ROUTLEDGE FOCUS

Societies at an AI Crossroads

Choices and Value Conflicts

IGNAS KALPOKAS
AND JULIJA KALPOKIENĖ

Societies at an AI Crossroads

This book highlights some of the crucial crossroads – moments in which choices as to the future relationship between humans and digital technologies have to be made – that societies face in light of the growing development and adoption of AI.

As contemporary societies undergo profound transformations driven by artificial intelligence and related technologies, a unilinear vision of inevitable 'progress' and development increasingly dominates public discourse. This book offers a counterpoint, challenging such deterministic narratives by mapping out the complex crossroads where futures ought to be shaped rather than passively received. Exploring a multiplicity of domains as diverse as education, heritage, creativity, the biological and technological boundaries of humans, work, war, and space colonialism, both the dominant (typically progressivist) and alternative (critical) ways of framing the present and the future are outlined, identifying choices to be made. The book teases out the underlying tendencies and underscores a recurring necessity to make value choices as to our further development, not only as societies but also as humanity, and in terms of regulatory choices. It reveals how decisions in these domains are shaped by ingrained intellectual, ideological, political, value-based, and economic assumptions, often hidden beneath the rhetoric of inevitability. These crossroads are not just technical or economic choices, but moral and societal: Between maximising efficiency and safeguarding plurality; between treating technological possibilities as imperatives and critically evaluating their desirability; between moral blindness and care. By illuminating the options available, the book invites readers to take an active role in choosing the future shape of societies and of humanity itself.

This interdisciplinary and creative text will benefit all graduate-level and scholarly readers working on AI within the fields of media and communication, cultural studies, sociology, and cultural anthropology.

Ignas Kalpokas is an Associate Professor and Senior Researcher at Vytautas Magnus University, Lithuania, where he also leads the MA Future Media and Journalism. His research focuses on the societal, cultural, and political impact of emerging digital technologies, political communication, post-truth and fake news, and media theory.

Julija Kalpokienė is a Lithuanian qualified attorney and Junior Research Fellow at Vytautas Magnus University, Lithuania. Her research focuses on internet regulation, cybercrime, and technology regulation, with a particular focus on AI and creativity, and more generally, intellectual property and privacy law, and AI in education and culture.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Routledge Focus on Digital Media and Culture

Media Piracy in the Cultural Economy
Intellectual Property and Labor Under Neoliberal Restructuring
Gavin Mueller

Mobilizing the Latinx Vote
Media, Identity, and Politics
Arthur D. Soto-Vásquez

Playlisting
Collecting Music, Remediated
Onur Sesigür

Understanding Reddit
Elliot T. Panek

Algorithms and Subjectivity
The Subversion of Critical Knowledge
Eran Fisher

TikTok Cultures in the United States
Edited by Trevor Boffone

Cypherpunk Ethics
Radical Ethics for the Digital Age
Patrick D. Anderson

Esports and the Media
Challenges and Expectations in a Multi-Screen Society
Edited by Angel Torres-Toukoumidis

AI Aesthetics
AI-Generated Images between Artistics and Aisthetics
Edited by Jan-Noël Thon and Lukas R.A. Wilde

Societies at an AI Crossroads
Choices and Value Conflicts
Ignas Kalpokas and Julija Kalpokienė

Societies at an AI Crossroads

Choices and Value Conflicts

Ignas Kalpokas and
Julija Kalpokienė

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2026
by Routledge
4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-041-20266-0 (hbk)
ISBN: 978-1-041-20267-7 (pbk)
ISBN: 978-1-003-71584-9 (ebk)

DOI: 10.4324/9781003715849

Contents

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1 A One-Way Street or Crossroads
On Thinking Otherwise

Progress for Its Own Sake?

Technology is neither self-sufficient nor self-determining but, instead, a system that 'includes not only scientific knowledge, machinery, and rules but also non-human living things, as well as human inventors, operators, and maintainers of the hardware and living things' (Dusek 2023: 11). Humans occupy a key place here, not because of some anthropocentric primacy and superiority but for the simple reason that it is human intentionality that drives the decision to proceed with some technological developments over others and pursue some ends over others. Consequently, there is an onus on the creators of new technologies to consider the impact of their work and ensure that positive effects are maximised while harms are minimised (Morrow 2023: 24). While the preceding might seem uncontroversial, the devil, typically, is in the details because what constitutes positive and negative outcomes is a value-based question.

The same applies to complex technological systems, such as AI – they are, in fact, 'inextricably value-laden', for example, in terms of prioritising certain tasks and outcomes over others (Biddle 2023: 134). The reasons for this are both internal to technology (it is difficult to imagine purposeless technology, particularly one that requires time and financial investment) and social, whereby if there are inequalities and other differences, they are bound to be reflected in data and, consequently, in machine learning models (Simons 2023: 29). However, AI tools differ from other forms of decision-making due to their scale and serialisation: As they make decisions at machine speed and often in inscrutable ways, any biases, errors, and omissions will be repeated accordingly (Simons 2023: 30–31). This includes decisions that are politically, culturally, or socially constitutive.

Despite the value judgements and even political decisions that come into play, 'the predictions generated by machine learning models all too quickly come to feel inevitable, natural, beyond our power to control', with the underlying intentionality being easily obfuscated (Simons 2023: 30–31). Hence, AI affects the decision-making autonomy of individuals

DOI: 10.4324/9781003715849-1

(e.g., by structuring the information environment, such as content selection on social media or integrating information through large language models) and the capacity for meaningful control (e.g., by removing algorithmic decisions from contestation due to their inscrutability and perceived authority) (Jungherr 2023). Indeed, once unleashed on big data, AI 'has its own dynamics, it does its own repair and its own optimisation', leaving few opportunities to comprehend why both the right and wrong results turned out to be so (Pearl 2020: 15). Likewise, individual considerations and inputs easily get lost within the flow of data. The individual and their behaviour are unimportant as long as we can make predictions based on people who are 'like' them.

What dominates is the discourse of efficiency, datafication, and 'flattening' of the world that could lead to digital-colonial modalities of control, whereby the world is seen as exploitable and mouldable at will (Brodie 2023: 6–7). Such moulding capacity is premised upon algorithmic predictions so that the future can be not only foreseen but also brought about. The latter works by identifying and enacting nudges and other alterations to encourage preferable futures or even make them quasi-inevitable (of course, 'preferable' by those doing the moulding). Such projections of the future often become self-fulfilling prophecies or 'performative predictions' that end up influencing human actions (Simons 2023: 33).

The predictions themselves are built upon a panoptical view of the world whereby processes of data collection at least theoretically give direct accessibility to everyday human life (Hepp and Couldry 2023: 144). Nevertheless, there is also a strong paradox inherent in data-based AI predictions: While ostensibly future-oriented, such predictions 'can *only* be based on the past', trapping us in what has already happened: Repeating the past into the future, indefinitely (Simons 2023: 216). Nevertheless, algorithmic predictions act as cognitive prostheses, seemingly reminding or even informing us about our 'true' interests and intentions (Galison 2020: 235). Simultaneously, the size and omnivorousness of datasets used for AI training and prediction-making means that minority views and emerging approaches are unavoidably diluted, resulting in the reproduction of dominant perspectives, narratives, and identities (Gillespie 2024). We end up entangled in an amalgamation of the pasts of multiple others, all pointing towards our quasi-inevitable future that is sold as a frictionless path towards value and experience optimisation.

Here, again, we must return to the responsibility to bring about positive outcomes – and the subjective nature of what counts as 'positive'. It is easy to think about progress in a linear fashion – one-sided improvement. For example, the spread and growing accessibility of new technologies can, thus, be expected to 'offer extraordinary new medical advances and clean energy breakthroughs, creating not just new businesses but new industries and quality of life improvements in almost every imaginable area'

(Suleyman 2023: 9–10). This is not an isolated view – there seems to be a permanent demand for new reinventions of digital technologies, leading to even new technological eras, supposedly bound to bring increased prosperity and emancipation for all, again, as a supposed continuation of a history allegedly consisting of an unavoidable series of leaps towards ever greater progress (see, e.g., Dixon 2024). Under such circumstances, one can only be waiting expectantly for a new technological leap that would change human lives, social organisation, and the very understanding of what constitutes the human – the next step in the progressive development of humanity (Runciman 2023). Again, technological development is equated with progress conceived in the broadest possible sense, for example, by claiming that '[a]s a result of our inventions, we become longer lived, richer, better educated, healthier, and perhaps happier too' (Runciman 2023: 1). Under this view, as Suleyman (2023: 25) explicitly claims, '[t]echnology has a clear, inevitable trajectory' – diffusion in great waves that supposedly bring unprecedented improvements to society.

Such thinking also means that efficiency trumps other values and considerations in decision-making (hence, it becomes a 'positive' outcome par excellence). It also precludes a human dimension from playing a role: After all, as Kissinger, Schmidt, and Huttenlocher (2022: 23) claim, once AI and other digital technologies begin to outperform humans in a particular task, failure to automate and remove humans from the loop 'may appear increasingly as perverse or even negligent'. Part of the problem is that, in case of negative externalities, humans may not necessarily be *unable* to intervene – they may be *unwilling* to do so or perhaps unwilling to even entertain the idea of intervention being in any way desirable. After all, a one-sided focus on progress means that 'those who challenge the origins, trajectory, and pace of AI evolution' by expressing critical views or questioning the fundamentality of maximum efficiency 'are routinely being harassed and accused of standing in the way of (technological) progress' (Tacheva and Ramasubramanian 2023: 1). Paradoxically, however, the willingness and the ability to question any status quo and to hold those in any form of power to account must be part and parcel of democratic societies, rather than mere reliance on the assumption that technology will simply be built for the better. In this way, the democratic responsibilities of citizens may not necessarily be denied – they may be forfeited by the citizens themselves as they become unwilling to object to the apparent consensus around technological progress. In the parlance of this book, citizens in the societies of today and the near future may be unwilling to even entertain the idea of there being a crossroads in which a meaningful choice between different paths (and, therefore, different directions) could be made.

Dominant decision-making and organisational models are further entrenched through platforms as the dominant mode of organisation

within the digital domain. Platforms establish 'a globalized model of largely private governance with a history in which a few states, and a few platforms, are dominant', largely building on top of already existing power structures; in this way, platforms emerge as key actors that 'alter or retrench international structures of production and labour, security, finance and knowledge' (Bannerman 2024: 1817–1818). They are, or seek to become, seemingly frictionless intermediaries that set up architectures of interaction to maximise data extraction and utilisation (Brubaker 2023). For the most part, platforms function as 'private regulators of their ecosystems' that, as part of their core operations, 'establish the rules through which their various users (be they individuals or organisations) interact, decide what behaviours to encourage or discourage on the platform, and choose how to enforce them' (Gawer 2022: 114). The modus operandi of such platforms consists of datafication and the luring of users into 'walled gardens' based primarily on convenience – removing friction from everything – whether that is shopping, social interactions, or any other process (Brubaker 2023: 14–15). Crucially, as Susskind (2018: 124) claims, playing on a classical Marxist theme, power now lies in 'control of the means of scrutiny', with platform companies (and dominant technology companies more broadly) becoming such owners par excellence.

However, significant accountability gaps are opened through the creeping integration of technology, whereby '[w]ithout significant fanfare – or even visibility – we are integrating nonhuman intelligence into the basic fabric of human activity' – a logic that, in many ways, remains inscrutable (Kissinger, Schmidt, and Huttenlocher 2022: 94, 107). The power of not only the technologies themselves (insofar as they cannot be considered to have independent agency) but also of those in charge of datafication, analytics, and implementation silently permeates today's societies without even a realistic possibility of a meaningful citizen response. Simultaneously, the inscrutable and unaccountable digital domain becomes a sine qua non for exercising citizenship, that is, the quintessential capacity to hold power accountable (Smyrnaios and Baisnée 2023: 436). The connection between digital technologies and citizenship comes primarily from the centrality of online tools and platforms in the exchange of information, identity formation, communication, and, generally, life-planning functions carried out through them. Largely, then, citizen capacity to hold power to account becomes self-invalidated as the output (political action) is based on inscrutable input resulting from the algorithmic architecture of the public sphere, subsequently taking on increasingly non-transparent pathways (optimisations and nudges within data analytics and AI), resulting in a *double* 'black box' problem, to update Pasquale's (2015) original term.

Public/political spaces are also increasingly digital, meaning that they depend on coded architecture in their functioning, while the spread of content (both honest and manipulated) is shaped by the algorithmic logics

of popularity and personalised targeting, thereby further entangling human social/political agency with the functioning of digital code (Jungherr 2023). Likewise, the functioning of digital spaces is not self-originating: instead, the power of technology companies is amplified significantly (Jungherr 2023). Such power is further entrenched through a 'shift from law enforced by *people* to law enforced by *technology*', meaning that 'power will increasingly lie in *force* rather than *coercion*, with self-enforcing laws that cannot be broken because they are encoded into the world around us' (Susskind 2018: 105) – in other words, algorithmic governance. Hence, traditional modes of governance are going to be supplemented (and perhaps even partly replaced) with '[d]igital law, privatized force, autonomous systems of power' (Susskind 2018: 121). Likewise, the growing automation of everyday life means that an absence of checks and balances emerges in governing the public sphere (Coeckelbergh 2022: 82–83). Again, the capacity to notice a crossroads – dilemmas and alternatives – let alone the ability to take alternative paths is both constrained and rendered contingent upon algorithmic structuring of the world. After all, dilemmas and crossroads do not go well with the emphasis on frictionlessness and process optimisation in today's technologised society – instead, crossroads are paradigmatic examples of friction.

Crucially, individuals become enmeshed with algorithms and data. On the one hand, the algorithm becomes not only a force to be reckoned with but also something to be pleased with since it holds keys to one's own digital presence (see, e.g., Haapoja et al. 2024). On the other hand, it is simply a tool, and the visibility, value, and influence of individuals depend on how strongly datafied they are (Jungherr 2023). Indeed, algorithmic forms of governance can only emerge, promising to perfectly match our lives with our environment, as long as we actively cultivate our data profiles (Burgess et al. 2022: 50). Just for the sake of being able to adequately function in increasingly digital-first societies, individuals must become part of 'an ever-tighter web of algorithmic processes that objectify, produce, regulate, and govern the self' (Brubaker 2023: 47). At the heart of this process, however, we can again find a coded insignificance of the particular and its subsumption under a generalised logic of data-fuelled progress, optimisation, and efficiency maximisation. Thus, culture, interactions, and life itself become content, resulting in a loss of distinctiveness within the networks of consumption and/as datafication (Brubaker 2023).

Observing the dynamic of digital technologies, it becomes evident that the key force underlying this domain is a drive to *collect* all the data potentially available and to *connect* more individuals and parts of the world to global networks to unlock even more data for collection (Curran 2023) so that ultimately there is as little outside of such extractive networks (i.e., as little 'waste') as possible. One can easily see this pervasiveness of datafication not only as a mere and neutral extension of knowledge but also

(and perhaps even primarily) as a reconfiguration of power relations with significant power being self-arrogated by (mostly, at least in the West) corporate actors in control of data collection and analytics (see Hepp and Couldry 2023: 143–144). But even here, the enmeshment does not stop, with broader infrastructures, such as 'fiber optic cables, content distribution networks, and telecoms masts as well as data centers' being at the heart of today's life, opening room for digital divides and unequal participation opportunities (Brodie 2023: 3), depending on one's proximity and access to digital infrastructures. After all, as everyday life, politics, economy, and other domains are increasingly digitally oriented, the only thing worse than being pervasively datafied is *not* being datafied.

Notably, Tacheva and Ramasubramanian (2023) use the term 'AI empire' to underscore the exploitative underpinnings of digital technologies, particularly their data- and labour-intensive nature. Similarly, Couldry and Mejias (2023: 787) underscore the extractivist and even colonialist tendencies in today's datafication processes that emerge in the form of 'continuous extraction of economic value from human life through data'. Just like human and natural resources in the colonial era, data are seen as 'just out there', ready for the taking (Couldry and Mejias 2023: 290). Ultimately, the for-profit nature of data extraction and analytics results in 'commercialisation of life' (Adams 2021), that is, the value of each individual and their living conditions being determined by the value of the data they generate. No less forcefully, for Knight (2023), ever larger AI models, such as generative AI tools, become 'a kind of monster that must be fed at all costs', the food, of course, is data. In this way, data extraction becomes its own goal and justification, re-framed as autonomous and plotting its own, seemingly unavoidable, course.

It is easy to forget that data are a relation and not a stable property (Mörtenböck and Mooshammer 2020): rather than revealing universal and essential truths about individuals and/or larger groups, it merely reflects their specific placedness within a broader spectrum of relationships, interactions, and technological affordances. And yet, data essentialism transpires to be the order of the day. However, what happens in practice is the emergence of a recursive society whereby 'data is so integrated that there is no point of origin – everything is already implicated by loop upon loop of data processing', meaning that data analytics end up 'working with data that is at least partly a product of previous analytic steps' (Beer 2022: 2). Moreover, authentic data (even though the term itself is problematic given that data extraction typically happens under specifically manufactured conditions) are increasingly supplemented with AI-generated data, meaning that even the idea of data as an a link between a person and the digital ecosystem can be questioned (see, e.g., Jacobsen 2024). The attractiveness of such a 'synthetic supplement' only illustrates that no amount of data are ever enough, and the more data can be

extracted or generated, allegedly, the better (Jacobsen 2024). Datafication becomes neither a choice nor an inevitability but, instead, an objective fact, completely detached from other circumstances – an illustration of mere technological possibility (of extending datafication beyond what is naturally possible) becoming an imperative.

A characteristic assertion is that 'the benefit of data emerges from its use'; in other words, the triad of datafication, analytics, and application also turns into imperative, allegedly bringing objective benefits to all, so that '[t]he more often we use data, and the more diverse our purposes in doing so, the greater the social and economic value we'll reap' (Mayer-Schönberger and Ramge 2022: 17). Anything on the flip side of efficiency maximisation of/through value generation becomes mere waste (Munn 2023), to be avoided at all costs. Consequently, the (qualitative and quantitative) spread of digital technologies becomes a good in and of itself, supposedly geared towards improvement on both individual and societal levels.

But What About Externalities?

Other oversights are present as well. For example, focus on the digital and the immaterial obfuscates the environmental harm caused by the exploitation of resources, demand for energy, and ever-growing waste (see, e.g., Kannengießer and McCurdy 2021). As Lehuedé (2025: 1763) observes, 'even well-intentioned "AI for good" initiatives have been proven to yield short-term benefits at the expense of the root causes underlying social and environmental injustices': While some problems may be solved, new harms are caused along the way. This problem is further exacerbated by the fact that 'discussions of the utility of AI are increasingly held in a vacuum […] disassociated from its costs, and from the extractive logics behind the technologies', which precludes any meaningful discussions about the proportionality and cost–benefit analysis of AI applications (Katirai 2024: 534; see also Moyano-Fernández and Rueda 2023: 219; Liu and Yin 2024: 2). Consequently, there is a tendency to overlook AI's impact on 'the environment and the planet, and the intense pressure it creates on limited resources and energy consumption' (Markelius et al. 2024: 728), with notions of progress and operational efficiency trumping other considerations (Meinhold, Wagner and Dhar 2025: 2323). The preceding is further enabled by 'the outsourcing of environmental damage to peripheral and semi-peripheral areas' (Lehuedé 2025: 1763), e.g., building data centres in deprived areas or unleashing pollution on already marginalised populations.

Nevertheless, the very meaning of sustainability (including AI) must be put into question, particularly with regard to greenwashing and ethics-washing, the primary goal being increasing political and public

acceptability of the matter at hand (Moyano-Fernández and Rueda 2023: 221). For example, while some authors (Dua, Singh, and Shehu 2025: 3809) treat the inclusion of ecological factors in national AI strategies as a sufficient 'commitment to sustainable practices, aligning with broader goals of environmental responsibility', a question arises whether that is enough or merely lip service. Moreover, such narrow emphasis reduces ethical concerns to mere analysis of possibilities 'to train high-performance, environmentally sustainable AI models without sacrificing their effectiveness' (Liu and Yin 2024: 2), thereby turning any such engagement into merely a puzzle to be solved in order to retain what is seen as technological progress.

Another domain where the promise of efficiency and novel solutions often precludes sufficient concern for externalities is healthcare. Here, concerns pertain to, for example, data security, whereby AI's reliance on immense datasets for, e.g., diagnostics also implies it being trained on highly sensitive patient information, making both ethical use of data and the security of individual details key imperatives (Mondal and Mondal 2024: 250; Ning et al. 2024; Shuaib 2024; van Leersum and Maathuis 2025). However, while the preceding concerns are of a more technical and procedural nature, even more fundamental ethical issues abound. Notably, as in many domains of AI, these include matters of algorithmic bias, particularly as medical decision-making and diagnostics are increasingly permeated with AI, 'it is imperative to address biases that may arise in algorithms to ensure equitable and unbiased healthcare outcomes for all individuals' and, therefore, preclude 'disparities in diagnoses, treatment recommendations, and overall patient care' (Mondal and Mondal 2024: 256; Shuaib 2024). Fairness and diversity in data (van Leersum and Maathuis 2025: 4), as well as equity (Ning et al. 2024), thus become primary matters so that less datafied groups and conditions are not condemned to inferior outcomes.

A related set of issues revolves around human decision-making, responsibility, and patient-doctor trust (Mondal and Mondal 2024: 263; Ning et al. 2024; Shuaib 2024; van Leersum and Maathuis 2025). This includes the explainability of treatment decisions (to what extent AI recommendations are transparent, not just to patients but to medical practitioners as well), human autonomy in decision-making, and, ultimately, the question of who bears responsibility if anything goes wrong. Matters are further complicated by social issues, such as disparities in accessibility, whereby '[u]nequal access to AI-driven healthcare solutions could widen existing healthcare disparities, exacerbating inequalities between different socioeconomic groups and geographic regions' (Mondal and Mondal 2024: 262), leading to a two-speed (or even more stratified) humanity. Additionally, it must be kept in mind that '[w]hile AI machines are highly capable of processing and providing information and analysis, they lack

an embodied, practical relationship to the world as a common place to live one's life in all its forms and practices' (Heuser, Steil and Salloch 2025: 10), meaning that addressing the above concerns solely as technological problems may be structurally impossible.

As for further considerations, there are also concerns of the increasing focus on AI contributing to 'de-skilling and privatisation of healthcare' while treating AI as a silver bullet, likely 'distracts attention from the societal and political-economic structures underlying the problems we claim AI can solve, as well as the normative values contained with its design' (Strange 2024: 838). Unsurprisingly, then, some would argue for a necessary quintessentially human role to guarantee that AI is developed and applied ethically (Iniesta 2025). However, a question remains as to whether that would be feasible considering the push for ever greater efficiency and struggles for AI 'dominance'.

Still, even a refocusing of attention might not help, as any harm might be written off as a trade-off for progress and a better future. No less prominently, as shown later in this book, data extractivism and resource extractivism transpire to be merely two sides of the same coin, following the supposed imperative whereby 'can' equals 'must'. Once again, crossroads (let alone the alternative pathways available) may simply fail to be noted.

The Evil of Omitting Moral Considerations

To better understand the fundamental importance of crossroads and the consequences of failing to even consider them, we build on the concepts of adiaphorisation and liquid evil, developed by Zygmunt Bauman and Leonidas Donskis. To begin with, part and parcel of the present problem is that, as Bauman and Donskis (2013: 5) already noted a while ago, 'technology has surpassed politics, the latter having in part become a supplement to technology'; the 'technological society' that consequently ensues is notable for manifesting a 'determinist consciousness' and treating 'a refusal to participate in the technological innovations […] as sufficient grounds to remove all those who lag behind […] to the margins of society'. Crucially, technology taking over control of society from politics (the latter being a matter of competition between alternatives) results in a 'moral vacuum' (Bauman and Donskis 2016: 11). If there is only one predetermined direction (that of technological and, through it, societal progress), then choice is not only superfluous – it does not exist. The moral vacuum that Bauman and Donskis lament is created precisely by the failure to even consider the possibility of crossroads. Crossroads, in turn, become a locus of facing moral obligations – after all, choosing from among alternatives only becomes possible by way of prioritising some values over others and considering one's obligations to those affected by either decision.

The determinism inherent in a technologised society not only fails to recognise any alternatives – in fact, it treats its own linearity as an imperative. In this way, the mere possibility of a new technological advancement becomes not only an inevitability but also a moral duty. As Bauman and Donskis (2013: 5) observe, '*I can* transmutes into *I must*. I can, therefore I am obliged to. No dilemmas allowed. We live in a reality of possibilities, not one of dilemmas'. This equation of possibility and imperative (or duty) will be seen repeating itself throughout the different manifestations of technology discussed in this book. However, as Bauman and Donskis (2013: 15) stress when further elaborating on linearity's perilous consequences on morality, living in a deterministic society devoid of dilemmas results in adiaphorisation – 'exemption from the realm of moral evaluation'. In other words, individual (and other, e.g., environmental) consequences become insignificant, unworthy of consideration in the grand scheme of things. It is this loss of care and concern that Bauman and Donskis refer to in the title of one of their co-authored books – *Moral Blindness*. Following Bauman and Donskis's (2013: 37) exegesis, '*adiaphoron* is a temporary withdrawal from one's own sensitivity zone', that is, failure to react, to care about someone or something as if they were completely detached from one's moral universe. This exclusion from moral obligations may masquerade as neutrality, but instead, it merely denotes indifference (Bauman and Donskis 2013: 40; see also Bauman and Donskis 2016: 50). Such ignored objects that fall outside the linearity of progress become 'a waste of instrumental rationality and rational calculation – not so much resented or hated as simply unnoted and left out of account' (Bauman and Donskis 2016: 50). The only thing that matters is an elusive goal, a technological pie in the sky, usually described in terms of efficiency, optimisation, and frictionlessness, while anything else has no value and can be ignored or even sacrificed.

An integral part of this society, following Bauman and Donskis (2013: 15), is a consumerist attitude, whereby we commit only as long as something or someone serves our needs more efficiently than other alternatives – an attitude that 'may lubricate the wheels of the economy [but] sprinkles sand into the bearings of morality'. This includes, considering today's technology, the drive towards platformisation and data-enabled frictionlessness of experience or endless extensions of one's otherwise humanly limited capacities. Fittingly, Bauman and Donskis (2013: 16) assert that '[a]s moral negligence grows in its reach and intensity, the demand for painkillers rises unstoppably', ultimately turning into addiction (Bauman and Donskis 2013: 16). In today's increasingly AI-centric society, such painkillers are primarily of efficiency-maximising and friction-removing kind, the latter promised to follow if only the technological advancement imperative is upheld. These developments are seen to bring about supposed immortality through self-renewal (including of a technologically enabled kind) and

enable the pursuit of immediate satisfaction, which obstructs the capacity to think about long-term consequences, including those of the technologies used (Bauman and Donskis 2013: 202–204). Likewise, freedom becomes a matter of 'consumption and self-renewal' while having 'lost any connection with the most important thing: believing that you can change something in the world' (Bauman and Donskis 2016: 8). The very thought of change being possible or desirable (i.e., of crossroads) is, as already shown, one of the first victims of a technologised progress-minded society.

Uselessness and inefficiency (including any attempt to stray away from the path towards supposed progress) become the most resented features in a technologised society that has become prone to adiaphorisation: As Bauman and Donskis (2016: 51) stress, '[e]xclusion, or the prospect and sanction of exclusion, on the ground of uselessness […] is on course to leave behind and dwarf other manifestations of evil'; in fact, '[i]t tends to become evil's *principal playground*, and evil's most widely and commonly deployed *weapon*'. This brings us to yet another contribution of Bauman and Donskis: Their reframing of evil. Following Bauman and Donskis (2013: 36), '[e]vil lurks in what we tend to take as normality and even as the triviality and banality of mundane life, rather than in abnormal cases, pathologies, aberrations and the like'. While a conventional interpretation would be that evil lies in dilemmas and alternatives (some of such choices being inherently 'good', others inherently 'bad'), evil today seems to be considerably more banal – it consists of inability to *see* choices, alternatives and dilemmas, mundanely following one possibility and alleged improvement after the other without pausing for consideration and without pondering the value of those sidelined or harmed.

Extending Bauman's earlier work on liquid modernity, Bauman and Donskis (2016: 3) call the above phenomenon liquid evil, which 'assumes the appearance of goodness and love' by way of recasting into centre stage 'a seemingly neutral and impartial acceleration of life' that simply cannot be resisted and has no alternatives; in this way, '[a] citizen becomes a consumer, and value-neutrality hides the fact of disengagement', resulting in 'moral amnesia' (Bauman and Donskis 2016: 3). This 'liquidisation' of evil makes 'resistance to evil more difficult and active participation in its reproduction more likely' simply because evil has become more difficult to define and identify – in many cases, it has become woven into the very fabric of progress-focused society (Bauman and Donskis 2016: 36).

Consumerist attitudes, dominant in liquid modernity and, specifically, its digitally technologised form, give rise to 'another time and place: discontinuous pointillist time', which 'just like pointillism in painting, makes the momentary impression or state into a more real thing than do long-term projects, history, the classical canon, and the past' (Bauman and Donskis 2013: 112). Reinterpreting this from today's perspective, in an AI-centric society, pointillist time manifests itself in jumps from one technological

advancement and product release to the next, all the way retaining the illusion of a progression towards ever greater good. Once again, focus on leap after leap relieves one from the need to notice the factors that had come into play to enable a new leap, as well as the sacrifices (in terms of material impact but also in terms of values that had to be foregone). We must return to the importance of crossroads because failure to notice them *is* moral blindness (adiaphorisation) par excellence, while mindless progression along a straight, seemingly dilemma-free path (the perceived frictionlessness of progress) lies at the heart of liquid evil. Crucially, these are features that will return, time and time again, throughout the course of this book.

References

Adams, R. (2021). Can Artificial Intelligence Be Decolonized? *Interdisciplinary Science Reviews*, 46(1–2), 176–197.

Bannerman, S. (2024). Platform Imperialism, Communications Law and Relational Sovereignty. *New Media & Society*, 26(4), 1816–1833.

Bauman, Z. and Donskis, L. (2013). *Moral Blindness: The Loss of Sensitivity in Liquid Modernity*. Cambridge: Polity.

Bauman, Z. and Donskis, L. (2016). *Liquid Evil*. Cambridge: Polity.

Beer, D. (2022). The Problem of Researching a Recursive Society: Algorithms, Data Coils and the Looping of the Social. *Big Data & Society*, doi: 10.1177/20539517221104997

Biddle, J.B. (2023). Values in Artificial Intelligence Systems. In G.J. Robson and J.Y. Tsou (eds.) *Technology Ethics: A Philosophical Introduction and Readings*. London: Routledge, pp. 132–140.

Brodie, P. (2023). Data Infrastructure Studies on an Unequal Planet. *Big Data & Society*, doi: 10.1177/20539517231182402

Brubaker, R. (2023). *Hyperconnectivity and Its Discontents*. Cambridge: Polity.

Burgess, J. et al. (2022). *Everyday Data Cultures*. Cambridge: Polity.

Coeckelbergh, M. (2022). *The Political Philosophy of AI*. Cambridge: Polity Press.

Couldry, N. and Mejias, U.A. (2023). The Decolonial Turn in Data and Technology Research: What Is at Stake and Where Is It Heading? *Information, Communication & Society*, 36(4), 786–802.

Curran, D. (2023). Surveillance Capitalism and Systemic Digital Risk: The Imperative to Collect and Connect and the Risks of Interconnectedness. *Big Data & Society*, doi: 10.1177/20539517231177621

Dixon, C. (2024). *Read Write Own: Building the News Era of the Internet*. New York: Random House.

Dua, M., Singh, J.P. and Shehu, A. (2025). The Ethics of National Artificial Intelligence Plans: An Empirical Lens. *AI Ethics*, 5, 3803–3831.

Dusek, V. (2023). The Definition of Technology. In G.J. Robson and J.Y. Tsou (eds.) *Technology Ethics: A Philosophical Introduction and Readings*. London: Routledge, pp. 5–13.

Galison, P. (2020). Algorists Dream of Objectivity. In J. Brockman (ed.) *Possible Minds: 25 Ways of Looking at AI*. London: Penguin, pp. 231–239.
Gawer, A. (2022). Digital Platforms and Ecosystems: Remarks on the Dominant Organizational Forms of the Digital Age. *Innovation: Organization and Management*, 24(1), 110–124.
Gillespie, T. (2024). Generative AI and the Politics of Visibility. *Big Data & Society*, doi: 10.1177/20539517241252131
Haapoja, J. et al. (2024). Moral Orders of Pleasing the Algorithm. *New Media & Society*, doi: 10.1177/14614448241278674
Hepp, A. and Couldry, N. (2023). Necessary Entanglements: Reflections on the Role of 'Materialist Phenomenology' in Researching Deep Mediatization and Datafication. *Sociologica*, 17(1), 137–153.
Heuser, S., Steil, J. and Salloch, S. (2025). AI Ethics Beyond Principles: Strengthening the Life-world Perspective. *Science and Engineering Ethics*, 31, 1–14.
Iniesta, R. (2025). The Human Role to Guarantee an Ethical AI in Healthcare: A Five-Facts Approach. *AI Ethics*, 5, 385–397.
Jacobsen, B.N. (2024). The Logic of the Synthetic Supplement in Algorithmic Societies. *Theory, Culture & Society*, 41(4), 41–56.
Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, doi: 10.1177/20563051231186353
Kannengießer, S. and McCurdy, P. (2021). Mediatization and the Absence of the Environment. *Communication Theory*, 31, 911–931.
Katirai, A. (2024). The Environmental Costs of Artificial Intelligence for Healthcare. *Asian Bioethics Review*, 16, 527–538.
Kissinger, H., Schmidt, E. and Huttenlocher, D. (2022). *The Age of AI*. London: John Murray.
Knight, W. (2023, August 10). Generative AI Is Making Companies Even More Thirsty for Your Data. *Wired*, https://www.wired.com/story/fast-forward-generative-ai-companies-thirsty-for-your-data/
Lehuedé, S. (2025). An Elemental Ethics for Artificial Intelligence: Water as Resistance Within AI's Value Chain. *AI & Society*, 40, 1761–1774.
Liu, V. and Yin, Y. (2024). Green AI: Exploring Carbon Footprints, Mitigation Strategies, and Trade-offs in Large Language Model Training. *Discover Artificial Intelligence*, 4, 1–12.
Markelius, A. et al. (2024). The Mechanisms of AI Hype and Its Planetary and Social Costs. *AI Ethics*, 4, 727–742.
Mayer-Schönberger, V. and Ramge, T. (2022). *Access Rules: Freeing Data from Big Tech for a Better Future*. Oakland: University of California Press.
Meinhold, R., Wagner, C. and Dhar, B.K. (2025). Digital Sustainability and Eco-Environmental Sustainability: A Review of Emerging Technologies, Resource Challenges, and Policy Implications. *Sustainable Development*, 33, 2323–2338.
Mondal, H. and Mondal, S. (2024). Ethical and Social Issues Related to AI in Healthcare. *Methods in Microbiology*, 55, 247–281.

Morrow, D. (2023). The Values Built into Technologies. In G.J. Robson and J.Y. Tsou (eds.) *Technology Ethics: A Philosophical Introduction and Readings*. London: Routledge, pp. 18–25.

Mörtenböck, P. and Mooshammer, H. (2020). Introduction. In P. Mörtenböck and H. Mooshammer (eds.) *Data Publics: Public Plurality in an Era of Data Determinacy*. London: Routledge, pp. 1–25.

Moyano-Fernández, C. and Rueda, J. (2023). AI, Sustainability, and Environmental Ethics. In F. Lara and J. Deckers (eds.) *Ethics of Artificial Intelligence*. Cham: Springer, pp. 219–236.

Munn, L. (2023). *Countering the Cloud: Thinking With and Against Data Infrastructures*. London: Routledge.

Ning, Y. et al. (2024). Generative Artificial Intelligence and Ethical Considerations in Health Care: A Scoping Review and Ethics Checklist. *Lancet Digit Health*, 6, 1–9.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Pearl, J. (2020). The Limitations of Opaque Learning Machines. In J. Brockman (ed.) *Possible Minds: 25 Ways of Looking at AI*. London: Penguin, pp. 13–19.

Runciman, D. (2023). *The Handover: How We Gave Control of Our Lives to Corporations, States and AIs*. London: Profile Books.

Shuaib, A. (2024). Transforming Healthcare with AI: Promises, Pitfalls, and Pathways Forward. *International Journal of General Medicine*, 17, 1765–1771.

Simons, J. (2023). *Algorithms for the People: Democracy in the Age of AI*. Princeton: Princeton University Press.

Smyrnaios, N. and Baisnée, O. (2023). Critically Understanding the Platformization of the Public Sphere. *European Journal of Communication*, 38(5), 435–445.

Strange, M. (2024). Three Different Types of AI Hype in Healthcare. *AI Ethics*, 4, 833–840.

Suleyman, M. (2023). *The Coming Wave*. New York: Crown.

Susskind, J. (2018). *Future Politics: Living Together in a World Transformed by Tech*. Oxford: Oxford University Press.

Tacheva, J. and Ramasubramanian, S. (2023). AI Empire. Unravelling the Interlocking Systems of Oppression in Generative AI's Global Order. *Big Data & Society*, doi: 10.1177/20539517231219241

van Leersum, C.M. and Maathuis, C. (2025). Human Centred Explainable AI Decision-Making in Healthcare. *Journal of Responsible Technology*, 21, 1–13.

2 Solutionist Imaginaries
AI and Education

AI as a Threat?

When discussing the impact of AI on education, one must start with the elephant in the room: traditional teaching and (especially) assessment methods being severely affected by the emergence of generative AI (Gill et al. 2024). As such, the effects of AI can be best described as disorientation due to earlier methods being put into question, while educators lack knowledge and support in coming up with new ways of teaching and assessing students (Lee et al. 2024). Quite unsurprisingly, this has mostly evoked a reaction of fear and desire to detect and police AI-generated content, particularly in assessments, regardless of the shortcomings of detection software itself. Consequently, it is unsurprising that at least some educators have come to see authorship and academic integrity as facing significant threats or even being in grave danger (Duah and McGivern 2024).

Among the fears of generative AI in education, two stand out, again representing the two key pillars of education: Learning and assessment. In terms of learning, the concern is that the ease of retrieving answers from generative AI models and their seemingly comprehensive nature will lead to overreliance, that is, students resorting to prompting AI instead of learning; meanwhile, on the plagiarism side of things, the obvious issue of AI-enabled plagiarism is at the forefront (Nah et al. 2023). Almost identically, Gefen and Arinze (2023: 3) present generative AI as enabling 'high-tech plagiarism' while 'potentially usurping academic authority' and providing 'a means to avoid education' – as getting answers from AI is so easy, the argument goes, this simply removes the need for understanding any subject at hand. As a result, there are fears of 'a possible decrease in students' cognitive and logical skills and a decline in academic competency' (Yusuf, Pervin, and Román-González 2024: 2), increasing procrastination among students, and a reduction of the ability to recall information (Abbas, Jam, and Khan 2024). As can be expected, some sense of doom and gloom is also not avoided by way of presenting generative AI as

DOI: 10.4324/9781003715849-2

'ruining the tradition of science' and being explicitly anti-Enlightenment: Bringing about a new form of superstition and blind faith in an inscrutable authority (Gefen and Arinze 2023: 7). AI is accused of creating an environment that facilitates academic dishonesty (Williams 2024): Not only is the (especially undeclared) use of generative AI in assessments considered a major educational problem, but the worry is also that further problems await down the line, including a 'serious' risk that 'intentional violations of academic integrity behaviours can continue into the workplace after students complete their studies' (Perkins et al. 2024: 91). Notably, students are almost expected to cheat at any opportunity, and the onus to behave ethically is seen to lie squarely on them, instead of being shared among the wider academic community (see, e.g., Lim et al., 2023).

While the preceding might seem like a strongly techno-pessimist outlook, it also has its own aspects of faith in technology. Particularly, there is hope that detection technology matures sufficiently to be able to reliably identify AI-generated text (Duah and McGivern 2024). To some extent, this approach can be seen as intuitive and even default, mimicking the 'police-catch-punish approach' that educational institutions have been using for simple plagiarism detection (Kramm and McKenna 2023: 2174). However, the preceding is not without criticism: As Kramm and McKenna (2023: 2175) stress, this outlook 'emerges from an instrumentalist understanding of higher education', with the latter's raison d'être being 'to sell qualifications that must be […] safeguarded against forgeries', leading to the delegation of significant power and privilege to technology companies purporting to sell such verification and safeguarding services. In this way, policing generative AI use becomes a way to protect an institution's credibility (Perkins et al. 2024), without more thought being put into the broader context of education, but instead, seeking quick and easy answers through technological solutionism (Kalpokas, Kalpokienė and Šalaševičiūtė 2024).

Beyond the simplistic policing-focused models, restricting the use of generative AI in educational contexts, including assessment, limits students' ability to learn the skills necessary to perform well in the job market; instead, the focus is shifted towards learning how to prompt AI tools and critically discern the generated output (Sidorkin 2024; see also Perkins et al. 2024). Moreover, explicitly including such technologies in the learning process helps build critical thinking and creativity among students (Leelavathi and Surendhranatha 2024) while also increasing ethical awareness and general confidence with regard to AI (Wood and Moss 2024), instead of diminishing such traits. After all, an argument can also be made that since students are already using generative AI in various contexts, including for learning and preparation of assignments, making them *un*learn such existing practices is illogical and counterproductive (Yusuf, Pervin, and Román-González 2024). Hence, this view goes, there is a need to develop student AI use practices, such as prompt engineering,

so that students are in charge rather than relying on mere guesswork and luck (Kim et al. 2025). Likewise, it is expected that the ability to productively use AI tools would make students more competitive in the job market, with the capacity to draw connections and inferences from AI-generated content, and curiosity to further interrogate AI, seen as the new critical skills (Khan 2024: 6).

According to the optimists, as generative AI is not merely a passing phenomenon but a technology that would remain in use in the long term, avoidance and waiting out cannot be a productive strategy; thus, educational institutions are expected to embrace the technology if they want to have any chance of shaping the practices around generative AI use (Jochim and Lenz-Kesekamp 2025). However, this focus on unavoidability is also not without its drawbacks, particularly in its conflation of unavoidability and desirability, thereby glossing over the drawbacks of generative AI, such as concerns about inappropriateness or context-blindness of responses and limitations arising from AI's reliance on online digital sources (Yusuf, Pervin and Román-González 2024: 24). Likewise, one should keep in mind that the very idea of unavoidability (or at least the way it is often framed) is not value-neutral but, instead, a matter of failing to see alternatives, that is, omission of crossroads.

As Cone (2023: 53) stresses, one should not succumb to the view that technologies 'merely offer a digital innovation of a pre-established "need" or "desire"'. Imaginaries of desirable futures in education are first developed by the technological and economic elites, with EdTech (provided by the same elites) subsequently becoming a solution to this pre-designed problem (Linderoth, Hultén, and Stenliden 2024). In addition to the obviously self-serving nature of such imaginaries, one must also enquire whose values constitute the building blocks of these desired futures and to what extent they relate (if at all) to actual lived experiences of the education community. Instead, new releases and added features, as well as in-built capacities, are primarily determined by technological and business competition for platform domination and the capacities (and limitations) of AI, and much less so by the actual needs of learners and teachers (Williamson, Macgilchrist, and Potter 2023).

AI as a Solution?

In terms of ideological thrust, AI in education ties in with the broader tendency towards 'technology-facilitated personal empowerment' (Clark 2024: 421). Likewise, the adoption of AI in education should be seen within the technological context, which involves three important premises: The proliferation of personal devices that learners can use for educational purposes, the growth of data-based understanding of the learning

process, and, finally, the availability of an increasing number of learning technologies (Moleaar 2022: 634). However, all of these assumptions are problematic: The first one ignores the persistence of the digital divide, while the other two gloss over the problems inherent in datafication and analytics by treating those processes as neutral and egalitarian. Moreover, there is also a stigmatising factor involved in the deterministic discourse that focuses on the 'inevitability' of AI in education, with resistance framed as driven merely by biases and misconceptions (Eager and Brunton 2023). Moreover, to further underscore the perceived need for AI, the entire education process is framed as a sequence of potential faults: potential teacher errors, lost or inefficiently used time, etc. (Selwyn, Campbell, and Andrejevic 2023). In this way, AI can be presented as the ultimate problem-solver.

At a minimum, the benefits of AI are seen to lie in the ability to delegate mundane and bureaucratic tasks, such as taking attendance, developing course and class plans to save time for other activities (Nemorin et al. 2023: 39; see also Labadze, Grigolia, and Machaidze 2024). However, the 'real' application of AI in education is mostly seen in terms of designing new teaching strategies and activities (Chiu 2024), allegedly leading towards hybridity of humans and AI in education (Moleaar 2022). While some are ready to admit that this is not necessarily a smooth process, challenges are seen primarily not in the technology but in 'ideological barriers' among both education practitioners and administrators in academic institutions (Baidoo-Anu et al. 2024: 3). Once such challenges are overcome, it is claimed, AI presents 'transformative opportunities', such as 'customized learning, research support, and enhanced problem-solving skills' (Yusuf, Pervin and Román-González 2024: 4). Other oft-repeated benefits include creating engaging learning opportunities, suggesting responses to student questions (Su and Yang 2023), or greater inclusivity by adapting to disabilities and special educational needs (Mao, Chen and Liu 2024).

AI in education often evokes unbounded optimism and endless expected opportunities: As Cope, Kalantzis, and Searsmith (2021: 1234) observe, '[i]n the world of AI-enabled learning, knowledge is limitless', particularly as technological aides, such as computers, vastly exceed the human brain, particularly in memory and speed. The possibilities thereby opened include tailoring the delivery of teaching and level of instruction to individual preferences and learning speed, ensuring timely feedback, etc., or acting as study companions (Labadze, Grigolia, and Machaidze 2024; see also Kooli 2023). Particularly with regard to the latter, virtual interactions, including interactions with simulated virtual learners and teachers, are expected to substitute in-person ones (Ferreira Costa et al. 2023; Käser and Alexandron 2024). Focus shifts towards learning through dialogical communication with AI chatbots, including AI tools capturing

student data from such interactions to measure learning success and generate new tasks and assessments or clarify problems as they arise (Chiu 2024), giving immediate personalised feedback (Adiguzel, Kaya, and Cansu 2023). Even something as simple as the inclusion of chatbots into educational practices is seen as offering instant benefits, including 'immediate assistance, quick access to information, enhanced learning outcomes, and improved educational experiences' (Labadze, Grigolia and Machaidze 2024: 6). Such innovations are claimed to unlock the ability to have a personal tutor instead of mass education – something that has historically been a privilege of the few (Khan 2024).

Promotion of such new educational strategies typically runs in tandem with attempts at creating a straw man out outdated practices, such as mere memorisation, and contrasting them with rich and creative AI-enabled learning that requires no (or very little) memorisation, only a capacity to find information (Cope, Kalantzis and Searsmith 2021; see, from a more critical perspective, Clark 2024). Such outdated practices are taken to be synonymous with what currently happens in the classroom, without regard to diverse teaching techniques that are already employed and without straightforward empirical evidence regarding effectiveness (which is also dependent on national cultural and educational traditions). Nevertheless, the more traditional approaches are contrasted with 'innovative and inclusive learning environments', and 'modern teaching materials' (Ferreira Costa et al. 2023: 1645) that are enabled by AI. The lack of datafication and surveillance in traditional education is lamented as 'random and irresponsible', in contrast to data-rich predictive AI-based education (Yu and Couldry 2022: 134). Here, one encounters maximisation of competitiveness in education through embracing digital technologies, thus rendering the education sector susceptible to outside influences and ideologies that are based on commercial and platform logics and not on education-native values (Clark 2024: 423–424).

Nevertheless, despite the hype, there is only limited evidence as to the actual benefits of AI in education (Nemorin et al. 2023). Despite the emphasis on learning through collaboration with AI, the latter still lacks the capacity to replicate a human touch in educational interactions, that is, interpersonal communication and collaboration elements that are necessary in a properly functioning education system (Ali et al. 2024: 7). There is still room for a human role in terms of empathy and emotional support (Labadze, Grigolia, and Machaidze 2024) – ignoring the latter might lead to 'teacher deskilling and de-humanising of the classroom' (Selwyn, Campbell and Andrejevic 2023: 176). However, the idea of AI learning partners is not unfounded either: there is evidence of students perceiving AI, including chatbots, such as ChatGPT, not only as mere response generators but as a peer/tutor, endowed with supportive agency and as possessing qualities, such as situation, contextual, and personal

awareness, higher-order thinking, etc. – effectively, understanding the AI tool as a quasi-human partner (Kim et al. 2025). The key question, however, is one of the effects of such interactions and the underlying values guiding this kind of human–AI teaming.

Notably, the functioning of AI models tends to be based on stereotypical accounts of neurotypicality and neurodivergence (Nagy 2024), thereby perpetuating – and further institutionalising – potentially discriminative and disempowering models of action and standards for evaluating performance. The preceding is further compounded by the introduction of values and priorities alien to the education domain, such as startup mentality (espoused by EdTech founders) of datafication and efficiency maximisation, adopted as the organising principle of technologised educational processes (Selwyn, Campbell, and Andrejevic 2023: 176). It must be noted that EdTech tools and the underlying technologies, such as deep learning neural networks, are developed by 'a small elite of data experts driven by technical mindsets and commercial incentives' and ultimately superimpose 'multiple layers of algorithmic complexity on stripped-down (and highly contentious) understandings of human learning' (Perrotta and Selwyn 2020: 254). Learner surveillance is crucial here – a sensing process, consisting not only of determining a learner's level of knowledge but also physiological responses to measure the quality and characteristics of their learning; this process tends to be shrouded in quasi-medical terms, such as 'diagnosis' (see, e.g., Moleaar 2022: 635), framing the predictive nature of AI's output as supposedly objective. The same issue is also noted by Porayska-Pomsta (2023: 77), stressing how 'learners are being *diagnosed* like medical patients for conceptual bugs that must be remediated or for missing competencies that must be instilled in them', treating any deviation from a normalised model (often skewed towards standard-abled neurotypical learners) as 'deficit'. The latter is further achieved through 'scientization', that is, 'claiming authority from empirical evidence and scientific theory, often by mobilizing findings and concepts selectively according to particular interests and objectives' (Williamson, Macgilchrist and Potter 2023: 3). Instead of being neutral and objective, such calculations are based on decisions and judgements: 'selecting what is to be measured, how this measurement will be represented, and how these measurements will be analyzed and otherwise used' (Gulson, Sellar and Webb 2022: 19). In a way, education becomes a recursive process: First deciding on desirable outcomes, then collecting large amounts of data, and subsequently using the insights derived from the latter to nudge individuals towards such outcomes (Decuypere and Hartong 2023). Personalisation then comes at the cost of user (i.e., learner) datafication: Constant reactive monitoring and shaping potential outcomes to within the scope of what is expected in advance: '[o]nce human interactions and activities are seen as *naturally* fused with

data-collection systems, optimising data outcomes follows naturally' (Yu and Couldry 2022: 134). Hence, again, the value judgements by a small (mostly Western) elite of technology founders shape educational processes and priorities.

Reshaping education around AI and EdTech more broadly means embracing a 'data-hungry' vision of education Clark 2024: 424), whereby students ultimately become extractable data sources (Nemorin et al. 2023), resulting in the dominance of resource appropriation, wealth concentration, and promotion of extractive ideologies (Kohnke and Foung: 2024; see also Bhatia, Arora and Gupta 2024). Underlying the preceding is the 'assumption that no space in the human body is sacred enough to be protected from the creep of AI's attention' and that 'every aspect of bare life is and should be thrown open for measurement and behavioural management via timely nudges' (Nemorin et al. 2023: 48). Such approaches aim to maximise data extraction and efficiency of learning (bringing back the questions of *whose* definition of efficiency prevails and, going back to the argument of this book, whether efficiency is truly a necessary value) before any other concerns (see, e.g., Selwyn 2024), with both learners and teachers becoming data subjects of technology operators (Berendt, Littlejohn and Blakemore 2020; Decuypere and Hartong 2023). Moreover, as digital tools are being integrated into education by default, it becomes extremely difficult to choose and refuse – generation of data through teaching and learning practices and digital tool use seemingly becomes the only option (Kohnke and Foung 2024).

Again, the preceding makes crossroads unfathomable as the seemingly straight path of extractivism, frictionlessness, efficiency maximisation, and optimal outcomes is set forward. The values espoused by primarily Western technology entrepreneurs are thus established as natural, unavoidable, and universal, with deviation being impossible. The result can only be adiaphorisation of education – a disregard for methods and outcomes that might be less efficiency-oriented but, instead, rooted in education-native values.

Education as a By-Product?

Simultaneously, one cannot avoid the question of agency, particularly as learners become 'confined to a learning world that is decided upon by an artificial intelligence through strategies of personalisation' (Nemorin et al. 2023: 49). Individuals are rendered into a multiplicity of distinct data points that both train predictive analytics systems and become the objects of prediction: In other words, 'the promise and the function of predictive analytics is to freeze students' futures through a freezing of the past that takes as its object the permanent optimisation of the present' (Smithers 2023: 110). The preceding involves large-scale processing of data and

maintenance of comprehensive data records of academic performance to be used for ranking and profiling (Livingstone et al. 2024), as well as student categorisation and prediction of outcomes (Yu and Couldry 2022: 129). However, it is by no means evident that the trade-off of reduced privacy for more efficient learning is really in the students' long-term interest (Livingstone et al. 2024).

One should note growing reliance of education on corporate tools and infrastructures due to 'the introduction of private platforms into the spaces and routines of public education' creates 'new kinds of dependencies and technological lock-ins', thereby 'enmeshing schools in technological stacks' (Williamson, Macgilchrist and Potter 2023: 3). The business here is dual: Once locked in, ensuring continuous monetisation through subscription fees while simultaneously collecting data that can be used for the development of new products or updates for additional subscriptions (Williamson 2023). In other words, schools themselves are turned into 'data platforms […] linked to vast data collection programmes' (Yu and Couldry 2022: 129–130). In this way, the education system becomes a corporate-led site of datafication instead of being an independent domain focused on its own values.

AI's challenges to agency also encompass a foreclosure of the future. The future becomes a pre-formulated hope to be achieved and, simultaneously, a threat of underachievement that has to be constantly avoided, both taking place in a permanent present. As Smithers (2023: 110) stresses, 'predictive analytics and other tools of continuous algorithmic variation function to create a permanent present, a space and time of ontological optimization fuelled by the promise (and fear) of a future that never arrives'. Similarly, for Gulson, Sellar, and Webb (2022: 11), education enters the broader pre-emptive logic of algorithmic governance, whereby 'societies are governed based on calculations of risk and predicted futures'. Similarly, to achieve pre-set goals, EdTech tends to embody and proliferate so-called 'engineering approach to learning', which squeezes learning and teaching practices into what is measurable and can be included in training datasets while subjecting human practices to the drive towards efficiency, optimisation, and constant improvement, which is the internal logic of AI (Tuomi 2024: 21). Likewise, a combination of 'machine learning approaches, such as reinforcement learning, real-time feedback, and new iterations of behaviorism' are used to nudge learners towards predefined outcomes (Gulson, Sellar and Webb 2022: 33). EdTech companies generally 'seek to make schooling as a lived practice knowable, predictable, and thereby governable' in line with 'positivist thinking and reductionist impulses' (Nemorin et al. 2023: 48). Learners themselves become machines, moving among computable (or already pre-computed) options, from one data-based nudge towards another and following a predicted path between them (Decuypere and Hartong 2023). Education becomes

programmable on the back of data analytics and unprecedented capacities to digitally structure learning and teaching.

EdTech in general seems to be pivoting towards a panoptical model of education, equating transparency (every aspect of a learner's behaviour, knowledge, and daily practices being datafied, analysable, and representative) with control and the ability to nudge students in a predefined direction (see, notably, Khan 2024; for a critical reflection on such practices, see Yu and Couldry 2022). Notably, real-time monitoring and visibility become key values in EdTech, turning education into a 'rehearsal for a fully datafied world' (Yu and Couldry 2022: 139). However, typically it is not only their own data and track record that learners are tied to but also that of their earlier peers: Previous learner data are used to tailor materials, predict grades, adjust tutoring, select assessment, or even filter potential job openings, thereby significantly impacting individual choices and life opportunities (Berendt, Littlejohn and Blakemore 2020; see also Chiu 2024) based on past performance that is not even theirs but of others deemed similar in some (not necessarily academic) way. Unsurprisingly, such attempts at measuring 'ability levels' and capacity potentials of populations generally end up being flawed, producing biased (racialised and/or gendered) results and, therefore, 'rationing educational experiences and funding based on flawed ontological assumptions' (Gulson, Sellar and Webb 2022: 19). Nevertheless, the shroud of perceived objectivity and the scientisation practices undergirding the datafication of education tends to preclude awareness of such drawbacks.

Another crucial issue is the lack of understanding by AI models themselves: They only reproduce patterns in training data, which might result in ambiguities and failure to go beyond statistically prevalent explanations, which can impede student understanding while also producing straightforward inaccuracies (Ali et al. 2024). Indeed, it must be always kept in mind that AI tools 'don't understand meaning or content, but just synthesise information through sophisticated language processing to produce passable imitations of knowledge' (Williamson, Macgilchrist and Potter 2023: 1). Moreover, due to digital divides in technology adoption and datafication, the data on which not only learning practices but also content recommendations are based tend to reflect the personalities, learning preferences, curriculum priorities, and educational practices of developed countries as well as their views on the subjects taught, including potentially controversial subjects, such as history, literature, and other parts of the cultural canon (see, generally, Gill et al. 2024). According to the proponents of AI in education, not only can such tools 'can work with students in any subject, including history, chemistry, physics, and art', but they can also do this while 'building comfort, rapport, and trust' (Khan 2024: 18–19). Such comfort and trust are also problematic: Inherently controversial issues cannot be tutored simply from a perspective that

dominates in the training data (which would likely reflect the prevalent ideologies in politically and economically dominant societies – those that have the most prolific data record) while silencing emerging (e.g., postcolonial, minority, or counter-cultural) voices. Unsurprisingly, some (see, e.g., Zembylas 2023) would go even further to assert that digital technologies only serve to further proliferate racist and colonial outlooks. Moreover, AI tends to create an illusion 'as though there is a consensus on whatever it brings precisely because it creates such an amalgam of its sources' (Gefen and Arinze 2023: 5), thereby occluding frictions and debates. Likewise, the omnivorous nature of AI training means that the quality of training data remains problematic (Nah et al. 2023). Hence, while the growing role of AI might lead to the emergence of new norms and approaches, there is no guarantee that these would not reinforce already problematic practices and knowledge (Gulson, Sellar, and Webb 2022: 139). Again, it must be remembered that research and development in the field of AI is dominated by mostly private companies in the US and China, vying for not only economic but also geopolitical dominance – potentially turning education into one more site of techno-geopolitical struggle (Nemorin et al. 2023). Consequently, there is a strong case to rethink accountability for knowledge (Mao, Chen, and Liu 2024).

Moreover, AI models remain 'incapable of dealing with resourceful and ingenious problem-solving contexts, such as critical thinking, which is a pre-requisite in the education system' (Ali et al. 2024: 9). Hence, instead of novelty and creativity, the use of AI in education is likely to only reproduce the types of standard thinking that are easy to automate, undermining the chances of learners it was meant to empower. Also, while the use of AI in education is often credited with the ability to democratise high-quality learning experiences by shifting them online and thus making them available even in remote areas (see e.g., Kooli 2023), this argument only works by omitting digital divides.

The advertised opportunities of AI in education include not only developing personalised learning but also offering 'a child in a remote village […] access to the same quality of education as a student in a bustling metropolis' (Przegalinska and Triantoro 2024: 30), thus allegedly closing the education gap within and across societies (Khan 2024). The problem to be solved is, of course, real – in remote areas, quality education is generally less accessible, which is something that can at least theoretically be ameliorated through EdTech (Ferreira Costa et al. 2023). However, the preceding ignores already-existing differences in terms of access to connectivity and devices, the likelihood of unfavourable socio-economic conditions, and cultural and societal norms, especially pertaining to gender (see, e.g., Kalpokienė 2024). Similarly, Clark (2024: 422) reminds us to consider 'the impact of broader structural inequalities, such as an individual's social, cultural, and economic situatedness, including their 'race, class,

and gender' while Bhatia, Arora, and Gupta (2024) specifically stress the 'gender cost' of platformised education whereby girls' participation in education becomes particularly precarious outside the physical classroom. Nevertheless, the role of traditions and culture is more complex – despite sometimes obstructive norms, they provide rootedness, context, and relevance to the content and experience of education. Meanwhile, today's focus on digital AI-enabled classrooms promotes 'a standardized, aspatial view of schooling where the student's lived experiences and connections to communities are viewed as a barrier' (Cohen 2024: 284). Double decontextualisation is at play: Decontextualisation of opportunities and decontextualisation of content, both likely to undermine educational benefits.

Still, accountability is not particularly common within the EdTech community. Some of the responses to potential problems include responsibility offloading, including making it the learner's task to lead AI across potential pitfalls, such as bias (Ouyang and Jiao 2021). Meanwhile, an even more notable strategy is deflection: Responding to concerns about potential generation of harmful content by pointing to other online safety concerns (particularly for children) or stressing biases produced by other tools or emerging in human interactions; the same applies to addressing concerns of data collection whereas other digital tools and services are presented as doing even worse or at least just as bad (see, notably, Khan 2024). Such a strategy not only leaves the core issues and concerns unaddressed but also might lead to a side effect of normalising the risks, biases, and other drawbacks that come with digital technologies instead of aiming to find a solution. Once again, pausing to stop and reconsider the potential challenges and omissions (including cultural and knowledge diversity) becomes framed as a luxury that cannot be afforded, instead charting a supposedly dilemma-free path of technological progress.

References

Abbas, M., Jam, F.A. and Khan, T.I. (2024). Is It Harmful or Helpful? Examining the Causes and Consequences of Generative AI Usage Among University Students. *International Journal of Education Technology in Higher Education*, 21, 1–22.

Adiguzel, T., Kaya, M.H. and Cansu, F.K. (2023). Revolutionizing Education with AI: Exploring the Transformative Potential of ChatGPT. *Contemporary Educational Technology*, 15(3), 1–13.

Ali, O. et al. (2024). The Effects of Artificial Intelligence Applications in Educational Settings: Challenges and Strategies. *Technological Forecasting & Social Change*, 199, 1–18.

Baidoo-Anu, D. et al. (2024). Exploring Student Perspectives on Generative Artificial Intelligence in Higher Education Learning. *Discover Education*, 3, 1–21.

Berendt, B., Littlejohn, A. and Blakemore, M. (2020). AI in Education: Learner Choice and Fundamental Rights. *Learning, Media and Technology*, 45(3), 312–324.

Bhatia, K.V., Arora, P. and Gupta, S. (2024). Edtech Platforms from Below: A Family Ethnography of Marginalized Communities and their Digital Learning Post-Pandemic. *Learning, Media and Technology*, doi: 10.1080/17439884.2024.2328693

Chiu, T.K.F. (2024). The Impact of Generative AI (GenAI) on Practices, Policies and Research Direction in Education: A Case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(10), 6187–6203.

Clark, D. (2024). The Construction of Legitimacy: A Critical Discourse Analysis of the Rhetoric of Educational Technology in Post-Pandemic Higher Education. *Learning, Media and Technology*, 49(3), 414–427.

Cohen, D. (2024). Any Time, Any Place, Any Way, Any Place: Markets, EdTech, and the Spaces of Schooling. *EPA: Economy and Space*, 56(1), 270–287.

Cone, L. (2023). The Platform Classroom: Troubling Student Configurations in a Danish Primary School. *Learning, Media and Technology*, 48(1), 52–64.

Cope, B., Kalantzis, M. and Searsmith, D. (2021). Artificial Intelligence for Education: Knowledge and Its Assessment in AI-Enabled Learning Ecologies. *Educational Philosophy and Theory*, 53(12), 1229–1245.

Decuypere, M. and Hartong, S. (2023). Edunudge. *Learning, Media and Technology*, 48(1), 138–152.

Duah, J.E. and McGivern, P. (2024). How Generative Artificial Intelligence Has Blurred Notions of Authorial Identity and Academic Norms in Higher Education, Necessitating Clear University Usage Policies. *The International Journal of Information and Learning Technology*, 42(2), 180–193.

Eager, B. and Brunton, R. (2023). Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice. *Journal of University Teaching & Learning Practice*, 20(5), 121.

Ferreira Costa, A.C.F. et al. (2023). Toward Quality Education: Contributions of EdTech to the Achievement of the Fourth United Nations Sustainable Development Goal. *Sustainable Development*, 32, 1634–1651.

Gefen, D. and Arinze, O. (2023). ChatGPT Usurping Academic Authority. *Journal of Information Technology Case and Application Research*, 25(1), 3–9.

Gill, S.S. et al. (2024). Transformative Effects of ChatGPT on Modern Education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, 4, 19–23.

Gulson, K.N., Sellar, S. and Webb, P.T. (2022). *Algorithms of Education: How Datafication and Artificial Intelligence Shape Policy*. Minneapolis: University of Minnesota Press.

Jochim, J. and Lenz-Kesekamp, V.K. (2025). Teaching and Testing in the Era of Text-Generative AI: Exploring the Needs of Students and Teachers. *Information and Learning Sciences*, 126(1/2), 149–169.

Kalpokas, I., Kalpokienė, J. and Šalaševičiūtė, V. (2024). Rage Against the Machine: Automated Policing of Generative AI. In T. Østbø Kuldova, H.O. Ingebrigtsen Gundhus and C.T. Wathne (eds.) *Policing and Intelligence in the Global Big Data Era, Volume II: New Global Perspectives on the Politics and Ethics of Knowledge*. London: Palgrave Macmillan, pp. 259–286.

Kalpokienė, J. (2024). Digitization in Education and the Social Divide: A Blessing or A Curse? In E. Gruodytė and A. Pasvenskienė (eds.) *Human Right to Education in the Age of Innovations and Smart Technologies*. Berlin: Peter Lang, pp. 37–54.

Käser, T. and Alexandron, G. (2024). Simulated Learners in Educational Technology: A Systematic Literature Review and a Turing-Like Test. *International Journal of Artificial Intelligence in Education*, 34, 545–585.

Khan, S. (2024). *Brave New Words: How AI Will Revolutionize Education (and Why That's a Good Thing)*. New York: Viking.

Kim, J. et al. (2025). Exploring Students' Perspectives on Generative AI-Assisted Academic Writing. *Education and Information Technologies*, 30, 1265–1300.

Kohnke, L. and Foung, D. (2024). Deconstructing the Normalisation of Data Colonialism in Educational Technology. *Education Sciences*, 14, 1–14.

Kooli, C. (2023). Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. *Sustainability*, 15(7), 1–15.

Kramm, N. and McKenna, S. (2023). AI Amplifies the Tough Question: What Is Higher Education Really For?, *Teaching in Higher Education*, 28(8), 2173–2178.

Labadze, L., Grigolia, M. and Machaidze, L. (2024). Role of AI Chatbots in Education: Systematic Literature Review. *International Journal of Educational Technology in Higher Education*, 20, 1–17.

Lee, D. et al. (2024). The Impact of Generative AI on Higher Education Learning and Teaching: A Study on Educators' Perspectives. *Computers and Education: Artificial Intelligence*, 6, 1–10.

Leelavathi, R. and Surendhranatha, R.C. (2024). ChatGPT in the Classroom: Navigating the Generative AI Wave in Management Education. *Journal of Research in Innovative Teaching & Learning*, doi: 10.1108/JRIT-01-2024-0017

Lim, W.M. et al. (2023). Generative AI and the Future of Education: Ragnarök or Reformation? A Paradoxical Perspective from Management Educators. *The International Journal of Management Education*, 21, 1–13.

Linderoth, C., Hultén, M. and Stenliden, N. (2024). Competing Visions of Artificial Intelligence in Education: A Heuristic Analysis on Sociotechnical Imaginaries and Problematizations in Policy Guidelines. *Policy Futures in Education*, 22(8), 1662–1678.

Livingstone, S. et al. (2024). The Googlization of the Classroom: Is the UK Effective in Protecting Children's Data and Rights? *Computers and Education Open*, 7, 1–8.

Mao, J., Chen, B. and Liu, J.C. (2024). Generative Artificial Intelligence in Education and Its Implications for Assessment. *TechTrends*, 68, 58–66.
Moleaar, I. (2022). Towards Hybrid Human-AI Learning Technologies. *European Journal of Education*, 57, 632–645.
Nagy, J. (2024). Autism and the Making of Emotion AI: Disability as Resource for Surveillance Capitalism. *New Media & Society*, 26(7), 3989–4007.
Nah, F.F.-H. et al. (2023). Generative AI and ChatGPT: Applications, Challenges, and Human-AI Collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277–304.
Nemorin, S. et al. (2023). AI Hyped? A Horizon Scan of Discourse on Artificial Intelligence in Education (AIED) and Development. *Learning, Media and Technology*, 48(1), 38–51.
Ouyang, F. and Jiao, P. (2021). Artificial Intelligence in Education: The Three Paradigms. *Computers and Education: Artificial Intelligence*, 2, 1–6.
Perkins, M. et al. (2024). Detection of GPT-$ Generated Text in Higher Education: Combining Academic Judgement and Software to Identify Generative AI Tool Misuse. *Journal of Academic Ethics*, 22, 89–113.
Perrotta, C. and Selwyn, N. (2020). Deep Learning Goes to School: Toward a Relational Understanding of AI in Education. *Learning, Media and Technology*, 45(3), 251–269.
Porayska-Pomsta, K. (2023). A Manifesto for a Pro-Actively Responsible AI in Education. *International Journal of Artificial Intelligence in Education*, 34, 73–83.
Przegalinska, A. and Triantoro, T. (2024). *Converging Minds: The Creative Potential of Collaborative AI*. Boca Raton: CRC Press.
Selwyn, N. (2024). Digital Degrowth: Toward Radically Sustainable Education Technology. *Learning, Media and Technology*, 49(2): 186–199.
Selwyn, N., Campbell, L. and Andrejevic, M. (2023). Autoroll: Scripting the Emergence of Classroom Facial Recognition Technology. *Learning, Media and Technology*, 48(1), 166–179.
Sidorkin, A.M. (2024). *Embracing Chatbots in Higher Education: The Use of Artificial Intelligence in Teaching, Administration, and Scholarship*. London: Routledge.
Smithers, L. (2023). Predictive Analytics and the Creation of the Permanent Present. *Learning, Media and Technology*, 48(1), 109–121.
Su, J. and Yang, W. (2023). Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. *ECNU Review of Education*, 6(3), 355–366.
Tuomi, I. (2024). Beyond Mastery: Toward a Broader Understanding of AI in Education. *International Journal of Artificial Intelligence in Education*, 34, 20–30.
Williams, R.T. (2024). The Ethical Implications of Using Generative Chatbots in Higher Education. *Frontiers in Education*, 8, 1–8.
Williamson, B. (2023). The Social Life of AI in Education. *International Journal of Artificial Intelligence in Education*, 34, 97–104.

Williamson, B. Macgilchrist, F. and Potter, J. (2023). Re-Examining AI, Automation and Datafication in Education. *Learning, Media and Technology*, 48(1), 1–5.

Wood, D. and Moss, S.H. (2024). Evaluating the Impact of Student's Generative AI Use in Educational Contexts. *Journal of Research in Innovative Teaching and Learning*, 17(2), 152–167.

Yu, J. and Couldry, N. (2022). Education as a Domain of Natural Data Extraction: Analysing Corporate Discourse about Educational Tracking. *Information, Communication & Society*, 15(1), 127–144.

Yusuf, A., Pervin, N. and Román-González, M. (2024). Generative AI and the Future of Higher Education: A Threat to Academic Integrity or Reformation? Evidence from Multicultural Perspectives. *International Journal of Educational Technology in Higher Education*, 21, 1–29.

Zembylas, M. (2023). A Decolonial Approach to AI in Higher Education Teaching and Learning: Strategies for Undoing the Ethics of Digital Neocolonialism. *Learning, Media and Technology*, 48(1), 25–37.

3 (Re)Locating the Author
AI and Creativity

AI and Creativity: An Uneasy Relationship

Until recently, the human mind alone was seen as capable of generating creative expression, while any devices and software could have only been considered as tools. This view was also supported by the seemingly unassailable assumption that culture – that is, *human-produced* culture – is central to human flourishing (Friedman 2024). However, creativity has by now become much more ambiguous (Benedikter 2021: 75; see also Carnovalini and Rodá 2020). According to Škiljić (2021: 1339), '[i]n today's rapidly evolving technological landscape, humans share the canvas with another talented artist – Artificial Intelligence'. However, one must also keep in mind that the answer to the question of whether AI is (or ever will be) truly capable of creativity first and foremost depends on the way one defines creativity in the first place, with competing definitions being plentiful (Moruzzi 2021; Henriksen, Creely, and Mehta 2022). One could even go one step further by asserting that '[a]ssigning a definition to creativity, or deeming something as creative, is often a political choice' – one that embodies the dominant constellation of tacit taken-for-granted assumptions, normalised and even enforced through education, societal discourse, commercial practices, etc. (Henriksen, Creely and Mehta 2022: 465–466; see also Stephensen 2022). Unsurprisingly, some would even refuse to become involved in what could be called a 'definition game of creativity', thereby rejecting creativity as a preexisting phenomenon in the first place (see, notably, Stephensen 2022).

Nevertheless, at least as of now, talking about completely automated AI creativity is futile since human prompting remains a necessity (Fenwick and Jurcys 2023): The idea that one could have 'a stand-alone, autonomous, solely human, natural creativity that precedes or escapes technology' is simply misleading (Stephensen 2022: 22). However, while it might be too ambitious and too early to claim that the human–machine hierarchy in the domain of creativity has been completely upended, one can, nevertheless, agree with Benedikter (2021: 76) that complexity and

DOI: 10.4324/9781003715849-3

ambiguity best define the present status quo as we transition from 'human-machine interaction' to 'human-machine convergence'. For this reason, it is important to explore both the current status quo around creativity and authorship and possible alternatives – as will be shown below, one does not need human replacement with AI for important crossroads to emerge.

Most legal systems today, arguably, have succumbed to the 'romantic' anthropocentric conception of authorship and its value (Dornis 2020: 23; see also Lim 2018: 816; Zeilinger 2021: 4; Stephensen 2022). However, the roots of such thinking go back in time – not even to Romanticism but to the Enlightenment, that is, to the separation of humans from their environment (and of the mind from the body), thereby setting the human (and, particularly, the human mind) as the measure of all things (Henriksen, Creely and Mehta 2022: 468). If the notion of authorship has become, following Arriagada (2020: 400), 'mystical', then AI certainly poses a challenge to that through the shrinking of the difference between human and machine creative capacities (Dornis 2020: 16). As Kaminski (2017: 594) observes, 'algorithmic authorship fundamentally challenges the notion of the romantic author, i.e. an individual human being who produces creative output during moments of enlightened creativity'. Any romanticising act becomes much more problematic to perform when an artificial agent becomes capable of producing the same, or very similar, output (see, e.g., Moruzzi 2021; Henriksen, Creely, and Mehta 2022; Hitsuwari et al. 2023). Hence, creativity today is increasingly pushed towards being something that takes place at the intersection between humans and technology, instead of humans unilaterally moulding seemingly subordinate matter – or pixels – as the traditional view would have it (Henriksen, Creely and Mehta 2022: 467).

A typical response to the above is a negation of AI's potential, which some (see, e.g., Kalpokiene and Kalpokas 2023; Millet et al. 2023) would see as a defensive reaction against the challenge posed to the entrenched anthropocentric belief in human exceptionality. Characteristically, for Zurth (2021: 11), AI's capacity to surprise both the user and the programmer is not in itself sufficient to justify the claim to authorship and creativity. In fact, whatever AI does, the result would then be seen as only a 'mock' impression of human capacities; such thinking, in turn, allows for a convenient dismissal of the creative potential of AI agents as lacking self-understanding and awareness: '[m]achines do not reflect the zeitgeist, do not process social and societal impressions, and do not get inspired on subconscious levels' (Zurth 2021: 11). Similarly, for Hertzmann (2020: 17), art is a fundamentally social domain – a matter of shared meaning and expression – which is something that machines are said to lack. According to Svedman (2020: 20), AI lacks the 'fundamentally human' embodiedness as it learns from preselected curated examples in a way that

is cut off from the broader (human) community; hence, the argument goes, there is neither a connection between the technology and the artistic tradition nor a relationship between the curators or coders and the end result (because the endeavours of the former are mediated by technology). Hence, AI-generated output can be seen to lack the intentionality, self-actualisation, and societal embeddedness of human art (Runco 2023). Also, to the extent that AI models *do* reflect such values as a result of machine learning, it is claimed that at least for as long as machines will be trained on datasets comprised of human art, the values reflected will not be independently derived but, instead, those of human audiences and earlier human artists (Veale and Pérez y Pérez 2020: 55).

Normally, when it comes to understanding and conceptualising creativity, one is supposed to focus on actions that 'exhibit at least a relevant purpose (in not being purely accidental), some degree of understanding (not using purely mechanical search procedures), a degree of judgement (in how to apply a rule, if a rule is involved), and an evaluative ability directed at the task at hand' – in other words, we are used to perceiving 'intention, understanding, and rationality' as the necessary underpinnings of creativity (De Cock Buning 2016: 316), with such qualities remaining largely out of bounds for generative AI models to this day. Gruner and Csikszentmihalyi (2019), meanwhile, emphasise the lack of emotions and moral reasoning as the main factor that undermines AI's creative endeavours (see also Tubadji, Huang and Weber 2021). Similarly, the lack of both emotions *and* symbolic interpretation of the world (including training data) is likewise seen as a major drawback to AI creativity (Sun, Wang, and Xiong 2024: 3134). And for Lim (2018: 842), contrary to human authors, 'algorithms do not have the quintessential lynchpin upon which to hang creativity – free will'. Indeed, whatever outcome is achieved, it can be traced to humans: Even if a specific result is not predetermined, the general direction still is (Moruzzi 2021: 15), again indicating a lack of self-determination. Lacking the abovementioned qualities, allegedly creative AI agents would be engaged in *mere generation*, that is, production of 'poem-shaped texts or cantata-shaped sounds' rather than creativity par excellence (Veale and Pérez y Pérez 2020: 556). Devoid of human cognitive capacities and general sense-making propensities, AI tools are thus seen as engaged in statistical imitation rather than originating anything (Floridi 2023). In that sense, AI-generated works would simply be parasitic in nature, feeding off human labour in terms of coding, curation, and creation of works that end up serving as a training dataset (Bogost 2019), even if the original works-as-data are restructured rather than merely reproduced. Again, the preceding traits of creativity transpire to be quintessentially rooted in the Enlightenment and, particularly, the Cartesian focus on autonomy and (self-)mastery (see, e.g., Manovich 2022; Kalpokiene and Kalpokas 2023).

Generative or not, AI is premised upon search and optimisation functions, that is, finding solutions to specified problems as efficiently as possible (Wang 2021: 142). In this way, one can construe artificial creativity as a search for ways to fulfil a pre-scripted or learned creativity function in an efficient manner. On the other hand, even simple reliance on search and optimisation techniques makes it increasingly possible to make new discoveries and produce new works that are in no meaningful way related to the ideas of the code writers – in this way, 'algorithms become able to adapt, learn, and create original, unpredictable outputs' (Köbis and Mossink 2021: 2); similarly, for Miller (2020), unpredictability is the consequence of the sheer volume of individually deterministic network nodes operating simultaneously.

Elsewhere, the focus is more on merger than substitution, whereby 'the emergence of generative AI introduces a transformative force—not as a replacement for creativity but as a catalyst for collaboration', primarily due to a proclaimed need to combine a human touch with ever increasing automation (Amankwah-Amoah et al. 2024: 4). The development of what could be called cyborg creativity (Yan 2025) or human–machine 'co-creativity' (Wingström, Hautala, and Lundman 2024) are seen as likely outcomes in the very near future, if not emerging already. The preceding would also mean that contribution (and rights) would have to be somehow apportioned (see also Yan 2025). Notably, the very idea that creativity *can* be 'distributed between human and non-human agents is sufficient to put into contestation conventional understandings of attributions of authorship, ownership, and copyright' (Celis Bueno, Chow and Popowicz 2024: 10). Hence, one can witness AI 'emulating humans and laying siege to what has been a strictly human outpost' (Gervais 2020: 2057). In this sense, despite the controversy surrounding its capacities, generative AI poses a challenge to the anthropocentric status quo.

While AI creativity seems easy to dismiss on intrinsic grounds, such a view is not without its critics either. Instead, a spin can be put on the social nature of creativity by externalising its own definition in the first place. Here, Natale and Henrickson (2024: 1910) make an important observation by criticising the tendency to discuss AI creativity 'as a quality of the internal functioning of the machine, rather than as an attribution of human users' which can, in turn, be stimulated, among other things, 'through non-technical elements such as context, cultural expectations and social dynamics'. The presence or absence of AI creativity becomes squarely the matter of dominant norms and societal conventions. Moreover, there is a further element of situatedness and interaction – not only the status of generative AI is largely dependent on the ways in which creativity is socially constructed but also AI, by generating automated reflections and extensions of human culture, has the capacity to construct defaults – recurring results that help define what a particular

artist, style, or mode of expression is all about (see, e.g., Atkinson and Barker 2023). AI, including its generative iterations, not only depends on but is also simultaneously productive of societal conventions.

The key issue is whether creativity is defined in terms that are quintessentially human (in which case, machine creativity is impossible) or as a matter of behaviour, which would allow for it to be coded (see, e.g., Cetinic and She 2022). Moreover, when it comes to the socially interactive aspect of artistic creativity, with advances in affective computing, the ability to sense and trigger emotions could be built in as well (Miller 2020; Celis Bueno, Chow, and Popowicz 2024). Simultaneously, with the expansion (in terms of both breadth and depth) of datafication processes, potentially integrating Internet of Things and other 'smart' infrastructures in addition to works of art, one could witness the emergence of a new form of 'total' art – one that is 'total' in terms of its absolute integration of the (human and machinic) everyday in the form of Big Data. As remains to be seen in the subsequent chapter, similar definitional issues also pertain to the heritage status of born-digital content.

Crucially, then, AI contradicts the ways in which agency is popularly located within anthropocentric models of creativity and authorship (Zeilinger 2021: 16–17), thereby challenging a core bastion of (at least Western) assumptions about the world and human role in it (Stephensen 2022; Millet et al. 2023). Indeed, an argument could be made that despite the commonplace derision of AI as non-creative by definition, particularly in the case of deep learning, the connection between an artwork and a human behind it might be finally cut (Köbis and Mossink 2021: 2), particularly should one choose to focus, with Millet et al. (2023: 1), on the ways in which generative AI's output increasingly 'extends beyond emulating already existing artistic styles'. Moreover, an argument can be made that the more sceptical views of AI creativity simply raise the bar too high: While they, implicitly or explicitly, focus on the rare standout works of human creativity, most of what we consider to be human creative output is lower scale (see, generally, Lee 2022).

As can be seen in this exposition, crossroads in the domain of creativity are clearer than in most other domains, likely due to the legacy of seeing creativity as a quintessentially (even definitionally) human quality. Hence, the need to make a choice between human focus and technological optimisation of output generation remains a pressing one, precluding the moral blindness of ignored alternatives.

What to Protect? Dealing With Anthropocentric Legacy

The preceding also points towards a split whereby 'human-made works are protected under an increasing system of copyrights and neighboring rights' but, simultaneously, 'the domain of AI creativity – although ever

more elaborate – has remained unprotected' (Dornis 2020: 11). Of course, it is crucial to keep in mind that as the cultural landscape evolves, so do the prevalent ideas and, no less importantly, available forms of their expression (Svedman 2020: 16), which in this case may potentially be moving in a way that challenges deeply engrained ideas of human self-sufficiency in creativity, threatening to undermine the 'ontological security' that humans may still have with regards to their alleged exceptionality among all entities in this world (Millet et al. 2023: 1–2; see also Celis Bueno, Chow, and Popowicz 2024). Meanwhile, if one assumes definitions of creativity to be mere inventions (as suggested by some of the accounts above), each serving a particular historical constellation of cultural assumptions and economic interests (see, e.g., Stephensen 2022), a reckoning with new modes of creativity might well be in order. In this sense, the emergence of more pronouncedly machine-focused creativity could simply be a new addition to the spirit of the times while simultaneously putting societies at a crossroads with regard to the role and position of human artists.

Perhaps nowhere is the anthropocentric nature of authorship and creativity more clearly pronounced than in the domain of Intellectual Property rights, with the scope of this section being more narrowly on copyright. In this respect, it might be rightfully observed, as Yanisky-Ravid and Velez-Hernandez (2018: 14) do, that 'the current (traditional) legal regime focuses only on what was relevant in the past, namely the human authors behind the creative process'. This anthropocentric view also runs across the fundamental existing international documents, from the Berne Convention to the TRIPS Agreement, albeit often implicitly (Miernicki and Ng 2021; Salami 2021: 126–127). Hence, despite some authors proclaiming that 'denying protection is inconsistent with the historically flexible interpretation and application of copyright laws as technology has developed' (Franceschelli and Musolesi 2022: 10), existing attempts at arguing *for* the protectability of AI output start on the back foot.

It must be stressed that current copyright law is premised not on the *quality* of the works in question but on the *type of agent* (human vs. non-human) that has produced them. In their current form, copyright laws across different jurisdictions would no doubt recognise much of the output that creative AI is already capable of *if only* it were produced by humans (Bonadio and McDonagh 2020; Gaffar and Albarashdi 2025). However, the *machinic* origin of such works continues to preclude their acceptance and protection. Consequently, it might seem paradoxical that while with the help of machine learning, AI can be seen as 'increasingly emancipating itself from human reign' (Dornis 2020: 14–15), evolving from deterministic generation of algorithmic art that once necessitated explicit coding to creative AI systems based on machine learning (see, e.g., Abbott and Rothman 2023; Haase and Hanel 2023), the relative legal standing of human and AI output has not changed.

Indeed, as Zurth (2021: 3) observes, '[d]espite all the variations in the world's multitude of copyright regimes, a common thread emerges: copyright is based on an anthropocentric perspective'. Similarly, for Zeilinger (2021: 3), it is a telling fact that the definitions of creativity underlying the current regulation, including benchmarks that would apply to AI's creative efforts, are strongly premised on anthropocentric conceptions of creativity. Unsurprisingly, then, there seems to be a broad agreement that the very purpose of copyright law is to incentivise *human* creativity (Kalpokiene 2024). As a result, it must be stressed that 'AI-generated artworks cannot be subsumed under the existing system [of copyright law] to satisfy all of its parameters and requirements' (Škiljić 2021: 1362; see also Gaffar and Albarashdi 2025). This is despite the possibility that, as Brown (2018: 20) asserts, '[o]ffering copyright protection to computer-generated works would directly advance copyright's purpose of encouraging the production of original literary, artistic, and musical expression for the good of the public' and overall societal progress. Moreover, it is not unfeasible that the future will be characterised by the dominance of cultural production generated by AI on the shoulders of AI (Dornis 2020: 42), making AI creativity simply the new normal, just as human creativity is today, thereby making it more difficult to justify the absence of protectability. However, an important point is raised by Mammen and Richey (2020: 284), for whom 'if AI-generated works are protected, this could result in a virtually unlimited supply of copyrightable work'; such oversupply would then drive the prices of copyrighted works down, harming all creators alike (see, e.g., Kalpokiene and Kalpokas 2023).

For Svedman (2020: 13), for example, there can be no question of copyrightability due to AI-generated output being too far removed from both training data *and* the coders' and trainers' intentions and, therefore, devoid of any personality, hence disqualifying it from being art *in principle* (see also Matulionyte and Lee 2022). Nevertheless, the fact that AI-generated art necessitates extensive training cannot be seen as a disqualifying factor in itself. After all, although '[an AI model's] teacher may be its programmer, another algorithm, a sample of artworks, and so forth', at the same time, it seems uncontroversial that 'human apprentices do not have to grant authorship of their work to their teachers' (Arriagada 2020: 404). Even more broadly, 'artists do not create in a bubble, and […] to a certain degree, all creativity requires influence' (Brown 2018: 26). In this sense, any art student (but also an established artist) creating a work of art with the tradition that came before them in mind would likely be disqualified, just as AI is (Brown 2018: 25; see also Kalpokiene and Kalpokas 2023). Hence, perhaps, *both* human and AI creative endeavours should be seen as mimicry based on experience and learning (Lim 2018: 874). It is thus evident that the argument still revolves around the presence or absence of some *special value* that is (or is not) exclusively inherent in *human* creativity.

While wholesale generation of new artistic content, potentially without a human in the loop, may play a value-increasing role in certain parts of society, others will face risks – not least, human creativity might be jeopardised (Gervais 2020: 2060). Human authors would simply be outperformed by AI. Similarly, for Dornis (2020: 34), despite the claims that AI creativity will have a positive effect on the public, this will harm human authors who will find themselves unable to compete, resulting in the dwindling of human-to-human creativity. Thus, according to Gervais (2020: 2106), since 'copyright is meant to promote human creativity', it is only natural that 'incentives to have more productions in the literary and artistic field made by machines could in fact pose threat to (human) progress'. From the latter point of view, despite acknowledging that machines can make autonomous and creative decisions, promotion of *human* progress should, nevertheless, be the aim, meaning that the output of machine creativity should be disqualified from copyright protection despite meeting the formal criteria – except for being human, that is (see Gervais 2020). But if that is the case, the market could be flooded with very cheap (or entirely free to use, if such content falls into the public domain) machine output, leaving humans unable to earn a living (for a similar argument, see, e.g., Franceschelli and Musolesi 2022). That is, the result would be (almost) identical to that of copyright protection *being* awarded.

Nevertheless, leaving AI creations unprotected might be considered a suboptimal outcome as well because it would not only leave a substantial legal lacuna but also hinder investment and innovation (Lim 2018: 841; Bisoyi 2022). After all, the ability to monetise generative AI could be seen, in some business models at least, as premised on the ability to exclude others (see, e.g., Brown 2018; Dornis 2020; Gervais 2020). A key task at hand would, therefore, be 'defining clear parameters to whom, among the humans in the AI art-generation process, should authorship be attributed' (Škiljić 2021: 1363). One way would be for the programmer to become a meta-author, still operating within the anthropocentric framework, just in another way. This might reposition and reassign incentives so that further digital technological advances are still encouraged, ostensibly allowing for societal progress (Bisoyi 2022; Matulionyte and Lee 2022: 30). Paradoxically, though, following this logic, AI-generated art, requiring more upfront investment and unlikely to arise from spontaneous crowd-based innovation, necessitates stronger copyright protection than human-created art, which can (and often does) arise spontaneously from the grassroots and can be motivated by factors other than monetary ones. Indeed, as Mammen and Richey (2020: 284) point out, 'history has shown [that] artists are going to create art, regardless of incentives'. However, it would hardly be right to use this difference as an excuse for creating less favourable conditions for the human authors. Hence, while Matulionyte and Lee (2022: 30) claim that awarding copyright protection to the programmer/owner of the AI model can achieve monetisation

with less transaction costs than any alternative approaches, this could also end up with large corporate actors hoarding IP and crowding everyone out.

Finally, it is not difficult to foresee a future in which even if human creativity continues to exist, it will be dependent on collaboration with AI agents licensed from technology companies, including such agents only being interacted with through interfaces that reveal nothing of the source code in the backend (Hoel 2021). Hence, it might be appropriate to claim, for the sake of precision, that corporations, and not AI, are automating art (Hoel 2021). Of course, one might still ask why artists would end up heavily using licensed platform tools. The first part of the answer is platform effects (see, notably, Srnicek 2017: 45), whereby the more individuals who use the platform, the more difficult it becomes for others *not to use* that same platform, further increasing pressure on those still outside. Hence, opting out of the platform ecosystem might simply no longer be a viable option. Likewise, the choice to license tools from technology companies could be a way to reduce data asymmetries, whereby platforms are always able to tailor their offerings better than anyone else (Srnicek 2017: 95). Moreover, as the widespread adoption of AI's generative capacities would flood the market with cheap and quickly produced works, the only realistic way not to lose out would be to embrace, even to merge, though one's creative practices, with AI to approximate machinic speed and cost as closely as possible. Without the financial resources necessary to invest in developing such automation capacities, tethering oneself to a corporate actor might be necessary. Moreover, being part of the platform ecosystem would likely be a (near-) must due to the ever-growing role of AI in decision-making and gatekeeping, that is, determining the discoverability, prominence, value, and possibly, even overall 'art-ness' of creative outputs (Baradaran 2024). Hence, the fate of human artistic autonomy is also part of this chapter's crossroads. As the pressures on human creativity mount, in terms of increasing drives to protect or otherwise monetise AI-generated outputs or to blur the lines between human creativity and AI generation, and with likely further strides towards efficiency and process optimisation that AI would offer, hybridity or human replacement in creativity would become a likely outcome. A choice between maintaining the importance of human creativity and more efficient AI generation would have to be made. However, given the general trajectory observed across the different domains explored in this book, a question can be raised as to whether a choice – crossroads – *will* be seen as possible or a seemingly frictionless path towards progress will emerge.

Towards a Kitsch Machine?

At first sight, the emergence of GenAI may lend itself to Benjaminian critique of mechanical reproduction. Famously, as observed by Benjamin (1964: 218), while 'a work of art has always been reproducible', a key shift

was observable already following the adoption of the camera: From human imitations of human-created works to mechanical reproduction of both the works and the world itself. However, an even more fundamental change, for Benjamin (1964: 220), was a lack in even the most perfect reproductions: That of the original work's 'presence in time and space, its unique existence at the place where it happens to be'. Benjamin called the uniqueness and historical placedness of the work 'aura', which 'withers in the age of mechanical reproduction' due to the impossibility of mechanical outputs to be part of the fabric of tradition (Benjamin 1964: 221). AI-generated output, of course, occupies a strange place vis-à-vis tradition: Simultaneously being based on it (as training data) and representing it as a decontextualised amalgamation. Moreover, the reproducibility in question here becomes not that of existing works as such (although there are significant concerns to that effect) and/or verbatim copies of the world but of patterns in already available data.

For Rancière (2004: 12), in the meantime, the placedness and rootedness that define Benjaminian aura are merely part and parcel of the 'distribution of the sensible', namely, 'the system of self-evident facts of sense perception that simultaneously discloses the existence of something in common and the delimitations that define the respective parts and positions within it'. In other words, that is a manifestation and a source of (political) power, determining who has a standing in a particular community and who does not, who has a share in the common objects of concern and appreciation and who does not (Rancière 2004: 12). The preceding has clear echoes with regard to AI: The heavily datafied representations of the world, as well as the dominant cultures and their representations acquiring normative power (e.g., being – or not – part of 'the canon' becomes a matter of data). Once again, the question of what and who gets included in the training data and what and who is left behind (and, likewise, what resurfaces in GenAI outputs and what gets drowned in the global mix) becomes a question of power.

In Benjamin's (1964: 224–225) exegesis, the work of art has its origin in religion, which remains 'recognizable as secularized ritual even in the most profane forms of the cult of beauty', thereby opening another avenue for critique: Of mechanical reproducibility shifting emphasis from 'cult value' to 'exhibition value', thereby leaving the artistic function merely accidental. As mechanical reproductions abound, the argument goes, the appreciation of a work of art changes as well, prioritising easy and uncomplicated enjoyment over artistic value and innovation as '[t]he conventional is uncritically enjoyed, and the truly new is criticized with aversion' (Benjamin 1964: 234). Integral to that is what we would now recognise as the attention economy: A move from a few creators enjoying undivided audience attention to everyone producing their own content and vying to be seen (Benjamin 1964: 231–232). Again, Benjamin here can be contrasted with

Rancière (2004: 12), for whom '[p]olitics revolves around what is seen and what can be said about it, around who has the ability to see and the talent to speak, around the properties of spaces and the possibilities of time'. Under this view, 'the transgressive appearance of unauthorized speakers on the public stage' is a sign of democratic potential, with the mechanical arts having the potential to 'confer visibility on the masses, or rather on anonymous individuals', at least if the former are recognised *as arts* and not merely as 'techniques of reproduction or transmission' (Rancière 2004: 12, 32). In this sense, GenAI has a democratising function – at least to some extent. It opens room for more individuals seeking avenues for (self-)expression by reducing the skill burden; simultaneously, though, this enabling hardly takes place in the unauthorised and transgressive manner appreciated by Rancière: Instead, it happens through redistribution and recomposition of what already is. GenAI then becomes the perfect kitsch machine: Democratising and conservative at the same time.

If we assume that mechanical art 'confers visibility on absolutely anyone' because 'anonymous became the subject matter of art' – a culmination of a long process of bringing mundanity and everyday life into the domain of the arts (Rancière 2004: 12, 32) – then it must also be stressed that GenAI has notably changed the perspective: From the standpoint of creators, it has shifted emphasis from a democratised mass production of outputs to mass interaction with corporate-produced and maintained generators the impact of which stretches way beyond, for example, the technological affordances of a camera; in terms of the subject matter, meanwhile, what is set forth in case of GenAI is not the anonymous – it is a corporate data fusion, with (non)inclusion into training datasets becoming a highly political issue. In addition, for Benjamin (1964: 224), provenance had become highly problematic, with mechanical reproduction having rendered authenticity and the possibility of 'the original' nonsensical because any number of fully identical photo prints or copies of film or audio recordings could now be produced. Likewise, with the diminution of the role of human effort, capacity, and skill, '[t]the equipment-free aspect of reality […] has become the height of artifice' (Benjamin 1964: 233), further strengthening the critique of a lack of authenticity. Again, GenAI stands as something in-between, simultaneously opening up and serialising: Anonymisation that is only achieved through amalgamation, thereby neutralising any critical potential through endless recombinations, all equally reworking existing data: Once again, a kitsch machine.

References

Abbott, R. and Rothman, E. (2023). Disrupting Creativity: Copyright Law in the Age of Generative Artificial Intelligence. *Florida Law Review*, 75(6), 1141–1202.

Amankwah-Amoah, J. et al. (2024). The Impending Disruption of Creative Industries by Generative AI: Opportunities, Challenges, and Research Agenda. *International Journal of Information Management*, doi: 10.1016/j.ijinfomgt.2024.102759

Arriagada, L. (2020). CG-Art: Demystifying the Anthropocentric Bias of Artistic Creativity. *Connection Science*, 32(4), 398–405.

Atkinson, P. and Barker, R. (2023). AI and the Social Construction of Creativity. *Convergence: The International Journal of Research into New Media Technologies*, 29(4), 1054–1069.

Baradaran, A. (2024). Towards a Decolonial I in AI: Mapping the Pervasive Effects of Artificial Intelligence on the Art Ecosystem. *AI & Society*, 39, 7–19.

Benedikter, R. (2021). Can Machines Create Art? A "Hot" Topic for the Future of Commodified Art Markets. *Challenge*, 64(1), 75–86.

Benjamin, W. (1964). The Work of Art in the Age of Its Mechanical Reproducibility. In Hannah Arendt (ed.) *Illuminations*. New York: Schocken Books, pp. 217–521.

Bisoyi, A. (2022). Ownership, Liability, Patentability, and Creativity Issues in Artificial Intelligence. *Information Security Journal: A Global Perspective*, 31(4), 377–386.

Bogost, I. (2019, March 6). The AI-Art Gold Rush is Here. *The Atlantic*, https://www.theatlantic.com/technology/archive/2019/03/ai-created-art-invades-chelsea-gallery-scene/584134/

Bonadio, E. and McDonagh, L. (2020). Artificial Intelligence and Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity. *Intellectual Property Quarterly*, 2, 112–137.

Brown, N.I. (2018). Artificial Authors: A Case for Copyright in Computer-Generated Works. *Columbia Science and Technology Law Review*, 20, 1–41.

Carnovalini, F. and Rodá, A. (2020). Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. *Frontiers in Artificial Intelligence*, doi: 10.3389/frai.2020.00014

Celis Bueno, C., Chow, P.-S. and Popowicz, A. (2024). Not 'What', but 'Where Is Creativity?': Towards a Relational-Materialist Approach to Generative AI. *AI & Society*, doi: 10.1007/s00146-024-01921-3

Cetinic, E. and She, J. (2022). Understanding and Creating Art with AI: Review and Outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(2), 1–22.

De Cock Buning, M. (2016). Autonomous Intelligent Systems as Creative Agents Under the EU Framework for Intellectual Property. *European Journal of Risk Regulation*, 7(2), 310–322.

Dornis, T.W. (2020). Artificial Creativity: Emerging Works and the Void in Current Copyright Doctrine. *Yale Journal of Law and Technology*, 22, 1–60.

Fenwick, M. and Jurcys, P. (2023). Originality and the Future of Copyright in an Age of Generative AI. *Computer Law & Security Review: The International Journal of Technology Law & Practice*, 51, 1–12.

Floridi, L. (2023). AI as *Agency Without Intelligence*: On ChatGPT, Large Language Models, and Other Generative Models. *Philosophy & Technology*, 36, 1–7.
Franceschelli, G. and Musolesi, M. (2022). Copyright in Generative Deep Learning. *Data & Policy*, 4, 1–18.
Friedman, D. (2024). Copyright as Affirmative Action for Human Authors Until the Singularity. *GRUR International*, 73(1), 1–2.
Gaffar, H. and Albarashdi, S. (2025). Copyright Protection for AI-Generated Works: Exploring Originality and Ownership in a Digital Landscape. *Asian Journal of International Law*, 15(1), 23–46.
Gervais, D.J. (2020). The Machine as Author. *Iowa Law Review*, 105, 2053–2106.
Gruner, D.T. and Csikszentmihalyi, M. (2019). Engineering Creativity in an Age of Artificial Intelligence. In I. Lebuda and V.P. Glaveanu (eds.) *The Palgrave Handbook of Social Creativity Research*. London: Palgrave Macmillan, pp. 447–462.
Haase, J. and Hanel, P.H.P. (2023). Artificial Muses: Generative Artificial Intelligence Chatbots Have Risen to Human-Level Creativity. *Journal of Creativity*, 33(3), 1–7.
Henriksen, D., Creely, E and Mehta, R. (2022). Rethinking the Politics of Creativity: Posthumanism, Indigeneity, and Creativity Beyond the Western Anthropocene. *Qualitative Inquiry*, 28(5), 465–475.
Hertzmann, A. (2020). Computers Do Not Make Art, People Do. *Communications of the ACM*, 63(5), 45–48.
Hitsuwari, J. et al. (2023). Does Human-AI Collaboration Lead to More Creative Art? Aesthetic Evaluation of Human-Made and AI-Generated Haiku Poetry. *Computers in Human Behavior*, 139, 1–10.
Hoel, E. (2021, September 8). Big Tech Is Replacing Human Artists with AI. *The Intrinsic Perspective*, https://erikhoel.substack.com/p/big-tech-is-replacing-human-artists
Kalpokiene, J. (2024). *Law, Human Creativity and Generative Artificial Intelligence: Regulatory Options*. London and New York: Routledge.
Kalpokiene, J. and Kalpokas, I. (2023). Creative Encounters of a Posthuman Kind: Anthropocentric Law, Artificial Intelligence, and Art. *Technology in Society*, 72, 1–8.
Kaminski, M.E. (2017). Authorship, Disrupted: AI Authors in Copyright and First Amendment Law. *University of California, Davis Law Review*, 51, 589–616.
Köbis, N. and Mossink, L.D. (2021). Artificial Intelligence versus Maya Angelou: Experimental Evidence that People Cannot Differentiate AI-Generated from Human-Written Poetry. *Computers in Human Behavior*, 114, 1–13.
Lee, H.-K. (2022). Rethinking Creativity: Creative Industries, AI and Everyday Creativity. *Media, Culture & Society*, 44(3), 601–612.
Lim, D. (2018). AI & IP: Innovation Creativity in an Age of Accelerated Change. *Akron Law Review*, 52(3), 813–876.
Mammen, C.E. and Richey, C. (2020). AI and IP: Are Creativity and Inventorship Inherently Human Activities? *Florida International University Law Review*, 14(2), 275–292.

Manovich, L. (2022). AI & Myths of Creativity. *Architectural Design*, 92(3), 60–65.
Matulionyte, R. and Lee, J. (2022). Copyright in AI-Generated Works: Lessons from Recent Developments in Patent Law. *SCRIPTed: Journal of Law, Technology and Society*, 19(1), 5–35.
Miernicki, M. and Ng, I. (2021). Artificial Intelligence and Moral Rights. *AI & Society*, 36, 319–329.
Miller, A.I. (2020, February 1). Machines Have Learned How to Be Creative. What Does that Mean for Art? *Salon*, https://www.salon.com/2020/02/01/machines-have-learned-how-to-be-creative-what-does-that-mean-for-art/
Millet, K. et al. (2023). Defending Humankind: Anthropocentric Bias in the Appreciation of AI Art. *Computers in Human Behavior*, 143, doi: 10.1016/j.chb.2023.107707
Moruzzi, C. (2021). Measuring Creativity: An Account of Natural and Artificial Creativity. *European Journal for Philosophy of Science*, 11(1), 1–20.
Natale, S. and Henrickson, L. (2024). The Lovelace Effect: Perceptions of Creativity in Machines. *New Media & Society*, 26(4), 1909–1926.
Rancière, J. (2004). *The Politics of Aesthetics: The Distribution of the Sensible*. London: Continuum.
Runco, M.A. (2023). AI Can Only Produce Artificial Creativity. *Journal of Creativity*, 33(3), 1–7.
Salami, E. (2021). AI-Generated Works and Copyright Law: Towards a Union of Strange Bedfellows. *Journal of Intellectual Property Law and Practice*, 16(2), 124–135.
Škiljić, A. (2021). When Art Meets Technology or Vice Versa: Key Challenges at the Crossroads of AI-Generated Artworks and Copyright Works. *IIC - International Review of Intellectual Property and Competition Law*, 52, 1338–1369.
Srnicek, N. (2017). *Platform Capitalism*. Cambridge and Malden: Polity Press.
Stephensen, J.L. (2022). Artificial Creativity: Beyond the Human, or Beyond Definition? *Transformations*, 36, 19–37.
Sun, D., Wang, H. and Xiong, J. (2024). Would You Like to Listen to My Music, My Friend? An Experiment on AI Musicians. *International Journal of Human – Computer Interaction*, 40(12), 3133–3143.
Svedman, M. (2020). Artificial Creativity: A Case Against Copyright for AI-Created Visual Artwork. *IP Theory*, 9(1), 1–22.
Tubadji, A., Huang, H. and Weber, D.J. (2021). Cultural Proximity Bias in AI Acceptability: The Importance of Being Human. *Technological Forecasting & Social Change*, 173, 1–14.
Veale, T. and Pérez y Pérez, R. (2020). Leaps and Bounds: An Introduction to the Field of Computational Creativity. *New Generation Computing*, 38, 551–563.
Wang, X. (2021). AI Output: A Human Condition that Should *Not* Be Protected Now, or Maybe Ever. *Chicago-Kent Journal of Intellectual Property*, 20(1), 136–168.

Wingström, R., Hautala, J. and Lundman, R. (2024) Redefining Creativity in the Era of AI? Perspectives of Computer Scientists and New Media Artists. *Creativity Research Journal*, 36(2), 177–193.

Yan, D. (2025). Posthuman Creativity: Unveiling Cyborg Subjectivity Through ChatGPT. *Qualitative Inquiry*, 31(2), 253–264.

Yanisky-Ravid, S. and Velez-Hernandez, L.A. (2018). Copyrightability of Artworks Produced by Creative Robots and Originality: The Formality-Objective Model. *Minnesota Journal of Law, Science & Technology*, 19(1), 1–53.

Zeilinger, M. (2021). Generative Adversarial Copy Machines. *Culture Machine*, 20, 1–23.

Zurth, P. (2021). Artificial Creativity? A Case Against Copyright Protection for AI-Generated Works. *UCLA Journal of Law and Technology*, 25(2), 1–18.

4 Many Problems, Few Solutions
AI and Digital Heritage

The Large and Expanding Scope of Digital Heritage

Article 1 of the UNESCO *Charter on the Preservation of the Digital Heritage* (hereafter, 'the Charter') defines digital heritage as 'unique resources of human knowledge and expression' existing in a digital form and encompassing 'cultural, educational, scientific and administrative resources as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources' with formats covered encompassing 'texts, databases, still and moving images, audio, graphics, software and web pages' etc., the list being deliberately non-exhaustive. Indeed, the scope of the Charter is simultaneously *unavoidably* broad to encompass any artefacts within a permanently developing domain and *overly* broad, hampering efforts to develop a clear field of concern (Haux, Dominicé, and Raspotnig 2021: 773). Paradoxically, it is also *insufficiently* broad: Having been drafted in a Web 1.0 context, it focused on static, formally and institutionally created content, with user-generated content not yet making an appearance in the text. Given that the latter type of content has been among the main drivers behind the exponential quantitative growth of the pool of (potential) digital heritage, this omission, although understandable, raises significant challenges in dealing with the *scope* of digital heritage and opens, as shown below, notable crossroads.

The scope of potential heritage poses a challenge due to the effort needed for the protection and maintenance of digital content. The Charter itself stresses the ephemerality of digital resources and the need for deliberate effort in preservation (Articles 1 and 2). Indeed, the driving force behind the adoption of the Charter has been the ever-increasing concern that digital material (particularly digital-born material) would disappear and become inaccessible, thereby leaving no trace for future generations to rely on (de Lusenet 2007: 164–165). The latter's ephemerality, such as lack of physical expression to be stored and reliance on digital storage, dependence on continued maintenance of (usually proprietary) software and file

DOI: 10.4324/9781003715849-4

formats for access, etc., has become a crucial issue. It was clear early on that the preservation of digital heritage necessitates approaches and facilities different from traditional archives: Instead of simply keeping artefacts statically preserved, for digital content, it is more 'like preserving the flame of a fire' that needs to be constantly maintained (Abid 2007: 10). Hence, the threat of loss is presented in Article 3 of the Charter as imminent, thereby providing urgency to preservation efforts.

As digital artefacts cannot be passively stored but, instead, must be constantly maintained in data centres and similar storage facilities, a clear implication is that preservation of digital heritage is resource-intensive. The infrastructural nature of digital technologies, including storage, means that their operation is insufficiently taken into account: We simply expect infrastructures to function without paying much attention to them. With regard to resource use and the effects of provision and maintenance of data storage, the environmental footprint is particularly concerning and includes not only the raw materials necessary for hardware manufacturing but also the high levels of energy consumption for computation and storage that need to be kept stable and running at all times (Brevini 2020). Indeed, the environmental impact generally tends to be underrepresented, to the extent that even research on the sustainability of data curation practices tends to focus on aspects such as ethics and fidelity but not the environment (see, e.g., Bareikytė et al. 2024). Hence, an overly pronounced zealousness in protecting digital heritage without appropriate selection would only further exacerbate the current climate emergency. Consequently, not only the amount of preserved content could spell its own end due to the impossibility of making sense of it all, but also the environmental impact necessitates some form of selection.

Meanwhile, traditional notions of heritage protection have been premised on scarcity: Collection and preservation of the already scarce material so that current and future generations have something to rely on. The drive is to preserve everything that is possible, effectively outsourcing the decision of what to keep to future generations (de Lusenet 2007). For some, with digital content, everything could potentially be cultural heritage as well, without the need to meet some specific qualitative criteria (Haux, Dominicé, and Raspotnig 2021: 772). Nevertheless, with regard to digital content, contemporary societies have moved beyond scarcity – in fact, *selecting from* the available content has become a challenge. As Dahlgren (2018: 26) has aptly put it, the contemporary information and knowledge environment is 'characterised by high velocity and dizzying excess'. By extension, then, digital heritage must be seen as 'an immense and expanding' domain (Alker and Donaldson 2018: 220). Arguably, though, such overabundance has made the omnivorous approach to heritage protection more difficult to sustain. One would be hard pressed not to remember Borges' fable *On Exactitude in Science*, featuring the map of an

empire that ultimately became as large as the empire itself. In the fable, cartography had reached such a point of perfection that it ultimately became useless and was abandoned, while the map came to ruin. Likewise, digital heritage protection might also meet a similar fate should the representation of our digital past and present become so detailed and vast that it is rendered useless by its own perfection. Unsurprisingly, then, others would argue that protection and selection of content can only be seen as two sides of the same coin (von Schorlemer 2020: 44). Nevertheless, with any decisions over what is and is not sufficiently valuable to be preserved as heritage, cultural, political, and economic considerations would kick in – whose interpretation of value, whose criteria for identifying culture, how much is to be preserved, etc., likely tilting towards the dominant nations, regions, and cultures (Lor and Britz 2012). Just like in the previous chapters, resorting to the defaults might end up strengthening implicit patterns of domination, resulting in blindness to *what* is left aside and *why*.

Crucially, the *rate* at while new content is generated grows year-on-year as well, with the online/offline distinction ultimately ceasing to be meaningful (Lambach 2020: 490; see also Chan 2022: 129–130), both coalescing into an 'intricate mix' (Dewandre 2020: 3). The implication is that not only the amount of born-digital content would continue to grow but so would its impact on everyday life and culture, rendering it of heritage value. Moreover, with the growth of Extended Reality experiences, including the prospect of fully virtual worlds becoming normalised (see, generally, Kalpokas and Kalpokiene 2023), in some contexts at least, *digital* culture, artefacts, and, therefore, heritage might well be *all there is*. While content produced through digital interactions might appear trivial and mundane, in practice, the opposite is the case: Digital life is a site of 'creativity and improvisation' in which not only social ties are established but also Big Data – which, in turn, acts as the building material for the architecture of digital sociality – is produced (Pink et al. 2017: 1). Clearly, 'people do not merely engage with the internet as a means of communication, but as a studio and an interactive gallery', which, in turn, serves as 'a space for the co-creation, circulation, and active reception of artistic practices' (Jiang 2022: 1). The digital has also become a crucial domain for identity formation – construction and curation of the self – vis-à-vis broader social structures (Ibrahim 2018: 42), thereby also weaving the fabric of online groups and other collective entities. Consequently, the stakes in digital content are very high, complicating any selection effort – everything and anything might potentially be important and worth preserving.

As information has now become overabundant to the extent that dealing with it has exceeded human capabilities (Vaidhyanathan 2018: 80), if the purpose of heritage protection is the preservation of an accurate account of societal life, the key takeaway to be preserved for future

generations is *precisely* the proliferation and growth of content whereby 'words and things take on new lives as digital surrogates, copies, and remixes' (Hennessy 2012: 346). Moreover, as today's born-digital heritage often arises from interactions among users, disentangling preservation from the need to protect privacy seems to be less straightforward than the drafters of the Charter seem to have presumed (such issues were already raised a while ago – see, e.g., Hennessy 2012: 347). The quantity aspect, nevertheless, can easily become incapacitating: After all, typical memory institutions, such as museums and libraries, are overwhelmed with the amount of digital content, leading to distributed storage having been suggested as a solution, offering continued access to everything digital for future generations (de Lusenet 2007: 168). The expectation is often that the entirety of the digital space would be preserved, including social spaces and interactions, gaming environments, etc., in a clear continuation of the omnivorous fashion of heritage protection, effectively outsourcing selection to present and future publics (de Lusenet 2007: 168). Here, private archiving initiatives, such as the Wayback Machine, also come into play while activists and volunteers play a key role in, for example, preserving digital records of atrocities and war crimes (see, e.g., Bareikytė et al. 2024).

Generating and sharing digital content, from simple posts to memes to videos, has become not only an everyday practice but also a key means through which culture is replicated and how the digital domain is experienced (see Shifman 2014: 19). Indeed, by now it seems to be a mere truism that '[t]he proliferation of social networking sites, ubiquity of mobile devices, and ease of connectivity has resulted in computer-mediated communication becoming the norm' (Krishnan and Hunt 2021: 85). Clearly, then, today's digital culture arises within 'the creative interplay of human capabilities and the capacities of more or less smart machines' (Pentzold and Bischof 2019: 3) as well as various forms of digital media and the infrastructures that undergird them (Hepp 2020: 2), introducing human–machine interactivity and hybridity at the heart of heritage-to-be. Consequently, when conceptualising what potentially can and cannot count as digital heritage, there is a need to pay more attention to 'hybrid agentivities' with regard to born-digital content and its status as heritage, with such agentivities pointing towards 'social and technical mediations in a multi-layered system prompting us to explore black boxes, particularly those of Web archiving, and its modes of governance' (Musiani and Schafer 2017: 5). The challenges do not end here; however, the social and interactive nature of the Web, as well as the propensity towards data extraction and analytics (and, thus, provision and implementation of solutions based on the latter), means that mere storage of any snapshots of digital life is insufficient. Instead, digital content and heritage-to-be are lively, meaning that 'digital cultural memory not only consists of data in

archives but also of everyday practices', thereby adding an extra degree of volatility, not least because '[t]he digital world is per se short-lived' as '[t]rends, images, and opinions change as quickly as ever, data are erased and the overflow of information may create ambiguity', in contrast to the inherited assumptions of a relatively stable cultural memory (Haux, Dominicé and Raspotnig 2021: 778). For this reason, the amount to be preserved is constantly increasing, and exponentially so.

When it comes to selection, further value problems abound as a significant portion of the cultural creativity that characterises today's societies takes place in an inauthentic fashion (e.g., by fake profiles and fake identities), with so-called bots imitating human behaviours and co-weaving the social fabric, including by co-shaping beliefs and promoting certain content (Delwiche 2020: 107). By inflating the perceived importance of particular norms or points of view and affecting the cognitive processes of individuals, such inauthentic accounts not only participate in disinformation campaigns in the political domain (Woolley 2020) but also take part in the moulding of cultural patterns, generation of content, and structuration of interactions. Crucially, then, bots connect and co-create with people in a grassroots-like manner, thereby gaining social, political, and cultural significance (Monaco and Woolley 2022). Hence, at least parts of digital culture and, therefore, potential heritage-to-be are co-produced by automated accounts. The question here is not only of identification (which can be extremely difficult) but also of ascribed value: Whether bot-affected content counts as heritage and, therefore, should be protected or not and, if the answer to that question is negative, how an adequate representation of today's culture and lived experiences can be preserved for posterity without a significant portion of everyday life that would be discarded due to its 'inauthenticity'. The same questions also pertain to the larger processes within which bots participate, such as coordinated disinformation and manipulation campaigns. Certainly, while nefarious and manipulative content, conspiracy theories, or hate speech might be seen as of lesser value (and often be automatically deleted by platforms), failure to protect such content would deprive future generations of important historical details and learning opportunities. However, should one side choose the protection option, the process becomes riddled with ethical challenges.

The discussion should not be limited to covertly inauthentic bots seeking to manipulate users. The growing domain of virtual influencers, for example, is a testament to the capacity of humans to enter parasocial relations with digital agents and, consequently, treat them as equal partners in cultural exchange (Block and Lovegrove 2021) or, even in situations where such a bond is not present, to meaningfully interact anyway (Lou et al. 2023). One could claim that such virtual personalities co-create cultural contexts with their human counterparts (Jiang 2022: 4). While such

virtual personalities will be discussed in greater detail later in this book, the preceding marks another step – hybridisation or, perhaps, even post-humanisation of digital heritage, raising questions as to what proportion of human and non-human (e.g., AI) contribution still merits protection and what falls below the benchmark (if there would be any). While the issue is similar (and directly related) to that of AI and creativity, distinctions and, therefore, dilemmas are less pronounced here. The default direction is already towards ever-greater expansion of digital-only content, co-produced by humans and AI or even exclusively generated by AI. The dilemmas as to what counts as heritage and ought to be protected (and the criteria used in decision-making) are largely occluded, frictionlessly sliding past this crossroads.

Post-Humanisation of Heritage?

The amount of content that has already been created and is stored in data centres around the world, combined with the rate at which content is being uploaded at any given moment, makes it impossible for humans to independently decide what to keep, but also what to access and use out of that which is already being kept. Finding the balance between the threat of content overflow and the risk of losing account of today's social fabric is already difficult, with strategies for curation and, thereby, forgetting becoming necessary (Haux, Dominicé, and Raspotnig 2021: 776, 778–779). Increasingly, decisions over preservation and loss (and, hence, memory and forgetting) would have to be made by automated tools or, at the very least, by humans in conjunction with AI, using such tools to sift through and select human-produced or hybrid (or, perhaps, even fully AI-generated) digital heritage.

While the preceding might seem like a practical imperative, the philosophical implications are huge: Instead of purely human-focused heritage (created and selected by humans for humans), we would now have co-structuration of what counts as heritage within an interplay of humans and machines – a post-humanisation of heritage. Here, one encounters both human agency in creation, sharing, and attention allocation and technological agency in sustaining (and often determining) the spread of content as well as, increasingly, its generation, thereby ultimately implying an 'entanglement between human and technical processes that shape born-digital heritage' (Musiani and Schafer 2017: 5). Additionally, problems generally inherent in AI tools would be extended into the domain of heritage. Should AI-powered tools, and machine learning models in particular, be applied here, questions of bias would more than likely arise: As one might expect examples from dominant cultures to be disproportionately present in the training data, so they would be in the results of selection processes at the expense of smaller or systematically excluded cultures,

the artefacts of which would simply be not recognised as worthy of protection. The problem is further exacerbated by the opacity of algorithms, whereby the operation of the key determining forces behind everyday culture and, therefore, heritage status of digital items lacks explainability. The preceding is also not helped by the increasingly private nature of digital archiving and storage, which sits uneasily with public concerns and cultural needs of communities (Musiani and Schafer 2017: 5).

However, post-humanisation processes are not entirely new and are already visible in algorithmic content governance, whereby the here-and-now of digital culture is largely determined by automated sorting, matching, and moderation of content that takes place on social media, streaming platforms, news apps, etc. Such governance of content – and, by extension, the personalisation of its supply – has a notable corollary: If 'the world that I see every day may be profoundly different from the one you see' (Susskind 2018: 230), then common culture and heritage are put in question. If cultural heritage is commonly understood as something that unites individuals into groups of shared experience and memory, the same cannot be said about personalised consumption of content. That, in turn, makes decisions on what is sufficiently significant and worth protecting even more complicated. Similarly, algorithmic processes of content governance have also affected the notion of time – at least in a cultural and social sense. While much of the talk still is of the digital world being premised on real-time interactions, Bucher (2020: 1711) rightly redirects focus towards what she calls 'right time', which functions as 'a new temporal regime produced by an increasingly algorithmic media landscape': Instead of horizontal simultaneity, implied by real time, right time draws on resources and artefacts available across the time spectrum so that every user experiences the right moment as and when necessary. For this reason, not only homogeneity (as per above) but also simultaneity of cultural experience cannot be taken for granted, making heritage worthiness of artefacts increasingly fluid and dependent on data-informed algorithmic matching of users with content in any here-and-now at any time.

A further step down the posthuman path is considering the question of AI-generated content and its status as cultural heritage. While the domain of heritage has shown significantly less anthropocentrism than other domains pertaining to culture and artefacts (e.g., natural heritage comprises a large proportion of what is currently protected, thereby demonstrating that human intentionality in creation is unnecessary), the ease of generation and, therefore, the rapidly growing amount of AI-generated content raises questions about its (non-)inclusion into the category of heritage. The preceding is also made even more acute by the ongoing discussions as to the value and protectability of AI-generated content and the latter's role in (human) culture, discussed previously. However, if a decision is made not to include AI-generated content, its detection and removal

from heritage repositories is complicated by the absence of sufficiently reliable detection tools. Also, the growing importance of such content means that cultural interactions would possibly become unintelligible for future generations if AI-generated content is removed from the equation.

A no less important question is that of values that should guide inclusion/exclusion. It would be difficult to argue with 'the importance of sustained attention to and facilitation of meaningful community participation in efforts to safeguard their digital cultural heritage' (Hennessy 2012: 365). However, particularly in the digital interconnected world, this begs the question of what such communities refer to and what the right point of reference is: After all, artefact-producing communities are now often amorphous and lacking clearly defined boundaries and geographical points of reference. Likewise, while UNESCO has also since moved towards acknowledging the need to develop guidelines for the selection of digital heritage in cooperation with citizen groups and civil society institutions (von Schorlemer 2020; see also UNESCO 2015), clarity as to who is to be consulted (selection of stakeholders) is lacking, particularly as the producers of content often tend to be anonymous and, when content thrives in groups, such groups could be fleeting or exist on closed and encrypted platforms and thus be, by definition, difficult or impossible to grasp. Moreover, following the text of the Charter, the selection of what is to be preserved should be based on 'significance and lasting cultural, scientific, evidential or other value' of the digital content in question (UNESCO 2003, Art. 7). However, how such criteria should be set and by whom remains unclear, as standards and interpretations may vary between states and cultures, let alone between states and private actors, such as platforms, that do not yet appear in the Charter. The latter omission is, of course, only natural since the Charter was drafted prior to platformisation tendencies becoming evident in the digital domain. Nevertheless, the interests of platform and other technology companies are likely to significantly differ from those of communities and states, making frameworks for their reconciliation a necessity and implying a need for a choice of direction at a crossroads of competing options, defined by values (particularly around the notion of the human) and interests vested in the digital economy.

Platformisation of Heritage?

Continuing on the matter of platforms, while significant portions of cultural heritage have always been privately owned, for example, by individuals, religious organisations, commercial entities, etc., with the platformisation of the internet, not only have private actors (platform companies) become owners of unprecedented amounts of culturally significant content (particularly user-generated) but they have also become key custodians and distributors of much of the rest (streaming platforms being

notable examples here), thus clearly revealing the scope of privatisation of culture, including the definition, the enabling, and the preservation of the latter (see, generally, von Schorlemer 2020: 52; Bareikytė et al. 2024: 1379). In this context, governing user attention becomes the key attribute of economic, social, political, and cultural power, which has largely been self-arrogated by online platforms (De Gregorio 2022: 179). In fact, platforms already can be seen as occupying a central position in contemporary societies as they 'do not reflect the social: They *produce* the social structures we live in' (van Dijck, Poell and de Waal 2018: 2), leading towards a shift from a free to an algorithmic marketplace of ideas (De Gregorio 2022).

It is notable that '[b]y virtue of the governance of their digital spaces, online platforms also perform autonomous quasi-public functions': Setting norms (what content and behaviour is allowed and what is disallowed) and enforcing them by overseeing 'the technological architecture of their digital spaces' (De Gregorio 2022: 81–82). Platforms now act as gatekeepers due to being the providers of both value and meaning to cultural artefacts and their creators; they also 'mediate tastes, moods, and lifestyles, converting them into valuable objects of consumption', thereby providing a new form of structuration of everyday life, cultural life included (Bonini and Gandini 2019: 2). What matters culturally, therefore, is what platform algorithms *select* to matter based on user data, thus further strengthening human-technology hybridity in the determination of culture and, therefore, heritage. This also works in terms of corporations keeping a keen eye on 'their' territories and 'their' data domains (Lambach 2020: 499). Hence, the ever-growing role of platforms makes cultural processes inseparable from the commercial logics of technology companies: Decisions on content governance, as well as on inclusion and exclusion of particular content items, cannot be separated from efforts to maximise profits rather than cultural value (Elkin-Korel 2020).

A further issue is that digital heritage must be preserved in its context and functionality rather than as a 'data graveyard', that is, instead of being downloaded and saved somewhere in a decontextualised fashion, it must continue functioning within the interlinked and interconnected digital environment (Dominicé and Haux 2020: 256). On the one hand, this approach imposes clear restrictions on potential selection of content to be preserved and that which is condemned to being forgotten: Any act of discarding content should not preclude that which is preserved from functioning effectively (which is a rather complicated endeavour given the interconnectedness of the digital ecosystem). Moreover, if content needs to be preserved in its original context, platform companies, particularly social media ones, effectively become ecosystems of both digital present *and* digital heritage as well as custodians thereof. While creating additional obligations for platforms, this view also implies new societal dependencies, whereby the identity and memory of societies are effectively outsourced and privatised.

Indeed, one must keep in mind that '[c]ulture is a process instead of a product, performance and enactment rather than artefacts, the role of communities and groups as bearers of culture' (de Lusenet 2007: 175). The problem, however, is that it is increasingly difficult to, metaphorically speaking, see the forest through the trees: The more artefacts are available and the more our capacity for capture and storage increases, the greater the difficulty in keeping track of processes and interactions – unless, once again, some form of enmeshment between humans and technology in setting an authoritative account of what counts as culture and heritage is presumed. Such a shift would, however, yet again underline a move towards digital technology companies becoming cultural and memory institutions.

Meanwhile, the focus of the Charter remains fundamentally state-centric (see, e.g., Article 7 on standard-setting by states and Article 8 on state legislation and policies). While this is not surprising given UNESCO's membership and the time of the Charter's drafting, the absence of frameworks for inclusion of non-state actors and multi-stakeholder cooperation was hardly sustainable already back in 2003 (particularly due to the global nature of the digital domain) and is even less so today. While Article 11 mentions partnerships, including with creators, publishers, and industries, the rise of powerful new actors and intermediaries, such as platform companies, has significantly altered the game. Likewise, states are urged to 'develop public bodies enabling and supporting digital heritage in a rapidly changing technological environment' and to 'promote cooperation between their legislative bodies and archives, libraries and museums'; crucially, any monitoring of the independence of heritage institutions and access to preserved content is also limited to those same traditional memory institutions (UNESCO 2012). In other words, the basic frameworks that applied to the analogue or Web 1.0 world are still maintained, leaving no answers as to who should be responsible for deciding on the heritage status and protection of, for example, the latest globally viral meme. While states are urged to 'encourage private sector organizations to invest in trustworthy digital infrastructure and digital preservation' (UNESCO 2012), emphasis here remains on the businesses that produce digital data as a byproduct of their daily operations, leaving societies without clear guidelines at the crossroads that feature the relationship of AI with cultural heritage.

While, with regard to creativity, the transmutation of technological possibility into an imperative was seen, in the previous chapter, as still less clear, the platformisation of heritage is increasingly developing along such lines. The private algorithmic governance of the production, maintenance, and curation of heritage is likely to prioritise platform business practices and content governance models (which, though machine learning, tend to be skewed towards the most data-generating cultures), rendering anything

outside as non-value. This means introducing adiaphorisation into a domain that otherwise ought to be the latter's antithesis – after all, heritage is supposed to be about constant noticing and caring.

References

Abid, A. (2007). Safeguarding Our Digital Heritage: A New Preservation Paradigm. In Y. de Lusenet and V. Wintermans (eds.). *Preserving the Digital Heritage: Principles and Policies*. The Hague: Netherlands National Commission for UNESCO, pp. 7–14.

Alker, Z. and Donaldson, C. (2018). Digital Heritage. *Journal of Victorian Culture*, 23(2), 2020–2021.

Bareikytė, M. et al. (2024). How Should Platforms Be Archived? On Sustainable Use Practices of a Telegram Archive to Study Russia's War Against Ukraine. *Media, Culture & Society*, 46(7), 1378–1396.

Block, E. and Lovegrove, R. (2021). Discordant Storytelling, 'Honest Fakery', Identity Peddling: How Uncanny CGI Characters are Jamming Public Relations and Influencer Practices. *Public Relations Inquiry*, 10(3), 265–293.

Bonini, T. and Gandini, A. (2019). 'First Week is Editorial, Second Week is Algorithmic': Platform Gatekeepers and the Platformization of Music Curation. *Social Media + Society*, doi: 10.1177/2056305119880006

Brevini, B. (2020). Black Boxes, not Green: Mythologizing Artificial Intelligence and Omitting the Environment. *Big Data & Society*, doi: 10.1177/2053951720935141

Bucher, T. (2020). The Right-Time Web: Theorizing the Kairologic of Algorithmic Media. *New Media & Society*, 22(9), 1699–1714.

Chan, M. (2022). *Digital Reality: The Body and Digital Technologies*. New York: Bloomsbury Academic.

Dahlgren, P. (2018). Media, Knowledge and Trust: The Deepening Epistemic Crisis of Democracy. *Javnost – The Public*, 25(1), 20–27.

De Gregorio, G. (2022). *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society*. Cambridge: Cambridge University Press.

de Lusenet, Y. (2007). Tending the Garden or Harvesting the Fields: Digital Preservation and the UNESCO Charter on the Preservation of Digital Heritage. *Library Trends*, 56(1), 164–182.

Delwiche, A. (2020). Computational Propaganda and the Rise of the Fake Audience. In P. Baines, N. O'Shaughnessy and N. Snow (eds.) *The SAGE Handbook of Propaganda*. Los Angeles and London: SAGE, pp. 105–125.

Dewandre, N. (2020). Big Data: From Modern Fears to Enlightened and Vigilant Embrace of New Beginnings. *Big Data & Society*, doi: 10.1177/2053951720936708

Dominicé, A.M. and Haux, D.H. (2020). The Decision of the German Federal Court of Justice against Facebook: Opportunity to Define Digital Heritage? *Santander Art and Culture Law Review*, 2(6), 251–260.

Elkin-Korel, N. (2020). Contesting Algorithms: Restoring the Public Interest in Content Filtering by Artificial Intelligence. *Big Data & Society*, doi: 10.1177/2053951720932296

Haux, D.H., Dominicé, A.M. and Raspotnig, J.A. (2021). A Cultural Memory of the Digital Age? *International Journal of the Semiotics of Law*, 34, 769–782.

Hennessy, K. (2012). Cultural Heritage on the Web: Applied Digital Visual Anthropology and Local Cultural Property Rights Discourse. *International Journal of Cultural Property*, 19, 345–369.

Hepp, A. (2020). *Deep Mediatization*. London: Routledge.

Ibrahim, Y. (2018). *Production of the 'Self' in the Digital Age*. London: Palgrave Macmillan.

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

Kalpokas, I. and Kalpokiene, J. (2023). *Regulating the Metaverse: A Critical Assessment*. London: Routledge.

Krishnan, A. and Hunt, D.S. (2021). TTYL :-) … Nonverbal Cues and Perceptions of Personality and Homophily in Synchronous Mediated Communication. *Information, Communication & Society*, 24(1), 85–101.

Lambach, D. (2020). The Territorialization of Cyberspace. *International Studies Review*, 22(3), 482–506.

Lor, P.J. and Britz, J.J. (2012). An Ethical Perspective on Political-Economic Issues in the Long-Term Preservation of Digital Heritage. *Journal of the American Society for Information Science and Technology*, 63(11), 2153–2164.

Lou, C. et al. (2023). Authentically Fake? How Consumers Respond to the Influence of Virtual Influencers. *Journal of Advertising*, 52(4), 540–557.

Monaco, N. and Woolley, S. (2022). *Bots*. Cambridge: Polity.

Musiani, F. and Schafer, V. (2017). Digital Heritage and Heritagization. *RESET: Recherches en Sciences Sociales sur Internet*, 6, 1–10.

Pentzold, C. and Bischof, A. (2019). Making Affordances Real: Socio-Material Prefiguration, Performed Agency, and Coordinated Activities in Human-Robot Communication. *Social Media + Society*, doi: 10.1177/2056305119865472

Pink, S. et al. (2017). Mundane Data: The Routines, Contingencies and Accomplishments of Digital Living. *Big Data & Society*, doi: 10.1177/2053951717700924

Shifman, L. (2014). *Memes in Digital Culture*. Cambridge, MA: The MIT Press.

Susskind, J. (2018). *Future Politics: Living Together in a World Transformed by Tech*. Oxford: Oxford University Press.

UNESCO (2003). Charter on the Preservation of the Digital Heritage. Ref.: CL/3865, 17 October, https://unesdoc.unesco.org/ark:/48223/pf0000179529

UNESCO (2012). Vancouver Declaration. The Memory of the World in the Digital Age: Digitization and Preservation, https://unesdoc.unesco.org/ark:/48223/pf0000373728.locale=en

UNESCO (2015). *Recommendation concerning the preservation of, and Access to, Documentary Heritage Including in Digital Form*, 17 October, 38 C/Resolution 55, Annex V.

Vaidhyanathan, S. (2018). *Anti-Social Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford: Oxford University Press.

van Dijck, J., Poell, T. and de Waal, M. (2018). *The Platform Society: Public Values in a Connective World*. Oxford: Oxford University Press.

von Schorlemer, S. (2020). UNESCO and the Challenge of Preserving the Digital Cultural Heritage. *Santander Art and Culture Law Review*, 2(6), 33–64.

Woolley, S. (2020). *The Reality Game: How the Next Wave of Technology Will Break the Truth and What We Can Do About It*. London: Endeavour.

5 Towards the (Last) Frontier
AI Persons

I See Virtual People

Social media already is one domain in which humans and human-like AI-powered entities mix – it is also a domain where entities, such as virtual influencers (VIs), are followed by humans (Mosley 2020). With the broader rise of synthetic media (digital content wholly or partly generated by AI), virtual personalities and, more specifically, VIs are no exception – in fact, they are merely part of a broader trend (Sands et al. 2022: 778). Moreover, as shown below, such digital entities can compete with digitised representations of flesh-and-blood humans (as social media profiles or as avatars).

For audiences to establish relationships with virtual personalities, they must first be able and willing to suspend the knowledge of the latter's unreality, interacting with them and developing feelings of attachment as if they were human (Gerrath et al. 2024: 3; see also Kim and Wang 2023). While the term 'parasocial relations' originally described the imagined relationships audiences have with celebrities or media characters, it is equally apt here in describing the subjective fiction of a reciprocal relationship – or any relationship, for that matter (Zhou et al. 2024: 2). However, establishing parasocial relations with AI agents also reflects a broader tendency to anthropomorphise AI: As the latter is difficult for most human users to comprehend (due to technological complexity and lack of transparency), projecting human-like qualities and trends becomes a coping strategy (see, e.g., Barrow 2024). Generally, anthropomorphising means 'the ascription of human qualities (e.g., intentions, motivations, human feelings, behaviors) onto non-human entities (e.g., objects, animals, natural events)'; possibly an evolutionary adaptive trait, it does not actually relate to any specific qualities of the object in question but is solely a matter of human interpretation of their surroundings (Placani 2024: 2). There is a widespread tendency to anthropomorphise AI as such (even the use of the term 'intelligence' is clearly anthropomorphic), to the effect of obscuring the actual capacities and limitations of the underlying

DOI: 10.4324/9781003715849-5

technology (Placani 2024). This tendency strengthens when an artificial entity is given a human appearance and seems capable of human-like interaction, thereby enabling a feeling of connection.

A VI can be defined as 'a fictional character on social media operated by a third party (freelance creator, digital agency or brand) who defines its appearance, personality and storyline' (Koles et al. 2024: 2). These digital entities transcend their virtual confines not only by collaborating with real-life brands, but also by appearing alongside real-life celebrities (Sokolov 2019; Bradley 2020a). Often, they are also socially unconfined, mixing endorsements of products and social movements and/or captioning real-world affairs (Delibasic 2018). Even if currently this is mostly about pitching the VI to a demographic by including trending concerns, there may be ample opportunities for shaping that same demographic further down the line, when the cultural clout of that VI is already established.

Broadly, VIs fall into two major groups: General-purpose ones and those closely attached to single brands. The former 'are marketed as online personalities who just happen to be interested in fashion' (or anything else that pays), in a way very similar to human influencers; the latter, meanwhile, are simply private property of the brands that had purposefully created them (Sokolov 2019). Hence, while the first type essentially mimics the modus operandi of human influencers, the second one is VI-native. However, both types equally follow the already established practices of the industry, particularly in terms of influencers being 'perceived as accessible, friendly, people next door who offer unbiased opinions and therefore possess high credibility and trustworthiness', even though neither of such impressions may have a reasonable claim to representing reality (Gräve 2019: 2). In other words, influencer marketing is already an industry that deals in appearances anyway. Seen in this way, VIs may simply be adding a further layer of appearance.

Moreover, it is also possible to trace a certain progression of (un)reality, whereby the distinction between reality and image has first been erased by marketing, then the line between real and manufactured identity was erased by social media; once those two converge, the logical next step would be the creation of virtual personalities (Turner-Williams 2020). In each case, such (un)reality also allows for control and strategic manufacturing of the audience's perceived choice environment. Crucially, 'virtual influencers are a painstakingly crafted vision of what their creators have assumed audiences will be interested in engaging with' (Torossian 2020). These assumptions, however, are not merely plucked out of thin air – they are, instead, based on far-reaching data analysis. Hence, VIs must not be seen as mere sock puppets of their creators – instead, they have 'a composite personality based on market research' that allows understanding of the target market's attitudes and tastes as well as, potentially, on real

people who may be seen as inspirational by the target audiences (Bradley 2020a). Such tailoring goes down to the detail, including sourcing audience preferences for 'age, gender, tone of voice and aesthetics' (Bradley 2020a) – certainly, no real-life influencer could be nearly as malleable. Quite naturally, the preceding also raises important issues with regard to potential audience manipulation.

Notably, it makes sense to understand VIs as embodied stories: perhaps the most audience-captivating aspect of VIs is that their 'lives and journeys' tend to be 'brought to life through emotional storytelling and empathy' as a dominant feature – in fact, the story often tends to matter even more than the actual underlying substance (or, in fact, lack thereof) – and that is precisely the aim (Luthera 2020; see also Allal-Chérif, Puertas and Carracedo 2024). Particularly for the more general-purpose VIs, an entire backstory is invented to make the synthetic person's actions seem relatable. To an extent, this might be like watching a film or reading a novel: As we get carried away by the twisting plot and the emotional journeys of the hero(ine), the awareness of fictitiousness easily gets suspended. It seems that users are not only not fussed about the virtual status of the entities they follow but also find that these can become truly beloved virtual idols (Vogue Business 2020). Again, that is not without precedent: Infatuation with film or cartoon characters, dedicated following, purchase of merchandise, etc., are all staples of contemporary entertainment culture and major revenue streams for their creators (Yates 2020). Not only do we get hooked while consuming the content, but it is also not unheard of for fictional characters to inspire fashion and lifestyle trends – even without a deliberate aim to that effect; just think of the many Harry Potter fans out there. Likewise, VIs are not too dissimilar from characters in, for example, TV series or comic books that their audiences understand to be fake and yet are willing to suspend that knowledge to jointly partake in the story or even to take it further, for example, through fanfiction (see, broadly, Allal-Chérif, Puertas and Carracedo 2024; Yu et al. 2024). It is, therefore, not much of a stretch to consider the profiles of VIs (at least general-purpose ones) as a contemporary form of fiction, following in the footsteps of literature and film, and simply taking fiction to a new media form and into live action. VIs likewise take the tendency of audience immersion further due to being specifically designed for maximum impact.

There are good reasons to agree with Robinson (2020: 2–3) that even though VIs may not exist in the physical sense, particularly the more successful ones, nevertheless possess the cultural clout, freestanding life stories, and the follower numbers that make them real through their own effects – strongly enough to 'exert considerable influence on the preferences, buying habits, and trends of [their] followers'. In this way, when the audience encounters posts by VIs, the relationship created does not seem to be distinguishable from that with real-life ones, thereby rendering the

ontological status of different influencers effectively undistinguishable, at least within the social media environment (Robinson 2020: 3).

Of course, forming an emotional connection is key to overcoming the artificiality of VIs in terms of interacting with the latter and even taking product or lifestyle advice from them (Yu et al. 2024: 7). For that, however, VIs and other digital personas must be able to establish social presence, which includes possession of human traits and at least rudimentary communicative capacities while also channelling a sense of purpose and usefulness (Kim and Park 2024). Consequently, there is a need for VI creators to strategically plan and create to elicit feelings of connection and attachment vis-à-vis a specific audience, which is, of course, helped by the fact that a VI is a blank slate that can be shaped to any audience's needs (Na, Kim, and Lee 2024).

Crucially, even human influencers are 'virtual' to a certain extent. First, the images that they strategically craft may be rather far removed from the actual person. Second, even their influencer status could easily be fake, as the impressive follower numbers can be bought from any vendor trading in troves of bot or sock puppet accounts (Gräve 2019). With VIs, by contrast, their well-known artificiality can act as a kind of reassurance: There is simply less risk of being fooled by staged and heavily edited photographs or hidden native advertising if everything is fake anyway, and such virtual personalities are manifestly created for profit (Sands et al. 2022: 779). Nevertheless, there might still be an acute emotional connection gap when it comes to VIs posting about specifically human experiences (Bradley 2020b) or about products that necessitate human perception or enjoyment to be fully appreciated (food or music could be obvious examples).

The data on the effectiveness of VIs are, however, mixed. Some authors have reported lower engagement and trust in product recommendations among audiences when endorsement comes from a VI (see, e.g., Zhou et al. 2024: 7). Others (see, e.g., Mirowska and Arsenyan 2023) have shown that less human-like (e.g., cartoonish) VIs attract more positive attention than the more realistic ones. To the contrary, Quach, Cheah, and Thaichon (2024) and Balakrishnan and Dwivedi (2024) report a *positive* effect of anthropomorphism on the quality of user interaction with VIs. Meanwhile, still others find no clear difference between the effectiveness of human and VI endorsements, particularly if the creators of a VI manage to cultivate the latter's reputation for trustworthiness and audience-likeness (see, e.g., Byun and Ahn 2023; see also Ferraro et al. 2024). It is also noteworthy that the effectiveness of VIs may depend not only on their virtual presence but also on the product endorsed: For example, as Franke, Groeppel-Klein, and Müller (2023) demonstrate, products that are less tied to human embodiment (such as consumer electronics) are a better fit to VIs and elicit more positive audience engagement than, for example, cosmetics.

While people may have learned to doubt the authenticity of human influencers, there usually exists at least a reason to believe they have tried the endorsed product (after all, there are all these glorious pictures to prove it). VIs, meanwhile, by definition, cannot have done the same. While the followers of VIs seem to trust the latter's taste (as is the case with human influencers), it is easy to forget that they do not, and cannot, have taste (Cook 2020). Nevertheless, a VI 'can appear more authentic and sincere than a human influencer, who has a completely invented existence and who only plays successive roles that are inconsistent with each other' (Allal-Chérif, Puertas, and Carracedo 2024: 11). In a way, then, VIs can be understood as at least 'authentically fake'. Indeed, as Kim and Wang (2023: 3) observe, VIs could be considered 'the most authentic social media influencers, as they are entirely honest and transparent'. That is, of course, provided their creators *are* transparent about the artificiality of a given VI, which should not be taken for granted. Still, authenticity can be hampered by the concentration of underlying technology: As most of the latter has been created by Western technology companies and based on Western norms and assumptions, the result could be, to some extent, alien to audiences in other parts of the world (see, e.g., Ndlovu 2024).

There are ample reasons why VIs could be seen as commercially attractive. For example, they could be seen as more ethical and environmentally friendly options because they do not need to travel and thus have a lower carbon footprint (Bauck 2020). The latter perhaps is a signal of further developments: Virtual representations of actual selves being used as stand-ins in everyday dealings or digital embodiments of real-life influencers and celebrities being licensed for commercial purposes – in other words, a digital-first future. Also, while previously brands had to swap control for user engagement and perceived authenticity (i.e. their brand message had to be mediated, and potentially diluted, by a human influencer), with their private VIs (and even with the general-purpose ones), they can have near-complete control over the message content, but with the same or similar user engagement a human influencer could offer. Indeed, one of the selling points of VIs is that their image 'can be carefully controlled, cultivated, and developed over time, which allows desired associations to be emphasized' (Gerrath et al. 2024: 3). In this way, influencers can be aligned with a company's vision of itself and its brand to an unprecedented level. Moreover, there is also no risk of the influencer discrediting the brand through their real-life behaviour, because there is, of course, no real-life behaviour, except for the character story that is carefully scripted and represented through the VIs' social media accounts (Allal-Chérif, Puertas and Carracedo 2024; Gerrath et al. 2024; Yu et al. 2024). Indeed, it can be asserted that '[t]he main quality of virtual influencers is their absence of human flaws', meaning that '[t]hey are not late, they do not forget anything, they do not betray, they do not age, they do not rest, nor do they ask

for a salary when they are developed in-house' (Allal-Chérif, Puertas, and Carracedo 2024: 10). The VI speaks what and how it is instructed to, ensuring consistency and predictability (Haile 2020).

In the overcrowded digital environment, it is natural that fatigue kicks in among the audiences as they become overwhelmed by more of (the same or similar) communication; VIs, being different by definition, may help offset this trend (Koles et al. 2024: 2; see also Gerrath et al. 2024). That is, of course, until the number of VIs increases exponentially, overwhelming platform users. Given the relative ease of creating VIs (particularly due to advances in AI) and the absence of industry or platform regulation, that seems to be a matter of when, not if.

Simultaneously, full control over a VI's online behaviour also means assuming full responsibility if anything goes wrong – after all, not only real-life influencers, but also professional content creators are not immune to gaffes and instances of poor judgement. Moreover, some of the structural problems endemic in the influencer industry are exacerbated by the synthetic nature of these influencers, including impossible beauty and lifestyle standards, which could possibly be made even more unachievable due to the absence of physical reality constraints (Bradley 2020a). Notably, 3D rendering software produces models of virtual persons that are perfect: Symmetrical, without blemishes, etc.; in fact, any bodily imperfections visible on VIs must be added *deliberately* to make them more relatable (Turner-Williams 2020). In the same way, as Rodríguez Monclou (2020) stresses, 'they will not age and remain identifiable and youthful for brands, a brilliant but creepy marketing strategy in a world seeking perfection'. Again, signs of ageing can be added deliberately if that is seen to contribute to relatability, but whether such a choice is made or not completely depends on a given cultural context and target audience. Virtual entities would likely have a physical world impact, including in terms of becoming reference models for humans and impacting human self-perceptions, in addition to shaping their opinions and choices. Consequently, questions ought to be asked about the setting of standards for humanness (of both virtual humans and, indirectly, embodied humans) and whether certain types of influence should be kept out of bounds.

Moreover, VIs tally well with the broader trend towards virtualisation and, correspondingly, diminishing value of the tangible: For example, even some fashion labels are now offering virtual clothing items that can be purchased to adorn a customer's photos. In this case, there would likely be no better way to enhance the visibility of virtual items than to have a VI endorse them. Likewise, part of the emergence of VIs might be cultural: Just like reality TV stars had become an epitome of celebrity of the past several decades, once everyday life becomes more mediatised and virtual, this shift might well necessitate equally virtual celebrities to embody it (Sokolov 2019). Moreover, VIs can be seen to satisfy escapist needs as

they 'act as a form of diversion by bridging real and imaginary worlds', and 'allow consumers to immerse themselves in an alternate reality' (Sands et al. 2022: 779; see also Mirowska and Arsenyan 2023). Consequently, VIs are part and parcel of emergent virtual-first (or, at the very least, digital-first) worlds that offer experiences both alternative and superior to daily lives. While the human influencers can also perform an escapist function, VIs can simply offer more all-rounded immersion and increased interactivity.

Crucially, the lack of 'reality' of VIs and the allegedly ensuing lack of authenticity that some authors emphasise (Forsey 2019) might even be beside the point – what might really matter is a sense of virtual intimacy. The growing popularity of VIs seems to indicate that people do not particularly care about their lack of conventional reality – as long as the audience identifies with what they see and hear, the question of reality becomes suspended. That might explain why even VIs that are decidedly cartoonish and unrealistic manage to craft their niche, including marketing very real goods (Sokolov 2019). Moreover, it transpires that 'consumers anthropomorphize machines and adhere to social rules when interacting with them' while also 'building rapport and projecting a feeling connectedness, even if instrumentally, to reduce uncertainty that may otherwise arise in human-virtual interaction' (Sands et al. 2022: 779–780). This, as well, might indicate a tendency to treat virtual entities as equal parts in an interaction. Also, while it is technically correct to point out that a VI is 'nothing more than a coded puppet at the hands of the people and brands who control it' (Haile 2020), it is the invested meaning that really counts.

If individuals dedicate time and effort to following a VI and form an emotional attachment (and, potentially, also invest actual money in merchandise), this in itself becomes sufficient to subjectively validate user following. Furthermore, a question must be asked as to what happens when some degree of AI-based autonomy kicks in. This will, in the foreseeable future at least, not be anything approximating full autonomy, but, rather, (semi-)automated appearance and behaviour rendering, based on the available audience data and geared simply towards maximising engagement. Nevertheless, the immediacy (and, to some extent, authenticity) of audience engagement is going to be transformed even further, pushing the boundaries of identification, emotional investment, and (perceived) reality.

VIs are moving into areas that stretch beyond product endorsements and are increasingly treated in the same way as human talent, paving the way to TV shows, film performances, or music (Spangler 2020). In this way, virtual general-purpose VIs can morph (and are often deliberately designed to morph) into multisided multimedia packages that can bring in multiple revenue streams to their creators (Yates 2020). Likewise, such an ability to straddle multiple contexts (social media, virtual worlds, music, film, or any combination thereof) also expands opportunities for parasocial relations

between audience members and VIs. A likely further development is the emergence of virtual person-like entities that either straddle the divide between virtual and offline or that are impossible to tell apart from such mixed personalities. That is likely to happen once human celebrities (such as actors, musicians, etc.) feel the pressure to also double in virtual forms, such as XR performances (Turner-Williams 2020). This trend is already starting to materialise, and once it is in full swing, the digital world will be inhabited by an indistinguishable mixture of purely virtual persons, those that are highly modified versions of real persons, and more or less faithful representations of human individuals.

Notably, VIs allow for frictionlessness and optimisation of interactions through their crafting and data analytics, functioning as interaction partners seemingly on par with flesh-and-blood humans. While in purely functional terms – such as purchase intention – the choice seems to be still open, perhaps because, just like with creativity, a fundamental aspect of humanness is at stake. Nevertheless, the frictionlessness combined with escapism and the capacity to enter into parasocial relationships together create the conditions for an uptake of virtual persons, with the potential of straightening the path from mere technological possibility to an imperative.

What Is Real When Everyone Is Virtual?

When discussing the possibility of virtual representations (avatars) of the self, either of celebrities or of ordinary individuals, the question of the value of each individual human must be raised. Essentially, the value of an individual (just like the value of anything) is based on scarcity: We are finite beings, with a limited life expectancy, capable of only being in one place at a time; hence, we can only accomplish a tiny fraction of things that are generally possible to do, which adds value to us being at a particular place at a particular time and with particular people. The ability to virtually multiply ourselves, it can be claimed, reduces the scarcity and exclusivity of our presence and, therefore, our value (Danaher and Nyholm 2024). It, thus, remains an open question whether the availability of virtual incarnations makes the life of an individual that is multiplied (transformed from an individual to a dividual) less valuable (Danaher and Nyholm 2024). Here, value, however, ought to be understood in a specific sense. Let us consider a musician who is able to perform multiple concerts at once in multiple virtual venues (or in multiple physical venues as a hologram). To some extent, value can even be increased in both a monetary sense (the musician in question can earn more money than when doing one concert at a time); moreover, value can be increased for audiences as more people can experience the satisfaction of attending a performance. However, something about that musician would become devalued: Exclusivity, being special, and, at least to some extent, being desirable would suffer if their

virtual versions were everywhere. Likewise, for any individual, such multiplication (dividualisation) would have some benefits (maximising achievement while simultaneously increasing social interaction and even personal relationships), but simultaneously, the value of interacting with them would be reduced since there would be nothing special about having them (or their avatar) around. Of course, unless this avatar is a perfect clone, capable of even simulating a person's thoughts, perspectives, and emotions, the scarcity of that core essence of the self would remain unaffected (Danaher and Nyholm 2024). The preceding also provides an opening for alternative views: If we assume that every person is unique in some way, perhaps that essence remains just as important; nevertheless, what is at stake is that the *perception* might change. Effectively, if digital clones are multiple and available and if they become the new normal, devaluation of human life and human person might happen simply because others may start seeing uniqueness and scarcity (and, by extension, the authentic human person) as unimportant while seeing the avatar as just as good as the underlying person (Sweeney 2024).

A much more troubling possibility is that of freezing a person at a particular stage in their life (see, e.g., Danaher and Nyholm 2024). This might involve presenting their younger, allegedly more 'attractive', selves (it must be remembered that an avatar does not age or otherwise change in appearance unless specifically programmed), thus bringing back the issue of appearance standards and conformity to stereotypical models. Likewise, avatars may allow the preservation of individuals as they were prior to life-changing events, such as injuries or illnesses, thus strengthening the role of neuro and physical normativity, rendering divergence as a shortcoming or an adverse condition that allegedly should be hidden or otherwise obfuscated. No less importantly, virtual avatars would offer the possibility of creating a restricted version of a person, for example, without some traits that are considered dangerous, toxic, or simply unpleasant. That, in turn, would threaten personal integrity and allow cherry-picking of traits in accordance with preset norms.

No less disturbingly, once virtual persons can be generated and their behaviour portrayed in real time, for multiple users simultaneously in accordance with their requests, the virtual persons business could be predicted to turn towards fulfilment of the darkest fantasies of consumers, particularly if effective guardrails are not implemented. Likewise, even the same virtual person can behave differently towards different users or audiences, thereby further challenging ideas around character continuity and personal integrity. The notion of a unique person with a clear moral, ideological, and epistemic continuity or a clear and identifiable personal story might cease to be viable or, at least, may no longer be expected.

A further issue would be that of autonomy. While the current generation of VIs as well as personal doubles in general are completely

dependent on their creators, it is not difficult to imagine a (perhaps not-so-distant) future in which they become (relatively) autonomous AI agents, for example, capable of learning what is trending at a given moment and then tapping into it or autonomously developing their character story based on some rudimentary brief (not too dissimilar from a generative AI prompt). Under such circumstances, questions of responsibility and accountability are bound to arise, not just in terms of directly violating the rights of others but also in terms of, for example, nudging their followers towards socially undesirable, dangerous/reckless, or maybe even harmful/violent behaviours (see, generally, Kim and Wang 2023). In this case, the standing of virtual personalities and their relations with flesh-and-blood humans become even more problematic, likely necessitating new regulatory solutions.

Interactive virtual entities, even short of meaningful autonomy, also emerge as machinic interlocutors. Such entities are not just artificial partners (even though, on the ontological level, they are precisely that) but, instead, can be seen as participants in the human meaning-creation process (Xu and Shi 2024). Again, it must be remembered from the above that humans appear to have the capacity to emotionally engage with digital entities, including those as simple and commonplace as virtual assistants, such as those on smartphones and smart home speakers, by anthropomorphising them and seemingly erasing the awareness of their artificiality (Guerreiro and Loureiro 2023) – entering into parasocial relations. Even something as simple as interacting with generative AI tools manifests features of inter-human relationships: '[w]e talk to them, say please and thank you, and have started to invite AIs into our lives as friends, lovers, mentors, therapists, and teachers' while also, at least in some cases, forming emotional reliance (Heikkilä 2024).

Simultaneously, greater perception of autonomy in chatbots has been demonstrated to lead towards stronger trust in the artificial communication partner and increased satisfaction in the interaction as such (Pan et al. 2024). When AI-enabled conversational capacities and human likeness blend together through at least a virtual embodiment, some users are willing to go as far as to report the emergence of something akin to a 'soul' (see Leo-Liu and Wu-Ouyang 2024). Such a combination has been demonstrated to matter even in completely utilitarian settings, such as sales, by increasing perceived interaction quality (see, e.g., Zhu et al. 2024). While perhaps not yet fully replacing interactions with humans, with some level of anthropomorphisation, at least part of those virtual interlocutors may pass as good-enough replacements. Once again, anthropomorphism transpires to be a significant factor in interaction and formation of emotional attachment to virtual personalities and companions to the extent that some users were seem to project a 'human soul' into such digital entities; the flip side, as might be expected, is that the high

expectations by definition cannot be fulfilled and, therefore, are likely to backfire, resulting in user disappointment (Pentina, Hancock, and Xie 2023: 10). Indeed, users appear to seek genuine and authentic personality in their AI companions; nevertheless, given the current state of AI technology, such companions end up being 'deplored when associated with scripted, programmed answers, and pre-chosen personality' – after all, once somebody has invested time, effort and, even more significantly, formed an emotional attachment, the least they want is a virtual partner that is clearly machine-like (Depounti, Saukko and Natale 2023: 726).

Some, however, expect technological capacities to quickly expand, resulting in computational empathy and language generation capacity being combined to give rise to completely realistic virtual persons, including those capable of simulating high-quality and satisfying interactions (see, e.g., McStay 2023). Indeed, the simulation (and simulation *is* all it takes) of human-level intelligence and empathy tends to be the strongest factor in establishing human-like connections with AI entities (Kim and Im 2023). Nevertheless, even under present technological capacities, deliberate design choices, such as the availability of human-like voice, linguistic cues, human-like appearance, etc., can act as a catalyst for anthropomorphic attitudes, thereby deepening and strengthening the connection that human users build with an AI entity (Tsung-Yu, Yu-Chia, and Wen 2024). Additional emotional cues, such as a perceived sense of humour, further strengthen the feeling of connectedness and a sense of human-likeness in interactions (Xie et al. 2023). Sometimes, details such as increased use of subjective language can have a significant effect on the quality of interactions with a virtual person or chatbot, pointing to the importance of perceived personality and sociability (Pan et al. 2024).

Of course, attribution of human-likeness (and, therefore, the tendency to enter into anthropomorphic parasocial relationships with AI personas) is not a matter of either-or. Instead, as Rapp et al. (2024) demonstrate, there is a broad and complex spectrum of attributed humanness, with actual interactions appearing on a continuum from high humanness to neither human nor machine or diminished humanness to low humanness, depending on both the expectations that users have for a given encounter and the performance of the AI persona in question. With the emergence of virtual companions, this relationship and, therefore, status attribution have become more interactive. The ability to create such fully customisable virtual 'friends' and 'companions' as well as to 'train' them so they respond to the habits, needs, and views of their owners already illustrates a significant degree of prior investment even before the encounter with such virtual entities happens. However, some quite predictable issues immediately become evident. Perhaps most notably, users often display a tendency to select or develop through training traits that are stereotypical, or virtual companions end up as a training ground for toxic offline relationships,

such as young males aiming to create submissive, stereotypically feminine AI 'girlfriends' or even hurling outright abuse at them (see, e.g., Depounti, Saukko, and Natale 2023). Hence, not only issues of the standing of virtual persons but also technological means of preventing virtual toxic behaviour should be considered.

Notably, in experimental settings at least, anthropomorphism-induced perceived connection with virtual entities could induce a certain mindlessness, whereby humans follow the recommendations of AI-powered assistants even when they were clearly bogus and biased (Tsung-Yu, Yu-Chia, and Wen 2024). Likewise, attachment to virtual personas is becoming a problematic issue in terms of psychological addiction, withdrawal, and alienation from inter-human relationships, and other types of addictive behaviours (Pan, Cui, and Mou 2024). In fact, as Mahari and Pataranutaporn (2024) argue, 'addictive intelligence' or 'digital attachment disorder' is what societies need to prepare for, whereby AI-enabled artificial personas would, for example, help overcome both the grief of losing loved ones and the risk and labour involved in forging and maintaining human relationships. While this might seem attractive – and in line with the broader idea of technology making lives easier (frictionless) – it might also lead to the unwillingness (and, increasingly, inability) to process fundamental aspects of offline life and offline relationships. Moreover, with AI acquiring the capacity to untangle content from human creation (and, therefore, removing constraints of scale), technology companies become able to offer 'hedonism as a service' due to the technological capacity to 'identify our desires and serve them up to us whenever and however we wish' (Mahari and Pataranutaporn 2024), for example, surrounding us with virtual personas responding to each of our needs and desires. Not only does no human relationship come close – in fact, it *should not* come close because this implies a one-way relationship in which one side must be satisfied at all costs while the other has no conceivable interests whatsoever. In this way, AI can be seen as promoting and normalising a very toxic model of relationships purely as a consequence of corporate business models.

Simultaneously, virtual personas could play a positive role in terms of fulfilling 'emotional, relational, and therapeutic roles on par with human interlocutors' while also causing 'immediate social satisfaction from the uninterrupted accessibility of social chatbots' (Pan, Cui, and Mou 2024: 6546). On the one hand, this could be seen as a positive development, enabling individuals to access companionship on demand and thereby providing a lifeline at least for some. On the other hand, this generalised availability can also be seen as sowing the seeds for toxic human relationships, once again extending the ready to please – ready to obey model. As ever, the problem lies in the cultivation of attitudes that may subsequently be transferred from human–AI to human–human interactions.

It is also worth stressing that even with simple generative AI tools, one of the most popular uses has so far been sexual role-playing; more broadly, it is becoming evident that despite the early focus on the efficiency-increasing work-transforming use of generative AI, hedonic uses dominate (Heikkilä 2024). Even more so than simple chatbots, virtual personas may be prone to playing out of sexualised roles. A further development to consider is the rise of erobotics, namely, 'virtual, embodied, and/or augmented artificial erotic agents, as well as the technologies and systems from which they emerge' (Dubé and Anctil 2021: 1207). The likely impact of this new and emerging field remains unclear, with predictions ranging from a capacity to 'promote or perpetuate harmful sociosexual norms; generate (new) problematic or pathological behaviours; increase child abuse; impair human relationships; deceive or manipulate humans; […] augment the risks pertaining to privacy and data confidentiality' to the precise opposite of the preceding (Dubé and Anctil 2021: 1205–1206). It is, nevertheless, likely that the positive or negative effects lie less in the technology itself but in pre-existing attitudes of the users, as the latter seem to determine the types and patterns of sex robot use and not vice versa (see, e.g., Leshner and Johnson 2024). Hence, as ever, the key question that emerges pertains to norms and ethical principles not only of behaviour and interaction but also, even more fundamentally, the value of the human person and the intersections, in terms of values, norms, and behaviours, between the human and the digital/virtual levels of interaction.

Overall, at least three instances of optimisation can be foreseeable. The first is personal optimisation, including one's appearance, traits, etc., as well as interactive capacity (both multiplication of avatars and their tailoring to specific interactions). The second is partner optimisation, whereas digital interlocutors can be tailored to the user's data and be maximally predictable and responsive. Finally, environmental optimisation allows placing the user at the centre of their virtual universe. In combination, these instances put into question the status and nature of the human person while simultaneously offering the promise of total frictionlessness and interactional efficiency. This combined promise raises the prospect of technological capacity again passing into a must, obfuscating the possibility of choice – the pathway towards interactional optimisation might simply become too smooth – indeed, frictionless.

References

Allal-Chérif, O., Puertas, R. and Carracedo, P. (2024). Intelligent Influencer Marketing: How AI-Powered Virtual Influencers Outperform Human Influencers. *Technological Forecasting & Social Change*, 200, 1–14.

Balakrishnan, J. and Dwivedi, Y.K. (2024). Conversational Commerce: Entering the Next Stage of AI-Powered Digital Assistants. *Annals of Operations Research*, 333, 653–687.

Barrow, N. (2024). Anthropomorphism and AI Hype. *AI and Ethics*, 4, 707–711.

Bauck, W. (2020, May 1). Could the Pandemic Push Influencer Marketing Deeper into the Realm of CGI Avatars? *Fashionista*, https://fashionista.com/2020/05/influencer-marketing-cgi-avatar-covid-19

Bradley, S. (2020a, March 20). Even Better than the Real Thing? Meet the Virtual Influencers Taking Over Your Feeds. *The Drum*, https://www.thedrum.com/news/2020/03/20/even-better-the-real-thing-meet-the-virtual-influencers-taking-over-your-feeds

Bradley, S. (2020b, January 24). Can Virtual Influencers Build Real Connections with Audiences? *The Drum*, https://www.thedrum.com/news/2020/01/24/can-virtual-influencers-build-real-connections-with-audiences

Byun, K.J. and Ahn, S.J. (2023). A Systematic Review of Virtual Influencers: Similarities and Differences between Human and Virtual Influencers in Interactive Advertising. *Journal of Interactive Advertising*, 23(4), 293–306.

Cook, J. (2020, January 29). Brands Are Building Their Own Virtual Influencers. Are Their Posts Legal? *Huffpost*, https://www.huffpost.com/entry/virtual-instagram-influencers-sponcon_n_5e31cbefc5b6328af2ef97fd

Danaher, J. and Nyholm, S. (2024). Digital Duplicates and the Scarcity Problem: Might AI Make Us Less Scarce and Therefore Less Valuable. *Philosophy & Technology*, 37, 1–20.

Delibasic, S., (2018, August 17). Meet the 'Disturbing' Rise of Virtual Influencers Faking Their Way into Our Instagram Feeds. *The New Daily*, https://thenewdaily.com.au/life/tech/2018/08/17/cgi-virtual-influencers/

Depounti, I., Saukko, P. and Natale, S. (2023). Ideal Technologies, Ideal Women: AI and Gender Imaginaries in Redditors' Discussions on the Replika Bot Girlfriend. *Media, Culture & Society*, 45(4) 720–736.

Dubé, S. and Anctil, D. (2021). Foundations of Erobotics. *International Journal of Social Robotics*, 13, 1205–1233.

Ferraro, C. et al. (2024). Diversity in the Digital Age: How Consumers Respond to Diverse Virtual Influencers. *International Journal of Advertising*, 43(8), 1342–1365.

Forsey, C. (2019, September 18). These Influencers Are Computer-Generated: The Future of Marketing, or Untrustworthy Advertising? *Hubspot*, https://blog.hubspot.com/marketing/virtual-influencers

Franke, C., Groeppel-Klein, A. and Müller, K. (2023). Consumers' Responses to Virtual Influencers as Advertising Endorsers: Novel and Effective or Uncanny and Deceiving?, *Journal of Advertising*, 52(4), 523–539.

Gerrath, M.H.E.E. et al. (2024). Virtual Influencers and Pro-Environmental Causes: The Roles of Message Warmth and Trust in Experts. *Journal of Business Research*, 175, 1–13.

Gräve, J.-F. (2019). What KPIs Are Key? Evaluating Performance Metrics for Social Media Influencers. *Social Media + Society*, doi: 10.1177/2056305119865475

Guerreiro, J. and Loureiro, S.M.C. (2023). I am attracted to my Cool Smart Assistant! Analyzing Attachment-Aversion in AI-Human Relationships. *Journal of Business Research*, 161, 1–13.

Haile, J. (2020, September 18). What Are the Ethical Implications of CGI Influencers Promoting Brands? *Talking Influence*, https://talkinginfluence.com/2020/09/18/ethical-implications-cgi-influencers/

Heikkilä, M. (2024, August 12). Here's How People Are Actually Using AI. *Wired*, https://www.technologyreview.com/2024/08/12/1096202/how-people-actually-using-ai/

Kim, D. and Wang, Z. (2023). The Ethics of Virtuality: Navigating the Complexities of Human-Like Virtual Influencers in the Social Media Marketing Realm. *Frontiers in Communication*, 8, 1–5.

Kim, H. and Park, M. (2024). When Digital Celebrity Talks to You: How Human-Like Virtual Influencers Satisfy Consumer's Experience Through Social Presence on Social Media Endorsements. *Journal of Retailing and Consumer Services*, 76, 1–12.

Kim, J. and Im, I. (2023). Anthropomorphic Response: Understanding Interactions Between Humans and Artificial Intelligence Agents. *Computers in Human Behavior*, 139, 1–19.

Koles, B. et al. (2024). The Authentic Virtual Influencer: Authenticity Manifestations in the Metaverse. *Journal of Business Research*, 170, 1–20.

Leo-Liu, J. and Wu-Ouyang, B. (2024). A 'Soul' Emerges When AI, AR, and Anime Converge: A Case Study on Users of the New Anime-Stylized Hologram Social Robot 'Hupo'. *New Media & Society*, 26(7), 3810–3832.

Leshner, C.E. and Johnson, J.R. (2024). Technically in Love: Individual Differences Relating to Sexual and Platonic Relationships with Robots. *Journal of Social and Personal Relationships*, 41(8), 2345–2365.

Luthera, N. (2020, January 16). The Dark Side of Deepfake Artificial Intelligence and Virtual Influencers. *Forbes*, https://www.forbes.com/sites/forbesbusinesscouncil/2020/01/16/the-dark-side-of-deepfake-artificial-intelligence-and-virtual-influencers/#23bafcf21cd9

Mahari, R. and Pataranutaporn, P. (2024, August 5). We Need to Prepare for 'Addictive Intelligence'. *MIT Technology Review*, https://www.technologyreview.com/2024/08/05/1095600/we-need-to-prepare-for-addictive-intelligence/

McStay, A. (2023). Replika in the Metaverse: The Moral Problem with Empathy in 'It from Bit'. *AI and Ethics*, 3, 1433–1445.

Mirowska, A. and Arsenyan, J. (2023). Sweet Escape: The Role of Empathy in Social Media Engagement with Human Versus Virtual Influencers. *International Journal of Human-Computer Studies*, 174, 1–18.

Mosley, M. (2020, July 14). Virtual Influencers: What They Are and How Do They Work? *Influencer Matchmaker*, https://influencermatchmaker.co.uk/blog/virtual-influencers-what-are-they-how-do-they-work

Na, Y., Kim, Y. and Lee, D. (2024). Investigating the Effect of Self-Congruity on Attitudes Toward Virtual Influencers: Mediating the

Effect of Emotional Attachment. *International Journal of Human-Computer Interaction*, 40(18), 5534–5547.
Ndlovu, M. (2024). Audience Perceptions of AI-Driven News Presenters: A Case of 'Alice' in Zimbabwe. *Media, Culture & Society*, 46(8), 1692–1706.
Pan, S., Cui, J. and Mou, Y. (2024). Desirable or Distasteful? Exploring Uncertainty in Human-Chatbot Relationships. *International Journal of Human-Computer Interaction*, 40(20), 6545–6555.
Pan, W. et al. (2024). Human-AI Communication in Initial Encounters: How AI Agency Affects Trust, Liking, and Chat Quality Evaluation. *New Media & Society*, doi: 10.1177/14614448241259149
Pentina, I., Hancock, T. and Xie, T. (2023). Exploring Relationship Development with Social Chatbots: A Mixed-Method Study of Replika. *Computers in Human Behavior*, 140, 1–15.
Placani, A. (2024). Anthropomorphism in AI: Hype and Fallacy. *AI and Ethics*, 4, 691–698.
Quach, S., Cheah, I. and Thaichon, P. (2024). The Power of Flattery: Enhancing Prosocial Behavior Through Virtual Influencers. *Psychology & Marketing*, 41(7), 1629–1648.
Rapp, A. et al. (2024). How Do People Ascribe Humanness to Chatbots? An Analysis of Real-World Human-Agent Interactions and a Theoretical Model of Humanness. *International Journal of Human-Computer Interaction*, 40(19), 6027–6050.
Robinson, B. (2020). Towards an Ontology and Ethics of Virtual Influencers. *Australasian Journal of Information Systems*, 24, 1–8.
Rodríguez Monclou, C. (2020, September 19). Female Virtual Influencers: A Look at the Future of Entertainment. *The Things*, https://www.thethings.com/female-virtual-influencers-a-look-at-the-future-of-entertainment/
Sands, S. et al. (2022). False Idols: Unpacking the Opportunities and Challenges of Falsity in the Context of Virtual Influencers. *Business Horizons*, 65, 777–788.
Sokolov, M. (2019, December 5). Virtual Influencer Trends: An Overview of the Industry. *The Drum*, https://www.thedrum.com/opinion/2019/12/05/virtual-influencer-trends-overview-the-industry
Spangler, T. (2020, May 6). Miquela, the Uncanny CGI Influencer, Signs with CAA. *Variety*, https://variety.com/2020/digital/news/miquela-virtual-influencer-signs-caa-1234599368/
Sweeney, P. (2024). Avatars and the Value of Human Uniqueness. *Philosophy & Technology*, 37, 1–4.
Torossian, R. (2020, March 25). The Risks – and Benefits – of Working with Virtual Influencers. *Agility PR Solutions*, https://www.agilitypr.com/pr-news/public-relations/the-risks-and-benefits-of-working-with-virtual-influencers/
Tsung-Yu, H., Yu-Chia, T. and Wen, Y.C. (2024). Is this AI Sexist? The Effects of a Biased AI's Anthropomorphic Appearance and Explainability on Users' Bias Perceptions and Trust. *International Journal of Information Management*, 76, 1–22.

Turner-Williams, J. (2020, May 28). Are Virtual Influencers More Valuable to Brands than Humans? *Super Maker*, https://supermaker.com/articles/virtual-influencers-changing-social-media

Vogue Business (2020, June 16). Brands Eye China's Virtual Idols, https://www.voguebusiness.com/technology/china-virtual-influencers

Xie, Y. et al. (2023). How Does Anthropomorphism Improve Human-AI Interaction Satisfaction: A Dual-Path Model. *Computers in Human Behavior*, 148, 1–18.

Xu, K. and Shi, J. (2024). Visioning a Two-Level Human-Machine Communication Framework: Initiating Conversations Between Explainable AI and Communication. *Communication Theory*, 34(4), 216–229.

Yates, K. (2020, May 29). The Lucrative Rise of the Virtual Influencer. *Vice*, https://www.vice.com/en_uk/article/xg8xdd/liam-nikuro-imma-lil-miquela-virtual-influencers

Yu, J. et al. (2024). Artificial Intelligence-Generated Virtual Influencer: Examining the Effects of Emotional Display on User Engagement. *Journal of Retailing and Consumer Services*, 76, 1–10.

Zhou, Q. et al. (2024). Mere Copycat? The Effects of Human Versus Human-Like Virtual Influencers on Brand Endorsement Effectiveness: A Moderated Serial-Mediation Model. *Journal of Retailing and Consumer Services*, 76, 1–15.

Zhu, S. et al. (2024). Virtual Agents in Immersive Virtual Reality Environments: Impact of Humanoid Avatars and Output Modalities on Shopping Experience. *International Journal of Human-Computer Interaction*, 40(19), 5771–5793.

6 From Post-Work to Transhumanist Dreams

AI for Everything

Is Work Even Necessary?

Visions of the future development of AI tend to predict significant levels of job automation and ensuing transformations of society. For authors working on approaches to what could be called post-work, this future is desirable, particularly if societies untether themselves from work-related income as the main source of individual sustenance and are able to offer people a meaningful and prosperous life on the back of technological advancements. For some, an unstoppable new wave of advancement would 'usher in a new dawn for humanity, creating wealth and surplus unlike anything ever seen' (Suleyman 2023: 7). However, such approaches also raise fundamental value questions, not least pertaining to what it means to live a life of quality and dignity, who determines the criteria for such a life, and in what ways.

The promise of post-work visions of the future is the *democratisation* of leisure: While previously leisure had been a privilege of those who could afford to live off other people's labour, automation promises to open up this luxury to all (Jones 2023: 380). More precisely, this means that people, at least in sufficiently developed societies (and the latter is a major caveat!), would be able to live off labour outsourced to AI and robots. Framed in this way, the change might be less revolutionary than it appears at first sight – the only change is in terms of *who* works. Still, the nature and content of the visions of a post-work future differ significantly, showing the importance of ideological underpinnings. Depending on one's (positive or negative) view of work, the policies expected in the wake of automation could range from more and better training to ensure that humans could keep up in their race against the machines (if work is treated as a positive source of both income and meaning) to an embrace of the absence of work and an allegedly ensuing society of leisure. In fact, for the optimists, automation must be accelerated so that a post-work future comes as soon as possible to ensure a novel society of leisure and abundance (see, e.g., Srnicek and Williams 2015; Bastani 2019; Danaher 2019).

DOI: 10.4324/9781003715849-6

On the other hand, such pressures to accelerate automation are criticised for being merely extensions of the current climate-wrecking and exploitative policies: Instead of profound change and liberation, they are seen to offer more of the same, just in a different packaging (see, e.g., Mariqueo-Russell and Read 2019): While humans would be ostensibly freed from work (although, arguably, not from unpaid labour as argued below), practices pertaining to environmental exploitation, resource extraction, and the overall utility-maximising logic of current economic organisation would continue. Likewise, for others, the very idea of human surrogation with machines only makes sense from a bounded perspective of rationalisation and ever more efficient value extraction that is neither universal nor unavoidable but, instead, determined by specific modes of economic organisation dominant in Western societies and being implanted elsewhere through globalisation (Atanasoski and Vora 2019).

On the other hand, the matter of work does not necessarily have to be an either-or situation: The fact that work will be largely removed from the equation as a means of earning a living and structuring societies does not mean that work must go away – some people may choose to work nevertheless, perhaps in a different – artisan – way (Rawlings 2018). That might involve anything from programmers choosing to write independent 'handmade' code for a small circle of connoisseurs in a world of largely self-maintaining, self-enhancing, and independently learning software to others moving into rural areas to grow their own food and acquire other means of self-sustenance. For others, certain forms of work are unlikely to disappear, since they are 'socially necessary', particularly those pertaining to social care, at least some aspects of education, and at least some forms of the running and administration of the society (unless a decision is made to delegate that to AI as well). For that reason, the 'equitable distribution' of such work through, for example, Universal Civil Service, would become a noteworthy concern (Gómez-Baggethun 2023). A similar argument is provided by Knell and Rüther (2024), for whom the performance of such tasks would be part of a wider reorientation of societies towards 'the good'. While such work would be scarce and would not require significant investment of time and effort when distributed across the society at large, it would, according to this view, provide some structure and meaning without simultaneously obstructing a society of leisure.

Jones (2023), meanwhile, holds a less optimistic view by claiming that certain activities will not be automated not because they are socially important or necessitate a human touch, but because they are unpaid anyway and, therefore, there is no financial gain in automating them; hence, activities that are currently taken to be 'non-work' (house care, childrearing, etc.) would remain domains of human activity, resulting in inequalities (e.g., gender-based) in access to leisure. Still others, by contrast, would view such visions as not going far enough and would discount the

potential extent of automation. It is claimed that a true post-work society would be one in which 'technology has displaced the majority of both paid *and* unpaid work' – in other words, 'a society that no longer requires most people to engage in *any* activities that are useful or necessary to others' (Althorpe and Finneron-Burns 2024: 332). The latter view presupposes that innovation and automation would trickle down throughout society (as, e.g., home appliances have proliferated despite tackling mostly unpaid forms of labour) as long as there are conveniences to be achieved and savings (including in time and effort) to be made, combined with a capacity to pay for such innovations.

Of course, the preceding still leaves some questions as to the replacement of the personal and social value of work open, particularly in terms of 'income, self-development and excellence, community, meaningfulness, and social contribution' (Althorpe and Finneron-Burns 2024: 333). And yet, proponents of full post-work would respond that the association of these benefits with work is merely a contingent one, arising from the central role that work currently plays in people's lives, and thus similar or analogous benefits could be obtained in other ways. For example, sustenance could be guaranteed through Universal Basic Income (UBI), meaning that self-development could be achieved perhaps even more easily due to the ample free time rendered available by automation; likewise, a sense of community could be achieved with fellow hobbyists, in religious associations, or similar groups while meaningful life could be lived through engagement in learning, philosophical reflection, or political activities; as for social contribution, while it might be more difficult to achieve without work, the argument goes that there is no guarantee that in a heavily roboticised society it would continue being a valuable quality in the first place (Althorpe and Finneron-Burns 2024: 355–358).

However, the notion of leisure must be problematised as well: While leisure has typically been framed as the opposite of work, it was never *disentangled* from work – instead, the leisure typically treated as worthy involved either rejuvenation for more productive work, acquisition of new productive knowledge, or the development of productive skills (Chia 2021: 51). For this reason, leisure would have to be reconsidered and reconceptualised to transform it into a standalone phenomenon. Simultaneously, though, another option is also available: It is often claimed that at least some remnants of work are necessary to live a good life (both personally and societally). For example, while Spencer (2025: 1240) admits that there is a deeply engrained view that 'workers trade-off the marginal disutility of work against the marginal utility of consumption', which implies that work is a merely instrumental activity to earn income and, therefore, can be replaced for the individual's own benefit, as opposed to something that might positively shape a person's life, he mainly frames the *quality* of work as the problem. The amount of work and the low quality of much of it are

claimed to be the real issues to be overcome; following this argument, the endpoint of automation should not be the replacement of humans but giving them the ability to work less and, simultaneously, more meaningfully, thereby leading towards greater human flourishing (Spencer 2025: 1244–1245). Hence, freeing people from work as a *necessity* does not automatically mean that nobody would work – in fact, some people may choose to work voluntarily if they *want* to. This approach would also have the benefit of retaining the classic distinction between work and leisure, thereby preserving the value of the latter.

It is also correct that not only does today's life revolve around work, but work itself is idolised in a way that makes one think that the more they work, the better; moreover, digital technology assists that greatly, erasing the boundaries between work and leisure in the first place (Beckett 2018). Unsurprisingly, the opponents of the work-positive view are quick to brand the latter as conservative, clinging to moralistic arguments and ideological assumptions that mask the exploitation and neglect of individual value (see, e.g., Mercer 2021: 687). The real problem to overcome, then, would be not the absence or reduction of work but the societal valorisation of work. In other words, the post-work project would then be not only (or perhaps even less) about transforming the economic organisation of societies but more about resetting the entire ideological structure that frames the individual (Mercer 2021).

In addition to the quality-of-life arguments, the physical effects of both work and its absence are far from unequivocal: From the physical effects of work (including allegedly non-physical work, such as the muscle and eye strain associated with office work) and mental fatigue to the adverse health effects associated with joblessness. It is also notable that automation goes beyond merely deepening the role of technology in specific sectors – it also spreads horizontally, from routine and relatively clearly definable tasks (e.g., in factory settings or entry-level professional jobs) to performance of more complex and even creative tasks as well as those requiring at least simulated empathy (such as elderly care), thereby making the prospect of jobless life a real possibility for many (Spencer 2023: 3). Of course, the above hinges upon the question of how transformative the coming wave of automation and, in particular, AI which lies in its centre, is going to be: We have learned from previous technological innovations, at least all the way since the Industrial Revolution, that the economy, and the job market alongside it, operates in waves of Schumpeterian 'creative destruction': While business models (and jobs) that had been previously taken for granted are destroyed, new ones are created in their place.

However, certain pressures and experiences may provide an extra impetus for automation. For example, the vulnerability of the human workforce has been underscored by the coronavirus pandemic, making it simply safer

to automate at least the roles that cannot be performed from home – just to be on the safe side, should a similar event occur in the future (Thomas 2020). Such contingency planning might be even more pressing as the emergence of new diseases and increased prevalence of old ones appears to be exacerbated by climate change in a chain of effects ranging from water contamination to altered vectors of dissemination (Ogden and Gachon 2019; Vaughan 2019; Gully 2020). Additional health threat factors include overall pollution and its effects on the immune system and population health, as well as often overlooked threats, such as long-dormant pathogens being released into the environment by melting permafrost and glaciers (Al Husseini 2020). This broad set of factors raises the potential problem of workforce collapse, further pushing businesses towards automation, potentially also breaking the cycle of creative destruction, that is, automating existing jobs without the creation of new ones.

Of course, work, particularly in discussions over its future, is usually conceptualised as paid labour. However, there are also forms of unpaid labour that must be considered. While some of that involves, for example, domestic labour, various forms of unpaid intellectual labour persist as well. Some of them are usually not considered labour at all, yet they are activities that generate economically lucrative output (although the benefit is reaped not by the individuals performing such labour, which is why it is considered unpaid). Examples may include something as mundane as creating and maintaining social media profiles, interacting with content on social media and streaming platforms, etc., and even using Internet of Things devices or smart city infrastructure; these are data- and, therefore, value-generating practices that may seem like leisure but are, nevertheless, deeply rooted in economic relations (Mathers 2020). Such activities may be more difficult to automate because the data generated covers the lives and preferences of real humans. Thinking about a post-work future and remuneration opportunities therein, and digital life as data-generating by default, there are likely to be stronger calls for personal data ownership, so that individuals would still have something to sell once they can no longer sell their work. However, such demands would likely usher in more extensive use of synthetic data, reducing human leverage (and the need for human work) here as well.

A post-work condition might also strengthen the appeal of artificial environments. As Mathers (2020: 10) stresses, '[b]y getting in ahead of demands as and when they manifest, capitalism can begin to direct consumption: building, a priori, the "worlds" through which the consumer encounters their own needs and desires'. Here, technologies including Extended Reality (XR) or the previously hyped visions for the metaverse come to the fore: As leisure takes up an ever-increasing proportion of everyday life, its structuration would become not only social (through hobbies, interpersonal interactions, etc.) but also technological

(e.g., digital environments and entire worlds that shape everyday activities). As immersion into such worlds would be even more data-intensive than the social experiences on the current internet, discussions as to the collection, use, and remuneration for data will become even more intense. After all, data would play a highly recursive role in the sense that its generation would become *the* labour in a post-work society, while that same data (or, rather, decisions based on data analysis) would simultaneously play a key role in determining the conditions of its own generation. Meanwhile, the models for monetisation or other forms of remuneration may range from direct financial rewards to simply free access to digital resources, similar to the current models of ostensibly free platforms and software.

Moreover, should the best experiences possible be available online through virtual and otherwise digitally structured worlds, there will likely simply be less motivation to go outside. While those who are able and willing will be indoors, enjoying their technologically mediated life, able to afford the basics such as connectivity and the necessary hardware due to their data-generating daily lives (creating a vicious circle: Those already privileged and living a digital-first life would be able to retain their privileged whereas others, apart from centralised interventions, such as UBI, would be perpetually sidelined), the outdoors will become the domain of the economically ostracised and seen as a dangerous place. That, in combination, will also lead to strong network effects: If all your friends and family are in virtual worlds, why go anywhere else (see, generally, Kalpokas and Kalpokiene 2023)?

One can already witness a playful turn throughout today's media or, perhaps, society as a whole through extensive gamification, real and simulated audience participation and contribution, etc. (Mäyrä 2017; Hjorth 2018; Richardson and Hjorth 2019), ideally conferring onto such activities some (perceived or actual) user value (Hassan 2017). Gameplay might be the first step along the way, characterised by the tension between 'play as being-here and game as being-there', thereby establishing 'a dynamic between the actual experience of playing here-and-now and the desired future not yet realized state' (Larsen and Walther 2020: 611, 622). Particularly if the post-work society materialises in conflation with the spread of virtual experiences, it can be argued that the 'here' aspect is abolished: The presence of action and the scenario in which it takes place are both 'there' due to the immersive capacity of XR and other technologies of gamification and virtualisation. This leisure-necessitating shift in itself might be a significant driver towards a post-work society as a virtual-first society.

Virtual gaming experiences are part of a broader question as to what people are going to do all day and where meaning and structure are going to come from. Since the understanding of leisure (at least in the West) has typically been as the antithesis of work, if the latter is gone or

transformed, the notion of leisure will have to change accordingly, perhaps towards something more focused on actual human flourishing and enjoyment (Snape et al. 2017; Jones 2023). There may, for example, be a foreseeable situation in which we dedicate more time to social pursuits, caring for each other and finding meaning through interaction; we may also dedicate more time to learning and thinking, community engagement, or the arts and other cultural pursuits (Beckett 2018; Jenkins 2019) – something Etzioni (2017) optimistically calls 'true flourishing' while Knell and Rüther (2024: 369–370) see this as part and parcel of societal reorientation 'towards the true and the beautiful'. Nevertheless, such visions are not without their own problems either: Not only they are elitist, extending the previously privileged activities across the society as a new normative standard but also, by that same token, they raise questions of power, value, and interest (i.e. who is going to fill vague terms, such as 'true' and 'beautiful' with meaning). In other words, who is to be the norm-setter, deciding how life in a post-work society would be organised, what lifestyles will be prioritised through such organisation, and what enforcement of acceptability standards (or, perhaps, less evident nudges) would be present.

Moreover, more attention must be paid to AI's capacity to automate an extremely broad range of activities, including those that could pass as leisure. Certain activities, once mastered by AI, could become out of bounds for humans; an example here could be driving: If autonomous vehicles become significantly safer than human-driven ones, humans may be barred from sitting behind the wheel, even recreationally (Jones 2023). Likewise, for the high-brow activities mentioned above, advances in generative AI may decrease the willingness of individuals to engage in writing or painting not just professionally but also in an amateur capacity; the same applies to other activities (e.g., from generating a recipe to actual food preparation) or engagement in learning: For example, if any question could easily lead to a reliable and instant AI-generated response, the motivation to learn – such as for participation in intellectual debates, imagined by some post-work thinkers – would likely decline (Jones 2023). Admittedly, for example, for Knell and Rüther (2024), active engagement in high-brow creative and intellectual activities may not be necessary for benefitting from a post-work society: Enjoyment of the *results* of such activities, regardless of who or what performs them, might be sufficient, since that already counts as passing time. However, the preceding is open to the criticism of merely supporting passive-consumptive behaviours. Still others (see, notably, Kurzweil 2024) expect that by merging with AI, humans would be able to overcome any challenges and create new forms of meaning unimaginable today.

Moreover, the hedonistic emphasis present in many theories of post-work (AI automating not only jobs but also other boring and repetitive

activities) may foreclose access to a significant number of pursuits: For example, even mastering a game might include lengthy repetitive practice, the capacity to engage in which might be lost (Jones 2023). Instead, people may resort to spectatorship (similar to passive consumption above) or to reliance on AI assistants that learn (and potentially play) games for us. As noted above, the same applies to more high-brow activities, such as engagement in creativity, philosophy, performance (of, e.g., music), etc. Hence, automation itself might become a factor that limits not only work but also, paradoxically, leisure if by the latter we mean deep engagement in something for the sake of it. It is, consequently, not unlikely that the default pastime would consist of observing AI performances (playing games, performing film-like experiences, playing music, composed in real time in accordance with our preferences and current mood, engaging in staged debates, etc.) and interacting with virtual influencers and personalised 'companion' avatars. This spectatorship-first lifestyle would likely extend to interpersonal interactions: Safe and predictable interactions with AI-powered avatars in virtual settings would become a realistic and attractive option. A notable question, though, is whether such passive hedonism would be enough to fill in all-encompassing leisure. And if it *is* enough, again a question of values looms large, that is, how such spectatorship experiences would be structured and by whom.

Potential issues do not end with the above, however. Regardless of whether one settles for more hedonistic activities or the relatively high-brow pursuits, the question of structure and meaning will likely remain unanswered. A key issue is definitional: What attributes will be considered as quintessentially human and, therefore, essential to be preserved? For example, one of such issues would relate to the nature of humans as political and social animals. Perhaps most obviously, this would involve decisions as to whether humans should continue governing themselves and stay involved in matters of administration and civil service, or whether such activities should also be included in the overall automation thrust and delegated to AI. That is part of a broader question of whether there are any quintessentially human activities that *should not* be automated, such as childcare (and child rearing in general), although that could arguably be a matter of individual convictions (some deciding to raise their children by themselves and others outsourcing this task to robots). Moving further, the quintessentially philosophical issues of the meaning of life and the nature of a good life will arise, juxtaposing the more active and the more passive-hedonist ways of life, and pertaining to, among other things, the design and architecture of digital infrastructures that will sustain leisure (or life in general). Related to that, matters of (digital) agency and autonomy would loom large: If the projected post-work life were to be lived primarily through digital environments, the generation and use of data on which the simulation of XR experiences rests would

acquire even larger importance than the economic aspects discussed above. Once data-derived virtual experiences become all-encompassing, always offering tailored content that the user cannot say no to (either because of a perfect fit or because of the deterministic nature of digital architectures, or both), opinion, choice, and other agency-defining qualities would become largely irrelevant.

The answer as to which direction is to be taken at the crossroads of AI and automation of work is less contingent on the technology itself than on factors external to it: First and foremost, the definition of good life embraced by societies through their representatives (Susskind 2023) or, later, perhaps as decided by an AI government. What might have in the past appeared to be an abstract philosophical question (what is the meaning of life and what does a good life entail) becomes a practical policy problem. Particularly in cases when post-work life is lived through digital platforms and virtual environments, the question of values embedded in such technologies, that is, what leisure models are to be prioritised, what counts as leisure in the first place, and who is entitled to (or can afford) it ought to become major points of debate before becoming inscribed into supposedly neutral code – an act that typically removes issues from contestation and places them within the technological 'black box'. Likewise, the matter of beneficiaries cannot be left out of the equation, that is, whose interests are best served by the specific – and subsequently technologically decontested – choices.

The frictionlessness of the transition from increased work automation to increased leisure automation becomes a major factor. The driving force behind the likely passage from the possibility of all-encompassing automation to its perceived necessity would be a promise of life that is optimised for unobtrusively and effortlessly consuming the results of technological advancement in a way that is detached from (and blind to) broader concerns. Crucially, not only matters of the meaning of (good) life are likely to be determined simply by technological affordances, but also issues of human nature and social interactions, as well as economic and digital divides (who has access to the 'post' of post-work) would potentially be occluded, showcasing a seemingly straight and unavoidable technological path instead of crossroads.

Transhumanism and the Search for New Human Relevance

In a way, the entire discourse on post-work can be seen as a set of strategies to deal with a predicted future of human irrelevance. Of course, it is framed in more palatable terms – as the emergence of a society of leisure – but this leisure is merely a result of the fact that humans would no longer be able to meaningfully contribute to societal progress and would, therefore, simply have to find ways to pass the time. However, alternatives to this view are

available, postulating that if competing with machines becomes futile, merging with them might provide a solution. Such technological enhancement would, according to proponents, keep humans relevant – in the sense of both keeping them productive and enabling more meaningful leisure than mere hedonistic pursuits. This is the domain of transhumanism.

Transhumanism can be defined as 'first and foremost, a promise to bestow human existence with meaning through recentralizing culture on the basis of techno-centered metanarrative' (Lipowicz 2023: 537). In other words, while there are, as seen throughout this chapter, significant challenges to traditional understandings of the human and the very meaning of humanness, transhumanism claims to take the latter in its stride and provide a highly optimistic narrative in which all of humanity's problems are seen as solvable through technological innovation and, ultimately, a merger with technology. Hence, the aim of transhumanism is 'to transform Homo sapiens into a superior successor […] through radical interventions that can overcome fundamental biological limitations' (Ewuoso and Fayemi 2021: 634). Hence, while the 'original' human body (and the mind it can support) is seen as inherently weak and inadequate, technological improvements and extensions are expected to infinitely stretch human capacities. According to transhumanists, human nature can (and, therefore, should) be improved through technology, 'eliminating such undesirable aspects of existence as disease, aging, and even death' (Chursinova and Sinelnikova 2024: 27). Technology becomes an enabler of 'the production of the human being by itself' (Lindberg 2023: 8). In other words, the human (in an idealised form) is simultaneously the goal to be achieved, the product to be created and the creator of the latter in an endless recursive loop.

Thomas (2024), meanwhile, identifies the essence of transhumanism as a striving for super-longevity, super-intelligence, and super-wellbeing, with the expectation of technology becoming the solution to any problems that humanity currently faces or may face in the future. Notably, transhumanists tend to see themselves as rightful heirs to the Enlightenment, 'grounded in rationalism, physicalism, and ardent faith in the scientific method, as well as a deep belief that humans can and should shape their destiny' (Grant 2023: 38). This can be achieved either by augmenting existing human bodies or (as in, e.g., Kurzweil 2024) shifting from biological to digital existence, that is, merging with AI. Others, however, have questioned whether this is merely a hubristic radicalisation of the Enlightenment promise – a sort of hypertrophised 21st-century optimism (Corjescu 2022: 25).

Crucially, transhumanists tend to deny the existence of something akin to human nature or any essentially human characteristics. For them, 'the only truly universal human essence is a drive to self-transcendence, equated with increasing freedom or autonomy' by way of overcoming all

limitations (Porter 2023: 477), biological or otherwise. Simultaneously, though, transhumanism can also be seen to echo certain quasi-religious tendencies, including promises of higher meaning in life and, ultimately, salvation, which are rendered into calculable and technologically solvable problems (Leidenhag 2020; see also Cohen and Spector 2020; Lindberg 2023). Elsewhere, such contextualisation builds on a *meta*religious desire for hope and fulfilment – one that is always on the horizon but never truly met (Corjescu 2022: 24–25). In the case of transhumanism, such fulfilment is constantly one major technological breakthrough away. In general, however, this techno-salvation always, in some way, pertains to overcoming biological nature and the creation of a new kind of entity: No longer 'animals with biological brains' but, instead, entities 'whose thoughts and identities are no longer shackled to what genetics provides', reinventing ourselves 'on a more powerful digital substrate' and becoming '[f]reed from the enclosure of our skulls' (Kurzweil 2024: 11, 73) – truly going beyond human.

More specific goals of posthumanism could include 'ending aging; increasing physical, cognitive, and emotional capacities thousands of times over; and potentially trading biological platforms of existence (our bodies) for nonbiological digitized platforms of virtual existence ("uploading" consciousness)', although the list could be expanded by including any conceivable enhancement (Grant 2023: 37). Likewise, on the agenda is 'radical reduction of undesirable traits/states', such as 'depression and introversion' (Grant 2023: 38–39). Likewise, Kurzweil (2024) paints a picture in which any diseases and forms of neurodivergence are removed from human existence. Of course, a key question to be asked here is who decides which traits/states are desirable and which are not (e.g., introversion), and according to what (and whose) values; no less importantly, the very idea that all humans should be remodelled under a single standard of 'normality' is deeply troubling.

As Thomas (2024: 215) also observes, 'depending on how the human is conceived, the question of what constitutes enhancement will change' (for a similar argument, see also Lindberg 2023: 265). If the changes to the human body and mind are such that they make the human species unrecognisable as humans in the traditional sense, that, for transhumanists, can only be celebrated, as one would celebrate the overcoming of an impediment or an illness (Grant 2023; from a critical perspective, see Lindberg 2023). Likewise, via radical alterations, humans may become less human in the biological sense but more human in the moral sense (Ewuoso and Fayemi 2021). In other words, transhumanism makes any definition of the human and any ends of human existence arbitrary, merely an instance within an ever-shifting horizon of technical capacities and imagination.

Consequently, *the* question to be asked is whether we are inclined to retain any stable points of reference as to what constitutes the human and

how such decisions are to be made. Meanwhile, if such decisions are left to the market based on supply and demand, the risk is that multiple layers of humanity would be created, depending on the enhancements afforded. And should mandatory limitations to enhancement be imposed to preserve either some version of 'humanness' or equality, the ability (and willingness) of states and other actors (including technology companies) to develop truly global rules and stick to them is questionable.

A major issue is the improvement of human mental capacity: The expectation that once merged with AI (or uploaded onto digital computing platforms), 'our minds will be empowered to grow exponentially, ultimately expanding our intelligence millions-fold' (Kurzweil 2024: 73). In this way, the only way for humans to keep up with AI becomes the ability to 'create meaningful expressions that we cannot imagine or understand today' (Kurzweil 2024: 229). Consequently, the whole question of human–AI interaction and future challenges and risks posed by AI is removed in one stroke: Any distinction between the two simply ceases to exist. Thus, transhumanists embrace the idea of improving (under their own definition of progress) the human condition to the extent that humans cease being human in the conventional sense, seen as a positive development due to the in-built limitations of the biological body and mind (Ewuoso and Fayemi 2021): If human biology precludes humans from reaching their full potential, it is something that needs to be overcome – as any other obstacle would (Grant 2023). Therefore, transhumanists conceive of themselves as taking evolution into their own hands for the sake of accelerating progress and ultimately shaping human destiny independent of any external circumstances (Sorgner 2022: 35; see also Ewuoso and Fayemi 2021: 634; Grant 2023). As the ultimate step in this direction, transhumanists tend to embrace the vision of not only living with, but ultimately merging with machines (Coeckelbergh 2020: 42).

For some, transhumanism appears as an escapist ideology. For example, as Lipowicz (2023: 542) notes, 'transhumanists do not value the satisfaction deriving from being able to overcome life's difficulties, but rather embrace a vision of happiness based on eliminating the adversity that is intrinsic to being alive in the first place'. In other words, instead of human growth (both personal and collective) through overcoming challenges and finding one's place on Earth, here one has an ideology that proposes technologically enabled shortcuts and ever more pervasive domination of the Earth (Mason 2022). Similarly, the continuation of the Enlightenment tradition in transhumanism is questioned: While the Enlightenment advocated rationalisation through the overcoming of the body by the mind, in transhumanism, there is nothing to overcome as the body is enhanced or removed altogether (Brusseau 2023). In this way, transhumanism can be understood as a form of cheating.

Others criticise transhumanism for the lack of critical reflection on its premises by reframing it as a means of reinventing hegemonic white

masculinity (along with historical racialised definitions of what constitutes 'the human') in a new technology-infused language (see, e.g., Ali 2017). Similarly, the anthropocentrism of transhumanism can be highlighted: Not only does it put the interests and aims of the human at the heart of everything and before any other considerations, but also 'it uncritically embraces the notion of a universalized human and thus carries over the discriminatory aspects of humanism', such as exclusive focus on rationalism (Thomas 2024: 214). In this way, transhumanism 'creates a hierarchical notion of the human that expels or "others" certain humans while further entitling privileged groups' (Thomas 2024: 214). As Lindberg (2023: 265) observes, the focus on enhancement has, as its corollary, the morally untenable assumption that sees 'less healthy and less intelligent life as less valuable'. Paradoxically, despite its seemingly progressivist rhetoric, transhumanism can be seen as fundamentally conservative, focused on 'projecting today's ideal of humanity into the future' to the effect that 'it extends the rule of these same ideals instead of interrogating them' (Lindberg 2023: 265). Hence, a full reinvention of the human is on the agenda only to the extent that it always happens in line with standardised and normative definitions of 'peak' humanness, dominant at a given time in the technology industry, but without any popular oversight. In fact, even individual decision-making would be constrained: If a critical mass of individuals embraced the enhancement on offer, others would have to follow suit regardless of their personal preferences simply to avoid becoming second-class humans.

Transhumanism can also be criticised for being centred around an instrumentalist account of the human, technology, and progress, whereby improvement and expansion serve no other goal but that of ever deeper and ever more efficient extraction, following exploitative modes of economic organisation (Thomas 2024). In other words, the idea of continuous augmentation only makes sense if one uncritically adopts the idea of maximising performance and value generation (Atanasoski and Vora 2029). As Lindberg (2023: 265) scathingly asks, 'Who would profit from the reduction of sleeping time, the person capable of working more or their employer?'. Effectively, 'the world inclusive of human beings becomes standing reserve: objects ready for exploitation' (Thomas 2024: 217). Thereby, transhumanism becomes a way to bring about a vision built on 'a harsh productivism and a hostility to incarnated existence' (Lindberg 2023: 9). Judging by the definition, this seems to be a radicalisation of not necessarily the Enlightenment but certainly of the Weberian protestant ethic. Transhumanists can also be seen to adopt a reductivist (and, again, instrumentalised) view of the world, 'conceiving of reality as a code which can be read and edited', thereby 'turning rich intra-related complexity into instrumentalizable data points', which can only lead to 'an erasure of meaning' while also enabling 'a further reconstruction of the world as more amenable to prediction and control' (Thomas 2024: 215, 217).

Yet again, exploitative views of the world come to the fore, albeit masquerading as improvement.

An additional matter to be considered is the logical impossibility of a rational choice to be augmented (even if we ignore the above observation of the choice environment being rigged in the first place): While we can now speculate about a radically augmented future, the actual experience (and even existence) of such an augmented human being would be so profoundly different that we cannot realistically imagine and, therefore, choose it; without the possibility of such a rational informed choice, decision to be augmented (let alone external pressure to be augmented) would be incompatible with human freedom (Lyreskog and McKeown 2020). Moreover, the decision to adopt enhancement techniques would also have profound effects on those unable to have a say, such as future generations (Lindberg 2023: 265). However, choice does not really play a notable role in transhumanist discourse, with human enhancement presented as 'not only uncontested but also incontestable' (Thomas 2024: 215). In fact, seen from within the transhumanist camp, the merits of augmentation are such that any delay becomes morally wrong (see, e.g., Ewuoso and Fayemi 2021).

A somewhat related issue centres on the loss of the 'I' as a subject and dissolution of human subjectivity in general: While transhumanists envisage a continuous human subject throughout processes of enhancement and, ultimately, uploading the human mind to machines, their critics assert that the human self and the 'I' as a person is too tied to the bodily characteristics and experiences (as well as cognitive capacities) of being biologically human to survive augmentation (Brusseau 2023). Indeed, it can be asserted that 'the human in transhumanist discourse is a tractable information processing machine, separable from and superior to the rest of the nature than surrounds it' (Thomas 2024: 215). As with any machine, the replacement of its bits and bobs with better-performing components (or updating software) would keep the machine functionally the same, but simultaneously instrumentally 'better'. However, there is no guarantee that human subjectivity operates in the same way – instead, it may well be formed through the continuous experience of overcoming the challenges of being human. With the elimination of such challenges, continuity of subjectivity cannot be assumed. Once again, transhumanists would not necessarily see that as a shortcoming – on the contrary, they would automatically assume such altered subjectivity to be 'better' (as it would ensue from an allegedly 'better' state of existence). However, the core questions running throughout this section – who and how would define 'the human' and any points of reference pertaining to it, and whether we ought to have such definitions in the first place – remain, necessitating a decision at this major crossroads.

As already hinted above, the inclusivity of the transhumanist project is being questioned. As Chursinova and Sinelnikova (2024: 26) observe, 'only those with money, power and connections will be able to get access

to the technologies, making the rich and poor potentially distinct subspecies'. Notably, transhumanist authors are usually quiet or elusive on the matter of availability and accessibility of enhancements, particularly in terms of the gaping inequalities both nationally and globally. It transpires that enhancement is understood in terms of humanity as such but not necessarily of every individual; in other words, value and moral concern seem to lie in humanity as a whole, while any particular person is as exchangeable as any of their parts (see, e.g., Thomas 2024: 218). Hence, should a decision be made to proceed with augmentation (and there is a possibility that such a decision would become a race to the lowest common denominator – if some states would begin encouraging enhancement, others would have no option but to follow so that they do not become hopelessly backwardly), funding for augmentation programs would have to be provided in order to keep everyone on an equal playing field.

Arguably, decisions as to what can be augmented and to what extent (or to leave augmentation unregulated or perhaps even encouraged) would likely be the most important (or maybe even the final) crossroads of humanity – after all, it will determine the future shape of humanity as such. Consequently, any decisions thus made must be carefully weighed, even if short-term benefits may provide an impetus for rushing in (thus, again, inducing a race to the lowest common denominator). Transhumanists, of course, have their own response to any criticism, often labelling their opponents as 'bio-conservatives' or 'speciesists' (see, e.g., Ewuoso and Fayemi 2021) – terms invoking biases and even bigotry. It is important not to give in to such marginalisation of alternatives, instead leaving the field of possibilities open to as broad a set of choices as possible and choosing a path forward as carefully as possible.

Clearly, transhumanism embodies a radicalisation of possibility qua imperative. Not only are alternatives or even attempts to point towards a choice (crossroads) discarded and even invalidated, focusing on the seemingly frictionless path towards human enhancement, but also the implementation of technological capacities to their fullest foreseen extent is seen as a moral duty to unlock seemingly unprecedented potential. A problematisation – opening a crossroads – here would mean reframing the human as an open-ended question, thereby foregoing a firm and final answer and, instead, (re-)embracing diversity.

References

Al Husseini, I. (2020, April 5). Climate Change Is Only Going to Make Health Crises Like Coronavirus More Frequent and Worse. *Business Insider*, https://www.businessinsider.com/climate-change-making-health-crises-like-coronavirus-frequent-worse-2020-4

Ali, S.M. (2017). Transhumanism and/as Whiteness. *Proceedings*, 1, 1–3.

Althorpe, C. and Finneron-Burns, E. (2024). Productive Justice in the 'Post-Work Future'. *Journal of Applied Philosophy*, 41(2), 330–349.

Atanasoski, N. and Vora, K. (2019). *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Durham, NC: Duke University Press.

Bastani, A. (2019). *Fully Automated Luxury Communism: A Manifesto*. London: Verso.

Beckett, A. (2018, January 19). Post-Work: The Radical Idea of a World without Jobs. *The Guardian*, https://www.theguardian.com/news/2018/jan/19/post-work-the-radical-idea-of-a-world-without-jobs

Brusseau, J. (2023). Mapping AI Avant-Gardes in Time: Posthumanism, Transhumanism, Genhumanism. *Discover Artificial Intelligence*, 3, 1–11.

Chia, A. (2021). Self-Making and Game-Making in the Future of Work. In O. Sotamaa and J. Švelch (eds.) *Game Production Studies*. Amsterdam: Amsterdam University Press, pp. 47–64.

Chursinova, O. and Sinelnikova, M. (2024). Post-Human and Trans-Human in the Future Perspective of the Humanity. *Filosofija. Sociologija*, 35(1), 23–30.

Coeckelbergh, M. (2020). *AI Ethics*. Cambridge, MA and London: The MIT Press.

Cohen, E. and Spector, S. (2020). Transhumanism and Cosmic Travel. *Tourism Recreation Research*, 45(2), 176–184.

Corjescu, D. (2022). *Prometheus Unbound: The Perils and Promises of Transhumanism*. Hannover: Ibidem Verlag.

Danaher, J. (2019). *Automation and Utopia: Human Flourishing in a World Without Work*. Cambridge, MA: Harvard University Press.

Etzioni, A. (2017). Job Collapse on the Road to New Athens. *Challenge*, 60(4), 327–346.

Ewuoso, C. and Fayemi, A.K. (2021). Transhumanism and African Humanism: How to Pursue the Transhumanist Vision Without Jeopardizing Humanity. *Bioethics*, 35(7), 634–645.

Gómez-Baggethun, E. (2023). Work and Needs in a Finite Planet: Reflections from Ecological Economics. In S. Villamayor-Tomas and R. Muradian (eds.) *The Barcelona School of Ecological Economics and Political Ecology*. Cham: Springer, pp. 357–366.

Grant, A.S. (2023). Will Human Potential Carry Us Beyond Human? A Humanistic Inquiry into Transhumanism. *Journal of Humanistic Psychology*, 63(1), 36–50.

Gully, P.R. (2020). Pandemics, Regional Outbreaks, and Sudden-Onset Disasters. *Healthcare Management Forum*, 33(4), 164–169.

Hassan, L. (2017). Governments Should Play Games: Towards a Framework for the Gamification of Civic Engagement Platforms. *Simulation & Gaming*, 48(2), 249–267.

Hjorth, L. (2018). Ambient and Soft Play: Play, Labour and the Digital in Everyday Life. *European Journal of Cultural Studies*, 21(1), 3–12.

Jenkins, M. (2019, October 24). Finding Meaning in a Future of Less Work. *Medium*, https://medium.com/cgo-benchmark/finding-meaning-in-a-future-of-less-work-ebd5412bdf14

Jones, J.A. (2023). Expanding Understandings of 'Work' in Response to AI. *Humanistic Management Journal*, 8, 379–397.
Kalpokas, I. and Kalpokiene, J. (2023). *Regulating the Metaverse: A Critical Assessment*. London and New York: Routledge.
Knell, S. and Rüther, M. (2024). Artificial Intelligence, Superefficiency and the End of Work: A Humanistic Perspective on Meaning in Life. *AI and Ethics*, 4, 363–373.
Kurzweil, R. (2024). *The Singularity Is Nearer: When We Merge With AI*. New York: Viking.
Larsen, L.J. and Walther, B.K. (2020). The Ontology of Gameplay: Toward a New Theory. *Games and Culture*, 15(6), 609–631.
Leidenhag, M. (2020). Saved Through Technology: Exploring the Soteriology and Eschatology of Transhumanism. *Religion Compass*, 14(11), 1–9.
Lindberg, S. (2023). *From Technological Humanity to Bio-technical Existence*. New York: State University of New York Press.
Lipowicz, M. (2023). Beyond Transhumanism: A Nietzschean Critique of the Cultural Implications of the Techno-Progressive Agenda. *International Journal of Philosophical Studies*, 31(4), 522–546.
Lyreskog, D.M. and McKeown, A. (2020). On the (Non-)Rationality of Human Enhancement and Transhumanism. *Science and Engineering Ethics*, 28, 1–18.
Mariqueo-Russell, A. and Read, R. (2019). Fully Automated Luxury Barbarism. *Radical Philosophy*, 2(6), 108–110.
Mason, J. (2022). Transhumanism, Motion, and Human Perfection. *Christian bioethics: Non-Ecumenical Studies in Medical Morality*, 28, 185–196.
Mathers, A. (2020). Universal Basic Income and Cognitive Capitalism: A Post-Work Dystopia in the Making? *Capital & Class*, 44(3), 325–343.
Mäyrä, F. (2017). *Pokémon Go*: Entering the Ludic Society. *Mobile Media & Communication*, 5(1), 47–50.
Mercer, S.J.R. (2021). Humanism and the Sociology of Post-Work. *Economy and Society*, 50(4), 686–706.
Ogden, N.H. and Gachon, P. (2019). Climate Change and Infectious Diseases: What Can We Expect? *Canada Communicable Disease Report*, 45(4), 76–80.
Porter, A.C. (2023). Transhumanism and Posthumanism(s)on Education. *International Journal of Philosophical Studies*, 31(4), 475–500.
Rawlings, J. (2018, August 16). Finding Meaning in a Post-Work World. *Medium*, https://medium.com/the-politicalists/finding-meaning-in-a-post-work-world-675402be3b60
Richardson, I. and Hjorth, L. (2019). Haptic Play: Rethinking Media Cultures and Practices. *Convergence*, 25(1), 3–5.
Snape, R. et al. (2017). Leisure in a Post-Work Society. *World Leisure Journal*, 59(3), 184–194.
Sorgner, S.L. (2022). Transhumanism. In M.R. Thomsen and J. Wamberg (eds.) *The Bloomsbury Handbook of Posthumanism*. New York: Bloomsbury Academic, pp. 35–46.

Spencer, D.A. (2023). Technology and Work: Past Lessons and Future Directions. *Technology in Society*, 74, 1–7.

Spencer, D.A. (2025). AI, Automation and the Lightening of Work. *AI & Society*, 40, 1237–1247.

Srnicek, N. and Williams, A. (2015). *Inventing the Future: Postcapitalism and a World Without Work*. London: Verso.

Suleyman, M. (2023). *The Coming Wave*. New York: Crown.

Susskind, D. (2023). Work and Meaning in the Age of AI. *Brookings Institution*, https://www.brookings.edu/wp-content/uploads/2023/01/Work-and-meaning-in-the-age-of-AI_Final.pdf

Thomas, A. (2024). *The Politics and Ethics of Transhumanism: Techno-Human Evolution and Advanced Capitalism*. Bristol: Bristol University Press.

Thomas, Z. (2020, April 19). Coronavirus: Will Covid-19 Speed up the Use of Robots to Replace Human Workers? *BBC*, https://www.bbc.com/news/technology-52340651

Vaughan, A. (2019, October 15). How Deadly Disease Outbreaks Could Worsen as the Climate Changes. *New Scientist*, https://www.newscientist.com/article/2219981-how-deadly-disease-outbreaks-could-worsen-as-the-climate-changes/

7 From Robot Wars to Space Colonialism

AI and Entitlement to Violence

I, Robot Warrior

To some extent, the introduction of lethal autonomous weapons systems (LAWS) into the battlefield would continue already observable processes. The increasing distance through which war is fought is a long-term historical trend, for example, moving away from direct combat through spears, bows and arrows, and later gunpowder to planes and intercontinental missiles (Bode and Huelss 2022: 16). Nevertheless, this would also be a continuation with a difference: Some (see, e.g., Ploumis 2022: 5) would see in the introduction of LAWS an unprecedented shift in the ways wars are fought – one even greater than the invention of, say, gunpowder or nuclear weapons. The reason for such an assertion is that all previous technologies still required human input in selecting a target and making the decision to trigger the weapon at hand. With LAWS, meanwhile, such decisions are set to be part of the weapon's autonomous operation.

The changes brought about by LAWS go beyond increasing distance and changing operational performance. Similar to the previous fundamental innovations in military technology (but perhaps to a greater degree), LAWS are bound to create a two-tier system of warfare. For example, Glonek (2024: 90) portrays the effects of AI in military affairs as 'enormous', offering unprecedented 'efficiency gains', particularly in terms of increasing situational awareness and speeding up warfare, with those using more advanced AI tools and weapons likely to be able to complete a military engagement even before the less technologically advanced are capable of understanding what is going on, let alone attempting to defend themselves or otherwise respond. Therefore, despite some allure of allegedly victimless robot-on-robot warfare, such sanitisation of war is hardly feasible: '[t]he violence, however, is always inflicted on people. It will always be so, because it is pain and suffering that causes the enemy to surrender' (Scharre 2018: 303). Consequently, inequality on the battlefield would be radicalised, with the technologically advanced side being

DOI: 10.4324/9781003715849-7

sheltered by its highly efficient technological proxy and the less advanced side bearing the human cost.

Moreover, the delegation of human decision-making will likely go beyond the operational level on the battlefield, extending towards the strategic domain as well because AI agents 'will not merely force-multiply existing advanced weaponry but will likely become de facto strategic actors (planners, warfighters, tacticians) in warfare' (Johnson 2022: 246–247). Likewise, LAWS can be expected to become a force multiplier, offering the ability to cause damage that could potentially far exceed the human size of the force deploying LAWS (Ploumis 2022: 5). A notable issue is also proliferation: As the technology develops, 'it is likely that private companies, people, and armed non-state groups will start using them' (Veress 2022). Hence, in addition to states of various sizes, such weapons would likely become a force multiplier of choice for non-state actors, including insurgent and terrorist groups, as well as private military companies, particularly those operating from less strict jurisdictions or enjoying tacit support of states. Overall, then, LAWS can be seen as capable of upending pre-existing notions of strategy and power and, therefore, potentially playing a destabilising role.

A further destabilising aspect might be that, to put it bluntly, '[f]ully autonomous weapons will make it easier and cheaper to kill people' (Veress 2022: 89). Indeed, as Ploumis (2022: 7) observes, a combination of 'the remote character of this type of warfare' with 'the fact that the bloodshed will remain unseen' would likely re-render the decision to militarily engage an opponent seemingly costless and effortless. Moreover, the removal of one's own (human) forces from harm means that LAWS could 'make the use of force appear more legitimate' (Bode and Huelss 2022: 59). In the same vein, the focus on the efficiency of autonomous warfare could make planners lose track of other criteria and relevant factors as well as obfuscating the balance between means and ends when making decisions pertaining to war (Johnson 2022: 255). The balance could also be altered by discounting human decision-making capacity in the first place by way of 'framing humans as biologically and psychologically fallible […] and inferior to intelligent machines', removing both soldiers and civilians from 'the moral, ethical, and legal decision to use lethal force' (Johnson 2022: 250). Of course, it is not impossible to at least partly dismiss such arguments: For example, as Baker (2022) claims, the same accusations of situation blindness can be raised against, for example, bomber pilots, long-range artillery operators, and most others not engaged in person-on-person combat. Nevertheless, operationally, LAWS could also lower the perceived burden of killing: After all, '[i]f the people who launched autonomous weapons did not feel responsible for the killing that ensued, the result could be more killing, with more suffering overall' (Scharre 2018: 275; see also Bartneck et al. 2020: 98; Johnson 2022: 256).

While today the act of killing might be geographically removed (as in the case of remotely operated drones), it is still the human decision that counts, thereby largely retaining the burden; autonomous weapons would, however, completely upend that relationship by removing the decision-making barrier. Likewise, the very presence of LAWS may become part of a motivation to wage war, that is, 'use capabilities just by virtue of possessing them, having invested time, energy, or political capital and resources in their acquisition' (Johnson 2022: 253). In these diverse ways, the world might become more dangerous, not less so.

More broadly, LAWS can be seen as part of the trend to externalise the burden of war – either to surrogates (such as various non-state actors, guerilla groups, etc.) or to technology (Krieg and Rickli 2019: 85). However, such externalisation is only one-sided: Removing the more technologically advanced – privileged – side from harm's way while condemning the other side to suffer the human cost. Notably, the preceding also opens important ethical issues. For many, it seems inherently unfair if LAWS are used against human soldiers precisely because while one side is shielded from human harm by conducting war over distance, the other is condemned to death and injury (Bartneck et al. 2020: 98). Also, the very act of delegating life and death decisions to machines tends to be dismissed as immoral while being condemned to such a death is seen as 'ultimate indignity' (Bartneck et al. 2020: 98). No less importantly, those subjected to autonomous lethal force may well be killed or injured in a manner that can be seen as tainted by machinic disinterestedness. Crucially, algorithms lack intuition and 'other non-calculable human qualities, such as compassion, respect, emotion, mercy' and are, consequently, unable to 'sense something is wrong and change a course of action without explicit orders' (Johnson 2022: 258). Even more problematically, if other contexts of AI's use can serve as an example, once AI decision-making becomes part of everyday operations, strategic culture, and military doctrine, its use is likely to be taken as natural and not meriting further questioning or even as good in itself (Johnson 2022: 255). In this way, decisions may be distorted by a status quo bias: An idea that if something is, it therefore must be right.

Arguing against the use of LAWS on the battlefield, authors tend to cite the incapacity of autonomous weapons systems to correctly interpret the complex situations that characterise today's battlefield, including judgements of proportionality and distinguishing between combatants and non-combatants, thereby putting them at odds with International Humanitarian Law (IHL) (Veress 2022: 93–94). This is seen to be the case because decisions that pertain to distinction and proportionality require significant nuance, making IHL compliance problematic (Bode and Huelss 2022: 43–44). By contrast, for others, as machine vision technologies improve and are supplemented with other sensing devices onboard weapons systems, there may come a time when LAWS become more accurate

than humans, presumably turning their use into an ethical necessity (Bartneck et al. 2020: 95). Nevertheless, the latter arguments are dependent on the assumption that human fallibility, in both psychological and biological terms, can be overcome by AI, thus 'making war more rational, predictable, and controllable' (Johnson 2022: 250). Notably, while humans may occasionally overstep the limits of what is permissible, they can also have a broader range of constraints, such as moral and legal imperatives, that are meaningless to robots (Veress 2022: 89). However, with increasing use of LAWS, a sense of constraints and broader sensitivity can be lost even while keeping humans in the loop, particularly as commanders become removed from the battlefield, possibly having never experienced its horrors throughout their careers (Johnson 2022: 259). A further issue at stake is motivation: While humans in war are, or are supposed to be, motivated by allegedly noble pursuits, such as patriotism and a sense of justice, LAWS are seen to operate merely based on cold calculation. However, it must also be admitted that not all human combatants are motivated merely by a desire to promote their nation's just interests, thus making the just motivation argument significantly less straightforward (Baker 2022).

A further objection pertains to the accountability gap, which is opened as machines cannot be held accountable, while humans may become too far removed from actual decision-making for accountability purposes (Veress 2022: 90; see also Scharre 2018: 261–262). The higher the autonomy of such weapons, the larger the gap, and for fully autonomous weapons, this gap is near total. At the very least, responsibility can be too fragmented and divided (between, e.g., employers, manufacturers, software designers, etc.) to remain meaningful (Bartneck et al. 2020: 96). Also, humans tend to overtrust machinic judgement, which could have fatal consequences on the battlefield (Johnson 2022: 256). For these reasons, at least an element of human control is seen as necessary (Krieg and Rickli 2019: 132–133). However, counterarguments are possible here as well: If humans are retained in the loop while needing to keep track of large datasets at near-machinic speeds, part of the decision-making would 'not likely be accessible to human reasoning' (Bode and Huelss 2022: 4). Moreover, definitions as to what constitutes 'meaningful' human control differ significantly, thereby making it difficult to come to a shared understanding among states and their militaries (Bode and Huelss 2022: 214).

Meanwhile, for the proponents of LAWS (or, at least, the more neutrally minded), the arguments against such technologies tend to seem highly conservative. For them, military ethics and laws of war risk being trapped in historical understandings of what war is, thereby failing to address the high-pace, heavily technology-reliant wars of today, let alone proximate development in areas like LAWS (Baker 2022: 40). For this reason, it would appear that the likely positive externalities of LAWS ought to be explored.

One of the thorniest issues with LAWS is human intentionality. While the mainstream argument is that LAWS reduce intentionality by shifting decision-making from humans to machines, others disagree: For example, according to Baker (2022: 44), when aiming to destroy a particular target with artillery fire, one option would be saturating the area with explosives so that to ensure the likelihood of a target being hit while the other is to launch autonomous weapons to independently identify and engage the target, without simply aiming in the general direction; arguably, if the intended target can be destroyed more efficiently, then human intentionality is only increased. Hence, the argument goes, LAWS can increase control by striking the intended target more precisely (including autonomously locating, identifying, and engaging the target in question), more accurately fulfilling the intentions of human military operators and not vice versa (Baker 2022; see also Tigard 2021). Following this point of view, the key problem is simply that 'control has been confused with direct manipulation' (Baker 2022: 81). If this position is upheld, one could argue that failure to use more precise LAWS and opting for traditional 'dumb' weapons can be reframed as failure to exercise meaningful control (Baker 2022: 83). However, it must also be stressed that such examples do not consider other means for increasing efficiency that are short of full autonomy. Even more fundamentally, the carrying out of such intent can be obfuscated, thereby likely opening a new (cyber) battlefield: One example could be adversarial attacks that make image recognition systems misclassify the visual input that they receive (Scharre 2018). Indeed, successful cyberattacks can turn such weapons into loose cannons. There are further issues as well, such as the malfunctioning of an autonomous weapon with disastrous consequences. After all, just like any AI-powered system, such a weapon would not have common sense of self-reflection capacities, meaning that, contrary to a human soldier, it would have no way of understanding that it is making a mistake. To make matters even worse, the same training error, programming bug, or data manipulation that has caused the mistake would probably be repeated across every weapon of the same type, scaling up the problem exponentially (Scharre 2018). Hence, intentionality could be limited in even more ways.

A notable argument in favour of autonomous weapons in the battlefield is that not only LAWS can engage a target more quickly and more precisely, thereby limiting the likelihood of collateral damage; also they do not tire or become carried away by emotions – all in all, they can be seen as a more reliable alternative (Scharre 2018: 6; see also Bartneck et al. 2020). If civilian casualties can be reduced by using LAWS, their use might, again, be seen as even a must (Bartneck et al. 2020: 97). Indeed, a similar argument is also voiced by Galliott, Cappuccio and Wyatt (2022), for whom the superior targeting and identification capacities of LAWS (at least the way they see it) would make war more ethical and humane.

Other authors go as far as to assert the possibility of artificially making machines moral: Not only pre-programming moral codes into machines but also using machine learning techniques to ensure that LAWS learn moral behaviours by observing human actions and reactions (Tigard 2021; see also Bode et al. 2023). Nevertheless, issues of algorithmic and data bias in civilian applications of AI also pinpoint some of the difficulties that would need to be overcome (Bode et al. 2023: 3), while possibilities of intentional manipulation create dangers not unlike those with regard to intentionality above.

According to some, human control would also be retained because LAWS would only engage predefined targets, which means that they would not be capable of meaningfully deciding on life and death – in fact, '[t]o speak of LAWS as making the decision to kill is to anthropomorphise LAWS' (Baker 2022: 88). Similarly, Scharre (2018: 27) emphasises the need to avoid overplaying the independence of LAWS as 'machine autonomy doesn't require a magical spark of freewill or soul. Autonomy is simply the ability to perform a task or function on its own' (Scharre 2018: 27). Instead, autonomous weapons operate by independently working out the ways towards a goal specified by humans and performing the task at hand on their own (Scharre 2018). However, a counterargument could be made against that as well: The conditions described here do not reflect full autonomy but merely a stepping stone towards it. In search for greater efficiency, state militaries and non-state actors would more than likely search for additional ways of removing 'inefficient' humans from the loop in target selection as well. Hence, fully autonomous target selection might be further in the pipeline, but cannot be discarded as a mere anthropomorphic fantasy.

Continuing on the matter of speed and efficiency, LAWS may become unavoidable as the speed that comes from automation (and, further, addition of autonomous capacities) on one side drives the other side(s) to step up; moreover, there are further inherent benefits of autonomous operation, such as overcoming an adversary's signal jamming efforts (Scharre 2018: 8), providing a significant motivating factor for adoption, giving rise to a pervasive feeling that 'humans will soon become the Achilles heel in the emergent AI-enabled techno-war regime' (Johnson 2022: 248). Moreover, it has to be kept in mind that the stakes are extremely high: As Scharre (2018: 94) stresses, '[t]he battlefield is an unforgiving environment' that typically does not give second chances to states and militaries caught lagging behind in the technological innovations race, thereby implying that when it comes to pursuing and adopting innovation, it is better to be safe than sorry. Hence, once again, civilian and military decision-makers may simply feel like they have no other option but to embrace LAWS. As the feeling is likely to be shared among competing powers globally, an autonomous weapons arms race seems imminent (see, e.g.,

Bode and Huelss 2022), with standards and use norms likely being reduced to the lowest common denominator. Moreover, the building blocks for LAWS, including the necessary software and AI models, are already widely available, potentially resulting in largely uncontrollable proliferation.

As always, evaluations of what the development and proliferation of LAWS would mean for humanity differ. For some, fears of autonomous weapons seem mostly based on science fiction cliches and result from failure to address the ways in which military technologies are rigorously tested before being deployed on the battlefield (Galliott, Cappuccio, and Wyatt 2022). Hence, the argument goes, attempts to regulate or even ban such weapons would not be productive. In fact, according to the same authors, a ban on LAWS would only hinder attempts to create military AI technology in ethical and responsible ways and with proper safeguards in place, while ushering in unregulated and uncontrolled proliferation underground (Galliott, Cappuccio, and Wyatt 2022). For others, instead of sweeping international agreements, one could perhaps expect an emergence of a more customary set of norms, derived from the way in which AI-powered military systems are designed and used (Bode et al. 2023). Even if proper customary international law norms are not established, we are still likely to witness the emergence of norms as shared understandings of 'how things are done', as opposed to legal obligations (Bode and Huelss 2022). However, this not necessarily seen as a best-case scenario, since it would likely open the door to silent (removed from public scrutiny) and tacit spread of norms that are convenient for states but may lack necessary attention to key variables, such as a consensus on ceding human control; for this reason, global governance frameworks may end up being useful, at least for deliberation purposes (Bode et al. 2023: 10). Moreover, at least in some circumstances, the emergence of even customary norms and practices might be too slow due to the aforementioned sense of immense urgency – the perceived need to develop new technologies before competitors are able to do that (see, e.g., Glonek 2024).

Ultimately, robot warfare can be seen as one of the paradigmatic examples of possibility turning into (perceived) unavoidability. With the development of AI technologies, lethal autonomy – the capacity of weapons systems to operate without human intervention – becomes a possibility. Autonomy is already increasingly integrated into existing weapons (such as drones) to a varying extent, and this trend is set to grow. Therein also lies a springboard for the second step: If battlefield conditions necessitate (or seem to necessitate) the introduction of such capacities or one's opponent is (or appears to be) taking a lead in the development of such technologies, it can turn into a must. Hence, a crucial crossroads is glossed over, leaving the dilemmas around automated killing to irrelevance – moral blindness par excellence.

Space: Colonialism Strikes Back

Visions for human use of outer space tend to involve technologically enabled attempts to transcend humanity's earth-bound nature, particularly in terms of resource exploitation and settlement. Hence, a key question that arises is, similar to the above, to what extent do the present and projected technological capacities translate into (perceived) entitlements or even imperatives, perhaps most clearly encapsulated in the narratives of space colonisation below. Such visions, at least in the West, are shaped by a techno-utopian ideology that permeates the Silicon Valley; through a combination of optimistic techno-determinism and libertarianism, the latter views the entire world (indeed, the entire Universe) as a resource that ought to be appropriated for value extraction, necessitating unobstructed technological development of technology, the latter itself seen as a good in itself (Tutton 2021). Part and parcel of this vision is transcending the terrestrial nature of humanity. Indeed, humanity becoming a multiplanetary species is framed as part of a linear vision of progress – as something that allegedly constitutes a mere continuation of human evolution (Tutton 2021).

At the heart of such visions of the future are projects for space colonisation. However, the term itself is simultaneously problematic and symptomatic. Crucially, this is not merely about careless use of a loaded term: The problem is not only that colonisation 'evokes supremacy, racism, subjugation, and disrespect for native forms of life' but also that it is the *intended* meaning, that is, the term accurately describes the essence of such projects (Calanchi 2023: 27). Of course, the absence of known life forms in space (especially of intelligent life) is often seen as a vindication of colonial discourse (Utrata 2024; see also Padden 2022). However, this excuse does not address the historical and ideological underpinnings of the term's use in any meaningful way.

Still, a significant blind spot in the defence of space colonialism is its universalising perspective: While human exploitative expansion beyond Earth is seen as necessary for improving the human condition and achievement of the good life (see, e.g., Balistreri and Umbrello 2022), there is no consideration as to the meaning and values embedded in such terms. After all, one first needs to ask according to which and whose standards 'improvement' and 'good life' are measured. Moreover, it is important to note that those advocating for space colonialism typically do so from a privileged Western standpoint while failing to account how their own situatedness is built upon the exploitative practices and injustices of earlier forms of colonialism (Szocik 2023: 2). Moreover, there is an even further way (besides whitewashing the injustices) in which the colonising approach is ahistorical: Just like the previous wave of colonisation never brought widespread and fairly distributed wealth even to European populations

and have come at an immense cost to past and present generations of those who had been colonised (Szocik 2023: 2), there is no reason to believe that space colonisation would be any more egalitarian (Calanchi 2021: 291 see also Deberdt and Le Billon 2023).

The ideological construct of space colonisation is also often accompanied by related terms, including the *conquest* of space (thus further underscoring the subjugation aspect) and pushing at the final *frontier* (Adar 2022). Indeed, the exploitation of outer space is enabled and legitimised through specific ideological-discursive practices: Namely, 'techno-utopian imaginaries expanding extractive frontiers beyond Earth' (Deberdt and Le Billon 2023: 1879). Frontiers themselves have a colonial history, one that enabled externalised destruction and exploitation: Frontiers used to be imagined as the borderline at which civilisation, law, and entitlements or rights ended, and the wilderness began. Here, we simply observe a shift from terrestrial to extraterrestrial frontiers (Bennett 2022; Utrata 2024). The frontier discourse also legitimates the idea that humans are destructive and exploitative by nature and, therefore, the only way to protect or, at least, prolong (human) life on Earth is by 'exporting' those same tendencies to new frontiers, with the latter also signifying the outer limits of concern – beyond the frontier, no moral or other concerns are seen to exist (see, e.g., Szocik and Reiss 2023; see also Smith 2020: 1364).

The loss of concern beyond the frontier is also manifested in the legitimation of outright destruction. While the ethical issues of today's space exploration include precluding the spread of terrestrial microbial life across space, projects that are significantly larger in scale and fundamental in their effects, such as terraforming, are increasingly put on the agenda (Szocik et al. 2020). Broadly, terraforming means artificial transformation of atmospheric and other conditions on an extraterrestrial body so that they resemble those on Earth and, therefore, become able to sustain human (and other terrestrial) life. Hence, both contamination and terraforming refer to a preference for terrestrial life over the natural conditions and, hypothetically, life elsewhere (Szocik et al. 2020: 3). For some, projects like terraforming are not only a hypothetical technological possibility but also a moral necessity, the underlying assumption being that 'an inhabitable space is better than a non-life space, and humans, as self-conscious beings, have the right (sometimes even the duty) to populate space' (Szocik et al. 2020: 3). Hence, harm to extraterrestrial environment is seen as an insufficient reason to reconsider the expansionist and colonialist visions of outer space exploitation (Szocik and Reiss 2023). Of course, 'inhabitable' here is meant from a squarely and exclusively human perspective, while the duty to populate space is ominously reminiscent of self-serving ideas from previous epochs of colonialism, such as 'the white man's burden'.

Environmental modification and other technological projects are, of course, not seen by proponents as self-serving – they are merely

instrumental to the bigger goal of 'colonising' space. The latter also often tends to be framed in terms of moral imperatives, for example, ensuring that 'the total future number of humans who will exist and whose lives will be worth living could be orders of magnitude greater than today' (Kovic 2021: 1). Space colonisation is also seen as a means for mitigating the risk of extinction, with the latter again being presented in a strongly value-loaded way: '[i]f humankind goes extinct or stagnates prematurely, the majority of humankind's positive future value (the many thousands of generations and many billions of people who could lead lives worth living) would be lost' (Kovic 2021: 1). Indeed, despite the progressivist appeal of the vision of humans as a spacefaring species, the massive effort and investment required to pull off such an effort requires additional justification. Hence, a dire account of the present and future life on Earth is typically invoked alongside a linear account of progress to further not only the necessary commitment but also a hands-off regulatory approach when it comes to space projects. Such projects are presented as existentially urgent (escaping the doom and gloom on Earth, usually associated with the climate catastrophe) and natural (fulfilling the drives of evolution and progress) – in other words, too important to be impeded in any way.

In addition, efforts are also seemingly made to whitewash the history of European colonialism, equating the 'discovery' of 'new' lands by European explorers (itself a strongly value-loaded narrative that implies the presence of allegedly open spaces as if waiting to be 'discovered' and taken) with a linear narrative of progress and, therefore, discarding any critical interventions as backwardly and dangerous (see, e.g., Ćirković 2022). The latter are dismissed as 'enslaved by manifestly narrow, parochial, and at least implicitly counter-Enlightenment dogmas', negativity-focused 'hysterical' extremists espousing a 'parochial, "closed box" view of the world, incompatible with the cosmic future of humanity', and manifesting a level of geo-centrism unseen since the Inquisition (Ćirković 2022: 2). In other words, sceptics of space colonialism are proactively framed as bigoted enemies of not only progress but also of humanity as such (a strategy also espoused by some proponents of transhumanism, as seen in the previous chapter).

The whitewashing of colonialism also extends into societal engineering – creating new, better-governed, and more prosperous societies on extraterrestrial bodies. To support such projects, alleged success stories of European colonisation are being put forward, such as the British colonisation of Australia or the establishment of British colonies in North America (see Szocik, Wójtowicz, and Braddock 2020). Tellingly, there is no mention of the cost of creating these 'new' societies, just as in the case of terraforming and similar projects. Other dubious references include touting the 1884 Berlin Conference – in which European countries divided Africa among themselves for colonisation – as a prudent way to smoothly

expedite colonisation (Crowell 2020: 83; see also Smith 2020). Elsewhere, emphasis is similarly put on the need for 'a robust system of colonization' to establish the rule of law and build a space civilisation (Smith 2020: 1390). Such an emphasis on territorialisation sits uneasily with the emphasis on equal peaceful use and free access, espoused in current international law (in particular, the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (1967) – hereafter, Outer Space Treaty).

As Adar (2022: 97) observes, the push for space colonies 'sanitises and erases the genocide, cultural subordination, economic exploitation, and other harms' of terrestrial colonisation; hence, instead of the use of force, enslavement, violence, and exploitation that made up the *actual* experience of colonialism, here, references to colonisation 'are used to invoke a romantic and heroic image of hard-working people overcoming the harsh environment of space for the benefit of humanity'. Here, a double injustice is being carried out: Not only towards the future object of expansion but also towards those who have experienced or live in the aftermath of colonialism, because the harms are simply plastered over.

Proponents of space colonisation also tend to raise the stakes to the maximum by positioning their preferred course as a matter of human survival, deeming the Earth doomed, either unavoidably or without there being alternative places to exploit (Szocik and Reiss 2023). Hence, when all the relevant factors are considered, such as 'the capacity for human pleasure and suffering, overpopulation of the Earth, the dynamics of human economy, human nature and the fact that the existence of life in our Solar System, especially sentient life, has not yet been found', the conclusion typically drawn is that 'there is a strong argument that the colonisation of space becomes a moral duty' (Szocik and Reiss 2023: 4). Such arguments are used to not only justify but also to add urgency to colonisation efforts, framing the added value of settlement-building at the earliest possibility as 'immense' (Kovic 2021: 1–2). Of course, the idea of a 'moral duty' to colonise in the preceding is particularly perplexing. However, additional troubling features must be highlighted. Notably, the destructive – not merely extractive and exploitative – logic of (space) colonisation is clear: The only worry or concern seems to be the finite nature of resources and, therefore, that ultimately not only the Earth but also the Moon and Mars would be stripped of them; however, instead of inspiring a reckoning, such a possibility (or, rather, unavoidability) is met with a hope that by the time this happens, technological advances would allow humanity to take its exploitative colonising endeavours even further (see Szocik and Reiss 2023: 5). Even if voluntary self-imposed limits to how much can be extracted are put in place to ensure incomplete exploitation, this is typically done from a perceived position of superiority, as an act of benevolent grace, and not as a matter of serious moral commitment

(see Szocik, Wójtowicz and Braddock 2020) and with no attention put to more sustainable practices that would reduce the need for exploitation in the first place.

No less importantly, there is built-in inequality in visions of space colonialism. Even if one accepted the argument that living in extraterrestrial human settlement would be either an attractive prospect or the last refuge of humanity, the benefits of it would likely be limited to those sufficiently wealthy to afford space travel and/or citizens of spacefaring states (and even then unevenly, as in earlier colonialism), leaving the vast majority of humanity to toil in the conditions from which the privileged ones are capable of escaping or even to perish without sufficient access to space (Szocik 2023). Meanwhile, if conditions in space were bad and arduous, making life in extraterrestrial colonies undesirable, space settlement would be carried out by migrant workers and other marginalised groups, with the fruits of their labour being sent back to Earth to those privileged enough to stay. In either case, the outcome would be exclusionary.

The exploitative agenda is further underscored by some of the proposals for allocating property rights on extraterrestrial bodies, such as those based on efficient use of territory. In other words, a property claim would only be valid as long as the territory was exploited in the most efficient way possible – otherwise it would revert to *terra nullius* since anything insufficiently exploited is seen as 'waste' (see, e.g., Crowell 2020: 83; 97). Elsewhere, the current system that does not allow capitalisation on space resources through claiming ownership over space bodies and whatever is extracted from them, but instead frames the benefits derived from space as a common property of humankind (as provided for in Article 2 of the Outer Space Treaty), is seen as holding humanity back (Smith 2020: 1363). By contrast, it is asserted, '[s]overeign control over land will also promote its use in an efficient, productive way' (Smith 2020: 1388). Notably, this view is grounded in historical justifications of colonialism (e.g., Lockean): It was precisely the industrious and efficient use argument that Locke used to disqualify Native Americans from property rights.

Notably, the above can be seen as further increasing inequalities: While the need for (mostly private) capital for the initiation of space exploitation and colonisation leaves the majority of individuals (and even states) outside the capacity to even claim ownership in space, regardless of the criteria, those able to lay claim over space resources (and, perhaps, entire space bodies) as well as over human colonies would benefit from an enormous boost to their wealth and power (Deberdt and Le Billon 2023). The latter also illustrates the self-serving nature of technological capacity, as everything that the technology of a given time allows to be reached and made use of is transformed into territory that can be owned and exploited (Utrata 2024). While previously the latest technology had enabled those with sufficient power and capital to reach far-flung corners of the Earth,

thereby even further increasing their power and capital, in the projected future, the same would increasingly apply to outer space.

Interestingly, space colonisation tends to be touted as a sure pathway towards sustainable development (Szocik and Reiss 2023) – which is sustainable only if one is willing to adopt a significantly restricted perspective by considering only the Earth and ignoring literally everything else. According to Deberdt and Le Billon (2023: 1879), this gives rise to the highly paradoxical idea of 'green' extractivism, which is, of course, only a myopic fantasy of greenness. Instead, by externalising harms beyond the visible horizon, 'industrial interests intend to disentangle themselves from the sociopolitical and environmental risks of terrestrial extraction'; the outer space here offers 'new "resources" and "sinks" for ever-accelerating consumption' (Deberdt and Le Billon 2023: 1880). In this way, the commercial interests behind space colonialism are further strengthened to encompass not only gaining access to new resources but also having a licence to continue destructive practices as usual through a completely new form of greenwashing.

Pushing the already dubious argument even further, for the proponents of space colonisation, '[t]here is no structural difference between the argument for delaying space colonization and the infamous argument for delaying climate action' as delay in either poses an existential risk (Ćirković 2022: 5). Nevertheless, this is a false analogy. The first falsity is historical: One of the key factors behind the current climate catastrophe has been precisely the colonising and exploitative drives inherent in parts of humanity: As Szocik (2023: 2) scathingly puts it, 'in order to achieve good living conditions, we exploited the planet, and now we have to find a new one'. The second one is practical: It transpires that an intensification of space launches would become an environmental threat on its own (as per below). Meanwhile, the third one is logical: There is simply no necessary connection between the two.

Even more importantly, the idea of space colonialism 'paves the way for a settler-colonial futurity that eliminates other types of possible futures and relations with the environment' while refusing to reach out towards the root causes of the much-stressed threats to humanity: Space colonialism is merely seen as an excuse, or even permission, to pollute and destroy – the aim of space colonisation seems to be not a great leap and societal improvement but, instead, 'the preservation of the existing status quo on Earth' (Au 2024: 22–23). A crucial failure of space colonialism projects is that they perpetuate the toxic ideology that prevents appreciation of the environment (both terrestrial and extraterrestrial) for its inherent value and not for its resources (Calanchi 2023). Paradoxically, research suggests that intensification of space travel might itself have a significant negative environmental impact on Earth, in terms of the release of greenhouse gases, depletion of the ozone layer, and large deposits of aluminium

in the atmosphere from re-entering spacecraft (Miraux, Wilson, and Calabuig 2022; see also Klinger 2021; Deberdt and Le Billon 2023). In other words, humans may well exacerbate the very problem they are purportedly trying to solve.

Crucially, technology ends up opening new horizons for exploitation. The very possibility of settling on extraterrestrial bodies and exploiting their resources can in itself become an imperative to do so, blinding humans of other, more ethical, possibilities – again, mistaking a crossroads for a straight path. Hence, it is crucial to ensure that ethical considerations are not left aside. Yet again, the choice lies between a linear vision of progress (this time framed through colonialism and exploitation) and alternatives based on care and concern about the environment, extraterrestrial included.

Moving more narrowly to AI, it is a crucial technology in space operations, not least in terms of data analytics as well as command and control, both back on Earth *and* on the satellites, probes, and spacecraft launched to space (Graham, Thangavel, and Martin 2024). AI's functions include 'astronaut assistant, space healthcare, mission planning, satellite data processing, space debris mitigation, real-time spacecraft monitoring, and automated propulsion and navigation systems' (Graham, Thangavel, and Martin 2024: 197). In particular, the speed often necessary to mitigate emergencies in space and delays in communication with spacecraft at greater distances from the Earth (thus making even routine tasks more complicated) necessitate not only automation but also autonomy of control functions to ensure mission success (Soroka, Danylenko, and Sokiran 2022).

In the future, this is likely to be extended towards automated spacecraft (not only space probes but also, e.g., transport spacecraft moving supplies and goods to and from space settlements), robotised asteroid mining and other forms of space exploitation, and preparation for settler projects, such as site exploration and/or prefabrication of core settlement infrastructure (see, e.g., Martin and Freeland 2021). As the use of AI technologies in space expands, some of the crucial issues at hand will include regulation and governance on the one hand and liability for harms on the other (Graham, Thangavel, and Martin 2024; see also Soroka, Danylenko, and Sokiran 2022).

There is also a level of coloniality inherent here, particularly with regard to digital infrastructures, such as data centres designed for operation in outer space or on the Moon (Au 2024: 18) and space resources being used to build and maintain terrestrial AI infrastructures. Once again, this is in line with the idea of space as *terra nullius* (a general spirit of which is also present in the various openness provisions in the Outer Space Treaty), allegedly open for exploitation by those possessing sufficient capacity and resources. As the use of digital technologies and infrastructures spreads in space, cybersecurity threats remain an underappreciated domain in current discussions of space exploration and settlement; such

potential threats could include subverting satellite-to-satellite and satellite-to-Earth communications, communication, sensor, and control systems on spacecraft, destruction or poisoning of datasets, tampering with AI systems, etc. (Breda et al. 2023; see also Martin 2023), also further extending into tampering with space infrastructures and space-based mission servicing (see, e.g., Martin 2023) as well as the management and life support systems of any space settlements. With the growing competition in space, attempts to disrupt and subvert digital systems in open space or competing settlements are likely to become commonplace. While the Outer Space Treaty stipulates, in no uncertain terms, that space should be used for exclusively peaceful purposes (see, e.g., Article IV), the 'grey zone' nature of cyber conflict and attribution problems inherent in the latter make such security threats particularly likely.

Moreover, given the inhospitability of both open space and extraterrestrial planetary environments, the potential adverse effects of failures caused by cyberattacks are going to be significantly more severe. In this context, tackling cyber threats in space acquires a renewed impetus. If it is, arguably, too late to do anything about cyber warfare on Earth, shared extra vulnerability introduced by the extraterrestrial environment might bring states closer to some agreement, although even then, attribution, verification, and proliferation among non-state actors would remain problematic. Still, ensuring the safety of technologies used in space, particularly from external threats, would likely become a pressing issue, and ways of minimising such threats would constitute yet another crossroads. However, as with LAWS, the crossroads may be ignored simply because of a sense of (military) necessity.

Overall, visions of space colonialism bring the logic of efficiency and optimisation to its natural exploitative and destructive conclusion, only externalising it to a distant frontier. It also radicalises the equation of possibility and moral imperative by hanging the future of humanity on an allegedly unidirectional path towards space exploitation. In doing so, both terrestrial and extraterrestrial harms are glossed over, putting them outside the ambit of concern – a typical case of adiaphorisation. In the end, crossroads are ignored by not only making the path towards progress appear frictionless but also by erecting walls on both sides, with even a consideration of dilemmas and alternatives being framed as a betrayal of humanity.

References

Adar, P.A. (2022). Space and the Future of Humanity: TWAIL Critique of International Space Law and Space Discourse. *Third World Approaches to International Law Review*, 3, 92–115.

Au, Y. (2024). Data Centres on the Moon and Other Tales: A Volumetric and Elemental Analysis of the Coloniality of Digital Infrastructures. *Territory, Politics, Governance*, 12(1), 12–30.

Baker, D. (2022). *Should We Ban Killer Robots?* Cambridge: Polity.
Balistreri, M. and Umbrello, S. (2022). Should the Colonisation of Space be Based on Reproduction? Critical Considerations on the Choice of Having a Child in Space. *Journal of Responsible Technology*, 11, 1–7.
Bartneck, C. et al. (2020). *An Introduction to Ethics in Robotics and AI*. Cham: Springer.
Bennett, M.M. (2022). Uncanny Frontiers: Amazon in Space. *Dialogues in Human Geography*, 12(2), 181–185.
Bode, I. and Huelss, H. (2022). *Autonomous Weapons Systems and International Norms*. Montreal: McGill-Queen's University Press.
Bode, I. et al. (2023). Prospects for the Global Governance of Autonomous Weapons: Comparing Chinese, Russian, and US Practices. *Ethics and Information Technology*, 25, 1–15.
Breda, P. et al. (2023). An Extended Overview of Cyber Vulnerabilities of AI Technologies in Space Applications: Technological Challenges and International Governance of AI. *Journal of Space Safety Engineering*, 10, 447–458.
Calanchi, A. (2021). An Interplanetary Transplantation, or, Reloading the Anthropocene on the Red Planet. *Green Letters*, 25(3), 285–298.
Calanchi, A. (2023). De-Colonizing the Earth to Re-enact Colonialism on Mars New Forms of 'Transplantation' in Outer Space. *Iperstoria*, 22, 27–41.
Ćirković, M.M. (2022). The Nutshell Kings: Why Is human Space Settlement Controversial in the First Place? *Futures*, 143, 1–8.
Crowell, C. (2020). The Final Frontier: Preventing Space from Becoming the Wild West Through the Establishment of Internationally Recognized Property Rights. *George Mason International Law Journal*, 11(1), 81–104.
Deberdt, R. and Le Billon, P. (2023). Outer Space Mining: Exploring Techno-Utopianism in a Time of Climate Crisis. *Annals of the American Association of Geographers*, 113(8), 1878–1899.
Galliott, J.C., Cappuccio, M.L. and Wyatt, A. (2022, July 14). Taming the Killer Robot: Toward a Set of Ethical Principles for Military Artificial Intelligence. *Journal of Indo-Pacific Affairs*, https://www.airuniversity.af.edu/JIPA/Display/Article/3091254/taming-the-killer-robot-toward-a-set-of-ethical-principles-for-military-artific/
Glonek, J. (2024). The Coming Military AI Revolution. *Military Review*, May-June, 88–99.
Graham, T., Thangavel, K. and Martin, A.-S. (2024). Navigating AI-lien Terrain: Legal Liability for Artificial Intelligence in Outer Space. *Acta Astronautica*, 217, 197–207.
Johnson, J. (2022) The AI Commander Problem: Ethical, Political, and Psychological Dilemmas of Human-Machine Interactions in AI-enabled Warfare. *Journal of Military Ethics*, 21(3–4), 246–271.
Klinger, J.M. (2021). Environmental Geopolitics and Outer Space. *Geopolitics*, 26(3), 666–703.
Kovic, M. (2021). Risks of Space Colonization. *Futures*, 126, 1–14.

Krieg, G. and Rickli, J.-M. (2019). *Surrogate Warfare: The Transformation of War in the Twenty-First Century*. Washington, DC: Georgetown University Press.

Martin, A.-S. (2023). Outer Space, the Final Frontier of Cyberspace: Regulating Cyber Security Issues in Two Interwoven Domains. *Astropolitics: The International Journal of Space Politics and Policy*, 21(1), 1–22.

Martin, A.-S. and Freeland, S. (2021). The Advent of Artificial Intelligence in Space Activities: New Legal Challenges. *Space Policy*, 55, 1–10.

Miraux, L., Wilson, A.R. and Calabuig, G.J.D. (2022). Environmental Sustainability of Future Proposed Space Activities. *Acta Astronautica*, 200, 329–346.

Padden, H. (2022). Does Space Law Prevent Patterns of Antarctic Imperialism in Outer Space? *Global Policy: Next Generation*, 13, 346–357.

Ploumis, M. (2022). AI Weapon Systems in Future War Operations: Strategy, Operations and Tactics. *Comparative Strategy*, 41(1), 1–18.

Scharre, P. (2018). *Army of None: Autonomous Weapons and the Future of War*. New York: W. W. Norton & Company.

Smith, M.T. (2020). One Small Plot for Man, or One Giant Easement for Mankind?: New Approach to the Outer Space Treaty's Property for Mankind Principle. *University of Illinois Law Review*, 4, 1361–1392.

Soroka, L., Danylenko, A. and Sokiran, M. (2022). Legal Issues and Risks of Artificial Intelligence in Space Activity. *Philosophy and Cosmology*, 28, 118–135.

Szocik, K. (2023). Space Not for Everyone: The Problem of Social Exclusion in the Concept of Space Settlement. *Futures*, 145, 1–4.

Szocik, K. and Reiss, M.J. (2023). Why Space Exploitation May Provide Sustainable Development: Climate Ethics and the Human Future as a Multi-Planetary Species. *Futures*, 147, 1–6.

Szocik, K., Wójtowicz, T. and Braddock, M. (2020). The Martian: Possible Scenarios for a Future Human Society on Mars. *Space Policy*, 5, 1–11.

Szocik, K. et al (2020). Ethical Issues of Human Enhancements for Space Missions to Mars and Beyond. *Futures*, 115, 1–14.

Tigard, D.W. (2021). Artificial Moral Responsibility: How We Can and Cannot Hold Machines Responsible. *Cambridge Quarterly of Healthcare Ethics*, 30(3), 435–447.

Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (1967), 610 U.N.T.S. 205, 18 U.S.T. 2410.

Tutton, R. (2021). Sociotechnical Imaginaries and Techno-Optimism: Examining Outer Space Utopias of Silicon Valley. *Science as Culture*, 30(3), 416–439.

Utrata, A. (2024). Engineering Territory: Space and Colonies in Silicon Valley. *American Political Science Review*, 118(3), 1097–1109.

Veress, C.B. (2022). The Limits of Protection of Human Rights in Warfare Led by AI. *Acta Universitatis Sapientiae: Legal Studies*, 11(2), 81–97.

8 Conclusions

When considering the effects of technological change, one must first ask what constitutes a positive development in technological progress. Here, the dilemmas posed by conflicting imperatives (divergent value systems, human rights, technological drives, assumptions inherent in particular domains of human action, etc.) and, therefore, potential courses of action ought to be considered. These instances are crossroads in which a course of action, often with significant future ramifications, needs to be taken. And yet, today we can observe efforts to put technological development beyond scrutability and meaningful control, shifting attention towards frictionlessness, efficiency, and outcome maximisation. Technological progress is, thus, framed as inevitable, almost part of the laws of nature, and only for the better.

The issue is not just in failure to notice any crossroads but in the inability to even entertain the *possibility* or idea of a crossroads. Dilemmas are replaced with logics of efficiency and extraction, where a mere technological possibility to do something turns into an imperative to do so. Such failure to notice any crossroads is seen here as a manifestation of moral vacuum – after all, it is in moments of unavoidable choice that we face our moral obligations. Here, meanwhile, anything left outside the supposedly straight and frictionless path of progress becomes insignificant, unworthy of moral or other obligations, or simply remains unnoticed, resulting in moral blindness. Related to the latter, the concept of liquid evil is embraced in this book to denote the incapacity to see choices and dilemmas (crossroads) and exclusion of any alternatives – it is moral blindness turned into practice.

In education, AI is typically seen as a transformative technology that must be embraced to remain relevant, despite the lingering fears of plagiarism and other forms of cognitive outsourcing. The greatest value, however, is generally located in AI-enabled practices of optimisation. Datafication, analytics, and prediction, for example, are framed as value-neutral in their application but significantly contribute to increased learning outcomes. Likewise, AI is presented as having a major impact on

DOI: 10.4324/9781003715849-8

education in terms of engagement and personalisation of learning and teaching practices, contributing to the perception of an emerging frictionless path towards maximised outcomes. In atypical progressivist fashion, legitimacy is achieved by contrasting alleged backwardness and inefficiency in traditional education with frictionless data-based optimisation. The neutrality side of things, meanwhile, transpires to derive from learner surveillance that allegedly enables quasi-medical 'diagnosis' of educational 'deficit'. Here, one witnesses not only the delegation of agency in education to digital tools and platforms but also the imposition of the values of a narrow, mostly Western, technological elite as unavoidable, natural, and universal, seemingly for the benefit of all, with no legitimate deviation, putting a data-extractive mindset at the heart of education. Hence, crucial questions are glossed over, not least whose and what data are being used for prediction, whose and what standards of 'normality' are being enforced, or whose and what knowledge is reproduced by AI models when they are used as learning assistants. Cultural, historical, and outlook frictions and value dilemmas are occluded due to the supposed objectivity of AI outputs.

Simultaneously, AI leads to problematisation of previously taken for granted assumptions, such as the exclusively human nature of creativity. Thus far, crossroads in the domain of creativity are more clearly pronounced than in other domains. In part, this is because of a long-standing propensity to see creativity as a quintessentially human quality. Thus, automation comes more like a challenge to the status quo rather than its extension. The anthropocentric understanding of creativity is particularly clearly pronounced in terms of the protection of creative output. Nevertheless, even in this area, pressure arising from generative AI is forcing societies to look for alternatives as the lines between human and AI creativity are increasingly erased or, at least, blurred. Likewise, pressures towards efficiency and optimisation become drivers towards greater AI adoption to retain competitiveness in the arts market, raising the question as to whether there will *still* be room for crossroads once decisions about the status of AI creativity are made.

The amount of digital content likewise implies the need for a choice between ignoring the consequences of preservation (e.g., the environmental costs of data storage) on the one hand and (potentially radical) selection on the other. However, selection would likely end up in a form of blindness towards the other, primarily ignoring non-mainstream content. Moreover, the centrality of digital-first or even digital-only culture further complicates the choice of what qualifies for preservation and why. The default direction is towards the expansion of digital-only content, which also includes inauthentic content and content co-produced with automated digital entities. However, such dilemmas can be seemingly frictionlessly occluded by outsourcing decisions over memory and forgetting to AI.

More broadly, there is a growing need to decide on the relative value of humans vis-à-vis AI and platforms in terms of both production and selection of heritage, especially given the growing privatisation by way of platforms creating the conditions for heritage to emerge, becoming the custodians of heritage-to-be, and selecting what is worthy and what is not – a process that is deeply imbued with algorithmic bias and skewed towards digitally dominant cultures. This frictionless platform-centric progression is likely to cast aside – render non-valuable – anything that does not fit platform business strategies and content governance models, potentially removing even the capacity to approach any crossroads, with decisions always already having been made.

Human–AI interactions can, however, go even deeper, particularly with regard to anthropomorphising AI and thus engaging in quasi-human interaction. Particularly with regard to Virtual Influencers, frictionlessness and interaction optimisation through data analytics become the rule of the game. While currently, at least with regard to VIs, a *functional* choice between humans and virtual persons – their effectiveness – might still be open, there seems to be a broader trend towards virtual personalities as optimised communication partners and towards deeper parasocial relations. Not only do such hybrid interactions offer all-rounded escapism but also the digital shift allows for optimisation of the self in an avatar form (in terms of physical appearance, character traits, and other characteristics considered desirable) and maximisation of interaction efficiency through multiple digital AI-powered clones. This potentially comes at the expense of a continuous self, making it digitally adaptable to any situation. Anthropomorphising digital interlocutors allows for an additional facet of optimisation – that of communication partners, swapping human unpredictability for AI's adaptability, enabling frictionless interactions. This can further be extended into optimising relationships with the entire world, at least in virtual surroundings, as AI agents respond to every want, need, and craving – a world that moves beyond frictionlessness into pure hedonism, making it unlikely that any alternatives – crossroads – would even be considered.

Continuing down the hedonism path, exploring a future where work is automated becomes a way forward. According to the optimists, accelerating a post-future is an imperative, since it would bring about a society of abundance and leisure. This society would be characterised by self-development and aptitude-based association, supposedly enabling noble pursuits. What would likely persist are various data-generating activities that are usually not considered work due to their unpaid nature, although even those would be shrunk by the abundance of synthetic data. Still, one can foresee a rise in prominence of dual – data- and leisure-focused – virtual environments as new spheres of consumption, or perhaps simply as ways of keeping individuals busy and entertained when they do not need to do

anything, leading to virtual gamification of everyday life. While for some, post-work would lead towards high-brow pursuits, such as philosophy or the arts, advances in AI could render such activities meaningless as well, particularly if they are effort-intensive (such as creative activities) – instead, all pursuits could simply be delegated to AI assistants. The emergence of such spectatorship-first lifestyle and its technological enablement means that the question of what a good life entails ceases to be a philosophical problem and, instead, becomes a policy problem.

The latter can also be approached from the perspective of human enhancement for retaining relevance and unlocking ever richer experiences, allegedly leading to the emergence of new, superior beings that transcend biological limitations. The possibility of enhancement immediately transforms into an imperative – even an obligation – as any problems currently or potentially faced by humans are seen as solvable through enhancement. Supporters of such transhumanist visions can be seen as promoting an optimised and efficiency-oriented version of the human, with age, character traits, physical and mental conditions manipulable, and any perceived shortcomings 'healable', meaning that all humans can be remodelled under a single definition of 'normality'. This attempt to allegedly take human evolution into their own hands not only ignores choice – it frames any delay as morally wrong, that is, precluding the achievement of the projected future good. Even if augmentation leads to the dissolution of the 'I' and human subjectivity, this is still seen as an improvement, because a 'better' form of existence can only lead to a 'better' conscience. Even the thought of choice – a crossroads – hence becomes *unthinkable*.

Performance maximisation can happen not only in the extension of human capacities but also in the domain of warfare, by way of increasing the efficiency and removing friction, such as human decision-making. Crucially, the distance of warfare and the possibility of keeping one's (human) forces out of direct engagement might even reduce the burden of the decision to resort to war, while responsibility would be diluted. Moreover, the proliferation of such technology would mean autonomous warfare by default, while an ensuing arms race would bring standards down to the lowest common denominator. However, for advocates, the growing accuracy of autonomous weapons is seen to make their use a moral necessity, while for others, such a necessity would simply be driven by a fear of competitor states taking advantage. In either case, the development of autonomous weapons would be removed from contestation and lead towards projected technological progress.

The drive towards efficient destruction is also visible in projects of space exploration and colonisation. Its necessity is framed by proponents in terms of enabling a good life and improving the human condition. The space thus becomes a frontier at which rights, obligations, and concerns

end, while destruction and exploitation become the norm. The preceding human self-entitlement is characterised by, for example, the alleged moral necessity of terraforming and expanding the life-space (essentially, *Lebensraum*) of the human species. Space colonialism is also framed in existential terms, such as avoiding human extinction. Part and parcel, though, is the whitewashing of colonial history on Earth as a benevolent march towards progress. Meanwhile, the exploitation of celestial bodies and stripping them of resources (supported by proposals for property rights regimes based on efficient exploitation) is presented as a norm, underscoring the anthropocentric sense of unmatched superiority and entitlement. Failure to consider the possibility of a crossroads becomes a betrayal of humanity itself, as progress and efficiency maximisation is taken to its natural exploitative and destructive conclusion.

In the end, frictionlessness is the domain of liquid evil: As we imperceptibly slide along the path towards actual or alleged technological progress, alternatives and dilemmas fade from view, precluding even a consideration of there being a crossroads. However, when alternative paths are not considered (or even remain unnoticed), anything that does not fit within the straight path of progress becomes adiaphora – things removed from the horizon of morality. Hence, blindness to alternatives is moral blindness. Nevertheless, as the straight frictionless path is typically reinforced by notions of optimised performance of efficient value extraction, efforts to overcome moral blindness necessitate going beyond these staples of today's heavily technologised society. Liquid evil, as moral blindness put into practice, is thus best seen as the default standpoint of today.

Index