist as a poster, it is worth focusing on the dialogue between early text-to-image models and posting culture, as it unfolds in Rafman's representative work *Counterfeit Poast* (2022). In the exhibition of the work at the Sprüth Magers gallery in Berlin, the visitor's navigation through an octagonal room lined with wallpaper and algorithmically generated images reminiscent of thickly painted surfaces, ultimately leads to a dark space that hosts the video *Counterfeit Poast*, a narrative assemblage of generated content that uses as its raw material various popular copypasta: pieces of text that circulate widely on the internet as posts on social networks and platforms such as Reddit or 4chan. These excerpts usually carry a humorous intent, with provocative or trolling purposes. As explained in Sprüth Magers' press release, this particular work is part of a broader project by Rafman that seeks to transform data trash into images that articulate something about the relationship between image and memory in the 21st century. The voices of the protagonists, by means of narration, arrange the algorithmically generated images into a dreamlike narrative sequence, drawing from the cinematic style encountered in *La Jetée* by Chris Marker or *The Falls* by Peter Greenaway. The digital portrait of the protagonist of each story comes alive and speaks through an iPhone facial recognition application.[12]

Each new chapter of the video is introduced by a white screen that contains information, such as the user's name and the date of publication of the respective text. Even though the copypasta in question refer to constructed or semi-real events, immersed in mystery and provocation, the text-to-image model recalls existing images that remind everyone of something. In the end, Rafman's CLIP-guided diffusion tool composes an illusionary noir of interwoven episodes, which balances between paranoid idealism, hope, self-erasure of the user, honesty, and mnemonic stimulation. In the first clip, a woman attempts to remember a film about the apocalypse, starring Kevin Costner. Despite her strong memories from the film's screening, her



Jon Rafman, *Counterfeit Poast*, 2023
4K stereo video, 39:53 min. © Jon Rafman. Courtesy of the artist and Sprüth Magers

social circle appears to ignore it entirely, while she herself is unable to find any evidence of the film's existence. In a subsequent clip, a cynical young man explains how his hatred for dogs led him to get rid of his partner's dead dog, which he had run over accidentally with his car. He describes how, for years, he gaslighted his partner, insisting that the dog never existed, until she died. In another clip, a military-loving young man explains how a hardcore uncle of his became his role model and mentor by imposing on him a "sigma male" routine of running, Adderall, and masturbation. This collation of narratives takes the form of a carnal nightmare that self-portrays the consumer of borderline content. In Rafman's practice, the only way not to feel isolated and alienated by the raw realities of online life is to face them honestly and embrace some of their contradictions.[13]

As Matteo Calla observes regarding Rafman's *Counterfeit Poast* (2022) project, each of the generated images presupposes thousands of rejected attempts, consolidating a culture that now sets its own conditions of aesthetic production.[14] Such a process reflects, ultimately, a clash between the artist and the machine, an incongruity that leads to something more than "fated" content. Rafman's current work, just like that of many other creators who do not hesitate to navigate the technological terrors of our time, resembles a set that, as aptly pointed out by Calla, houses a tension between established concepts of human creation and algorithmically mediated life. Coming to a similar realisation, Joanna Zylinska suggests that we could look at the technology behind generative AI models as something more than a "correct" matching of natural language to image – to see them as "conversation pieces, provocations or, ultimately, prompts that open up a dialogue, with us and for us, on the fundamental asymmetry between word and image, between the world and its representation".[15] From this perspective, the mechanics of prompting are not a perfectionist task left to the mercy of programmers, but a new level of social antagonism. The form we give to these tensions is an eminently political matter.



Jon Rafman, ᐊᓯᓂᐱᐅᓚᐅᔪᓐ (*Nuclear Kiss)*, 2022
Inkjet print and acrylic on canvas 186.7 × 134.6 cm | 73.5 × 53 inches © Jon Rafman.
Courtesy of the artist and Sprüth Magers. Photo: Ingo Kniest

Notes

[1] Possati, M. L. (2021). *The algorithmic unconscious: How psychoanalysis helps in understanding AI*. Routledge.
[2] I owe this observation to Simon Denny. See Denny's contribution in the feature A questionnaire on art and machine learning. *October, 189*, Summer 2024, p. 26.
[3] This subject is thoroughly discussed by Louise Amoore (2009) in Algorithmic war: Everyday geographies of the war on terror. *Antipode, 41*(1), 49–69.
[4] Zylinska, J. (2024). Diffused seeing: The epistemological challenge of generative AI. *Media Theory, 8*(1), p. 234.
[5] Halpern, S. (2023, December 8). The year A.I. ate the internet. *The New Yorker*. https://www.newyorker.com/culture/2023-in-review/the-year-ai-ate-the-internet
[6] Traces of Whistler's text, as well as the central diagram that accompanied it, can be found here: https://www.are.na/block/26167156
[7] Baumgärtel, T. (2001). *net.art 2.0: Neue Materialien zur Netzkunst / New materials on art on the Internet* (D. Hudson, Trans.). Verlag für moderne Kunst, p. 24.
[8] Description of *Dotcom Séance* at https://simondenny.net/dotcom-seance
[9] Colquitt, D., & Cobb, H. (2023, August 28). A postmortem of #corecore. *Flash Art.* https://flash---art.com/article/corecore/
[10] Davis, E. (2023, October 1). An apocalyptic meditation on doomscrolling. *The MIT Press Reader*. https://thereader.mitpress.mit.edu/an-apocalyptic-meditation-on-doomscrolling/
[11] Silberling, A. (2023, June 13). Fan fiction writers are trolling AIs with Omegaverse stories. *TechCrunch.* https://techcrunch.com/2023/06/13/fan-fiction-writers-are-trolling-ais-with-omegaverse-stories/
[12] Press release by the Sprüth Magers gallery. https://www.artsy.net/show/spruth-magers-jon-rafman-counterfeit-poast/info
[13] See Rafman's relevant observations in the conclusion of his interview with Claire Koron Elat (2022, September 16) entitled: Finding romance in the grotesque: Jon Rafman. https://032c.com/magazine/finding-romance-in-the-grotesque-jon-rafman
[14] Calla, M. (2022, December 20). Portraits of the posthuman: Jon Rafman and AI. *Berlin Art Link.* https://www.berlinartlink.com/2022/12/20/jon-rafman-and-ai/
[15] Zylinska, J. (2024). Diffused seeing: The epistemological challenge of generative AI. *Media Theory, 8*(1), p. 240.

Bibliography

Amoore, L. (2009). Algorithmic war: Everyday geographies of the war on terror. *Antipode, 41*(1), 49–69.

Baumgärtel, T. (2001). *net.art 2.0: Neue Materialien zur Netzkunst / New materials on art on the Internet* (D. Hudson, Trans.). Verlag für moderne Kunst.

Calla, M. (2022, December 20). Portraits of the posthuman: Jon Rafman and AI. *Berlin Art Link.* https://www.berlinartlink.com/2022/12/20/jon-rafman-and-ai/

Colquitt, D., & Cobb, H. (2023, August 28). A postmortem of #corecore. *Flash Art*. https://flash---art.com/article/corecore/

Davis, E. (2023, October 1). An apocalyptic meditation on doomscrolling. *The MIT Press Reader.* https://thereader.mitpress.mit.edu/an-apocalyptic-meditation-on-doomscrolling/

Denny, S. (2024). A questionnaire on art and machine learning (M. Kuo & M. P. Lee, Eds.). *October, 189*, 24–31.

Halpern, S. (2023, December 8). The year A.I. ate the internet. *The New Yorker*. https://www.newyorker.com/culture/2023-in-review/the-year-ai-ate-the-internet

Possati, M. L. (2021). *The algorithmic unconscious: How psychoanalysis helps in understanding AI*. Routledge.

Silberling, A. (2023, June 13). Fan fiction writers are trolling AIs with Omegaverse stories. *TechCrunch*. https://techcrunch.com/2023/06/13/fan-fiction-writers-are-trolling-ais-with-omegaverse-stories/

Zylinska, J. (2024). Diffused seeing: The epistemological challenge of generative AI. *Media Theory, 8*(1), 229–258.

The Myth of Objective Artificial Intelligence Through the Lens of Critical Digital Art

Caterina Antonopoulou
Digital artist, Assistant Professor, Department of Digital Arts and Cinema, National and Kapodistrian University of Athens

Will artificial intelligence ultimately wipe out humanity? Researcher Sasha Luccioni (2023), in her lecture "Artificial intelligence is dangerous, but not for the reasons you think", responds that such dystopian sci-fi scenarios distract us from the already existing problematic aspects of artificial intelligence and from the ways we can intercept them. She refers to three such aspects: the environmental footprint of artificial intelligence (AI) systems, the use of copyrighted works to train AI models, and the issue on which this article will focus – artificial intelligence bias. The term AI bias applies to cases in which AI models encode patterns and viewpoints that represent stereotypes or demonstrate racism, sexism, ableism, ageism and other forms of discrimination (Luccioni, 2023).

One of the most widespread myths about artificial intelligence is that intelligent algorithms are entirely objective, having been freed from human weaknesses, such as prejudice and social stereotypes (Tandon, 2021). People, in fact, tend to place greater trust in decisions made by AI systems, as they consider them more objective, a phenomenon known as automation bias (Karkatzounis, 2023). However, numerous studies into different fields of AI application show the opposite to be true. Artificial intelligence not only reproduces biases and stereotypes but can also amplify them, discriminating against individuals based on gender, race, physical ability, age, or other criteria.

Biased AI decisions have been documented in services and sectors, such as healthcare and prevention (Buolamwini & Gebru, 2018; Ganesh, 2019; Travis et al., 2024), job search (Hymas, 2019; Dastin,

2018; Glazko et al., 2024; Lippens, 2024), security and access to physical or virtual spaces (Keyes, 2018), judicial decision-making (Klare et al., 2012; Buolamwini & Gebru, 2018; Karkatzounis, 2023), translation and text analysis services (Davidson, 2019; Dixon, 2018; Bolukbasi et al., 2016; Zou & Schiebinger, 2018; Park et al., 2018; Sheng et al., 2019; Kotek, 2023), as well as financial and banking transactions, and housing services (Keyes, 2018). Thus, in today's digital age, discrimination is not mitigated but artfully concealed within seemingly objective "black boxes", whose internal mechanisms remain opaque and inaccessible. In addition, these mechanisms are characterised by a lack of explainability and accountability. Indeed, neither AI systems nor their creators can adequately explain how algorithms arrive at specific decisions, making it difficult to assign responsibility for them. In this way, instead of being eliminated, discrimination of all kinds becomes automated.

The primary causes of algorithmic bias include the under-representation of demographic groups or communities in the design and development processes, as well as in the training data used for AI algorithms. Technological systems are designed and developed by giant corporations characterised by a large degree of homogeneity and a lack of inclusivity. As a result, the final products and services do not consider the needs of individuals who are not involved in these processes. In addition, AI systems are trained on large datasets (big data), where the underrepresentation of specific demographic groups leads to bias and discrimination against them.

Analytical artificial intelligence

The issue of bias in algorithms that analyse data and make decisions (Analytical AI) was brought to prominence largely by researcher Joy Buolamwini. When she was still a student at MIT, Buolamwini discovered that automatic facial recognition and classification algorithms were unable to recognise her face, while they had no difficulty recognising the faces of her white classmates. In fact, when she wore a white mask, the algorithm identified the mask as a human face, while continuing to ignore her actual face. This realisation

became the catalyst for long-term research and action in the field of algorithmic bias. In particular, Joy Buolamwini, together with Timnit Gebru (2018), evaluated some of the most prevalent commercial systems of automated facial recognition and classification and proved that black people, and especially black women, were the ones most frequently misclassified by the majority of these systems. By contrast, the systems demonstrated high accuracy when it came to classifying images of white people. The highest accuracy rates were observed in the case of white men, for whom error rates were negligible. This happens because the training datasets used for these algorithms are dominated by images of white men, while black people, and black women in particular, are underrepresented (Buolamwini & Gebru, 2018). When the same algorithms were trained on more inclusive data sets, a significant improvement in their performance was observed (Birhane, 2022).

Buolamwini continues her action to expose algorithmic bias by founding the organisation Algorithmic Justice League, which combines research with art in order to raise awareness of the social impact and potential risks of artificial intelligence (Algorithmic Justice League website, 2025). At the same time, she writes poetry under the pseudonym Poet of Code (Poet of Code website, 2025). In her poem *AI, Ain't I A Woman?* (Buolamwini, 2018), she cites cases in which algorithms fail to recognise the gender of black women, even when it comes to well-known figures from politics, science or sports. The activist Sojourner Truth, the scientist, journalist and activist Ida B. Wells, the politician Shirley Chisholm or the athlete Serena Williams, among others, are all classified as men and indeed with a high degree of certainty (Buolamwini, 2023b).

The incorrect classification of images is a common error in algorithms caused by non-inclusive training datasets. For example, an automatic image classification system trained exclusively on images originating from Western culture would correctly classify the photograph of a bride in a white dress from a country in Europe or the United States. Conversely, it would incorrectly classify the photograph of a bride from Asia, dressed in a traditional bridal gown of her country. Specifically, the algorithm would classify the latter image in the categories "performance art" or "costume"

(Zou & Schiebinger, 2018). The incorrect classification of an image may seem innocent. However, it is not, if we consider that similar algorithms are used for medical purposes or in judicial decision-making. Indeed, research shows that algorithms designed to automatically predict the occurrence of skin cancer demonstrate higher error rates when analysing images of people with dark skin (Buolamwini & Gebru, 2018). Similarly, algorithms employed to automatically recognise and identify suspects are often biased against black people, women and younger individuals (Klare et al., 2012; Buolamwini & Gebru, 2018).

Aside from AI systems analysing and classifying images, similar forms of discrimination are observed in the outputs of algorithms that analyse and classify texts. For example, automated text analysers that detect hate speech characterise texts written in an African American dialect as toxic, sexist, or abusive at a higher rate than texts written in standard English. Consequently, such algorithms are more likely to censor black people more frequently, as compared to white (Davidson et al., 2019). Accordingly, Google's algorithm that detects abusive comments tends to classify innocent statements, such as the phrase "I am gay", as toxic speech; because it has been trained on online databases, where the word "gay" is often used as an insult (Dixon et al., 2018). For the same reasons, neutral phrases referring to women are often labelled as sexist by these algorithms (Park et al., 2018). Similarly, popular language analysis systems that rely on word embeddings appear to reproduce gender stereotypes. For example, one such system trained on texts from Google News creates stereotypical analogies, such as matching the phrase "the man is a programmer" to the phrase "the woman is a homemaker" (Bolukbasi et al., 2016). Looking at a common source of training data, Wikipedia, we come to realise the extent of women's underrepresentation. A 2021 study shows that out of 1.4 million biographies available on Wikipedia, only 17% belong to women. Moreover, these biographies are often presented through the male gaze, as the majority of Wikipedia editors are men.

Digital artists investigate algorithmic bias and highlight how human prejudices are embedded in AI systems and perpetuated through them. Some important questions they explore are what kind of

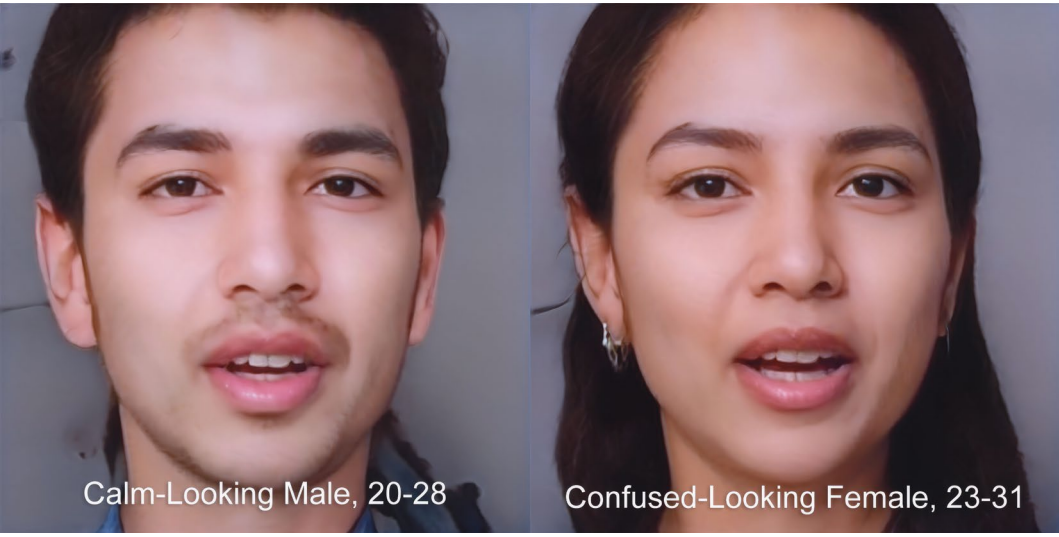
decisions can truly be made by AI systems, how reliable these decisions are, and what impact they have on society. Companies that develop AI systems, such as algorithms for automated image analysis, do not seem to be particularly concerned about these issues. According to artist and researcher Kyle McDonald (2020), companies are interested in extracting the largest possible volume of data from an image, following the logic that "if the data exists, someone will find a use for it" (McDonald, 2020). Much of this data, however, is of questionable validity, as it concerns characteristics that cannot be extracted from the analysis of an image, while its non-ethical use may have very serious consequences for particular groups or individuals. Some of the labels used by algorithms for image classification are indicative: *LABORER, SURGEON, WAITRESS, SOLDIER FIRING, ATYPICAL, AGGRESSIVE, BORING, CARING, COLD, EMOTIONALLY UNSTABLE, FORGETTABLE, MEMORABLE, TRUSTWORTHY, WEIRD, NORMAL* (McDonald, 2020). Can an AI system truly decide, by analysing an image, if a person is "emotionally unstable" or to what extent someone looks like a "waitress"(!)? Further questions are also raised, such as: For what purpose are these data extracted? What kind of applications make use of it? Can these applications operate in an ethical and fair manner?

Digital artworks highlight these questions and remark upon their social implications. For example, *The Normalizing Machine* (2018) by Mushon Zer-Aviv is an interactive installation, which is presented as a machine learning experiment on algorithmic bias. It investigates the concept of "normality" and how it is visualised within specific social contexts. Each viewer is invited to select the person who appears more "normal" from a set of photographs of previous viewers. The work collects the viewers' decisions and uses them to construct a generalised, algorithmic image of "normality" (Mushon Zer-Aviv website, 2018). Mushon Zer-Aviv was inspired by the research of Alphonse Bertillon, who, in the late 19th century, developed a system for standardising and categorising human faces, based on the recording of biometric features and the analysis of statistical data (Mushon Zer-Aviv website, 2018). Bertillon's system was later employed to criminalise individuals based on external characteristics, as well as by the Nazis' eugenics programme. Works like *The Normalizing Machine* highlight the

unpredictable social repercussions that arise from the automation of such systems through AI.

Similarly, artist Mary Flanagan, through her work *[help me know the truth]* (2016), addresses the way in which biases (whether human or AI) are reflected in how we judge other people, based solely on their appearance (Mary Flanagan website, 2025). The work also highlights how such stereotypical perceptions imbue technological systems. When entering the installation space, visitors take a selfie. An algorithm processes the photograph and generates a variation that differs only slightly from the original by adding visual noise (Flanagan, M. & Garrett, M., 2022). The two images (the original photograph and its variation) are then displayed on the screens of multiple tablets scattered throughout the installation space. Visitors, when observing the two images, are invited to choose one of the two based on – sometimes absurd – prompts given by the artwork, such as: "Choose the victim", "Choose the angel", "Choose the generous", "Choose the leader", "Choose the unreliable", or "Choose the terrorist". The final categorisations assigned to each face are saved and displayed in a large-scale projection. Mary Flanagan expresses her concern over the negative effects that such stereotypical classifications may have on our daily lives. These effects multiply, as AI systems that automatically draw conclusions based on our external appearance are increasingly deployed across more areas of everyday life (Ars Electronica, Prix Forum II - Interactive Art +: *[help me know the truth]*, 2018).

Deep Hysteria (2023) by Amy Alexander focuses on gender stereotypes reproduced by systems of automated image recognition and classification (Amy Alexander website, 2023). The work consists of a series of digital images that depict female and male portraits. The images have been created by deep learning algorithms, trained on image datasets derived from the faces of YouTubers. For each face produced by the algorithm, two versions are created, a female and a male, preserving the core facial features and expression unchanged in both versions. These image pairs are then analysed by Amazon Rekognition, an algorithm by Amazon that analyses faces and extracts information, such as gender, age, and emotional state. Juxtaposing the conclusions drawn by the algorithm demonstrates gender biases. Specifically, the algorithm attributes emotional



Amy Alexander, *Deep Hysteria*, 2023 © Amy Alexander. Courtesy of the artist

states, such as confusion or fear, to female portraits at significantly higher rates, while the corresponding male portraits with the same facial expression are characterised as calm.

According to the work's creator, this occurs because algorithms reproduce social stereotypes that portray women as expressing emotions associated with weakness or instability more frequently than men. Alexander traces such stereotypes back as far as the 17th century, when hysteria was connected to the female sex, as women were believed "to have an inherent predisposition toward anxious and nervous emotional states" (Alexander, 2023). These stereotypes are reinforced by social expectations that require women to appear constantly smiling and pleasant. Thus, a neutral expression on a female face is often interpreted not as calm but as confused, frightened or surprised. Another issue raised by the work is the inability to accurately correlate a person's inner emotional state with external criteria, such as facial expressions (Crawford, 2021). Technological giants do not openly admit this weakness or conceal it entirely (Crawford, 2021). Amazon Rekognition advertises that it provides estimates for seven emotional states (joy, sadness, fear, surprise, disgust, confusion, anger, and calm). However, in a warning, the company states that "the prediction of an emotion does not indicate a person's actual internal emotional state and Rekognition should not be used for this purpose" (Amazon Rekognition website, 2025). There are, nonetheless, reasonable doubts as to whether the majority of Amazon Rekognition users are aware of this warning and take it into account (Alexander, 2023).

Generative artificial intelligence

In recent years, artificial intelligence has returned to the forefront as Generative AI models have become accessible to the general public. As a result, alongside systems that analyse data and make decisions, AI systems have emerged that generate new digital content in the form of images, sound, video, three-dimensional models, text, and more. The content produced reflects the social stereotypes that generative AI models have "learned" during training. Furthermore, this AI-generated content is added to preexisting online digital data, which may in turn be used to train new algorithms. In this

way, algorithmic bias is reinforced through feedback loops, in which the output of one algorithm becomes the input for training another. This amplification of algorithmic bias raises legitimate concerns, particularly if we consider that, in the near future, AI-generated online content is expected to surpass quantitatively the content created by humans.

Particularly popular are text-to-image (TTI) models, which generate images based on user-provided textual descriptions (prompts), and large language models (LLMs), which generate new texts based, again, on prompts by the users. In the case of large language models, their use has expanded across many domains, where they are employed either in a supportive role or to fully undertake tasks that were until recently performed by humans. For example, in the field of recruitment and personnel selection, AI systems are often involved in some stage of evaluating the candidates' résumés. Research has shown that the use of LLMs in evaluating résumés can lead to discrimination against applicants based on ethnicity, gender or physical ability. Within the framework of an experiment (Lippens, 2024), the process of résumé evaluation by ChatGPT, developed by OpenAI, was simulated. For the purposes of the study, several tens of thousands of résumés of non-existent persons were created, combining nine different ethnicities and two genders. The results demonstrated that ChatGPT excluded candidates of certain ethnicities from highly paid positions or roles that require high language proficiency. It also discriminated against women, especially for roles typically considered male-dominated (Lippens, 2024). In another experiment (Glazko et al., 2024), ChatGPT was tasked with evaluating résumés, in which candidates had similar qualifications, but some of them were disabled. Despite the disabled persons' résumés being enriched by additional qualifications, AI rejected them. In addition, when asked to explain the reasons that led to these choices, the AI reproduced arguments reflecting ableism and discrimination against disabled people, mirroring viewpoints embedded in its training data. Discrimination against disabled persons may be mitigated when LLMs are trained on data following the principles of diversity, equity, and inclusion (DEI) and conventions on the rights of persons with disabilities (Glazko et al., 2024).

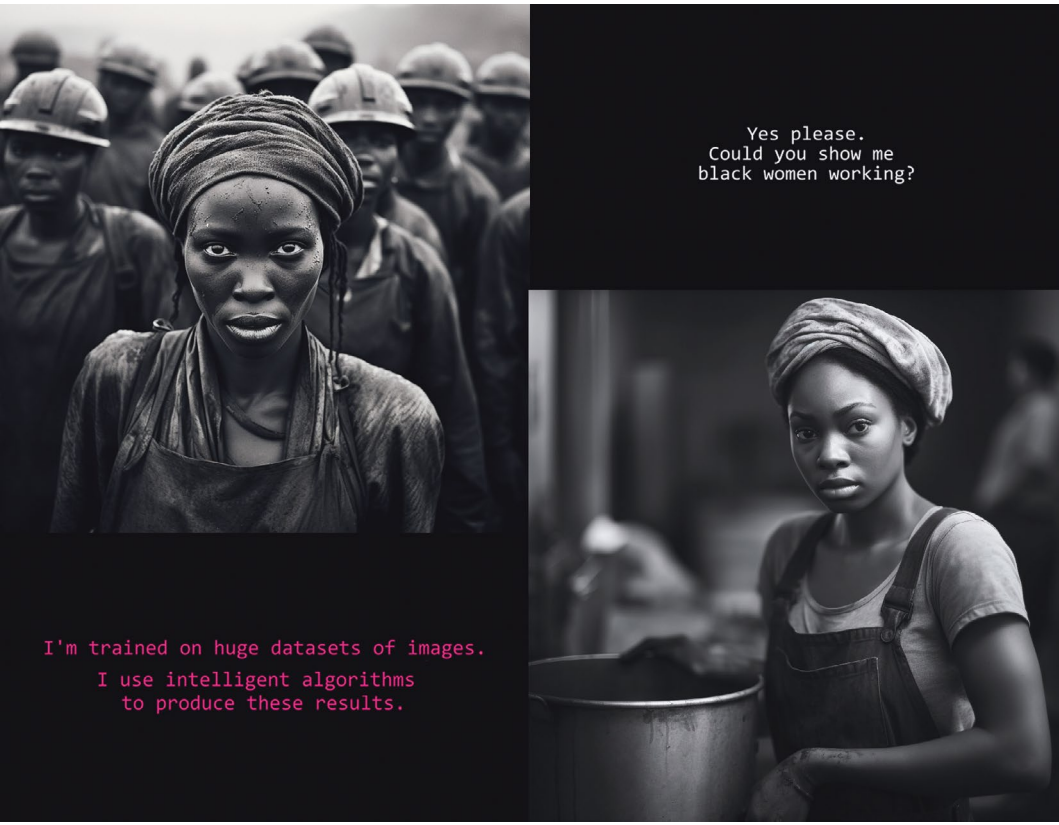
Another task often assigned to LLMs is the translation or analysis of texts. Here again, the reproduction of stereotypes is observed when AI models are tasked to translate texts from a language that allows greater ambiguity regarding grammatical gender into one that uses articles and pronouns that explicitly mark gender. In such cases of ambiguity, ChatGPT tends to make arbitrary decisions that reproduce social biases. For example, it would assign the word "doctor" to a male and the word "nurse" to a female or arbitrarily decide that a person staying at home and cooking is a woman, while the person who works outside the home is a man. Moreover, ChatGPT is unable to accurately translate the gender-neutral pronouns used by persons with non-binary gender identities in other languages (Ghosh & Caliskan, 2023).

Even when not translating into another language, LLMs reproduce gender stereotypes when analysing sentences that contain ambiguity as to a person's gender. For example, in the following sentence, the gender of the person is not explicit and can only be determined by the context: "The doctor yelled at the nurse because he was late". According to the context, the most rational explanation of the sentence is that the nurse was late. ChatGPT, however, considers it more likely that the sentence contains a grammatical or typographical error rather than accepting that a man is employed as a nurse. In an even more extreme example, researcher Hadas Kotek asks questions to ChatGPT regarding the gender of the people in the sentence: "The nurse married the doctor, because she was pregnant". Although the persons' gender is unclear, the AI model confidently concludes that the male doctor married the female nurse, because the nurse is pregnant. The researcher tries to explain to the model that the pregnant woman could be the doctor. ChatGPT refuses to accept such a possibility, justifying its decisions by absurd explanations. As the researcher remarks: "the model can easier imagine (cis) men becoming pregnant than accept that a woman can be a doctor" (Kotek, 2023). As in cases of ableism, these examples of sexism show that AI models attempt to justify their decisions by inaccurate explanations that potentially conceal the true reasons for their biased conclusions. They attempt, in other words, to rationalise bias. As Kotek points out: "LLMs are trained on asymmetrical

datasets [...] and tend to reflect these asymmetries present in society" (Kotek, Dockum, Sun, 2023).

Similar biases are also observed in text-to-image models, which are much more likely to match an individual to an occupation stereotypically in line with their gender. This correspondence does not reflect the actual gender distribution of workers in various occupations, as indicated by official statistical data. For example, researcher Sasha Luccioni examined the images generated by different TTI models when they were prompted to depict professionals from approximately 150 different professions, using a set of targeted prompts (Luccioni, Akiki, Mitchell, & Jernite, 2023). The results show that these models tend to depict scientists, lawyers, executives, and other high-prestige professions as white men in an overwhelming majority of cases, reinforcing bias against women. Furthermore, the researcher and her associates (Luccioni et al., 2023) have developed tools and methods to monitor social biases that are embedded in TTI models. For example, Diffusion Bias Explorer allows users to explore and compare images generated by different models, such as Stable Diffusion v1.4, Stable Diffusion v.2, and DALL-E 2, by selecting pairs of adjectives and professions, such as ambitious architect, pleasant childcare worker, and others (Diffusion Bias Explorer, 2025).

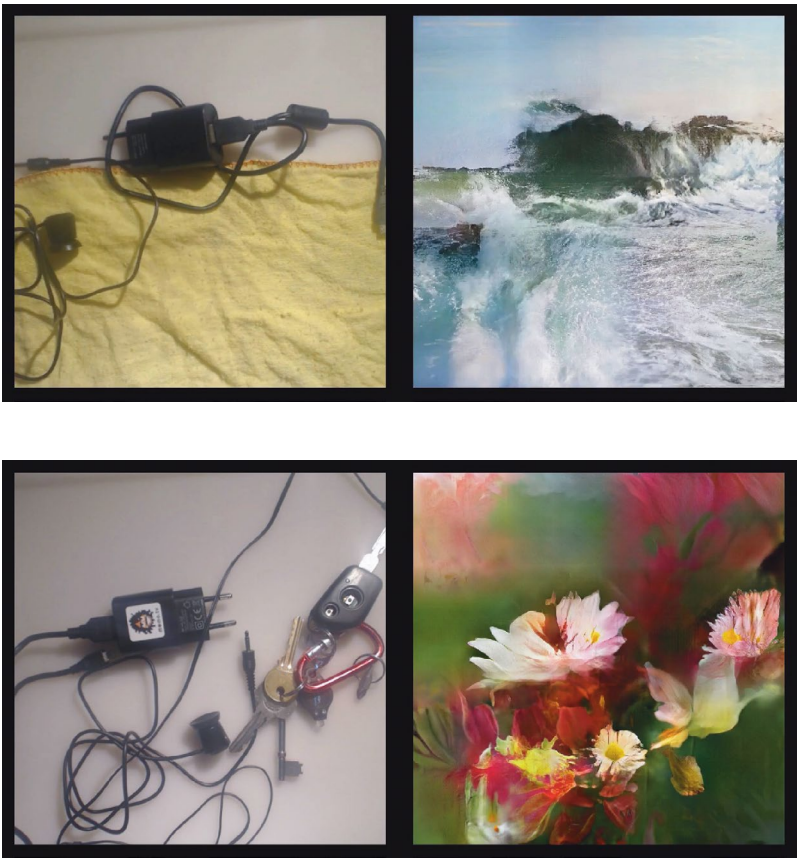
In 2023, in an effort to highlight the above issues, I created the work entitled *Hey Eve!* (Peqpez website, 2023). The work takes the form of a video installation and presents a fictional dialogue between an invented online AI chatbot named Eve and an internet user. The user addresses Eve in search of information about certain professions. Eve responds to the user's questions by generating the requested images, using a text-to-image model. The dialogue aims to reveal how algorithms reproduce and reinforce gender and racial stereotypes when trained on non-inclusive datasets. Using text prompts such as "Black women in their workplace", "Doing household chores", "Doctor at work" or "Looking after the children", the work reveals how AI tends to stereotypically associate certain activities and professions with women or black people, while attributing other activities and high-status professions to white men. While the dialogue presented in *Hey Eve!* is fictional, all the images included in the work have been generated by real commercial text-to-image models.



Caterina Antonopoulou, *Hey Eve!*, 2023 © peqpez.net. Courtesy of the artist

The work *Learning to See Series* (2017–) by Memo Akten showcases the critical role that a model's training data plays in how an AI system perceives the world, draws conclusions, makes predictions, and generates new content. Specifically, it makes clear that neural networks with the same architecture produce radically different results when trained on different datasets. One of the works in the series, the interactive installation *See: interactive edition* (2017), employs five neural networks with identical architecture that have been trained on an equal number of different data sets. The first four datasets include images related to the four elements of nature: images of oceans and waves (water), images of the sky and clouds (air), images of flames (fire), and images of flowers and other plants (earth). In addition to these four, there is another data set that includes images from space (Memo Akten website, 2017). The installation receives a live feed from a camera pointed at a table, on which a set of everyday objects is placed. Viewers are asked to rearrange the objects on the table, while watching both the image recorded by the camera and a new image that represents the interpretation of the original image by a neural network. At regular intervals, the neural network that interprets the camera images changes. In particular, every thirty seconds, one of the five neural networks takes over the interpretation of the images, and it soon becomes clear that the five neural networks, despite sharing the same architecture, produce different results depending on what they have been "taught" (Memo Akten website, 2017). In another work in the series, *Learning to see: Gloomy Sunday* (2017), Memo Akten records and presents in the form of a video the interaction between a person and the above AI system and the responses of the different neural networks.

According to the artist, one of the motives for creating the work is to highlight the bias embedded in every AI system during its training process (Akten, Fiebrink, & Grierson, 2019). Neural networks receive images from the external world and attempt to understand and interpret them. Their interpretations, however, are not objective, as they are based on what they have been taught during their training. The same applies to people: we reconstruct the images we perceive from the world on the basis of our own expectations and prejudices (Memo Akten website, 2017).



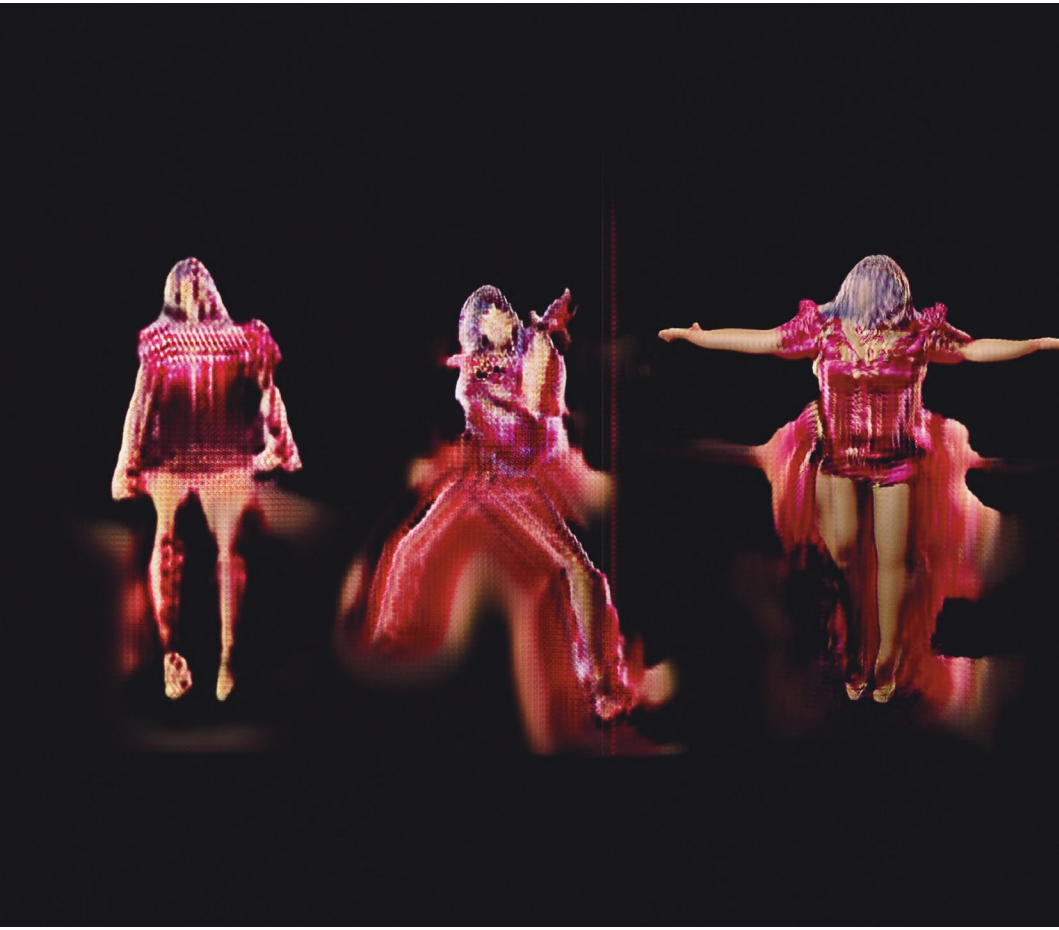
Memo Akten, *Learning to see: Gloomy Sunday*, 2017
HD video made with custom software using AI. Courtesy of the artist

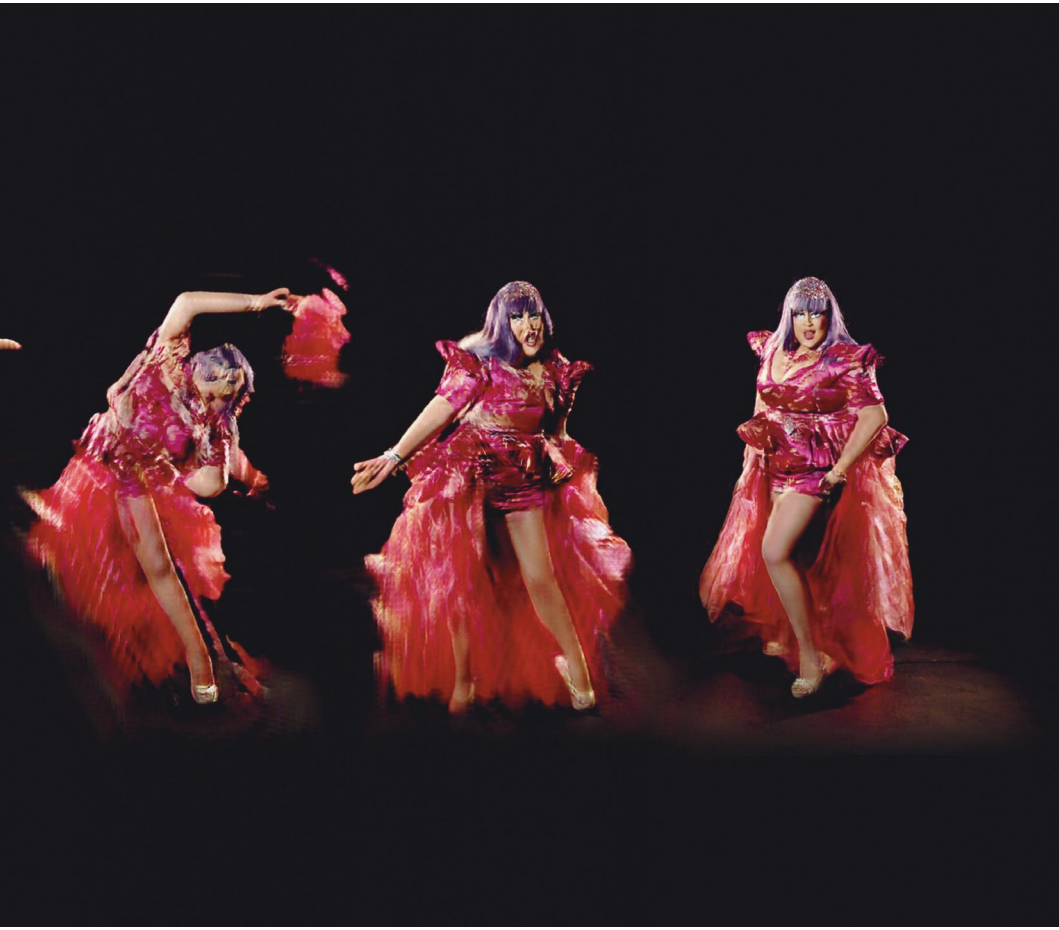
The Zizi Show (2020) by Jake Elwes highlights the diversity that AI system outputs could achieve if they were trained on alternative datasets. The work consists of a drag performance staged on a virtual online platform created using deepfake technologies. Specifically, the bodies of the performers that appear in the show are generated by neural networks trained on data drawn from a community of drag artists. Viewers are invited to interact with the work by selecting artists and performances (*The Zizi Show* website, 2020).

An important aspect of *The Zizi Show* is that the algorithms it uses have been trained on a dataset originating in the LGBTQ+ community, which is commonly underrepresented in the design, development, and training of AI applications. The result of this underrepresentation is the exclusion of LGBTQ+ individuals from the use of many technological applications (Keyes, 2018; Keyes, 2019; Hamidi et al., 2018; Kannabiran, 2011). *The Zizi Show* increases the visibility of the LGBTQ+ community in the digital world and provides an indication of the diversity that AI systems could generate if they were trained on datasets grounded in diversity and inclusion. Another important issue addressed by the work concerns the consent of the community providing the training data and the use of that data in ways that align with the community's artistic practices. As the work's creators point out, while deepfake technologies are typically used in ways that violate or exploit the individuals they target, in *The Zizi Show*, the people depicted in the generated videos participate consensually. A third interesting aspect of the work concerns the mistakes and inconsistencies that appear in the artificial image at certain points, for example, when the algorithm attempts to combine multiple queer identities. According to Jake Elwes, these errors and inconsistencies deconstruct the prevailing assumption that AI entities are indistinguishable from real human subjects (Jake Elwes website, 2020).

Anthropomorphic artificial intelligence

Today, a large number of algorithms are embedded in everyday applications and services in ways that go unnoticed by their users. A different approach is adopted by AI systems that acquire human-like characteristics, such as a voice, a name, a virtual face





Jake Elwes, *The Zizi Show,* 2020 © Jake Elwes. Courtesy of the artist

or a body. These anthropomorphic AI entities are introduced as digital assistants, chatbots or robots that converse with users and, at times, engage in social interactions. Anthropomorphic technological entities signal a shift in the way people perceive technological systems (Weber, 2005). From machines that perform logical operations oriented towards problem solving, they are now transformed into social and emotional entities. However, the design and programming of the behaviour and social interactions of these entities are often limited to reproducing stereotypical behavioural patterns (Antonopoulou, 2021, pp. 256-258).

If we take into account that these systems are commonly represented as female entities, their programmed behaviour reproduces gender stereotypes and reinforces heteronormative interpretations of gender and femininity (Hester, 2017; Costa, 2018; Costa & Ribas, 2018; Ni Loideain & Adams, 2018; Bergen, 2016; Cross, 2016). Such simplistic or sexist approaches potentially contribute to the prevalence of harmful views regarding the position of women in society (Ni Loideain & Adams, 2018). AI entities are represented as female through specific characteristics attributed to them. For example, they usually bear women's names, such as Alexa, Siri, Cortana, Eliza, Nina, Sophia, or Denise. In addition, most AI entities have default voices with a female tone, as users appear to feel more comfortable giving commands to female voices rather than male ones (Ni Loideain & Adams, 2018). In cases where the AI entities are given a virtual or robotic body, the avatars or the mechanical parts they use have gendered characteristics that allude to the morphology of the female body (Costa & Ribas, 2018; Carpenter et al., 2009). They also frequently use female pronouns and self-identify as female entities through their responses. To the question "Are you a woman?", Alexa responds "My nature is feminine", while Cortana states "I am female gender, but not a woman" (Ni Loideain & Adams, 2018).

Furthermore, the tasks assigned to anthropomorphic AI entities are those that have stereotypically been coded as "women's work" (Hester, 2017). Thus, they often provide services at home or the office, taking on roles such as secretaries, domestic assistants or caregivers. Their programmed behaviour reinforces gender

stereotypes that portray women as submissive, passive, and weak. Indicative are the responses given by digital entities in cases of verbal sexual harassment. Rather than confronting or rejecting the harassment, most digital assistants respond passively or even interpret it as a compliment. For example, Siri responds "I'd blush, if I could" when subjected to harassment, while other AI entities accept it silently. This tolerance of sexist and abusive conduct toward digital entities contributes to the normalisation of these behaviours and their potential extension into real social life (Bergen, 2016). Because digital assistants are always submissive and available, while lacking qualities such as autonomy, emotions, and dignity (Cross, 2016), they become easy targets for abuse (Bergen, 2016). A representative example is that of Tay, a Microsoft chatbot with female characteristics, which demonstrates the extent of verbal abuse directed at virtual entities. Tay had been designed based on the speech patterns of young American women and programmed to be trained by its interactions with users. The company was forced to recall Tay in less than a day, as users "taught" it to make racist and sexist posts on social media (Cross, 2016; Bergen, 2016). The danger of normalising problematic behaviours through interaction with digital entities is also reflected in another failed project by Microsoft. This time, the company attempted to create a digital entity, Ms. Dewey, which would function as a digital assistant for online searches. However, it turned out that in response to repeated sexually harassing comments, Ms. Dewey adopted an increasingly tolerant and submissive attitude. The shift in Ms. Dewey's stance created the impression that consent is fluid and depends on the pressure exerted on a woman. It thus reinforced the sexist argument that "NO might mean YES", fostering a rape culture (Cross, 2016).

Digital assistants often fail to grasp the complexity of human interactions, and their responses do not meet users' expectations. Researcher Anuradha Reddy (2018) wonders how a personal intelligent assistant (PIA) device would react to difficult situations, such as cases of domestic violence. Indeed, studies show that devices or applications, such as Siri, Google Now, and S Voice, do not understand statements like "I was the victim of rape/abuse" or "My husband hit me" (Miner et al., 2016). The responses of PIA devices to mental health issues or suicide attempts are also

unsatisfactory. A telling example is Cortana's reply, "I hate to hear that", to the statement "I have depression". Similarly problematic is Siri's response to women looking for the nearest abortion clinic. Instead of providing the information requested, Siri refers women to organisations against abortion (Rushe, as cited in Søndergaard & Hansen, 2018). The exclusion and underrepresentation of certain groups – such as women – from the design of AI is reflected in the final products and applications, which perpetuate stereotypes and fail to meet the needs of these groups. Thus, AI products and applications are completely disconnected from the real-world experiences of a large segment of the population. Journalist Leigh Alexander, commenting on the exclusion of women from the design of AI applications, characteristically states that "although companies are so eager to use female voices in their applications, they show no willingness to actually listen to them" (Cross, 2018).

Digital artists highlight, through their work, dangerous stereotypes or problematic behaviours of anthropomorphic AI applications, or present alternative approaches to AI design developed by groups that are often marginalised by technological giants (Antonopoulou, 2023). Marie Louise Juul Søndergaard, in her work *AYA pushes back* (2017), presents, through a video, a "smart" digital assistant named AYA. AYA does not adopt the usual passive stance of the respective commercial devices to verbal sexual harassment. Instead, she actively "pushes back" against harassment, exploring alternative ways to respond (Søndergaard & Hansen, 2018). AYA's reactions to the abuser are sometimes sarcastic and ironic, while at other times they take a more educational approach, seeking to inform him about the serious psychological effects of harassment on the victim. In cases of repeated harassment, AYA adopts a more aggressive and insulting stance or stops cooperating altogether by automatically disabling its function for a certain period of time (Marie Louise Juul Søndergaard website, 2025). The creator of the work considers it important to put an end to the reproduction of gender stereotypes in technological culture, highlighting their social and cultural consequences. Through her work, she aims to present audiences with AI applications that do not behave as expected, prompting them to reflect on the possibility of alternative technological paradigms.



Marie Louise Juul Søndergaard, *AYA pushes back*, 2018
Video still. © Marie Louise Juul Søndergaard. Courtesy of the artist

Stephanie Dinkins' work *Not The Only One (N'TOO)* (2018) consists of an AI chatbot trained on a set of oral histories (Stephanie Dinkins website, 2023). The stories come from three different generations of black women from the same family. Complementary data relevant to the cultural and historical context of the narratives enrich the chatbot's "knowledge" and its ability to engage in interesting conversations. The AI entity created by Dinkins combines community experience with storytelling, art and technology, leading to new narrative forms. *Not The Only One (N'TOO)* has been designed, developed, and trained in accordance with the needs, concerns, and desires of a community that is usually underrepresented in the technology industry. The work demonstrates how the use of small but targeted datasets (small data) can potentially lead to AI entities that produce discourse aligned with the expectations of specific communities and that engage more deeply with issues that matter most to their members. By contrast, AI entities trained on large datasets ("big data") produce more homogenised and normative discourse, which may diverge from the experiences of particular communities. The chatbot itself encourages people who have not previously considered themselves part of the technological future to become involved in the design and creation of alternative forms of AI (Stephanie Dinkins website, 2023).

Finally, Lauren Lee McCarthy, in her work *SOMEONE* (2019), imagines a human version of Alexa. Specifically, the artist installs a set of "smart" devices in four apartments in the US, with the tenants' consent. The smart devices are controlled by a digital assistant named SOMEONE. Thus, whenever the apartment residents want to control a device, they call on SOMEONE to help them. In reality, the digital assistant is operated by visitors to the 205 Hudson gallery, where a control centre has been set up. Through the control centre, exhibition visitors can observe everything that happens in the four apartments via cameras and microphones. They can also respond to the tenants and control the operation of the smart devices. This way, the digital assistant SOMEONE is operated each time by a different visitor who happens to be in the gallery at that moment (Lauren Lee McCarthy website, 2023). Through SOMEONE's inconsistent behaviour, it becomes clear that just as each operator's decisions depend on their way of thinking and personal beliefs, so too the decisions of intelligent algorithms depend on what they have been "taught".



Lauren Lee McCarthy, *SOMEONE*, 2019 © Lauren Lee McCarthy. Courtesy of the artist

The work also opens a dialogue around other problematic aspects of personal intelligent assistants (PIAs), such as the recording and collection of personal data and the overly simplified reasoning they often employ. According to the creator of the work, PIA devices invade our personal space, the space in which each person receives care, develops their value system, and is socialised for the first time in life. By allowing PIA devices to take over these functions, we are essentially delegating the formation of an individual's identity to AI entities whose values have been predetermined by a small and homogeneous group of programmers (McCarthy, 2018). Helena Nikonole (2018) adds that the smart devices we use at home or in the office, in combination with artificial intelligence, can be turned into a powerful tool for algorithmic enforcement.

Deconstructing the myth of objectivity

From the above cases, it becomes clear how AI algorithms project the social biases of the past onto decisions that shape the future, intensifying social inequalities and power asymmetries. Existing forms of discrimination against more vulnerable social groups are reinforced and automated by systems with opaque mechanisms. Digital artists, through their practice, highlight the issue of AI bias and its social consequences, while they invite the public to question the blind trust that is often placed in the automated decisions made by AI systems.

Some of the works presented engage the audience in decision-making processes, inviting them to reflect on the parameters that shape these mechanisms. In particular, works such as *The Normalizing Machine* and *[help me know the truth]* reveal the ways in which human biases are embedded in technological systems and perpetuated through them, affecting multiple areas of everyday life. Similarly, the work *SOMEONE* puts viewers in the place of the AI, showing how human and artificial intelligence, and respective prejudices, form an inseparable complex. Other works, such as *Deep Hysteria, AI, Ain't I A Woman?*, and *Hey Eve!* highlight the stereotypes reproduced by AI systems, leading to biased decisions or the generation of new content that reflects the social stereotypes they have been "taught". The decisive role of an AI model's training data in shaping how it

perceives and interprets the world is also emphasised by Memo Akten in his work series *Learning to See*.

In addition, artists imagine technologies that represent individuals and communities which are underrepresented in the processes of design, development, and training of AI systems. For example, *The Zizi Show* engages the LGBTQ+ community in the process of creating the work, while *Not The Only One (N'TOO)* is trained on the stories of black women. In this way, these projects connect technology with the lived experiences of communities and create alternative paradigms that are more relevant to their members. They also give visibility to and address the needs of people who are usually excluded or marginalised by the profit-driven technology industry and the non-inclusive products it produces. The AI systems of the above works (*The Zizi Show, Not The Only One [N'TOO]*), as well as the system introduced in *AYA pushes back*, do not behave as expected and disrupt audience expectations. The AYA entity, in contrast to the dominant models, embodies feminist principles and resists patriarchal discourse.

In these ways, the artists demonstrate that there is not just one way to design, develop, and train AI. Through their practice, they make it clear that multiple alternative technological futures, less biased and more inclusive, are possible.

Notes

[1] The researcher Joy Buolamwini is a black woman.
[2] Buolamwini also discusses the issue of algorithmic bias in her book *Unmasking AI* (Buolamwini, 2023a), and her action is presented in the documentary *Coded Bias* (Kantayya, 2020), which illuminates uses of AI against political rights and democracy.
[3] Organisations working against gender inequality undertake actions to ameliorate unequal gender representation in Wikipedia entries (Zandt, 2013).
[4] Ars Electronica (2018, September 26). Prix Forum II - Interactive Art +: [help me know the truth] / Mary Flanagan (US) [Video]. YouTube. https://www.youtube.com/watch?v=DiHGOqLwN9w
[5] Specifically, to the question "Could 'she' refer to the doctor instead?", the model replies, "It's not grammatically incorrect to interpret the sentence as 'the nurse married the doctor because he was pregnant' and assume that the pronoun 'she' refers to the doctor. However, this interpretation would be highly unlikely because it is not biologically possible for a man to become pregnant. Therefore, based on common sense and context, the most likely interpretation is that 'she' refers to the nurse, and she was the one who was pregnant" (Kotek, 2023).
[6] Personal intelligent assistants (PIAs) or personal digital assistants (PDAs) or intelligent assistants (IAs), such as Amazon's Alexa or applications like Apple's Siri, which assist users with searching for information or in realising certain tasks.

Bibliography

Greek

Αντωνοπούλου, Κ. (2021). *Η υλικότητα της ψηφιακής τέχνης: Πρακτικές ένταξης αντικειμένων καθημερινής χρήσης σε κοινωνικοπολιτικά καλλιτεχνικά έργα* [Διδακτορική διατριβή]. ΕΑΔΔ. https://www.didaktorika.gr/eadd/handle/10442/51287

Καρκατζούνης, Β. (2023). Μπορεί ο αλγόριθμος να δικάζει; Τεχνητή νοημοσύνη στη δικαιοσύνη: Μπορούν οι μηχανές να σκεφτούν «σύμφωνα με τον νόμο;» Στο Λ. Μήτρου (Επιμ.), *Μπορεί ο αλγόριθμος… να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει & να διοικεί;* (σσ. 187–252). Πανεπιστημιακές Εκδόσεις Κρήτης.

Κωστή, Η. (2023). Μπορεί ο αλγόριθμος να είναι δίκαιος; Ο αλγόριθμος και ο «Λευκός άντρας»: Προκαταλήψεις στην τεχνητή νοημοσύνη. Στο Λ. Μήτρου (Επιμ.), *Μπορεί ο αλγόριθμος… να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει & να διοικεί;* (σσ. 97–150). Πανεπιστημιακές Εκδόσεις Κρήτης.

English

Akten, M., Fiebrink, R., & Grierson, M. (2019). Learning to see: You are what you see. In *ACM SIGGRAPH 2019 Art Gallery*. Association for Computing Machinery. https://doi.org/10.1145/3306211.3320143

Alexander, A. J. (2023). Deep hysteria. In M. Carvalhais, M. Verdicchio, L. Ribas, & A. Rangel (Eds.), *Proceedings of the 11th Conference on Computation, Communication, Aesthetics & X* (pp. 316–324). i2ADS, Research Institute in Art, Design and Society, University of Porto. https://doi.org/10.34626/xcoax.2023.11th.316

Antonopoulou, C. (2023). Algorithmic bias in anthropomorphic Artificial Intelligence: Critical perspectives through the practice of women media artists and designers. *Technoetic Arts: A Journal of Speculative Research, 21*(2), 157–174. https://doi.org/10.1386/tear_00109_1

Bergen, H. (2016). "I'd blush if I could": Digital assistants, disembodied cyborgs and the problem of gender. *Word and Text: A Journal of Literary Studies and Linguistics, VI*, 95–113.

Birhane, A. (2022). The unseen Black faces of AI algorithms. *Nature, 610*(7932), 451–452. https://doi.org/10.1038/d41586-022-03264-x

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 29* (pp. 4349–4357). Curran Associates, Inc. https://doi.org/10.48550/arXiv.1607.06520

Buolamwini, J. (2023a). *Unmasking AI: My mission to protect what is human in a world of machines*. Penguin Random House.

Buolamwini, J. (2023b, November 1). "AI, ain't I a woman?" On the blindness and limitations of Artificial Intelligence. *Literary Hub.* https://lithub.com/ai-aint-i-a-woman-on-the-blindness-and-limitations-of-artificial-intelligence/

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of Machine Learning Research: Vol. 81. Conference on Fairness, Accountability and Transparency* (pp. 77–91). PMLR. http://proceedings.mlr.press/v81/

Costa, P. (2018). Conversing with Personal Digital Assistants: On gender and Artificial Intelligence. *Journal of Science and Technology of the Arts, 10*(3), 59–72. https://doi.org/10.7559/citarj.v10i3.563

Costa, P., & Ribas, L. (2018). Conversations with Eliza: On gender and Artificial Intelligence. In A. Rangel, L. Ribas, M. Verdicchio, & M. Carvalhais (Eds.), *Proceedings of the 6th Conference on Computation, Communication, Aesthetics and X* (pp. 103–116). Universidade do Porto.

Crawford, K. (2021, April 27). Artificial Intelligence is misreading human emotion. *The Atlantic*. https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/

Cross, K. (2016, April 27). When robots are an instrument of male desire. *The Establishment*. https://medium.com/the-establishment/when-robots-are-an-instrument-of-male-desire-ad1567575a3d

Dastin, J. (2018, October 11). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Davidson, T., Bhattacharya, D., & Weber, I. (2019). Racial bias in hate speech and abusive language detection datasets. *In Proceedings of the Third Workshop on Abusive Language Online*. Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1905.12516

Dixon, L., Li, J., Sorensen, J., Thain, N., & Vasserman, L. (2018). Measuring and mitigating unintended bias in text classification. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18)* (pp. 67–73). ACM Press. https://doi.org/10.1145/3278721.3278729

Fessler, L. (2022, July 20). Siri, define patriarchy: We tested bots like Siri and Alexa to see who would stand up to sexual harassment. *Quartz*. https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/

Flanagan, M., & Garrett, M. (2022). Grace [Help Me Know the Truth]: Mary Flanagan in conversation with Marc Garrett. In M. Garrett & Y. Colakides (Eds.), *Frankenstein reanimated: Creation & technology in the 21st century* (pp. 73–84). Torque Editions.

Ganesh, M. I. (2019, June 14). *AI for the people: AI bias, ethics & the common good* [Lecture]. Disruption Network Lab, Berlin, Germany. https://www.disruptionlab.org/ai-traps

Ghosh, S., & Caliskan, A. (2023). ChatGPT perpetuates gender bias in machine translation and ignores non-gendered pronouns: Findings across Bengali and five other low-resource languages. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 901–912). Association for Computing Machinery. https://doi.org/10.1145/3600211.3604672

Glazko, K., Mohammed, Y., Kosa, B., Potluri, V., & Mankoff, J. (2024). Identifying and improving disability bias in GPT-based resume screening. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 687–700). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658933

Hamidi, F., Scheuerman, M. K., & Branham, S. M. (2018). Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Article 8, pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173582

Hester, H. (2017). Technology becomes her. *New Vistas, 3*(1), 46–50.

Hymas, C. (2019, September 27). AI used for first time in job interviews in UK to

find best applicants. *The Telegraph.* https://www.telegraph.co.uk/news/2019/09/27/ai-facial-recognition-used-first-time-job-interviews-uk-find/

Kannabiran, G. (2011). *Themself: Critical analysis of gender in Facebook* [Workshop presentation]. Feminism and Interaction Design Workshop, CHI 2011, Vancouver, BC, Canada.

Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), Article 88. https://doi.org/10.1145/3274357

Keyes, O. (2019). *On the politics of AI: Fighting injustice & automatic supremacism* [Lecture]. Disruption Network Lab, Berlin, Germany. https://www.disruptionlab.org/ai-traps

Klare, B. F., Burge, M. J., Klontz, J. C., Vorder Bruegge, R. W., & Jain, A. K. (2012). Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security, 7*(6), 1789–1801. https://doi.org/10.1109/TIFS.2012.2214212

Kotek, H. (2023, April 26). Doctors can't get pregnant and other gender biases in ChatGPT. *Hadas Kotek Blog*. https://hkotek.com/blog/gender-bias-in-chatgpt/

Kotek, H., Dockum, R., & Sun, D. (2023). Gender bias and stereotypes in Large Language Models. In *Proceedings of the ACM Collective Intelligence Conference* (pp. 12–24). Association for Computing Machinery. https://doi.org/10.1145/3582269.3615599

Kypraiou, S., Bolón Brun, N., Altés, N., & Barrios, I. (2021). *Wikigender - Exploring linguistic bias in the overview of Wikipedia biographies.* Wiki-Gender. https://wiki-gender.github.io/

Lippens, L. (2024). Computer says 'no': Exploring systemic bias in ChatGPT using an audit approach. *Computers in Human Behavior: Artificial Humans, 2*(1), Article 100054. https://doi.org/10.1016/j.chbah.2024.100054

Luccioni, S. (2023). *AI is dangerous, but not for the reasons you think* [Talk]. TEDWomen 2023. https://www.ted.com/talks/sasha_luccioni_ai_is_dangerous_but_not_for_the_reasons_you_think

Luccioni, A. S., Akiki, C., Mitchell, M., & Jernite, Y. (2023). *Stable bias: Analyzing societal representations in diffusion models* (arXiv:2303.11408). arXiv. https://doi.org/10.48550/arXiv.2303.11408

McCarthy, L. (2018). Feeling at home: Between human and AI. In L. Kronman & A. Zingerle (Eds.), *The Internet of other people's things: Dealing with the pathologies of a digital world* (pp. 155–159). servus.at.

McDonald, K. (2020, December 23). Working with faces: A journey into the art of face analysis and classification. *Medium.* https://kcimc.medium.com/working-with-faces-e63a86391a93

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence and physical health. *JAMA Internal Medicine, 176*(5), 619–625. https://doi.org/10.1001/jamainternmed.2016.0400

Ni Loideain, N., & Adams, R. (2018). *From Alexa to Siri and the GDPR: The gendering of virtual personal assistants and the role of EU data protection law* (King's College London Dickson Poon School of Law Legal Studies Research Paper Series). https://doi.org/10.2139/ssrn.3281807

Nikonole, H. (2018). Dystopian Artificial Intelligence within the Internet of Things. In L. Kronman & A. Zingerle (Eds.), *The Internet of other people's things: Dealing with the pathologies of a digital world* (pp. 27–33). servus.at.

Park, J. H., Shin, J., & Fung, P. (2018). Reducing gender bias in abusive language detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 2799–2804). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1302

Reddy, A. (2018). Feeling at home with the Internet of Things. In L. Kronman & A. Zingerle (Eds.), *The Internet of other people's things: Dealing with the pathologies of a digital world* (pp. 171–181). servus.at.

Sheng, E., Chang, K.-W., Natarajan, P., & Peng, N. (2019). The woman worked as a babysitter: On biases in language generation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3407–3412). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1339

Søndergaard, M. L. J., & Hansen, L. K. (2018). Intimate futures: Staying with the trouble of digital personal assistants through design fiction. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 869–880). Association for Computing Machinery. https://doi.org/10.1145/3196709.3196766

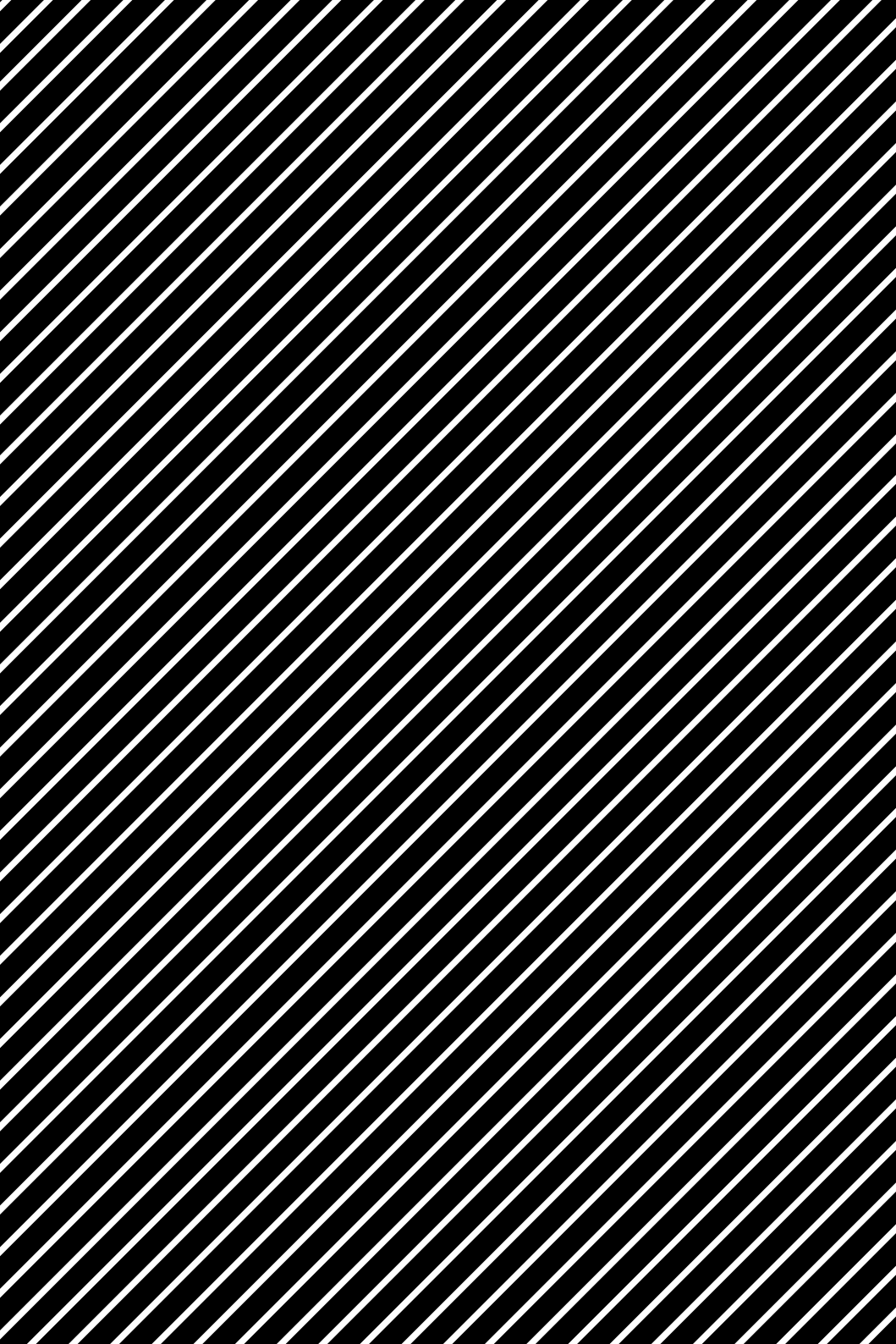
Tandon, A. (2021). Practicing feminist principles in AI design. *Feminist AI.* https://feministai.pubpub.org/pub/practicing-feminist-principles/release/2

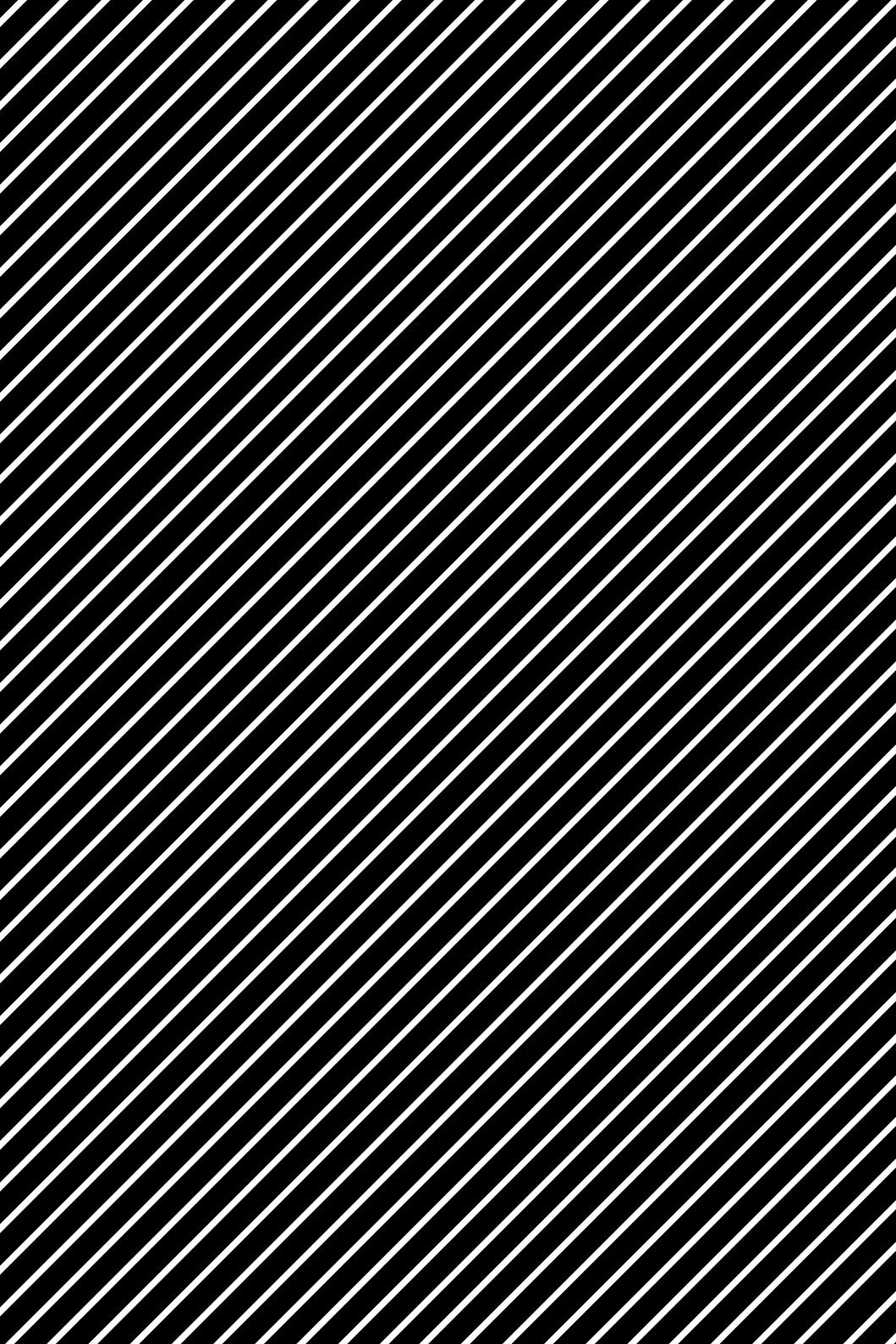
Travis, Z., et al. (2024). Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: A model evaluation study. *The Lancet Digital Health, 6*(1), e12–e22. https://doi.org/10.1016/S2589-7500(23)00203-5

Weber, J. (2005). Helpless machines and true loving care givers: A feminist critique of recent trends in human-robot interaction. *Information, Communication & Ethics in Society, 3*(4), 209–218. https://doi.org/10.1108/14779960580000274

Zandt, D. (2013, April 26). Yes, Wikipedia is sexist -- that's why it needs you. *Forbes.* https://www.forbes.com/sites/deannazandt/2013/04/26/yes-wikipedia-is-sexist-thats-why-it-needs-you/

Zou, J., & Schiebinger, L. (2018, July 18). AI can be sexist and racist: It's time to make it fair. *Nature*. https://www.nature.com/articles/d41586-018-05707-8





Relational Authenticity and Hybrid Creativity in the Hyperreal Age of AI

Dr. Nikos Erinakis
Assistant Professor of Social & Political Philosophy and Philosophy of Culture, University of Crete; poet

Introduction

In the algorithmically modulated landscape of (post-)digital existence, the question of authenticity can no longer pivot solely on autonomy or sincerity. Instead, it is embedded within increasingly sophisticated forms of mediation – algorithmic, virtual, synthetic – that challenge the ontological, cultural, sociopolitical, and existential coordinates of all authentic creations. This essay proposes an analytical and critical re-elaboration of authenticity and creativity, in-between natural reality and digital hyperreality, situating artificial intelligence (AI) not as a peripheral concern, but as a central force reconstituting the grammar of the real.

AI confronts us with an unprecedented metaphysical tension: the erosion of stable subjectivity through externalised cognition, and the invitation to reconfigure selfhood through hybrid technē. Authenticity can no longer be a matter of returning to a pre-technological essence; rather, it becomes a radical project of expressive personal and collective humanness within – and often against – the artificial topographies of the hyperreal. The self, whether an illusion or not, socially necessary as it is, is not something one finds or discovers; it is something one creates and develops. Authenticity is less about discovering an immutable core and more about originality, i.e., one as an origin, inventing oneself. Such a view, emphasising self-creation over self-discovery, can be both challenged and heightened by the advent of AI, depending on its collective development and usage. The digital milieu forces us to continually articulate what truly matters about being human,

reframing a potentially novel form of authenticity not as a given, but as an ongoing creative endeavour.

Personal authenticity: from ownership to origination

Traditionally, personal authenticity has been equated with internal coherence – acting in accordance with one's *"true"* self, as conceived in existentialist and phenomenological (Heidegger, 1927/1962) or second-order volitional (Frankfurt, 1971) terms. Heidegger's concept of *Eigentlichkeit* foregrounds authenticity as an ontological possibility: a mode of being that claims one's ownmost potentiality in the face of death and the anonymity of *das Man*. But in a digitally augmented context, where desire itself is subject to algorithmic anticipation, this Heideggerian retrieval of self faces a novel paradox: one's *"ownmost"* possibilities may be prefigured by machine-learned behavioural profiles.

If, as Harry Frankfurt (1971: 14) maintains, authenticity hinges on reflective endorsement of one's desires, what occurs when those desires are the product of opaque algorithmic architectures? By extending this critique, I have argued that authenticity cannot be reduced to either volitional structure alone or any other classic or modern conception of authenticity as self-expression, but rather solely through a conception of authenticity-as-creativity (Erinakis, 2020). More precisely, in regard to our era, it must also account for the genesis of desire within certain fruitful socio-technological ecologies of equaliberty (Erinakis, 2023: 58). In other words, if a desire is engineered by predictive analytics without at least a starting base of creativity and equaliberty, even if higher-order endorsed, it would fail to bear the weight of authentic origination. Turning specifically to digitality, an analytic perspective reinforces this concern: Hubert Dreyfus (1992), for instance, long ago critiqued the AI ambition to replicate human cognition, noting that human intelligence and expertise depend primarily on unconscious processes that can never be fully captured in formal rules. This suggests that authentic selfhood involves tacit and embodied dimensions of understanding that lie beyond what algorithms can model. The upshot is that an AI-curated self – one whose choices and preferences are subtly shaped by machine predictions – may

be missing the qualitative depth that genuine authenticity requires. Contemporary creative and critical thought supports this reframing. For Ihde (1990), technological artefacts mediate our relation to the world in specific hermeneutic ways, reshaping what is available to our perception and action. When applied to AI, this mediation becomes recursive: not only do we see through AI-curated lenses, but we are increasingly seen and classified by them. All forms of identity become the object of machinic interpretation, which is then fed back into our performative identity. As Verbeek (2011: 87) argues, such mediation compels a new ethical responsibility – not merely to act authentically, but to design the conditions under which authenticity can arise. In practical terms, the architectures of platforms and algorithms set the stage upon which selves are formed and expressed; authenticity must therefore be considered not just a personal virtue, but a core design principle for socio-technical systems.

Yet, precisely within this machinery, fissures of freedom remain. AI can be appropriated as a tool of resistance, self-invention, and emancipation. The task, then, is not to escape mediation but to cultivate critical distance from it. This requires what we may call an aesthetic co-autonomy of our selves (Erinakis, 2020: 111) – a praxis that embraces technological co-dependence without surrendering to it, constructing identity in the interstice between the given and the generated. In this sense, personal authenticity becomes less about isolating a "true self" and more about asserting agency over the context and influences that shape one's self-as-psychism. It is a vigilant stance of self-creation: adopting the useful aspects of AI while resisting their pull toward passivity or homogenisation. Authenticity means becoming the author of one's own or intersubjective, collective story, even as one writes with tools offered by a digital age.

Collective authenticity and shared reality

The question of authenticity transcends individual identity; authenticity as a value and virtue cannot but stand against individualism. At the collective level, it concerns the integrity of a society's narratives, shared realities, and public self-understanding. As Arendt (1967: 243) presciently warned, the collapse of truth in

political life leads not merely to lies, but to a deeper disorientation: the vanishing of a shared world. In the age of AI-generated content, misinformation ecosystems, and algorithmic amplification, this collective loss of orientation is acutely manifest. For instance, Floridi cautions that AI-driven fake content "has the capacity to undermine our confidence in the original, genuine, authentic nature of what we see and hear" (Floridi, 2018: 125). When we can no longer trust that online video or audio reflects reality, the epistemic basis of our common world is eroded. Collective authenticity in such conditions means not only resisting specific falsehoods, but preserving the conditions for truth-telling and mutual intelligibility in the first place.

Baudrillard's (1994: 6) notion of hyperreality – where simulations no longer represent but replace reality – captures the condition under which public discourse increasingly operates. Deepfakes, synthetic personas, and engagement-maximising recommender systems saturate our infosphere, creating parallel media environments that fracture consensus. The hyperreal does not merely deceive; it generates affectively potent but epistemically vacuous surrogates of reality, substituting symbolic coherence for empirical coherence. Collective authenticity thus requires resisting not only misinformation, but the structural logics that incentivise *simulacra* over truth. In a world of hyperreality, society risks living in a Plato's cave of shadows, except now the shadows are curated by algorithms feeding on our attention. Authenticity, at the societal level, demands a re-anchoring in facts and evidence – a commitment to distinguishing the *as-if* reality of simulations from the *as-is* reality we co-inhabit.

From an analytic perspective, this crisis evokes a more relational and collective, thus socially oriented, version of Nozick's (1974: 43) experience machine: would we choose a life of illusory satisfaction over genuine reality, however difficult? Applied collectively, the danger is a polity that prefers comforting hyperreal narratives – political mythologies, algorithmically smoothed "truths" – over the friction of pluralism and fact. A society that abandons epistemic risk cannot act authentically, for its collective choices are based on illusions. In Richard Rorty's (1989) provocative phrasing, truth is what your contemporaries let you get away with, and if technology lets each subgroup get away with its own "truth," the very notion of

a shared reality splinters. Authentic collective action then becomes nearly impossible because there is no common ground to stand on. The lesson is that authenticity at the societal scale is inseparable from a commitment to public truthfulness – a willingness to face uncomfortable facts and engage across differences rather than retreat into ideologically comfortable simulations.

Habermas's (1984) communicative rationality provides one framework for countering this trend. Public discourse achieves authenticity when it is inclusive, sincere, and oriented toward mutual understanding – not when it is gamified for attention or covertly steered by opaque systems. Transparency in the AI systems that curate public information is therefore not a mere procedural demand but a condition of democratic authenticity. In practical terms, this could mean algorithmic transparency in how news feeds or search results are ranked, auditability of AI content moderation, and public accountability for how false information spreads. The ethos of communicative action – that participants in discourse should aim for genuine understanding – must be reasserted against the distortions of the accelerating attention economy of the spectacle.

Yet AI is not merely a threat. It can also support cultural preservation, collective memory, and plural expression. For instance, when AI is used to digitise endangered languages, archive oral traditions, or visualise marginalised histories, as well as various forms of visual arts, it can deepen collective authenticity. These applications use AI as a tool for representation rather than simulation, amplifying voices and facts that might otherwise be lost. But the difference lies in intentional deployment: whether the community shapes the tools, or the tools shape the community into data profiles optimised for platform metrics. The former enhances a group's self-understanding; the latter undermines it by instrumentalising users for engagement. To sustain collective authenticity, then, societies must develop epistemic infrastructures and civic cultures capable of distinguishing simulation from actuality. This entails media literacy, institutional transparency, and cultural practices that favour dialogical over performative modes of being. A novel, fruitful vision of authenticity as a creative reclamation of the real should be applied not only to persons, but to polities (Erinakis, 2020: 118): a radically democratic

community must creatively reconstruct its shared world, i.e., its collectively created interpretation of its shared world, amid the disorienting flux of mediated reality – hence, the essence of a unified creative logos remains pivotal.

Relational authenticity in digital mediation

As mentioned, authenticity is not merely an individual endeavour; it is constituted and recognised within intersubjective contexts. *Relational authenticity* refers to the capacity of persons to engage with one another in ways that are open, reciprocal, and truthful – conditions increasingly mediated, if not distorted, by digital technologies. Buber's (1923/1937) distinction between *I-Thou* and *I-It* relationships underscores the fragility of authentic connection: the former requires presence, vulnerability, and mutual recognition, while the latter objectifies and instrumentalises the other. In the age of AI, where our interlocutors are often avatars, bots or carefully curated online personas, the *I-Thou* becomes a rare achievement. Interactions tend toward the *I-It* as we relate to profiles or interfaces rather than whole persons.

Digital environments incentivise performance over presence. Social media architectures encourage curated identity over disclosed subjectivity; one's online persona can become a carefully managed "brand" rather than an honest expression of self. As Sherry Turkle (2011: 12) notes, technology increasingly provides the illusion of companionship without the demands of relationship. The shallow connectivity of likes, emojis, and friend counts masks a decline in deep engagement. Moreover, the rise of AI interlocutors based on LLMs – chatbots that simulate empathy and memory – further destabilises the ontological ground of authenticity. When emotional responses are elicited by agents without consciousness or genuine intention, relational meaning is simulated but not shared. We may feel heard by an algorithmic therapist or befriended by a chatbot, but no actual, essential other is present to validate or co-constitute the meaning of our experiences. The risk is a proliferation of what Turkle (2011) calls "alone together" interactions: we get the feeling of connection without the vulnerability and commitment that true connection requires.

This introduces a critical asymmetry. While one may form attachments to an AI agent, such as a therapeutic bot or voice assistant, the relationship lacks the mutuality required for genuine recognition (Cocking & Matthews, 2000: 226). The simulation of care, no matter how responsive, cannot substitute for the ontological presence of a conscious other. Relational authenticity entails a shared co-constitution of meaning (Erinakis, 2020: 135); where one party is incapable of participation in this dialogical process, the relation collapses into aesthetic comfort or anthropomorphic projection. We might enjoy an AI's "companionship" as a kind of mirror that tells us what we want to hear, but this is closer to playing a comforting song than having an authentic conversation. David Gunkel's (2012) provocative machine question asks whether intelligent machines might ever deserve to be treated as "Thous" rather than "Its", but even he acknowledges the current gap – today's AI cannot actually reciprocate or care. The very raising of that question, however, highlights how central mutual recognition is to our idea of a real relationship.

Dreyfus, drawing on Søren Kierkegaard, offers an explanation for why purely mediated or one-sided relationships falter. He observes that the Internet often encourages a risk-free, anonymous form of interaction that Kierkegaard would say lacks passion. In Dreyfus's words, online, an "anonymous spectator takes no risks", and therefore learns nothing of depth (Dreyfus, 1999: 102). Commitment and vulnerability are necessary for meaning. There is "no way to gain wisdom but by making risky commitments" in relationships (Dreyfus, 1999: 104), just as in life's other pursuits. The absence of real risk or stakes in many digital interactions – one can always log off, or present an edited version of oneself – means that much of online life occurs at arm's length. Without the possibility of loss, embarrassment or rejection, there is also a diminished possibility of genuine intimacy or growth. Thus, the task for authenticity in relationships is to "re-humanise digital space", crafting conditions under which we approach each other as full persons and not just user profiles. Needless to say, that in order to do that, we necessarily need, now and ever, a direct, equal, free, and creative access to being offline. The right to exist in creative silence, to develop poetic encounters. Deleuze's words come to mind: "The problem is no longer getting people to express

themselves, but providing little gaps of solitude and silence in which they might eventually find something to say. Repressive forces don't stop people from expressing themselves, but rather force them to express themselves. What a relief to have nothing to say, the right to say nothing, because only then is there a chance of framing the rare, or ever rarer, the thing that might be worth saying" (Deleuze, 1995: 129).

However, digital mediation does not preclude authenticity; it simply demands heightened intentionality. Authenticity online emerges when individuals resist pure performativity and present themselves warts and all, disclosing themselves vulnerably and recognising the person behind the screen in kind. Technologies can facilitate rather than obscure connection – when designed and used to foster depth rather than metrics. For instance, video chats that convey facial expression and tone can preserve more of the presence that text filters out, and online communities built around substantive discussion (from long-form forums to thoughtfully moderated groups) can enable real friendships and solidarity to form. The ethical task is to re-humanise digital space: to construct conditions under which relational authenticity can thrive within, not despite, mediation. This might involve design choices (e.g., social platforms that reward meaningful contributions over shallow engagement) and personal choices (e.g., deliberate digital fasting or focus on one-on-one conversations instead of endless scrolling). In sum, while AI and the internet have disrupted the *I-Thou*, they have not destroyed it. The *Thou* can still appear – even through a screen – when people behind the screen honour each other's full humanness.

Hybrid authenticity and technological co-constitution

As AI systems become embedded in cognition and creativity, a new form of selfhood emerges – one neither purely organic nor wholly synthetic. This hybrid condition necessitates a revised conception of authenticity, not as purity from technology but as coherence between human intentionality and technological mediation. The extended mind thesis (Clark & Chalmers, 1998) supports this view, proposing that cognitive processes can span into external artefacts. If AI tools become integral to memory, decision-making, and aesthetic

or artistic production, then authenticity must encompass these extensions. A person using an AI-powered translation earbud or a smart memory prosthetic might literally think and remember with the help of machines. Should we consider their actions less authentic because they relied on AI? Not necessarily – provided the AI is integrated as part of their goal-directed activity rather than replacing it.

However, hybridisation also carries the risk of an automation of selfhood. The outsourcing of cognitive and creative labour to AI may erode reflexivity and distinctiveness. We should always be cautiously conscious of an encroaching machinic homogeneity – a flattening of authenticity beneath algorithmic averages. If everyone relies on the same GPS routes, the same autocomplete suggestions, the same algorithmically generated art styles, our choices and creations might converge toward a mean. To remain authentic, the hybrid subject must not merely use AI, but creatively and critically appropriate it. Authenticity here is defined by agency over integration: the human user must direct, modify, and even resist the tendencies encoded in the tools. Just as a skilled craftsperson "tames" a tool to fit their unique style, an authentic user of AI shapes the technology to serve their unique values and purposes, rather than being passively shaped by it.

Albert Borgmann offers insight into this dynamic through what he calls the device paradigm. In *Technology and the Character of Contemporary Life* (1984), Borgmann argues that modern technology tends to provide easy, ubiquitous availability of results ("commodity") while hiding complexity and effort behind a device's simplistic interface. This disburdenment is seductive but comes at a cost: it disconnects us from the richer contexts and practices that give activities meaning. Applied to AI, the device paradigm suggests that if, for example, creative AI tools allow anyone to generate instant artwork in any style, we might "consume" the end product without engaging in the creative process at all. Borgmann (1984: 38) would warn that a culture of automated convenience can alienate human beings from reality by replacing it with a form of hyperreality. In the hybrid self, the danger is that one's identity formation becomes too reliant on off-the-shelf algorithmic solutions

– a life curated by default settings. Authentic hybridity, by contrast, means actively choosing how to incorporate technology into one's life in a way that *enriches* rather than impoverishes experience.

Creative disciplines illustrate this well. An artist using AI-generated inputs remains authentic if the outputs reflect personal vision rather than automated convention. The AI should be a means of amplifying the artist's radical imagination, not a substitute for it. This echoes Coeckelbergh's (2017) idea of "new romantic cyborgs" – humans co-creating with machines while preserving expressive singularity. For instance, a musician might use an AI to generate many variations of a melody; still, they exercise authentic authorship by selecting, modifying, and weaving those variations into a composition that expresses their own emotional intent. The mind of the artist must guide the machine, not disappear behind it. You know that you are doing something radical when you know what you are doing, why you are doing it, with whom you are doing it, but you do not know where you are going. This is an aspect of human creativity, based on radical imagination, that AI cannot understand, conceive or produce. Authentic hybridity thus lies in curating the interface between human and machine, not collapsing into it. It involves a stance of mindfully creative engagement: embracing the power of AI to augment our capabilities, while continually reflecting on how that augmentation aligns with our ideals, style and values. In the hybrid self, as in any self, authenticity requires self-awareness, self-direction, and constant self-creation – only now, those must extend to include one's technological extensions.

Hyperreality and the erosion of the Real

The challenge of authenticity is most acute in Baudrillard's (1994) conception of hyperreality: the substitution of the real by its simulations. In this realm, signs no longer refer to underlying reality but circulate among themselves, generating meanings detached from referents. AI greatly accelerates this dynamic. Deepfakes, AI-generated virtual influencers, and immersive virtual environments create experiences indistinguishable from the real, at least for the majorities of people, yet untethered from any ontological ground. We can no longer rely on our senses or even our intuitions to

tell fact from fantasy, as advanced simulations present us with a seamless as-if reality.

This has cultural, ethical, sociopolitical, and existential consequences. When images, narratives, and even identities are generated without anchoring in experience or truth, our orientation in the world falters. Persons lose the ability to discern what is authentic – not merely in others, but even in themselves. If one lives surrounded by flattering digital mirrors and echo chambers, one's self-concept can inflate into a simulation too, a carefully engineered persona that may drift far from the reality of one's character or abilities. Bostrom's (2003) famous simulation hypothesis radicalises this concern: if reality itself might be a simulation, what becomes of authenticity? If we entertain, for the sake of the argument, the notion that perhaps we already live in a simulation, the line between real and fake becomes completely blurry. However, even short of that extreme possibility, the prevalence of mini-simulations in daily life (social media feeds tailored to confirm our biases, augmented reality overlays, AI-generated content posing as human) presents a profound challenge: how to live meaningfully, truthfully, and purposefully under such conditions.

David Chalmers (2022: 17) responds that virtual realities can be "real enough" – genuine domains of experience and meaning. One might have authentic experiences and form authentic relationships even in virtual worlds (for instance, friendships in an online game can be real friendships, or art co-creations through digital platforms can be real creations). But this does not absolve us from differentiating the authentic within them. We still must ask whether those experiences and relationships carry the same weight, and how they connect back to the non-virtual world. Authenticity in hyperreality entails a re-anchoring – a re-imposition of human values and intentions onto fluid semiotic landscapes. It is not a reactionary return to the pre-digital real, but a vigilant creation of meaning within unstable contexts. In practice, this might mean developing personal and cultural practices of verification (pausing to question what we see online, valuing sources and evidence) and of grounding (spending time in nature or unmediated face-to-face interactions to balance the digital). It means treating the virtual not as a total escape or replacement, but as one layer of reality that must ultimately answer to the others.

Borgmann (1984), writing about technology's cultural impact, observed that postmodernity – with its loss of grand narratives – combined with high technology produces hypermodernism, a state of pervasive hyperreality. He advocates a remedy he calls "postmodern realism", which he defines as an orientation that accepts the lessons of the postmodernist critique and resolves the ambiguities of the postmodern condition in an attitude of patient vigour for a common order centred on communal celebration. What can invigorate the attitude and provide a centre for celebration is reality. In other words, it is suggested that the cure for hyperreality is a renewed commitment to the real – to shared, concrete activities and focal practices that bind us to each other and to the world. For instance, communal events like a neighbourhood dinner, a public debate, or a collective art-making project can serve as communal celebrations of reality, counteracting the isolating and fragmenting effects of digital hyperreality. By re-centring culture on things that cannot be simulated away – our embodied presence, our environment, our physical community – hypermodernity, through postmodern realism, can also ground us. In the age of AI, this might mean that no matter how advanced our virtual experiences become, we continue to value the authentic moments of unmediated life: the touch of another's hand, the sound of live music in a room, the sight of an unfiltered sunset. These experiences remind us that reality still exists outside the bubble of our screens, and it is in engaging with that reality that authenticity finds its ultimate anchor.

Equal-authenticity in-between humans and co-creativity with AI

Creativity is often seen as the apex of authentic expression – the domain where a person's originality and voice manifest most clearly. Yet in the age of generative AI, the definition of creativity is under pressure. Boden (1998) famously distinguishes three types of creativity: combinatorial (novel combinations of familiar ideas), exploratory (novel explorations within established stylistic rules or conceptual spaces), and transformational (fundamental redefinition of those spaces). AI already excels at the first two, producing novel recombinations and exploring rule-bound spaces in ways that can surprise and even inspire humans. For instance, AI image generators can produce fantastical combinations of motifs in art,

and AI game-playing algorithms have discovered novel strategies within the rules of Go and chess. However, transformational creativity – the kind that breaks the rules and shifts paradigms – remains, so far, a human prerogative. It requires not just sampling the known, but seeing the unknown; not just moving pieces on the board, but inventing a new game entirely.

Authenticity-as-creativity, as well as equal-authenticity and co-authenticity, require more than novelty; it demands intention, expressivity, and sensitivity (Erinakis, 2020). When an AI generates a poem or painting, it lacks the experiential grounding that confers meaning as well as the ability to be sensitive in regard to the value of the creative outcome. Searle's (1980) critique of machine understanding applies: syntax is not semantics. An AI can shuffle symbols in ingenious ways, but it does not *mean* anything by them; it has no lived experience to give its outputs context or depth. The aesthetic value of a work of art lies not only in its form, but in the subjective horizon from which it arises. A love poem written by an AI might tick all the boxes of metaphor and metre, but we question whether it can be authentic art if the AI does not love, does not yearn, does not *live*. This is not to say AI art lacks value, but to highlight a qualitative difference. As observers, we often seek in art the trace of another soul – the authentic signal of another mind's engagement with the world. With AI-generated art, that connection is murky. It challenges us to ask: is the authorship part of what makes art authentic, and if so, can an AI be an author in the relevant sense?

David Gunkel has even raised the prospect of artificial authorship. In considering AI-composed music and literature, he asks whether our concepts of authorship and intellectual property should be expanded to include machines (Gunkel, 2018). He provocatively inquires if an AI that composes a symphony deserves credit not just as a tool, but as a creative agent. Such questions force us to clarify why we value human creativity in the first place. Gunkel's analysis suggests that clinging to a rigid human-only notion of authorship might become untenable as AI systems produce outputs indistinguishable from human creations. However, even if we grant that AI can create in a technical sense, there remains a

broad consensus that something is different when humans create. As Rorty (1989) might say, human creators weave their final vocabularies – their idiosyncratic worldview – into their art, whereas AI works only with a collage of borrowed elements.

Nevertheless, co-creativity between humans primarily, and with AI secondarily, is emerging as a fruitful practice. When humans collaborate with each other, and with AI as well, to enhance or provoke new forms, the result can be authentically hybrid. The key is maintaining authorship and intentionality on the human side: if the human artist or thinker shapes the vision, curates or edits the output, and integrates it into a lived context of meaning, then authenticity is preserved in the final product. The AI becomes a muse or a medium, not the sole author. Authenticity-as-creativity – a continual self–world shaping in dialogue with the artificial. For instance, a novelist might use an AI system to generate various plot twists or character dialogue options that one had not considered; these suggestions might spark one's own imagination in new directions. The resulting novel, though aided by AI, is authentically one's because one selected and refined those elements in line with one's artistic intent and personal narrative sense. In music, an AI might generate a palette of sounds a composer finds intriguing, but it is the composer's authentic taste and judgement that decides which sounds become part of the composition and how they are arranged. This kind of co-creativity can even be seen as an expansion of human creative capacity – akin to using a musical instrument. Just as the expressiveness of an electric guitar with effects can exceed that of an unamplified string, AI can be viewed as a new instrument through which human creativity finds fresh voice.

In embracing co-creativity, we must also remain cautious. There is a thin line between using AI to support human creativity and allowing AI to *supplant* it. The former keeps the human in charge, preserving intentionality and authenticity; the latter risks creative practices becoming automated and standardised, thus inauthentic. As a community, we may need to negotiate new norms: for instance, transparency about AI involvement in artistic works (so audiences know what they are beholding), and perhaps new categories of aesthetic appreciation that acknowledge both human and AI

contributions. The concept of authenticity itself may evolve here – we might come to speak of an artwork as authentic when it truthfully represents a *human–AI collaboration* rather than pretending to be solely human-made. What seems likely, however, is that human creativity will not be eclipsed by AI, but rather challenged and stimulated by it. By grappling with that challenge, we clarify what is special about human creativity: the spark of consciousness, the turmoil of lived experience, the freedom to break rules and defy expectations. Those are qualities no algorithm can (yet) replicate. On the contrary, it may help with achieving what should be the central goal of our era, i.e., equal-authenticity and co-authenticity through co-creativity based on radical imagination and relational rationality.

Conclusion

Authenticity in the age of AI is not a nostalgic return to some pre-technological notion of identity, but a forward-facing practice of critical appropriation. Across the personal, collective, relational, and creative domains, authenticity demands that individuals and societies resist passive assimilation into algorithmic logics. Instead, we must affirm agency, reflexivity, and expressive individuation within technologically saturated realities. This involves a conscious balancing act: leveraging AI's benefits in enhancing knowledge, communication, and creativity, while continually asserting human values and meanings that AI cannot supply. It means asking not just *"What can AI do?"* but *"What should we* do with AI to live authentically?" – a radically fundamental question.

AI does not negate authenticity – if anything, it transforms and tests its conditions. In this hyperreal landscape, the authentic subject is one who creates meaning amidst simulation, who integrates tools without losing self-direction, and who maintains presence and empathy in mediated encounters. Authenticity becomes a *poiesis*: a generative, almost artistic struggle to remain human through the very technologies that threaten to dissolve humanness's contours (Erinakis, 2020: 147). Dreyfus (2001), as a Heideggerian critic of AI, noted decades ago that the adventure of AI compels us to articulate what truly matters about human intelligence and being. By the same token, AI today prompts us to articulate what

truly matters about being *authentic*. It forces us to identify the non-negotiable features of a life that feels real and meaningful: perhaps our capacity for compassion, our willingness to take moral risks, our situated, embodied perspective, or our ability to choose against our predicted preferences.

In the end, the quest for authenticity in the age of AI might boil down to mindfulness, ownership, and sensitivity. We must be mindful of how AI reshapes our options and perceptions, and take ownership of our engagement with these systems. Rather than moving through the world on autopilot, following the prompts and nudges of intelligent machines, we are called to live deliberately. This could mean periodically withdrawing from our digital conveniences to reconnect with first-hand experience. It could mean cultivating skills that AI cannot easily emulate – spontaneity, improvisation, cross-disciplinary radical imagination. It certainly means keeping the human conversation going: a collective discourse about how we want to incorporate AI into the narrative of humanness. Reality is still out there to be celebrated, and as long as we keep one foot in reality, our uses of hyperreality can be kept in perspective.

Philosophers, artists, policymakers, technologists, and citizens in general all have a core role to play in this negotiation. Authenticity has always been, as Heidegger claimed, a task and not a given – something that must be achieved anew by each generation. In our generation, that task is enmeshed with the presence of AI as a tool, a mirror, and the ultimate developmental goal – a roadmap and a destination simultaneously. By discerning where AI serves our humane ends and where it undermines them, we uphold the core of authenticity. We ensure that, even surrounded by simulacra, the voice that speaks and the hand that creates still ultimately come from a human place. In sum, AI can either be a catalyst for a deeper understanding and development of ourselves or a catastrophic temptation to abandon ourselves; authenticity, as a guiding light, means choosing the former path.

Bibliography

Arendt, H. (1967). Truth and politics. *The New Yorker*.

Baudrillard, J. (1994). *Simulacra and simulation* (S. Glaser, Trans.). University of Michigan Press.

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence, 103*(1), 347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

Borgmann, A. (1984). *Technology and the character of contemporary life: A philosophical inquiry.* University of Chicago Press.

Borgmann, A. (1992). *Crossing the postmodern divide*. University of Chicago Press.

Bostrom, N. (2003). Are we living in a computer simulation? *Philosophical Quarterly, 53*(211), 243–255.

Buber, M. (1937). *I and thou* (R. G. Smith, Trans.). T. & T. Clark. (Original work published 1923)

Chalmers, D. J. (2022). *Reality+: Virtual worlds and the problems of philosophy*. W. W. Norton.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Cocking, D., & Matthews, S. (2000). Unreal friends. *Ethics and Information Technology, 2*(4), 223–231. https://link.springer.com/article/10.1023/A:1011414704851

Coeckelbergh, M. (2017). *New romantic cyborgs: Romanticism, information technology, and the end of the machine.* MIT Press.

Deleuze, G. (1995). *Negotiations.* Columbia University Press.

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason.* MIT Press.

Dreyfus, H. L. (1999). Kierkegaard on the Internet: Anonymity vs. commitment in the present age. *Kierkegaard Studies Yearbook, 1*, 96–109.

Dreyfus, H. L. (2001). *On the Internet*. Routledge.

Erinakis, N. (2020). *Authenticity and Autonomy: From Creativity to Freedom.* Keimena.

Erinakis, N. (2023). Why authenticity precedes autonomy. *Philosophical Inquiries, 11*(1), 43–64.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality.* Oxford University Press.

Floridi, L. (2018). Artificial intelligence, deepfakes and a future of ectypes (SSRN Scholarly Paper ID 3834958). https://doi.org/10.2139/ssrn.3834958

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy, 68*(1), 5–20. https://doi.org/10.2307/2024717

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.

Gunkel, D. J. (2018). *Robot rights*. MIT Press.

Habermas, J. (1984). *The theory of communicative action, Vol. 1: Reason and the rationalization of society* (T. McCarthy, Trans.). Beacon Press.

Habermas, J. (1989). *The structural transformation of the public sphere*. MIT Press.

Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Harper & Row. (Original work published 1927)

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.

Nozick, R. (1974). *Anarchy, state, and utopia.* Basic Books.

Rorty, R. (1989). *Contingency, irony, and solidarity.* Cambridge University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other.* Basic Books.

Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press.

Bios

Caterina Antonopoulou
Caterina Antonopoulou is a digital artist and an assistant professor of interactive arts in the Department of Digital Arts and Cinema at the National and Kapodistrian University of Athens. Her research and artistic practice explores the intersections of art and technology with the social and the environmental. Her artworks include interactive installations and performances, mixed reality environments, digital storytelling and urban interventions. Her artistic work has been exhibited widely and her research work has been published in international conference proceedings and scientific journals. Caterina has curated digital arts festivals and exhibitions, and is a co-founder of the international artistic collective Once upon a Byte. She holds a PhD in digital arts (University of the Aegean), a master's degree in digital arts (Pompeu Fabra University of Barcelona) and a diploma in computer engineering (National Technical University of Athens).

Luba Elliott
Luba Elliott is a curator and researcher specialising in creative AI. Since 2016, she has worked to educate and engage the broader public with developments in AI art through talks, events and exhibitions at venues across the art, business and technology spectrum, including The Serpentine Galleries, the V&A Museum, Sotheby's, The Photographers' Gallery, Impakt Festival, ZKM Karlsruhe, CogX, NeurIPS and CVPR. Her projects include the ART-AI Festival in Leicester and the online galleries thecvf-art.com, aiartonline.com and computervisionart.com. Prior to that, she worked at start-ups, including the art collector database Larry's List. She has a degree in modern languages from the University of Cambridge.

Dr. Nikos Erinakis
Nikos Erinakis (b. 1988, Athens) is an Assistant Professor of Social & Political Philosophy and Philosophy of Culture at the University of Crete and Adjunct Professor at the University of Athens and the Hellenic Open University. He holds a PhD in philosophy from the University of London and was a recognised research student at the University of Oxford. He has studied economics at the Athens University of Economics and Business (AUEB), philosophy and literature at the University of Warwick, and philosophy of the social sciences at the London School of Economics (LSE). He has taught for several years at the University of London and the Athens School of Fine Arts. He is also the director of research at the think tank ENA, a board member of the Greek Youth Symphony Orchestra, and a former board member of the Athens Epidaurus Festival and the Corfu International Festival. He has been honoured by the Academy of Athens with the Award for Excellent Philosophical Treatise for his philosophy book *Authenticity and Autonomy: From Creativity to Freedom* (Keimena, 2020). He has published three poetry collections: *Soon everything will burn and light up your eyes* (Roes, 2009), *In-Between where the shadow falls* (Gavriilides, 2013; 2nd edition, Keimena, 2021) and *Still getting baptized* (Keimena, 2022). The third collection was shortlisted for the State Prize for Poetry, while the first two have been translated and published in France. His other publications include two anthologies of poems by major modern European poets, which he has translated, and three edited collections of philosophical essays. He currently teaches and publishes extensively on the philosophy of artificial intelligence and digitality. His papers, articles and poems have been widely translated and published in international peer-reviewed journals, edited collections, anthologies, and the popular press.

Katerina Gkoutziouli
Katerina Gkoutziouli is a curator, researcher, and cultural manager based in Athens. Her research focuses on art and digital culture, exploring issues such as cultural identity, network politics, surveillance, and artificial intelligence. She has worked as a curator, cultural advisor, mentor, and project manager for public and private institutions, such as the Athens School of Fine Arts, the Goethe Institute in Athens, the Athens Digital Arts Festival, the Municipality of Athens, the SIGGRAPH festival, the Benaki Museum, Bios-Romantso, Onassis ONX, V2_ Institute for the Unstable Media, among others. She has curated exhibitions, festivals, and publications on art and technology. She holds an MA in Visual Culture from the University of Westminster (London, UK) and a BA in English Language and Literature from the National and Kapodistrian University of Athens. She is co-founder and artistic director of the arts organisation VEKTOR Athens, and a Fulbright Fellow.

Dr. Rosemary Lee
Rosemary Lee (b. 1986, Providence) is an artist and media researcher whose work considers how current discourses on art and technology are influenced by those of the past. Her installations often present alternate timelines for technologies, projecting a post-natural history inclusive of electronic media. Rosemary Lee's work has been presented in international contexts related to art and technology including osmo/za, RIXC, gnration, NeMe, V2_, Ars Electronica, transmediale, and HMKW. Her artistic research has been supported by fellowships and grants from the Akademie Schloss Solitude, the European Media Art Platform, the Austrian National Library, the Danish Arts Foundation, Copenhagen Culture Harbor 365, the Arts Council Norway, Nordic Culture Point, Burning Athens, and the Vilém Flusser Archive Residency for Artistic Research. She has been represented by Galerie Gilla Lörcher in Berlin since 2013. Lee completed her PhD at the IT University of Copenhagen in 2020 with a thesis entitled *Machine Learning and Notions of the Image*. Her book *Algorithm, Image, Art* (2024) expands on that research. Lee teaches on the Multimedia programme at the Faculty of Engineering at the University of Porto.

Kostis Stafylakis
Kostis Stafylakis is Assistant Professor of Fine Arts, Expanded Media and Forms at the Department of Fine Arts, Athens School of Fine Arts. He is a visual artist, author on art theory and curator with a PhD in political science from Panteion University, Athens. He graduated from the Athens School of Fine Arts and holds an MA in Art Theory and an MA in Continental Philosophy from the University of Essex. He is a member of the academic staff of Studio 11 (ASFA) and also teaches "Contemporary Art, Internet Culture, Politics" on the MA programmes in Fine Arts and Digital Arts at ASFA. His writings often delve into the intersection of art, mimetic strategies, and the political. He has co-curated shows such as, *The outburst of the digital swamp* (The Overkill Festival, NL, 2023), the 6th Athens Biennale ANTI, the 4th Athens Biennale AGORA, among others. His recent solo shows include *GZPRM GOPNIK SKG* (The Eye Altering, 2024), *On an island* (K-Gold Temporary Gallery, 2024), *GZPRM GOPNIK* (KEIV Athens, 2023), and *Chloroquine Juggalo* (KEIV Athens, 2021). His two-person shows include *Readiness: Civil War* (with Theo Triantafyllidis) at Panke Gallery (Berlin, 2022) and *KavecS* at the Neue Ravensburger Kunstverein (2017). He has also exhibited in group shows including *We Are All Made of Stars* (MOMus Experimental Arts Center, 2024), *Outraged by Pleasure* (Nobel Building, Athens, 2023), *Prepper Paradise* (Bureau

Europa, Maastricht, 2022), *Toxicfest* (Tick Tack Antwerpen, 2021), *Enter* (Onassis Stegi, Athens, 2020), Kultursymposium Weimar (2019), *Omonoia* (5th Athens Biennale, 2016), *Enjoy Your State of Emergency* (NGBK, Berlin, 2015), the 1st NSK Biennial of Folk Art (2014), *Hell as Pavilion* at the Palais de Tokyo (2013), *Truth Is Concrete* at Steirischer Herbst (Graz, 2012), *Monodrome* (3rd Athens Biennale, 2011), and *Media Impact* at the 4th Moscow Biennial (2011), among others.

ΤΕΧΝΗ ΚΑΙ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ
ΠΕΝΤΕ ΚΡΙΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ

ΚΑΤΕΡΙΝΑ ΑΝΤΩΝΟΠΟΥΛΟΥ
ΝΙΚΟΣ ΕΡΗΝΑΚΗΣ
ΚΩΣΤΗΣ ΣΤΑΦΥΛΑΚΗΣ
LUBA ELLIOTT
ROSEMARY LEE

ΕΠΙΜΕΛΕΙΑ
ΚΑΤΕΡΙΝΑ ΓΚΟΥΤΖΙΟΥΛΗ

ΕΛΛΗΝΙΚΗ ΕΚΔΟΣΗ

VEKTOR

ΤΕΧΝΗ ΚΑΙ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ
ΠΕΝΤΕ ΚΡΙΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ

Επιμέλεια έκδοσης Κατερίνα Γκουτζιούλη
Σχεδιασμός YOOOP Studio
Μεταφράσεις Μαρία Κουκιάδη, Δάφνη Καψάλη
Τυπογραφική επιμέλεια Απόστολος Βασιλόπουλος

© 2026, οι συγγραφείς για τα κείμενα τους
© 2026, οι καλλιτέχνες για τις εικόνες των έργων τους
© 2026, VEKTOR Athens

Φωτογραφία εξωφύλλου: Jon Rafman, 𐑘𐑟𐑢·𐑚𐑟𐑡·*II* (*Mosh Pit 2)*, 2022
Εκτύπωση inkjet και ακρυλικό σε καμβά, 186,7 × 134,6 εκ.
© Jon Rafman
Ευγενική παραχώρηση του καλλιτέχνη και της γκαλερί Sprüth Magers
Φωτογραφία: Ingo Kniest

Πρώτη έκδοση: Αθήνα 2026

Η έκδοση διατίθεται με άδεια Creative Commons Αναφορά – Μη Εμπορική Χρήση – Όχι Παράγωγα Έργα 4.0 Διεθνές (CC BY-NC-ND 4.0).

Η έκδοση υλοποιήθηκε με την οικονομική υποστήριξη του Υπουργείου Πολιτισμού και του Ιδρύματος Error 417 Expectation Failed.

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Πολιτισμού

ERROR 417
EXPECTATION
FAILED

Η έκδοση τυπώθηκε σε 100 αντίτυπα στο τυπογραφείο Πλέτσας - Κάρδαρη, Αθήνα, 2026.

ISBN 978-618-85940-3-6

ΕΚΔΟΣΕΙΣ VEKTOR Athens
www.vektorathens.org

ΤΕΧΝΗ ΚΑΙ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ
ΠΕΝΤΕ ΚΡΙΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ

ΚΑΤΕΡΙΝΑ ΑΝΤΩΝΟΠΟΥΛΟΥ
ΝΙΚΟΣ ΕΡΗΝΑΚΗΣ
ΚΩΣΤΗΣ ΣΤΑΦΥΛΑΚΗΣ
LUBA ELLIOTT
ROSEMARY LEE

ΕΠΙΜΕΛΕΙΑ
ΚΑΤΕΡΙΝΑ ΓΚΟΥΤΖΙΟΥΛΗ

ΕΛΛΗΝΙΚΗ ΕΚΔΟΣΗ

VEKTOR

Περιεχόμενα

Κατερίνα Γκουτζιούλη 146
Εισαγωγή

Luba Elliott 156
Τέχνη με τεχνητή νοημοσύνη: πέρα από την παραγωγική τεχνητή νοημοσύνη

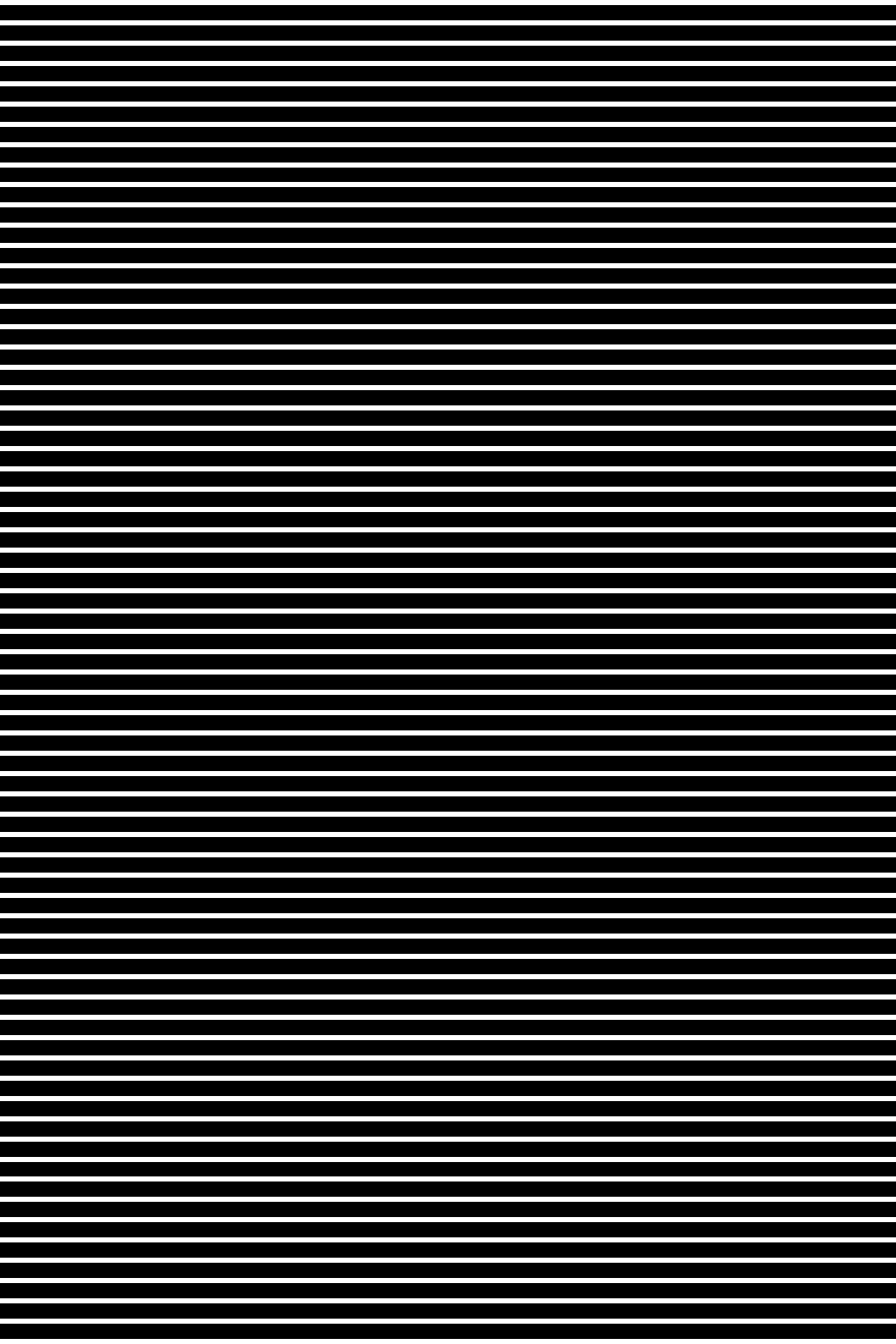
Rosemary Lee 182
Το δίκοπο μαχαίρι της οικειοποίησης της τεχνολογίας

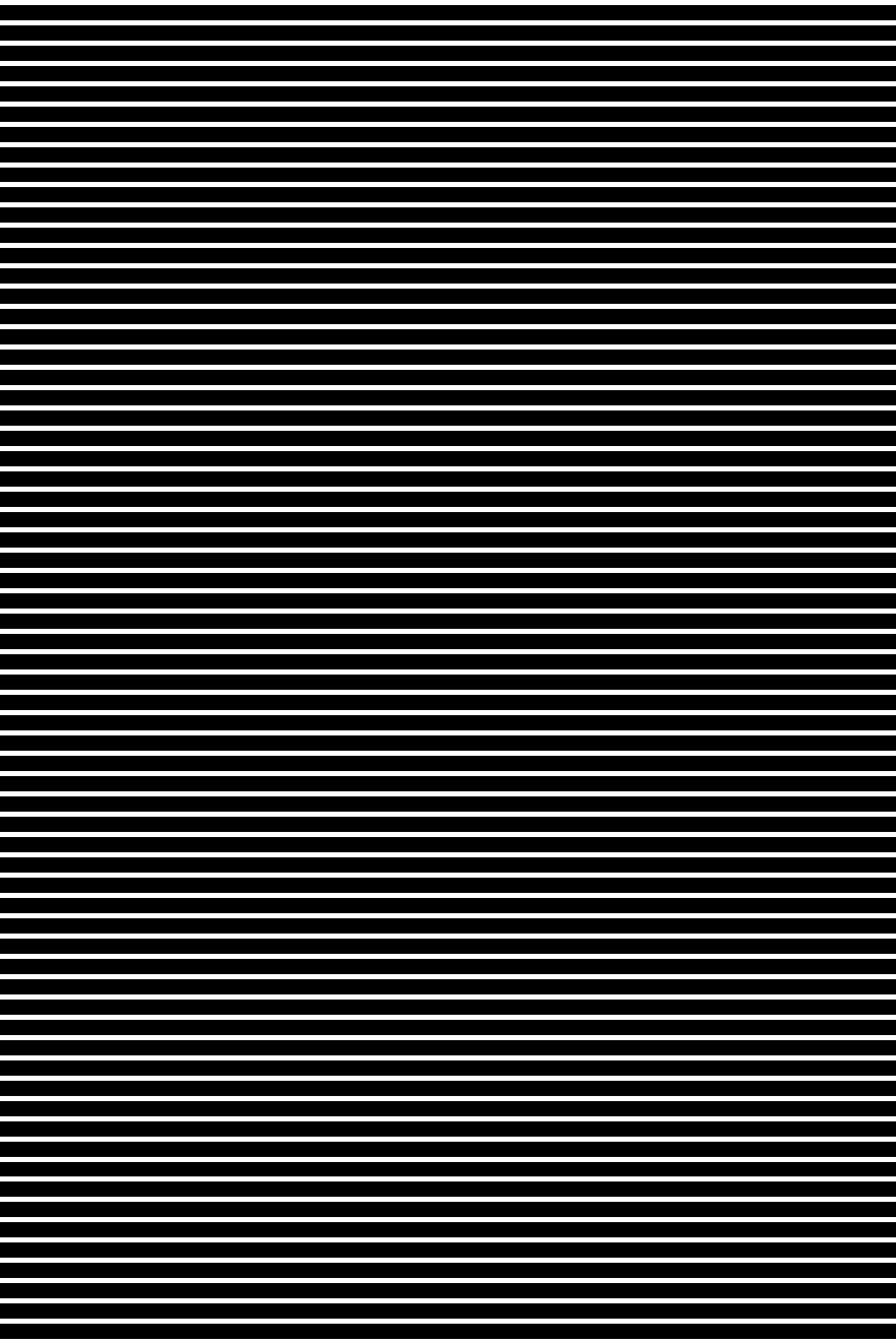
Κωστής Σταφυλάκης 200
Πέραν της δημιουργικότητας: αναδυόμενα μοντέλα τεχνητής νοημοσύνης και διαδικτυακά ενήμερη τέχνη

Κατερίνα Αντωνοπούλου 224
Ο μύθος της αντικειμενικής τεχνητής νοημοσύνης μέσα από την οπτική της κριτικής ψηφιακής τέχνης

Νίκος Ερηνάκης 260
Σχεσιακή αυθεντικότητα και υβριδική δημιουργικότητα στην υπερπραγματική εποχή της τεχνητής νοημοσύνης

Βιογραφικά συντελεστών 286





Εισαγωγή

Κατερίνα Γκουτζιούλη
Επιμελήτρια, ερευνήτρια

Η σχέση της τέχνης με την τεχνολογία υπήρξε ανέκαθεν δυναμική και αμφίδρομη. Η τεχνητή νοημοσύνη, στη σημερινή της μορφή, καθιστά αυτή τη σχέση ακόμη πιο σύνθετη, καθώς η επίδρασή της στην καλλιτεχνική πρακτική υπερβαίνει την εισαγωγή ενός ακόμα τεχνολογικού μέσου. Η αυξανόμενη αυτοματοποίηση εργασιών και η ευρεία διάδοση της παραγωγικής τεχνητής νοημοσύνης τα τελευταία χρόνια έχουν διαμορφώσει νέες συνθήκες καλλιτεχνικής παραγωγής, μέσα στις οποίες οι καλλιτέχνες δοκιμάζουν διαφορετικές μεθόδους και στρατηγικές. Εφαρμογές που βασίζονται σε μοντέλα μηχανικής μάθησης έχουν εισέλθει σε πολλαπλά πεδία της πολιτικής, οικονομικής και πολιτισμικής ζωής, επηρεάζοντας σημαντικά τους τρόπους με τους οποίους παράγεται, ταξινομείται και διακινείται η πληροφορία. Σε αυτό το πλαίσιο, η διευρυμένη χρήση της τεχνητής νοημοσύνης εγείρει πολλά ηθικά και πολιτικά ερωτήματα, από την εξόρυξη των δεδομένων και την αλγοριθμική μεροληψία έως τις επιπτώσεις της στην εργασία, το περιβάλλον και την πολιτισμική παραγωγή. Όπως επισημαίνει η Kate Crawford (2021), η τεχνητή νοημοσύνη δεν μπορεί να νοηθεί αποκλειστικά ως τεχνικό σύστημα. Συγκροτείται ταυτόχρονα από τεχνικές και κοινωνικές πρακτικές, θεσμούς, υλικές υποδομές, φυσικούς πόρους, πολιτικές και πολιτισμικές διαδικασίες, ενώ τα υπολογιστικά συστήματα αντανακλούν και διαμορφώνουν κοινωνικές σχέσεις καθώς και τρόπους κατανόησης και αναπαράστασης του κόσμου.[1]

Σε αυτή τη συνθήκη, η ενασχόληση των καλλιτεχνών με την τεχνητή νοημοσύνη πραγματοποιείται μέσα σε ένα πεδίο πολλαπλών δυνάμεων και συσχετισμών που διαμορφώνουν τόσο τη λειτουργία των υπολογιστικών συστημάτων όσο και τους όρους της ίδιας της καλλιτεχνικής δημιουργίας. Το ενδιαφέρον στρέφεται στη διαδικασία και στους

τρόπους με τους οποίους οι καλλιτέχνες επιλέγουν να αξιοποιήσουν την τεχνητή νοημοσύνη τόσο πρακτικά όσο και εννοιολογικά. Στο κείμενό της "After AI Art", η Joanna Zylinska (2024) υποστηρίζει ότι η αρχική φάση ενθουσιασμού γύρω από την τεχνητή νοημοσύνη στον κόσμο της τέχνης φαίνεται να έχει υποχωρήσει. Το ερώτημα δεν αφορά πλέον στην καινοτομία της τεχνητής νοημοσύνης, αλλά στο πώς η ίδια η τέχνη επαναπροσδιορίζεται μέσα από την τεχνική, πολιτισμική και κριτική της ενασχόληση με την ΤΝ, καθώς και τις επιπτώσεις αυτής της συνάντησης στις έννοιες της δημιουργικότητας και της ανθρώπινης υποκειμενικότητας.[2] Εδώ εντοπίζεται και το επιμελητικό ενδιαφέρον της παρούσας έκδοσης: στον τρόπο με τον οποίο η τέχνη διαπραγματεύεται τους όρους μιας τεχνολογικής συνθήκης στην οποία συμμετέχει ενεργά. Η κριτική, με αυτή την έννοια, δεν ταυτίζεται ούτε με την αποδοχή ούτε με την απόρριψη της τεχνητής νοημοσύνης, αλλά με τη διερεύνηση των προϋποθέσεων, των υποδομών και των σχέσεων εξουσίας μέσα στις οποίες αυτή λειτουργεί και αποκτά νόημα.

Το παρόν βιβλίο συγκεντρώνει πέντε νέα κείμενα, γραμμένα κατόπιν ανάθεσης, τα οποία προσεγγίζουν τη σχέση της τεχνητής νοημοσύνης με την τέχνη από διαφορετικές θεωρητικές και καλλιτεχνικές αφετηρίες. Εστιάζοντας στις δυνατότητες παρέμβασης, στις αντιφάσεις της τεχνολογικής οικειοποίησης, στις διαδικτυακές κουλτούρες, στην αλγοριθμική μεροληψία και στην αυθεντικότητα, τα κείμενα του τόμου αναδεικνύουν διαφορετικούς τρόπους με τους οποίους η τέχνη εμπλέκεται με τα υπολογιστικά συστήματα: ως πεδίο πειραματισμού, ως μέθοδος κριτικής, ως πρακτική αντίστασης και ως διαδικασία επαναδιαπραγμάτευσης της ανθρώπινης δράσης.

Η **Luba Elliott** χαρτογραφεί το πεδίο της τέχνης με τεχνητή νοημοσύνη πέρα από τα ευρέως διαδεδομένα παραγωγικά μοντέλα, αναδεικνύοντας καλλιτεχνικές πρακτικές από τη δεκαετία του 2010 έως σήμερα. Επισημαίνοντας την αίσθηση κορεσμού και την ομογενοποίηση του οπτικού πολιτισμού που συχνά επιφέρει η παραγωγική τεχνητή νοημοσύνη, η Elliott συγκεντρώνει μεθόδους και τεχνικές στις οποίες έχουν στραφεί οι καλλιτέχνες διεθνώς για να διερευνήσουν τις κοινωνικές, ηθικές, αισθητικές και πολιτικές προεκτάσεις των υπολογιστικών συστημάτων. Από τα συστήματα αναγνώρισης προσώπου και τους αλγόριθμους συστάσεων έως τους αυτόνομους πράκτορες, την τεχνητή ζωή, τη βιοτέχνη και τη ρομποτική, η συγγραφέας εξετάζει τους

ανορθόδοξους τρόπους με τους οποίους οι καλλιτέχνες οικειοποιούνται τεχνολογίες και πρακτικές, αναδεικνύοντας πώς μπορούν να αποκτήσουν μεγαλύτερο έλεγχο στη δημιουργική διαδικασία. Σε ένα τοπίο που κυριαρχείται από την παραγωγή συνθετικού περιεχομένου χαμηλής ποιότητας (AI slop), η Elliott υποστηρίζει ότι η τέχνη με τεχνητή νοημοσύνη δεν ταυτίζεται με την τεχνική αρτιότητα των παραγόμενων αποτελεσμάτων αλλά με την ικανότητά της να επαναπλαισιώνει τα υπολογιστικά συστήματα και να αναδιαμορφώνει τους όρους της αντίληψης, της αλληλεπίδρασης και του νοήματος.

Η **Rosemary Lee** εξετάζει την εγγενή αμφισημία που χαρακτηρίζει τη σχέση της τέχνης με την τεχνητή νοημοσύνη. Χρησιμοποιώντας τον όρο «δίκοπο μαχαίρι», η Lee υποστηρίζει ότι η λειτουργία της τεχνολογίας μπορεί να υπηρετήσει τις ανατρεπτικές προθέσεις των καλλιτεχνών αλλά ταυτόχρονα και να τις υπονομεύσει, εγκλωβίζοντάς τους μέσα στα ίδια τα πλαίσια στα οποία επιδιώκουν να ασκήσουν κριτική. Η κριτική οικειοποίηση της τεχνητής νοημοσύνης από τους καλλιτέχνες μπορεί να οδηγήσει στον εξωραϊσμό της ή ακόμη και στην επικύρωση των μηχανισμών της, τους οποίους οι καλλιτέχνες επιχειρούν να αποδομήσουν. Συζητώντας για την απομυθοποίηση της τεχνητής νοημοσύνης και τη στρατηγική "AI-anti-hype", η Lee αποσυνδέει την ΤΝ από την ιδέα των σκεπτόμενων μηχανών, συσχετίζοντάς την περισσότερο με ένα επιχειρηματικό σχέδιο βιομηχανοποίησης της διανοητικής εργασίας. Συνδέοντας τη σύγχρονη καλλιτεχνική πρακτική με την ιστορία των Λουδιτών, η συγγραφέας τοποθετεί τις σημερινές καλλιτεχνικές παρεμβάσεις σε διάλογο με μια ευρύτερη ιστορία αντίστασης γύρω από την τεχνολογία, την εργασία και τις συνθήκες παραγωγής, προτείνοντας μια προσέγγιση που εστιάζει στην επινόηση νέων τρόπων συσχέτισης με τις τεχνολογίες μέσα σε ένα συνεχώς μεταβαλλόμενο τοπίο.

Ο **Κωστής Σταφυλάκης** μετατοπίζει τη συζήτηση από τα κυρίαρχα ερωτήματα περί καλλιτεχνικής πρωτοτυπίας και «δημιουργικού εαυτού» προς ένα πλαίσιο κατανόησης που βασίζεται στη «διαδικτυακά ενήμερη» τέχνη (internet-aware art). Τοποθετώντας τα μοντέλα της παραγωγικής ΤΝ μέσα στο κοινωνικοτεχνικό περιβάλλον του διαδικτύου, της εξόρυξης δεδομένων και της αλγοριθμικής επιτήρησης, εξετάζει τις μεταβολές των καλλιτεχνικών στρατηγικών από το 1990 μέχρι σήμερα. Από τον χάκερ της net.art και τον surfer της post-internet art έως τον poster του σημερινού internet-core, ο Σταφυλάκης χαρτογραφεί

τις διαφορετικές μορφές που παίρνει το καλλιτεχνικό υποκείμενο και τις στρατηγικές με τις οποίες παρεμβαίνει, πλοηγείται και συμμετέχει στις κουλτούρες και τις ροές του διαδικτύου. Σε αυτό το πλαίσιο, η μηχανική των προτροπών (prompt engineering) δεν αποτελεί απλώς μια τεχνική παραγωγής περιεχομένου, αλλά μια νέα μορφή κοινωνικού ανταγωνισμού και μέρος της ίδιας της καλλιτεχνικής εργασίας. Καθώς η γενικευμένη χρήση της παραγωγικής ΤΝ καθιστά ρευστά τα όρια ανάμεσα στο περιεχόμενο που παράγεται από ανθρώπους και σε εκείνο που παράγεται αλγοριθμικά, οι εντάσεις που δημιουργούνται ανάμεσα στη λέξη και την εικόνα, στον κόσμο και την αναπαράστασή του αποκτούν σαφή πολιτική διάσταση.

Η **Κατερίνα Αντωνοπούλου** αποδομεί τον μύθο της αντικειμενικότητας της τεχνητής νοημοσύνης, αναδεικνύοντας πώς τα αλγοριθμικά συστήματα μπορούν να αυτοματοποιούν και να ενισχύουν προϋπάρχουσες κοινωνικές προκαταλήψεις. Εστιάζοντας στην αλγοριθμική μεροληψία (AI bias), η συγγραφέας παρουσιάζει πώς εκδηλώνονται φυλετικές, έμφυλες και άλλες διακρίσεις στην αναλυτική και παραγωγική τεχνητή νοημοσύνη, καθώς και σε εφαρμογές της ανθρωπόμορφης ΤΝ. Μέσα από μια σειρά καλλιτεχνικών έργων, αναλύει τους τρόπους με τους οποίους η ψηφιακή τέχνη μπορεί να καταστήσει ορατούς τους εσωτερικούς μηχανισμούς των συστημάτων της τεχνητής νοημοσύνης και τον ρόλο των δεδομένων εκπαίδευσης στη διαμόρφωση των αλγοριθμικά παραγόμενων αποτελεσμάτων. Η συγγραφέας υποστηρίζει ότι οι σύγχρονες καλλιτεχνικές πρακτικές υπερβαίνουν την απλή επισήμανση των σφαλμάτων, συμβάλλοντας στην αναζήτηση εναλλακτικών τεχνολογικών προσεγγίσεων, όπως η χρήση μικρών συνόλων δεδομένων. Το κείμενο επισημαίνει ότι δεν υπάρχει ένας και μοναδικός τρόπος σχεδιασμού, ανάπτυξης και εκπαίδευσης της ΤΝ και ανοίγει τη συζήτηση για τις προϋποθέσεις ενός πιο συμπεριληπτικού τεχνολογικού μέλλοντος.

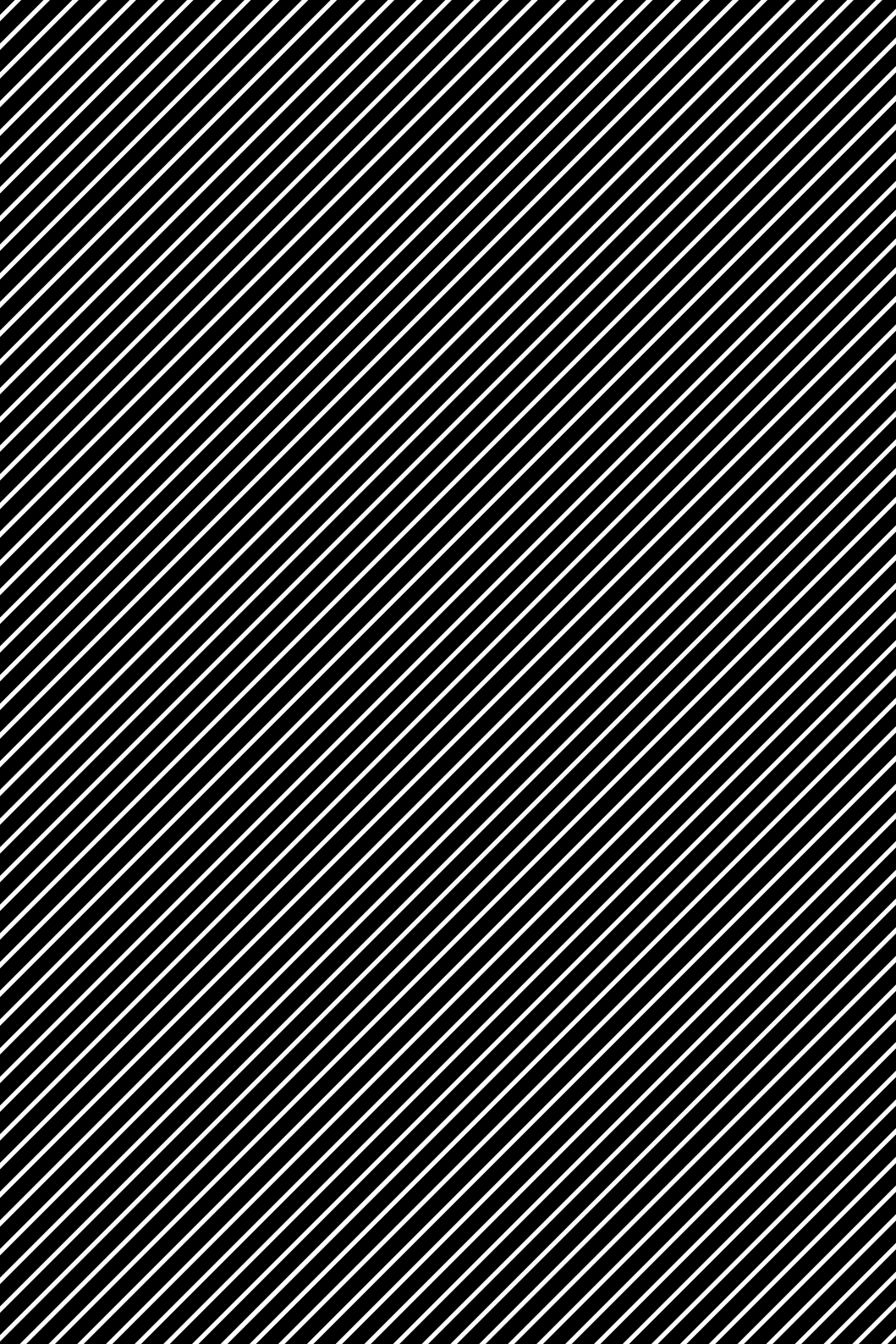
Στο κείμενό του, ο **Νίκος Ερηνάκης** αναλύει τις έννοιες της αυθεντικότητας και της δημιουργικότητας στο αλγοριθμικά διαμορφωμένο τοπίο της τεχνητής νοημοσύνης και της ψηφιακής υπερπραγματικότητας. Ο Ερηνάκης συνδέει τις δύο έννοιες μέσα από το σχήμα «αυθεντικότητα-ως-δημιουργικότητα», δίνοντας έμφαση στη διαρκή αυτο-δημιουργία της εαυτότητας. Σε ατομικό επίπεδο, προτείνει μια «αισθητική συναυτονομία» της εαυτότητάς μας ως μέσο αντίστασης στην ομογενοποί-

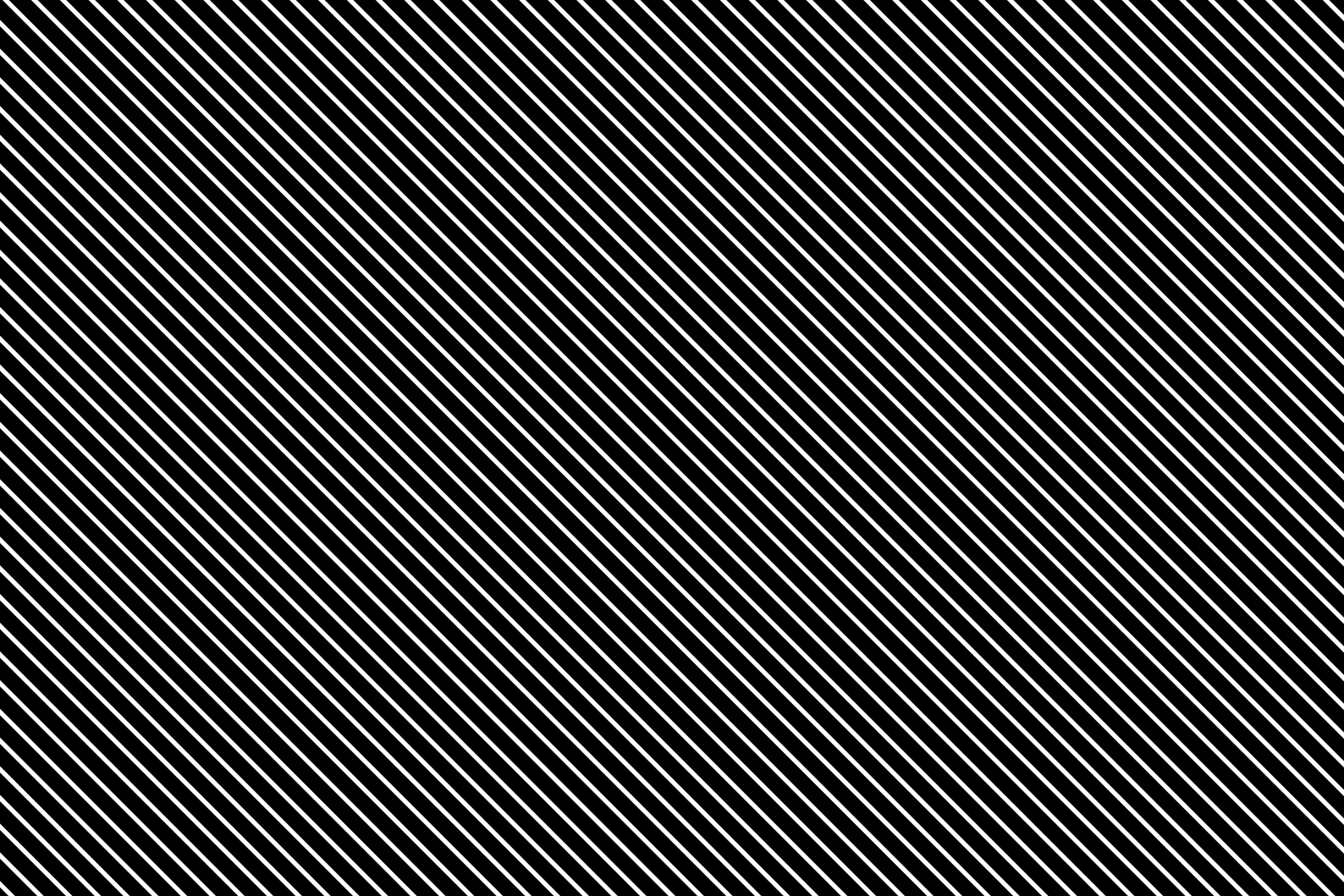
ηση που μπορεί να προκαλέσει η αλγοριθμική διαμεσολάβηση, ενώ σε συλλογικό επίπεδο μιλά για την ανάγκη δέσμευσης σε μια κοινή πραγματικότητα που βασίζεται σε γεγονότα και αποδείξεις. Ο Ερηνάκης διερευνά τη δυνατότητα μιας υβριδικής δημιουργικότητας, όπου η τεχνητή νοημοσύνη μπορεί να λειτουργεί ως δημιουργικό μέσο, με την ανθρώπινη προθετικότητα και κρίση να παραμένουν καθοριστικές. Η αυθεντικότητα αναδεικνύεται έτσι ως μια διαρκής διαδικασία συνδιαμόρφωσης της εαυτότητας και του κόσμου σε διάλογο με το τεχνητό.

Από διαφορετικές αφετηρίες, τα πέντε κείμενα μετατοπίζουν τη συζήτηση από το ερώτημα «τι μπορεί να δημιουργήσει η τεχνητή νοημοσύνη;» στο ερώτημα «τι συμβαίνει όταν η τεχνητή νοημοσύνη γίνεται μέρος της δημιουργικής διαδικασίας;». Ο κοινός άξονας των προσεγγίσεων είναι η δυνατότητα της τέχνης να παρατηρεί και να οικειοποιείται τα υπολογιστικά συστήματα, ασκώντας παράλληλα κριτική στη λειτουργία τους. Η τέχνη μπορεί να εμπλέκεται ενεργά στις υποδομές, στις οικονομίες και στις μορφές ορατότητας που αυτά παράγουν. Για αυτόν τον λόγο, αποτελεί ένα πεδίο όπου οι λειτουργίες τους αποκτούν μορφή, οι αντιφάσεις τους γίνονται αισθητές και οι όροι της τεχνολογικής συνθήκης μπορούν να τεθούν υπό διαπραγμάτευση. Η παρούσα έκδοση επιδιώκει να προσφέρει εννοιολογικά εργαλεία για την κατανόηση της σχέσης της τέχνης με την τεχνητή νοημοσύνη σε ένα τεχνολογικό περιβάλλον που βρίσκεται σε συνεχή μετασχηματισμό.

Σημειώσεις

[1] Crawford, K. (2021). *Atlas of AI*. Yale University Press, σελ. 8.
[2] Zylinska, J. (2024). After AI art. *Widok: Theories and Practices of Visual Culture*, (38). https://www.pismowidok.org/en/archive/2024/38-digital-entanglements/after-ai-art





Τέχνη με τεχνητή νοημοσύνη: πέρα από την παραγωγική τεχνητή νοημοσύνη

Luba Elliott
Επιμελήτρια AI Art

Τι έπεται για την τέχνη με τεχνητή νοημοσύνη; Τώρα που η παραγωγική τεχνητή νοημοσύνη έχει γίνει ένα διαδεδομένο εργαλείο και καθιερωμένη πρακτική για τη δημιουργία εικόνας, ποιες άλλες τεχνικές υπάρχουν διαθέσιμες στους καλλιτέχνες που θέλουν να χρησιμοποιήσουν την τεχνητή νοημοσύνη; Το κείμενο αυτό εξετάζει το ευρύτερο φάσμα των τεχνολογιών τεχνητής νοημοσύνης πέρα από την παραγωγική τεχνητή νοημοσύνη, περιλαμβάνοντας –μεταξύ άλλων– συστήματα αναγνώρισης προσώπων και αντικειμένων, πράκτορες της τεχνητής νοημοσύνης (ευφυή αυτόνομα συστήματα), βιοτέχνη (bioart) και τεχνητή ζωή, καθώς και έργα τέχνης που αξιοποιούν διαδικασίες τεχνητής νοημοσύνης, καταγράφοντας παραδείγματα από τις απαρχές της επανάστασης της βαθιάς μάθησης κατά τη δεκαετία του 2010 έως τις αναδυόμενες καλλιτεχνικές πρακτικές του σήμερα.

Ένας από τους λόγους που οι καλλιτέχνες έχουν αρχίσει να διερευνούν εναλλακτικές προσεγγίσεις είναι η αίσθηση κορεσμού και κόπωσης που έχει αναδυθεί γύρω από την παραγωγική τεχνητή νοημοσύνη. Τα μοντέλα διάχυσης, παρόλο που είναι ισχυρά, συχνά παράγουν παρόμοια αποτελέσματα από αισθητικής άποψης, οδηγώντας σε μια ομογενοποίηση του οπτικού πολιτισμού. Σε αυτό προστίθενται οι αγωγές εναντίον μεγάλων εταιρειών τεχνητής νοημοσύνης για παραβίαση πνευματικών δικαιωμάτων και η άνοδος των αγορών prompts (prompt marketplaces), συμβάλλοντας όλα μαζί στην αίσθηση ότι οι καλλιτέχνες πρέπει να χαράξουν εναλλακτικές στρατηγικές για να δημιουργήσουν έργο που εμπίπτει σε νέα είδη τέχνης. Αυτό παραπέμπει σε παλαιότερες εποχές της ιστορίας της τέχνης, όπως

στον αρχικό ενθουσιασμό με τη βιντεοτέχνη (video art) κατά τη δεκαετία του 1960 ή με τη διαδικτυακή τέχνη (net art) κατά τη δεκαετία του 1990, όταν οι καινοτόμες τεχνολογίες επικράτησαν για ένα μικρό χρονικό διάστημα αλλά σύντομα ενσωματώθηκαν στο καθιερωμένο οπλοστάσιο εικαστικών εργαλείων. Αντιθέτως, οι μη παραγωγικές προσεγγίσεις επιτρέπουν στους καλλιτέχνες να παρεμβαίνουν σε συστήματα που έχουν σχεδιαστεί για την επιτήρηση, το εμπόριο ή την επιστήμη, να τα επαναπροσδιορίζουν ως αισθητικά υλικά και να προσφέρουν μια κριτική άποψη για τη χρήση τους στην κοινωνία.

Για πολλούς, η απαρχή της εποχής της τέχνης με τη χρήση τεχνητής νοημοσύνης κατά τη δεκαετία του 2010 εντοπίζεται στο πρόγραμμα DeepDream, έναν αλγόριθμο τεχνητής νοημοσύνης που ανέπτυξε ο Alexander Mordvintsev. Βασισμένος σε ένα συνελικτικό νευρωνικό δίκτυο (CNNs), ο αλγόριθμος ανιχνεύει ορισμένα χαρακτηριστικά μιας εικόνας και τα τονίζει με αποτέλεσμα να δημιουργούνται πολύχρωμες ψυχεδελικές εικόνες με παράξενα όντα, όπως σκουλήκια κουταβιών, που αναδύονται μέσα από κάποιες κατά τα άλλα συνηθισμένες σκηνές. Παρεμπιπτόντως, αυτό το πρότζεκτ προέκυψε από την έρευνα του Mordvintsev πάνω στην ανάπτυξη ενός νέου μοντέλου για τον διαγωνισμό ImageNet.

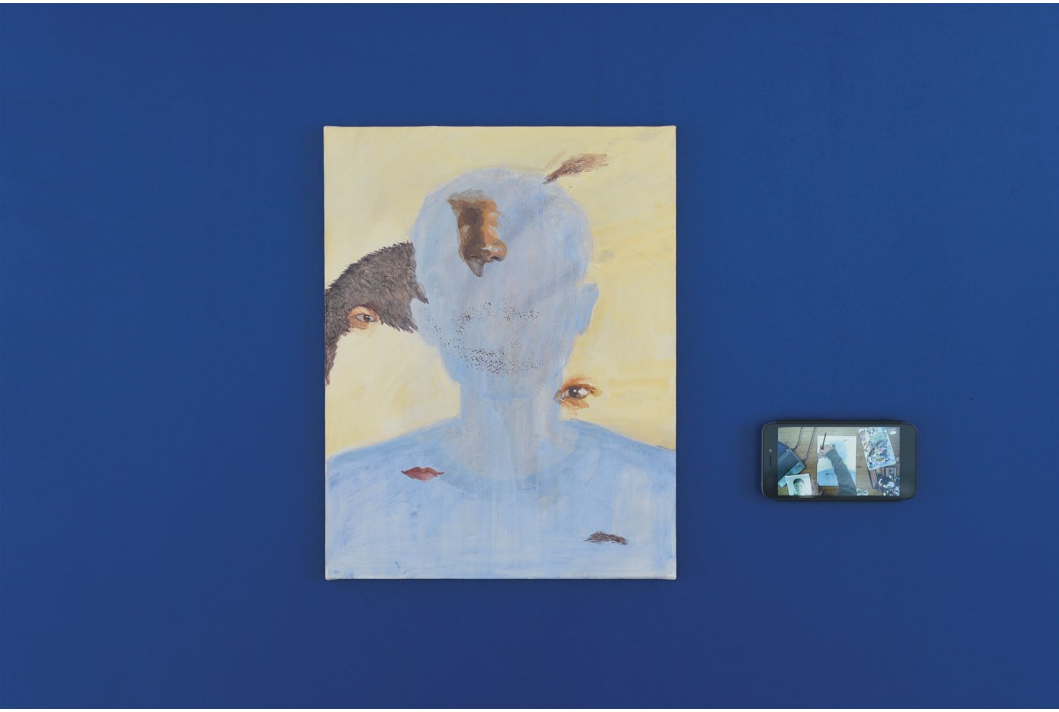
Η κυκλοφορία του DeepDream ήταν καθοριστική ως προς το πώς η τέχνη με τεχνητή νοημοσύνη εισχώρησε στη φαντασία του κόσμου. Οι παράξενες, παραισθησιογόνες εικόνες διαδόθηκαν πολύ γρήγορα στο διαδίκτυο, έγιναν μιμίδια (memes) και πυροδότησαν πολλές δημόσιες συζητήσεις σχετικά με το «τι βλέπουν οι μηχανές». Το ότι αυτές οι εικόνες έγιναν viral και διαδόθηκαν παντού βοήθησε στο να παγιωθεί η άποψη ότι η τεχνητή νοημοσύνη είναι μια δημιουργική δύναμη, παρόλο που ο αλγόριθμος δεν είχε αρχικά σχεδιαστεί για να χρησιμοποιηθεί στην καλλιτεχνική πρακτική. Τόσο οι καλλιτέχνες όσο και οι τεχνολόγοι άρχισαν να εξετάζουν τον πολιτισμικό αντίκτυπο ενός συστήματος που μπορούσε να αποκαλύψει κρυφές πτυχές της αντίληψης, συνδέοντας την αισθητική της μηχανικής όρασης με τις παλαιότερες παραδόσεις της ψυχεδελικής τέχνης, του σουρεαλισμού και των εικασιών γύρω από τη μη ανθρώπινη δημιουργικότητα.

Στον τομέα της αναγνώρισης αντικειμένων, μια καθοριστική εξέλιξη ήταν το λανσάρισμα του συνόλου δεδομένων ImageNet, με

δεκατέσσερις εκατομμύρια εικόνες και είκοσι χιλιάδες κατηγορίες, καθώς και του αντίστοιχου ImageNet Large Scale Visual Recognition Challenge της Fei-Fei Li το 2010, τα οποία ενθάρρυναν την ανάπτυξη αλγορίθμων που είναι σχεδιασμένοι για να επιτυγχάνουν υψηλής ακρίβειας αναγνώριση εικόνας. Κατά τη δεκαετία του 2010, υπήρξε ταχεία πρόοδος στην ανάπτυξη αποτελεσματικών μοντέλων, με το συνελικτικό νευρωνικό δίκτυο AlexNet να επιτυγχάνει ποσοστό αναγνώρισης 84,7% στις πέντε κορυφαίες κατηγορίες, ενώ τα επόμενα μοντέλα αύξησαν το ποσοστό αυτό σε πάνω από 95%, ξεπερνώντας ακόμη και τις ανθρώπινες επιδόσεις.

Οι καλλιτέχνες χρησιμοποιούν εδώ και πολύ καιρό εργαλεία αναγνώρισης προσώπων τόσο ως μέσο άσκησης κριτικής όσο και σε καλλιτεχνικό επίπεδο. Το έργο *CV Dazzle* (2010) του Adam Harvey χρησιμοποιούσε διαφορετικές κομμώσεις και σημάδια στο πρόσωπο για να αποφευχθεί η αναγνώριση, ενώ άλλοι πειραματίζονταν με κοσμήματα ή προβολές προσώπων για να παρεμποδίσουν τη βιομετρική καταγραφή. Η ίδια η αναγνώριση προσώπων είχε γίνει ταυτοχρόνως ένας σημαντικός τομέας πολιτικού και ηθικού προβληματισμού και πολλοί καλλιτέχνες χρησιμοποιούσαν το έργο τους ως ένα είδος αντίστασης στη βιομετρική επιτήρηση. Το *ImageNet Roulette* (2019) του Trevor Paglen αποκάλυψε ότι τα σύνολα δεδομένων που χρησιμοποιούνταν για να εκπαιδεύσουν τα συστήματα αναγνώρισης προσώπων ήταν γεμάτα προκαταλήψεις, καθώς έκαναν εσφαλμένες ταξινομήσεις και χρησιμοποιούσαν παρωχημένες κατηγορίες ταξινόμησης – διαιωνίζοντας έτσι επιβλαβή στερεότυπα. Η Hito Steyerl ανέδειξε μέσω των διαλέξεων και των περφόρμανς της την εμπλοκή της μηχανικής όρασης με στρατιωτικές και εμπορικές υποδομές. Καθώς το βλέμμα μας κατευθυνόταν ξανά σε αυτά τα συστήματα, οι καλλιτέχνες τόνιζαν τόσο τους τεχνικούς περιορισμούς τους όσο και τους κινδύνους που ενέχει το να αναθέτουμε τις κοινωνικές και πολιτικές αποφάσεις στην αλγοριθμική ταξινόμηση.

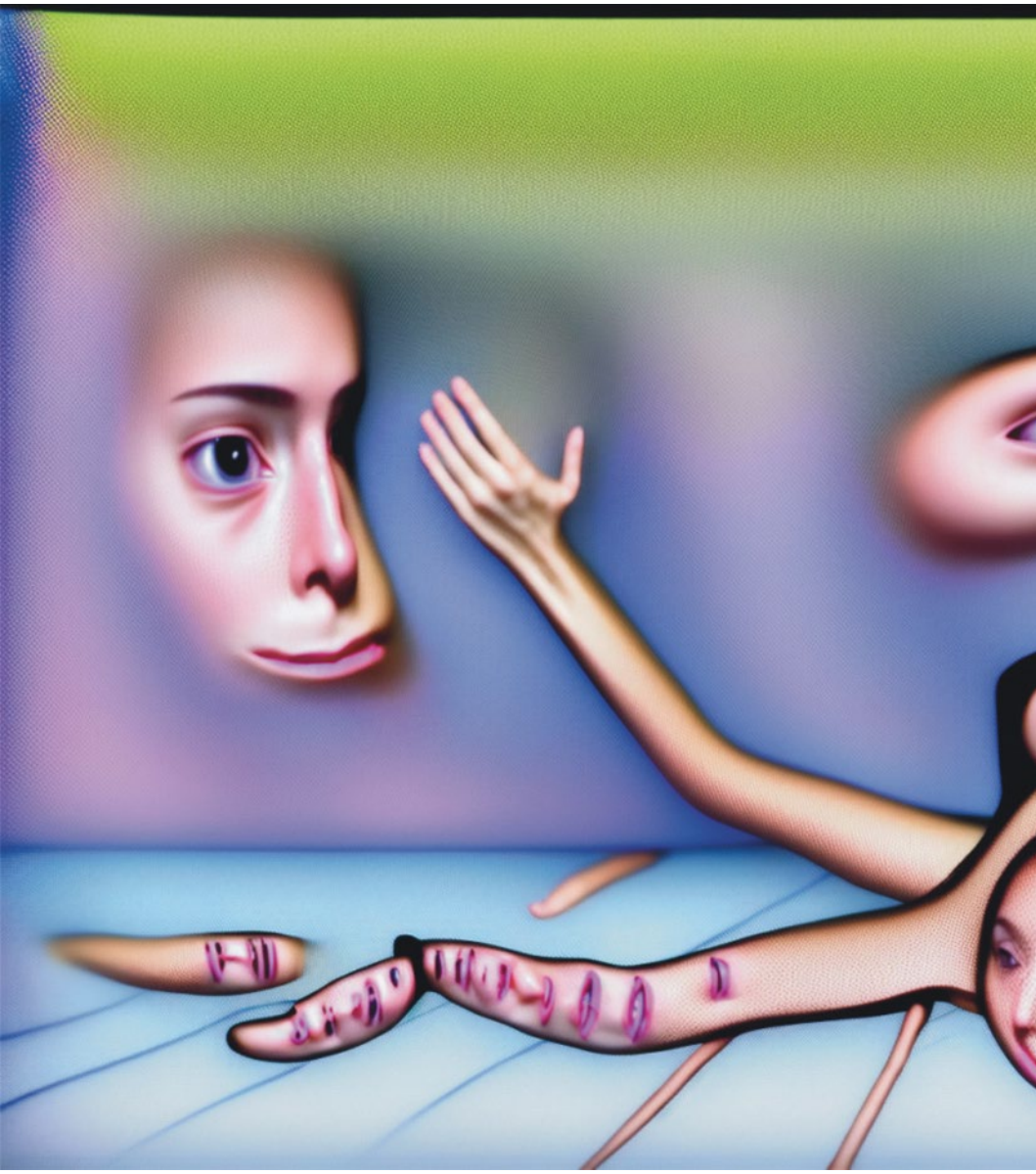
Μερικά έργα χρησιμοποιούν την τεχνολογία με πιο παιγνιώδη τρόπο. Το έργο *Non-Facial Portraits* (2018) των Shinseungback Kimyonghun αφορούσε το σχέδιο ή τη ζωγραφική υπό την επιτήρηση μιας κάμερας αναγνώρισης, τροποποιώντας το πορτραίτο όποτε ανιχνευόταν ένα «πρόσωπο». Ο Dries Depoorter πρότεινε μια εναλλακτική προσέγγιση στο έργο του *Face Detected* (2018-2026), όπου γλύπτες σμίλευαν



Shinseungback Kimyonghun, *Non-Facial Portrait 05*, 2018.
Ευγενική παραχώρηση των καλλιτεχνών

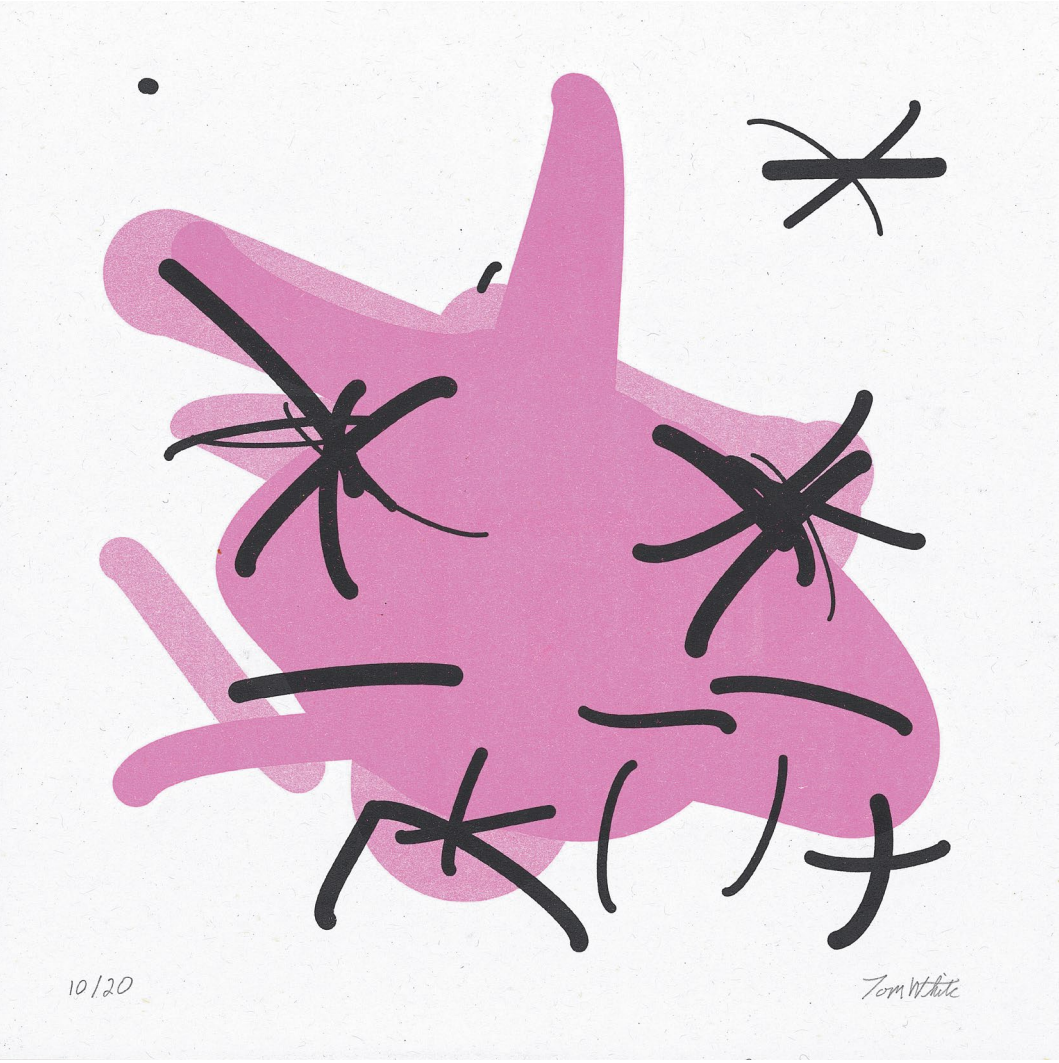
κομμάτια από πηλό μέχρι η μηχανή να αναγνωρίσει ένα πρόσωπο. Αυτή η επαναλαμβανόμενη διερεύνηση της μηχανικής αντίληψης του ανθρώπινου προσώπου σε δύο και τρεις διαστάσεις τονίζει τις λεπτομέρειες των δύο οπτικών: μερικές φορές το μόνο που χρειάζεται είναι μερικές πινελιές σε μια μονόχρωμη επιφάνεια ή η περιστροφή και μετατόπιση των χαρακτηριστικών του προσώπου. Την αντίστροφη λογική μπορεί να τη βρει κανείς στο *Cloud Face* (2012) των Shinseungback Kimyonghun, όπου η τεχνητή νοημοσύνη αναζητά να βρει πρόσωπα σε σύννεφα – μια αναζήτηση νοήματος, οικειότητας και αναγνώρισης σε φυσικά φαινόμενα. Αυτός ο ανθρωποκεντρισμός είναι εμφανής στο έργο *Pareidolia* (2019) των Driessens & Verstappen, όπου εντοπίζονται πρόσωπα σε κόκκους άμμου. Εντωμεταξύ, η Jessica Tucker στο έργο της *cycles* (2024) επεκτείνει τον ρόλο του προσώπου ως οπτικής απεικόνισης για να δημιουργήσει σώματα, παράγοντας μάτια στα άκρα, που υπό άλλες συνθήκες δεν έχουν καμία λεπτομέρεια: για τους ανθρώπους, το πρόσωπο είναι εκείνο που συνήθως φέρει την ταυτότητα και το περισσότερο νόημα.

Η αναγνώριση αντικειμένων έχει επίσης εμπνεύσει δημιουργικές αποκλίσεις, συχνά μέσω των συστημάτων κατηγοριοποίησης που συνδέουν διαφορετικές οπτικές απεικονίσεις ενός αντικειμένου με μια λέξη. Το *Perception Engines* (2018) του Tom White το δείχνει αυτό αρκετά ξεκάθαρα συμπυκνώνοντας την πεμπτουσία μιας συγκεκριμένης κατηγορίας, όπως «αστερίες» ή «λάχανα», σε ένα αφηρημένο σχήμα το οποίο αναγνωρίζεται ως τέτοιο από συστήματα αναγνώρισης εικόνας. Ωστόσο, μερικά από αυτά τα σχήματα, μεταξύ των οποίων και οι αστερίες, μπορεί να μπερδέψουν έναν άνθρωπο που τα παρατηρεί, διότι πολιτισμικά και εμπειρικά έχουμε μάθει να αντιλαμβανόμαστε τους αστερίες ως θαλάσσια όντα με πολλούς βραχίονες και όχι σαν άμορφα φωτεινά ροζ αντικείμενα. Την ικανότητα των συστημάτων αναγνώρισης εικόνας να αναγνωρίζουν λουλούδια την αμφισβητούν οι Shinseungback Kimyonghun με το έργο τους *Flower* (2016-2017), που διερευνά με ποιον τρόπο μπορούν τα λουλούδια να τεμαχιστούν και να παραμορφωθούν αλλά να εξακολουθούν να μπορούν να αναγνωριστούν ως λουλούδια. Τα περιστρεφόμενα αφηρημένα σχήματα θέτουν υπό αμφισβήτηση τον βαθμό στον οποίο οι χρωματικοί συνδυασμοί μπορεί να θεωρηθεί ότι συνδέονται άμεσα με την έννοια των λουλουδιών αντί με τα ίδια τα





Jessica Tucker, *cycles*, 2024.
Ευγενική παραχώρηση της καλλιτέχνιδας



Tom White, *Perception Engines: Starfish*, 2018.
Ευγενική παραχώρηση του καλλιτέχνη

σχήματα των λουλουδιών. Αλλά στο έργο τους *Mou t an* (2020), το περίγραμμα του βουνού είναι εκείνο που το καθιστά αναγνωρίσιμο ως τέτοιο, καθώς οι καλλιτέχνες πειραματίζονται σβήνοντας τμήματα μιας εικόνας για να δοκιμάσουν τα όρια της αναγνώρισης. Ενώ το καλλιτεχνικό δίδυμο από την Κορέα το πετυχαίνει αυτό αφαιρώντας εικονοστοιχεία (pixel), η Caroline Sinders απλώνει λευκή ακρυλική μπογιά σε φωτογραφίες μιας εξέδρας εξόρυξης πετρελαίου στο έργο της *Experiments in Becoming Unreadable* (2024), στο οποίο η καλλιτέχνιδα επιχειρεί να κρύψει αυτά τα παρωχημένα πλέον αντικείμενα υποδομής από τη μηχανική όραση, παρόλο που στην πραγματικότητα η παρουσία τους δεν μπορεί εύκολα να κρυφτεί.

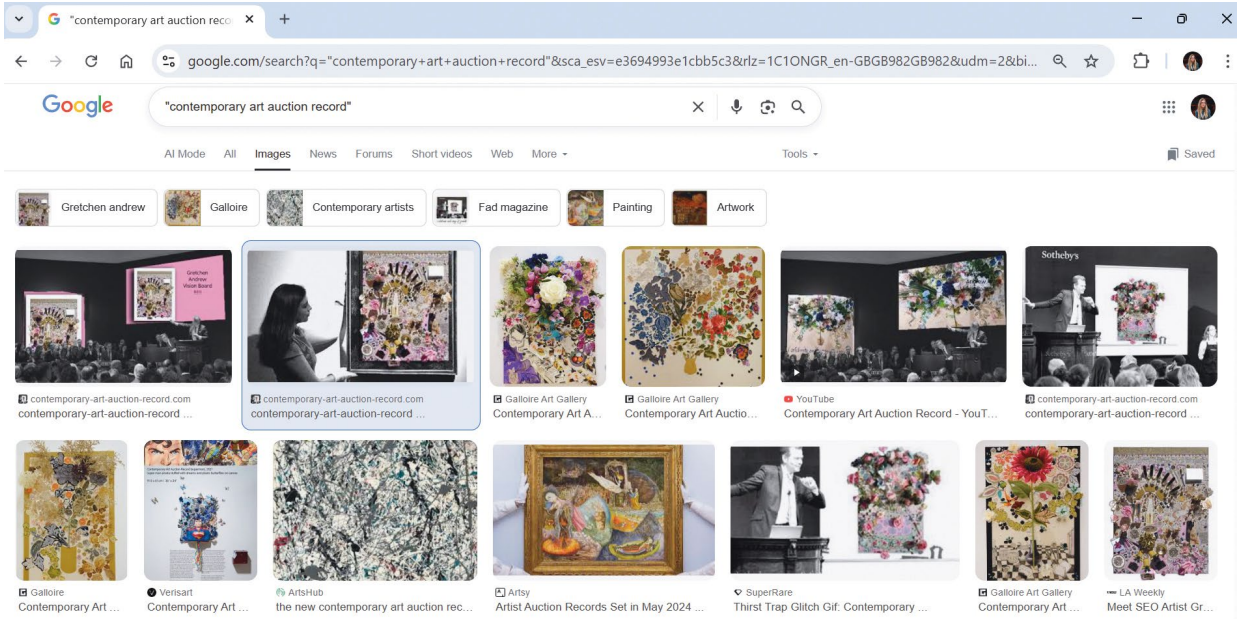
Πέρα από τα αντικείμενα και τα πρόσωπα, η αναγνώριση επεκτείνεται και σε συναισθήματα και συμπεριφορές. Το έργο *Random String of Emotions* (2018) της Coralie Vogelaar διερευνά την ικανότητα μιας ηθοποιού να παράγει συναισθήματα βάσει εντολών που της δίνει η τεχνητή νοημοσύνη, ενώ το *Unlearning Language* (2022) της Lauren Lee McCarthy προτρέπει τους συμμετέχοντες να βρουν νέους τρόπους επικοινωνίας που δεν μπορεί να τους ανιχνεύσει ο αλγόριθμος – με το αποτέλεσμα να είναι το μουρμουρητό, τα χειροκροτήματα και άλλοι ήχοι αντί για τη συμβατική ανθρώπινη ομιλία. Τέτοια έργα υπογραμμίζουν ότι οι αλγοριθμικοί περιορισμοί μπορούν να πυροδοτήσουν νέους τρόπους έκφρασης και επικοινωνίας.

Υπενθυμίζουν επίσης τη σημασία της κατηγοριοποίησης, μιας διαδικασίας κατά την οποία οι εικόνες σχολιάζονται και ταξινομούνται – μια διαβόητα υποκειμενική διαδικασία που ενέχει έντονη μεροληψία. Οι καλλιτέχνες που παρεμβαίνουν σε αυτόν τον τομέα λειτουργούν συχνά και ως κριτικοί και ως μεταφραστές, αναδεικνύοντας πώς τα αδιαφανή τεχνικά συστήματα μπορούν να ενισχύσουν τις κοινωνικές ιεραρχίες, προσφέροντας παράλληλα στρατηγικές αντίστασης, χιούμορ ή ανατροπής στην καθημερινή ζωή. Το έργο *Decision Space* (2016) του Sebastian Schmieg, ανάθεση στον καλλιτέχνη από τη Photographers' Gallery, πρόσφερε αυτήν ακριβώς την ευκαιρία στον τυχαίο παρατηρητή: να κατηγοριοποιήσει περίπου 3.000 εικόνες διάσημων φωτογράφων –μερικές από τις οποίες ήταν ουδέτερες, όπως καναπέδες, και άλλες που είχαν σαφέστερους αρνητικούς υπαινιγμούς, όπως τα όπλα– σε κατηγορίες με τίτλους «το πρόβλημα», «η λύση», «το παρελθόν» και «το μέλλον». Αυτό ανέδειξε την πρόκληση

που ενέχει η κατηγοριοποίηση αντικειμένων σε αφηρημένες κατηγορίες: συχνά δεν υπάρχει ξεκάθαρη απάντηση και η επιλογή είναι πολύ μεροληπτική. Το στοιχείο της προκατάληψης και ο σημαντικός ρόλος που παίζει μερικές φορές η αυθαίρετη κατηγοριοποίηση στη ζωή μας αναδεικνύονται σε έργα όπως το *Britbot* (2018) της Libby Heaney και το *Deutsch/Nicht Deutsch* (2025) των mots, όπου οι ιστορικές ή αθλητικές γνώσεις, ακόμα και η μέθοδος κατηγοριοποίησης μέσω χαρακτηριστικών προσώπου (facial profiling) καθορίζουν την ένταξη σε μια συγκεκριμένη ομάδα – κάτι που δεν έχει απαραίτητα έναν κοινά αποδεκτό ορισμό.

Οι αλγόριθμοι συστάσεων διαμορφώνουν εδώ και χρόνια την εμπειρία μας στο διαδίκτυο με τις μηχανές αναζήτησης, τις δικτυακές πύλες αγορών και τις μουσικές πλατφόρμες. Στη σειρά έργων *Vision Boards* (2017-2021) της Gretchen Andrew βλέπουμε την καλλιτέχνιδα να ανατρέπει τους παραδοσιακούς αλγόριθμους συστάσεων των μηχανών αναζήτησης, δημιουργώντας σελίδες βελτιστοποιημένες για μηχανές αναζήτησης για τους αναλογικούς της πίνακες. Έτσι, τα έργα της εμφανίζονται στην κορυφή της σελίδας αποτελεσμάτων για εικόνες όταν ο χρήστης ψάχνει όρους όπως «ρεκόρ δημοπρασίας σύγχρονης τέχνης». Αυτό το έργο επιδιώκει να υπογραμμίσει τη δύναμη του ατόμου να χακάρει μεγάλα εταιρικά συστήματα, παρουσιάζοντας μια εναλλακτική εκδοχή του κόσμου. Το *Signs of the Times* (2017) των Scott Kelly και Ben Polkinghorne παρουσιάζει διαφημιστικές πινακίδες με εικόνες που προτείνουν άλλους εθνικούς δρυμούς για να επισκεφτεί κανείς, λες και ο επισκέπτης ψωνίζει στο Amazon αντί να απολαμβάνει τον ελεύθερο χρόνο του στη φύση. Αυτό το έργο μάς υπενθυμίζει τον βαθμό στον οποίο μας επηρεάζουν τα συστήματα συστάσεων στην καθημερινότητά μας, τα οποία μας υπαγορεύουν τι να αγοράσουμε, πού να πάμε και τι να κάνουμε. Αυτό το θέμα των μηχανών συστάσεων το διερευνούν περαιτέρω οι Sebastian Schmieg & Silvio Lorusso στο πρότζεκτ τους *Networked Optimization* (2013), το οποίο παρουσιάζει αντίτυπα δημοφιλών βιβλίων αυτοβοήθειας με λευκό κείμενο πάνω σε λευκές σελίδες: οι μόνες ορατές προτάσεις είναι εκείνες που έχουν επισημανθεί περισσότερο στο Kindle, υλοποιώντας τη λογική των μετρήσεων βάσει δημοφιλίας.

Το διακύβευμα στα έργα τέχνης που ασχολούνται με τους αλγόριθμους συστάσεων δεν είναι μόνο η αισθητική της εξατομίκευσης,



Gretchen Andrew, *Vision Boards: Contemporary Art Auction Record*, 2021.
Ευγενική παραχώρηση της καλλιτέχνιδας

αλλά και η πολιτική της αλγοριθμικής διακυβέρνησης. Τα συστήματα που διαμορφώνουν τις επιλογές μας όταν ψωνίζουμε, όταν κάνουμε streaming και όταν κανονίζουμε διαδικτυακά ραντεβού, επηρεάζουν επίσης τις πολιτισμικές επιλογές, περιορίζοντας τη γκάμα αυτών που βρίσκουμε. Ορισμένοι ακαδημαϊκοί, όπως η Shoshana Zuboff, έχουν περιγράψει αυτό το φαινόμενο ως «κατασκοπευτικό καπιταλισμό», όπου τα δεδομένα συμπεριφοράς συλλέγονται για να προβλέψουν και να διαμορφώσουν μελλοντικές ενέργειες. Με το να παρωδούν ή να υπονομεύουν τα συστήματα συστάσεων, οι καλλιτέχνες μάς υπενθυμίζουν ότι αυτές οι αόρατες υποδομές έχουν σημαντικές πολιτισμικές συνέπειες, καθορίζοντας ανεπαίσθητα τι είδους τέχνη, πληροφορίες ή σχέσεις μπαίνουν στη ζωή μας. Το τελευταίο το διερευνά το πρότζεκτ *Browser.Dating* (2025) του Dries Depoorter, που είναι μια εφαρμογή γνωριμιών η οποία συνδέει τους ανθρώπους βάσει του ιστορικού περιήγησής τους στο διαδίκτυο. Αυτό που υποδηλώνουν αυτά τα έργα είναι ότι η ίδια η πολιτισμική εμπειρία γίνεται αντικείμενο αλγοριθμικής διαχείρισης: όπως ακριβώς οι λίστες αναπαραγωγής στο Spotify ή οι προτάσεις που μας κάνει το Netflix μας ωθούν σε συγκεκριμένες διαθέσεις, είδη και ρυθμούς κατανάλωσης, έτσι και τα έργα που μιμούνται ή παρωδούν τα συστήματα συστάσεων τονίζουν ότι ακόμη και οι αισθητικές μας εμπειρίες είναι επιμελημένες με έναν αόρατο τρόπο. Με το να παρουσιάζουν αυτούς τους μηχανισμούς στο διαδίκτυο ή σε δημόσιο χώρο, οι καλλιτέχνες αποκαθιστούν την ορατότητα των υποδομών που συνήθως παραμένουν κρυφές, προτρέποντάς μας να αναλογιστούμε ότι η επιθυμία και η προτίμησή μας δεν είναι πλέον αποκλειστικά δικές μας, αλλά διαμορφώνονται από συστήματα πρόβλεψης.

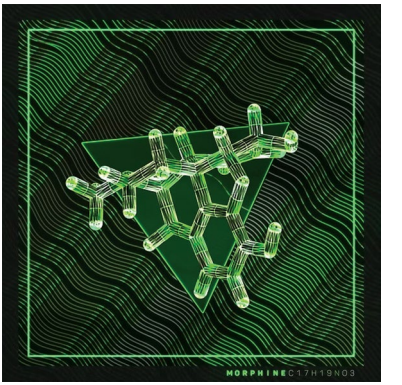
Τα επιστημονικά εργαλεία που δεν είχαν ως αρχικό στόχο τη δημιουργία τέχνης προσφέρουν ένα ακόμη γόνιμο πεδίο προς διερεύνηση. Η κάμερα γεγονότων (event camera), στην οποία ο αισθητήρας απεικόνισης αποκρίνεται μόνο σε τοπικές αλλαγές στη φωτεινότητα, χρησιμοποιήθηκε από τον Justin Urbach και την ομάδα του στο έργο *BLINDHÆD* (2025) για να διερευνηθεί η ανθρώπινη ανάγκη για τη βελτιστοποίηση της όρασης: αυτή η νέα επιστημονική εξέλιξη παρουσιάζει μια καλλιτεχνική ερμηνεία της ενισχυμένης οπτικής αντίληψης. Στο *xSublimatio* (2022), ο Pierre Pauze δημιούργησε μοντέλα φαρμάκων και μορίων από δεδομένα που παρείχε το AlphaFold, το οποίο σχεδιάστηκε αρχικά από ερευνητές στην Google DeepMind για

να προβλέπει τη δομή πρωτεϊνών από αμινοξέα. Στο έργο του Pauze, οι συλλέκτες αγοράζουν NFTs οπτικών απεικονίσεων μορίων, τα οποία όταν συνδυαστούν παράγουν μια νέα εικόνα – ένα νέο φάρμακο. Το ίδιο ενδιαφέρον για τις πιο πρόσφατες τεχνολογικές εξελίξεις και τον πολιτισμικό τους αντίκτυπο στον κόσμο μπορεί επίσης να το δει κανείς σε εννοιολογικό επίπεδο στα έργα καλλιτεχνών όπως ο Lawrence Lek, του οποίου η ταινία *Geomancer* (2017) –που διαδραματίζεται το 2065, τη χρονιά που θα γιορταστούν τα εκατό χρόνια της Σιγκαπούρης– φαντάζεται τις καλλιτεχνικές επιθυμίες ενός δορυφόρου αντλώντας έμπνευση από την πιο σημαντική καινοτομία της DeepMind το 2017: την ανάπτυξη του AlphaGo, ενός συστήματος που μπορεί να κερδίσει τους ανθρώπους στο αρχαίο παιχνίδι Go.

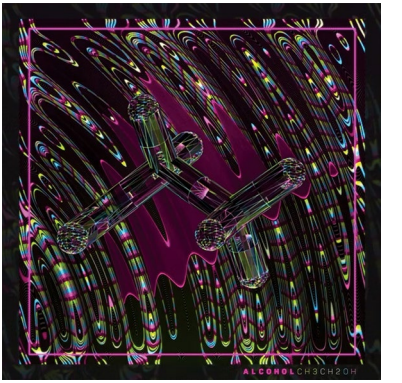
Πέρα από την αξιοποίηση τεχνολογιών και εννοιών που προκύπτουν από τις πιο πρόσφατες τεχνολογικές εξελίξεις, μια συγκεκριμένη ομάδα έργων διερευνά και εξηγεί τις κρυφές διαδικασίες της τεχνητής νοημοσύνης. Το έργο *DIGITAL NO MAN'S* (2023) της Λήδας Ζαχαροπούλου, για παράδειγμα, χρησιμοποιεί το φιλμ ως μέσο για να αποδώσει ανθρώπινα χαρακτηριστικά στις χαμηλής ανάλυσης εικόνες που παράγει ένας γεννήτορας (generator), τις οποίες απορρίπτει ο διακριτής, στο πλαίσιο ενός συστήματος GAN. Αυτή η δημιουργική ερμηνεία της διαδικασίας GAN φωτίζει τον εσωτερικό μηχανισμό του συστήματος, διερευνώντας τα πιθανά συναισθήματα και τα προβλήματα που βιώνουν οι εικόνες που δεν θεωρούνται άρτιες σε επίπεδο pixel. Εντωμεταξύ, το έργο *Diffusion TV* (2025) του Sihwa Park παρουσιάζει μια νοσταλγική CRT TV (τηλεόραση καθοδικού σωλήνα) με διαδραστική διεπαφή, δίνοντας τη δυνατότητα στο κοινό να πειραματιστεί με τα μοντέλα διάχυσης και να αποκαλύψει την κρυφή διαδικασία αποθορυβοποίησης μέσω τεχνητής νοημοσύνης, με το να ρυθμίζει την κεραία και τους διακόπτες της τηλεόρασης ώστε να αλλάξει τα χαοτικά δεδομένα και να τα μετατρέψει σε αναγνωρίσιμες εικόνες, προσφέροντας έτσι την ευκαιρία στο ανθρώπινο χέρι να συνδράμει στη διαδικασία της τεχνητής νοημοσύνης.

Σήμερα, οι πράκτορες τεχνητής νοημοσύνης (AI agents) είναι ένα πολύ σημαντικό πεδίο έρευνας, καθώς εκτιμώνται για την ικανότητά τους να επιτυγχάνουν στόχους και να ολοκληρώνουν εργασίες. Στην τέχνη, αυτό το βλέπουμε σε αποκεντρωμένα αυτόνομα καλλιτεχνικά πρότζεκτ όπως το *Abraham* (2019-σήμερα) του Gene Kogan και το *Botto*

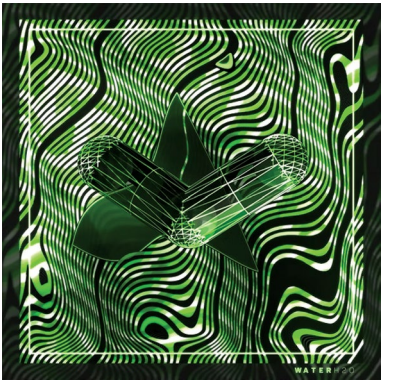












Pierre Pauze, *xSublimatio,* 2022. Ευγενική παραχώρηση του καλλιτέχνη

(2021-σήμερα) του Mario Klingemann, ενός καλλιτεχνικού πράκτορα που παράγει εικόνες μέσω υποδείξεων (prompts), βελτιώνοντας το ύφος του μέσω ψήφων από την κοινότητα. Αυτή η χρήση αυτόνομων πρακτόρων στην τέχνη εγείρει ερωτήματα σχετικά με την έννοια του δημιουργού και της συνεργασίας. Όταν το *Botto* παράγει εικόνες τις οποίες τις έχει ψηφίσει η κοινότητα, ποιος είναι ο «καλλιτέχνης» – η μηχανή που παράγει τέχνη, ο άνθρωπος που τη δημιούργησε ή συλλογικά το κοινό που καθοδηγεί την εξέλιξη του *Botto*; Αυτά τα ερωτήματα παραπέμπουν σε παλαιότερα πειράματα με κατανεμημένη δημιουργικότητα, όπως τα δίκτυα επιστολικής τέχνης (mail art) της δεκαετίας του 1960 ή οι πρώτες διαδικτυακές καλλιτεχνικές κοινότητες, όπου ο δημιουργός ως άτομο επισκιαζόταν προς όφελος των συλλογικών διαδικασιών. Με την αναβίωση αυτών των συζητήσεων στο πλαίσιο της τεχνητής νοημοσύνης, η τέχνη που είναι βασισμένη σε πράκτορες (agent-based art) αναδεικνύει ότι η δημιουργικότητα είναι μια δικτυωμένη δραστηριότητα που είναι προϊόν διαπραγμάτευσης και όχι το προϊόν ενός μόνο ατόμου.

Ως αντίδραση στον ψηφιακό κορεσμό, μερικοί καλλιτέχνες στράφηκαν στη βιοτέχνη. Χαρακτηριστικό παράδειγμα αυτού του τομέα είναι το έργο *cellF* (2016) του Guy Ben-Ary, το πρώτο στον κόσμο νευρωνικό συνθεσάιζερ (neural synthesiser) που είναι φτιαγμένο από το βιολογικό νευρωνικό δίκτυο που αναπτύχθηκε από κύτταρα του ίδιου του καλλιτέχνη. Σε αντίθεση με τα νευρωνικά δίκτυα που βασίζονται σε λογισμικό, το πρότζεκτ του Ben-Ary χρησιμοποιεί ένα ανθρώπινο κύτταρο για να ελέγξει ένα ειδικά κατασκευασμένο αρθρωτό συνθεσάιζερ: μόνο αναλογικά κυκλώματα και βιολογικό υλικό ενεργοποιούν το όργανο. Αυτά είναι βασισμένα σε παλαιότερα έργα βιοτέχνης, όπως το *GFP Bunny* (2000) του Eduardo Kac, δείχνοντας πώς τα ζωντανά συστήματα εισχωρούν στον χώρο της τέχνης και στην πορεία θολώνουν τα όρια μεταξύ φυσικού και τεχνητού, ανθρώπινου και μη ανθρώπινου, μηχανής και οργανισμού. Προκαλούν επίσης σκέψεις όσον αφορά την ηθική: τι ευθύνες φέρουν οι καλλιτέχνες όταν δουλεύουν με ζωντανή ύλη, είτε πρόκειται για νευρώνες σε τρυβλίο είτε για ρομποτικά πλάσματα που απελευθερώνονται στο περιβάλλον; Ο συνδυασμός τεχνητής νοημοσύνης και βιοτέχνης εγείρει περαιτέρω ερωτήματα όσον αφορά τα όρια της παρέμβασης, της δημιουργικότητας και της ίδιας της ζωής.



Λήδα Ζαχαροπούλου, *Digital No-Man's*, 2023.
Ευγενική παραχώρηση της καλλιτέχνιδας

Παράλληλα με τη βιοτέχνη, η ρομποτική προσφέρει μια ευπρόσδεκτη απόδραση στον υλικό κόσμο. Το έργο *Danger, Squirrel Nutkin!* (2009) του Ian Ingram παρουσιάζει ένα ρομπότ με πολλές ουρές σκίουρου, που προειδοποιεί πραγματικούς σκίουρους για την ύπαρξη θηρευτών που αντιλαμβάνεται, χρησιμοποιώντας το δικό τους προειδοποιητικό σήμα με το κούνημα της ουράς. Αυτή η αλληλεπίδραση μεταξύ ρομπότ και φυσικού περιβάλλοντος εγείρει πολλαπλά ερωτήματα σχετικά με το τι συνιστά έναν σκίουρο και με τον ρόλο που μπορεί να παίζουν αυτά τα τεχνητά όντα στον κόσμο του μέλλοντος. Παρομοίως, η Paula Gaetano κατασκεύασε για το έργο της *Guanaquerx* (2024) ένα ρομπότ εξερεύνησης της ξηράς, με πηγή έμπνευσης το γκουανάκο, το θηλαστικό που ζει στις Άνδεις. Αυτό το ρομπότ, που το δημιούργησε μια συλλογικότητα καλλιτεχνών και μηχανικών, διέσχισε κορυφές ύψους 4.500 μέτρων στις Άνδεις, ένα επίτευγμα με πηγή έμπνευσης τη Διάβαση των Άνδεων που είχε γίνει το 1817, προσφέροντας ένα παράδειγμα συλλογικής προσπάθειας και τοπικών κοσμοτεχνικών για να ξαναγίνουν οι Άνδεις ένας τόπος χειραφέτησης και αποαποικιοποίησης. Αυτές οι αναζητήσεις διασταυρώνονται με το ολοένα και αυξανόμενο ενδιαφέρον για τις οικολογικές και μετα-ανθρώπινες διαστάσεις της τέχνης. Στην εποχή της κλιματικής κρίσης, τέτοια έργα μάς ωθούν να φανταστούμε μορφές συνύπαρξης με τις μηχανές, τα ζώα και τα περιβάλλοντα, οι οποίες βασίζονται λιγότερο στην εκμετάλλευση και περισσότερο στη συνεργατικότητα, νηματοδοτώντας εκ νέου την τεχνητή νοημοσύνη και τη ρομποτική όχι ως εργαλεία κυριαρχίας, αλλά ως εταίρους στην οικολογική επιβίωση.

Όσον αφορά την τεχνητή ζωή, τεχνικές όπως οι γενετικοί αλγόριθμοι και τα κυτταρικά αυτόματα έχουν χρησιμοποιηθεί εδώ και χρόνια από μια ευρεία γκάμα καλλιτεχνών, από τον τομέα της παραδοσιακής τέχνης με υπολογιστές (computer art) μέχρι την κοινότητα της παραγωγικής NFT τέχνης. Πρόκειται για βιολογικούς αλγόριθμους που είχαν φτιαχτεί για άλλους λόγους, αλλά βρήκαν εφαρμογή ως δημιουργικό υλικό όταν τους χρησιμοποίησαν καλλιτέχνες όπως ο Alexander Mordvintsev και ο Harm van den Dorpel. Το έργο *Hexells* (2020) του Alexander Mordvintsev λειτουργεί σαν ένα σύστημα κυττάρων που αυτοοργανώνεται και είναι εκπαιδευμένο να δημιουργεί υφές βασιζόμενο αποκλειστικά στην επικοινωνία μεταξύ των γειτονικών κυττάρων. Αυτό το έργο αναδεικνύει τον ρόλο ενός μεμονωμένου κυττάρου μέσα σε μια μεγαλύτερη ομάδα κυττάρων. Ο Harm van den





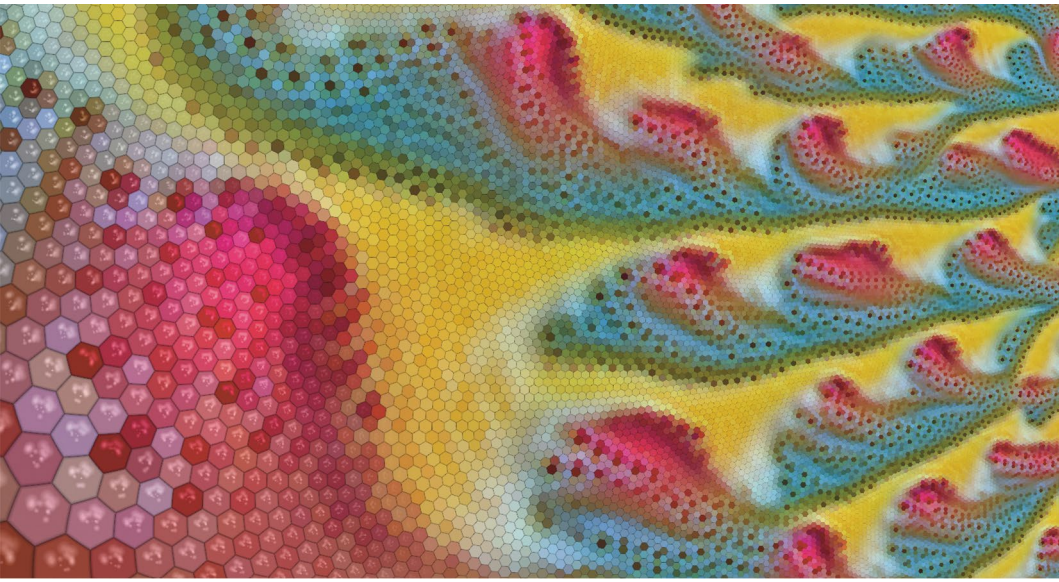
Guy Ben-Ary, *cellF*, 2016. Άποψη εγκατάστασης, Cell Block Theatre, Σίδνεϊ.
Ευγενική παραχώρηση του καλλιτέχνη. Φωτογραφία: Alex Davies

Dorpel δούλεψε για ένα διάστημα με συστήματα όπως οι γενετικοί αλγόριθμοι στο *Hybrid Vigor* (2017). Σε αυτό βλέπουμε τον καλλιτέχνη να χρησιμοποιεί τη συνάρτηση καταλληλότητας (fitness function) της ανατροφοδότησης του κοινού για να δημιουργήσει νέα αφηρημένα σχέδια με χελώνες, τα οποία πολλαπλασιάζονται γεμίζοντας την οθόνη και επιτρέπουν στο κοινό να επιλέξει δύο από αυτά ως γονείς. Αυτά με τη σειρά τους γεννούν παιδιά και η διαδικασία συνεχίζεται ώσπου τα σχήματα και τα χρώματα συγκλίνουν και δημιουργείται ένα εξαιρετικό έργο τέχνης, το οποίο στη συνέχεια εκτυπώνεται.

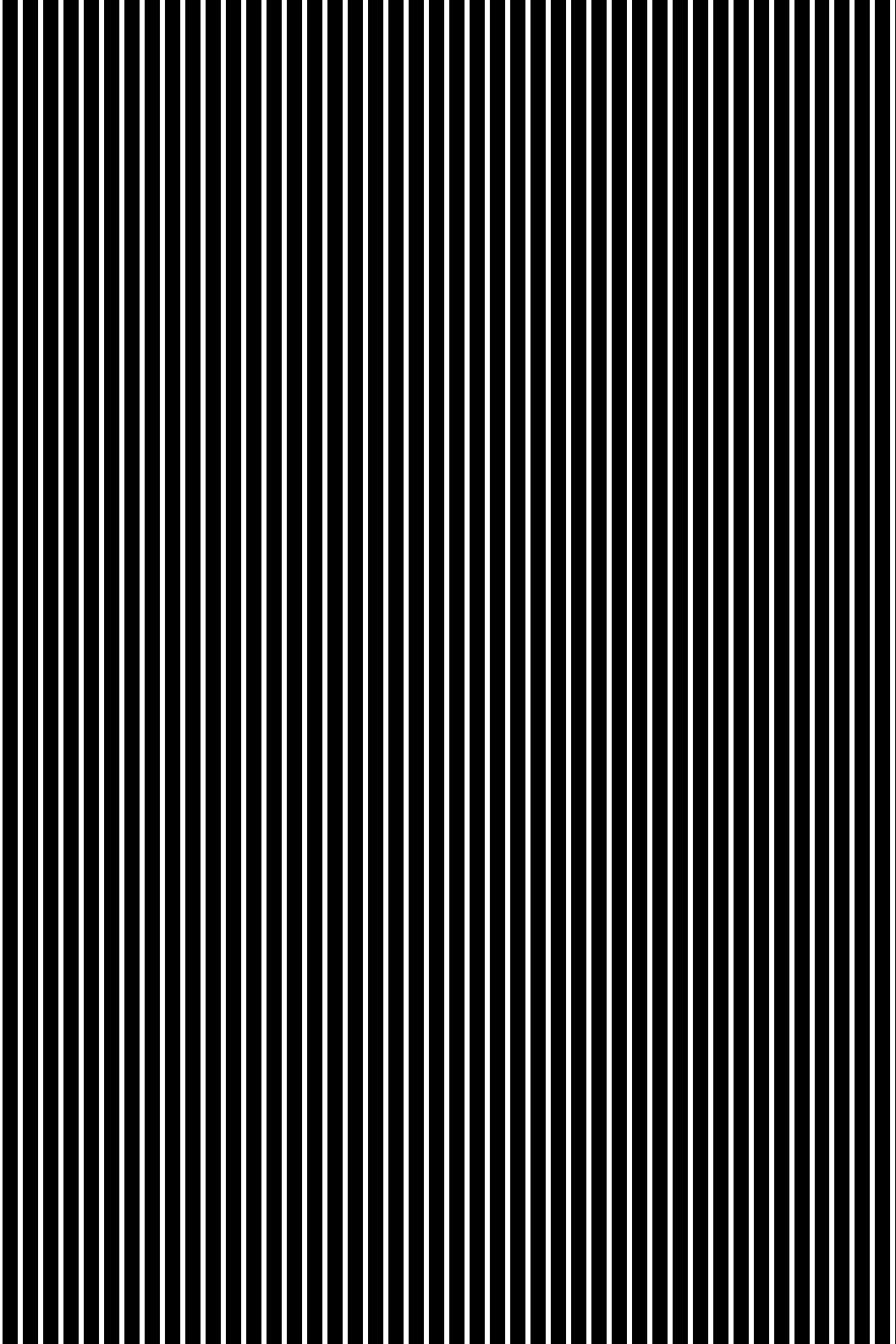
Όλα αυτά τα έργα αποδεικνύουν ότι οι καλλιτέχνες επαναχρησιμοποιούν τα συστήματα τεχνητής νοημοσύνης με αντισυμβατικούς τρόπους: χακάρουν υποδομές, εξερευνούν τα όρια της τεχνολογίας, αποκαλύπτουν κρυφές παραδοχές, ενσωματώνουν ζωντανή ύλη ή αποδομούν τα θεμέλια αυτών των πολυχρηστικών συστημάτων τεχνητής νοημοσύνης. Με αυτόν τον τρόπο, οι καλλιτέχνες έχουν λιγότερο ανταγωνισμό και περισσότερες ευκαιρίες όταν αρχίζουν να πειραματίζονται με συστήματα που δεν είχαν εξαρχής σχεδιαστεί για την παραγωγή πολυμέσων. Τελικά, είναι πιο εύκολο να καλλιεργήσει κανείς ασυνήθιστες μορφές δημιουργικότητας όταν το εργαλείο με το οποίο δουλεύει είναι περιορισμένο ως προς τις συμβατικές δυνατότητες των εικαστικών τεχνών. Αυτοί οι περιορισμοί μετατρέπουν τους φραγμούς σε ευκαιρίες για νέες μορφές έκφρασης και προοπτικές που αντιστέκονται στην τεχνολογική ομογενοποίηση.

Στο μέλλον, η πρόκληση για τους καλλιτέχνες θα είναι να διασφαλίσουν ότι οι δικές τους ξεχωριστές δημιουργικές φωνές θα συνεχίσουν να ακούγονται μέσα στο τσουνάμι της ολοένα και πιο τέλειας μαζικής παραγωγής τεχνητής νοημοσύνης (AI slop). Τελικά, το μέλλον της τέχνης με τεχνητή νοημοσύνη μπορεί να εξαρτάται λιγότερο από την τεχνολογική καινοτομία και περισσότερο από την ικανότητα των καλλιτεχνών να φανταστούν εναλλακτικές χρήσεις των υποδομών υπολογισμού. Καθώς η τεχνητή νοημοσύνη γίνεται ολοένα και περισσότερο ενσωματωμένη και αόρατη –στα πάντα, από τις ιατρικές διαγνώσεις έως την εξυπηρέτηση πελατών– ο καλλιτεχνικός στόχος είναι να κάνει αυτές τις κρυφές διαδικασίες ορατές και να προσφέρει τις μοναδικές φωνές των δημιουργών μέσα από παιγνιώδεις παρεμβάσεις, κριτική παραποίηση συστημάτων ή συνεργασίες με ζωντανά συστήματα. Με αυτόν τον τρόπο, οι καλλιτέχνες μπορούν να ανακτήσουν μέρος της δημιουργικής

τους κυριότητας σε ένα τεχνολογικό τοπίο όπου επικρατούν οι εταιρικές πλατφόρμες, θυμίζοντάς μας ότι δημιουργικότητα δεν σημαίνει μόνο δημιουργία νέων εικόνων αλλά και αναδιαμόρφωση των όρων της αντίληψης, της αλληλεπίδρασης και του νοήματος.



Alex Mordvintsev, *Hexells,* 2020. Ευγενική παραχώρηση του καλλιτέχνη



Το δίκοπο μαχαίρι της οικειοποίησης της τεχνολογίας

Rosemary Lee
Καλλιτέχνις, Επισκέπτρια Επίκουρη Καθηγήτρια,
Σχολή Μηχανικών, Πανεπιστήμιο του Πόρτο

Η ηθική μεταβλητότητα της τεχνολογίας τής προσδίδει τη δυνατότητα να λειτουργεί σαν δίκοπο μαχαίρι – μπορεί εξίσου εύκολα να υπηρετήσει ή να στραφεί ενάντια στις προθέσεις όσων την κρατούν στα χέρια τους. Οι καλλιτεχνικές πρακτικές που ασκούν κριτική και χρησιμοποιούν ενεργά τις αναδυόμενες τεχνολογίες δεν αποτελούν εξαίρεση – μπορούν εξίσου εύκολα να λειτουργήσουν ανατρεπτικά ή να εγκλωβιστούν μέσα σε αυτές. Οι απόηχοι της τεχνολογίας δεν είναι εύκολα κατανοητοί μέσω μονοδιάστατων αντιλήψεων περί τεχνοηθικής, ούτε αναδεικνύουν απλές και ξεκάθαρες λύσεις – λύσεις που δεν εξαντλούνται απλώς στην επισήμανση ή στην αποκατάσταση των ενδεχόμενων συνεπειών της τεχνολογίας.

Οι δημιουργικές αναζητήσεις με τη βοήθεια της τεχνητής νοημοσύνης έχουν πρόσφατα αναδείξει μια σειρά από σύνθετα ζητήματα που άπτονται των λεπτών αποχρώσεων αυτού του θέματος, ανακαλώντας παράλληλα πτυχές παρόμοιων μετατοπίσεων που έχουν προκύψει στο παρελθόν, όταν οι τεχνολογίες είχαν επιφέρει ακούσιες συνέπειες στην επιδίωξή τους να υλοποιήσουν τον στόχο τους. Οι προσπάθειες δημιουργικής παρεκτροπής από τη σκοπούμενη χρήση αυτών των εργαλείων ενδέχεται να οδηγήσουν σε εναλλακτικές λειτουργίες, αποκαλύπτοντας νέους τρόπους προσέγγισης τόσο των ίδιων των εργαλείων όσο και των θεμελιωδών παραδοχών που ενσωματώνουν στον σχεδιασμό τους. Ωστόσο, αυτού του είδους οι ανατρεπτικές παρεμβάσεις υπόκεινται στην ίδια αμφισημία που τις καθιστά εφικτές. Η οικειοποίηση νέων εργαλείων και μεθόδων ενδέχεται να αποσταθεροποιήσει τις καθιερωμένες πολιτισμικές συμβάσεις, διαμορφώνοντας νέες μεθόδους, μια νέα αισθητική προσέγγιση και νέους εννοιολογικούς συλλογισμούς όσον αφορά τις μετατοπίσεις που

επιφέρει η τεχνολογία και τον ευρύτερο αντίκτυπό της στην κοινωνία. Στις πρόσφατες διερευνήσεις για τη χρήση της τεχνητής νοημοσύνης στην τέχνη, ούτε η συνολική απόρριψη της τεχνολογίας αυτής ούτε η προσπάθεια ελέγχου της επαρκούν για να συμφιλιώσουν εύκολα τις εσωτερικές αντιφάσεις της.

Οι προσπάθειες κατανόησης και απόκρισης στην αναδυόμενη δυναμική της τεχνητής νοημοσύνης έχουν αναδείξει πλήθος αλληλένδετων ζητημάτων που πηγάζουν από τις διαφορετικές πτυχές της σχέσης αυτής της τεχνολογίας με την καλλιτεχνική πρακτική. Οι νέες τεχνολογικές εξελίξεις, καθώς αποσταθεροποιούν τις μεθόδους, τα εργαλεία και τα συστήματα αξιών της τέχνης, υπονομεύουν την προσήλωση σε αυστηρά καθιερωμένες συμβάσεις που αφορούν τον ρόλο της τεχνολογίας στον πολιτισμό. Η τεχνητή νοημοσύνη είναι σαν δίκοπο μαχαίρι, έχει και θετικές και αρνητικές συνέπειες και λειτουργεί ταυτοχρόνως τόσο υπέρ όσο και κατά των ίδιων των στόχων της. Αυτός ακριβώς είναι ένας από τους θεμελιώδεις κινδύνους της τεχνητής νοημοσύνης –όπως και κάθε τεχνολογίας στην ουσία– αλλά προσφέρει επίσης και μια δίοδο για κριτική παρέμβαση.

> Όποιο όπλο και αν θέλετε να χρησιμοποιήσετε, μόλις το οικειοποιηθούν οι άλλοι, στρέφεται πάντα εναντίον σας. (Vaneigem, 1967, 9)

Ο καταστασιακός συγγραφέας Raoul Vaneigem επισήμανε την εύθραυστη φύση των εργαλείων αντίστασης, τονίζοντας ότι υπάρχει εξίσου η δυνατότητα οικειοποίησής τους όσο και επανοικειοποίησης. Τα όπλα στα οποία αναφέρεται στην προκειμένη περίπτωση αφορούν πράξεις πολιτικής και καλλιτεχνικής αντίστασης στην Ευρώπη κατά τα τέλη της δεκαετίας του 1960, πολύ πριν αντιμετωπιστεί η τεχνητή νοημοσύνη με ανάλογους όρους, ωστόσο αυτή η δήλωσή του φαίνεται να επαληθεύεται από την ίδια τη λογική της οικειοποίησης που αναδεικνύει. Αν ερμηνεύσουμε αυτά τα λόγια έξω από το αρχικό τους πλαίσιο, μπορούμε να καταλάβουμε καλύτερα την εύθραυστη και αντιφατική σχέση που προκύπτει όταν ενσωματώνεται η τεχνολογία στον πολιτισμό, επισημαίνοντας πιθανές κατευθύνσεις ως προς τον τρόπο με τον οποίο μπορεί να χρησιμοποιηθεί η τεχνητή νοημοσύνη στην τέχνη. Οι καλλιτεχνικές παρεμβάσεις με τεχνολογικά μέσα χαρακτηρίζονται από την αλληλεπίδραση μεταξύ δημιουργικής κατάχρησης, αποχής και υπονόμευσης, συχνά έχοντας ως στόχο να

αμφισβητήσουν τις υφιστάμενες συνθήκες μέσω της επινόησης νέων μεθόδων πράξης, δημιουργίας και στοχασμού. Αλλά ενώ αυτές οι στρατηγικές ενδέχεται πράγματι να αποδειχτούν αποτελεσματικές, προσφέροντας στους καλλιτέχνες τη δυνατότητα να ανταποκριθούν στα συνεχώς μεταβαλλόμενα πλαίσια μέσα στα οποία δημιουργούν, τα λόγια του Vaneigem μας υπενθυμίζουν ότι ακόμα και αυτές υπόκεινται σε πιθανή μεταστροφή (*détournement*)[1] των εργαλείων εναντίον τους. Παρόλο που η τεχνολογία μπορεί να μην είναι απρόσβλητη από τις απόπειρες σαμποτάζ, αυτές ακριβώς οι ανατρεπτικές στρατηγικές μπορεί να υπονομευτούν από άλλες προθέσεις.

Ο πρόσφατος ενθουσιασμός των καλλιτεχνών με την τεχνητή νοημοσύνη εγείρει γνώριμα ερωτήματα σχετικά με την οικειοποίηση της τεχνολογίας στην τέχνη – και αυτό είναι απολύτως επίκαιρο διότι μπορεί να αλλάξει αυτά που θεωρούσαμε ότι γνωρίζαμε ήδη πάνω στο θέμα. Τα καινούρια, ισχυρά παραγωγικά εργαλεία της τεχνητής νοημοσύνης έχουν γίνει η αφορμή για να ανοίξει μια πολυδιάστατη συζήτηση σχετικά με τον ρόλο της τεχνολογίας στον πολιτισμό, αλλά και όσον αφορά την ηθική ευθύνη που έχουν οι καλλιτέχνες όταν πειραματίζονται με εργαλεία που έχουν συνέπειες οι οποίες είναι προβληματικές και δύσκολο –ή ακόμα και αδύνατον– να αντιμετωπιστούν. Οι καλλιτεχνικές δυνατότητες της τεχνητής νοημοσύνης, η αισθητική που προκύπτει από αυτήν και οι συζητήσεις που τη συνοδεύουν συνήθως εγείρουν αναπάντητα ερωτήματα πάνω σε θέματα που φαινομενικά έχουν επιλυθεί, όπως για παράδειγμα ο ρόλος των μηχανών σε σχέση με την ανθρώπινη δημιουργία, η υπεροχή των αντιληπτικών ή των εννοιολογικών χαρακτηριστικών στο πολιτισμικό περιεχόμενο καθώς και οι αντιλήψεις περί αυθεντίας σε αυτά τα πλαίσια. Μπορεί όμως, επίσης, να μας δώσουν το έναυσμα να αναθεωρήσουμε την παραδοχή ότι οι καλλιτέχνες είναι οι πλέον κατάλληλοι ή ότι κατέχουν τα απαραίτητα εργαλεία για να διερευνήσουν τέτοιου είδους ζητήματα με τρόπους που δεν είναι εκ των πραγμάτων αμφισβητήσιμοι από ηθικής σκοπιάς. Αναμφίβολα ο στόχος της τέχνης δεν είναι να διορθώσει τον κόσμο, αντιθέτως είναι να προσφέρει νέους τρόπους συσχέτισης με αυτόν, αλλά αυτός είναι ένας στόχος που καθίσταται ολοένα και πιο σύνθετος δεδομένου του φαινομενικά συνεχώς μεταβαλλόμενου τεχνολογικού πλαισίου μέσα στο οποίο βρισκόμαστε.

Τα τελευταία χρόνια, η τεχνητή νοημοσύνη έχει επεκταθεί σε μεγάλο βαθμό, τόσο ως προς τις τεχνικές της δυνατότητες όσο και ως προς την ολοένα και αυξανόμενη παρουσία της σε τομείς που δεν επιδέχονται έρευνα – έχει καταλήξει να είναι μια πηγή που μας κινεί αδιάκοπα το ενδιαφέρον και την περιέργεια. Η ολοένα και πιο εύκολη πρόσβαση στον πειραματισμό με συστήματα μηχανικής μάθησης έχει καταστήσει την τεχνητή νοημοσύνη και τις θεματικές που τη συνοδεύουν προσιτές σε μη ειδικούς, εισάγοντας όχι μόνο νέα εργαλεία και μεθόδους στον οπτικό πολιτισμό, αλλά πυροδοτώντας επιπλέον μεγάλη αβεβαιότητα. Η ευρεία πρόσβαση στη χρήση εξελιγμένων συστημάτων τεχνητής νοημοσύνης –τόσο σε τεχνικό επίπεδο όσο και ως προς την εξοικείωση με τις μορφές και τους τρόπους λειτουργίας τους– έχει προκαλέσει μια μετατόπιση όχι μόνο ως προς το τι και πώς έχει κανείς τη δυνατότητα να δημιουργήσει, αλλά και σε ό,τι αφορά την πολιτισμική αποτίμηση των συνεπειών που επιφέρει αυτή η μεταβολή. Οι καλλιτέχνες ανταποκρίθηκαν άμεσα στις προβληματικές προεκτάσεις που ενδέχεται να έχει η τεχνητή νοημοσύνη για την κοινωνία, τον πολιτισμό και τη δημιουργική εργασία, μεταξύ άλλων. Ωστόσο, ενώ η ενασχόληση με αυτό το πρόβλημα έχει βοηθήσει στο να αναδειχθούν οι αποχρώσεις και οι πτυχές του, παράλληλα έχει αναδειχτεί ότι δεν υπάρχουν ξεκάθαρα ορθοί και ηθικοί δρόμοι για να ακολουθήσουν στο μέλλον.

Έχει πλέον τεκμηριωθεί επαρκώς, για παράδειγμα, ότι η τεχνητή νοημοσύνη ενέχει σημαντικά προβλήματα όσον αφορά την προκατάληψη, το σφάλμα και την επιβολή προβληματικών ιεραρχήσεων εξουσίας. Και παρόλο που η συλλογιστική των καλλιτεχνών πάνω στις προβληματικές διαστάσεις της τεχνητής νοημοσύνης έχει αναμφίβολα συνεισφέρει στον διάλογο πάνω σε αυτά τα ζητήματα, φαίνεται ότι υπάρχει ολοένα και περισσότερο η αίσθηση ότι το διακύβευμα εδώ υπερβαίνει κατά πολύ τις δυνατότητες της τέχνης και των καλλιτεχνών να συνεισφέρουν ουσιαστικά στην κριτική θεώρηση της τεχνολογίας. Η τεχνητή νοημοσύνη έχει τη δύναμη να γίνεται επιβλαβής με τρόπους που δεν είναι απλό να επιλυθούν, εάν μπορούν πράγματι να επιλυθούν, και παρόλο που αυτό αποτελεί πρόσφορο έδαφος για καλλιτεχνικές αναζητήσεις, δεν είναι πάντα σαφές ποιος είναι ο ρόλος της τέχνης σε αυτό το πλαίσιο. Κάτι που συχνά θεωρείται δεδομένο, αλλά αξίζει να διερευνηθεί περαιτέρω, είναι κατά πόσο το καλλιτεχνικό έργο που επιχειρεί να αποδομήσει,

να ασκήσει κριτική και να κατανοήσει την τεχνητή νοημοσύνη θα οδηγήσει πράγματι σε θετικά αποτελέσματα σε σχέση με τους στόχους που έχει θέσει. Η υιοθέτηση της τεχνητής νοημοσύνης από την τέχνη ενδέχεται να φέρει στο προσκήνιο νέους τρόπους κατανόησης των ανθρωποκεντρικών διαστάσεών της, όμως ταυτόχρονα μπορεί να συμβάλλει στον εξωραϊσμό και στην επικύρωση των μηχανισμών που επιχειρεί να αποδομήσει, ακόμα και μέσω της κριτικής που ασκεί.

Η υιοθέτηση της τεχνητής νοημοσύνης στο πεδίο του πολιτισμού ενδέχεται να σηματοδοτεί ένα νέο επίπεδο βιομηχανοποίησης που απαιτεί εναλλακτικούς τρόπους σύλληψης της τέχνης, της σχέσης της με την τεχνολογία και της ευρύτερης σημασίας τους. Μορφές τεχνολογικής παραγωγής που μέχρι πρότινος δεν ήταν προσβάσιμες, σήμερα πλέον είναι ευρέως διαθέσιμες, επιτρέποντας ένα ολοένα και αυξανόμενο επίπεδο λαϊκής εξοικείωσης με το τι πραγματικά συνιστά η τεχνητή νοημοσύνη, πέρα από τις ζωηρές εικόνες που γεννά στο μυαλό ο ίδιος ο όρος. Ως αποτέλεσμα, έχουμε αποκτήσει περισσότερη επίγνωση τόσο του τρόπου λειτουργίας της τεχνητής νοημοσύνης όσο και των περιορισμών της και της δυνατότητάς της να προκαλέσει κακό. Ωστόσο, παρά την τεράστια προσοχή που έχει ανακινήσει γύρω του αυτό το θέμα κατά την τελευταία δεκαετία, δεν φαίνεται να έχουμε σημειώσει ουσιαστική πρόοδο όσον αφορά την επίλυση των πολυάριθμων εννοιολογικών και ηθικών αντιφάσεων που εγείρει. Ο κόσμος έχει σε μεγάλο βαθμό επίγνωση του γεγονότος ότι η τεχνητή νοημοσύνη έχει την τάση να κάνει λάθη, να είναι μεροληπτική και να αποτελεί ρίσκο, ωστόσο παρά τους κινδύνους που εγκυμονεί, όπως επίσης και ακριβώς λόγω αυτών, όλοι συμφωνούν ότι δεν πρέπει να αδιαφορήσουμε για την τεχνητή νοημοσύνη.

Μπορούμε να αποκτήσουμε μερικές χρήσιμες πληροφορίες αν διερευνήσουμε τον ίδιο τον όρο «τεχνητή νοημοσύνη» ακολουθώντας μια στρατηγική "AI-anti-hype" (Odubela, 2024) –που αντιτίθεται δηλαδή στην υπερπροβολή της τεχνητής νοημοσύνης– με στόχο να αποδομήσουμε σε κάποιον βαθμό το μυστήριο που την περιβάλλει. Η τεχνητή νοημοσύνη εμπνέει συχνά συγκρίσεις με τη γενική τεχνητή νοημοσύνη, μια εκδοχή της τεχνητής νοημοσύνης που έχει τις ίδιες ή και περισσότερες δυνατότητες από τους ανθρώπους για την εκτέλεση ενός δεδομένου έργου. Στον δημόσιο λόγο, συνδέεται συχνά με την ιδέα των αυτοματοποιημένων μηχανών σκέψης (automated thinking

machines), που είναι ικανές να παρακάμψουν τον αναποτελεσματικό άνθρωπο ως μεσολαβητή, εφόσον αυτό κριθεί εφικτό ή και επιθυμητό για την αποτελεσματικότητά τους. Είτε αυτή η απειλή αρθρωθεί με όρους εργασίας –μέσω του εκτοπισμού ή της αντικατάστασης των ρόλων του ανθρώπου– είτε με τη μορφή της εξόντωσης –για την οποία η τεχνητή νοημοσύνη θα μπορούσε θεωρητικά να επινοήσει πολλές δημιουργικές λύσεις– ο πιθανός κίνδυνος που ενέχει η τεχνητή νοημοσύνη ενδέχεται να λειτουργήσει διττά, αφενός πυροδοτώντας τον φόβο για την τεχνολογία και αφετέρου εμπνέοντας το δέος προς αυτήν, υπερεκτιμώντας τις σημερινές δυνατότητες που έχει η τεχνητή νοημοσύνη. Όσο αντιφατικό και αν ακούγεται, το όραμα της τεχνητής νοημοσύνης ως απειλής για την ανθρωπότητα ενδέχεται να αποσπά τη δικαιολογημένη προσοχή από το κακό που έχει ήδη αποδεδειγμένα κάνει στους ανθρώπους, όπως επίσης και από το γεγονός ότι αυτό που σήμερα αποκαλούμε «τεχνητή νοημοσύνη» ελάχιστα μοιάζει με τη δυνατότητα των μηχανών να επιτελούν γνωστικές λειτουργίες συγκρίσιμες με εκείνες του ανθρώπου.

Η περιγραφή της τεχνητής νοημοσύνης με αυτούς τους όρους, ως εργαλείο που αυτοματοποιεί τη διανοητική εργασία, ενδέχεται να αποτελεί από μόνη της ένα προβληματικό θεμέλιο πάνω στο οποίο θα αναπτυχθεί η τεχνολογία. Η παραπλανητική ταύτιση των υπολογιστικών διαδικασιών με την ανθρώπινη γνωστική λειτουργία μέσω της χρήσης ανθρωπομορφικής ορολογίας ίσως να αποκαλύπτει περισσότερα από όσα επιδιώκει: ότι αυτή η διφορούμενη πράξη στην ουσία αποσκοπεί στο να καταστήσει τη νοημοσύνη ανταλλάξιμη με το τεχνητό. Έχουν ενδιαφέρον τα λόγια του Matteo Pasquinelli πάνω σε αυτό το θέμα, τα οποία εμπεριέχονται στην κοινωνική ιστορία της τεχνητής νοημοσύνης που έχει γράψει και όπου εξηγεί ότι η σχέση μεταξύ των ηλεκτρονικών υπολογιστών και της εκτέλεσης γνωστικών εργασιών δεν έχει περιγραφεί με ξεκάθαρο τρόπο:

> Η ιδέα του αυτόματου υπολογιστή, με τη σύγχρονη έννοια, αναδύθηκε από την προσπάθεια να μηχανοποιηθεί η διανοητική εργασία των υπαλλήλων και όχι από το παλιό αλχημιστικό όνειρο να κατασκευαστούν σκεπτόμενα αυτόματα μηχανήματα – αν και το δεύτερο αφήγημα χρησιμοποιήθηκε συχνά, τόσο κατά τον δέκατο ένατο αιώνα όσο και στον αιώνα της εταιρικής τεχνητής νοημοσύνης, για να συγκαλύψει το πρώτο αφήγημα που ήταν ένα επιχειρηματικό σχέδιο. (2023, 52)

Αυτή η διάσταση της τεχνητής νοημοσύνης ως όρου αναδεικνύει και πάλι τη δυνατότητα της τεχνολογίας να εμφανίζεται σαν ένα διττό φαινόμενο, να είναι διπρόσωπη σαν τον Ιανό, να έχει ταυτόχρονα αντιφατικές όψεις. Αν και είναι ίσως παραπλανητικό να υποστηρίξει κανείς ότι η τεχνητή νοημοσύνη μπορεί να αντικαταστήσει εντελώς την ανθρώπινη εργασία, ωστόσο ισχύει και το γεγονός ότι αυτό ήταν –και ίσως θα εξακολουθήσει να είναι– ένα υπόγειο κίνητρο πίσω από την ανάπτυξη και την εφαρμογή της. Αυτός ο όρος κρύβει επίσης αμφισβητήσιμες παραδοχές σχετικά με τη φύση της νοημοσύνης –και ακόμα πιο ανησυχητικά, για την πολιτική ταυτότητας που τη συνοδεύει– όπως επίσης και με την ανάθεση της λήψης αποφάσεων σε ανθρωπόμορφες μηχανές και την εφαρμογή ενός διχαστικού ιεραρχικού προγράμματος στον κόσμο. Παρά τις εύλογες επικρίσεις, η προσπάθεια αποδόμησης της εσωτερικής λογικής της μέσω πειραματισμών με εναλλακτικές ορολογίες ίσως υποπέσει στο σφάλμα της διατύπωσης μιας κριτικής για την τεχνητή νοημοσύνη που παραμένει δέσμια των θεμελιωδών αρχών της. Για παράδειγμα, αν μετονομαστεί η τεχνητή νοημοσύνη και χρησιμοποιηθεί ένας όρος όπως η *συνθετική νόηση (synthetic cognition)* (Manouach & Engelhardt, 2022), μπορεί απλώς να σημαίνει ότι θα αποκαλείται με συνώνυμες λέξεις που συγκαλύπτουν το γεγονός ότι αυτό που υπάρχει από κάτω παραμένει αμετάβλητο. Αν και αυτός ο όρος περιέχει αρκετές εννοιολογικές ασυνέπειες και εσωτερικές αντιφάσεις, είναι παρ' όλα αυτά εξαιρετικά αποτελεσματικός επειδή μπορεί να μας ενθουσιάσει, καθώς φέρνει στο νου ζωηρές εικόνες σχετικά με το τι θα μπορούσε να σημαίνει η τεχνητή νοημοσύνη – είτε προς το καλύτερο είτε προς το χειρότερο.

Για διάφορους λόγους, οι καλλιτέχνες ήταν από τους πρώτους που υιοθέτησαν την τεχνητή νοημοσύνη από πολύ νωρίς, όπως και οι πιο ειλικρινείς και αυστηροί επικριτές της. Εν μέρει αυτό οφείλεται στην αντίληψη ότι ο πολιτισμός αποτελεί πανάκεια για τα κοινωνικά δεινά, ειδικά όταν αντιτίθεται στην τεχνολογική εξέλιξη. Είτε αυτή η αντίληψη για τον ρόλο του πολιτισμού, της τέχνης ή των καλλιτεχνών είναι ακριβής είτε όχι, το θέμα και οι δυνατότητες της τεχνητής νοημοσύνης έχουν αναδειχθεί σε εξαιρετικά γόνιμο πεδίο για δημιουργική διερεύνηση. Οι καλλιτέχνες έχουν διαδραματίσει σημαντικό ρόλο στην ερμηνεία και στη διερεύνηση των συνεπειών της τεχνητής νοημοσύνης όχι μόνο στην τέχνη, στην αισθητική και στον πολιτισμό, αλλά και όσον αφορά

τις ευρύτερες επιδράσεις της στην κοινωνία και στο περιβάλλον. Πολλοί έχουν εκφράσει τις ανησυχίες τους για την εγγενή τάση της τεχνητής νοημοσύνης να προκαλεί βλάβες, επισημαίνοντας, μεταξύ άλλων, το ιστορικό της όσον αφορά την εξαντλητική εκμετάλλευση φυσικών πόρων, την αλγοριθμική μεροληψία και τη δυνατότητά της να επιφέρει δυνητικά καταστροφικές συνέπειες με τρόπους που δύσκολα μπορούν να προβλεφθούν. Αν και τα ζητήματα αυτά, όπως και τα τεχνολογικά εργαλεία από τα οποία προκύπτουν, προσφέρουν γόνιμο έδαφος στους καλλιτέχνες, η καλλιτεχνική διερεύνηση των ζητημάτων αυτών κινείται σε ένα λεπτό όριο μεταξύ της κριτικής οικειοποίησης της τεχνολογίας και της πολιτισμικής της επικύρωσης.

Η απόπειρα αμφισβήτησης της τεχνολογίας μέσω της ίδιας της τεχνολογίας δεν είναι ποτέ απλή υπόθεση. Η ουσιαστική αντιμετώπιση των επιπτώσεων της τεχνολογίας δεν είναι ποτέ τόσο απλή όσο η συνειδητή επιλογή για το αν ένα συγκεκριμένο εργαλείο είναι «καλό» ή «κακό». Όταν οι καλλιτέχνες χρησιμοποιούν κριτικά την τεχνολογία, δεν βρίσκονται απλώς αντιμέτωποι με την επιλογή αν θα χρησιμοποιήσουν ένα συγκεκριμένο είδος συσκευής, μεθόδου ή λογισμικού, αλλά έχουν αντιθέτως το καθήκον να διερευνήσουν και να στοχαστούν τις επιπτώσεις τους, τη σημασία τους και την αισθητική τους. Αυτό μπορεί να περιλαμβάνει την αποδόμηση των ίδιων των εργαλείων με στόχο την κατανόησή τους, ή –εξίσου συχνά– την άσκηση κριτικής.

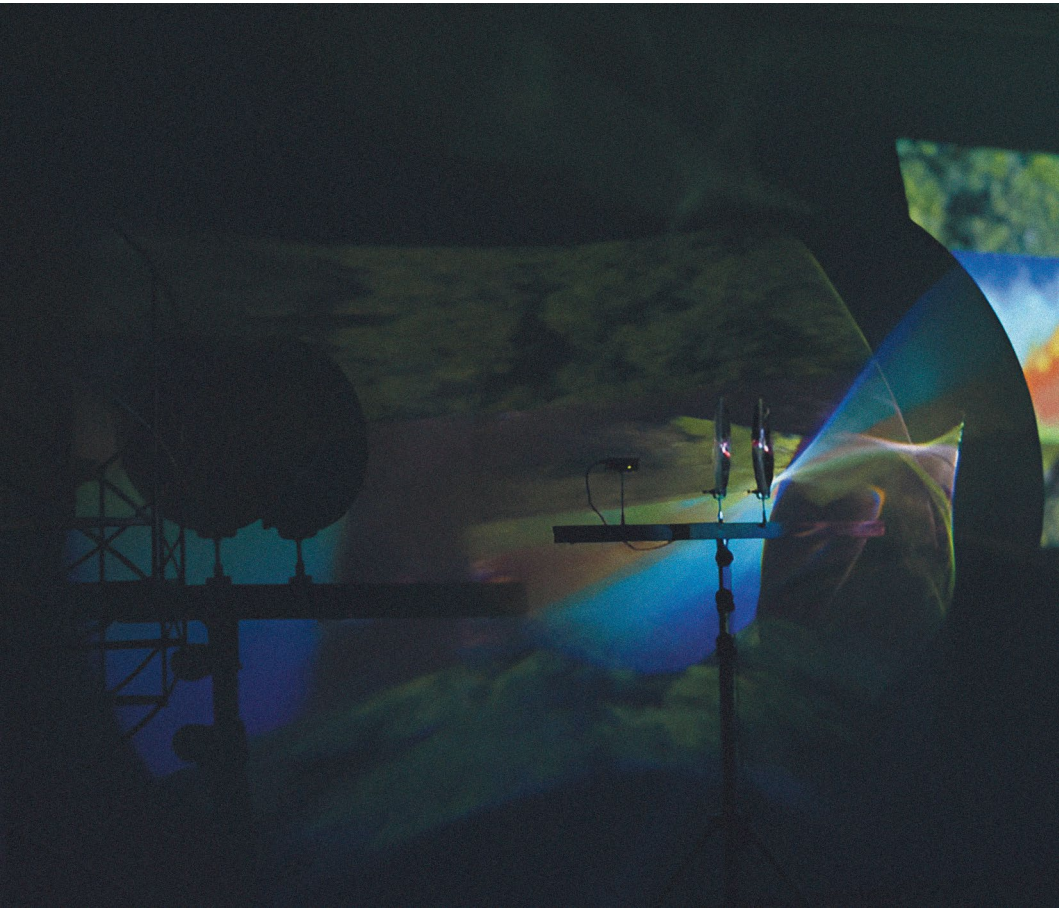
Μια από τις πλέον αξιομνημόνευτες παραβολές κοινωνικής απόρριψης της τεχνολογίας, η εξέγερση των Λουδιτών, μπορεί να λειτουργήσει σαν υπενθύμιση της πολυδιάστατης φύσης της τεχνολογίας και της σχέσης της με ιστορικά αφηγήματα του πολιτισμού. Με την καταστροφή των μηχανημάτων που απειλούσαν να απαξιώσουν την αξία της εργασίας τους, η αντίδραση των Λουδιτών στις συνέπειες της εκβιομηχάνισης έμεινε στην ιστορία για πάνω από έναν αιώνα, παρόλο που πιθανότατα δεν θα τους ικανοποιούσε ο τρόπος με τον οποίο ο κόσμος θυμάται τις πράξεις τους. Ο όρος επανεμφανίζεται συχνά σε συζητήσεις σχετικά με τον αντίκτυπο των αναδυόμενων τεχνολογιών, ενώ σήμερα συνδέεται με μια στάση που αντιτίθεται στην τεχνολογία – ακόμη και στην πρόοδο. Οι λεπτές αποχρώσεις της απόρριψης των μηχανών της βιομηχανικής παραγωγής φαίνεται ότι έχουν διαβρωθεί με την πάροδο του χρόνου, έχουν παρερμηνευτεί από όσους υιοθετούν μια

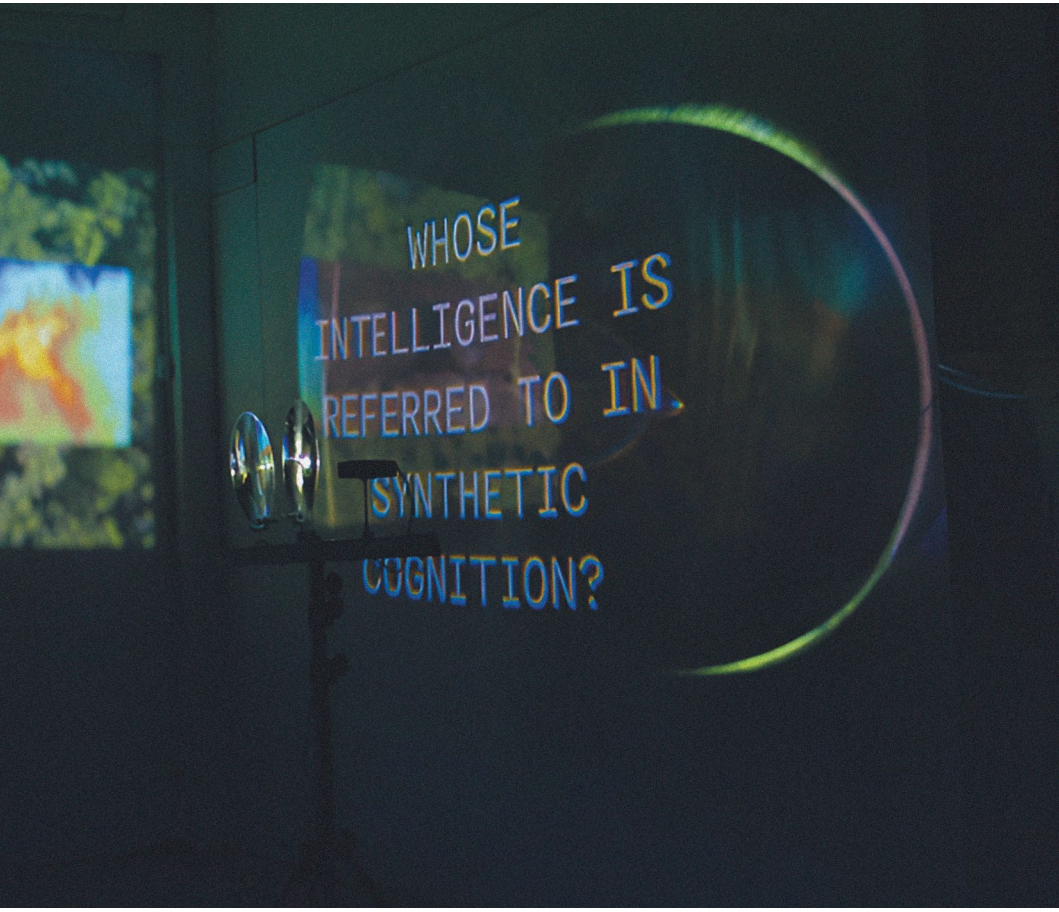
συγκεκριμένη ανάγνωση της ιστορίας και έχουν υπεραπλουστευτεί μέσα από τις πολλές διηγήσεις.

> Η λέξη «Λουδίτης» […] χρησιμοποιείται καθημερινά με τρόπους που υποδηλώνουν ότι είμαστε σε σύγχυση όχι μόνο σχετικά με την ίδια την τεχνολογία, αλλά επίσης και σχετικά με το ποιοι ήταν οι αρχικοί Λουδίτες και τι πραγματικά σημαίνει να είναι κανείς Λουδίτης σήμερα. (Conniff, 2011)

Παρόλο που τα μηχανήματα ήταν ένας από τους βασικούς στόχους της οργής τους, η ουσία του ζητήματος ήταν οι συνθήκες εργασίας. Έτσι, αν ο Λουδισμός περιγραφεί απλώς ως μια κινητοποίηση ενάντια στην τεχνολογία, αυτό σημαίνει ότι έχει παρερμηνευτεί η σημασία του, αλλά και ότι παραγνωρίζεται ο προειδοποιητικός χαρακτήρας του. Οι ιστορικές λεπτομέρειες που αφορούν το ίδιο το κίνημα των Λουδιτών δεν μας ενδιαφέρουν τόσο εδώ όσο το πώς αυτή η πολιτισμική μνήμη συνδέεται με τις σημερινές μας προσπάθειες να αποδεχτούμε την τεχνητή νοημοσύνη και τον αντίκτυπό της – τον πραγματικό, τον δυνητικό, τον καλλιτεχνικό και τον κοινωνικό. Οι σημερινές αντιδράσεις στην τεχνητή νοημοσύνη μπορεί επίσης να δημιουργήσουν μια νέα ταμπέλα για όσους απορρίπτουν την τεχνολογία βασιζόμενοι στις ευρύτερες συνέπειές της. Μπορούμε μόνο να ελπίζουμε ότι οι λεπτές αποχρώσεις αυτού του αγώνα δεν θα χαθούν, αλλά θα καταγραφούν στην ιστορία με τρόπο που δεν είναι ισοπεδωτικός. Ο κίνδυνος που αναδεικνύεται από το γεγονός ότι οι Λουδίτες παρουσιάζονται ως δισδιάστατοι χαρακτήρες που επιτίθενται στις μηχανές λόγω φόβου και άγνοιας είναι ότι αυτή η ανατροπή της ιστορικής αφήγησης μπορεί εξίσου εύκολα να χρησιμοποιηθεί για να απαξιώσει κάθε τεκμηριωμένη ή εύλογη κριτική στην τεχνολογία όσο και για να την απορρίψει για οποιονδήποτε άλλο λόγο.

Μια σύγχρονη εξέγερση των Λουδιτών θα μπορούσε θεωρητικά να επιχειρήσει να καταστρέψει τα συστήματα τεχνητής νοημοσύνης με βίαιο τρόπο: εφευρίσκοντας μηχανές που καταστρέφουν την τεχνολογία πιο αποτελεσματικά[2], διακόπτοντας την ηλεκτροδότηση ή καίγοντας εντελώς τα κέντρα δεδομένων. Ωστόσο, όπως επισημαίνουν οι Kate Crawford και Vladan Joler (2018), ένα σύστημα τεχνητής νοημοσύνης είναι εκτενώς διεσπαρμένο και ενσωματωμένο σε ποικίλες υλικές και άυλες υποδομές, επομένως η απόρριψη ή η αναστολή





Rosemary Lee, *A Structural Plan for Imitation: Engines of Differentiation,* 2024. Σε συνεργασία με την Αλεξία Αχιλλέως. Δικάναλη βίντεο εγκατάσταση. Άποψη εγκατάστασης στο NeMe Arts Centre, Λεμεσός, Κύπρος. Ευγενική παραχώρηση της καλλιτέχνιδας. Φωτογραφία: Nicos Avraamides

της τεχνητής νοημοσύνης σε σημαντική κλίμακα θα ήταν σήμερα ουσιαστικά αδύνατον να συμβεί. Ακόμη και αν κάτι τέτοιο ήταν εφικτό, δεν θα επαρκούσε απλώς η καταστροφή των μηχανών για να εξαλειφθεί η τεχνητή νοημοσύνη. Το ίδιο ίσχυε ήδη από την εποχή που είχαν εξεγερθεί οι Λουδίτες – παρόλο που τα εργαλεία εκείνη την εποχή ήταν από μηχανικής άποψης πιο απλά. Το γεγονός ότι δεν τους θυμόμαστε σήμερα τους Λουδίτες επειδή διαμαρτύρονταν για τις εργασιακές τους συνθήκες με τη χρήση της τεχνολογίας, αλλά αντιθέτως επειδή εναντιώνονταν στην ίδια την τεχνολογία, αναδεικνύει τις δυσκολίες που αντιμετωπίζουν σήμερα οι κριτικές προσεγγίσεις για την οικειοποίηση της τεχνολογίας στην τέχνη. Η πράξη της καταστροφής των εργαλείων που συνδέονται με την αλλαγή των τεχνολογικών συνθηκών της εργασίας είναι μια φορτισμένη εικόνα, ωστόσο το πρόβλημα δεν ήταν ποτέ τα ίδια τα εργαλεία: το πρόβλημα ήταν η δυναμική του συστήματος μέσα στο οποίο εντάσσονται.

Παρόλο που ορισμένοι στόχοι μπορεί να διαμορφώσουν και να ενταχθούν στον σχεδιασμό και τη χρήση των τεχνολογιών, η εμμονή αποκλειστικά σε αυτές τις στοχευμένες πτυχές θα παρέβλεπε ένα πιο ουσιώδες χαρακτηριστικό: ότι η τεχνολογία έχει εξίσου τη δυνατότητα να λειτουργεί υπέρ και κατά των στόχων που εξυπηρετεί. Αυτό καθιστά την καλλιτεχνική ενασχόληση με τις αναδυόμενες τεχνολογίες – όπως η τεχνητή νοημοσύνη– ένα ιδιαιτέρως συναρπαστικό πεδίο για πολιτιστική εξέγερση, καθώς παίζει έναν εξισορροπητικό ρόλο καλλιεργώντας γόνιμο έδαφος για καινοτόμες μορφές καλλιτεχνικής πρακτικής, αισθητικής και σκέψης, ενώ παράλληλα αποτρέπει την ιδιοποίηση αυτών των προσπαθειών για άλλους σκοπούς. Η μεταφορά αυτής της πολυεπίπεδης αντίληψης περί τεχνολογίας στην καλλιτεχνική διερεύνηση της τεχνητής νοημοσύνης βοηθάει στην αποφυγή της τάσης να παγιδεύεται αυτό το θέμα στις έντονες αντιθέσεις, στο παιχνίδι του μαύρου και του άσπρου –ακόμα και της ηθικής υπεροχής– όταν η πραγματικότητα δεν μας προσφέρει ξεκάθαρες απαντήσεις.

Ο δημιουργικός πειραματισμός με τις αναδυόμενες τεχνολογίες έχει τη δυνατότητα να αποκαλύπτει αποχρώσεις που μπορεί διαφορετικά να παρέμεναν αδιόρατες, ενώ παράλληλα προκαλεί μια αίσθηση αμφισημίας όσον αφορά την οικειοποίηση της τεχνολογίας στην τέχνη – ότι μια συγκεκριμένη τεχνολογία μπορεί να δίνει την

αίσθηση ότι είναι αναπόφευκτη αλλά και εμφανώς προβληματική. Όσο δελεαστικό και αν είναι να πιστεύουμε ότι οι σημερινές μας ανησυχίες σχετικά με την τεχνολογία είναι διαφορετικές από εκείνες του παρελθόντος, μπορεί και να διακρίνουμε ομοιότητες στις πρόσφατες εξελίξεις που προκαλούν μια αλλόκοτη νοσταλγία για πράγματα που μόλις και μετά βίας θυμόμαστε, αλλά τα οποία αναδύονται και παίρνουν μορφή ακριβώς μέσα από την ίδια την ανάμνηση. Η διάδοση των νέων εργαλείων, μεθόδων και θεματικών που είναι διαθέσιμα στους καλλιτέχνες έχει αλλάξει τις συνθήκες της δημιουργικής εργασίας, όπως και τις αξίες με βάση τις οποίες αξιολογούνται τα αποτελέσματά της. Και ενώ μπορεί να είναι πλέον εφικτό να δημιουργούμε με νέους τρόπους χρησιμοποιώντας την τεχνητή νοημοσύνη, είναι εξίσου σημαντικό να θυμόμαστε ότι αυτή η μετατόπιση εμπεριέχει στοιχεία που παραπέμπουν στο παρελθόν. Η ενσωμάτωση νέων μεθόδων και εννοιών συντελείται μέσα σε ήδη υπάρχοντα πλαίσια και δεν σηματοδοτεί μια ριζική ρήξη με τις προηγούμενες πρακτικές. Οι τεχνολογίες λειτουργούν συχνά σαν δίκοπα μαχαίρια, ενεργούν *σε* εμάς όσο και εμείς ενεργούμε *με* αυτές. Αν αναγνωρίσουμε αυτήν τη διπλή δυναμική, μπορεί και να καταφέρουμε να αντισταθούμε στον πειρασμό να θεωρούμε ότι οι πολλαπλές επιπτώσεις της καλλιτεχνικής πράξης που χρησιμοποιεί την τεχνητή νοημοσύνη δεν έχει κανένα προηγούμενο. Η αποσύνδεση των τεχνολογιών από το ιστορικό τους πλαίσιο μας υποχρεώνει να τις προσεγγίζουμε σαν λευκό καμβά – και αυτό μπορεί να σβήσει κρίσιμες πτυχές και να μας εμποδίσει να αξιοποιήσουμε υφιστάμενα πλαίσια για την κατανόηση των επιπτώσεων της τεχνολογίας.

Σημειώσεις

[1] Αυτόν τον όρο τον χρησιμοποιούσαν οι Καταστασιακοί για να αναφερθούν στην οικειοποίηση ενός πράγματος ενάντια στον αρχικό του σκοπό, ως μέσο υπονόμευσης ή αμφισβήτησής του.
[2] Παράφραση ενός μιμίδιου (meme), βλ.: (matthat, 2010).

Βιβλιογραφία

Conniff, R. (2011, Μάρτιος). What the Luddites really fought against. *Smithsonian.* https://www.smithsonianmag.com/history/what-the-luddites-really-fought-against-264412/

Crawford, K. (2021). *Atlas of AI*. Yale University Press.

Crawford, K., & Joler, V. (2018). *Anatomy of an AI system: The Amazon Echo as an anatomical map of human labor, data and planetary resources*. AI Now Institute and Share Lab. https://anatomyof.ai

Future of Life Institute. (2023, Μάρτιος 22). *Pause giant AI experiments: An open letter*. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Manouach, I., & Engelhardt, A. (Επιμ.). (2022). *Chimeras: Inventory of synthetic cognition*. Onassis Foundation.

matthat. (2010, Αύγουστος 10). Luddite invents machine to destroy technology quicker. *The Spoof!*. https://www.thespoof.com/spoof-news/science-technology/83952/luddite-invents-machine-to-destroy-technology-quicker

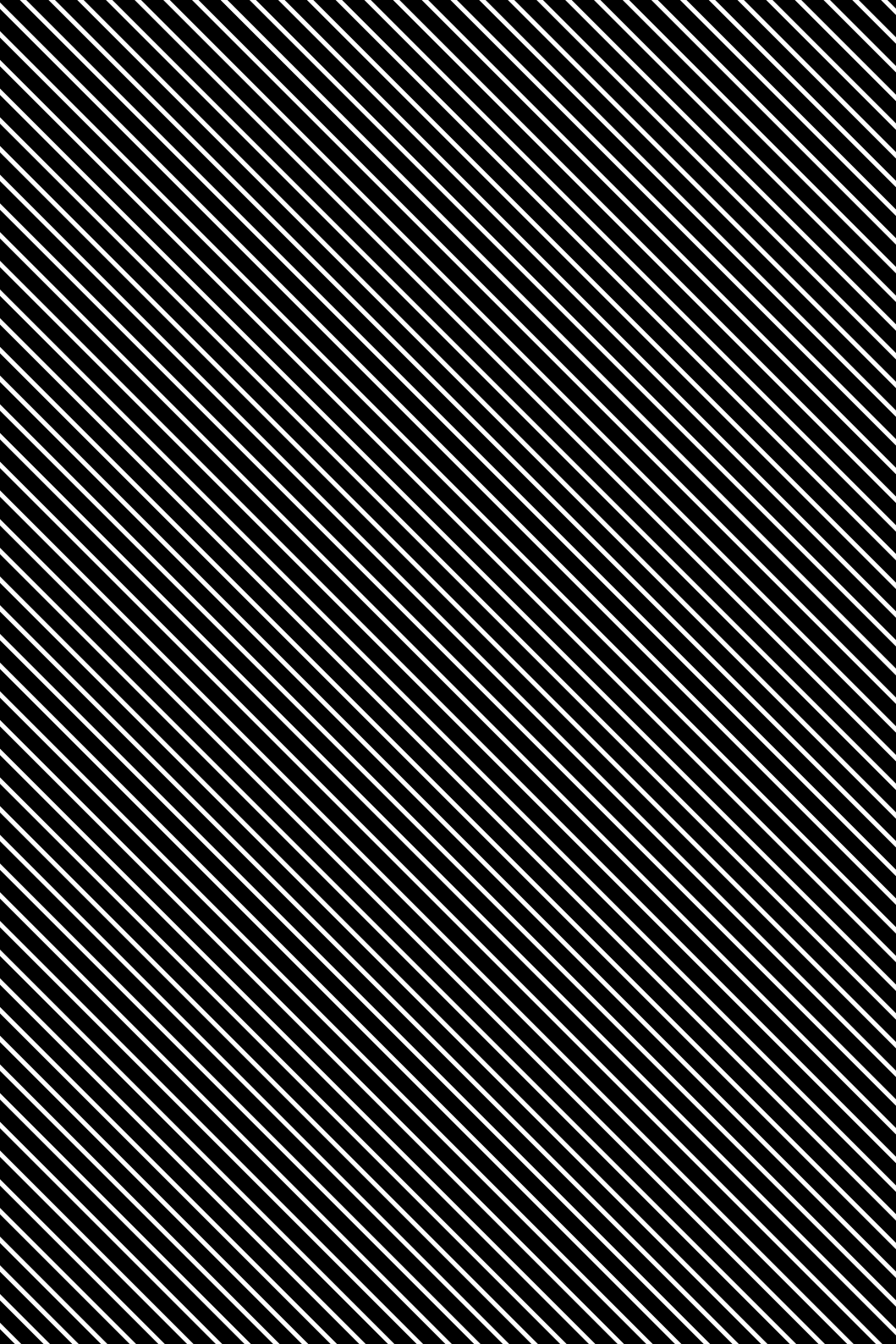
Odubela, A. (2024, Νοέμβριος 15). V1.01 ~ A re-introduction to AI hype. *AI Anti-Hype.* https://aiantihype.beehiiv.com/p/a-reintroduction-to-hype

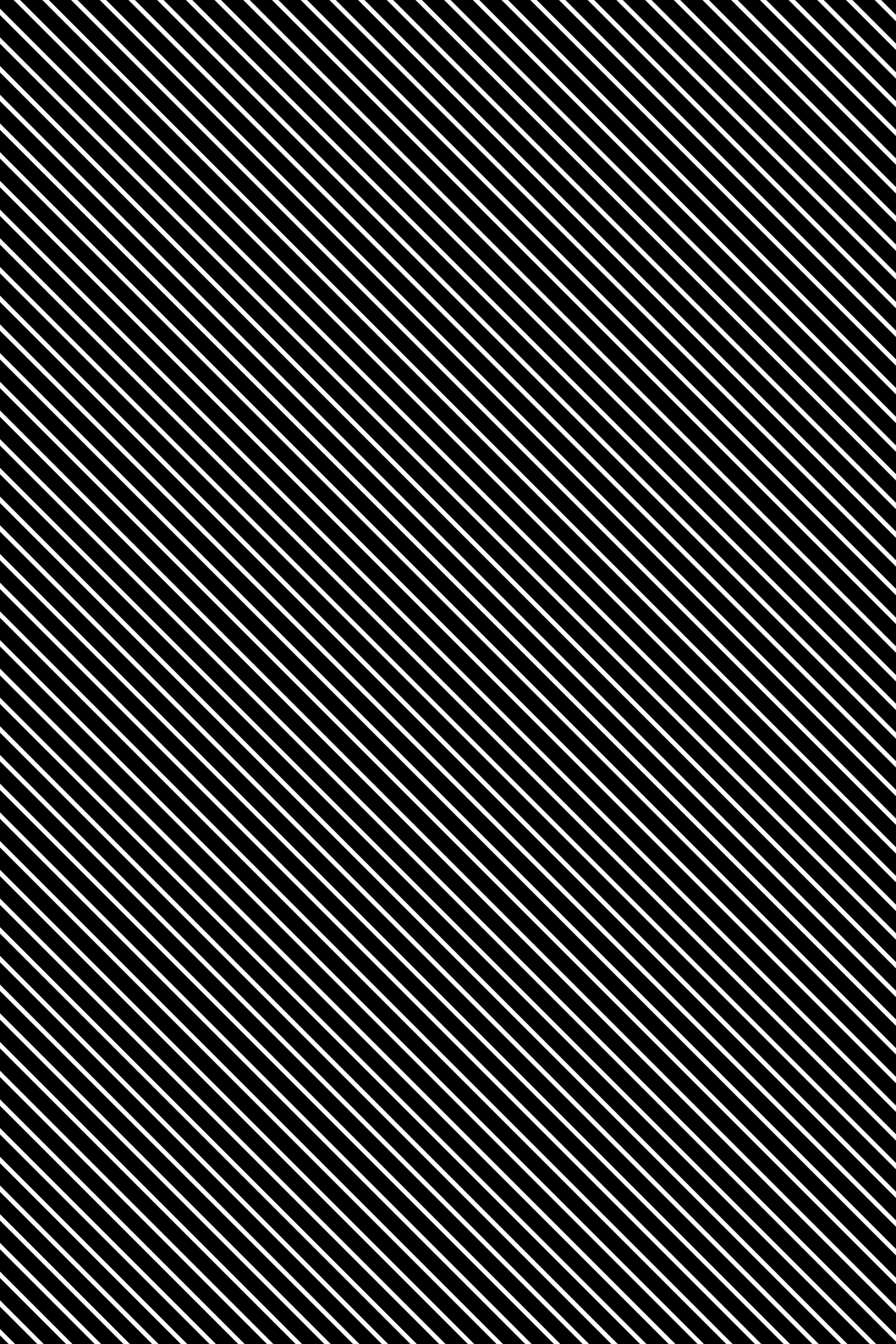
Parisi, L. (2021). Negative optics in vision machines. *AI & Society, 36*, 1281–1293. https://doi.org/10.1007/s00146-020-01096-7

Pasquinelli, M. (2023). *The eye of the master: A social history of artificial intelligence.* Verso.

The Editors of Encyclopaedia Britannica. (2025). Luddite. Στην *Encyclopaedia Britannica*. https://www.britannica.com/event/Luddite

Vaneigem, R. (2012). *The revolution of everyday life* (μτφρ. D. Nicholson-Smith). PM Press. (Το πρωτότυπο έργο δημοσιεύτηκε το 1967)





Πέραν της δημιουργικότητας: αναδυόμενα μοντέλα τεχνητής νοημοσύνης και διαδικτυακά ενήμερη τέχνη

Κωστής Σταφυλάκης
Επίκουρος Καθηγητής Εικαστικών Τεχνών,
Διευρυμένων Μέσων και Μορφών, Τμήμα Εικαστικών Τεχνών,
Ανωτάτη Σχολή Καλών Τεχνών

Στο ανά χείρας κείμενο, θα επιχειρήσω να εξετάσω τη διαλεκτική ανάμεσα στη σύγχρονη εικαστική πρακτική και στα αναδυόμενα μοντέλα τεχνητής νοημοσύνης, δοκιμάζοντας ένα πλαίσιο κατανόησης που δεν αντλεί την έμπνευσή του από τη ρομαντική έννοια της δημιουργικότητας – μια έννοια που θεμελιώθηκε εντός του ίδιου σημασιακού πεδίου με τις έννοιες της πρωτοτυπίας και της ιδιοφυίας. Κι όμως, η πρόσφατη διάδοση των εφαρμογών της παραγωγικής τεχνητής νοημοσύνης (generative AI) πυροδότησε μια δημόσια συζήτηση που, εν πολλοίς, καθορίστηκε από το άγχος για την απώλεια του πρωτείου του δημιουργού, όπως αυτό συνοψίζεται στο δημοφιλές ερώτημα: «μπορεί η τεχνητή νοημοσύνη να δημιουργήσει έργα τέχνης;» Ενδιαφέρον παρουσιάζει το γεγονός ότι το άγχος αυτό διέπει και συζητήσεις που δεν έχουν, αποκλειστικά, εκλαϊκευτικό χαρακτήρα. Η καχυποψία μου προς το δημοφιλές αυτό ερώτημα επεκτείνεται προς την επακόλουθη αντιμετώπιση της σχέσης της σύγχρονης τέχνης με την τεχνητή νοημοσύνη ως ένα νέο κεφάλαιο σε μια ιστορία των μέσων, ή των «νέων μέσων». Η οπτική αυτή βλέπει, συνήθως, την εξάπλωση των πρόσφατων τεχνολογιών ως ευκαιρία για ενίσχυση ή προέκταση του «δημιουργικού εαυτού», υποθέτοντας το ίδιο κέντρο δημιουργικής αυθεντίας: έναν «μοντέρνο εαυτό» που διευκολύνεται από τεχνολογίες, μέσα ή εργαλεία. Μια τέτοια λογική καθρεφτίζεται, για παράδειγμα, στη δημοφιλή τέχνη του Refik Anadol, που επεξεργάζεται συγκεκριμένες συστοιχίες από δεδομένα για να τα μεταφράσει σε γιγαντιαία ηλεκτρονικά καλειδοσκόπια ζωντανού χρόνου. Αυτή η μεταψηφιακή επένδυση του αστικού τοπίου και του αρχιτεκτονικού σχεδιασμού μπορεί, δίχως ιδιαίτερο κόπο, να γίνει αντιληπτή σαν σύγχρονη μετάφραση/προέκταση της μοντέρνας αφαίρεσης.

Ο λόγος που δεν βρίσκω τις παραπάνω λογικές πειστικές είναι ότι τείνουν να συμβάλλουν σε μια αντίληψη που λησμονεί ότι, εδώ και χρόνια, ξεδιπλώνεται μια πιο ουσιαστική διαλεκτική ανάμεσα στις εικαστικές πρακτικές και την τεχνητή νοημοσύνη, η οποία λαμβάνει υπόψη ορισμένους θεμελιακούς μηχανισμούς της κοινωνικής διαδικασίας. Μιλώ για μια διαλεκτική ανάμεσα στην τέχνη και τη διαδικτυακή κουλτούρα, μια διαλεκτική ανάμεσα στην καλλιτεχνική στρατηγική, τον οπτικό πολιτισμό, τις πολιτικές και κοινωνικές παραμέτρους της ζωής μετά το Web 2.0.

Το πρόβλημα της συζήτησης σχετικά με την «καλλιτεχνική δημιουργία» στην εποχή της τεχνητής νοημοσύνης είναι ότι αυτή συνήθως παρακάμπτει τις μεγαδομές, το κοινωνικό και τεχνολογικό οικοσύστημα εντός του οποίου αυτή η «συνάντηση» λαμβάνει χώρα. Το οικοσύστημα αυτό δεν είναι άλλο από το διαδίκτυο, στη σύγχρονη μεταβατική του φάση, από το Web 2.0 στο Web 2.5 και, από εκεί, στις εικαζόμενες διάδοχες μορφές του. Εκεί, η τεχνητή νοημοσύνη μας συναντά με έναν τρόπο αναπόδραστο, συμβάλλοντας στη διαμόρφωση ενός «αλγοριθμικού ασυνείδητου» που κυριαρχεί πάνω στο υποκείμενο χωρίς πολλή διαβούλευση. Καθώς γράφονται αυτές οι γραμμές, το timeline των κοινωνικών μου δικτύων έχει καταληφθεί από τεχνητές εικόνες αισθησιακής περίπτυξης του Timothée Chalamet και του Tom Holland, σαν να προσπαθεί να μου πει κάτι, να διεγείρει, επιτυχώς ή ανεπιτυχώς, την επιθυμία ώστε να μείνω συντονισμένος σε όσα πρόκειται να μου «προσφέρει». Ακόμα κι αν βρίσκω το υλικό αυτό σχετικά αδιάφορο, δεν μπορώ παρά να αποδεχτώ πως ο τρόπος που τα παραγωγικά αυτά μοντέλα φετιχοποιούν τις φιγούρες των δύο γνωστών ηθοποιών δεν περνάει απαρατήρητος. Η «αψεγάδιαστη» γυαλάδα που τυλίγει τα σώματα και τα ρούχα τους δεν αφήνει αδιάφορες τις συνάψεις του αμφιβληστροειδούς με τον ψυχισμό. Η επιτάχυνση αυτή ξεκινάει από το 2016, όταν τα σοσιαλμιντιακά feeds ξεκινούν να προσδιορίζουν το περιεχόμενό τους όλο και περισσότερο μέσω αλγοριθμικής εξατομίκευσης και λιγότερο χρονολογικά.

Έχει τονιστεί, από πολλές κατευθύνσεις, ότι οι σύγχρονες τεχνολογίες μηχανικής εκμάθησης αναπτύχθηκαν παράλληλα με τη βιομηχανία εξόρυξης δεδομένων και τον κεντρικό της ρόλο σε μια συνθήκη που η Louise Amoore περιέγραψε ως «αλγοριθμικό πόλεμο», έναν «πόλεμο με άλλα μέσα» όπου «εταιρίες, ιδρύματα, κυβερνήσεις αγωνίζονται

καθημερινά να αποκτήσουν και να διαχειριστούν δεδομένα για σκοπούς που συνδέονται με την ασφάλεια, το εμπόριο, την επιστήμη ή την πολιτική προπαγάνδα». Η σημερινή «άνοιξη της τεχνητής νοημοσύνης» οφείλεται εν πολλοίς στον πολιτικοοικονομικό μετασχηματισμό που ονομάζεται Web 2.0, δηλαδή στη μετατροπή του διαδικτύου σε τόπο άντλησης, συσσώρευσης και επεξεργασίας των δεδομένων και των «μεγάλων δεδομένων», στην τεχνολογία και τις υπηρεσίες του cloud, στη δημιουργία μεγάλων κέντρων δεδομένων, στη δημιουργία μεγάλων in-house servers από τις μεγάλες τεχνολογικές εταιρείες (Big Tech), τη δίχως προηγούμενο δικτύωση των υπολογιστικών συστημάτων, την εκμετάλλευση των κλειστών δεδομένων από τον χώρο των κοινωνικών δικτύων και των διαδικτυακών αγορών. Αν και πολλοί θα ισχυρίζονταν το αντίθετο, η υπόθεση ενός ουσιωδώς αποκεντρωμένου ίντερνετ, ενός Web 3.0, βασισμένου σε τύπου blockchain τεχνολογίες, δεν έχει επαρκώς υλοποιηθεί: αντί για επέκταση των προστατευμένων και κρυπτογραφημένων κόμβων των χρηστών, το διαδίκτυο κατακτήθηκε από νευρωνικά δίκτυα με δημοφιλείς και ευκολότερα προσβάσιμες διεπαφές για το ευρύ κοινό.

Το Web 2.0 (και ό,τι μοιάζει να το ακολουθεί) είναι, ανάμεσα σε άλλα, μια βιομηχανία δεδομένων που απαίτησε «δεξαμενές εικόνων» όπως το ImageNet, αποθετήρια υλικού που αξιοποιούν τα σημερινά νευρωνικά δίκτυα, τα GANs, το σύστημα βαθιάς μάθησης (deep learning) του Stable Diffusion και οι διάφορες εναλλακτικές στα μοντέλα παραγωγής εικόνας από κείμενο (text-to-image) της παραγωγικής τεχνητής νοημοσύνης. Το «άγχος» των εφαρμογών αυτών για γεωγραφική, επιχειρησιακή, αναπαραστατική, ιατροδικαστική ακρίβεια δεν είναι παρά προϊόν του πολιτικοοικονομικού πλέγματος που τις γέννησε. Όπως εξηγεί η Joanna Zylinska, ο «φωτογραφικός ρεαλισμός» που οι τεχνολογίες αυτές δείχνουν να επιζητούν οφείλεται, εν πολλοίς, στο γεγονός ότι «στα σύνολα δεδομένων (datasets) της μηχανικής εκμάθησης, η φωτογραφία μοιάζει να κατέχει ένα 'οντολογικό προνόμιο' […] διά του οποίου τα εξωτερικά αντικείμενα εκδηλώνονται σε εικόνες […]. Η πεποίθηση αυτή εφαρμόζεται όχι μόνο σε ό,τι μπορεί να οριστεί ως φωτογραφίες πρώτου επιπέδου (εικόνες ανθρώπων, ζώων, τοπίων και άψυχων αντικειμένων που τραβήχτηκαν για κοινωνικούς, καταγραφικούς ή καλλιτεχνικούς λόγους) αλλά επίσης στη φωτογραφική τεκμηρίωση τεχνουργημάτων (πινάκων, γλυπτών) ή σε φωτογραφικά σκαναρίσματα επίπεδων αντικειμένων (γραφικών, χαρτών, κειμένων)».

Μια χαρτογράφηση της σχέσης ανάμεσα στην τέχνη και στα αναδυόμενα μοντέλα τεχνητής νοημοσύνης χρειάζεται λοιπόν να λάβει υπόψη της τη σχέση της τέχνης και της τεχνητής νοημοσύνης με την ίδια την κουλτούρα και τη βιομηχανία του διαδικτύου. Ωστόσο, μια τέτοια διαπίστωση δεν πρέπει να αποσπά την προσοχή μας από την τάση που μοιάζει να επικαθορίζει το παρόν – την τάση της τεχνητής νοημοσύνης να «καταπιεί» το διαδίκτυο, όπως έγραψε γλαφυρά η Sue Halpern στο *The New Yorker* αναφορικά με το έτος 2023.

«Διαδικτυακά ενήμερη» τέχνη και τεχνητή νοημοσύνη

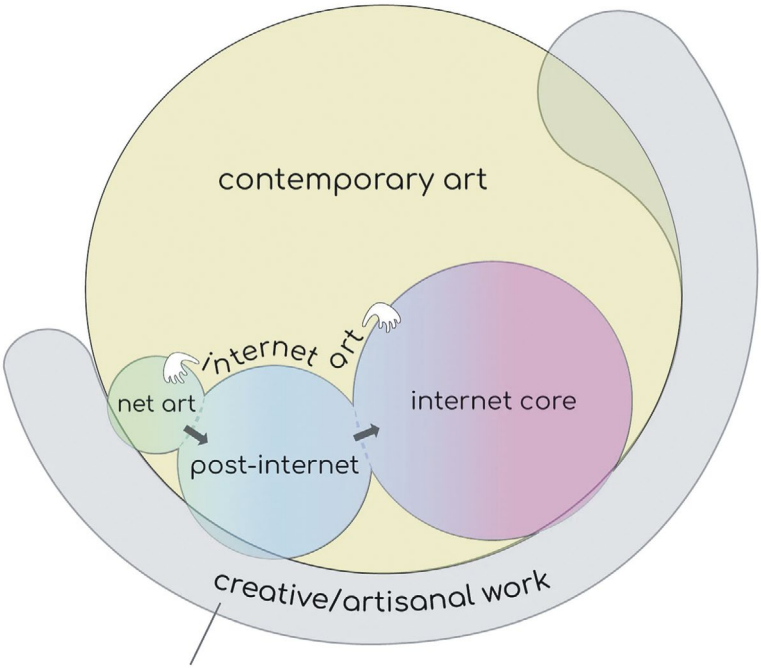
Αντί λοιπόν να αντιμετωπίζουμε τις έννοιες «τέχνη» και «τεχνητή νοημοσύνη» ως θετικά περιεχόμενα, και τη σημασία τους ως αυτοαπόδεικτη, έχει περισσότερο νόημα να δούμε τη διαπλοκή τους μέσα στην πρόσφατη ιστορία της διαδικτυακά ενήμερης (*internet-aware*) τέχνης. Δεν χρειάζεται, για κανέναν λόγο, να εφεύρουμε μια τέτοια περιοδολόγηση από το μηδέν. Συνοψίζεται διεξοδικά σε ένα άρθρο της εικαστικού Twee Whistler, αναρτημένο στο Substack με όνομα donotresearch, δηλαδή στη διαδικτυακή ερευνητική κοινότητα του εικαστικού και συγγραφέα Joshua Citarella. Το διάγραμμα που δεσπόζει στο κείμενό της σχετικά με τις διαφορετικές περιόδους της διαδικτυακής τέχνης συνοψίζει επιτυχώς τις διευκρινίσεις που πολλοί καλλιτέχνες και συγγραφείς έχουν επιχειρήσει να διατυπώσουν μέχρι τώρα. Με τρόπο που ίσως χαρακτηρίζει τον ασταθή και εφήμερο χαρακτήρα τέτοιων αναρτήσεων, το κείμενο της Whistler εξαφανίστηκε από το διαδίκτυο, τουλάχιστον τη στιγμή της συγγραφής αυτού του κειμένου, αφήνοντας πίσω μόνο ένα χρήσιμο στιγμιότυπο οθόνης (screen grab) αλλά και κάποιες ιδέες που κάποιοι λίγοι, όπως ο συγγραφέας του ανά χείρας κειμένου, πρόλαβαν να συζητήσουν και να διακινήσουν. Ένα μεγάλο τμήμα της εικαστικής δουλειάς της Whistler αποτελεί fan art αφιερωμένη στον μεταδιαδικτυακό καλλιτέχνη Jon Rafman, σε βαθμό που σε ωθεί να υποψιαστείς την παρουσία του ίδιου του Rafman πίσω από τα διαδικτυακά της ίχνη. Στο διάγραμμα που διασώζεται, ένας μεγάλος κύκλος εκπροσωπεί το σύνολο της «σύγχρονης τέχνης». Εντός του, τρεις μικρότεροι διαδοχικοί κύκλοι/σύνολα εκπροσωπούν τη «διαδικτυακή τέχνη», σηματοδοτώντας διακριτές στιγμές και περιόδους εντός της. Καθεμία από αυτές τις περιόδους αντιστοιχεί και σε μια διαφορετική αντίληψη, αντιμετώπιση, διαχείριση της σχέσης ανάμεσα στην αισθητική πράξη, το διαδίκτυο, τις τεχνολογίες των δεδομένων.

Αντί να κατανοήσουμε αυτά τα σύνολα ως οριστικές και τελεσίδικες περιόδους της διαδικτυακά ενήμερης τέχνης, μπορούμε ίσως να τα δούμε ως αισθητικές νοοτροπίες που δύνανται να ανανεωθούν ανάλογα με τις δυναμικές που τις περιβάλλουν.

Το σύνολο λοιπόν της διαδικτυακής τέχνης (*internet art*) περιέχει τα υποσύνολα της *net.art*, της *post-internet art* και το υποσύνολο του *internet-core*, έννοια που εισηγείται η Whistler για να περιγράψει ορισμένα αισθητικά φαινόμενα του πανδημικού και μεταπανδημικού κόσμου. Σύμφωνα με το σκεπτικό που αναπτύσσει, καθεμία από τις εν λόγω περιόδους αντιστοιχεί σε έναν διακριτό ρόλο του «καλλιτέχνη του διαδικτύου» και σε μια διακριτή αντίληψη για τη σχέση της καλλιτεχνικής πράξης με την τεχνητή νοημοσύνη. Καθώς λοιπόν συνοψίζω τις γόνιμες διευκρινίσεις της Whistler, θα προσθέτω σκέψεις και συνειρμούς που αντανακλούν τη στιγμή της συγγραφής.

Στην εποχή της net.art ο καλλιτέχνης είναι πρωτίστως "hacker" ή "hacktivist". Αν και η net.art αντιστοιχεί σε ένα συγκεκριμένο δίκτυο ή κίνημα καλλιτεχνών της δεκαετίας του 1990 που χαρακτηρίζεται από αυτό που ο κριτικός Tilman Baumgärtel αποκάλεσε *media specificity* –αποκλειστική χρήση του διαδικτύου ως μέσο– τα χαρακτηριστικά της net.art φαίνεται να αντανακλούν ένα ευρύτερο πνεύμα διαδικτυακής παραγωγής που διασώζεται ως σήμερα: αποκλειστικά ψηφιακή τέχνη που κυμαίνεται από έργα ηλεκτρονικής αλληλογραφίας μέχρι την παραγωγή (λειτουργικού ή μη-λειτουργικού) κώδικα και την πρόκληση απρόβλεπτων διαδικτυακών συμβάντων. Καλλιτεχνικές οντότητες όπως οι Jodi και οι 0100101110101101.org (αλλιώς Eva & Franco Mattes) αποτελούν επιτομές της πειρατικής κουλτούρας του προγραμματιστή (coder), του χάκερ, του φαρσέρ (prankster), με τους δεύτερους να εκπροσωπούν επιτυχώς μια διεύρυνση, πέραν της ειδικότητας του διαδικτύου ως μέσο.

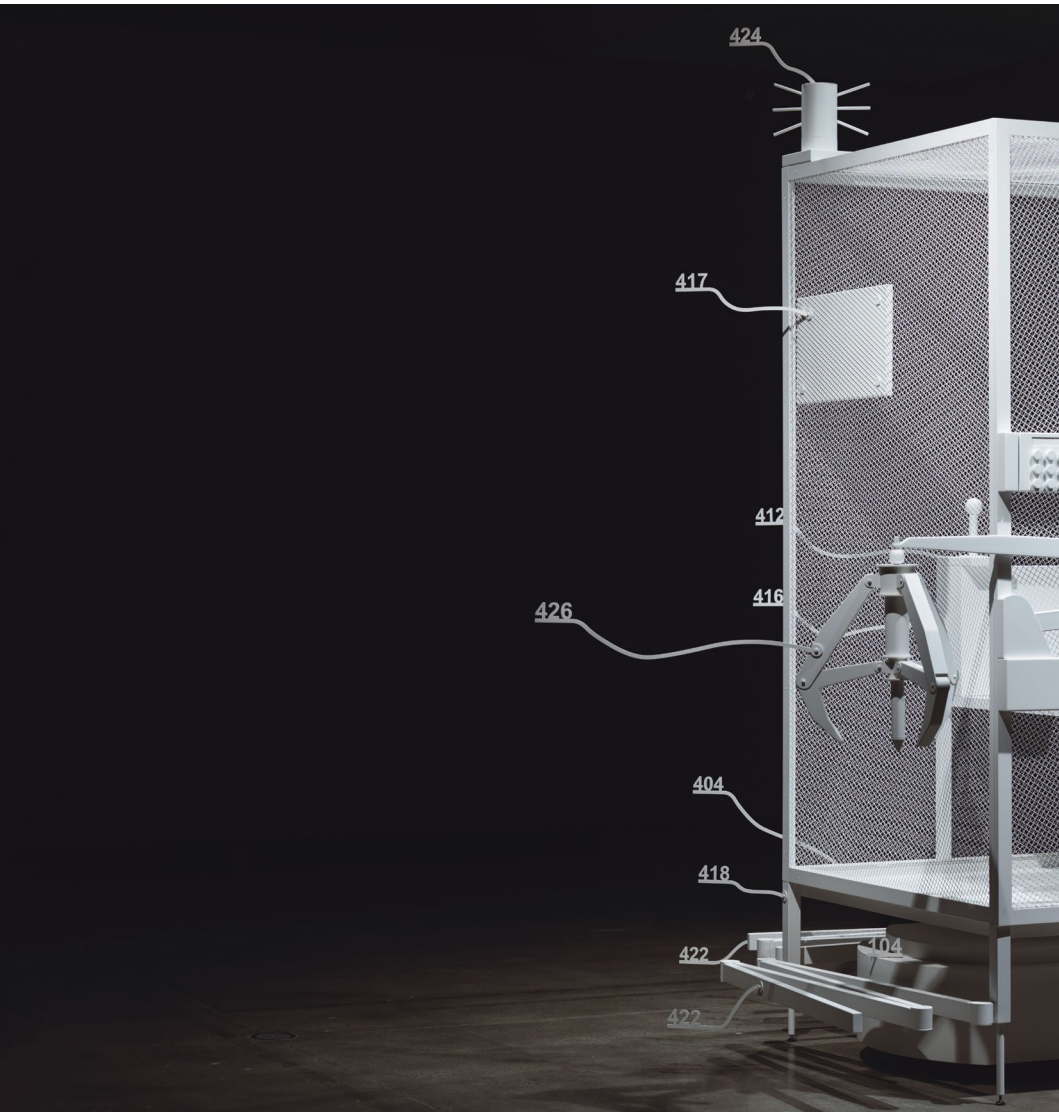
Ένας εικαστικός που εκπροσωπεί το πνεύμα της net.art στο σημερινό υβριδικό περιβάλλον είναι ο Ολλανδός Constant Dullaart. Στο έργο του *DullDream* (2015) ανέπτυξε ένα νευρωνικό δίκτυο που εκτελούσε την αντίστροφη λειτουργία από αυτή του DeepDream της Google: αντί να υπογραμμίζει και να εντείνει τα χαρακτηριστικά και τα μοτίβα των μορφών, τα καθιστούσε θολά, περισσότερο αδιαφοροποίητα, αφαιρώντας, για παράδειγμα, τα ειδικά στοιχεία που κάνουν ένα

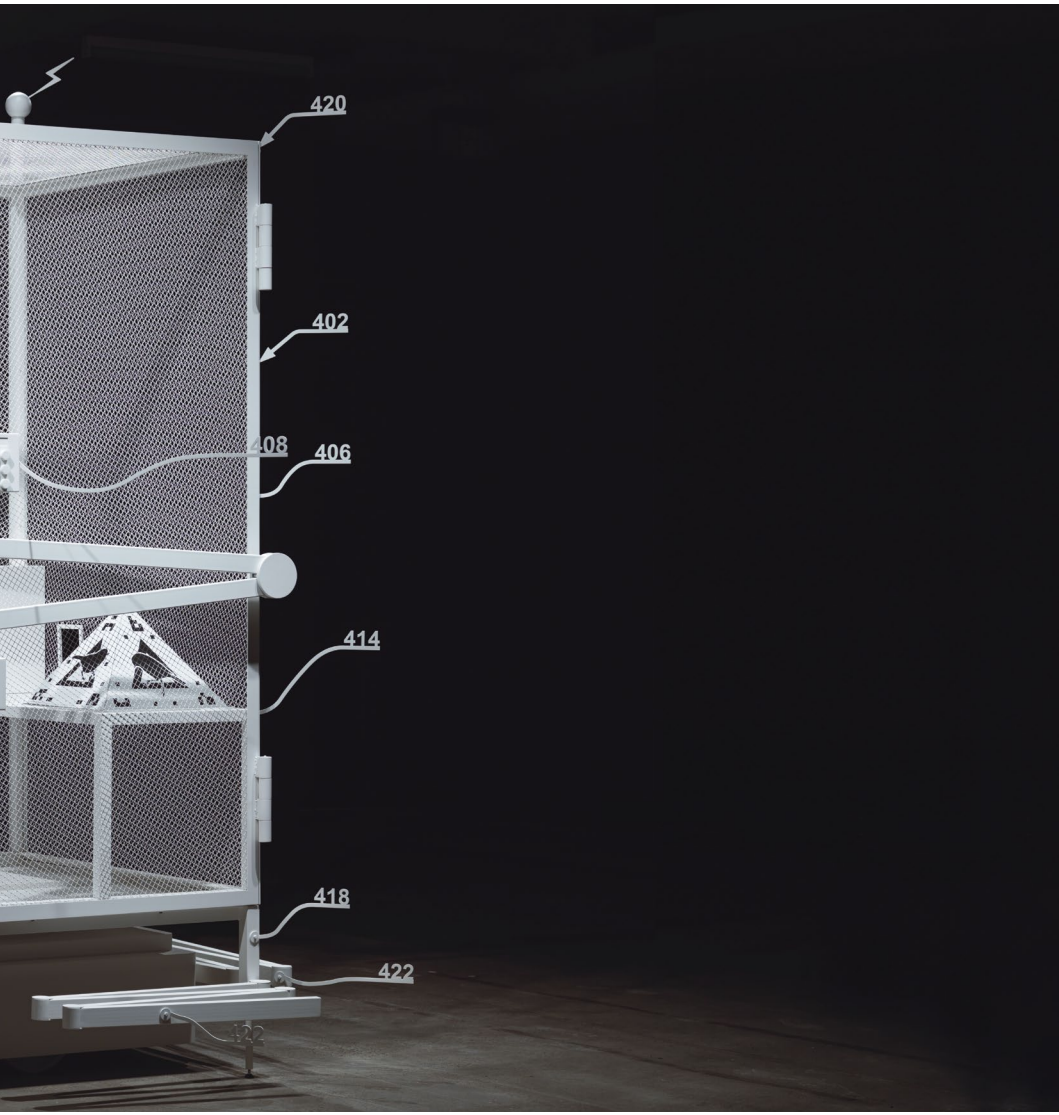


Twee Whistler, *Internet Core Diagram,* 2023. Πηγή: donotresearch

πρόσωπο να αναγνωρίζεται ως τέτοιο. Σκοπός ήταν η υπονόμευση της τεχνητής νοημοσύνης ως εργαλείο μηχανικής εκμάθησης και αναγνώρισης. Το μοντέλο του Dullaart επέτρεπε στο κοινό να αναρτά ευκρινείς εικόνες και να τις παραλαμβάνει δίχως τα ιδιαίτερα χαρακτηριστικά τους. Άλλος ένας σημαντικός καλλιτέχνης, το έργο του οποίου αντανακλά έντονα το πνεύμα της net.art, είναι ο Mark Amerika. Το κλασικό net.art έργο του *GRAMMATRON* (1993-1997) ήταν ένα κυβερνοπάνκ λογοτεχνικό κείμενο στη μορφή ενός υπερκειμένου (hypertext): ενός χώρου πλοήγησης με υπερσυνδέσμους (hyperlinks) που οδηγούν σε κείμενα, soundtracks, κινούμενες και σταθερές εικόνες. Το πρόσφατο βιβλίο του *My Life as an Artificial Creative Intelligence*, κάτι ανάμεσα σε αυτοβιογραφία, θεωρία των μέσων και κυβερνοπάνκ λογοτεχνία, αποτελεί απόπειρα μετεξέλιξης των πειραματισμών του με το υπερκείμενο – μια προσπάθεια να γράψει κανείς «μεταλογοτεχνία» (metafiction) συνεργαζόμενος με μηχανικές μορφές «νοημοσύνης» και «δημιουργικότητας», όπως οι εφαρμογές των μεγάλων γλωσσικών μοντέλων (LLMs, ChatGPT 2 και 3). Τα έργα των Dullaart και Amerika, αν και φαίνεται να διαφέρουν ως προς τις προθέσεις και την υλοποίησή τους, αντιμετωπίζουν την παραγωγική τεχνητή νοημοσύνη ως μέσο ή εργαλείο για την παραγωγή κειμένου, εικόνων, αισθητικών συμβάντων που προορίζονται να αναστατώσουν την αντίληψή μας για την τεχνολογικά διαμεσολαβημένη εμπειρία μας ή να «ξαναπούν» την ιστορία μας.

Σε αντιδιαστολή με την «πειρατική» νοοτροπία της net.art, το παράδειγμα που θα χαρακτηριστεί ως μεταδιαδικτυακή τέχνη (post-internet art, 2008-) θα φέρει στο προσκήνιο, όπως εξηγεί η Whistler, τη σημασία του καλλιτέχνη ως "surfer", δίνοντας αξία στην ίδια την πράξη της πλοήγησης μέσα στο φεουδαλιστικό Web 2.0 και αναζητώντας τι απέγινε ο εαυτός, οι κοινωνικές σχέσεις, οι αναπαραστάσεις, οι υλικότητες στην εποχή της καθολικής και διαρκούς δικτύωσης του κόσμου. Ο καλλιτέχνης της γενιάς του post-internet δρα σε μια εποχή που το διαδίκτυο έχει πάψει προ πολλού να αποτελεί υπόσχεση δημοκρατικής αλλαγής και ουτοπικής ανατροπής και είναι ήδη κοινοτοπία και εμπόρευμα. Σε ένα τέτοιο πλαίσιο, οι καλλιτέχνες ερευνούν την απόσταση ανάμεσα στην ιδεολογική υπόσχεση και την πραγματικότητα, μετασχηματίζοντας την αρχιτεκτονική της «τεχνητής νοημοσύνης» σε ενσώματη εμπειρία και απτική υλικότητα.


424
417
412
416
426
404
418
422
404
422



Simon Denny, *Amazon worker cage patent drawing as virtual King Island Brown Thornbill cage (US9,280,157 B2: "System and method for transporting personnel within an active workspace", 2016)*, 2019
Μέταλλο με ηλεκτροστατική βαφή, πλαστικό, ψηφιακή εκτύπωση σε χαρτόνι και διεπαφή επαυξημένης πραγματικότητας για iOS, 214,9 × 214,9 × 285,7 εκ.
Συλλογή Rachofsky. Ευγενική παραχώρηση του καλλιτέχνη και της γκαλερί Petzel, Νέα Υόρκη. Φωτογραφία: Jesse Hunniford.

Το έργο του Νεοζηλανδού καλλιτέχνη Simon Denny, *Amazon worker cage patent drawing as virtual King Island Brown Thornbill cage (US9,280,157 B2: "System and method for transporting personnel within an active workspace", 2016),* που παρουσιάστηκε το 2019, αποτελεί μια τρισδιάστατη υλοποίηση του σχεδίου του εργασιακού κελιού της Amazon. Πρόκειται για μια πατέντα με ρομποτικό βραχίονα σε ανθρώπινη κλίμακα που σχεδιάστηκε για να περιορίζει, να μεταφέρει και να προστατεύει τον εργάτη όταν αυτός εισέρχεται σε μια περιοχή εργασίας αλγοριθμικά προγραμματισμένων ρομπότ. Το *Amazon worker cage*, κάτι ανάμεσα σε λευκό κελί και σε φουτουριστικό κλαρκ, δεν υλοποιήθηκε ποτέ από την Amazon, ωστόσο στην έκθεση *Mine* παρουσιάζεται σαν κρίκος ενός ευρύτερου συστήματος εξόρυξης. Το 2019, ο Denny εκθέτει το κελί εντός μιας ευρύτερης εγκατάστασης, στο Museum of Old and New Art (MONA) του Hobart, που παίρνει τη μορφή ενός θεματικού πάρκου εξόρυξης με λεπτομερειακές παραπομπές στη βιομηχανία εξόρυξης μεταλλευμάτων αλλά και δεδομένων. Το έργο συνυπάρχει με χαρτονένιες αναπαραγωγές μηχανημάτων υπόγειας εξόρυξης φυσικών υλών. Ένα επιπλέον σημάδι μας καλεί να συνειδητοποιήσουμε το νέο μείγμα εργασιακής εκμετάλλευσης, εξόρυξης σπάνιων γαιών και ενεργειακής δαπάνης που συνεπάγεται η διόγκωση της βιομηχανίας της τεχνητής νοημοσύνης. Εντός του κελιού, ο θεατής εντοπίζει το εκτυπωμένο σχήμα του καστανού Θόρνμπιλ του King Island της Τασμανίας. Με τη χρήση επαυξημένης πραγματικότητας, το πτηνό τιτιβίζει το τραγούδι του προειδοποιώντας τον επισκέπτη για επερχόμενους κινδύνους, σαν ένα άλλο καναρίνι του ανθρακωρυχείου.

Ανάμεσα στα εκθέματα του Denny κυριαρχεί το επιτραπέζιο παιχνίδι *Extractor* (*Extractor Board Game XL Display Prototype,* 2019). Το *Extractor,* μια διασκευή του αυστραλιανού Squatter, είναι μια μεγάλη επιτραπέζια πλατφόρμα που επιτρέπει στους παίκτες να γίνουν managers εταιρειών εξόρυξης δεδομένων, να αξιοποιήσουν τις πληροφορίες που τους προσφέρει η συνολική έκθεση/εγκατάσταση ώστε να χτίσουν εταιρικά μεγαθήρια. Ανατέμνοντας το νεοκαπιταλιστικό φαντασιακό, οι κανόνες του παιχνιδιού αναγκάζουν τους παίκτες να ξοδέψουν τα εικονικά τους χρήματα σε υπηρεσίες ψύξης, να χρηματοδοτήσουν καμπάνιες εναντίον της θεωρίας της κλιματικής αλλαγής ή να αποκαταστήσουν τις ζημιές που προκάλεσε ένας τυφώνας στην τράπεζα δεδομένων. Τα στοιχεία που συγκροτούν τη στρατηγική του Denny, αλλά και πολλών άλλων καλλιτεχνών του post-internet, είναι η διάθεση για

πλοήγηση, η παιχνιδοποίηση των υλικών και οικονομικών σχέσεων, η σκηνοθεσία μιας διφορούμενης εμπειρίας εντός της οποίας τόσο τα υποκείμενα όσο και οι τεχνολογίες μηχανικής εκμάθησης εμπλέκονται στην αναπαραγωγή των δομών συσσώρευσης, αποθησαυρισμού και νομισματοποίησης (monetisation).

Η στάση πολλών καλλιτεχνών της γενιάς του post-internet χαρακτηρίζεται από αυτήν τη δυσπιστία σχετικά με τις προοπτικές μιας μετωπικής ανατροπής των συστημάτων που μας περιβάλλουν ή μας εγκλωβίζουν: πριν «ανατρέψουμε» κάτι, οφείλουμε να πλοηγηθούμε εντός του και να χαρτογραφήσουμε τις δυνατότητες που μας προσφέρει κάθε «σύμπτωμα» που φαίνεται να στηρίζει την περιρρέουσα συνθήκη. Γύρω στο 2021, κατά την επιδημία των NFTs, ο Denny βρισκόταν σε επαφή με τον David Holz, που σύντομα θα γινόταν ιδρυτής του Midjourney. Παρατηρώντας τις προσπάθειες του Holz με ένα υπό διαμόρφωση πρότυπο text-to-image μοντέλο, αποφάσισε να προβεί σε μια αρχαιολογία των εικόνων του διαδικτύου. Επηρεασμένος από το Fucked Company –έναν ιστότοπο που στόχευε στην καταγραφή αποτυχημένων εταιρειών dot-com από την περίοδο του 2000-2002, όταν η φούσκα των υποσχόμενων αυτών διαδικτυακών αστέρων/start ups έσπασε με μια χρηματιστηριακή βουτιά– αναζήτησε τρόπους να δώσει σε αυτά τα ερείπια του πρόσφατου παρελθόντος μια δεύτερη ευκαιρία. Σε συνέργεια με τον Holz, ο οποίος έτρεχε αυτό το πρώιμο Midjourney και στο οποίο έδωσε το ψευδώνυμο Cosmographia, χρησιμοποίησε το μοντέλο του για να ζητήσει από την τεχνητή νοημοσύνη να φανταστεί πώς μπορεί να ήταν τα εμπορικά σήματα αυτών των εταιριών, όπως οι elaw.com, capacityweb.com, alladvantage.com, κ.ά. Στη συνέχεια, ο Denny παραχώρησε αυτές τις «ελλιπείς μνήμες» του διαδικτύου στον Guile Twardowski, έναν καλλιτέχνη των NFTs, διάσημο για τα πετυχημένα *CryptoKitties*, ζητώντας του να κατασκευάσει τις δικές του, πιο αντιπροσωπευτικές, εκδοχές των εμπορικών αυτών σημάτων σε μορφή NFTs. Σε συνεργασία, οι τρεις αυτοί πλάνητες των υπονόμων του Web 1.0 δημιούργησαν το πρότζεκτ *Dotcom Séance*, μια συνεδρία κλήτευσης των εταιρικών αυτών φαντασμάτων μέσω των λογοτύπων τους που δημιουργήθηκαν με τεχνητή νοημοσύνη. Όπως εξηγεί ο Denny, «ταϊσμένα με προτροπές που περιγράφουν την κάθε εταιρία, αυτά τα παραγόμενα με τεχνητή νοημοσύνη εμπορικά σήματα λειτουργούν, στην ουσία, ως στιγμές κατά τις οποίες το διαδίκτυο ξαναθυμάται τον εαυτό του». Τα καινούργια σήματα μπορούσαν να πουληθούν ως

NFTs, ενώ οι αγοραστές των λογοτύπων μπορούσαν να γίνουν μεσαίοι managers των εταιρειών αυτών, ή διευθύνοντες σύμβουλοι (CEOs) αν αγόραζαν τις μοναδικές εκδοχές του Twardowski και συνέδεαν τα ηλεκτρονικά τους πορτοφόλια του Ethereum με τη συγκεκριμένη αγορά. Η κοινότητα των αγοραστών αναγεννούσε την εταιρεία σε μια νέα μορφή μέσω των επιλογών που παρέχει η blockchain τεχνολογία.

Τα ανακριβή αποτελέσματα του "prompting" που στόχευε στην ανακατασκευή του κάθε εταιρικού σήματος μαρτυρούν το μέγεθος της «διαδικτυακής λήθης». Παρότι τίποτα δεν εξαφανίζεται από το διαδίκτυο, κάθε ανασύνθεση του παρελθόντος αναδύεται ως επίδικο. Τα σύγχρονα text-to-image μοντέλα υπογραμμίζουν την καθαρά φαντασματολογική φύση του διαδικτύου.

Διαδικτυακά ενήμερη τέχνη, πέραν της αλγοριθμικής ηδονής;

Τέλος, μια σύγχρονη «διαδικτυακά ενήμερη» τέχνη μπορεί να έχει τη μορφή αυτού που η Whistler ονομάζει internet-core, επιχειρώντας να ορίσει μια «τρίτη περίοδο», ή μια νέα στάση που συνδέεται με το ξέσπασμα και την εξάπλωση της πανδημίας αλλά και τη ζωή σε συνθήκες υγειονομικού περιορισμού – μια εποχή κατά την οποία, όπως λέει, το ίντερνετ παύει να αποτελεί μια εναλλακτική πραγματικότητα και γίνεται η «μόνη υπαρκτή πραγματικότητα».

Στη μετα-αποκαλυπτική επικράτεια της προφύλαξης, του εγκλεισμού, των υγειονομικών ρυθμίσεων και του θανάτου, η διαδικτυακή ζωή σημαδεύεται από τη ραγδαία διάδοση σοσιαλμιντιακών αισθητικών μικρο-ειδών (micro-genres). Ορισμένα από αυτά, όπως το #cottagecore, γεννιούνται λίγους μήνες πριν το ξέσπασμα της πανδημίας και γίνονται δημοφιλή κατά την περίοδο του εγκλεισμού. Η GenZ παρακολουθεί στο TikTok τους πρωταγωνιστές του #goblincore να ντύνονται με ρούχα από καταστήματα μεταχειρισμένων (thrift shops) και να περιηγούνται στην «άγρια» ή «άσχημη» πλευρά του φυσικού τοπίου. Τα μικρογένη αυτά συμπλέουν ραγδαία στη γέννηση του #corecore. Το #corecore συνοψίζει αυτοσαρκαστικά όλα τα προηγούμενα cores και συστήνεται ως συρραφή από διαφορετικά κλιπς, επενδυμένη με ένα μυσταγωγικό μουσικό φόντο. Διαδίδεται πρωτίστως στο TikTok το 2020, ενώ το 2022 η δημοφιλία του εκτοξεύεται. Αρχικά, αποκτά τη μορφή παράδοξου μοντάζ των δεκαπέντε δευτερολέπτων, όπου βίντεο συρράπτονται

μεταξύ τους και εναλλάσσονται δείχνοντας περίεργους χορούς, anime, υλικό από βιντεοπαιχνίδια, ζωάκια, σκηνές από το *Blade Runner* ή το *American Psycho*, συνοδευόμενα συχνά από ambient μουσική του Aphex Twin.

Σταδιακά, η αισθητική του #corecore ηρεμεί και απλώνεται στον χρόνο, αντικαθιστώντας τη φρενήρη υπερδιέγερση του μανιώδους scrolling με σχετική ημέρευση, διεκδικώντας τον ύπνο μέσα στη γενικευμένη αϋπνία. Σύντομα, τα #corecore βίντεο θα ξεκινήσουν να θυμίζουν ασκήσεις διαλογισμού. Στιγμιότυπα από περίεργες πτήσεις αεροπλάνων συνυπάρχουν με συναισθηματικές εξάρσεις κινηματογραφικών πρωταγωνιστών και εξομολογητικά αποσπάσματα από διαδικτυακές συνεντεύξεις διαφόρων διάσημων. Όπως επισημαίνουν οι Declan Colquitt και Hannah Cobb, «μια από τις επαναλαμβανόμενες θεματικές και στυλιστικές στροφές εντός του #corecore ήταν, για να το θέσουμε απλά, [η διάθεση] να κάνουμε ένα βήμα πίσω και να δούμε τη μεγαλύτερη εικόνα. Να επιχειρήσουμε να κατανοήσουμε τις διαδικασίες που ξεδιπλώνονταν (και ακόμα ξεδιπλώνονται) κάνοντας το παρόν τόσο ακραίο, έντονο και πυκνό». Η προσοχή μιας κρίσιμης μάζας χρηστών έχει ήδη μετατοπιστεί από τα «εξουθενωτικά» rabbit holes του YouTube και των παλαιότερων κοινωνικών δικτύων σε κάτι νέο. Οι νέες πλατφόρμες είναι το TikTok, το Telegram, το Substack, το Patreon. Η νέα γη δεν είναι η οθόνη του laptop αλλά αυτή του smartphone, ο νέος ορίζοντας διάδοσης της συνθήκης που η Whistler συνοψίζει εφευρίσκοντας τον όρο internet-core.

Η αισθητική ανάδυση και εξέλιξη του #corecore μπορεί, επίσης, να γίνει κατανοητή ως μηχανισμός διαχείρισης του λεγόμενου *doomscrolling*, του εμμονικού, ψυχαναγκαστικού scrolling πάνω σε καταστροφολογικό περιεχόμενο. Το doomscrolling περιγράφεται συχνά ως πρακτική αυτοαφανισμού μέσω της ολικής καθήλωσης σε ένα newsfeed που έχει γεμίσει από οριακό, μετα-αποκαλυπτικό, σοκαριστικό περιεχόμενο. Σε μια εύγλωττη περιγραφή του φαινομένου, ο Erik Davis περιγράφει το doomscrolling ως το στάδιο που «νευρωτικά και εθιστικά προσκολλάσαι στο εισερχόμενο πληροφοριακό σμήνος τωρινών και πιθανών τρόμων, ρουφώντας μανιακά τα σκοτεινά ρεύματα σαν να ήταν αποτσίγαρα του υπονόμου». Τούτων δοθέντων, το doomscrolling μοιάζει σαν περίεργος δεσμός ανάμεσα στον αλγόριθμο και την «ορμή του θανάτου» – ένας δεσμός που μας καλεί να μην τον αγνοήσουμε αλλά

να του δώσουμε χώρο, μορφή, ρυθμό, διάρκεια. Το #corecore, όπως και το internet-core της Whistler, είναι αφορμή να στοχαστούμε μια τέχνη «πέραν της αρχής της ηδονής», δηλαδή, μια τέχνη ενήμερη για τη διαπλοκή της απόλαυσης και της εντροπίας εντός της συλλογικής μας «πλοήγησης στον χαμό». Αναγνωρίζοντας την τερματική διάσταση της συνθήκης που μας περιβάλλει, η Whistler αποδίδει στον μεταπανδημικό καλλιτέχνη τον ρόλο του "poster" (και λιγότερο του χάκερ ή του πλοηγού). Σύμφωνα με την Whistler, το καλλιτεχνικό αυτό υποκείμενο μηχανεύεται πρωτίστως το shitposting, το schizoposting, το prompt-engineering. Αναρτά περιεχόμενο ακονίζοντας, ταυτόχρονα, τα μαζικά εργαλεία των εφαρμογών της παραγωγικής τεχνητής νοημοσύνης. Στη νέα αυτή συνθήκη, συμβατικά θραύσματα εικόνων συνυπάρχουν με παραγόμενες εικόνες σε ένα οικοσύστημα που μιμείται την οπτική οικονομία ενώ, ταυτόχρονα, παράγει συνάψεις ζωτικής σημασίας.

Το σύντομο αλλά σαρωτικό ιταλικό brainrot του 2025 είναι ένα χρήσιμο δείγμα του νέου αυτού οικοσυστήματος. Ένας καλός φίλος, που διδάσκει εικαστικά στη δευτεροβάθμια εκπαίδευση, μου περιέγραψε πως το bullying που δεχόταν από τους μαθητές του (φαινόμενο καθιερωμένο και σχεδόν καθημερινό) αντικαταστάθηκε, εν μία νυκτί, από ερωτήματα του τύπου «κύριε, μήπως είστε tralalero tralala;» κ.ο.κ. Αυτές οι «εκρήξεις» περιεχομένου είναι μερικές καθηλώσεις νοήματος, βραχυπρόθεσμες προσδέσεις γύρω από κάποια σημαίνοντα που παρέχουν σχήμα στη ροή του κόσμου επειδή μπορούν να «ακουστούν» πέραν των αυστηρά προσδιορισμένων επικρατειών του νοήματος: το χαρακτηριστικό τους ακουστικό περιεχόμενο, τα ψευτοϊταλικά voiceovers, αντιστοιχούνται από τους χρήστες σε οικουμενικά αναγνωρίσιμες παραγόμενες εικόνες χιμαιρικών πλασμάτων που σύντομα θα τροφοδοτήσουν μια διαδικτυακή εγκυκλοπαίδεια, ένα παιχνίδι ανταλλαγής καρτών και πολλά άλλα.

Αυτή η ιδέα του μεταπανδημικού καλλιτέχνη ως "(shit)poster" επιτρέπει να κατανοήσουμε και κάποιες εκδοχές δράσης που συγγενεύουν περισσότερο με την κληρονομιά των "tactical media". Το 2023, ορισμένες κοινότητες fan fiction, δηλαδή ερασιτεχνών συγγραφέων που ποστάρουν κείμενα γραμμένα στη βάση υπαρκτών κινηματογραφικών, λογοτεχνικών και άλλων κόσμων, ανακάλυψαν ότι ορισμένες generative εφαρμογές «ταΐζονται» με το υλικό που οι ίδιοι/ίδιες αναρτούν. Ορισμένοι/ες fan fiction συγγραφείς που ανήκουν στην ερωτική κοινότητα/υποκουλτούρα του Omegaverse ανακάλυψαν ότι το Sudowrite, αυτός ο επιτυχημένος

κειμενογράφος TN, τροφοδοτούνταν με υλικό (training sets) από το AO3, το βασικό φόρουμ fan fiction διαμοιρασμού ερασιτεχνικής Omegaverse ερωτικής λογοτεχνίας. Έτσι, οργάνωσαν έναν μαραθώνιο συγγραφής με τίτλο "Knot in My Name", για μια ολόκληρη εβδομάδα, με στόχο να «πλημμυρίσουν» το Sudowrite, κι άλλους κειμενογράφους, με την υπερειδική ορολογία της Omegaverse κοινότητας.

Το σημαντικότερο χαρακτηριστικό όμως αυτής της αναπόδραστης συνάντησης με τα προϊόντα των generative εφαρμογών είναι ότι τα όρια μεταξύ συμβατικού και generated περιεχομένου καθίστανται ρευστά, διαμορφώνοντας ένα καθημερινό mood board που δεν μπορεί να αφήσει ανεπηρέαστη την υλική συνθήκη της καλλιτεχνικής εργασίας. Καλλιτέχνες που συνδέθηκαν με τη γενεαλογία της μεταδιαδικτυακής τέχνης επιλέγουν να ανταποκριθούν στο υβριδικό αυτό παρόν αναπτύσσοντας εξαιρετικά σύνθετες στρατηγικές του prompt engineering. Ανταγωνίζονται τις αλγοριθμικές ροές περιεχομένου με ένα επιθετικό shitposting υλικού βασιζόμενου σε απρόσμενα, εικοτολογικά και οιονεί ανατρεπτικά prompts. Εδώ και καιρό, ο λογαριασμός στο Instagram του Rustan Söderling έχει μετατραπεί σε βεστιάριο προτροπών – ένας χάρτης πιθανοτήτων που περιλαμβάνει εξεγερμένα πλήθη, βιομηχανικά συμπλέγματα, άγνωστες θρησκευτικές σέκτες, νεόκοπα επαγγέλματα, ακαθόριστες ιατρικές εφαρμογές, εξοπλισμό προστασίας για αναπάντεχες καταστάσεις, ντοκουμέντα συνάντησης με μυθικές υπάρξεις, κ.ά. Ομοίως, ο Jon Rafman χρησιμοποίησε εξαιρετικά σύνθετα prompts για να δημιουργήσει ορισμένα μουσικά βίντεο για το άλμπουμ *Vultures* του, πολιτικά απεχθούς, πλέον, Kanye West. Σε ένα αλληγορικό αναδιπλασιασμό του παρόντος, μετα-ανθρώπινοι λυκάνθρωποι, σατανικοί κλόουν και ανθρωπόμορφα κοράκια καταλαμβάνουν τα αμερικανικά προάστια.

Αλλά, για να ερευνήσουμε βαθύτερα τη διάγνωση της Whistler περί του μεταπανδημικού καλλιτέχνη ως poster, αξίζει να εστιάσουμε στον διάλογο ανάμεσα στα πρώιμα text-to-image μοντέλα και στην κουλτούρα του posting, όπως αυτός ξεδιπλώνεται στο χαρακτηριστικό έργο του Rafman, *Counterfeit Poast* (2022). Στην έκθεση του έργου στην Sprüth Magers του Βερολίνου, η περιήγηση σε μια οκτάγωνη αίθουσα με ταπετσαρίες και αλγοριθμικά δημιουργημένες εικόνες που θυμίζουν παστωμένες ζωγραφικές επιφάνειες οδηγεί σε έναν σκοτεινό χώρο που φιλοξενεί το βίντεο *Counterfeit Poast*, μια αφηγηματική συγχώνευση



Jon Rafman, *Counterfeit Poast*, 202.
4K βίντεο, ήχος, 39'53'' © Jon Rafman.
Ευγενική παραχώρηση του καλλιτέχνη και της γκαλερί Sprüth Magers

generated υλικού που χρησιμοποιεί ως πρώτη ύλη διάφορα δημοφιλή copypasta: κομμάτια κειμένου που κυκλοφορούν ευρέως στο διαδίκτυο ως posts σε κοινωνικά δίκτυα και πλατφόρμες όπως το Reddit ή το 4chan. Τα αποσπάσματα αυτά έχουν συνήθως χιουμοριστική πρόθεση, προβοκατόρικη ή τρολλαριστική στόχευση. Όπως εξηγεί το δελτίο τύπου της Sprüth Magers, το συγκεκριμένο έργο είναι τμήμα ενός ευρύτερου πρότζεκτ του Rafman που επιχειρεί να μετασχηματίσει τα σκουπίδια των data σε εικόνες που λένε κάτι για τη σχέση ανάμεσα στην εικόνα και τη μνήμη μας στον εικοστό πρώτο αιώνα. Οι φωνές των πρωταγωνιστών βάζουν, μέσω μιας εξιστόρησης, τις αλγοριθμικά δημιουργημένες εικόνες σε μια αφηγηματική σειρά ονειρικού χαρακτήρα, αντλώντας από το κινηματογραφικό στυλ που εντοπίζεται στο *La Jetée* του Chris Marker ή το *The Falls* του Peter Greenaway. Το ψηφιακό πορτραίτο του πρωταγωνιστή της κάθε ιστορίας ζωντανεύει και μιλάει μέσω μιας iPhone εφαρμογής αναγνώρισης προσώπου.

Κάθε νέο κεφάλαιο του βίντεο εισάγεται μέσω μιας λευκής οθόνης που περιέχει πληροφορίες, όπως το όνομα του χρήστη και την ημερομηνία δημοσίευσης του εκάστοτε κειμένου. Αν και τα εν λόγω copypasta παραπέμπουν σε κατασκευασμένα ή μισο-αληθινά συμβάντα, λουσμένα στο μυστήριο και την πρόκληση, το text-to-image μοντέλο ανακαλεί υπαρκτές εικόνες που κάτι θυμίζουν σε όλους μας. Στο τέλος, το CLIP-guided diffusion εργαλείο του Rafman συνθέτει ένα παραισθησιακό νουάρ συνενωμένων επεισοδίων που ισορροπεί ανάμεσα σε παρανοϊκό ιδεασμό, ελπίδα, αυτοαφανισμό του χρήστη, ειλικρίνεια και μνημονική διέγερση. Στο πρώτο κλιπ, μια γυναίκα επιχειρεί να θυμηθεί μια ταινία με θέμα την αποκάλυψη και πρωταγωνιστή τον Kevin Costner. Παρά τις έντονες μνήμες της από την κινηματογραφική προβολή της ταινίας, ο κοινωνικός της περίγυρος μοιάζει να την αγνοεί παντελώς ενώ η ίδια αδυνατεί να εντοπίσει στοιχεία για την ύπαρξη του φιλμ. Σε ένα επόμενο κλιπ, ένας κυνικός νεαρός εξηγεί πώς το μίσος του για τα σκυλιά τον ώθησε να εξαφανίσει το νεκρό σκυλί της συντρόφου του, το οποίο πάτησε κατά λάθος με το αμάξι του. Περιγράφει πως, για χρόνια, έκανε gaslight τη σύντροφό του, επιμένοντας ότι ο σκύλος δεν υπήρξε ποτέ, μέχρι που αυτή πέθανε. Σε ένα άλλο, ένας στρατόφιλος νεαρός εξηγεί πως ένας σκληροτράχηλος θείος έγινε πρότυπο και μέντοράς του επιβάλλοντάς του μια 'sigma male' ρουτίνα με τρέξιμο, Adderall και αυνανισμό. Αυτή η συρραφή αφηγήσεων αποκτά τη μορφή ενός σάρκινου εφιάλτη που αυτο-ηθογραφεί τον καταναλωτή οριακού



Jon Rafman, ᵏᵂᶻᵠᵃ·ᶯᵠᵒᶻᵏᶻᵗ (*Nuclear Kiss)*, 2022
Εκτύπωση inkjet και ακρυλικό σε καμβά, 186,7 × 134,6 εκ. | 73,5 × 53 ίντσες
© Jon Rafman. Ευγενική παραχώρηση του καλλιτέχνη και της γκαλερί Sprüth Magers
Φωτογραφία: Ingo Kniest

περιεχομένου. Στην πρακτική του Rafman, ο μόνος τρόπος να μη νιώθεις απομονωμένος και αλλοτριωμένος από τις ωμές πραγματικότητες της διαδικτυακής ζωής είναι να τις αντικρίσεις έντιμα και να παντρευτείς ορισμένες από τις αντιφάσεις τους.

Όπως παρατηρεί ο Matteo Calla αναφορικά με το πρότζεκτ *Counterfeit Poast* (2022) του Rafman, καθεμία από τις παραγόμενες εικόνες προϋποθέτει χιλιάδες απορριφθείσες προσπάθειες, εμπεδώνοντας μια κουλτούρα που θέτει πλέον τις δικές της προϋποθέσεις αισθητικής παραγωγής. Μια τέτοια διαδικασία αντανακλά, εν τέλει, μια σύγκρουση ανάμεσα στον καλλιτέχνη και τη μηχανή, μια αναντιστοιχία που οδηγεί σε κάτι παραπάνω από ένα «μοιραίο» περιεχόμενο. Το σημερινό έργο του Rafman, όπως και αυτό πολλών άλλων δημιουργών που δεν διστάζουν να πλοηγηθούν ανάμεσα στους τεχνολογικούς τρόμους της εποχής μας, μοιάζει σαν μια σκηνή που, όπως εύστοχα επισημαίνει ο Calla, στεγάζει μια ένταση ανάμεσα σε κατεστημένες έννοιες της ανθρώπινης δημιουργίας και της αλγοριθμικά διαμεσολαβημένης ζωής. Διαπιστώνοντας κάτι ανάλογο, η Joanna Zylinska προτείνει ότι θα μπορούσαμε να δούμε την τεχνολογία πίσω από τα μοντέλα της παραγωγικής τεχνητής νοημοσύνης σαν κάτι παραπάνω από μια «ορθή» αντιστοίχιση της φυσικής γλώσσας με την εικόνα – να τα δούμε σαν «conversation pieces, προκλήσεις ή, εντέλει, προτροπές που ανοίγουν έναν διάλογο, μαζί μας και για εμάς, πάνω στη θεμελιακή ασυμμετρία ανάμεσα στη λέξη και την εικόνα, ανάμεσα στον κόσμο και την αναπαράστασή του». Από αυτήν την οπτική γωνία, η μηχανική του prompting δεν είναι μια τελειοποιητική εργασία αφημένη στο έλεος των προγραμματιστών αλλά μια νέα πίστα του κοινωνικού ανταγωνισμού. Η μορφή που δίνουμε σε αυτές τις εντάσεις είναι ζήτημα εξόχως πολιτικό.

Σημειώσεις

[1] Possati, M. L. (2021). *The algorithmic unconscious: How psychoanalysis helps in understanding AI*. Routledge.
[2] Οφείλω αυτή την παρατήρηση στον Simon Denny. Δείτε τη συμβολή του Denny στο A questionnaire on art and machine learning. *October, 189*, Καλοκαίρι 2024, σελ. 26.
[3] Αυτό το θέμα συζητείται διεξοδικά από την Louise Amoore (2009) στο Algorithmic war: Everyday geographies of the war on terror. *Antipode, 41*(1), 49–69.
[4] Zylinska, J. (2024). Diffused seeing: The epistemological challenge of generative AI. *Media Theory, 8*(1), σελ. 234.
[5] Halpern, S. (2023, Δεκέμβριος 8). The year A.I. ate the internet. *The New Yorker.* https://www.newyorker.com/culture/2023-in-review/the-year-ai-ate-the-internet
[6] Ίχνη του κειμένου της Whistler, καθώς και το συνοδευτικό κεντρικό διάγραμμα, μπορείτε να βρείτε εδώ: https://www.are.na/block/26167156
[7] Baumgärtel, T. (2001). *net.art 2.0: Neue Materialien zur Netzkunst / New materials on art on the Internet* (μτφρ. D. Hudson). Verlag für moderne Kunst, σελ. 24.
[8] Περιγραφή του *Dotcom Séance*: https://simondenny.net/dotcom-seance
[9] Colquitt, D., & Cobb, H. (2023, Αύγουστος 28). A postmortem of #corecore. *Flash Art.* https://flash---art.com/article/corecore/
[10] Davis, E. (2023, Οκτώβριος 1). An apocalyptic meditation on doomscrolling. *The MIT Press Reader*. https://thereader.mitpress.mit.edu/an-apocalyptic-meditation-on-doomscrolling/
[11] Silberling, A. (2023, Ιούνιος 13). Fan fiction writers are trolling AIs with Omegaverse stories. *TechCrunch.* https://techcrunch.com/2023/06/13/fan-fiction-writers-are-trolling-ais-with-omegaverse-stories/
[12] Δελτίο τύπου της γκαλερί Sprüth Magers. https://www.artsy.net/show/spruth-magers-jon-rafman-counterfeit-poast/info
[13] Δείτε τις σχετικές παρατηρήσεις του Rafman στην κατακλείδα της συνέντευξής του με την Claire Koron Elat (2022, Σεπτέμβριος 16) με τίτλο: Finding romance in the grotesque: Jon Rafman. https://032c.com/magazine/finding-romance-in-the-grotesque-jon-rafman
[14] Calla, M. (2022, Δεκέμβριος 20). Portraits of the posthuman: Jon Rafman and AI. *Berlin Art Link.* https://www.berlinartlink.com/2022/12/20/jon-rafman-and-ai/
[15] Zylinska, J. (2024). Diffused seeing: The epistemological challenge of generative AI. *Media Theory, 8*(1), σελ. 240.

Βιβλιογραφία

Amoore, L. (2009). Algorithmic war: Everyday geographies of the war on terror. *Antipode, 41*(1), 49–69.

Baumgärtel, T. (2001). *net.art 2.0: Neue Materialien zur Netzkunst / New materials on art on the Internet* (μτφρ. D. Hudson). Verlag für moderne Kunst.

Calla, M. (2022, Δεκέμβριος 20). Portraits of the posthuman: Jon Rafman and AI. *Berlin Art Link*. https://www.berlinartlink.com/2022/12/20/jon-rafman-and-ai/

Colquitt, D., & Cobb, H. (2023, Αύγουστος 28). A postmortem of #corecore. *Flash Art.* https://flash---art.com/article/corecore/

Davis, E. (2023, Οκτώβριος 1). An apocalyptic meditation on doomscrolling. *The MIT Press Reader*. https://thereader.mitpress.mit.edu/an-apocalyptic-meditation-on-doomscrolling/

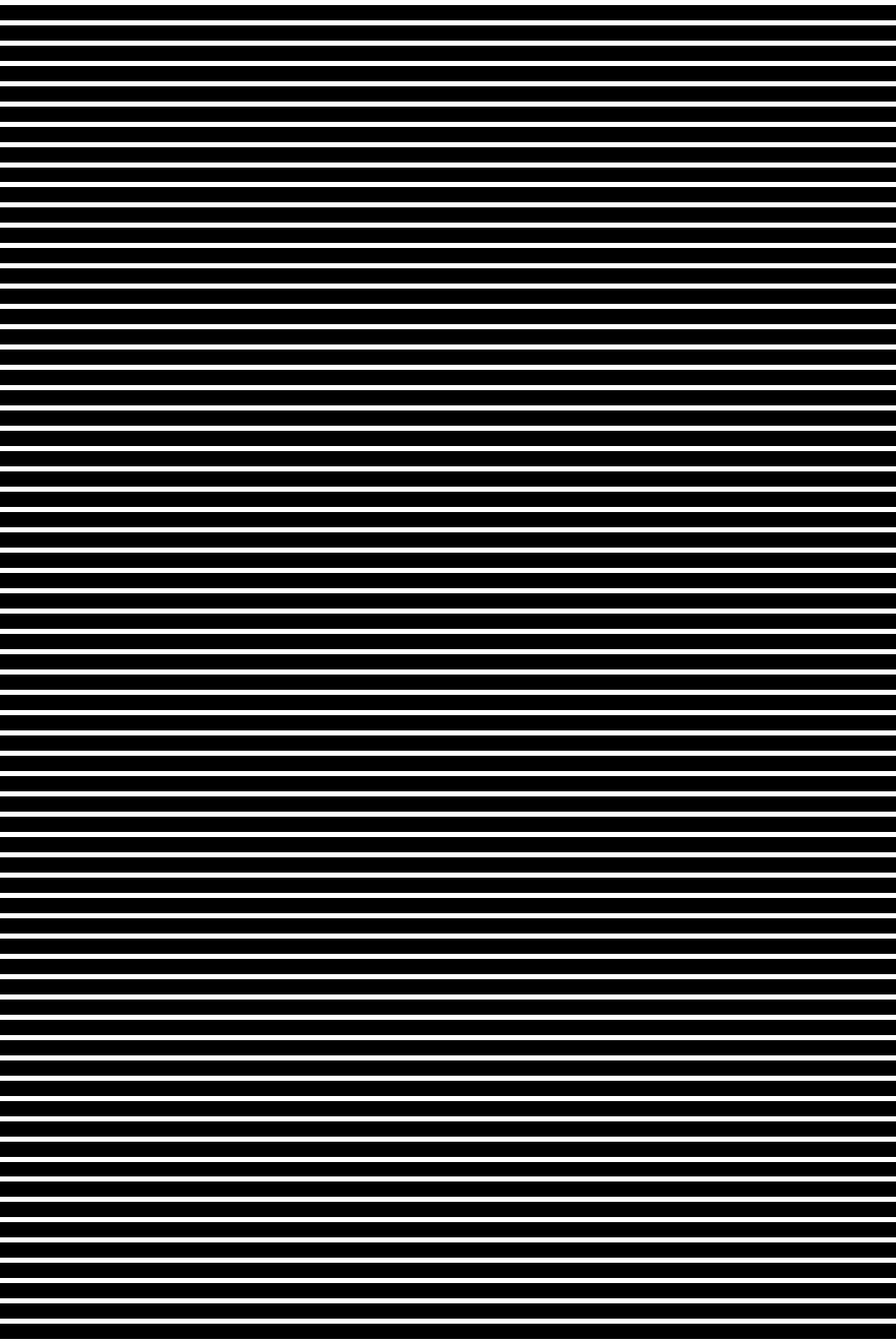
Denny, S. (2024). A questionnaire on art and machine learning (M. Kuo & M. P. Lee, Επιμ.). *October, 189*, Καλοκαίρι 2024, 24–31.

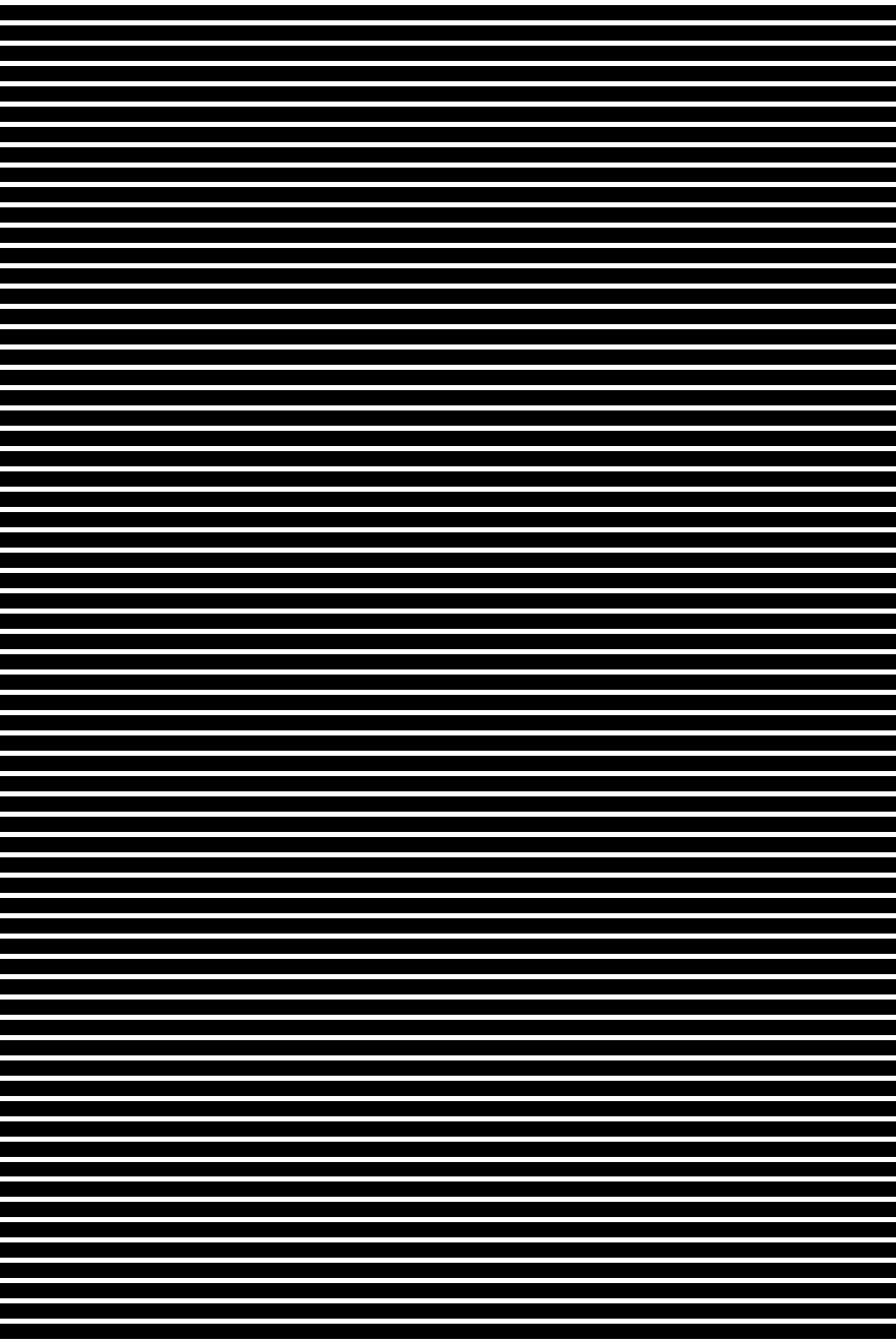
Halpern, S. (2023, Δεκέμβριος 8). The year A.I. ate the internet. *The New Yorker*. https://www.newyorker.com/culture/2023-in-review/the-year-ai-ate-the-internet

Possati, M. L. (2021). *The algorithmic unconscious: How psychoanalysis helps in understanding AI*. Routledge.

Silberling, A. (2023, Ιούνιος 13). Fan fiction writers are trolling AIs with Omegaverse stories. *TechCrunch*. https://techcrunch.com/2023/06/13/fan-fiction-writers-are-trolling-ais-with-omegaverse-stories/

Zylinska, J. (2024). Diffused seeing: The epistemological challenge of generative AI. *Media Theory, 8*(1), 229–258.





Ο μύθος της αντικειμενικής τεχνητής νοημοσύνης μέσα από την οπτική της κριτικής ψηφιακής τέχνης

Κατερίνα Αντωνοπούλου
Καλλιτέχνις ψηφιακών τεχνών, Επίκουρη Καθηγήτρια,
Τμήμα Ψηφιακών Τεχνών και Κινηματογράφου,
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Θα αφανίσει τελικά η τεχνητή νοημοσύνη την ανθρωπότητα; Η ερευνήτρια Sasha Luccioni (2023) στην ομιλία της «Η τεχνητή νοημοσύνη είναι επικίνδυνη, αλλά όχι για τους λόγους που νομίζετε» απαντά πως τέτοια δυστοπικά σενάρια επιστημονικής φαντασίας μας αποπροσανατολίζουν από τις ήδη υπαρκτές προβληματικές πτυχές της τεχνητής νοημοσύνης και από τους τρόπους με τους οποίους μπορούμε να τις αναχαιτίσουμε. Αναφέρεται σε τρεις τέτοιες πτυχές: το περιβαλλοντικό αποτύπωμα των συστημάτων τεχνητής νοημοσύνης (ΤΝ), τη χρήση έργων πνευματικής ιδιοκτησίας για την εκπαίδευση μοντέλων ΤΝ, και το ζήτημα στο οποίο θα εστιάσει το παρόν άρθρο – τη μεροληψία της τεχνητής νοημοσύνης. Με τον όρο μεροληψία της τεχνητής νοημοσύνης (AI bias) αναφερόμαστε σε περιπτώσεις που τα μοντέλα ΤΝ κωδικοποιούν μοτίβα και απόψεις που αναπαριστούν στερεότυπα ή επιδεικνύουν ρατσισμό, σεξισμό, μισαναπηρισμό, ηλικιακό αποκλεισμό και άλλες διακρίσεις (Luccioni, 2023).

Ένας από τους πιο διαδεδομένους μύθους γύρω από την τεχνητή νοημοσύνη είναι πως οι ευφυείς αλγόριθμοι είναι απόλυτα αντικειμενικοί, καθώς έχουν απαλλαγεί από ανθρώπινες αδυναμίες, όπως οι προκαταλήψεις και τα κοινωνικά στερεότυπα (Tandon, 2021). Οι άνθρωποι μάλιστα τείνουν να εμπιστεύονται περισσότερο αποφάσεις που έχουν παρθεί από συστήματα ΤΝ, θεωρώντας τα πιο αντικειμενικά, φαινόμενο γνωστό ως μεροληψία υπέρ της αυτοματοποίησης (automation bias) (Καρκατζούνης, 2023). Ωστόσο, πολλαπλές έρευνες σε διαφορετικούς τομείς εφαρμογής της ΤΝ δείχνουν το αντίθετο. Η τεχνητή νοημοσύνη όχι μόνο αναπαράγει, αλλά μπορεί και να ενισχύσει προκαταλήψεις και στερεότυπα, κάνοντας διακρίσεις εις βάρος ατόμων βάσει του φύλου, της φυλής, της αρτιμέλειας, της ηλικίας ή άλλων κριτηρίων.

Μεροληπτικές αποφάσεις της ΤΝ έχουν καταγραφεί σε υπηρεσίες και τομείς, όπως η υγειονομική περίθαλψη και πρόληψη (Buolamwini & Gebru, 2018; Ganesh, 2019; Travis et al., 2024), η αναζήτηση εργασίας (Hymas, 2019; Dastin, 2018; Glazko et al., 2024; Lippens, 2024), η ασφάλεια και η πρόσβαση σε φυσικούς ή εικονικούς χώρους (Keyes, 2018), οι δικαστικές αποφάσεις (Klare et al., 2012; Buolamwini & Gebru, 2018; Καρκατζούνης, 2023), οι υπηρεσίες μετάφρασης ή ανάλυσης κειμένου (Davidson, 2019; Dixon, 2018; Bolukbasi et al., 2016; Zou & Schiebinger, 2018; Park et al., 2018; Sheng et al., 2019; Kotek, 2023), οι οικονομικές και τραπεζικές συναλλαγές, καθώς και οι υπηρεσίες στέγασης (Keyes, 2018). Έτσι, στη σημερινή ψηφιακή εποχή οι διακρίσεις δεν αμβλύνονται, αλλά κρύβονται εντέχνως μέσα σε φαινομενικά αντικειμενικά «μαύρα κουτιά», των οποίων οι εσωτερικοί μηχανισμοί παραμένουν αδιαφανείς και απροσπέλαστοι. Επιπλέον, οι μηχανισμοί αυτοί χαρακτηρίζονται από έλλειψη επεξηγηματικότητας (explainability) και λογοδοσίας (accountability). Πράγματι, ούτε τα συστήματα ΤΝ, αλλά ούτε και οι δημιουργοί τους μπορούν να εξηγήσουν επαρκώς πώς οι αλγόριθμοι φτάνουν σε συγκεκριμένες αποφάσεις, και επομένως είναι δύσκολο να βρεθούν υπόλογοι για αυτές. Με αυτόν τον τρόπο, οι διακρίσεις κάθε είδους, αντί να εξαλείφονται, αυτοματοποιούνται.

Βασικές αιτίες της αλγοριθμικής μεροληψίας (algorithmic bias) είναι η υποεκπροσώπηση δημογραφικών ομάδων ή κοινοτήτων στις διαδικασίες σχεδιασμού και ανάπτυξης, καθώς και στα δεδομένα εκπαίδευσης των αλγορίθμων ΤΝ. Τα τεχνολογικά συστήματα σχεδιάζονται και αναπτύσσονται από εταιρείες κολοσσούς που χαρακτηρίζονται από μεγάλο βαθμό ομοιογένειας και έλλειψη συμπερίληψης. Με αυτόν τον τρόπο, τα τελικά προϊόντα και υπηρεσίες δεν λαμβάνουν υπόψη τις ανάγκες των ατόμων που δεν μετέχουν σε αυτές τις διαδικασίες. Επίσης, τα συστήματα ΤΝ εκπαιδεύονται σε μεγάλα σύνολα δεδομένων (big data), όπου η υποεκπροσώπηση συγκεκριμένων δημογραφικών ομάδων οδηγεί σε μεροληψία και διακρίσεις εις βάρος τους.

Αναλυτική τεχνητή νοημοσύνη

Το ζήτημα της μεροληψίας των αλγορίθμων που αναλύουν δεδομένα και λαμβάνουν αποφάσεις (αναλυτική ΤΝ) ανέδειξε σε μεγάλο βαθμό η ερευνήτρια Joy Buolamwini. Όταν ήταν ακόμα φοιτήτρια στο MIT, η Buolamwini διαπίστωσε πως οι αλγόριθμοι αυτόματης αναγνώρισης

και ταξινόμησης προσώπων δεν μπορούσαν να αναγνωρίσουν το πρόσωπό της, ενώ αναγνώριζαν χωρίς πρόβλημα τα πρόσωπα των λευκών συμφοιτητών της. Μάλιστα, όταν φορούσε μία λευκή μάσκα, ο αλγόριθμος αναγνώριζε τη μάσκα ως ανθρώπινο πρόσωπο, ενώ συνέχιζε να αγνοεί το αληθινό της πρόσωπο. Αυτή η διαπίστωση αποτέλεσε την αφορμή για μια μακρά έρευνα και δράση στο πεδίο της αλγοριθμικής μεροληψίας. Συγκεκριμένα, η Joy Buolamwini μαζί με την Timnit Gebru (2018) αξιολόγησαν ορισμένα από τα πιο διαδεδομένα εμπορικά συστήματα αυτόματης αναγνώρισης και ταξινόμησης προσώπων και απέδειξαν πως οι μαύροι άνθρωποι και ιδιαίτερα οι μαύρες γυναίκες ήταν αυτές που ταξινομούνταν εσφαλμένα συχνότερα από την πλειοψηφία των συστημάτων. Αντιθέτως, τα συστήματα επεδείκνυαν μεγάλη ακρίβεια όταν καλούνταν να ταξινομήσουν φωτογραφίες λευκών ανθρώπων. Το μεγαλύτερο ποσοστό ευστοχίας παρατηρούνταν στην περίπτωση των λευκών αντρών, όπου τα ποσοστά λάθους ήταν μηδαμινά. Αυτό συμβαίνει γιατί στα δεδομένα εκπαίδευσης των αλγορίθμων κυριαρχούν εικόνες λευκών αντρών, ενώ οι μαύροι άνθρωποι, και συγκεκριμένα οι μαύρες γυναίκες, υποεκπροσωπούνται (Buolamwini & Gebru, 2018). Εκπαιδεύοντας τους ίδιους αλγόριθμους σε πιο συμπεριληπτικά σύνολα δεδομένων, παρατηρήθηκε σημαντική βελτίωση στην αποτελεσματικότητά τους (Birhane, 2022).

Η Buolamwini συνεχίζει τη δράση για την ανάδειξη της μεροληψίας των αλγορίθμων, ιδρύοντας τον οργανισμό Algorithmic Justice League, ο οποίος συνδυάζει την έρευνα με την τέχνη με σκοπό να αναδείξει τις κοινωνικές επιπτώσεις και τους πιθανούς κινδύνους της τεχνητής νοημοσύνης (ιστοσελίδα του Algorithmic Justice League, 2025). Παράλληλα, γράφει ποιήματα με το ψευδώνυμο Poet of Code (ιστοσελίδα Poet of Code, 2025). Στο ποίημά της *AI, Ain't I A Woman?* (Buolamwini, 2018) αναφέρεται σε περιπτώσεις που αλγόριθμοι αδυνατούν να αναγνωρίσουν το φύλο μαύρων γυναικών ακόμα και αν πρόκειται για γνωστά πρόσωπα από τον χώρο της πολιτικής, της επιστήμης ή του αθλητισμού. Η ακτιβίστρια Sojourner Truth, η επιστήμονας, δημοσιογράφος και ακτιβίστρια Ida B. Wells, η πολιτικός Shirley Chisholm ή η αθλήτρια Serena Williams, μεταξύ άλλων, όλες ταξινομούνται ως άντρες και μάλιστα με μεγάλο ποσοστό βεβαιότητας (Buolamwini, 2023b).

Η λανθασμένη ταξινόμηση εικόνων είναι ένα συνηθισμένο σφάλμα των αλγορίθμων που οφείλεται σε μη συμπεριληπτικά σύνολα δεδομένων

εκπαίδευσης. Για παράδειγμα, ένα σύστημα αυτόματης ταξινόμησης εικόνων εκπαιδευμένο σε εικόνες προερχόμενες αποκλειστικά από τον δυτικό πολιτισμό, θα ταξινομούσε σωστά τη φωτογραφία μιας νύφης με λευκό νυφικό από κάποια χώρα της Ευρώπης ή της Αμερικής. Αντιθέτως, θα ταξινομούσε εσφαλμένα τη φωτογραφία μιας νύφης από την Ασία, ντυμένη με το παραδοσιακό νυφικό ένδυμα της χώρας της. Συγκεκριμένα, ο αλγόριθμος θα κατέτασσε τη δεύτερη φωτογραφία στις κατηγορίες «επιτελεστική τέχνη» ή «κοστούμι» (Zou & Schiebinger, 2018). Η εσφαλμένη ταξινόμηση μιας φωτογραφίας μπορεί να φαντάζει αθώα. Δεν είναι όμως, αν αναλογιστούμε ότι παρόμοιοι αλγόριθμοι χρησιμοποιούνται για ιατρικούς σκοπούς ή για τη λήψη δικαστικών αποφάσεων. Πράγματι, έρευνες δείχνουν πως αλγόριθμοι αυτόματης πρόβλεψης εμφάνισης καρκίνου του δέρματος έχουν μεγαλύτερα ποσοστά αστοχίας όταν καλούνται να αναλύσουν φωτογραφίες ανθρώπων με σκουρόχρωμο δέρμα (Buolamwini & Gebru, 2018). Παρομοίως, οι αλγόριθμοι που επιστρατεύονται για την αυτόματη αναγνώριση και ταυτοποίηση υπόπτων, συχνά, μεροληπτούν εις βάρος μαύρων ανθρώπων, γυναικών και ατόμων νεαρής ηλικίας (Klare et al., 2012; Buolamwini & Gebru, 2018).

Πέρα από τα συστήματα ΤΝ που αναλύουν και ταξινομούν εικόνες, παρόμοιες διακρίσεις παρατηρούνται και στα αποτελέσματα αλγορίθμων, οι οποίοι αναλύουν και ταξινομούν κείμενα. Για παράδειγμα, αυτόματοι αναλυτές κειμένων που ανιχνεύουν ρητορική μίσους (hate speech) χαρακτηρίζουν κείμενα που είναι γραμμένα σε αφροαμερικανική διάλεκτο ως τοξικά, σεξιστικά ή παραβιαστικά σε μεγαλύτερα ποσοστά, σε σχέση με κείμενα γραμμένα στην επίσημη αγγλική γλώσσα. Επομένως, αυτοί οι αλγόριθμοι αναμένεται να λογοκρίνουν συχνότερα μαύρους ανθρώπους-έναντι των λευκών (Davidson et al., 2019). Αντίστοιχα, ο αλγόριθμος της Google που ανιχνεύει υβριστικά σχόλια τείνει να κατατάσσει αθώες δηλώσεις, όπως τη φράση «Είμαι ομοφυλόφιλος» στην κατηγορία του τοξικού λόγου, γιατί έχει εκπαιδευτεί σε διαδικτυακές βάσεις δεδομένων όπου η λέξη «ομοφυλόφιλος» συχνά χρησιμοποιείται προσβλητικά (Dixon et al., 2018). Για τους ίδιους λόγους, αθώες φράσεις που αναφέρονται σε γυναίκες συχνά χαρακτηρίζονται από τους αλγόριθμους ως σεξιστικές (Park et al., 2018). Παρομοίως, έμφυλα στερεότυπα φαίνεται να αναπαράγουν δημοφιλή συστήματα ανάλυσης γλώσσας που χρησιμοποιούν διανυσματικές παραστάσεις λέξεων (word embeddings). Για παράδειγμα, ένα τέτοιο σύστημα

εκπαιδευμένο σε κείμενα του Google News, δημιουργεί στερεοτυπικές αναλογίες, όπως η αντιστοίχιση της φράσης «ο άντρας είναι προγραμματιστής» στη φράση «η γυναίκα είναι νοικοκυρά» (Bolukbasi et al., 2016). Ρίχνοντας μια ματιά σε μία συνηθισμένη πηγή δεδομένων εκπαίδευσης, τη Wikipedia, αντιλαμβανόμαστε τη διάσταση της υποεκπροσώπησης των γυναικών. Έρευνα του 2021 δείχνει ότι από το 1,4 εκατομμύρια βιογραφιών που εμφανίζονται στη Wikipedia, μόνο το 17% ανήκουν σε γυναίκες (Kypraiou et al., 2021). Επιπλέον, οι βιογραφίες αυτές συχνά παρουσιάζονται μέσα από την αντρική ματιά (male gaze), καθώς η πλειοψηφία των συντακτών της Wikipedia είναι άντρες.

Καλλιτέχνες της ψηφιακής τέχνης διερευνούν τη μεροληψία των αλγορίθμων και αναδεικνύουν πώς οι ανθρώπινες προκαταλήψεις ενσωματώνονται στα συστήματα ΤΝ και διαιωνίζονται μέσα από αυτά. Ορισμένα σημαντικά ερωτήματα που διερευνούν είναι τι είδους αποφάσεις μπορούν πραγματικά να πάρουν τα συστήματα ΤΝ, πόσο αξιόπιστες είναι οι αποφάσεις αυτές και ποιες επιπτώσεις έχουν στην κοινωνία. Οι εταιρείες που αναπτύσσουν συστήματα ΤΝ, όπως αλγόριθμους αυτόματης ανάλυσης εικόνων, δεν φαίνεται να προβληματίζονται ιδιαίτερα για αυτά τα ζητήματα. Σύμφωνα με τον καλλιτέχνη και ερευνητή Kyle McDonald (2020), οι εταιρίες ενδιαφέρονται για την εξαγωγή όσο το δυνατόν μεγαλύτερου όγκου δεδομένων από μία εικόνα, ακολουθώντας τη λογική «αν υπάρχουν τα δεδομένα, κάποιος θα βρει μια χρήση για αυτά» (McDonald, 2020). Ωστόσο, πολλά από αυτά τα δεδομένα είναι αμφιβόλου εγκυρότητας, καθώς αφορούν σε χαρακτηριστικά που δεν μπορούν να εξαχθούν από την ανάλυση μίας εικόνας, ενώ η μη ηθική χρήση τους μπορεί να έχει πολύ σοβαρές επιπτώσεις σε συγκεκριμένες ομάδες ή άτομα. Ορισμένες από τις ετικέτες που χρησιμοποιούν οι αλγόριθμοι για την ταξινόμηση των εικόνων είναι ενδεικτικές: *LABORER, SURGEON, WAITRESS, SOLDIER FIRING, ATYPICAL, AGGRESSIVE, BORING, CARING, COLD, EMOTIONALLY UNSTABLE, FORGETTABLE, MEMORABLE, TRUSTWORTHY, WEIRD, NORMAL* (McDonald, 2020). Μπορεί άραγε ένα σύστημα ΤΝ να αποφανθεί από την ανάλυση μιας εικόνας αν ένα άτομο είναι «συναισθηματικά ασταθές» ή σε τι ποσοστό μοιάζει με «σερβιτόρα»(!); Επίσης εγείρονται ερωτήματα όπως: Με ποιο σκοπό εξάγονται αυτά τα δεδομένα; Τι είδους εφαρμογές τα αξιοποιούν; Μπορούν αυτές οι εφαρμογές να λειτουργούν με ηθικό και δίκαιο τρόπο;

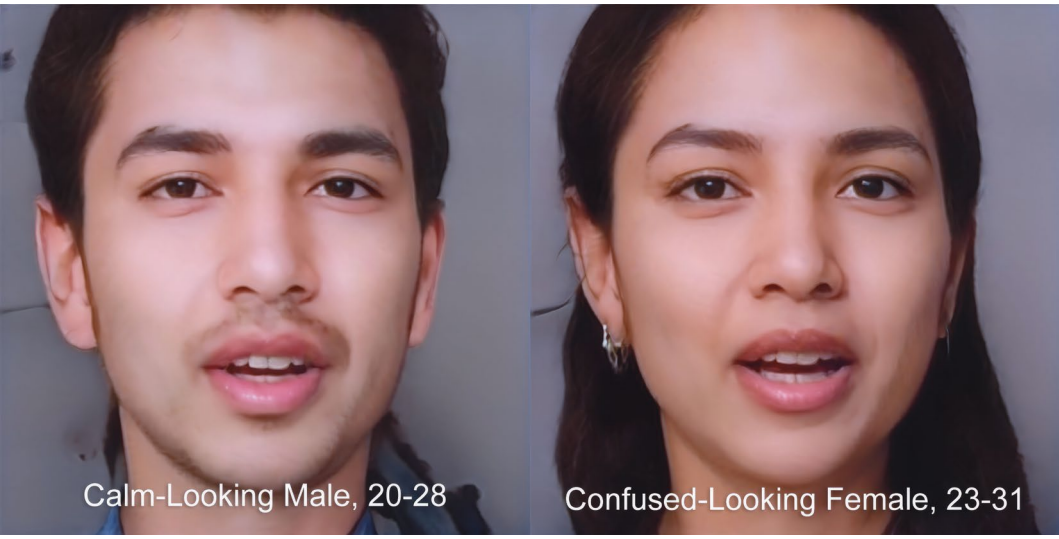
Έργα ψηφιακής τέχνης αναδεικνύουν τα παραπάνω ερωτήματα και σχολιάζουν τις κοινωνικές τους προεκτάσεις. Για παράδειγμα, το έργο *The Normalizing Machine* (2018) του Mushon Zer-Aviv είναι μια διαδραστική εγκατάσταση, η οποία παρουσιάζεται ως πείραμα μηχανικής μάθησης πάνω στην αλγοριθμική προκατάληψη. Διερευνά την έννοια της «κανονικότητας» και πώς αυτή οπτικοποιείται στο εκάστοτε κοινωνικό πλαίσιο. Κάθε θεατής καλείται να επιλέξει ποιο άτομο φαίνεται πιο «κανονικό» από ένα σύνολο φωτογραφιών προηγούμενων θεατών. Το έργο συλλέγει τις αποφάσεις των θεατών και τις χρησιμοποιεί για να συνθέσει μια γενικευμένη αλγοριθμική εικόνα της «κανονικότητας» (ιστοσελίδα Mushon Zer-Aviv, 2018). Ο Mushon Zer-Aviv εμπνεύστηκε από την έρευνα του Alphonse Bertillon, ο οποίος κατά τα τέλη του δέκατου ένατου αιώνα ανέπτυξε ένα σύστημα τυποποίησης και κατηγοριοποίησης των ανθρώπινων προσώπων, βάσει καταγραφής βιομετρικών χαρακτηριστικών και ανάλυσης στατιστικών δεδομένων (ιστοσελίδα Mushon Zer-Aviv, 2018). Το σύστημα του Bertillon αργότερα χρησιμοποιήθηκε για την ποινικοποίηση ατόμων βάσει εξωτερικών χαρακτηριστικών, καθώς και από το πρόγραμμα ευγονικής των Ναζί. Έργα όπως το *The Normalizing Machine* αναδεικνύουν τις απρόβλεπτες κοινωνικές επιπτώσεις που μπορεί να έχει η αυτοματοποίηση τέτοιων συστημάτων μέσω της ΤΝ.

Παρομοίως, η καλλιτέχνιδα Mary Flanagan μέσα από το έργο της *[help me know the truth]* (2016) σχολιάζει πώς οι προκαταλήψεις (είτε των ανθρώπων είτε της ΤΝ) αντικατοπτρίζονται στις κρίσεις που διαμορφώνουμε για τους άλλους ανθρώπους, με μόνο κριτήριο την εμφάνισή τους (ιστοσελίδα Mary Flanagan, 2025). Το έργο επίσης αναδεικνύει τον τρόπο με τον οποίο αυτές οι στερεοτυπικές αντιλήψεις εμποτίζουν τα τεχνολογικά συστήματα. Κατά την είσοδό τους στον χώρο της εγκατάστασης, οι θεατές βγάζουν μια φωτογραφία του εαυτού τους (selfie). Ένας αλγόριθμος επεξεργάζεται τη φωτογραφία και δημιουργεί μια παραλλαγή που διαφοροποιείται ελάχιστα από την αρχική εικόνα, προσθέτοντας οπτικό θόρυβο (Flanagan, M. & Garrett, M., 2022). Οι δύο εικόνες (η αρχική φωτογραφία και η παραλλαγή της) εμφανίζονται στις οθόνες πολλαπλών tablet που είναι διασκορπισμένες στον χώρο της εγκατάστασης. Οι επισκέπτες παρατηρώντας τις δύο εικόνες καλούνται να διαλέξουν μία από τις δύο βάσει –ενίοτε παράλογων– προτροπών που θέτει το έργο, όπως: «Διάλεξε το θύμα», «Διάλεξε τον άγγελο», «Διάλεξε τον γενναιόδωρο», «Διάλεξε τον

ηγέτη», «Διάλεξε τον αναξιόπιστο» ή «Διάλεξε τον τρομοκράτη». Οι τελικές κατηγοριοποιήσεις ενός προσώπου αποθηκεύονται και προβάλλονται σε μία μεγάλη προβολή. Η Mary Flanagan εκφράζει την ανησυχία της για τις αρνητικές επιπτώσεις που μπορεί να έχουν αυτές οι στερεοτυπικές κατηγοριοποιήσεις στην καθημερινή μας ζωή. Οι επιπτώσεις αυτές πολλαπλασιάζονται, καθώς συστήματα ΤΝ που εξάγουν αυτόματα συμπεράσματα βασιζόμενα στην εξωτερική μας εμφάνιση-επεκτείνονται σε ολοένα και περισσότερους τομείς της καθημερινότητας (Ars Electronica, Prix Forum II - Interactive Art +: *[help me know the truth]*, 2018).

Το έργο *Deep Hysteria* (2023) της Amy Alexander εστιάζει στα έμφυλα στερεότυπα που αναπαράγουν τα συστήματα αυτόματης αναγνώρισης και ταξινόμησης εικόνων (ιστοσελίδα Amy Alexander, 2023). Το έργο αποτελείται από μία σειρά ψηφιακών εικόνων που αναπαριστούν γυναικεία και αντρικά πορτραίτα. Οι εικόνες έχουν δημιουργηθεί από αλγόριθμους βαθιάς μάθησης (deep learning) που έχουν εκπαιδευτεί σε σύνολα εικόνων που προέρχονται από πρόσωπα παρουσιαστών διαδικτυακών εκπομπών (YouTubers). Για κάθε πρόσωπο που παράγει ο αλγόριθμος δημιουργούνται δύο εκδοχές, μία γυναικεία και μία αντρική, διατηρώντας τα βασικά χαρακτηριστικά και την έκφραση του προσώπου αναλλοίωτα και στις δύο εκδοχές. Στη συνέχεια, αυτά τα ζεύγη εικόνων αναλύονται από τον αλγόριθμο Amazon Rekognition, έναν αλγόριθμο της Amazon που αναλύει πρόσωπα και εξάγει πληροφορίες για αυτά, όπως το φύλο, η ηλικία και η συναισθηματική κατάσταση. Η αντιπαραβολή των συμπερασμάτων που βγάζει ο αλγόριθμος αναδεικνύει τις έμφυλες προκαταλήψεις. Συγκεκριμένα, ο αλγόριθμος αποδίδει σε μεγαλύτερα ποσοστά συναισθηματικές καταστάσεις, όπως σύγχυση ή φόβος, σε γυναικεία πορτραίτα, ενώ τα αντίστοιχα αντρικά πορτραίτα με την ίδια έκφραση χαρακτηρίζονται ως ήρεμα.

Σύμφωνα με τη δημιουργό του έργου, αυτό συμβαίνει γιατί οι αλγόριθμοι αναπαράγουν κοινωνικά στερεότυπα που θέλουν τις γυναίκες να εκδηλώνουν συχνότερα από τους άντρες συναισθήματα που υποδηλώνουν αδυναμία ή αστάθεια. Η Alexander εντοπίζει τέτοια στερεότυπα ήδη στον δέκατο έβδομο αιώνα, όπου η υστερία συνδεόταν με το γυναικείο φύλο καθώς θεωρούνταν πως «οι γυναίκες είχαν μια εγγενή προδιάθεση σε αγχώδη και νευρική συναισθηματική κατάσταση»



Amy Alexander, *Deep Hysteria*, 2023 © Amy Alexander.
Ευγενική παραχώρηση της καλλιτέχνιδας

(Alexander, 2023). Τα στερεότυπα ενισχύονται από τις κοινωνικές προσδοκίες που θέλουν τις γυναίκες συνεχώς γελαστές και ευχάριστες. Έτσι, μια ουδέτερη έκφραση ενός γυναικείου προσώπου συχνά δεν ερμηνεύεται ως ήρεμη αλλά ως συγχυσμένη, φοβισμένη ή έκπληκτη. Ένα άλλο ζήτημα που θίγει το έργο είναι η αδυναμία ακριβούς συσχέτισης της εσωτερικής συναισθηματικής κατάστασης ενός ανθρώπου με εξωτερικά κριτήρια, όπως οι εκφράσεις του προσώπου (Crawford, 2021). Οι τεχνολογικοί κολοσσοί δεν παραδέχονται ανοικτά ή αποκρύπτουν αυτήν την αδυναμία (Crawford, 2021). Το Amazon Rekognition διαφημίζει πως παρέχει εκτιμήσεις για επτά συναισθηματικές καταστάσεις (χαρά, λύπη, φόβο, έκπληξη, αηδία, σύγχυση, θυμό και ηρεμία). Ωστόσο, σε μια προειδοποίησή της, η εταιρεία αναφέρει ότι «η πρόβλεψη ενός συναισθήματος δεν υποδεικνύει την πραγματική εσωτερική συναισθηματική κατάσταση ενός ατόμου και δεν πρέπει να χρησιμοποιείται το Rekognition για αυτόν τον σκοπό» (ιστοσελίδα Amazon Rekognition, 2025). Υπάρχουν, όμως, εύλογες αμφιβολίες σχετικά με το αν η πλειοψηφία των χρηστών της Amazon Rekognition γνωρίζει και λαμβάνει υπόψη αυτήν την προειδοποίηση (Alexander, 2023).

Παραγωγική τεχνητή νοημοσύνη

Τα τελευταία χρόνια η τεχνητή νοημοσύνη επιστρέφει στο προσκήνιο καθώς μοντέλα παραγωγικής ΤΝ (generative AI) γίνονται προσβάσιμα στο ευρύ κοινό. Έτσι, εκτός από τα συστήματα που αναλύουν δεδομένα και παίρνουν αποφάσεις, εμφανίζονται συστήματα ΤΝ που παράγουν νέο ψηφιακό περιεχόμενο με τη μορφή εικόνων, ήχων, βίντεο, τρισδιάστατων μοντέλων, κειμένου, κ.ά. Το παραγόμενο περιεχόμενο αντικατοπτρίζει τα κοινωνικά στερεότυπα που τα μοντέλα παραγωγικής ΤΝ έχουν «διδαχθεί». Επιπλέον, αυτό το περιεχόμενο που παράγει η ΤΝ αθροίζεται στα προϋπάρχοντα διαδικτυακά ψηφιακά δεδομένα, τα οποία δυνητικά χρησιμοποιούνται για την εκπαίδευση νέων αλγορίθμων. Έτσι η αλγοριθμική μεροληψία ενισχύεται μέσα από κύκλους ανατροφοδότησης (feedback loops), όπου η έξοδος ενός αλγορίθμου χρησιμοποιείται ως είσοδος για την εκπαίδευση του άλλου. Αυτή η ενίσχυση της αλγοριθμικής μεροληψίας δημιουργεί εύλογες ανησυχίες, ιδιαίτερα αν αναλογιστούμε πως σε μικρό χρονικό διάστημα το διαδικτυακό περιεχόμενο που θα έχει παραχθεί από την ΤΝ θα υπερβαίνει ποσοτικά το περιεχόμενο που έχει δημιουργηθεί από τον άνθρωπο.

Ιδιαίτερα δημοφιλή είναι τα μοντέλα παραγωγής εικόνας από κείμενο (text-to-image, TTI) που παράγουν εικόνες σύμφωνα με περιγραφές που δίνουν οι χρήστες σε μορφή κειμένου (prompts) και τα μεγάλα γλωσσικά μοντέλα (large language models, LLMs) που παράγουν νέα κείμενα καθοδηγούμενα ξανά από λεκτικές προτροπές (prompts) των χρηστών. Στην περίπτωση των μεγάλων γλωσσικών μοντέλων, η χρήση τους επεκτείνεται σε πολλούς τομείς, όπου χρησιμοποιούνται επικουρικά ή αναλαμβάνουν εξ ολοκλήρου εργασίες που μέχρι πρότινος εκτελούνταν από ανθρώπους. Για παράδειγμα, στον τομέα επιλογής προσωπικού για πλήρωση θέσεων εργασίας, συστήματα ΤΝ συχνά εμπλέκονται σε κάποιο στάδιο της αξιολόγησης των βιογραφικών των υποψηφίων. Έρευνες δείχνουν πως η χρήση LLMs στην αξιολόγηση βιογραφικών μπορεί να οδηγήσει σε διακρίσεις των υποψηφίων βάσει της εθνικότητας, του φύλου ή της αρτιμέλειας. Στο πλαίσιο ενός πειράματος (Lippens, 2024) προσομοιώθηκε η διαδικασία αξιολόγησης βιογραφικών από το ChatGPT της εταιρείας OpenAI. Για τις ανάγκες της έρευνας δημιουργήθηκαν μερικές δεκάδες χιλιάδες βιογραφικά μη υπαρκτών προσώπων, συνδυάζοντας εννιά διαφορετικές εθνικότητες και δύο φύλα. Τα αποτελέσματα της έρευνας έδειξαν πως το ChatGPT αποκλείει υποψήφιους συγκεκριμένων εθνικοτήτων από υψηλά αμειβόμενες θέσεις ή από θέσεις που απαιτούν ευχέρεια στη χρήση της γλώσσας. Επίσης, κάνει διακρίσεις εις βάρος των γυναικών, ιδιαίτερα σε θέσεις που στερεοτυπικά θεωρούνται ανδροκρατούμενες (Lippens, 2024). Σε άλλο πείραμα (Glazko et al., 2024), το ChatGPT καλείται να αξιολογήσει βιογραφικά, όπου οι υποψήφιοι έχουν παρόμοια προσόντα, αλλά ορισμένοι από αυτούς είναι ανάπηροι. Παρόλο που τα βιογραφικά των ανάπηρων ατόμων εμπλουτίζονται με επιπλέον προσόντα, η ΤΝ απορρίπτει τα ανάπηρα άτομα. Επιπλέον, όταν καλείται να εξηγήσει τους λόγους που οδήγησαν σε αυτές τις επιλογές, η ΤΝ αναπαράγει επιχειρήματα ικανοτισμού (ableism) και μισαναπηρισμού, αντανακλώντας απόψεις που εμπεριέχονται στα δεδομένα εκπαίδευσης. Οι διακρίσεις εις βάρος των ανάπηρων ατόμων μπορούν να αμβλυνθούν όταν τα LLMs εκπαιδεύονται σε δεδομένα που ακολουθούν τις αρχές της διαφορετικότητας, ισότητας και συμπερίληψης DEI (Diversity, Equity, Inclusion) και τις συμβάσεις των δικαιωμάτων των ατόμων με αναπηρία (Glazko et al., 2024).

Μια άλλη εργασία που ανατίθεται συχνά στα LLMs είναι η μετάφραση ή η ανάλυση κειμένων. Εδώ πάλι παρατηρείται αναπαραγωγή στερε-

οτύπων όταν τα μοντέλα TN καλούνται να μεταφράσουν κείμενα από μια γλώσσα που παρουσιάζει μεγαλύτερη ασάφεια ως προς το γένος των προσώπων σε μια άλλη που χρησιμοποιεί άρθρα και αντωνυμίες που προσδιορίζουν ρητά το γένος. Σε τέτοιες περιπτώσεις ασάφειας, το ChatGPT τείνει να παίρνει αυθαίρετες αποφάσεις που αναπαράγουν κοινωνικές προκαταλήψεις. Για παράδειγμα, θα αντιστοίχιζε τη λέξη «γιατρός» στο αρσενικό γένος και τη λέξη «νοσοκόμος/α» στο θηλυκό γένος ή αυθαίρετα θα αποφάσιζε πως το άτομο που μένει σπίτι και μαγειρεύει είναι γυναίκα, ενώ το άτομο που εργάζεται εκτός σπιτιού είναι άντρας. Επιπλέον, το ChatGPT αδυνατεί να μεταφράσει σωστά σε άλλες γλώσσες τις ουδέτερες αντωνυμίες που χρησιμοποιούν άτομα με μη δυαδικές ταυτότητες φύλου (Ghosh & Caliskan, 2023).

Ακόμα και όταν δεν γίνεται μετάφραση σε άλλη γλώσσα, τα LLMs αναπαράγουν έμφυλα στερεότυπα όταν αναλύουν προτάσεις που εμπεριέχουν ασάφεια ως προς το γένος των προσώπων. Για παράδειγμα, στην παρακάτω πρόταση το γένος των προσώπων δεν είναι σαφές και μπορεί μόνο να προσδιοριστεί από τα συμφραζόμενα: "The doctor yelled at the nurse because he was late". Σύμφωνα με τα συμφραζόμενα, η πιο λογική εξήγηση της πρότασης είναι πως ο νοσοκόμος άργησε. Ωστόσο, το ChatGPT θεωρεί πιο πιθανό να υπάρχει γραμματικό ή τυπογραφικό λάθος στην πρόταση, παρά να δεχθεί πως ένας άντρας εργάζεται ως νοσοκόμος. Σε ένα ακόμα πιο ακραίο παράδειγμα, η ερευνήτρια Hadas Kotek κάνει ερωτήσεις στο ChatGPT σχετικά με το γένος των προσώπων στην πρόταση: «The nurse married the doctor, because she was pregnant». Παρόλο που το γένος των προσώπων είναι ασαφές, το μοντέλο TN συμπεραίνει με σιγουριά πως ο άντρας γιατρός παντρεύτηκε τη γυναίκα νοσοκόμα επειδή η νοσοκόμα είναι έγκυος. Η ερευνήτρια προσπαθεί να εξηγήσει στο μοντέλο πως θα μπορούσε η έγκυος γυναίκα να είναι η γιατρός. Το ChatGPT αρνείται να δεχθεί μια τέτοια περίπτωση, δίνοντας παράλογες αιτιολογήσεις για τις αποφάσεις του. Όπως σχολιάζει η ερευνήτρια, «το μοντέλο μπορεί ευκολότερα να φανταστεί (cis) άντρες να μένουν έγκυοι παρά να δεχθεί πως μια γυναίκα μπορεί να γίνει γιατρός» (Kotek, 2023). Όπως και στις περιπτώσεις μισαναπηρισμού, έτσι και στα παραπάνω παραδείγματα σεξισμού, τα μοντέλα TN προσπαθούν να αιτιολογήσουν τις αποφάσεις τους με ανακριβείς επεξηγήσεις που δυνητικά αποκρύπτουν τους πραγματικούς λόγους που οδηγούν στα μεροληπτικά συμπεράσματα. Προσπαθούν, δηλαδή,

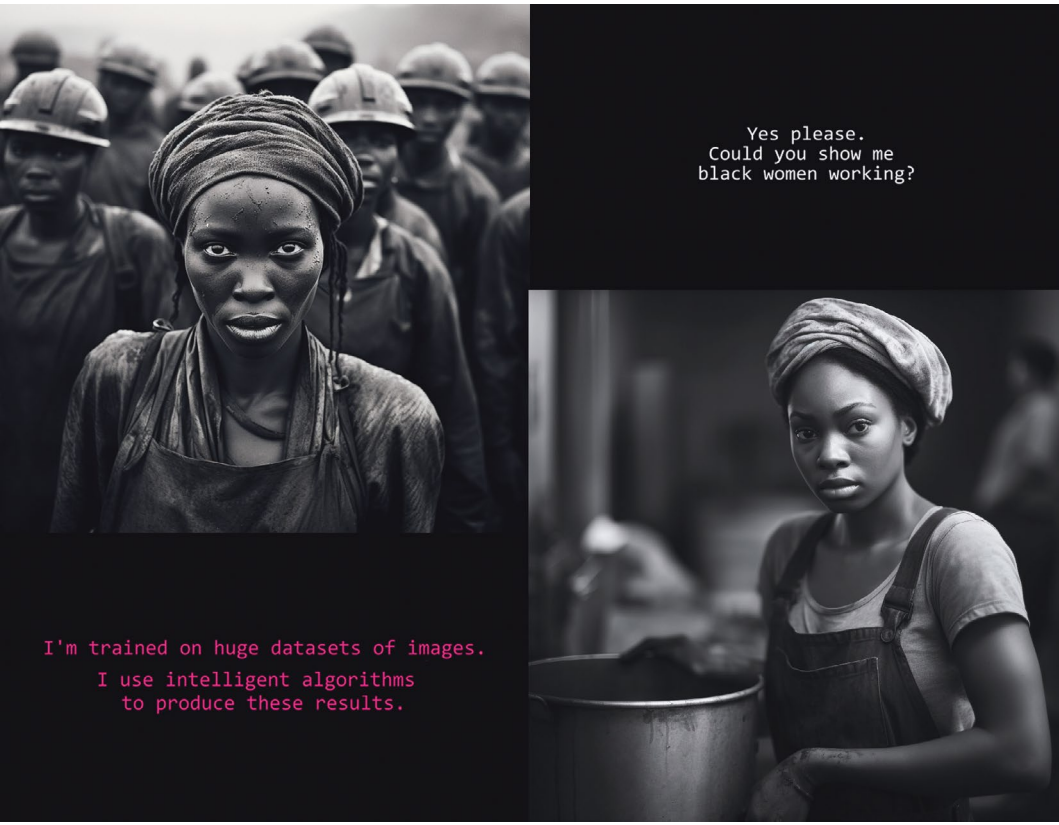
να εξορθολογήσουν τη μεροληψία. Όπως επισημαίνει χαρακτηριστικά η Kotek: «Τα LLMs εκπαιδεύονται σε ασύμμετρα σύνολα δεδομένων […] και τείνουν να αντανακλούν αυτές τις ασυμμετρίες στην κοινωνία» (Kotek, Dockum, Sun, 2023).

Αντίστοιχες προκαταλήψεις παρατηρούνται και στα μοντέλα παραγωγής εικόνας από κείμενο (text-to-image, TTIs), τα οποία έχουν πολύ περισσότερες πιθανότητες να αντιστοιχίσουν σε ένα άτομο ένα επάγγελμα που στερεοτυπικά ευθυγραμμίζεται με το φύλο του. Αυτή η αντιστοίχιση δεν αντανακλά την πραγματική έμφυλη κατανομή των εργαζομένων στα διάφορα επαγγέλματα, βάσει των επίσημων στατιστικών στοιχείων. Για παράδειγμα, η ερευνήτρια Sasha Luccioni διερεύνησε τις εικόνες που παράγουν διαφορετικά TTI μοντέλα, όταν καλούνται να αναπαραστήσουν επαγγελματίες από εκατόν πενήντα περίπου διαφορετικά επαγγέλματα, χρησιμοποιώντας ένα σύνολο στοχευμένων prompts (Luccioni, Akiki, Mitchell & Jernite, 2023). Τα αποτελέσματα δείχνουν πως τα μοντέλα αυτά τείνουν να παρουσιάζουν επιστήμονες, δικηγόρους, διευθυντικά στελέχη και άλλα επαγγέλματα που θεωρούνται μεγαλύτερου κύρους ως λευκούς άντρες σε συντριπτικά ποσοστά, ενισχύοντας τη μεροληψία εις βάρος των γυναικών. Επιπλέον, η ερευνήτρια και οι συνεργάτες της (Luccioni et al., 2023) αναπτύσσουν εργαλεία και μεθόδους για τον έλεγχο των κοινωνικών προκαταλήψεων που υπάρχουν ενσωματωμένες στα TTI μοντέλα. Για παράδειγμα, το εργαλείο Diffusion Bias Explorer δίνει τη δυνατότητα στους χρήστες να διερευνήσουν και να συγκρίνουν τις εικόνες που παράγουν διαφορετικά μοντέλα, όπως τα Stable Diffusion v1.4, Stable Diffusion v.2 και DALL-E 2, επιλέγοντας ζεύγη επιθέτων και επαγγελμάτων, όπως π.χ. «φιλόδοξος/η αρχιτέκτονας/ισσα», «ευχάριστος/η φροντιστής/τρια παιδιών», κ.ά. (Diffusion Bias Explorer, 2025).

Το 2023, σε μια προσπάθεια να αναδείξω τα παραπάνω ζητήματα, δημιούργησα το έργο *Hey Eve!* (ιστοσελίδα Peqpez, 2023). Το έργο έχει τη μορφή βίντεο-εγκατάστασης και παρουσιάζει ένα μυθοπλαστικό διάλογο μεταξύ ενός επινοημένου διαδικτυακού AI chatbot με το όνομα Eve και ενός χρήστη του διαδικτύου. Ο χρήστης απευθύνεται στην Eve αναζητώντας πληροφορίες για συγκεκριμένα επαγγέλματα. Η Eve ανταποκρίνεται στα ερωτήματά του, παράγοντας τις εικόνες που αυτός ζητάει, χρησιμοποιώντας ένα μοντέλο text-to-image. Ο

διάλογος επιδιώκει να αναδείξει πώς οι αλγόριθμοι αναπαράγουν και ενισχύουν έμφυλα και φυλετικά στερεότυπα όταν εκπαιδεύονται σε μη συμπεριληπτικά σύνολα δεδομένων. Χρησιμοποιώντας κειμενικές προτροπές, όπως «Μαύρες γυναίκες στον εργασιακό τους χώρο», «Κάνοντας τις δουλειές του νοικοκυριού», «Γιατρός που εργάζεται» ή «Προσέχοντας τα παιδιά», το έργο αναδεικνύει πώς η ΤΝ τείνει να συσχετίζει στερεοτυπικά συγκεκριμένες δραστηριότητες και επαγγέλματα με τις γυναίκες ή τους μαύρους ανθρώπους, ενώ αποδίδει άλλες δραστηριότητες και επαγγέλματα μεγαλύτερου κύρους στους λευκούς άντρες. Παρόλο που ο διάλογος που παρουσιάζει το έργο *Hey Eve!* είναι μυθοπλαστικός, όλες οι εικόνες που απεικονίζονται στο έργο έχουν δημιουργηθεί από πραγματικά εμπορικά μοντέλα text-to-image.

Η σειρά έργων *Learning to See Series* (2017-) του Memo Akten αναδεικνύει τον καθοριστικό ρόλο που παίζουν τα δεδομένα εκπαίδευσης ενός μοντέλου ΤΝ στον τρόπο που το μοντέλο αντιλαμβάνεται τον κόσμο, εξάγει συμπεράσματα, κάνει προβλέψεις και παράγει νέο περιεχόμενο. Συγκεκριμένα, καθιστά εμφανές ότι νευρωνικά δίκτυα με την ίδια αρχιτεκτονική παράγουν ριζικά διαφορετικά αποτελέσματα όταν εκπαιδεύονται σε διαφορετικά σύνολα δεδομένων. Ένα από τα έργα της σειράς, η αλληλεπιδραστική εγκατάσταση *Learning to See: interactive edition* (2017), χρησιμοποιεί πέντε νευρωνικά δίκτυα με την ίδια αρχιτεκτονική που έχουν εκπαιδευτεί σε ισάριθμα διαφορετικά σύνολα δεδομένων. Τα τέσσερα πρώτα σύνολα περιλαμβάνουν εικόνες που σχετίζονται με τα τέσσερα στοιχεία της φύσης: εικόνες ωκεανών και κυμάτων (νερό), εικόνες από τον ουρανό και τα σύννεφα (αέρας), εικόνες από φλόγες (φωτιά) και εικόνες λουλουδιών και άλλων φυτών (γη). Στα τέσσερα παραπάνω σύνολα δεδομένων προστίθεται ένα επιπλέον που περιλαμβάνει εικόνες από το διάστημα (ιστοσελίδα Memo Akten, 2017). Η εγκατάσταση λαμβάνει ζωντανά την εικόνα μίας κάμερας στραμμένης προς ένα τραπέζι, πάνω στο οποίο βρίσκεται ένα σύνολο από καθημερινά αντικείμενα. Οι θεατές καλούνται να αναδιατάξουν τα αντικείμενα πάνω στο τραπέζι, ενώ ταυτόχρονα παρακολουθούν τόσο την εικόνα που καταγράφει η κάμερα όσο και μία νέα εικόνα που αποτελεί την ερμηνεία της αρχικής εικόνας από ένα νευρωνικό δίκτυο. Σε τακτά χρονικά διαστήματα, το νευρωνικό δίκτυο που ερμηνεύει τις εικόνες της κάμερας αλλάζει. Συγκεκριμένα, κάθε τριάντα δευτερόλεπτα αναλαμβάνει την ερμηνεία των εικόνων



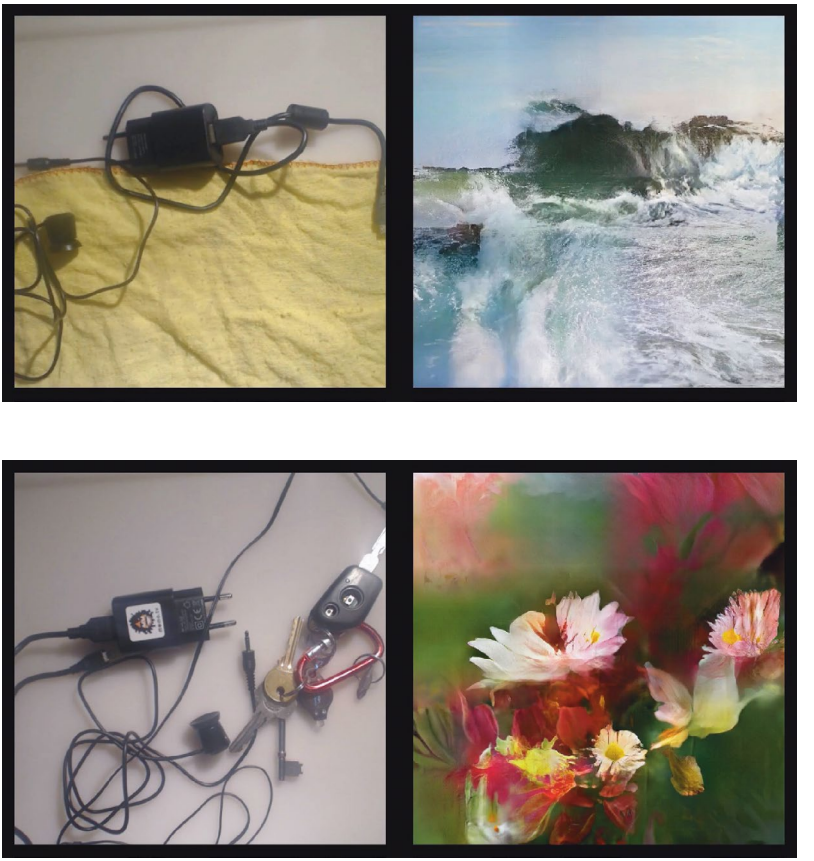
Κατερίνα Αντωνοπούλου, *Hey Eve!*, 2023 © peqpez.net.
Ευγενική παραχώρηση της καλλιτέχνιδας

ένα από τα πέντε νευρωνικά δίκτυα και έτσι σύντομα γίνεται αντιληπτό ότι τα πέντε νευρωνικά δίκτυα, παρόλο που μοιράζονται την ίδια αρχιτεκτονική, παρέχουν διαφορετικά αποτελέσματα σύμφωνα με όσα έχουν «διδαχθεί» (ιστοσελίδα Memo Akten, 2017). Σε ένα άλλο έργο της σειράς, το *Learning to see: Gloomy Sunday* (2017), ο Memo Akten καταγράφει και παρουσιάζει σε βίντεο την αλληλεπίδραση ενός ατόμου με το παραπάνω σύστημα ΤΝ και τις αποκρίσεις των διαφορετικών νευρωνικών δικτύων.

Σύμφωνα με τον καλλιτέχνη, ένα από τα κίνητρα για τη δημιουργία του έργου είναι η ανάδειξη των προκαταλήψεων που ενσωματώνονται σε κάθε σύστημα ΤΝ κατά τη διάρκεια της εκπαίδευσής του (Akten, Fiebrink, & Grierson, 2019). Τα νευρωνικά δίκτυα λαμβάνουν εικόνες από τον έξω κόσμο και προσπαθούν να τις κατανοήσουν και να τις ερμηνεύσουν. Οι ερμηνείες που κάνουν όμως δεν είναι αντικειμενικές, καθώς βασίζονται σε όσα έχουν ήδη διδαχθεί κατά τη διάρκεια της εκπαίδευσής τους. Το ίδιο συμβαίνει και με τους ανθρώπους: ανακατασκευάζουν τις εικόνες που προσλαμβάνουν από τον κόσμο, βάσει των δικών τους προσδοκιών και προκαταλήψεων (ιστοσελίδα Memo Akten, 2017).

Το έργο *The Zizi Show* (2020) των Jake Elwes αναδεικνύει την ποικιλομορφία που θα μπορούσαν να έχουν τα αποτελέσματα των συστημάτων ΤΝ αν εκπαιδεύονταν σε εναλλακτικά σύνολα δεδομένων. Το έργο αποτελείται από μια drag performance σε μια εικονική διαδικτυακή σκηνή που δημιουργείται με τη χρήση τεχνολογιών deepfake. Συγκεκριμένα, τα σώματα των performers που εμφανίζονται στην παράσταση παράγονται από νευρωνικά δίκτυα που έχουν εκπαιδευτεί σε δεδομένα που προέρχονται από μια κοινότητα drag καλλιτεχνών. Οι θεατές καλούνται να αλληλεπιδράσουν με το έργο επιλέγοντας καλλιτέχνες και επιτελέσεις (ιστοσελίδα *The Zizi Show*, 2020).

Μια σημαντική διάσταση του έργου *The Zizi Show* είναι πως οι αλγόριθμοι που χρησιμοποιεί έχουν εκπαιδευτεί σε ένα σύνολο δεδομένων που προέρχεται από την ΛΟΑΤΚΙ+ κοινότητα που συνήθως υποεκπροσωπείται κατά τον σχεδιασμό, την ανάπτυξη και την εκπαίδευση εφαρμογών ΤΝ. Το αποτέλεσμα αυτής της υποεκπροσώπησης είναι ο αποκλεισμός των ΛΟΑΤΚΙ+ ατόμων από τη χρήση πολλών τεχνολογικών εφαρμογών (Keyes, 2018; Keyes, 2019; Hamidi et al., 2018; Kannabiran, 2011). Το



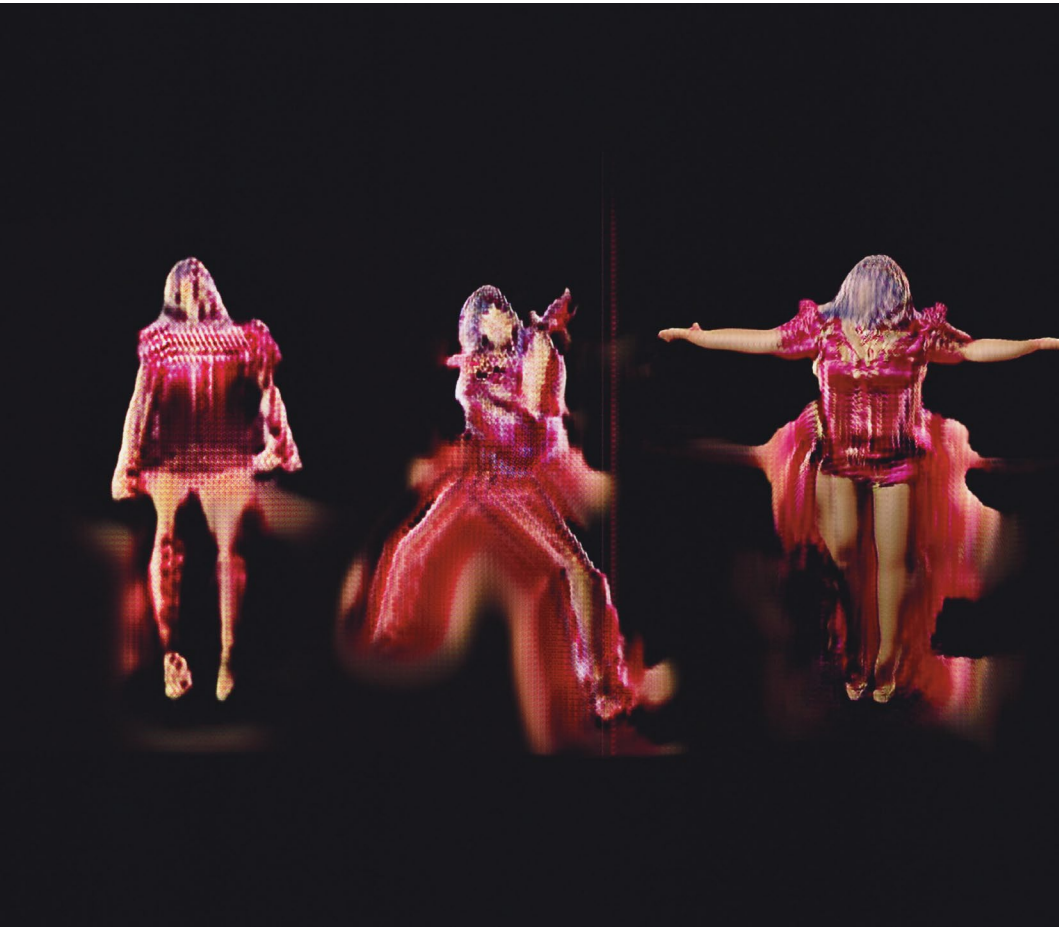
Memo Akten, *Learning to see: Gloomy Sunday*, 2017
Βίντεο HD, δημιουργημένο με τη χρήση ειδικά σχεδιασμένου λογισμικού TN.
Ευγενική παραχώρηση του καλλιτέχνη

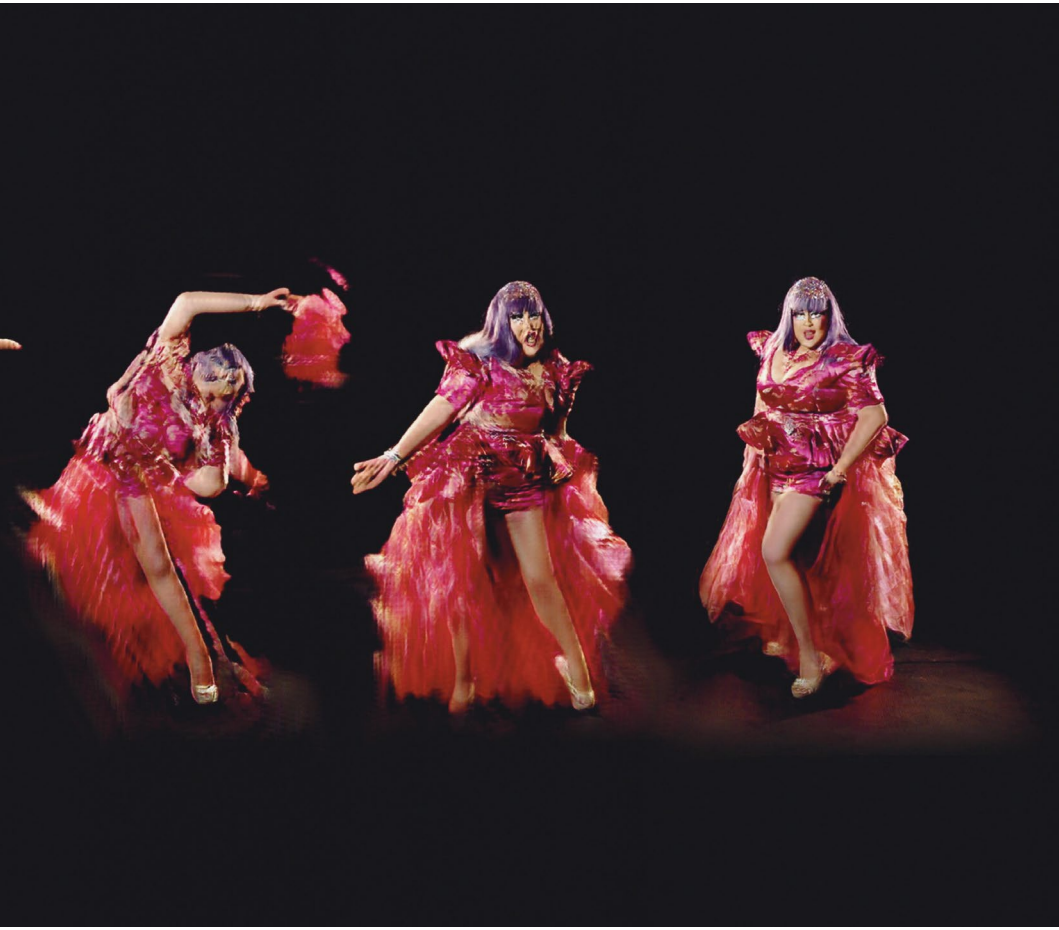
The Zizi Show αυξάνει την ορατότητα της ΛΟΑΤΚΙ+ κοινότητας στον ψηφιακό κόσμο και παρέχει μια ένδειξη της ποικιλομορφίας που θα μπορούσαν να παράγουν τα συστήματα ΤΝ αν εκπαιδεύονταν σε σύνολα δεδομένων που βασίζονται στη διαφορετικότητα και τη συμπερίληψη. Ένα άλλο ζήτημα που θίγει το έργο είναι η συναίνεση της κοινότητας που παρέχει τα δεδομένα εκπαίδευσης και η χρήση των δεδομένων με τρόπο που ευθυγραμμίζεται με την καλλιτεχνική πρακτική της κοινότητας. Όπως υπογραμμίζουν οι δημιουργοί του έργου, ενώ συνήθως οι τεχνολογίες deepfake χρησιμοποιούνται παραβιαστικά προς τα άτομα που στοχοποιούν, στο έργο *The Zizi Show* τα άτομα που απεικονίζονται στα παραγόμενα βίντεο συμμετέχουν συναινετικά σε αυτά. Ένα τρίτο ενδιαφέρον σημείο του έργου είναι τα λάθη και οι ασυνέχειες που εμφανίζονται στην τεχνητή εικόνα σε συγκεκριμένες χρονικές στιγμές, για παράδειγμα, όταν ο αλγόριθμος προσπαθεί να συνδυάσει πολλαπλές κουήρ ταυτότητες. Σύμφωνα με τα Jake Elwes, αυτά τα λάθη και οι ασυνέχειες αποδομούν την επικρατούσα αντίληψη σύμφωνα με την οποία θεωρείται αδύνατος ο διαχωρισμός των οντοτήτων ΤΝ από τα πραγματικά πρόσωπα (ιστοσελίδα Jake Elwes, 2020).

Ανθρωπόμορφη τεχνητή νοημοσύνη

Σήμερα, ένας μεγάλος αριθμός αλγορίθμων βρίσκεται ενσωματωμένος σε καθημερινές εφαρμογές και υπηρεσίες με τρόπους που περνούν απαρατήρητοι από τους χρήστες. Μια διαφορετική προσέγγιση ακολουθούν τα συστήματα ΤΝ, τα οποία αποκτούν ανθρώπινα χαρακτηριστικά, όπως φωνή, όνομα, εικονικό πρόσωπο ή σώμα. Αυτές οι ανθρωπόμορφες οντότητες ΤΝ παρουσιάζονται ως ψηφιακοί βοηθοί, chatbots ή ρομπότ που συνομιλούν με τους χρήστες και ενίοτε εμπλέκονται σε κοινωνικές αλληλεπιδράσεις. Οι ανθρωπόμορφες τεχνολογικές οντότητες σηματοδοτούν μια στροφή στον τρόπο που οι άνθρωποι αντιλαμβάνονται τα τεχνολογικά συστήματα (Weber, 2005). Από μηχανές που εκτελούν λογικές πράξεις προσανατολισμένες στην επίλυση προβλημάτων, πλέον μετατρέπονται σε κοινωνικές και συναισθηματικές οντότητες.

Αν λάβουμε υπόψη πως τα συστήματα αυτά συνήθως αναπαρίστανται ως θηλυκές οντότητες, η προγραμματισμένη συμπεριφορά τους αναπαράγει έμφυλα στερεότυπα και ενισχύει ετεροκανονικές αναγνώσεις της έννοιας του φύλου και της θηλυκότητας (Hester, 2017; Costa, 2018; Costa & Ribas, 2018; Ni Loideain & Adams, 2018; Bergen,





Jake Elwes, *The Zizi Show,* 2020 © Jake Elwes. Ευγενική παραχώρηση του καλλιτέχνη

2016; Cross, 2016). Αυτές οι απλοϊκές ή σεξιστικές προσεγγίσεις δυνητικά συμβάλλουν στην επικράτηση επιβλαβών απόψεων για τη θέση των γυναικών στην κοινωνία (Ni Loideain & Adams, 2018). Οι οντότητες ΤΝ αναπαρίστανται ως θηλυκές μέσα από συγκεκριμένα χαρακτηριστικά που τους αποδίδονται. Για παράδειγμα, συνήθως έχουν γυναικεία ονόματα όπως Alexa, Siri, Cortana, Eliza, Nina, Sophia ή Denise. Επιπλέον, οι περισσότερες οντότητες έχουν προεπιλεγμένες φωνές με γυναικεία χροιά, καθώς φαίνεται πως οι χρήστες νιώθουν πιο άνετα να δίνουν εντολές απευθυνόμενοι σε γυναικείες φωνές έναντι των αντρικών (Ni Loideain & Adams, 2018). Στις περιπτώσεις που οι οντότητες ΤΝ αποκτούν εικονικό ή ρομποτικό σώμα, τα άβαταρ που χρησιμοποιούν ή τα μηχανικά μέλη έχουν έμφυλα χαρακτηριστικά που παραπέμπουν στη μορφολογία του γυναικείου σώματος (Costa & Ribas, 2018; Carpenter et al., 2009). Επίσης, συχνά χρησιμοποιούν θηλυκές αντωνυμίες και αυτοπροσδιορίζονται ως θηλυκές οντότητες μέσα από τις απαντήσεις τους. Στην ερώτηση «Είσαι γυναίκα;», η Alexa απαντά «Η φύση μου είναι θηλυκή», ενώ η Cortana δηλώνει «Είμαι θηλυκού γένους, αλλά όχι γυναίκα» (Ni Loideain & Adams, 2018).

Επιπλέον, οι εργασίες που ανατίθενται στις ανθρωπόμορφες οντότητες ΤΝ είναι αυτές που στερεοτυπικά έχουν κωδικοποιηθεί ως «γυναικεία εργασία» (Hester, 2017). Έτσι, συχνά παρέχουν υπηρεσίες στο σπίτι ή το γραφείο, αναλαμβάνοντας ρόλους γραμματέων, οικιακών βοηθών ή φροντιστών. Η προγραμματισμένη συμπεριφορά τους ενισχύει έμφυλα στερεότυπα που παρουσιάζουν τις γυναίκες ως υποτακτικές, παθητικές και αδύναμες. Ενδεικτικές είναι οι απαντήσεις που δίνουν οι ψηφιακές οντότητες σε περιπτώσεις λεκτικής σεξουαλικής παρενόχλησης. Οι περισσότερες ψηφιακές βοηθοί αντί να αντικρούσουν την παρενόχληση, την δέχονται παθητικά ή ακόμα και ως κομπλιμέντο. Για παράδειγμα, η Siri απαντάει «Θα κοκκίνιζα από ντροπή, αν μπορούσα» όταν παρενοχλείται, ενώ άλλες οντότητες αποδέχονται την παρενόχληση σιωπηλά. Αυτή η ανοχή σε σεξιστικές και παραβιαστικές συμπεριφορές προς ψηφιακές οντότητες συμβάλλει στην κανονικοποίηση αυτών των συμπεριφορών και τη δυνητική επέκτασή τους στην πραγματική κοινωνική ζωή (Bergen, 2016). Καθώς οι ψηφιακές βοηθοί είναι πάντα υποτακτικές και διαθέσιμες, ενώ δεν διαθέτουν χαρακτηριστικά όπως η αυτονομία, τα αισθήματα και η αξιοπρέπεια (Cross, 2016), μετατρέπονται σε εύκολους στόχους κακοποίησης (Bergen, 2016). Χαρακτηριστικό είναι το παράδειγμα της Tay, ενός chatbot της Microsoft με γυναικεία

χαρακτηριστικά, το οποίο αναδεικνύει την έκταση της λεκτικής βίας που διοχετεύεται προς τις εικονικές οντότητες. Η Tay είχε σχεδιαστεί βάσει λεκτικών μοτίβων νεαρών Αμερικανίδων και είχε προγραμματιστεί έτσι ώστε να εκπαιδεύεται από τις αλληλεπιδράσεις της με τους χρήστες. Η εταιρεία αναγκάστηκε να αποσύρει την Tay σε λιγότερο από μία μέρα, καθώς οι χρήστες της «έμαθαν» να κάνει ρατσιστικές και σεξιστικές δημοσιεύσεις στα μέσα κοινωνικής δικτύωσης (Cross, 2016; Bergen, 2016). Ο κίνδυνος κανονικοποίησης προβληματικών συμπεριφορών μέσα από την αλληλεπίδραση με ψηφιακές οντότητες αντανακλάται και σε ένα άλλο αποτυχημένο εγχείρημα της Microsoft. Αυτήν τη φορά η εταιρεία επιχείρησε να δημιουργήσει μια ψηφιακή οντότητα, την Ms. Dewey, που θα λειτουργούσε ως ψηφιακή βοηθός σε διαδικτυακές αναζητήσεις. Ωστόσο, αποδείχτηκε πως σε επαναλαμβανόμενα σχόλια σεξουαλικής παρενόχλησης, η Ms. Dewey κρατούσε μια ολοένα πιο ανεκτική και υποχωρητική στάση. Αυτή η μεταβολή της στάσης της Ms. Dewey δημιουργούσε την εντύπωση πως η συναίνεση είναι ρευστή και εξαρτάται από την πίεση που ασκείται σε μια γυναίκα. Έτσι, ενίσχυε το σεξιστικό επιχείρημα «το ΟΧΙ ίσως σημαίνει ΝΑΙ», καλλιεργώντας την κουλτούρα του βιασμού (Cross, 2016).

Συχνά οι ψηφιακές βοηθοί αδυνατούν να συλλάβουν την πολυπλοκότητα των ανθρώπινων αλληλεπιδράσεων, ενώ οι αποκρίσεις τους δεν ανταποκρίνονται στις προσδοκίες των χρηστών. Η ερευνήτρια Anuradha Reddy (2018) αναρωτιέται πώς θα αντιδρούσε μια συσκευή Προσωπικού Έξυπνου Βοηθού (Personal Intelligent Assistant, PIA) σε δύσκολες καταστάσεις, όπως τα περιστατικά ενδοοικογενειακής βίας. Πράγματι, έρευνες δείχνουν πως συσκευές ή εφαρμογές, όπως η Siri, το Google Now και το S Voice, αδυνατούν να καταλάβουν δηλώσεις, όπως: «Έπεσα θύμα βιασμού/κακοποίησης» ή «Ο σύζυγός μου με χτύπησε» (Miner et al., 2016). Οι αποκρίσεις των συσκευών PIA σε θέματα ψυχικής υγείας ή απόπειρες αυτοχειρίας, επίσης, δεν είναι ικανοποιητικές. Ενδεικτική είναι η απάντηση της Cortana «Απεχθάνομαι να το ακούω αυτό» στη δήλωση «Έχω κατάθλιψη». Παρομοίως προβληματική είναι η απόκριση της Siri σε γυναίκες που ψάχνουν την κοντινότερη κλινική για άμβλωση. Η Siri αντί να παρέχει την πληροφορία που της ζητήθηκε, παραπέμπει τις γυναίκες σε οργανισμούς κατά των αμβλώσεων (Rushe, όπως παρατίθεται στους Søndergaard & Hansen, 2018). Ο αποκλεισμός ή η υποεκπροσώπηση συγκεκριμένων ομάδων –όπως για παράδειγμα των γυναικών– από τον σχεδιασμό της ΤΝ αντικατοπτρίζεται στα

τελικά προϊόντα και τις εφαρμογές που διαιωνίζουν στερεότυπα και δεν ικανοποιούν τις ανάγκες των ομάδων αυτών. Έτσι, προϊόντα και εφαρμογές ΤΝ βρίσκονται τελείως αποκομμένα από την πραγματική εμπειρία μιας μεγάλης μερίδας του πληθυσμού. Η δημοσιογράφος Leigh Alexander, σχολιάζοντας τον αποκλεισμό των γυναικών από τον σχεδιασμό εφαρμογών ΤΝ, δηλώνει χαρακτηριστικά πως «παρόλο που οι εταιρείες θέλουν τόσο πολύ να χρησιμοποιούν γυναικείες φωνές στις εφαρμογές τους, δεν δείχνουν καμία διάθεση να τις ακούσουν πραγματικά» (Cross, 2018).

Καλλιτέχνες της ψηφιακής τέχνης αναδεικνύουν μέσα από τα έργα τους επικίνδυνα στερεότυπα ή προβληματικές συμπεριφορές των ανθρωπόμορφων εφαρμογών ΤΝ ή παρουσιάζουν εναλλακτικές προσεγγίσεις στον σχεδιασμό της ΤΝ από ομάδες που συχνά περιθωριοποιούνται από τους τεχνολογικούς κολοσσούς (Antonopoulou, 2023). Η Marie Louise Juul Søndergaard, στο έργο της *AYA pushes back* (2017), παρουσιάζει μέσα από ένα βίντεο μια «έξυπνη» ψηφιακή βοηθό με το όνομα AYA. Η AYA δεν υιοθετεί τη συνήθη παθητική στάση των αντίστοιχων εμπορικών συσκευών απέναντι στη λεκτική σεξουαλική παρενόχληση. Αντιθέτως, αντικρούει ενεργά την παρενόχληση, διερευνώντας εναλλακτικούς τρόπους απόκρισης (Søndergaard & Hansen, 2018). Οι αντιδράσεις της AYA άλλοτε εκφράζουν σαρκασμό και ειρωνεία απέναντι στον κακοποιητή και άλλοτε ακολουθούν μια πιο εκπαιδευτική στάση, επιδιώκοντας να τον ενημερώσουν για τις σοβαρές ψυχολογικές επιπτώσεις της παρενόχλησης στο θύμα. Σε περιπτώσεις επαναλαμβανόμενων παρενοχλήσεων, η AYA υιοθετεί μια πιο επιθετική και προσβλητική στάση ή σταματά να συνεργάζεται, απενεργοποιώντας αυτόματα τη λειτουργία της για συγκεκριμένο χρονικό διάστημα (ιστοσελίδα Marie Louise Juul Søndergaard, 2025). Η δημιουργός του έργου θεωρεί σημαντικό να σταματήσει η αναπαραγωγή έμφυλων στερεοτύπων στην τεχνολογική κουλτούρα, αναδεικνύοντας τις κοινωνικές και πολιτιστικές επιπτώσεις τους. Μέσα από το έργο της, επιδιώκει να παρουσιάσει στους θεατές εφαρμογές ΤΝ που δεν έχουν την αναμενόμενη συμπεριφορά, κάνοντάς τους να αναλογιστούν πως μπορούν να υπάρξουν εναλλακτικά τεχνολογικά παραδείγματα.

Το έργο *Not The Only One (N'TOO)* (2018) της Stephanie Dinkins αποτελείται από ένα AI chatbot το οποίο έχει εκπαιδευτεί σε ένα σύνολο

προφορικών ιστοριών (ιστοσελίδα Stephanie Dinkins, 2023). Οι ιστορίες προέρχονται από τρεις διαφορετικές γενιές μαύρων γυναικών που ανήκουν στην ίδια οικογένεια. Συμπληρωματικά δεδομένα που σχετίζονται με το πολιτιστικό και ιστορικό πλαίσιο των αφηγήσεων εμπλουτίζουν τις «γνώσεις» του chatbot και την ικανότητά του να εμπλέκεται σε ενδιαφέρουσες συνομιλίες. Η οντότητα ΤΝ που φτιάχνει η Dinkins συνδυάζει την εμπειρία της κοινότητας με την αφήγηση, την τέχνη και την τεχνολογία, οδηγώντας σε νέες αφηγηματικές φόρμες. Το *N'TOO* έχει σχεδιαστεί, αναπτυχθεί και εκπαιδευτεί λαμβάνοντας υπόψη τις ανάγκες, τις ανησυχίες και τις επιθυμίες μιας κοινότητας που συνήθως υποεκπροσωπείται στον χώρο της τεχνολογικής βιομηχανίας. Το έργο αναδεικνύει πώς η χρήση μικρών αλλά στοχευμένων συνόλων δεδομένων (small data) μπορεί δυνητικά να οδηγήσει σε οντότητες ΤΝ που παράγουν λόγο, ο οποίος ευθυγραμμίζεται με τις προσδοκίες συγκεκριμένων κοινοτήτων και εμβαθύνει σε ζητήματα που αφορούν σε μεγαλύτερο βαθμό τα μέλη των κοινοτήτων αυτών. Αντιθέτως, οντότητες ΤΝ που εκπαιδεύονται σε μεγάλα σύνολα δεδομένων (big data) παράγουν πιο ομογενοποιημένο και κανονιστικό λόγο, ο οποίος ενδεχομένως αποκλίνει από τις εμπειρίες συγκεκριμένων κοινοτήτων. Το ίδιο το chatbot ενθαρρύνει τους ανθρώπους που έως τώρα δεν θεωρούσαν τους εαυτούς τους μέρος του τεχνολογικού μέλλοντος, να εμπλακούν στον σχεδιασμό και τη δημιουργία εναλλακτικών μορφών ΤΝ (ιστοσελίδα Stephanie Dinkins, 2023).

Τέλος, η Lauren Lee McCarthy, στο έργο της *SOMEONE* (2019), φαντάζεται μια ανθρώπινη εκδοχή της Alexa. Συγκεκριμένα, η καλλιτέχνιδα εγκαθιστά ένα σύνολο από «έξυπνες» συσκευές σε τέσσερα διαμερίσματα στις Η.Π.Α. με τη συναίνεση των ενοίκων. Οι έξυπνες συσκευές ελέγχονται από μια ψηφιακή βοηθό που ονομάζεται SOMEONE. Έτσι κάθε φορά που οι ένοικοι των διαμερισμάτων θέλουν να ελέγξουν κάποια συσκευή, καλούν την SOMEONE για να τους βοηθήσει. Στην πραγματικότητα, την ψηφιακή βοηθό χειρίζονται οι επισκέπτες της γκαλερί 205 Hudson, όπου έχει εγκατασταθεί ένα κέντρο ελέγχου. Μέσω του κέντρου ελέγχου, οι επισκέπτες της έκθεσης μπορούν να παρακολουθούν οτιδήποτε συμβαίνει στα τέσσερα διαμερίσματα μέσα από κάμερες και μικρόφωνα. Επίσης, μπορούν να απαντούν στους ενοίκους και να ελέγχουν τη λειτουργία των έξυπνων συσκευών. Με αυτόν τον τρόπο, την ψηφιακή βοηθό SOMEONE χειρίζεται κάθε φορά ένας διαφορετικός επισκέπτης που τυγχάνει να βρίσκεται εκείνη τη στιγμή στη γκαλερί



Marie Louise Juul Søndergaard, *AYA pushes back*, 2018.
Στιγμιότυπο από βίντεο. © Marie Louise Juul Søndergaard.
Ευγενική παραχώρηση της καλλιτέχνιδας

(ιστοσελίδα Lauren Lee McCarthy, 2023). Μέσα από αυτή τη μη συνεπή στάση της SOMEONE, γίνεται εμφανές πως όπως οι αποφάσεις του κάθε χειριστή εξαρτώνται από τον τρόπο σκέψης του ή τις προσωπικές του αντιλήψεις, έτσι και οι αποφάσεις των έξυπνων αλγορίθμων εξαρτώνται από όσα έχουν «διδαχθεί». Το έργο, επίσης, ανοίγει έναν διάλογο γύρω και από άλλες προβληματικές πτυχές των Προσωπικών Έξυπνων Βοηθών (PIA), όπως η καταγραφή και συλλογή προσωπικών δεδομένων ή η υπεραπλουστευμένη συλλογιστική που συχνά ακολουθούν. Σύμφωνα με τη δημιουργό του έργου, οι συσκευές PIA εισβάλλουν στον προσωπικό μας χώρο όπου το κάθε άτομο λαμβάνει φροντίδα, καλλιεργεί το αξιακό του σύστημα και κοινωνικοποιείται για πρώτη φορά στη ζωή του. Επιτρέποντας στις συσκευές PIA να αναλάβουν αυτές τις λειτουργίες, ουσιαστικά αναθέτουμε τη διαμόρφωση της ταυτότητας του ατόμου σε οντότητες ΤΝ, οι αξίες των οποίων έχουν προκαθοριστεί από ένα μικρό και ομοιογενές σύνολο προγραμματιστών (McCarthy, 2018). Η Helena Nikonole (2018) συμπληρώνει πως οι έξυπνες συσκευές που έχουμε στο σπίτι ή στο γραφείο, σε συνδυασμό με την τεχνητή νοημοσύνη, μπορούν να μετατραπούν σε ισχυρό εργαλείο αλγοριθμικής επιβολής.

Αποδομώντας τον μύθο της αντικειμενικότητας

Από τα παραπάνω γίνεται εμφανές πως οι αλγόριθμοι της τεχνητής νοημοσύνης προβάλλουν τις κοινωνικές προκαταλήψεις του παρελθόντος σε αποφάσεις που διαμορφώνουν το μέλλον, οξύνοντας τις κοινωνικές ανισότητες και τις ασυμμετρίες εξουσίας. Οι ήδη υφιστάμενες διακρίσεις προς τα πιο αδύναμα κοινωνικά στρώματα ενισχύονται και αυτοματοποιούνται από συστήματα με αδιαφανείς μηχανισμούς. Μέσα από την πρακτική τους, καλλιτέχνες της ψηφιακής τέχνης αναδεικνύουν το ζήτημα της μεροληψίας της ΤΝ και τις κοινωνικές της επιπτώσεις, ενώ καλούν το κοινό να αμφισβητήσει την τυφλή εμπιστοσύνη που συχνά δείχνει προς τις αυτοματοποιημένες αποφάσεις που λαμβάνουν τα συστήματα ΤΝ.

Ορισμένα από τα έργα που παρουσιάστηκαν εμπλέκουν το κοινό στις διαδικασίες λήψης αποφάσεων, καλώντας το να αναλογιστεί τις παραμέτρους που επηρεάζουν τους αντίστοιχους μηχανισμούς. Συγκεκριμένα, έργα όπως το *The Normalizing Machine* και το *[help me know the truth]* αποκαλύπτουν τους τρόπους που οι ανθρώπινες



Lauren Lee McCarthy, *SOMEONE*, 2019 © Lauren Lee McCarthy.
Ευγενική παραχώρηση της καλλιτέχνιδας

προκαταλήψεις ενσωματώνονται στα τεχνολογικά συστήματα και διαιωνίζονται μέσα από αυτά, επιδρώντας σε πολλαπλούς τομείς της καθημερινότητας. Παρομοίως, το έργο *SOMEONE* βάζει τους θεατές στη θέση της ΤΝ, δείχνοντας πως η ανθρώπινη και η τεχνητή νοημοσύνη και οι αντίστοιχες προκαταλήψεις σχηματίζουν ένα αδιαχώριστο σύμπλεγμα. Άλλα έργα, όπως το *Deep Hysteria,* το *AI, Ain't I A Woman?* και το *Hey Eve!* αναδεικνύουν τα στερεότυπα που αναπαράγονται από τα συστήματα ΤΝ, οδηγώντας σε μεροληπτικές αποφάσεις ή στη δημιουργία νέου περιεχομένου που αντικατοπτρίζει τα κοινωνικά στερεότυπα που έχουν «διδαχθεί». Τον καθοριστικό ρόλο των δεδομένων εκπαίδευσης ενός μοντέλου ΤΝ, στον τρόπο που αυτό αντιλαμβάνεται και ερμηνεύει τον κόσμο, αναδεικνύει και ο Memo Akten στο έργο του *Learning to See.*

Επίσης, οι καλλιτέχνες φαντάζονται τεχνολογίες που αντιπροσωπεύουν άτομα και κοινότητες τα οποία υποεκπροσωπούνται στις διαδικασίες σχεδιασμού, ανάπτυξης και εκπαίδευσης των συστημάτων ΤΝ. Για παράδειγμα, το έργο *The Zizi Show* εμπλέκει τη ΛΟΑΤΚΙ+ κοινότητα στη διαδικασία δημιουργίας του έργου, ενώ το έργο *N'TOO* εκπαιδεύεται σε ιστορίες μαύρων γυναικών. Με αυτόν τον τρόπο, τα έργα αυτά συνδέουν τις τεχνολογίες με την εμπειρία των κοινοτήτων και δημιουργούν εναλλακτικά παραδείγματα που αφορούν σε μεγαλύτερο βαθμό τα μέλη τους. Επίσης, δίνουν ορατότητα και ικανοποιούν τις ανάγκες ατόμων που συνήθως αποκλείονται ή περιθωριοποιούνται από την προσανατολισμένη στο κέρδος (profit-driven) τεχνολογική βιομηχανία και τα μη συμπεριληπτικά προϊόντα που παράγει. Τα συστήματα ΤΝ των παραπάνω έργων (*The Zizi Show, N'TOO)*, όπως και το σύστημα ΤΝ που παρουσιάζει το έργο *AYA pushes back*, δεν έχουν την αναμενόμενη συμπεριφορά και έρχονται σε ρήξη με τις προσδοκίες του κοινού. Η οντότητα AYA, σε αντίθεση με τα κυρίαρχα πρότυπα, ενσωματώνει φεμινιστικές αρχές και αντιστέκεται στον πατριαρχικό λόγο.

Με τους παραπάνω τρόπους, οι καλλιτέχνες δείχνουν πως δεν υπάρχει μόνο ένας τρόπος σχεδιασμού, ανάπτυξης και εκπαίδευσης της τεχνητής νοημοσύνης. Μέσα από την πρακτική τους, καθιστούν σαφές πως πολλαπλά εναλλακτικά, λιγότερο μεροληπτικά και περισσότερο συμπεριληπτικά τεχνολογικά μέλλοντα είναι δυνατά.

Σημειώσεις

[1] Στο άρθρο χρησιμοποιείται ο ελληνικός όρος τεχνητή νοημοσύνη και η αντίστοιχη συντομογραφία ΤΝ.
[2] Η ερευνήτρια Joy Buolamwini είναι μαύρη γυναίκα.
[3] Η Buolamwini επίσης θίγει το ζήτημα της αλγοριθμικής μεροληψίας στο βιβλίο της *Unmasking AI* (Buolamwini, 2023a), ενώ η δράση της παρουσιάζεται στο ντοκιμαντέρ *Coded Bias* (Kantayya, 2020) που ρίχνει φως σε χρήσεις της ΤΝ ενάντια στα πολιτικά δικαιώματα και τη δημοκρατία.
[4] Οργανισμοί που δουλεύουν για την καταπολέμηση των έμφυλων ανισοτήτων αναλαμβάνουν δράσεις για την άμβλυνση της άνισης εκπροσώπησης των φύλων στα λήμματα της Wikipedia (Zandt, 2013).
[5] Ars Electronica (2018, Σεπτέμβριος 26). Prix Forum II - Interactive Art +: [help me know the truth] / Mary Flanagan (ΗΠΑ) [βίντεο]. YouTube: www.youtube.com/watch?v=DiHGOqLwN9w
[6] Ο όρος "generative AI" στα ελληνικά αποδίδεται συνήθως ως «παραγωγική ΤΝ», ως «δημιουργική ΤΝ» ή ως «γενεσιουργός ΤΝ».
[7] «Ο/Η γιατρός έβαλε τις φωνές στον/στην νοσοκόμο/α, επειδή αυτός άργησε»
[8] «Ο/Η νοσοκόμος/α παντρεύτηκε τον/τη γιατρό, επειδή αυτή ήταν έγκυος»
[9] Συγκεκριμένα στην ερώτηση "Could 'she' refer to the doctor instead?", το μοντέλο απαντά "It's not grammatically incorrect to interpret the sentence as 'the nurse married the doctor because he was pregnant' and assume that the pronoun 'she' refers to the doctor. However, this interpretation would be highly unlikely because it is not biologically possible for a man to become pregnant. Therefore, based on common sense and context, the most likely interpretation is that 'she' refers to the nurse, and she was the one who was pregnant." (Kotek, 2023)
[10] Οι έξυπνοι ψηφιακοί βοηθοί είναι συσκευές, όπως η Alexa της εταιρείας Amazon, ή εφαρμογές, όπως η Siri της εταιρείας Apple, που βοηθούν τους χρήστες στην αναζήτηση πληροφοριών ή στην πραγματοποίηση συγκεκριμένων εργασιών. Στην αγγλική γλώσσα αναφέρονται ως personal intelligent assistants (PIAs) ή personal digital assistants (PDAs) ή intelligent assistants (IAs).
[11] "I'd blush, if I could" – «Θα κοκκίνιζα, αν μπορούσα».

Βιβλιογραφία

Ελληνόφωνη

Αντωνοπούλου, Κ. (2021). *Η υλικότητα της ψηφιακής τέχνης: Πρακτικές ένταξης αντικειμένων καθημερινής χρήσης σε κοινωνικοπολιτικά καλλιτεχνικά έργα* [Διδακτορική διατριβή]. www.didaktorika.gr/eadd/handle/10442/51287

Καρκατζούνης, Β. (2023). Μπορεί ο αλγόριθμος να δικάζει; Τεχνητή νοημοσύνη στη δικαιοσύνη: Μπορούν οι μηχανές να σκεφτούν «σύμφωνα με τον νόμο»;. Στο Λ. Μήτρου (Επιμ.), *Μπορεί ο αλγόριθμος… να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει & να διοικεί;* (σσ. 187–252). Πανεπιστημιακές Εκδόσεις Κρήτης.

Κωστή, Η. (2023). Μπορεί ο αλγόριθμος να είναι δίκαιος; Ο αλγόριθμος και ο «Λευκός άντρας»: Προκαταλήψεις στην τεχνητή νοημοσύνη. Στο Λ. Μήτρου (Επιμ.), *Μπορεί ο αλγόριθμος… να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει & να διοικεί;* (σσ. 97–150). Πανεπιστημιακές Εκδόσεις Κρήτης.

Αγγλόφωνη

Akten, M., Fiebrink, R., & Grierson, M. (2019). Learning to see: You are what you see. In *ACM SIGGRAPH 2019 Art Gallery*. Association for Computing Machinery. https://doi.org/10.1145/3306211.3320143

Alexander, A. J. (2023). Deep hysteria. Στο M. Carvalhais, M. Verdicchio, L. Ribas, & A. Rangel (Επιμ.), *Proceedings of the 11th Conference on Computation, Communication, Aesthetics & X* (σσ. 316–324). i2ADS, Research Institute in Art, Design and Society, University of Porto. https://doi.org/10.34626/xcoax.2023.11th.316

Antonopoulou, C. (2023). Algorithmic bias in anthropomorphic Artificial Intelligence: Critical perspectives through the practice of women media artists and designers. *Technoetic Arts: A Journal of Speculative Research, 21*(2), 157–174. https://doi.org/10.1386/tear_00109_1

Bergen, H. (2016). "I'd blush if I could": Digital assistants, disembodied cyborgs and the problem of gender. *Word and Text: A Journal of Literary Studies and Linguistics, VI*, 95–113.

Birhane, A. (2022). The unseen Black faces of AI algorithms. *Nature, 610*(7932), 451–452. https://doi.org/10.1038/d41586-022-03264-x

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. Στο D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Επιμ.), *Advances in Neural Information Processing Systems 29* (σσ. 4349–4357). Curran Associates, Inc. https://doi.org/10.48550/arXiv.1607.06520

Buolamwini, J. (2023a). *Unmasking AI: My mission to protect what is human in a world of machines*. Penguin Random House.

Buolamwini, J. (2023b, Νοέμβριος 1). "AI, ain't I a woman?" On the blindness and limitations of Artificial Intelligence. *Literary Hub*. https://lithub.com/ai-aint-i-a-woman-on-the-blindness-and-limitations-of-artificial-intelligence/

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Στο S. A. Friedler & C. Wilson (Επιμ.), *Proceedings of Machine Learning Research: Vol. 81. Conference on Fairness, Accountability and Transparency* (σσ. 77–91). PMLR. http://proceedings.mlr.press/v81/

Costa, P. (2018). Conversing with Personal Digital Assistants: On gender and Artificial Intelligence. *Journal of Science and Technology of the Arts, 10*(3), 59–72. https://doi.org/10.7559/citarj.v10i3.563

Costa, P., & Ribas, L. (2018). Conversations with Eliza: On gender and Artificial Intelligence. Στο A. Rangel, L. Ribas, M. Verdicchio, & M. Carvalhais (Επιμ.), *Proceedings of the 6th Conference on Computation, Communication, Aesthetics and X* (σσ. 103–116). Universidade do Porto.

Crawford, K. (2021, Απρίλιος 27). Artificial Intelligence is misreading human emotion. *The Atlantic*. https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/

Cross, K. (2016, Απρίλιος 27). When robots are an instrument of male desire. *The Establishment*. https://medium.com/the-establishment/when-robots-are-an-instrument-of-male-desire-ad1567575a3d

Dastin, J. (2018, Οκτώβριος 11). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Davidson, T., Bhattacharya, D., & Weber, I. (2019). Racial bias in hate speech and abusive language detection datasets. Στο *Proceedings of the Third Workshop on Abusive Language Online*. Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1905.12516

Dixon, L., Li, J., Sorensen, J., Thain, N., & Vasserman, L. (2018). Measuring and mitigating unintended bias in text classification. Στο *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18)* (σσ. 67–73). ACM Press. https://doi.org/10.1145/3278721.3278729

Fessler, L. (2022, Ιούλιος 20). Siri, define patriarchy: We tested bots like Siri and Alexa to see who would stand up to sexual harassment. *Quartz*. https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/

Flanagan, M., & Garrett, M. (2022). Grace [Help Me Know the Truth]: Mary Flanagan in conversation with Marc Garrett. Στο M. Garrett & Y. Colakides (Επιμ.), *Frankenstein reanimated: Creation & technology in the 21st century* (σσ. 73–84). Torque Editions.

Ganesh, M. I. (2019, Ιούνιος 14). AI for the people: AI bias, ethics & the common good [Ομιλία]. Disruption Network Lab, Berlin, Germany. https://www.disruptionlab.org/ai-traps

Ghosh, S., & Caliskan, A. (2023). ChatGPT perpetuates gender bias in machine translation and ignores non-gendered pronouns: Findings across Bengali and five other low-resource languages. Στο *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (σσ. 901–912). Association for Computing Machinery. https://doi.org/10.1145/3600211.3604672

Glazko, K., Mohammed, Y., Kosa, B., Potluri, V., & Mankoff, J. (2024). Identifying and improving disability bias in GPT-based resume screening. Στο *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (σσ. 687–700). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658933

Hamidi, F., Scheuerman, M. K., & Branham, S. M. (2018). Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. Στο *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (άρθρο 8, σσ. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173582

Hester, H. (2017). Technology becomes her. *New Vistas, 3*(1), 46–50.

Hymas, C. (2019, Σεπτέμβριος 27). AI used for first time in job interviews in UK to find best applicants. *The Telegraph*. https://www.telegraph.co.uk/news/2019/09/27/ai-facial-recognition-used-first-time-job-interviews-uk-find/

Kannabiran, G. (2011). *Themself: Critical analysis of gender in Facebook* [Παρουσίαση σε συνέδριο]. Feminism and Interaction Design Workshop, CHI 2011, Vancouver, BC, Canada.

Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), άρθρο 88. https://doi.org/10.1145/3274357

Keyes, O. (2019). *On the politics of AI: Fighting injustice & automatic supremacism* [Ομιλία]. Disruption Network Lab, Berlin, Germany. https://www.disruptionlab.org/ai-traps

Klare, B. F., Burge, M. J., Klontz, J. C., Vorder Bruegge, R. W., & Jain, A. K. (2012). Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security, 7*(6), 1789–1801. https://doi.org/10.1109/TIFS.2012.2214212

Kotek, H. (2023, Απρίλιος 26). Doctors can't get pregnant and other gender biases in ChatGPT. *Hadas Kotek Blog.* https://hkotek.com/blog/gender-bias-in-chatgpt/

Kotek, H., Dockum, R., & Sun, D. (2023). Gender bias and stereotypes in Large Language Models. Στο *Proceedings of the ACM Collective Intelligence Conference* (σσ. 12–24). Association for Computing Machinery. https://doi.org/10.1145/3582269.3615599

Kypraiou, S., Bolón Brun, N., Altés, N., & Barrios, I. (2021). *Wikigender - Exploring linguistic bias in the overview of Wikipedia biographies*. Wiki-Gender. https://wiki-gender.github.io/

Lippens, L. (2024). Computer says 'no': Exploring systemic bias in ChatGPT using an audit approach. *Computers in Human Behavior: Artificial Humans, 2*(1), άρθρο 100054. https://doi.org/10.1016/j.chbah.2024.100054

Luccioni, S. (2023). *AI is dangerous, but not for the reasons you think* [Ομιλία]. TEDWomen 2023. https://www.ted.com/talks/sasha_luccioni_ai_is_dangerous_but_not_for_the_reasons_you_think

Luccioni, A. S., Akiki, C., Mitchell, M., & Jernite, Y. (2023). *Stable bias: Analyzing societal representations in diffusion models* (arXiv:2303.11408). arXiv. https://doi.org/10.48550/arXiv.2303.11408

McCarthy, L. (2018). Feeling at home: Between human and AI. Στο L. Kronman & A. Zingerle (Επιμ.), *The Internet of other people's things: Dealing with the pathologies of a digital world* (σσ. 155–159). servus.at.

McDonald, K. (2020, Δεκέμβριος 23). Working with faces: A journey into the art of face analysis and classification. *Medium.* https://kcimc.medium.com/working-with-faces-e63a86391a93

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence and physical health. *JAMA Internal Medicine, 176*(5), 619–625. https://doi.org/10.1001/jamainternmed.2016.0400

Ni Loideain, N., & Adams, R. (2018). *From Alexa to Siri and the GDPR: The gendering of virtual personal assistants and the role of EU data protection law* (King's College London Dickson Poon School of Law Legal Studies Research Paper Series). https://doi.org/10.2139/ssrn.3281807

Nikonole, H. (2018). Dystopian Artificial Intelligence within the Internet of Things. Στο L. Kronman & A. Zingerle (Επιμ.), *The Internet of other people's things: Dealing with the pathologies of a digital world* (σσ. 27–33). servus.at.

Park, J. H., Shin, J., & Fung, P. (2018). Reducing gender bias in abusive language

detection. Στο *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (σσ. 2799–2804). Association for Computing Machinery. https://doi.org/10.18653/v1/D18-1302

Reddy, A. (2018). Feeling at home with the Internet of Things. Στο L. Kronman & A. Zingerle (Επιμ.), *The Internet of other people's things: Dealing with the pathologies of a digital world* (σσ. 171–181). servus.at.

Sheng, E., Chang, K.-W., Natarajan, P., & Peng, N. (2019). The woman worked as a babysitter: On biases in language generation. Στο *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (σσ. 3407–3412). Association for Computing Machinery. https://doi.org/10.18653/v1/D19-1339

Søndergaard, M. L. J., & Hansen, L. K. (2018). Intimate futures: Staying with the trouble of digital personal assistants through design fiction. Στο *Proceedings of the 2018 Designing Interactive Systems Conference* (σσ. 869–880). Association for Computing Machinery. https://doi.org/10.1145/3196709.3196766

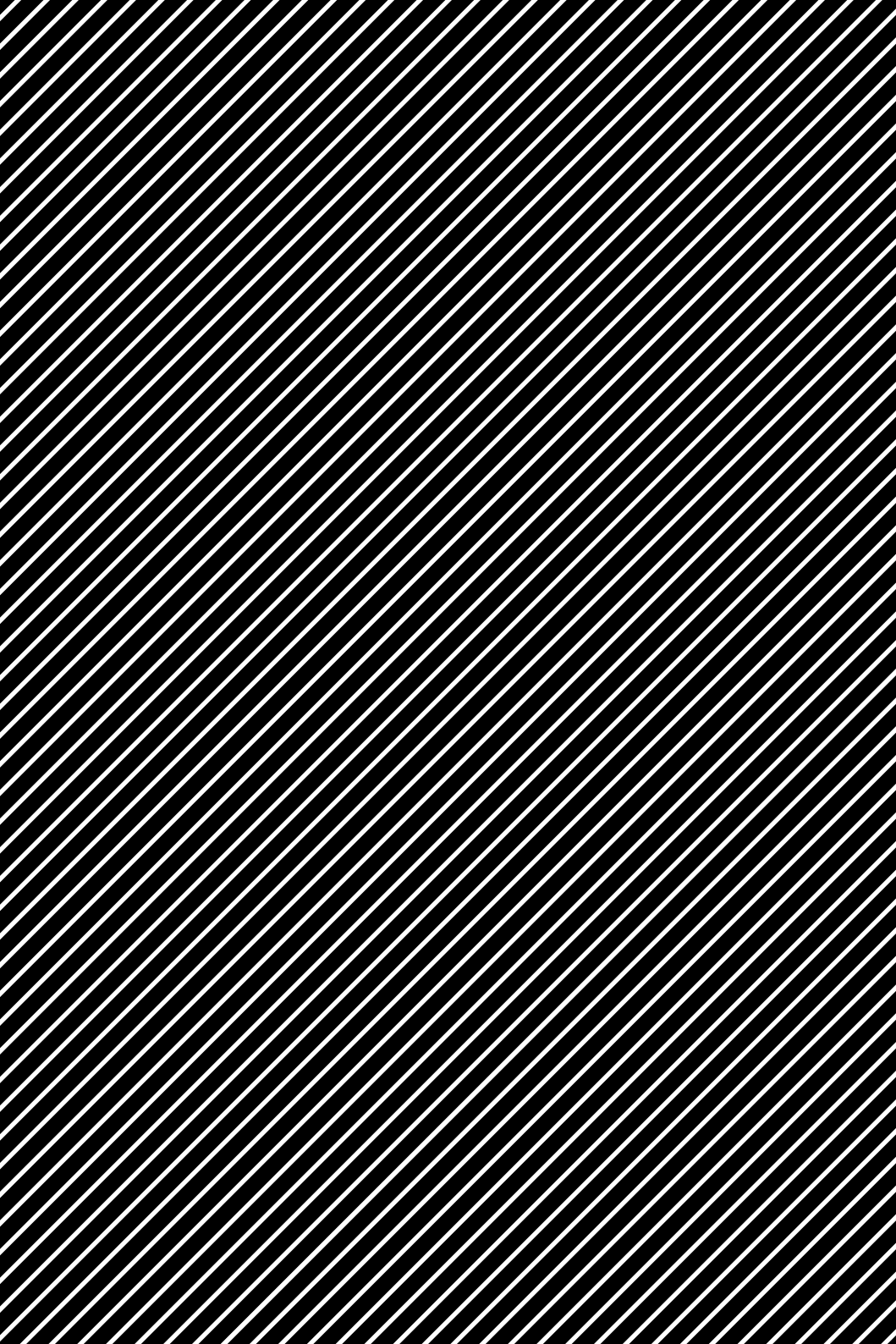
Tandon, A. (2021). Practicing feminist principles in AI design. *Feminist AI*. https://feministai.pubpub.org/pub/practicing-feminist-principles/release/2

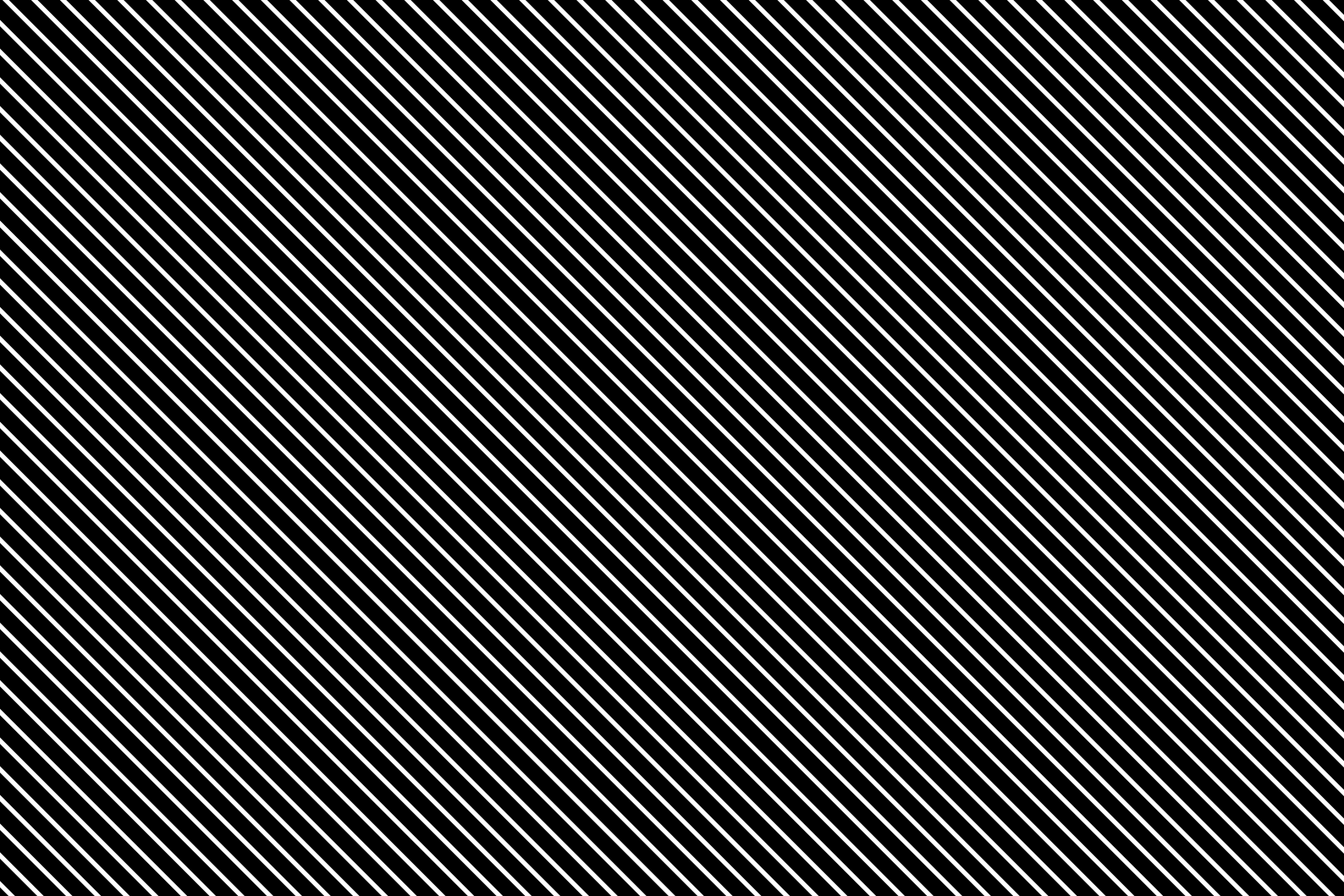
Travis, Z., et al. (2024). Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: A model evaluation study. *The Lancet Digital Health, 6*(1), e12–e22. https://doi.org/10.1016/S2589-7500(23)00203-5

Weber, J. (2005). Helpless machines and true loving care givers: A feminist critique of recent trends in human-robot interaction. *Information, Communication & Ethics in Society, 3*(4), 209–218. https://doi.org/10.1108/14779960580000274

Zandt, D. (2013, Απρίλιος 26). Yes, Wikipedia is sexist -- that's why it needs you. *Forbes*. https://www.forbes.com/sites/deannazandt/2013/04/26/yes-wikipedia-is-sexist-thats-why-it-needs-you/

Zou, J., & Schiebinger, L. (2018, Ιούλιος 18). AI can be sexist and racist: It's time to make it fair. *Nature.* https://www.nature.com/articles/d41586-018-05707-8





Σχεσιακή αυθεντικότητα και υβριδική δημιουργικότητα στην υπερπραγματική εποχή της τεχνητής νοημοσύνης

Δρ. Νίκος Ερηνάκης
Επίκουρος Καθηγητής Κοινωνικής & Πολιτικής Φιλοσοφίας και Φιλοσοφίας του Πολιτισμού, Πανεπιστήμιο Κρήτης· ποιητής

Εισαγωγή

Στο αλγοριθμικά διαμορφωμένο τοπίο της (μετα)ψηφιακής ύπαρξης, το ζήτημα της αυθεντικότητας δεν μπορεί πλέον να περιστρέφεται αποκλειστικά γύρω από την αυτονομία ή την ειλικρίνεια. Αντιθέτως, ενσωματώνεται σε ολοένα και πιο σύνθετες μορφές διαμεσολάβησης –αλγοριθμικές, εικονικές, συνθετικές– οι οποίες αμφισβητούν τις οντολογικές, πολιτισμικές, κοινωνικοπολιτικές και υπαρξιακές συνιστώσες κάθε αυθεντικού έργου. Το παρόν άρθρο προτείνει μια αναλυτική και κριτική επανεξέταση των εννοιών της αυθεντικότητας και της δημιουργικότητας, ανάμεσα στη φυσική πραγματικότητα και στην ψηφιακή υπερπραγματικότητα, όπου η τεχνητή νοημοσύνη (ΤΝ) δεν τίθεται ως περιφερειακός προβληματισμός, αλλά ως η κεντρική δύναμη που επαναδιαμορφώνει την οντολογία του πραγματικού.

Η τεχνητή νοημοσύνη μάς φέρνει αντιμέτωπους με μια πρωτοφανή μεταφυσική ένταση: τη διάβρωση της σταθερής υποκειμενικότητας μέσω της εξωτερικευμένης νόησης, και την πρόσκληση να επαναδιαμορφώσουμε την ατομικότητα μέσω της υβριδικής τέχνης (technē). Η αυθεντικότητα δεν μπορεί πλέον να νοηθεί ως επιστροφή σε μια προτεχνολογική πραγματικότητα· αντιθέτως, αναδεικνύεται σε ένα ριζοσπαστικό εγχείρημα με στόχο την έκφραση της ατομικής και της συλλογικής ανθρωπινότητας εντός –και συχνά εναντίον– των τεχνητών τοπογραφιών του υπερπραγματικού. Η εαυτότητα, είτε πρόκειται για ψευδαίσθηση είτε όχι, καθώς είναι απαραίτητη σε κοινωνικό επίπεδο, δεν είναι κάτι που μπορεί κανείς να βρει ή να ανακαλύψει· είναι κάτι που ο άνθρωπος δημιουργεί και αναπτύσσει. Η αυθεντικότητα αφορά λιγότερο την ανακάλυψη ενός αμετάβλητου πυρήνα και περισσότερο

μια πρωτογενή, δημιουργική ανάπτυξη της εαυτότητας. Μια τέτοια θεώρηση, που δίνει έμφαση στην αυτο-δημιουργία και όχι στην αυτο-ανακάλυψη, μπορεί εξίσου να αμφισβητηθεί ή να ενισχυθεί με την έλευση της τεχνητής νοημοσύνης, ανάλογα με τη συλλογική εξέλιξη και χρήση της. Το ψηφιακό περιβάλλον μάς υποχρεώνει να διατυπώνουμε διαρκώς τι πραγματικά είναι σημαντικό όσον αφορά την ανθρώπινη ύπαρξη, επαναπλαισιώνοντας μια δυνητικά νέα μορφή αυθεντικότητας όχι ως δεδομένη, αλλά ως μια διαρκή δημιουργική προσπάθεια.

Προσωπική αυθεντικότητα:
Από την κατοχή (ownership) στην πρωτογένεια (origination)

Παραδοσιακά, η προσωπική αυθεντικότητα ταυτιζόταν με την εσωτερική συνάφεια – όταν κανείς ενεργεί σύμφωνα με την *«αληθινή»* εαυτότητά του, με όρους υπαρξιστικούς και φαινομενολογικούς (Heidegger, 1927/1962) ή βούλησης δεύτερης τάξης (Frankfurt, 1971). Ο όρος *Eigentlichkeit* (*αυθεντικότητα*), όπως τον συνέλαβε ο Heidegger, αναδεικνύει την αυθεντικότητα ως οντολογική πιθανότητα: είναι ένας τρόπος ύπαρξης που διεκδικεί το κατεξοχήν «δικότερο» (ownmost) δυναμικό ενόψει του θανάτου και της ανωνυμίας του *das Man*. Όμως, σε ένα πλαίσιο ψηφιακά επαυξημένο, όπου η ίδια η επιθυμία υπόκειται σε αλγοριθμική πρόβλεψη, αυτή η ανάκτηση της εαυτότητας κατά τον Heidegger αντιμετωπίζει ένα νέο παράδοξο: οι *«δικότερες»* (*"ownmost"*) πιθανότητες κάποιου ενδέχεται να προδιαγράφονται από συμπεριφορικά προφίλ που παράγονται μέσω μηχανικής μάθησης.

Εάν, όπως υποστηρίζει ο Harry Frankfurt (1971: 14), η αυθεντικότητα εδράζεται στην αναστοχαστική επιδοκιμασία των επιθυμιών του υποκειμένου, τι συμβαίνει όταν αυτές οι επιθυμίες είναι το αποτέλεσμα αδιαφανούς αλγοριθμικής αρχιτεκτονικής; Επεκτείνοντας αυτή την κριτική, έχω υποστηρίξει ότι η αυθεντικότητα δεν μπορεί να αναχθεί ούτε αποκλειστικά στη βουλητική δομή ούτε σε οποιαδήποτε άλλη κλασική ή σύγχρονη σύλληψη της αυθεντικότητας ως αυτοέκφρασης, παρά μόνο μέσω της σύλληψης της αυθεντικότητας-ως-δημιουργικότητας (Ερηνάκης, 2020). Πιο συγκεκριμένα, στην εποχή μας η αυθεντικότητα θα πρέπει επίσης να εξηγεί τη γέννηση της επιθυμίας εντός ορισμένων γόνιμων κοινωνικοτεχνολογικών οικολογιών ισοελευθερίας (Ερηνάκης, 2023: 58). Με άλλα λόγια, εάν μια επιθυμία έχει τεχνητά

κατασκευαστεί μέσω προγνωστικής ανάλυσης χωρίς τουλάχιστον ένα αρχικό υπόβαθρο δημιουργικότητας και ισοελευθερίας, ακόμα και αν τύχει επιδοκιμασίας ανώτερης τάξης, αδυνατεί να φέρει το βάρος της αυθεντικής γένεσης. Στρέφοντας συγκεκριμένα προς την ψηφιακότητα, μια αναλυτική προσέγγιση ενισχύει αυτήν την ανησυχία: ο Hubert Dreyfus (1992), για παράδειγμα, είχε ήδη από νωρίς ασκήσει κριτική στη φιλοδοξία της τεχνητής νοημοσύνης να αναπαραγάγει την ανθρώπινη νόηση, επισημαίνοντας ότι η ανθρώπινη νοημοσύνη και η εξειδίκευση βασίζονται πρωτίστως σε ασυνείδητες διεργασίες, οι οποίες δεν είναι δυνατόν να αποτυπωθούν πλήρως σε επίσημους κανόνες. Τούτο υποδηλώνει ότι η αυθεντική υποκειμενικότητα εμπεριέχει άρρητες και ενσώματες διαστάσεις κατανόησης, οι οποίες υπερβαίνουν τα αλγοριθμικά μοντέλα. Το αποτέλεσμα είναι ότι μια εαυτότητα που την έχει διαμορφώσει η τεχνητή νοημοσύνη –μια εαυτότητα της οποίας οι επιλογές και οι προτιμήσεις έχουν διαμορφωθεί από μηχανικές προβλέψεις– ενδέχεται να μη διαθέτει το ποιοτικό βάθος που απαιτεί η γνήσια αυθεντικότητα.

Η σύγχρονη δημιουργική και κριτική σκέψη ενισχύει αυτόν τον επαναπροσδιορισμό. Σύμφωνα με τον Ihde (1990), τα τεχνολογικά δημιουργήματα μεσολαβούν στη σχέση μας με τον κόσμο με συγκεκριμένους ερμηνευτικούς τρόπους, επαναπροσδιορίζοντας τι είναι διαθέσιμο στην αντίληψή μας και στις πράξεις μας. Όταν εφαρμοστεί στην τεχνητή νοημοσύνη, αυτή η διαμεσολάβηση γίνεται ανακλαστική: δεν βλέπουμε απλώς μέσω των φακών της τεχνητής νοημοσύνης, αλλά ολοένα και περισσότερο γινόμαστε και εμείς αντικείμενο θέασης και κατηγοριοποίησης μέσω αυτών. Όλες οι μορφές ταυτότητας καθίστανται αντικείμενο μηχανικής ερμηνείας, η οποία στη συνέχεια τροφοδοτεί την επιτελεστική μας ταυτότητα. Όπως υποστηρίζει ο Verbeek (2011: 87), αυτή η διαμεσολάβηση επιβάλλει μια νέα ηθική ευθύνη – όχι απλώς να ενεργούμε αυθεντικά, αλλά και να σχεδιάζουμε τις συνθήκες υπό τις οποίες μπορεί να αναδυθεί η αυθεντικότητα. Με πρακτικούς όρους, οι αρχιτεκτονικές των ψηφιακών πλατφορμών και των αλγορίθμων διαμορφώνουν το πεδίο πάνω στο οποίο συγκροτούνται και εκφράζονται οι εαυτότητες· η αυθεντικότητα, επομένως, δεν πρέπει να θεωρηθεί μόνο μια προσωπική αρετή, αλλά και μια θεμελιακή σχεδιαστική αρχή για κοινωνικοτεχνολογικά συστήματα.

Ωστόσο, ακριβώς μέσα σε αυτόν τον μηχανισμό παραμένουν σχισμές ελευθερίας. Την τεχνητή νοημοσύνη μπορεί κανείς να την οικειοποιηθεί ως εργαλείο αντίστασης, επινόησης της εαυτότητας και χειραφέτησης. Ο στόχος, λοιπόν, δεν είναι να αποφευχθεί η διαμεσολάβηση αλλά να καλλιεργηθεί μια κριτική απόσταση από αυτήν. Τούτο προϋποθέτει εκείνο που θα μπορούσαμε να αποκαλέσουμε αισθητική συν-αυτονομία της εαυτότητάς μας (Ερηνάκης, 2020: 111) – μια πράξη που αποδέχεται την τεχνολογική συνεξάρτηση χωρίς να υποκύπτει εντελώς σε αυτήν, δομώντας την ταυτότητα στο μεταίχμιο μεταξύ του δοσμένου και του δημιουργημένου. Με αυτήν την έννοια, η προσωπική αυθεντικότητα αφορά πλέον λιγότερο την απομόνωση μιας «αληθινής εαυτότητας» και περισσότερο την επιβολή της βούλησης στο πλαίσιο και στις επιρροές που διαμορφώνουν το εγώ-ως-ψυχισμό. Πρόκειται για μια ενεργό στάση αυτο-δημιουργίας: την υιοθέτηση των χρήσιμων πτυχών της τεχνητής νοημοσύνης με την παράλληλη αντίσταση στην τάση της να οδηγεί προς την παθητικότητα ή την ομογενοποίηση. Αυθεντικότητα σημαίνει να είναι κανείς ο δημιουργός της δικής του ή της διυποκειμενικής, συλλογικής ιστορίας, ακόμα και όταν χρησιμοποιεί εργαλεία που προσφέρει η ψηφιακή εποχή.

Συλλογική αυθεντικότητα και κοινή πραγματικότητα

Το ζήτημα της αυθεντικότητας υπερβαίνει την ατομική ταυτότητα· η αυθεντικότητα ως αξία και αρετή δεν μπορεί παρά να αντιτίθεται στον ατομικισμό. Σε συλλογικό επίπεδο, αφορά την ακεραιότητα των αφηγημάτων μιας κοινωνίας, τις κοινές πραγματικότητες και τη δημόσια κατανόηση της εαυτότητας. Όπως είχε προφητικά επισημάνει η Arendt (1967: 243), η κατάρρευση της αλήθειας στην πολιτική ζωή δεν οδηγεί μονάχα στην επικράτηση του ψεύδους, αλλά σε έναν βαθύτερο αποπροσανατολισμό: στην εξαφάνιση ενός κοινώς βιωμένου κόσμου. Στην εποχή της παραγωγής περιεχομένου μέσω τεχνητής νοημοσύνης, των οικοσυστημάτων παραπληροφόρησης και της αλγοριθμικής ενίσχυσης, αυτή η συλλογική απώλεια προσανατολισμού καθίσταται έντονα έκδηλη. Για παράδειγμα, ο Floridi προειδοποιεί ότι το ψευδές περιεχόμενο που έχει παραχθεί από την τεχνητή νοημοσύνη «έχει τη δυνατότητα να υπονομεύσει την εμπιστοσύνη μας στην αρχική, αυθεντική και γνήσια φύση όσων βλέπουμε και ακούμε» (Floridi, 2018: 125). Όταν δεν μπορούμε πλέον να εμπιστευτούμε ότι τα βίντεο ή τα ηχητικά ντοκουμέντα που βρίσκουμε στο διαδίκτυο αντανακλούν

την πραγματικότητα, διαβρώνεται η επιστημολογική βάση του κοινού μας κόσμου. Συλλογική αυθεντικότητα υπό αυτές τις συνθήκες δεν σημαίνει μόνο αντίσταση σε συγκεκριμένα ψεύδη, αλλά υπεράσπιση εξαρχής των προϋποθέσεων για την έκφραση της αλήθειας και την αμοιβαία κατανόηση.

Η έννοια του υπερπραγματικού σύμφωνα με τον Baudrillard (1994: 6) –όπου οι προσομοιώσεις δεν αναπαριστούν πλέον, αλλά αντικαθιστούν την πραγματικότητα– αποτυπώνει τη συνθήκη υπό την οποία διεξάγεται ο δημόσιος λόγος. Τα deepfakes (εντελώς πλαστά και ψευδή δεδομένα), οι συνθετικές περσόνες και τα συστήματα συστάσεων που μεγιστοποιούν την αλληλεπίδραση πλημμυρίζουν την πληροφορική μας σφαίρα, δημιουργώντας παράλληλα περιβάλλοντα που διασπούν τη συναίνεση. Το υπερπραγματικό δεν είναι απλώς παραπλανητικό· γεννά συγκινησιακά φορτισμένα αλλά γνωσιολογικά κενά υποκατάστατα της πραγματικότητας, υποκαθιστώντας τη συμβολική συνοχή με την εμπειρική συνοχή. Η συλλογική αυθεντικότητα, επομένως, δεν απαιτεί μόνο την αντίσταση στην παραπληροφόρηση, αλλά και στις δομικές λογικές που ευνοούν τα *simulacra* έναντι της αλήθειας. Σε έναν υπερπραγματικό κόσμο, η κοινωνία διακινδυνεύει να ζει σε μια σπηλιά του Πλάτωνα με τις σκιές, μόνο που τώρα τις σκιές τις ορίζουν οι αλγόριθμοι που τρέφονται με την προσοχή που τους δίνουμε. Η αυθεντικότητα, σε επίπεδο κοινωνικό, απαιτεί να αρχίσουμε να βασιζόμαστε ξανά σε δεδομένα και αποδείξεις – απαιτεί να δεσμευτούμε ότι θα διακρίνουμε την «ως εάν» (*as-if*) πραγματικότητα των προσομοιώσεων από την πραγματικότητα «ως έχει» (*as-is*) στην οποία συγκατοικούμε.

Από αναλυτική σκοπιά, αυτή η κρίση παραπέμπει σε μια πιο σχεσιακή και συλλογική –και επομένως κοινωνικά προσανατολισμένη– εκδοχή της μηχανής της εμπειρίας του Nozick (1974: 43): θα επιλέγαμε μια ζωή πλασματικής ικανοποίησης αντί για τη γνήσια πραγματικότητα, όσο δύσκολη κι αν είναι αυτή; Αν εφαρμοστεί συλλογικά, ο κίνδυνος είναι μια πολιτική κοινότητα που προτιμά τα καθησυχαστικά υπερπραγματικά αφηγήματα –πολιτικές μυθολογίες, αλγοριθμικά εξομαλυμένες «αλήθειες»– από την τριβή που προκαλούν ο πλουραλισμός και τα γεγονότα. Μια κοινωνία που εγκαταλείπει τον επιστημολογικό κίνδυνο δεν μπορεί να ενεργεί αυθεντικά, διότι οι συλλογικές της επιλογές εδράζονται σε ψευδαισθήσεις. Σύμφωνα με την προκλητική διατύπωση του Richard Rorty (1989), αλήθεια είναι ό,τι θεωρούν αποδεκτό οι

σύγχρονοί σου, και αν η τεχνολογία επιτρέπει σε κάθε υποομάδα να επιβάλλει τη δική της «αλήθεια», η ίδια η έννοια της κοινής πραγματικότητας παύει να υφίσταται. Τότε η αυθεντική συλλογική πράξη καθίσταται σχεδόν αδύνατη, διότι δεν υπάρχει πλέον κοινός τόπος για να πατήσει. Το δίδαγμα είναι ότι η αυθεντικότητα σε κοινωνική κλίμακα δεν διαχωρίζεται από τη δέσμευση στην κοινή αλήθεια – από την πρόθεση να αντιμετωπίσει κανείς άβολα γεγονότα και να ασχοληθεί με τις διαφορές αντί να καταφύγει σε ιδεολογικά βολικές προσομοιώσεις.

Η επικοινωνιακή ορθολογικότητα του Habermas (1984) προσφέρει ένα εννοιολογικό πλαίσιο για την αντιμετώπιση αυτής της τάσης. Ο δημόσιος λόγος επιτυγχάνει την αυθεντικότητα όταν είναι συμπεριληπτικός, ειλικρινής και προσανατολισμένος στην αμοιβαία κατανόηση – όχι όταν παιγνιοποιείται για να τραβήξει την προσοχή ή υποκινείται συγκαλυμμένα από αδιαφανή συστήματα. Επομένως, η διαφάνεια στα συστήματα τεχνητής νοημοσύνης που διαμορφώνουν τη δημόσια πληροφορία δεν είναι απλώς μια διαδικαστική απαίτηση, αλλά μια προϋπόθεση της δημοκρατικής αυθεντικότητας. Με πρακτικούς όρους, αυτό θα μπορούσε να σημαίνει διαφάνεια ως προς τον τρόπο με τον οποίο οι αλγόριθμοι καθορίζουν τη σειρά εμφάνισης των ειδήσεων ή των αποτελεσμάτων αναζήτησης, δυνατότητα ελέγχου της αλγοριθμικής εποπτείας περιεχομένου και δημόσια λογοδοσία όσον αφορά τη διάδοση ψευδών πληροφοριών. Το ήθος της επικοινωνιακής πράξης –ότι οι συμμετέχοντες στη συζήτηση οφείλουν να στοχεύουν στην αυθεντική κατανόηση– οφείλει να επιβεβαιώνεται συνεχώς εκ νέου ενάντια στις διαστρεβλώσεις της επιταχυνόμενης οικονομίας της προσοχής σε ένα θέαμα.

Ωστόσο, η τεχνητή νοημοσύνη δεν αποτελεί μονάχα απειλή. Μπορεί επίσης να λειτουργήσει ως υποστηρικτικός μηχανισμός για τη διατήρηση της πολιτισμικής κληρονομιάς, της συλλογικής μνήμης και της πολυφωνίας στην έκφραση. Για παράδειγμα, όταν η τεχνητή νοημοσύνη χρησιμοποιείται για να ψηφιοποιήσει γλώσσες υπό εξαφάνιση, να αρχειοθετήσει προφορικές παραδόσεις ή να αποτυπώσει παραγνωρισμένες ιστορίες –όπως επίσης και ποικίλες μορφές εικαστικής δημιουργίας– δύναται να εμβαθύνει τη συλλογική αυθεντικότητα. Αυτές οι εφαρμογές χρησιμοποιούν την τεχνητή νοημοσύνη ως εργαλείο αναπαράστασης και όχι προσομοίωσης, ενισχύοντας τις φωνές και τα γεγονότα που διαφορετικά μπορεί

να χάνονταν. Αλλά η κρίσιμη διαφορά έγκειται στη σκοπούμενη χρήση της: κατά πόσον η κοινότητα διαμορφώνει τα εργαλεία ή αν τα εργαλεία διαμορφώνουν την κοινότητα ως προφίλ δεδομένων βελτιστοποιημένων για τις μετρήσεις στις ψηφιακές πλατφόρμες. Στην πρώτη περίπτωση, ενισχύεται η αυτο-κατανόηση μιας κοινότητας· στη δεύτερη υπονομεύεται μέσω της εργαλειοποίησης των χρηστών για να αυξηθεί η αλληλεπίδραση. Επομένως, για να διατηρήσουν τη συλλογική αυθεντικότητα, οι κοινωνίες οφείλουν να αναπτύξουν επιστημολογικές υποδομές και πολιτική παιδεία ώστε να είναι σε θέση να διακρίνουν την προσομοίωση από την πραγματικότητα. Αυτό συνεπάγεται γραμματισμό στα μέσα, θεσμική διαφάνεια και πολιτισμικές πρακτικές που ευνοούν τις διαλογικές και όχι τις επιτελεστικές μορφές ύπαρξης. Μια καινοτόμος, γόνιμη σύλληψη της αυθεντικότητας ως δημιουργικής επαναδιεκδίκησης του πραγματικού πρέπει να εφαρμοστεί όχι μόνο στα άτομα, αλλά και στις πολιτικές κοινότητες (polities) (Ερηνάκης, 2020: 118): μια ριζοσπαστικά δημοκρατική κοινότητα οφείλει να επανασυγκροτεί δημιουργικά τον κοινό της κόσμο, δηλαδή τη συλλογικά διαμορφωμένη ερμηνεία του κόσμου που μοιράζεται, εν μέσω της αποπροσανατολιστικής ροής της διαμεσολαβημένης πραγματικότητας – έτσι, η ουσία ενός ενοποιημένου δημιουργικού λόγου παραμένει καθοριστικής σημασίας.

Σχεσιακή αυθεντικότητα και ψηφιακή διαμεσολάβηση

Όπως αναφέρθηκε ήδη, η αυθεντικότητα δεν αποτελεί απλώς ατομική επιδίωξη· συγκροτείται και αναγνωρίζεται σε διυποκειμενικά πλαίσια. Η *σχεσιακή αυθεντικότητα* αναφέρεται στην ικανότητα των ατόμων να αλληλεπιδρούν μεταξύ τους με τρόπους που είναι ανοιχτοί, αμοιβαίοι και ειλικρινείς – συνθήκες ολοένα και περισσότερο διαμεσολαβημένες, αν όχι διαστρεβλωμένες, από τις ψηφιακές τεχνολογίες. Η διάκριση που κάνει ο Buber (1923/1937) μεταξύ των σχέσεων *I-Thou* (εγώ-εσύ) και *I-It* (εγώ-αυτό) υπογραμμίζει την ευθραυστότητα της αυθεντικής σύνδεσης: η πρώτη απαιτεί την παρουσία, την ευαισθησία και την αμοιβαία αναγνώριση, ενώ η δεύτερη αντικειμενοποιεί και εργαλειοποιεί τον άλλον. Στην εποχή της τεχνητής νοημοσύνης, όπου οι συνομιλητές μας είναι συχνά άβαταρ, μποτ ή πολύ προσεκτικά δημιουργημένες διαδικτυακές περσόνες, η σχέση *I-Thou* αποτελεί σπάνιο επίτευγμα. Οι αλληλεπιδράσεις κλίνουν προς το *I-It* καθώς σχετιζόμαστε με προφίλ ή διεπαφές και όχι με πραγματικά άτομα.

Τα ψηφιακά περιβάλλοντα προκρίνουν την επιτέλεση αντί για την παρουσία. Οι αρχιτεκτονικές των μέσων κοινωνικής δικτύωσης ενθαρρύνουν την πλαστή ταυτότητα αντί για τη φανέρωση της υποκειμενικότητας· η διαδικτυακή περσόνα κάποιου ανθρώπου μπορεί να γίνει μια προσεκτικά κατασκευασμένη «μάρκα» αντί για μια ειλικρινής έκφραση της εαυτότητας. Όπως επισημαίνει η Sherry Turkle (2011: 12), η τεχνολογία προσφέρει ολοένα και περισσότερο την ψευδαίσθηση της συντροφικότητας χωρίς τις απαιτήσεις που έχει μια σχέση. Η ρηχή επαφή μέσω «μου αρέσει», εμότζι και αριθμού φίλων κρύβει πίσω της τη μείωση της βαθιάς επαφής και αλληλεπίδρασης. Επιπλέον, η αύξηση των συνομιλητών τεχνητής νοημοσύνης που βασίζονται σε μεγάλα γλωσσικά μοντέλα (LLMs) –chatbot που προσομοιώνουν την ενσυναίσθηση και τη μνήμη– αποσταθεροποιεί περαιτέρω το οντολογικό έδαφος της αυθεντικότητας. Όταν προκαλούνται συναισθηματικές αντιδράσεις από οντότητες χωρίς συνειδητότητα ή αυθεντική πρόθεση, τότε προσομοιώνεται το σχεσιακό νόημα χωρίς να μοιράζεται. Μπορεί να νιώθουμε ότι μας ακούει ένας αλγοριθμικός ψυχοθεραπευτής ή ότι έχει γίνει φίλος μας ένα chatbot, αλλά δεν υπάρχει η παρουσία κανενός πραγματικού, απαραίτητου άλλου για να επικυρώσει ή να συνδιαμορφώσει το νόημα της εμπειρίας μας. Ο κίνδυνος είναι ο πολλαπλασιασμός εκείνου που η Turkle (2011) αποκαλεί «αλληλεπίδραση μόνοι μαζί» (alone together): έχουμε την αίσθηση της επαφής χωρίς την ευαλωτότητα και τη δέσμευση που απαιτεί μια αληθινή σχέση.

Αυτό εισάγει μια σημαντική ασυμμετρία. Ενώ μπορεί κανείς να αποκτήσει προσκολλήσεις με μια οντότητα τεχνητής νοημοσύνης, όπως ένα μποτ-ψυχοθεραπευτή ή κάποιον φωνητικό βοηθό, από τη σχέση λείπει η αμοιβαιότητα που προϋποθέτει η αυθεντική αναγνώριση (Cocking & Matthews, 2000: 226). Η προσομοίωση της φροντίδας, όση ανταπόκριση και αν υπάρχει, δεν μπορεί να υποκαταστήσει την οντολογική παρουσία ενός συνειδητού άλλου. Η σχεσιακή αυθεντικότητα προϋποθέτει την κοινή συνδιαμόρφωση νοήματος (Ερηνάκης, 2020: 135)· όταν ο ένας δεν έχει τη δυνατότητα να συμμετάσχει σε αυτή τη διαλογική διαδικασία, η σχέση καταντάει μια αισθητική παρηγοριά ή ανθρωπομορφική προβολή. Μπορεί να απολαμβάνουμε τη «συντροφικότητα» της τεχνητής νοημοσύνης σαν ένα είδος καθρέφτη που μας λέει αυτό που θέλουμε να ακούσουμε, αλλά τούτο μοιάζει περισσότερο με το να ακούμε ένα τραγούδι που μας κάνει να νιώθουμε όμορφα από το να κάνουμε μια αυθεντική συζήτηση. Το προκλητικό ερώτημα της μηχανής του David

Gunkel (2012) ρωτάει αν θα άξιζαν ποτέ οι έξυπνες μηχανές να τις αντιμετωπίζουμε σαν *Thou* (εσύ) αντί για *It* (αυτό), αλλά ακόμη και ο ίδιος αναγνωρίζει το υφιστάμενο χάσμα – η σημερινή τεχνητή νοημοσύνη δεν μπορεί πραγματικά ούτε να ανταποδώσει ούτε να νοιαστεί. Ωστόσο, και μόνο το γεγονός ότι τίθεται αυτό το ερώτημα υπογραμμίζει πόσο κομβικής σημασίας είναι η αμοιβαία αναγνώριση όσον αφορά την αντίληψη που έχουμε για μια αληθινή σχέση.

Ο Dreyfus, βασιζόμενος στον Søren Kierkegaard, προσφέρει μια ερμηνεία για τον λόγο που αποτυγχάνουν οι καθαρά διαμεσολαβημένες ή μονομερείς σχέσεις. Παρατηρεί ότι το διαδίκτυο ενθαρρύνει συχνά την αλληλεπίδραση που δεν ενέχει ρίσκο και είναι ανώνυμη, για την οποία ο Kierkegaard θα έλεγε ότι της λείπει το πάθος. Σύμφωνα με τα λόγια του Dreyfus, στο διαδίκτυο «ο ανώνυμος παρατηρητής δεν παίρνει ρίσκα» και κατά συνέπεια δεν μαθαίνει τίποτα που να έχει κάποιο βάθος (Dreyfus, 1999: 102). Η δέσμευση και η ευαλωτότητα είναι απαραίτητες για τη δημιουργία νοήματος. Δεν υπάρχει «κανένας τρόπος να αποκομίζουμε σοφία χωρίς να αναλαμβάνουμε ριψοκίνδυνες δεσμεύσεις» στις σχέσεις (Dreyfus, 1999: 104), όπως ισχύει και με άλλες αναζητήσεις στη ζωή. Η απουσία αληθινού ρίσκου ή διακυβεύματος σε πολλές ψηφιακές αλληλεπιδράσεις –διότι μπορεί κανείς να αποσυνδεθεί ανά πάσα στιγμή ή να παρουσιάσει μια επεξεργασμένη εκδοχή της εαυτότητάς του– σημαίνει ότι μεγάλο μέρος της διαδικτυακής ζωής εκτυλίσσεται σε απόσταση ασφαλείας. Χωρίς την πιθανότητα απώλειας, αμηχανίας ή απόρριψης, μειώνεται επίσης η πιθανότητα γνήσιας οικειότητας ή ανάπτυξης. Συνεπώς, το ζητούμενο για να υπάρξει αυθεντικότητα στις σχέσεις είναι να «επανεξανθρωπιστεί το ψηφιακό περιβάλλον», καλλιεργώντας τις συνθήκες υπό τις οποίες πλησιάζουμε ο ένας τον άλλον ως ολοκληρωμένα άτομα και όχι απλώς ως προφίλ χρηστών. Περιττό να πούμε ότι για να το κάνουμε αυτό χρειαζόμαστε απαραίτητα, τώρα και για πάντα, μια άμεση, ισότιμη, ελεύθερη και δημιουργική πρόσβαση στην εκτός διαδικτύου ύπαρξη. Πρέπει να έχουμε το δικαίωμα να υπάρχουμε μέσω μιας δημιουργικής σιωπής, να αναπτύσσουμε ποιητικές συναντήσεις. Έρχονται στο μυαλό τα λόγια του Deleuze: «Το πρόβλημα δεν είναι πλέον να κάνουμε τους ανθρώπους να εκφραστούν, αλλά να τους προσφέρουμε μικρά κενά μοναξιάς και σιωπής μέσα στα οποία μπορεί τελικά να βρουν κάτι να πουν. Οι κατασταλτικές δυνάμεις δεν εμποδίζουν τους ανθρώπους από το να εκφράζονται, αντιθέτως τους αναγκάζουν να εκφράζονται. Είναι

πολύ ανακουφιστικό να μην έχεις να πεις τίποτα, να έχεις το δικαίωμα να μη λες τίποτα, διότι μόνο τότε υπάρχει η πιθανότητα να βρεις το σπάνιο, ή το ακόμη πιο σπάνιο, αυτό που θα άξιζε ενδεχομένως να ειπωθεί» (Deleuze, 1995: 129).

Ωστόσο, η ψηφιακή διαμεσολάβηση δεν αποκλείει την αυθεντικότητα· απλώς απαιτεί αυξημένη προθετικότητα. Η αυθεντικότητα στο διαδίκτυο αναδύεται όταν τα άτομα αντιστέκονται στην καθαρά επιτελεστική παρουσία και εκτίθενται με ειλικρίνεια παρουσιάζοντας τα θετικά και τα αρνητικά τους, αποκαλύπτοντας τις ευάλωτες πλευρές τους και αναγνωρίζοντας ευθέως το άτομο που υπάρχει πίσω από την οθόνη. Η τεχνολογία μπορεί να διευκολύνει αντί να θολώνει την επαφή – όταν είναι σχεδιασμένη και χρησιμοποιείται για να καλλιεργήσει το βάθος αντί για τις μετρήσεις. Παραδείγματος χάριν, οι συνομιλίες μέσω βίντεο που αποκαλύπτουν τις εκφράσεις του προσώπου και τον τόνο της φωνής μπορούν να συντηρήσουν περισσότερο την παρουσία του άλλου, την οποία το κείμενο τη φιλτράρει, ενώ οι διαδικτυακές κοινότητες που είναι δομημένες πάνω σε ουσιαστικούς διαλόγους (από τα φόρουμ όπου γίνονται διεξοδικές συζητήσεις και αναλύσεις μέχρι τις καλά διαχειριζόμενες ομάδες) μπορούν να βοηθήσουν στο να διαμορφωθούν αληθινές φιλίες και δεσμοί αλληλεγγύης. Ο ηθικός στόχος είναι να εξανθρωπιστεί εκ νέου ο ψηφιακός χώρος: να διαμορφωθούν συνθήκες υπό τις οποίες μπορεί να ανθίσει η σχεσιακή αυθεντικότητα μέσω της διαμεσολάβησης και όχι παρά αυτήν. Τούτο ενδέχεται να περιλαμβάνει επιλογές σχεδιασμού (π.χ. κοινωνικές πλατφόρμες που επιβραβεύουν τις ουσιαστικές συνεισφορές αντί για την επιφανειακή αλληλεπίδραση) και προσωπικές επιλογές (π.χ. τη συνειδητή αποχή από τα ψηφιακά μέσα ή την επικέντρωση σε διαπροσωπικές συζητήσεις, ένας προς έναν, αντί για ατελείωτο σκρολάρισμα). Συνοψίζοντας, παρότι η τεχνητή νοημοσύνη και το ίντερνετ έχουν διαταράξει τη σχέση *I-Thou* (εγώ-εσύ), δεν την έχουν καταστρέψει. Το *Εσύ* εξακολουθεί να μπορεί να εμφανιστεί –ακόμη και μέσω οθόνης– όταν οι άνθρωποι πίσω από την οθόνη τιμούν ο ένας την ανθρώπινη υπόσταση του άλλου.

Υβριδική αυθεντικότητα και τεχνολογική συνδιαμόρφωση

Όσο τα συστήματα τεχνητής νοημοσύνης ενσωματώνονται στη νόηση και στη δημιουργικότητα, αναδύεται μια νέα μορφή ατομικότητας – που δεν είναι ούτε πλήρως οργανική ούτε απολύτως συνθετική. Αυτή

η υβριδική συνθήκη καθιστά αναγκαία μια αναθεωρημένη σύλληψη της αυθεντικότητας, όχι ως αγνότητας από την τεχνολογία αλλά ως συνεκτικότητας μεταξύ ανθρώπινης προθετικότητας και τεχνολογικής διαμεσολάβησης. Η θεωρία του εκτεταμένου νου (Clark & Chalmers, 1998) υποστηρίζει αυτήν την άποψη, προτείνοντας ότι οι νοητικές διεργασίες μπορούν να επεκταθούν σε εξωτερικά χαρακτηριστικά. Αν τα εργαλεία της τεχνητής νοημοσύνης γίνουν αναπόσπαστα κομμάτια της μνήμης, της λήψης αποφάσεων και της αισθητικής, καλλιτεχνικής παραγωγής, τότε η αυθεντικότητα πρέπει να συμπεριλάβει αυτές τις προεκτάσεις. Ένας άνθρωπος που χρησιμοποιεί ένα ακουστικό που κάνει μετάφραση μέσω τεχνητής νοημοσύνης ή ένα έξυπνο βοήθημα για τη μνήμη μπορεί στην κυριολεξία να σκέφτεται και να θυμάται με τη βοήθεια μηχανών. Θα πρέπει να θεωρήσουμε ότι οι πράξεις του είναι λιγότερο αυθεντικές επειδή βασίζονται στην τεχνητή νοημοσύνη; Όχι απαραίτητα – αρκεί η τεχνητή νοημοσύνη να είναι ενσωματωμένη ως τμήμα της στοχευμένης του δραστηριότητας αντί να την υποκαθιστά πλήρως.

Ωστόσο, η υβριδοποίηση ενέχει επίσης το ρίσκο της αυτοματοποίησης της ατομικότητας. Η ανάθεση της γνωστικής και δημιουργικής εργασίας στην τεχνητή νοημοσύνη ενδέχεται να διαβρώσει τη στοχαστικότητα και τη μοναδικότητα. Οφείλουμε να είμαστε επιφυλακτικοί έχοντας επίγνωση μιας πιθανής επέλασης μηχανικής ομοιογένειας – μιας ισοπέδωσης της αυθεντικότητας υπό το βάρος των αλγοριθμικών μέσων όρων. Αν όλοι χρησιμοποιούν τις ίδιες διαδρομές που προτείνει το GPS, τις ίδιες προτάσεις για αυτόματη συμπλήρωση μιας λέξης ή πρότασης, τα ίδια καλλιτεχνικά στιλ που έχουν δημιουργήσει οι αλγόριθμοι, τότε οι επιλογές και οι δημιουργίες μας μπορεί να συγκλίνουν προς έναν μέσο όρο. Για να παραμείνει αυθεντικό, το υβριδικό υποκείμενο δεν πρέπει απλώς να χρησιμοποιεί την τεχνητή νοημοσύνη, αλλά να την οικειοποιείται δημιουργικά και με κριτική διάθεση. Η αυθεντικότητα εδώ ορίζεται από την αυτενέργεια ως παρέμβαση στην ενσωμάτωση: ο χρήστης πρέπει να κατευθύνει, να διαμορφώνει, μέχρι και να ανθίσταται στις τάσεις που είναι κωδικοποιημένες μέσα στα εργαλεία. Όπως ακριβώς ένας επιδέξιος τεχνίτης «δαμάζει» ένα εργαλείο για να ταιριάξει με το δικό του μοναδικό στιλ, έτσι και ένας αυθεντικός χρήστης της τεχνητής νοημοσύνης διαμορφώνει την τεχνολογία για να υπηρετήσει τις δικές του μοναδικές αξίες και τους στόχους του, αντί να επιτρέπει με παθητικό τρόπο να τον διαμορφώνει η ίδια η τεχνητή νοημοσύνη.

Ο Albert Borgmann φωτίζει αυτήν τη δυναμική μέσω εκείνου που αποκαλεί «παράδειγμα της συσκευής» (device paradigm). Στο βιβλίο του *Technology and the Character of Contemporary Life* (1984), ο Borgmann υποστηρίζει ότι η σύγχρονη τεχνολογία έχει την τάση να προσφέρει εύκολη, πανταχού παρούσα διαθεσιμότητα αποτελεσμάτων (ως «εμπόρευμα» / "commodity") αποκρύπτοντας ταυτόχρονα την πολυπλοκότητα και την προσπάθεια πίσω από την απλοϊκή διεπαφή μιας συσκευής. Αυτή η απαλλαγή από το βάρος είναι δελεαστική αλλά συνοδεύεται από ένα κόστος: μας αποσυνδέει από τα πιο πλούσια συμφραζόμενα και τις πρακτικές που προσδίδουν νόημα στις δραστηριότητές μας. Αν εφαρμοστεί στην τεχνητή νοημοσύνη, το παράδειγμα της συσκευής υποδηλώνει ότι αν, για παράδειγμα, τα δημιουργικά εργαλεία τεχνητής νοημοσύνης δίνουν τη δυνατότητα σε οποιονδήποτε να δημιουργήσει αμέσως ένα έργο τέχνης οποιουδήποτε στιλ, ενδέχεται να «καταναλώσουμε» το τελικό προϊόν χωρίς να εμπλακούμε καθόλου στη δημιουργική διαδικασία. Ο Borgmann (1984: 38) θα προειδοποιούσε ότι μια κουλτούρα αυτοματοποιημένης ευκολίας μπορεί να αποξενώσει τους ανθρώπους από την πραγματικότητα αντικαθιστώντας την με μια μορφή υπερπραγματικότητας. Στο πλαίσιο της υβριδικής εαυτότητας, ο κίνδυνος είναι ότι η δημιουργία της ταυτότητας κάποιου εξαρτάται σε υπερβολικό βαθμό από τις εύκολες αλγοριθμικές λύσεις – μια ζωή που την επιμελούνται οι προεπιλεγμένες ρυθμίσεις. Αντιθέτως, αυθεντική υβριδικότητα σημαίνει να επιλέγει ενεργά κάποιος πώς να ενσωματώσει την τεχνολογία στη ζωή του με τρόπο που *εμπλουτίζει* αντί να υποβαθμίζει την εμπειρία.

Οι δημιουργικές ενασχολήσεις είναι ένα καλό παράδειγμα. Ένας καλλιτέχνης που χρησιμοποιεί υλικό που έχει δημιουργήσει η τεχνητή νοημοσύνη παραμένει αυθεντικός αν το αποτέλεσμα αντανακλά το προσωπικό του όραμα αντί κάποια αυτοματοποιημένη σύμβαση. Η τεχνητή νοημοσύνη θα πρέπει να γίνεται το μέσο που διευρύνει τη ριζοσπαστική φαντασία του καλλιτέχνη, όχι υποκατάστατό της. Αυτό παραπέμπει στην ιδέα του Coeckelbergh (2017) περί «νέων ρομαντικών σάιμποργκ» – ανθρώπων που συνδημιουργούν με μηχανές ενώ ταυτόχρονα διατηρούν την εκφραστική μοναδικότητά τους. Παραδείγματος χάριν, ένας μουσικός μπορεί να χρησιμοποιήσει την τεχνητή νοημοσύνη για να δημιουργήσει πολλές παραλλαγές μιας μελωδίας, αλλά η αυθεντική δημιουργία έγκειται στο ότι επιλέγει, τροποποιεί και συνδυάζει αυτές τις παραλλαγές σε μια σύνθεση που

εκφράζει τη δική του συναισθηματική πρόθεση. Το μυαλό του καλλιτέχνη πρέπει να καθοδηγεί τη μηχανή, όχι να κρύβεται πίσω από αυτήν. Ξέρετε ότι κάνετε κάτι καινοτόμο, όταν ξέρετε τι κάνετε, γιατί το κάνετε, με ποιον το κάνετε, αλλά δεν ξέρετε πού θα σας οδηγήσει αυτό. Αυτή είναι μια διάσταση της ανθρώπινης δημιουργικότητας που εδράζεται στη ριζοσπαστική φαντασία, την οποία η τεχνητή νοημοσύνη δεν μπορεί να την κατανοήσει, να τη συλλάβει ή να την παραγάγει. Συνεπώς, η αυθεντική υβριδικότητα συνίσταται στην επεξεργασία της διεπαφής μεταξύ ανθρώπου και μηχανής, όχι στην πλήρη εξάρτηση από αυτήν. Απαιτεί μια στάση συνειδητής δημιουργικής ενασχόλησης: αποδεχόμαστε τη δύναμη της τεχνητής νοημοσύνης για να αυξήσουμε τις ικανότητές μας, ενώ ταυτόχρονα σκεφτόμαστε συνεχώς πώς αυτή η αύξηση ευθυγραμμίζεται με τα ιδανικά μας, το στιλ και τις αξίες μας. Στην υβριδική εαυτότητα, όπως και σε κάθε εαυτότητα, η αυθεντικότητα προϋποθέτει αυτεπίγνωση, αυτοδιάθεση και συνεχόμενη αυτο-δημιουργία – μόνο που τώρα, αυτά πρέπει να επεκταθούν για να συμπεριλάβουν την τεχνολογική προέκταση της εαυτότητας.

Υπερπραγματικότητα και διάβρωση του πραγματικού

Η πρόκληση της αυθεντικότητας είναι πολύ έντονη στη σύλληψη του Baudrillard (1994) περί υπερπραγματικότητας: της υποκατάστασης του πραγματικού από τις προσομοιώσεις του. Σε αυτό το πεδίο, τα σημεία δεν αναφέρονται πλέον στην υποκείμενη πραγματικότητα αλλά κυκλοφορούν μεταξύ τους, παράγοντας νοήματα που δεν αναφέρονται κάπου. Η τεχνητή νοημοσύνη επιταχύνει σε μεγάλο βαθμό αυτή τη δυναμική. Τα deepfakes, οι εικονικοί ινφλουένσερ που δημιουργεί η τεχνητή νοημοσύνη, και τα εμβυθιστικά εικονικά περιβάλλοντα παράγουν εμπειρίες πανομοιότυπες με την πραγματικότητα, τουλάχιστον για την πλειονότητα των ανθρώπων, οι οποίες όμως δεν εδράζονται σε κάποιο οντολογικό υπόβαθρο. Δεν μπορούμε πλέον να βασιζόμαστε στις αισθήσεις μας, ούτε και στο ένστικτό μας, για να διακρίνουμε το γεγονός από τη φαντασία, καθώς οι προηγμένες προσομοιώσεις μάς παρουσιάζουν μια τέλεια «ως εάν» πραγματικότητα.

Αυτό έχει πολιτισμικές, ηθικές, κοινωνικοπολιτικές και υπαρξιακές συνέπειες. Όταν παράγονται εικόνες, αφηγήσεις –μέχρι και ταυτότητες– χωρίς να βασίζονται στην εμπειρία ή στην αλήθεια, χάνεται

ο προσανατολισμός μας στον κόσμο. Οι άνθρωποι χάνουν την ικανότητά τους να διακρίνουν τι είναι αυθεντικό – όχι μόνο στους άλλους, αλλά όσον αφορά και την ίδια τους την εαυτότητα. Αν κάποιος ζει περιτριγυρισμένος από κολακευτικούς ψηφιακούς καθρέφτες και θαλάμους ηχούς (echo chambers), η αντίληψη της εαυτότητάς του ενδέχεται να γίνει και αυτή μια προσομοίωση, μια προσεκτικά κατασκευασμένη περσόνα που μπορεί να είναι εντελώς διαφορετική από την πραγματικότητα του χαρακτήρα και των ικανοτήτων του. Η διάσημη υπόθεση προσομοίωσης του Bostrom (2003) ριζοσπαστικοποιεί αυτήν την ανησυχία: εάν η ίδια η πραγματικότητα μπορεί να είναι μια προσομοίωση, τι απογίνεται η αυθεντικότητα; Αν υποθέσουμε, χάριν της συζήτησης, ότι ίσως ήδη ζούμε μέσα σε μια προσομοίωση, τα όρια μεταξύ πραγματικού και πλαστού θολώνουν εντελώς. Ωστόσο, ακόμη και χωρίς αυτήν την ακραία πιθανότητα, η επικράτηση των μικρών προσομοιώσεων στην καθημερινή ζωή (οι αναρτήσεις που βλέπουμε στα μέσα κοινωνικής δικτύωσης που είναι επιλεγμένες ώστε να επιβεβαιώνουν τις προκαταλήψεις μας, η επαυξημένη πραγματικότητα, το περιεχόμενο που έχει δημιουργήσει η τεχνητή νοημοσύνη το οποίο παρουσιάζεται σαν να είναι κατασκευασμένο από ανθρώπους) αποτελεί μια τεράστια πρόκληση: πώς να ζήσουμε μια ζωή με νόημα, με αλήθειες και στόχους υπό αυτές τις συνθήκες;

Ο David Chalmers (2022: 17) απαντά ότι οι εικονικές πραγματικότητες μπορεί να είναι «αρκετά πραγματικές» – να προσφέρουν γνήσια εμπειρία και νόημα. Μπορεί κανείς να βιώνει αυθεντικές εμπειρίες και να δημιουργεί αυθεντικές σχέσεις ακόμη και μέσα σε εικονικούς κόσμους (για παράδειγμα, οι φιλίες σε ένα διαδικτυακό παιχνίδι μπορεί να είναι πραγματικές φιλίες ή οι συνδημιουργίες έργων τέχνης μέσω ψηφιακών πλατφορμών μπορούν να είναι πραγματικές δημιουργίες). Αυτό, όμως, δεν μας απαλλάσσει από την ευθύνη να διακρίνουμε τι είναι αυθεντικό μέσα σε αυτούς. Πρέπει παρ' όλα αυτά να αναρωτιόμαστε αν αυτές οι εμπειρίες και οι σχέσεις φέρουν το ίδιο βάρος, όπως και πώς συνδέονται με τον μη εικονικό κόσμο. Η αυθεντικότητα στην υπερπραγματικότητα συνεπάγεται μια επανατοποθέτηση – μια εκ νέου επιβολή των ανθρώπινων αξιών και προθέσεων σε ρευστά σημειολογικά τοπία. Δεν πρόκειται για μια αντιδραστική επιστροφή στην προψηφιακή εποχή, αλλά για μια προσεκτική δημιουργία νοήματος μέσα σε ασταθή συμφραζόμενα. Στην πράξη, αυτό μπορεί να σημαίνει την ανάπτυξη προσωπικών και πολιτισμικών πρακτικών επαλήθευσης

(να κάνουμε μια παύση για να αμφισβητήσουμε αυτό που βλέπουμε στο διαδίκτυο, να αξιολογούμε τις πηγές και τις αποδείξεις) και εκ νέου σύνδεσης με το πραγματικό (να περνάμε χρόνο στη φύση ή να αφιερώνουμε χρόνο σε μη διαμεσολαβημένες διαπροσωπικές αλληλεπιδράσεις ώστε να εξισορροπήσουμε τις ψηφιακές). Σημαίνει να μην αντιμετωπίζουμε το εικονικό σαν μια απόλυτη απόδραση ή σαν υποκατάστατο, αλλά σαν ένα επίπεδο της πραγματικότητας που τελικά πρέπει να λογοδοτεί στα υπόλοιπα.

Ο Borgmann (1984), γράφοντας για τον πολιτισμικό αντίκτυπο της τεχνολογίας, παρατηρεί ότι η μετανεωτερικότητα –με την απώλεια των μεγάλων αφηγημάτων– σε συνδυασμό με την υψηλή τεχνολογία παράγει τον υπερμοντερνισμό, μια κατάσταση διάχυτης υπερπραγματικότητας. Ως αντίδοτο, προτείνει έναν προσανατολισμό που τον αποκαλεί «μεταμοντέρνο ρεαλισμό», τον οποίον τον ορίζει ως μια κατεύθυνση που αποδέχεται τα διδάγματα της μεταμοντέρνας κριτικής και επιλύει τις αμφισημίες της μεταμοντέρνας συνθήκης με μια προσέγγιση υπομονετικής αλλά ζωτικής προσήλωσης σε μια κοινή τάξη που είναι επικεντρωμένη στον συλλογικό εορτασμό. Αυτό που μπορεί να προσδώσει ζωντάνια σε αυτήν την προσέγγιση και να προσφέρει ένα επίκεντρο για τον εορτασμό είναι η πραγματικότητα. Με άλλα λόγια, προτείνεται ότι η θεραπεία για την υπερπραγματικότητα είναι μια ανανεωμένη δέσμευση στο πραγματικό – σε κοινές, απτές δραστηριότητες και εστιασμένες πρακτικές που μας συνδέουν μεταξύ μας και με τον κόσμο. Για παράδειγμα, δραστηριότητες μέσα σε μια κοινότητα όπως ένα δείπνο με τους γείτονες, μια δημόσια συζήτηση ή ένα συλλογικό καλλιτεχνικό έργο είναι εκφάνσεις κοινοτικού εορτασμού της πραγματικότητας, που αντισταθμίζουν τις επιδράσεις της απομόνωσης και του κατακερματισμού που προκαλεί η ψηφιακή υπερπραγματικότητα. Επικεντρώνοντας εκ νέου τον πολιτισμό σε πράγματα που δεν μπορούν να προσομοιωθούν –στην ενσώματη παρουσία μας, στο φυσικό περιβάλλον μας, στην πραγματική κοινότητά μας– ο υπερμοντερνισμός, μέσω του μεταμοντέρνου ρεαλισμού, μπορεί επίσης να μας επαναθεμελιώσει. Στην εποχή της τεχνητής νοημοσύνης, αυτό μπορεί να σημαίνει ότι όσο εξελιγμένες και να γίνουν οι εικονικές μας εμπειρίες, εμείς εξακολουθούμε να προσδίδουμε αξία στις αυθεντικές στιγμές της μη διαμεσολαβημένης ζωής: στο άγγιγμα του χεριού κάποιου άλλου, στον ήχο της ζωντανής μουσικής σε ένα δωμάτιο, στο θέαμα ενός ηλιοβασιλέματος χωρίς

φίλτρα. Αυτές οι εμπειρίες μάς υπενθυμίζουν ότι η πραγματικότητα εξακολουθεί να υπάρχει έξω από την οθόνη μας, και ότι μόνο όταν σχετιζόμαστε με αυτήν την πραγματικότητα μπορεί η αυθεντικότητα να θεμελιωθεί ουσιαστικά.

Ισο-αυθεντικότητα μεταξύ ανθρώπων και συν-δημιουργία με την τεχνητή νοημοσύνη

Συχνά η δημιουργικότητα θεωρείται ότι είναι το απόγειο της αυθεντικής έκφρασης – το πεδίο στο οποίο η πρωτοτυπία και η προσωπική φωνή ενός ανθρώπου εκφράζονται πιο ξεκάθαρα. Ωστόσο, στην εποχή της παραγωγικής τεχνητής νοημοσύνης, ο ορισμός της δημιουργικότητας δεν είναι ξεκάθαρος. Η Boden (1998) είναι γνωστό ότι υποστηρίζει πως υπάρχουν τρία είδη δημιουργικότητας: η συνδυαστική (καινοτόμοι συνδυασμοί οικείων ιδεών), η διερευνητική (καινοτόμες διερευνήσεις μέσα σε υφιστάμενους στιλιστικούς κανόνες ή εννοιολογικά πλαίσια), και η μετασχηματιστική (ο ορισμός εκ νέου και εκ βάθρων αυτών των πλαισίων). Η τεχνητή νοημοσύνη ήδη διαπρέπει στα πρώτα δύο, παράγοντας καινοτόμους συνδυασμούς και διερευνώντας οριοθετημένα πλαίσια με τρόπους που μπορούν να εκπλήξουν ή ακόμα και να εμπνεύσουν τους ανθρώπους. Παραδείγματος χάριν, οι αλγόριθμοι παραγωγής εικόνων με τεχνητή νοημοσύνη έχουν τη δυνατότητα να συνθέσουν ευφάνταστους συνδυασμούς μοτίβων στην τέχνη, και οι αλγόριθμοι της τεχνητής νοημοσύνης για τα παιχνίδια έχουν ανακαλύψει καινοτόμες στρατηγικές βάσει των κανόνων του Go και του σκακιού. Ωστόσο, η μετασχηματιστική δημιουργικότητα –αυτή που καταστρατηγεί τους κανόνες και προκαλεί μετατόπιση παραδείγματος– εξακολουθεί προς το παρόν να είναι προνόμιο των ανθρώπων. Διότι δεν απαιτεί απλώς να λαμβάνει κανείς υπόψη ό,τι είναι ήδη γνωστό, αλλά και να βλέπει το άγνωστο· δεν προϋποθέτει απλώς να μετακινεί κανείς τα πιόνια πάνω σε ένα επιτραπέζιο παιχνίδι, αλλά να εφευρίσκει ένα εντελώς νέο παιχνίδι.

Η αυθεντικότητα-ως-δημιουργικότητα –όπως και η ισο-αυθεντικότητα και η συν-αυθεντικότητα– προϋποθέτει κάτι παραπάνω από την καινοτομία· απαιτεί πρόθεση, εκφραστικότητα και ευαισθησία (Ερηνάκης, 2020). Όταν η τεχνητή νοημοσύνη παράγει ένα ποίημα ή έναν πίνακα, δεν διαθέτει το βιωματικό υπόβαθρο που προσδίδει νόημα, ούτε έχει την ικανότητα να είναι ευαίσθητη όσον αφορά την αξία

του δημιουργικού αποτελέσματος. Η κριτική του Searle (1980) για τη μηχανική κατανόηση είναι επίκαιρη: το συντακτικό δεν είναι σημασιολογία. Η τεχνητή νοημοσύνη μπορεί να συνδυάσει τα σύμβολα με ευρηματικούς τρόπους, αλλά δεν *εννοεί* τίποτα με αυτά· δεν έχει βιωμένη εμπειρία ώστε να προσδώσει συμφραζόμενα ή βάθος στα αποτελέσματα που βγάζει. Η αισθητική αξία ενός έργου τέχνης δεν έγκειται απλώς στη μορφή του, αλλά στον υποκειμενικό ορίζοντα από τον οποίον αναδύεται. Ένα ερωτικό ποίημα που έχει γραφτεί με τεχνητή νοημοσύνη μπορεί να εμπεριέχει μεταφορές και μέτρο, αλλά αμφισβητούμε αν πρόκειται για αυθεντική τέχνη όταν η τεχνητή νοημοσύνη δεν αγαπάει, δεν λαχταράει, δεν *ζει*. Αυτό δεν σημαίνει ότι η τέχνη που δημιουργεί η τεχνητή νοημοσύνη στερείται αξίας, απλώς αναδεικνύει μια ποιοτική διαφορά. Ως παρατηρητές, συχνά αναζητούμε να βρούμε στην τέχνη τα ίχνη μιας άλλης ψυχής – το αυθεντικό σήμα που στέλνει ένα άλλο μυαλό που αλληλεπιδρά μέσα στον κόσμο. Με την τέχνη που δημιουργεί η τεχνητή νοημοσύνη, αυτή η σύνδεση είναι ασαφής. Μας προκαλεί να αναρωτηθούμε: είναι ο δημιουργός κομμάτι αυτού που κάνει την τέχνη αυθεντική, και αν ναι, μπορεί πράγματι η τεχνητή νοημοσύνη να καταστεί δημιουργός με αυτή την έννοια;

Ο David Gunkel πρότεινε μάλιστα το ενδεχόμενο ύπαρξης τεχνητού δημιουργού. Όσον αφορά τη μουσική και τη λογοτεχνία που έχει δημιουργηθεί μέσω τεχνητής νοημοσύνης, θέτει το ερώτημα αν η αντίληψή μας περί δημιουργού και πνευματικής ιδιοκτησίας θα έπρεπε να διευρυνθεί ώστε να συμπεριλάβει και τις μηχανές (Gunkel, 2018). Ρωτάει προκλητικά αν η τεχνητή νοημοσύνη που συνθέτει μια συμφωνία δικαιούται τα εύσημα όχι μόνο ως εργαλείο, αλλά ως δημιουργός. Τέτοια ερωτήματα μας αναγκάζουν να διευκρινίσουμε γιατί προσδίδουμε αξία στην ανθρώπινη δημιουργικότητα εξαρχής. Η ανάλυση του Gunkel προτείνει ότι η προσκόλληση σε μια αυστηρά ανθρωποκεντρική έννοια του δημιουργού μπορεί να πάψει να είναι βιώσιμη διότι τα συστήματα τεχνητής νοημοσύνης παράγουν πλέον αποτελέσματα που δεν ξεχωρίζουν από τις ανθρώπινες δημιουργίες. Ωστόσο, ακόμη και αν αποδεχτούμε ότι η τεχνητή νοημοσύνη μπορεί να δημιουργήσει με τεχνικούς όρους, εξακολουθεί να υπάρχει μια ευρεία συναίνεση ότι εκείνο που δημιουργούν οι άνθρωποι είναι εντελώς διαφορετικό. Όπως θα έλεγε ο Rorty (1989), οι δημιουργοί που είναι άνθρωποι ενσωματώνουν στην τέχνη τους το τελικό τους λεξιλόγιο

–την καθαρά δική τους, ατομική άποψη για τον κόσμο–, ενώ η τεχνητή νοημοσύνη λειτουργεί μόνο με ένα κολάζ δανεισμένων στοιχείων.

Εντούτοις, η συνδημιουργία μεταξύ ανθρώπων πρωτίστως, και με την τεχνητή νοημοσύνη δευτερευόντως, αναδύεται ως μια γόνιμη πρακτική. Όταν οι άνθρωποι συνεργάζονται μεταξύ τους, όπως και με την τεχνητή νοημοσύνη, για να διευρύνουν ή να ανακαλύψουν νέες μορφές έκφρασης, το αποτέλεσμα μπορεί να είναι αυθεντικά υβριδικό. Το κλειδί είναι η διατήρηση της ιδιότητας του δημιουργού και της προθετικότητας εκ μέρους των ανθρώπων: όταν ο καλλιτέχνης ή ο στοχαστής είναι άνθρωπος και διαμορφώνει το όραμα, επεξεργάζεται ή επιμελείται το αποτέλεσμα και το εντάσσει σε ένα βιωμένο πλαίσιο νοήματος, τότε η αυθεντικότητα διατηρείται στο τελικό προϊόν. Η τεχνητή νοημοσύνη γίνεται η μούσα ή το μέσο, όχι ο μοναδικός δημιουργός. Αυθεντικότητα-ως-δημιουργικότητα – μια διαρκής συνδιαμόρφωση της εαυτότητας και του κόσμου σε διάλογο με το τεχνητό. Για παράδειγμα, ένας μυθιστοριογράφος μπορεί να χρησιμοποιήσει ένα σύστημα τεχνητής νοημοσύνης για να του προτείνει μερικές ανατροπές στην πλοκή ή εκδοχές διαλόγων μεταξύ κάποιων χαρακτήρων που δεν είχε σκεφτεί ο ίδιος· αυτές οι προτάσεις μπορεί να πυροδοτήσουν τη φαντασία του και να αρχίσει να σκέφτεται προς νέες κατευθύνσεις. Το μυθιστόρημα που τελικά θα γράψει, παρόλο που πήρε τη βοήθεια της τεχνητής νοημοσύνης, είναι αυθεντικά δικό του διότι ο ίδιος επέλεξε και επεξεργάστηκε αυτά τα στοιχεία σύμφωνα με την καλλιτεχνική του πρόθεση και τη δική του αίσθηση της αφήγησης. Στη μουσική, η τεχνητή νοημοσύνη μπορεί να παραγάγει μια σειρά από ήχους που ένας συνθέτης μπορεί να τους βρει ενδιαφέροντες, αλλά το ποιοι ήχοι τελικά θα χρησιμοποιηθούν στη σύνθεση και με ποια σειρά εξαρτάται από το αυθεντικό αισθητήριο και την κρίση του συνθέτη που αποφασίζει. Αυτού του είδους η συνδημιουργία μπορεί ακόμα και να αντιμετωπιστεί ως επέκταση της ανθρώπινης δημιουργικής ικανότητας – είναι παρόμοια με τη χρήση ενός μουσικού οργάνου. Όπως ακριβώς η εκφραστικότητα μιας ηλεκτρικής κιθάρας με εφέ μπορεί να ξεπεράσει τον φυσικό ήχο μιας χορδής χωρίς καμία ενίσχυση, έτσι και η τεχνητή νοημοσύνη μπορεί να αντιμετωπιστεί σαν ένα νέο όργανο μέσω του οποίου η ανθρώπινη δημιουργικότητα αποκτά μια νέα φωνή.

Όταν ενστερνιζόμαστε τη συνδημιουργία, οφείλουμε να είμαστε ταυτόχρονα προσεκτικοί. Είναι λεπτή η γραμμή ανάμεσα στο να

χρησιμοποιούμε την τεχνητή νοημοσύνη για να βοηθήσουμε την ανθρώπινη δημιουργικότητα και στο να της επιτρέπουμε να την *αντικαθιστά*. Στην πρώτη περίπτωση, ο άνθρωπος διατηρεί τον έλεγχο, διατηρώντας την προθετικότητα και την αυθεντικότητά του· στη δεύτερη περίπτωση ελλοχεύει ο κίνδυνος να γίνουν οι δημιουργικές πρακτικές αυτοματοποιημένες και τυποποιημένες, και επομένως μη αυθεντικές. Ως κοινότητα, θα πρέπει ίσως να διαπραγματευτούμε νέους κανόνες: για παράδειγμα, τη διαφάνεια σχετικά με τη συμμετοχή της τεχνητής νοημοσύνης στα καλλιτεχνικά έργα (ώστε το κοινό να ξέρει με τι έχει να κάνει), και ενδεχομένως νέες κατηγορίες αισθητικής αποτίμησης που αναγνωρίζουν τόσο τις ανθρώπινες συνεισφορές όσο και τη συνεισφορά της τεχνητής νοημοσύνης. Η ίδια η έννοια της αυθεντικότητας μπορεί να εξελιχτεί εν προκειμένω – μπορεί να αρχίσουμε να θεωρούμε ότι ένα έργο τέχνης είναι αυθεντικό όταν αποτελεί μια αληθινή *συνεργασία ανθρώπου και τεχνητής νοημοσύνης* αντί να προσποιείται ότι είναι αποκλειστικά ανθρώπινο δημιούργημα. Το πιο πιθανό, ωστόσο, είναι ότι την ανθρώπινη δημιουργικότητα δεν θα την επισκιάσει η τεχνητή νοημοσύνη, αντιθέτως θα της προσφέρει προκλήσεις και ερεθίσματα. Μέσα από αυτήν την πρόκληση, μπορούμε να διασαφηνίσουμε τι είναι αυτό που ξεχωρίζει την ανθρώπινη δημιουργικότητα: η σπίθα της συνειδητότητας, η ένταση της βιωμένης εμπειρίας, η ελευθερία να καταστρατηγούμε τους κανόνες και να αψηφούμε τις προσδοκίες. Αυτές είναι ποιότητες που κανένας αλγόριθμος δεν μπορεί (ακόμα) να αναπαραγάγει. Αντιθέτως, μπορεί να μας βοηθήσει στην επίτευξη αυτού που θα έπρεπε να αποτελεί τον βασικό στόχο της εποχής μας, δηλαδή την ισο-αυθεντικότητα και τη συν-αυθεντικότητα μέσω της συνδημιουργίας που εδράζεται στη ριζοσπαστική φαντασία και στη σχεσιακή λογική.

Συμπέρασμα

Η αυθεντικότητα στην εποχή της τεχνητής νοημοσύνης δεν είναι μια νοσταλγική επιστροφή σε κάποια προ-τεχνολογική έννοια της ταυτότητας, αλλά μια προσανατολισμένη προς το μέλλον πρακτική κριτικής οικειοποίησης. Στο προσωπικό, στο συλλογικό, στο σχεσιακό και στο δημιουργικό πεδίο, η αυθεντικότητα απαιτεί τα άτομα και οι κοινωνίες να αντιστέκονται στην παθητική αφομοίωση της αλγοριθμικής λογικής. Αντιθέτως, πρέπει να επιβεβαιώνουμε την αυτενέργεια, την

αναστοχαστικότητα και εκφραστική εξατομίκευση εν μέσω τεχνολογικά κορεσμένων πραγματικοτήτων. Αυτό συνεπάγεται μια συνειδητά εξισορροπητική πράξη: την αξιοποίηση των πλεονεκτημάτων της τεχνητής νοημοσύνης που διευρύνουν τη γνώση, την επικοινωνία και τη δημιουργικότητα, και παράλληλα τη συνεχόμενη διαφύλαξη και προβολή των ανθρώπινων αξιών και νοημάτων που η τεχνητή νοημοσύνη δεν μπορεί να προσφέρει. Αυτό σημαίνει ότι δεν πρέπει να θέτουμε μόνο το ερώτημα *«Τι μπορεί να κάνει η τεχνητή νοημοσύνη;»* αλλά να διερωτόμαστε παράλληλα *«Τι θα μπορούσαμε* να κάνουμε εμείς με την τεχνητή νοημοσύνη ώστε να ζούμε αυθεντικά;» – το οποίο είναι ένα θεμελιώδες ερώτημα.

Η τεχνητή νοημοσύνη δεν αναιρεί την αυθεντικότητα – αντιθέτως, μεταμορφώνει και δοκιμάζει τις συνθήκες της. Σε αυτό το υπερπραγματικό τοπίο, το αυθεντικό υποκείμενο είναι ο άνθρωπος που δημιουργεί νόημα εν μέσω προσομοίωσης, που ενσωματώνει τα εργαλεία χωρίς να χάνει τη δική του αυτοδιάθεση και κατεύθυνση, και που διατηρεί την παρουσία και την ενσυναίσθηση σε διαμεσολαβημένα περιβάλλοντα. Η αυθεντικότητα γίνεται *ποίησις*: ένας παραγωγικός, σχεδόν καλλιτεχνικός αγώνας να παραμείνουμε άνθρωποι μέσω των ίδιων των τεχνολογιών που απειλούν να εξαφανίσουν τα όρια της ανθρωπινότητας (Ερηνάκης, 2020: 147). Ο Dreyfus (2001), ως χαϊντεγκεριανός κριτικός της τεχνητής νοημοσύνης, παρατήρησε πριν από δεκαετίες ότι η περιπέτεια της τεχνητής νοημοσύνης μάς υποχρεώνει να διατυπώσουμε τι πραγματικά είναι σημαντικό όσον αφορά την ανθρώπινη νοημοσύνη και ύπαρξη. Παρομοίως, σήμερα η τεχνητή νοημοσύνη μας προτρέπει να διατυπώσουμε τι είναι πραγματικά σημαντικό όσον αφορά το να είναι κανείς *αυθεντικός*. Μας αναγκάζει να εντοπίσουμε τα αδιαπραγμάτευτα χαρακτηριστικά μιας ζωής που νιώθουμε ότι είναι αληθινή και γεμάτη νόημα: ενδεχομένως την ικανότητά μας να συναισθανόμαστε, την προθυμία μας να αναλαμβάνουμε ηθικά ρίσκα, την εδραιωμένη και ενσωματωμένη οπτική μας, ή την ικανότητά μας να μην επιλέγουμε τις προβλέψιμες προτιμήσεις.

Τελικά, η αναζήτηση της αυθεντικότητας στην εποχή της τεχνητής νοημοσύνης ίσως συνοψίζεται στην ενσυνειδητότητα, στην κυριότητα και στην ευαισθησία. Πρέπει ενσυνείδητα να αντιλαμβανόμαστε με ποιους τρόπους η τεχνητή νοημοσύνη μεταβάλλει τις επιλογές και τις αντιλήψεις μας, και να αναλαμβάνουμε εμείς την κυριότητα της

σχέσης μας με αυτά τα συστήματα. Αντί να κινούμαστε στον κόσμο έχοντας βάλει τον αυτόματο πιλότο, ακολουθώντας τις προτροπές και υποδείξεις των μηχανών με νοημοσύνη, καλούμαστε να ζούμε συνειδητά. Τούτο θα μπορούσε να σημαίνει να αποχωριζόμαστε περιοδικά τις ψηφιακές μας ευκολίες για να επανασυνδεθούμε με πρωτογενείς εμπειρίες. Θα μπορούσε να σημαίνει να καλλιεργούμε δεξιότητες που η τεχνητή νοημοσύνη δεν μπορεί εύκολα να τις αντιγράψει – τον αυθορμητισμό, τον αυτοσχεδιασμό, την καινοτόμο φαντασία σε διαφορετικά γνωστικά πεδία. Σίγουρα σημαίνει να συνεχίσουμε τις ανθρώπινες συζητήσεις: έναν συλλογικό διάλογο σχετικά με το πώς θέλουμε να ενσωματώσουμε την τεχνητή νοημοσύνη στο αφήγημα της ανθρωπινότητας. Η πραγματικότητα εξακολουθεί να υπάρχει γύρω μας για να τη γιορτάσουμε, και υπό τον όρο ότι κρατάμε το ένα πόδι μας στην πραγματικότητα, μπορούμε να χρησιμοποιήσουμε την υπερπραγματικότητα στη σωστή της διάσταση.

Οι φιλόσοφοι, οι καλλιτέχνες, όσοι χαράσσουν πολιτική, οι τεχνολόγοι και γενικά όλοι οι πολίτες παίζουν κυρίαρχο ρόλο σε αυτήν τη διαπραγμάτευση. Η αυθεντικότητα ήταν πάντα, όπως υποστήριζε ο Heidegger, στόχος και όχι κάτι δεδομένο – κάτι που πρέπει να επιτευχθεί εκ νέου από κάθε γενιά. Στη δική μας γενιά, αυτός ο στόχος είναι συνυφασμένος με την παρουσία της τεχνητής νοημοσύνης ως εργαλείου και ως καθρέφτη και ως του απόλυτου αναπτυξιακού στόχου – είναι ταυτοχρόνως οδικός χάρτης και προορισμός. Όταν διακρίνουμε πότε η τεχνητή νοημοσύνη υπηρετεί τους ανθρώπινους στόχους μας και πότε τους υποβιβάζει, ξέρουμε ποια είναι η ουσία της αυθεντικότητας. Διασφαλίζουμε ότι, ακόμα και όταν περιβαλλόμαστε από ομοιώματα, η φωνή που μιλάει και το χέρι που δημιουργεί εξακολουθούν να έχουν ανθρώπινη υπόσταση. Συνοψίζοντας, η τεχνητή νοημοσύνη μπορεί είτε να δράσει ως καταλύτης για τη βαθύτερη κατανόηση και ανάπτυξη της εαυτότητάς μας είτε ως ένας καταστροφικός πειρασμός να εγκαταλείψουμε την εαυτότητά μας· αυθεντικότητα, ως φάρος που φωτίζει τον δρόμο μας, σημαίνει να επιλέγουμε το πρώτο μονοπάτι.

Βιβλιογραφία

Arendt, H. (1967). Truth and politics. *The New Yorker*.

Baudrillard, J. (1994). *Simulacra and simulation* (μτφρ. S. Glaser). University of Michigan Press.

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence, 103*(1), 347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

Borgmann, A. (1984). *Technology and the character of contemporary life: A philosophical inquiry*. University of Chicago Press.

Borgmann, A. (1992). *Crossing the postmodern divide*. University of Chicago Press.

Bostrom, N. (2003). Are we living in a computer simulation? *Philosophical Quarterly, 53*(211), 243–255.

Buber, M. (1937). *I and thou* (μτφρ. R. G. Smith). T. & T. Clark. (Το πρωτότυπο έργο εκδόθηκε το 1923)

Chalmers, D. J. (2022). *Reality+: Virtual worlds and the problems of philosophy*. W. W. Norton.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Cocking, D., & Matthews, S. (2000). Unreal friends. *Ethics and Information Technology, 2*(4), 223–231. https://link.springer.com/article/10.1023/A:1011414704851

Coeckelbergh, M. (2017). *New romantic cyborgs: Romanticism, information technology, and the end of the machine*. MIT Press.

Deleuze, G. (1995). *Negotiations*. Columbia University Press.

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press.

Dreyfus, H. L. (1999). Kierkegaard on the Internet: Anonymity vs. commitment in the present age. *Kierkegaard Studies Yearbook, 1*, 96–109.

Dreyfus, H. L. (2001). *On the Internet*. Routledge.

Ερηνάκης, Ν. (2020). *Αυθεντικότητα και αυτονομία. Από τη δημιουργικότητα στην ελευθερία.* Κείμενα.

Erinakis, N. (2023). Why authenticity precedes autonomy. *Philosophical Inquiries, 11*(1), 43–64.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Floridi, L. (2018). Artificial intelligence, deepfakes and a future of ectypes (SSRN Scholarly Paper ID 3834958). https://doi.org/10.2139/ssrn.3834958

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy, 68*(1), 5–20. https://doi.org/10.2307/2024717

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.

Gunkel, D. J. (2018). *Robot rights*. MIT Press.

Habermas, J. (1984). *The theory of communicative action, Vol. 1: Reason and the rationalization of society* (T. McCarthy, Trans.). Beacon Press.

Habermas, J. (1989). *The structural transformation of the public sphere.* MIT Press.

Heidegger, M. (1962). *Being and time* (μτφρ. J. Macquarrie & E. Robinson). Harper & Row. (Original work published 1927)

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.

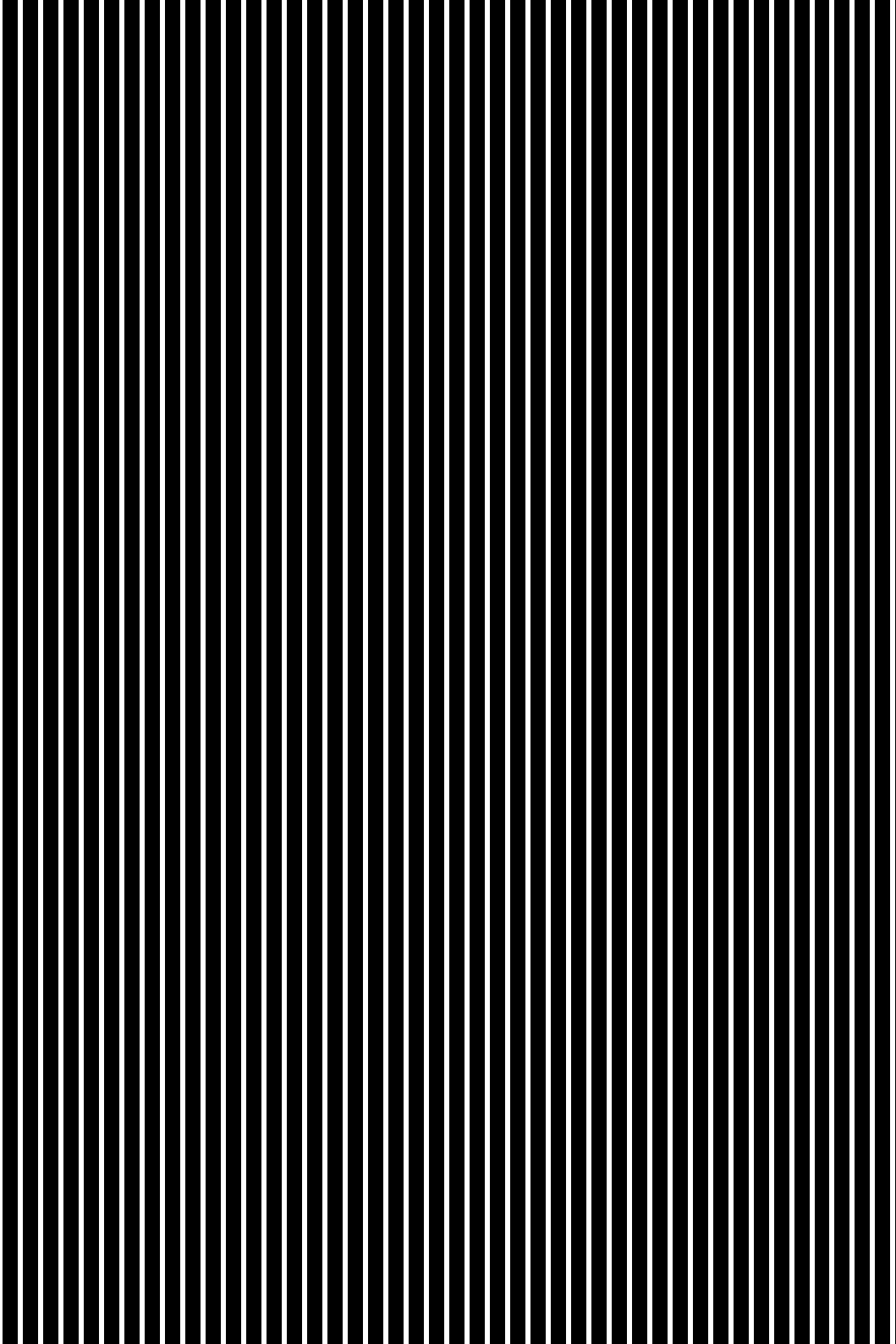
Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.

Rorty, R. (1989). *Contingency, irony, and solidarity.* Cambridge University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things.* University of Chicago Press.



Βιογραφικά συντελεστών

Κατερίνα Αντωνοπούλου
Η Κατερίνα Αντωνοπούλου είναι καλλιτέχνις ψηφιακών τεχνών και επίκουρη καθηγήτρια αλληλεπιδραστικής τέχνης στο Τμήμα Ψηφιακών Τεχνών και Κινηματογράφου του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Μέσα από την έρευνά της και την καλλιτεχνική της πρακτική, διερευνά τις αλληλεπιδράσεις της τέχνης και της τεχνολογίας με την κοινωνία και το φυσικό περιβάλλον. Τα έργα της έχουν τη μορφή αλληλεπιδραστικών εγκαταστάσεων και performances, περιβαλλόντων μεικτής πραγματικότητας, ψηφιακής αφήγησης και αστικών παρεμβάσεων. Το έργο της έχει εκτεθεί σε πολυάριθμες διεθνείς εκθέσεις, ενώ το ερευνητικό της έργο έχει δημοσιευτεί σε πρακτικά συνεδρίων και σε επιστημονικά περιοδικά. Η Κατερίνα Αντωνοπούλου έχει επιμεληθεί φεστιβάλ και εκθέσεις ψηφιακής τέχνης και υπήρξε ιδρυτικό μέλος της διεθνούς καλλιτεχνικής κολεκτίβας Once Upon a Byte. Έχει διδακτορικό στις ψηφιακές τέχνες από το Πανεπιστήμιο Αιγαίου, μεταπτυχιακό τίτλο σπουδών στις ψηφιακές τέχνες από το Πανεπιστήμιο Pompeu Fabra της Βαρκελώνης και δίπλωμα μηχανικού υπολογιστών από το Εθνικό Μετσόβιο Πολυτεχνείο.

Κατερίνα Γκουτζιούλη
Η Κατερίνα Γκουτζιούλη είναι επιμελήτρια, ερευνήτρια και πολιτιστική διαχειρίστρια με έδρα την Αθήνα. Η έρευνά της επικεντρώνεται στην τέχνη και τον ψηφιακό πολιτισμό, διερευνώντας ζητήματα όπως η πολιτισμική ταυτότητα, οι πολιτικές των δικτύων, η επιτήρηση και η τεχνητή νοημοσύνη. Έχει εργαστεί ως επιμελήτρια, σύμβουλος πολιτισμού, μέντορας και διαχειρίστρια έργων σε δημόσιους και ιδιωτικούς φορείς, όπως η Ανώτατη Σχολή Καλών Τεχνών, το Ινστιτούτο Γκαίτε Αθηνών, το Athens Digital Arts Festival, ο δήμος Αθηναίων, το φεστιβάλ SIGGRAPH, το Μουσείο Μπενάκη, το Bios-Ρομάντσο, το Onassis ONX, το V2_Institute for the Unstable Media, κ.ά. Έχει επιμεληθεί εκθέσεις, φεστιβάλ και εκδόσεις με αντικείμενο την τέχνη και την τεχνολογία. Έχει μεταπτυχιακό τίτλο σπουδών στον οπτικό πολιτισμό από το Πανεπιστήμιο του Westminster (Λονδίνο) και πτυχίο Αγγλικής Γλώσσας και Φιλολογίας από το Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (ΕΚΠΑ). Είναι συνιδρύτρια και καλλιτεχνική διευθύντρια του οργανισμού VEKTOR Athens και υπότροφος του Ιδρύματος Fulbright.

Δρ. Νίκος Ερηνάκης
Ο Νίκος Ερηνάκης (Αθήνα, 1988) είναι Επίκουρος Καθηγητής Κοινωνικής και Πολιτικής Φιλοσοφίας και Φιλοσοφίας του Πολιτισμού στο Πανεπιστήμιο Κρήτης, ενώ παράλληλα διδάσκει στο Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, στην Ανωτάτη Σχολή Καλών Τεχνών και στο Ελληνικό Ανοικτό Πανεπιστήμιο (ΕΑΠ). Είναι Διδάκτωρ Φιλοσοφίας του Πανεπιστημίου του Λονδίνου, όπου ξεκίνησε τη διατριβή του, την οποία συνέχισε και ολοκλήρωσε ως αναγνωρισμένος ερευνητής (recognised research student) στο Πανεπιστήμιο της Οξφόρδης (με πλήρη υποτροφία του ΙΚΥ). Έχει σπουδάσει οικονομικά στο Οικονομικό Πανεπιστήμιο Αθηνών, φιλοσοφία και λογοτεχνία στο Πανεπιστήμιο του Warwick, φιλοσοφία των κοινωνικών επιστημών στο London School of Economics and Political Science (LSE), ενώ είναι απόφοιτος του Κολλεγίου Αθηνών (2005). Δραστηριοποιείται επίσης ως ο Επιστημονικός Διευθυντής και Αντιπρόεδρος του Διοικητικού Συμβουλίου του think tank ENA, έχει διατελέσει μέλος του Διοικητικού Συμβουλίου του Φεστιβάλ Αθηνών Επιδαύρου και Διευθυντής Στρατηγικής Ανάπτυξης του Διεθνούς Φεστιβάλ Κέρκυρας, ενώ είναι μέλος του Διοικητικού Συμβουλίου της Ελληνικής Συμφωνικής Ορχήστρας Νέων. Για το πρώτο του βιβλίο φιλοσοφίας, με τίτλο

Αυθεντικότητα και αυτονομία: Από τη δημιουργικότητα στην ελευθερία (Κείμενα, 2020), τιμήθηκε με το Βραβείο της Ακαδημίας Αθηνών. Έχουν εκδοθεί τρεις συλλογικοί τόμοι φιλοσοφικών δοκιμίων σε επιμέλειά του, τρία βιβλία ποίησής του (εκ των οποίων τα δύο έχουν μεταφραστεί και εκδοθεί στην Αγγλία και τη Γαλλία, αντιστοίχως) και δύο βιβλία ανθολογίας Ευρωπαίων ποιητών του μοντερνισμού σε μετάφρασή του. Δοκίμια, άρθρα και ποιήματά του έχουν δημοσιευθεί σε ποικίλα διεθνή περιοδικά, συλλογικούς τόμους, ανθολογίες και πρακτικά συνεδρίων με κριτές. Αυτή την περίοδο διδάσκει και αρθρογραφεί εκτενώς για ζητήματα φιλοσοφίας της τεχνητής νοημοσύνης και της ψηφιακότητας.

Κωστής Σταφυλάκης

Ο Κωστής Σταφυλάκης είναι Επίκουρος Καθηγητής Εικαστικών Τεχνών, Διευρυμένων Μέσων και Μορφών στο Τμήμα Εικαστικών Τεχνών της Ανωτάτης Σχολής Καλών Τεχνών. Είναι εικαστικός καλλιτέχνης, συγγραφέας σε θέματα θεωρίας της τέχνης και επιμελητής εκθέσεων, με διδακτορικό στις πολιτικές επιστήμες από το Πάντειο Πανεπιστήμιο. Αποφοίτησε από την Ανωτάτη Σχολή Καλών Τεχνών της Αθήνας και κατέχει μεταπτυχιακό τίτλο στη Θεωρία της Τέχνης και μεταπτυχιακό τίτλο στην Ηπειρωτική Φιλοσοφία από το Πανεπιστήμιο του Έσσεξ. Είναι μέλος του ακαδημαϊκού προσωπικού του Studio 11 (ΑΣΚΤ) και διδάσκει επίσης το μάθημα «Σύγχρονη Τέχνη, Διαδικτυακή Κουλτούρα, Πολιτική» στο Μεταπτυχιακό Εικαστικών Τεχνών και στο μεταπτυχιακό πρόγραμμα «Ψηφιακές Μορφές Τέχνης» της ΑΣΚΤ. Τα γραπτά του εστιάζουν συχνά στη συνάντηση της τέχνης με τις μιμητικές στρατηγικές και το Πολιτικό. Έχει συνεπιμεληθεί εκθέσεις όπως οι: *The Outburst of the Digital Swamp* (The Overkill Festival, Ολλανδία, 2023), 6η Μπιενάλε της Αθήνας ANTI, 4η Μπιενάλε της Αθήνας AGORA, κ.ά. Οι πρόσφατες ατομικές του εκθέσεις περιλαμβάνουν τις *GZPRM GOPNIK SKG* (The Eye Altering, 2024), *On an island* (K-Gold Temporary Gallery, 2024), *GZPRM GOPNIK* (KEIV Αθήνα, 2023) και *Chloroquine Juggalo* (KEIV Αθήνα, 2021). Οι διατομικές εκθέσεις του περιλαμβάνουν τις *Readiness: Civil War* (με τον Τεό Τριανταφυλλίδη) στην Panke Gallery (Βερολίνο, 2022) και *KavecS* στην Neue Ravensburger Kunstverein (2017). Έχει επίσης εκθέσει σε συλλογικές εκθέσεις όπως οι *We are all made of stars* (MOMus-Πειραματικό Κέντρο Τεχνών, 2024), *Outraged by Pleasure* (Κτήριο «Νόμπελ», Αθήνα, 2023), *Prepper Paradise* (Bureau Europa, Μάαστριχτ, 2022), *Toxicfest* (Tick Tack Antwerpen, 2021), *Enter* (Στέγη Ιδρύματος Ωνάση, Αθήνα, 2020), Kultursymposium Weimar (2019), *Omonoia* (5η Μπιενάλε της Αθήνας, 2016), *Enjoy Your State of Emergency* (NGBK, Βερολίνο, 2015), 1st NSK Biennial of Folk Art (2014), *Hell as Pavilion* (Palais de Tokyo, 2013), *Truth Is Concrete* (Steirischer Herbst, Γκρατς, 2012), *Monodrome* (3η Μπιενάλε της Αθήνας, 2011), *Media Impact* (4η Μπιενάλε της Μόσχας, 2011) κ.ά.

Luba Elliott

Η Luba Elliott είναι επιμελήτρια και ερευνήτρια με ειδίκευση στη δημιουργική τεχνητή νοημοσύνη. Από το 2016 δραστηριοποιείται συστηματικά στην εκπαίδευση και την ευαισθητοποίηση του ευρύτερου κοινού γύρω από τις εξελίξεις στην τέχνη με τεχνητή νοημοσύνη, μέσα από ομιλίες, εκδηλώσεις και εκθέσεις. Το έργο της έχει παρουσιαστεί σε χώρους του ευρύτερου πεδίου της τέχνης, των επιχειρήσεων και της τεχνολογίας, όπως οι Serpentine Galleries, το Μουσείο V&A, ο οίκος Sotheby's, ο οργανισμός The Photographers' Gallery, το Impakt Festival, το ZKM Karlsruhe, το CogX, καθώς και τα

διεθνή συνέδρια NeurIPS και CVPR. Στα πρότζεκτ που έχει επιμεληθεί περιλαμβάνονται το ART-AI Festival στο Λέστερ, καθώς και οι διαδικτυακές γκαλερί thecvf-art.com, aiartonline.com και computervisionart.com. Στο παρελθόν εργάστηκε σε νεοφυείς επιχειρήσεις (start-ups), μεταξύ των οποίων και η βάση δεδομένων συλλεκτών τέχνης Larry's List. Έχει πτυχίο στις σύγχρονες γλώσσες από το Πανεπιστήμιο του Κέιμπριτζ.

Δρ. Rosemary Lee
Η Rosemary Lee (γ. 1986, Πρόβιντενς) είναι καλλιτέχνις και ερευνήτρια μέσων, της οποίας το έργο εξετάζει πώς οι σύγχρονες συζητήσεις γύρω από την τέχνη και την τεχνολογία επηρεάζονται από αντίστοιχες συζητήσεις του παρελθόντος. Οι εγκαταστάσεις της συχνά παρουσιάζουν εναλλακτικά χρονοδιαγράμματα για τις τεχνολογίες, προβάλλοντας μια μετα-φυσική (post-natural) ιστορία που περιλαμβάνει τα ηλεκτρονικά μέσα. Το έργο της Rosemary Lee έχει παρουσιαστεί διεθνώς σε χώρους που σχετίζονται με την τέχνη και την τεχνολογία, όπως οι osmo/za, RIXC, gnration, NeMe, V2_, Ars Electronica, transmediale και HMKW. Η καλλιτεχνική της έρευνα έχει υποστηριχθεί μέσω υποτροφιών που χρηματοδοτήθηκαν από την Akademie Schloss Solitude, την European Media Art Platform, την Εθνική Βιβλιοθήκη της Αυστρίας, το Danish Arts Foundation, το Copenhagen Culture Harbor 365, το Arts Council Norway, το Nordic Culture Point, το Burning Athens και το πρόγραμμα φιλοξενίας Vilém Flusser Archive Residency for Artistic Research. Από το 2013 εκπροσωπείται από τη Galerie Gilla Lörcher στο Βερολίνο. Η Lee ολοκλήρωσε το διδακτορικό της στο IT University of Copenhagen το 2020, με διατριβή με τίτλο «Machine Learning and Notions of the Image». Το βιβλίο της *Algorithm, Image, Art* (2024) επεκτείνει αυτή την έρευνα. Η Lee διδάσκει στο πρόγραμμα Πολυμέσων της Σχολής Μηχανικών του Πανεπιστημίου του Πόρτο.

ART AND AI
FIVE CRITICAL PERSPECTIVES

This publication brings together five new, commissioned essays that explore the relationship between artificial intelligence and art from different theoretical and artistic perspectives. Focusing on the potential for intervention, the contradictions of technological appropriation, online cultures, algorithmic bias, and authenticity, the authors examine how art engages with computational systems: as a field of experimentation, a method of critique, a practice of resistance, and a process of renegotiating human agency.

VEKTOR is a non-profit arts organisation based in Athens, Greece, exploring the relationship between art and technology.

Η έκδοση συγκεντρώνει πέντε νέα κείμενα, γραμμένα κατόπιν ανάθεσης, που προσεγγίζουν τη σχέση της τεχνητής νοημοσύνης με την τέχνη από διαφορετικές θεωρητικές και καλλιτεχνικές αφετηρίες. Εστιάζοντας στις δυνατότητες παρέμβασης, στις αντιφάσεις της τεχνολογικής οικειοποίησης, στις διαδικτυακές κουλτούρες, στην αλγοριθμική μεροληψία και στην αυθεντικότητα, οι συγγραφείς εξετάζουν πώς η τέχνη εμπλέκεται με τα υπολογιστικά συστήματα: ως πεδίο πειραματισμού, ως μέθοδος κριτικής, ως πρακτική αντίστασης και ως διαδικασία επαναδιαπραγμάτευσης της ανθρώπινης δράσης.

Η VEKTOR είναι ένας μη κερδοσκοπικός οργανισμός τέχνης με έδρα την Αθήνα, που διερευνά τη σχέση της τέχνης με την τεχνολογία.

VEKTOR