AI and the illusion of control

Nowotny, Helga [1]

[1] *European Research Council*

DOI 10.5281/zenodo.13588582

TO CITE

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582

PUBLICATION DATE

04/07/2024

ABSTRACT

The paper explores the impact of the dazzling performance of Generative AI on the sense of being in control and the Illusion that may come with it. Control of technology as a hallmark of modernity was accompanied by hubris and often the illusion of being in control. Now our anthropomorphic tendencies to attribute human-like features to AI exposes human vulnerability anew. Control of technology cannot be restricted to its mere technical functioning and has successively expanded since industrialization. After first guaranteeing the safety and health of workers, at least in highly developed countries, gradually a 'safety culture' emerged. We expect control of technology to include impact on health and safety conditions and the protection of the natural environment. The illusion of control sets in when CEOs of major international corporations deny the necessity to extend control of AI (to foreseeable, and even unforeseeable consequences that it has on cognitive and mental abilities). The paper then retraces the history of outsourcing knowledge operations, from the invention of writing to the printing press and mass media, raising the question of agency and responsibility. It concludes by asking whether our ancestors who believed that they shared an immanent cosmic order with 'meta-persons' lived in an illusion and what it might mean when we must learn to live together with the digital Others.

Acknowledgements

This paper was written during a 1-month residence at the Paris Institute for Advanced Study under the "Paris IAS Ideas" program *. It is a greatly expanded and restructured synthesis. Some of the ideas presented here overlap with ideas in the following publications:

Nowotny, H. (2022). Digital Humanism: Navigating the Tensions Ahead. In: Werthner, H., Prem, E., Lee, E.A., Ghezzi, C., (eds.) *Perspectives on Digital Humanism*. (pp. 317-322). Springer.

Nowotny, H. (2024a). The Re-Enchanted Universe of AI: the Place for Human Agency. In: Ghezzi, C, Kramer, J., Nida-Rümelin, J., Nuseibeh, B., Prem, E.,

Stenger, A., Werthner H., (eds.) *Introduction to Digital Humanism. A Textbook*. (pp.197-209). Springer.

Nowotny, H. (2024b). The Illusion of Control: Living with the digital Others. In: van der Leeuw, S.E., Galaz, V., Vasbinder. J.W., (eds.) *Global Perspectives, special issue "Illusion of Control"*.

Nowotny, H., Van Hoyweghen, I., & Vandewalle, J. (2023). *AI as an agent of change, KVAB Thinker's report 2023*. KVAB Standpunt 85 b.

1. Generative AI and the illusion of control

The amazing feats of LLMs, Large Language Models, to generate texts and images that have been trained with data from the web as well as being synthetically produced, have surprised even experts. New opportunities to render our lives and our future even better and brighter were celebrated. Likewise, numerous concerns were raised ranging from threats that targeted disinformation will undermine liberal democracies to the challenges of likely job losses, this time affecting mainly professionals and artists. By unleashing ChatGPT, OpenAI, financed by Microsoft, many organizations experimented on millions of users without asking for anyone’s consent. Huge investments and fierce competition between the tech giants followed, accompanied by promises and rising expectations. Arguably, the enthusiasm generated this time exceeded previous ones.

Underlying even the brightest prospects of the benefits and opportunities that Generative AI undoubtedly harbors, one persisting question surfaces repeatedly: are we still in control of the digital machines we produce? When experts working on the most recent digital developments admit that they do not fully understand how the output designed and programmed by them actually is generated; when the specter of 'existential risk' is gleefully raised by the very same corporations that claim to be at the forefront of reaching the ultimate goal of Artificial General Intelligence (AGI); and when governments in liberal democracies are scrambling to erect safeguarding barriers for

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

their proclaimed 'technological' or 'digital' sovereignty -- how are ordinary citizens expected to believe that the new technology is still under control?

Concerns about AI are nothing new. All of us have become accustomed to, and continue to worry about, the darker sides of the Internet, where cybercrime and child abuse thrive that seem to be beyond the reach of governments and civil society. Fears about further loss of privacy and surveillance, described in Zuboff's 'Surveillance capitalism', have hardly lessened, but we continue to hand over ever more data in return for the convenience of services or because we practically have no choice (Zuboff, 2019). During the COVID-19 pandemic, distrust in their governments has prevented citizens from using the tracking apps that were available with sufficient assurances on the technical and legal side to guarantee their privacy. We continue to live with the alleged impact of social media on furthering the fragmentation and polarization in our societies. They are accused of facilitating hate speech and driving ever more wedges into already divided communities. Bias, present in every one of us and ubiquitous in society, and the discrimination that might result from it, are easily perpetuated, and reinforced through the indiscriminate proliferation and use of data. Although there is no lack of appeals for an ethical, responsible, fair, and transparent AI, governments are struggling to hold large corporations accountable for content and the social harm it causes. Regulation is still in its infancy and although the EU is internationally at the forefront with an impressive legislative package, its full implementation is still to come.

So, what is new about ChatGPT and the unprecedented adoption rate that has led to more than one million users in one week? We do not know whether the hype that greeted it will be followed by a downturn, as has happened many times before following Gardener's hype curve. What speaks against it are the massive investments that keep pouring in, betting on 'too big to fail', and the technological progress that continues to advance at a quick pace, in the direction projected by those who invest. Meanwhile, governments undertake efforts to reach an international agreement on minimum safety standards, like the recent setting up of the AI Safety Institute Consortium led by the UK. Collective imaginaries about the future impact of AI on people's lives, nurtured through sci-fi, receive a public confirmation when the fears about 'existential risk' are openly voiced by leading CEOs from Silicon Valley, mixed with deliberately ambiguous messages that only they are the ones to contain these risks.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

The fear of losing control is a deep-seated human worry. Losing control of machines that are more powerful, clever, and intelligent than humans is a recurrent trope in film, literature, and the arts. It is now re-activated, for everyone to see with their own eyes when prompting a Generative AI on their screen. The effect is a bedazzling performance, followed by a relieving laugh when it makes a stupid mistake. Yet, it is difficult to escape the uncanny feeling that results from not knowing how it works and who is in charge. Losing out to machines was first demonstrated to great effect in 1997 when DeepBlue, an IBM supercomputer, beat the world's chess master Garry Kasparov.

This was followed by the even more spectacular victory of AlphaGo over the world's best Go player, Lee Sodol, in 2016. Never mind that the machine was masterminded by a fabulous team of AI researchers and developers at DeepMind, who were physically located one floor above the public arena in which the contest was staged for the entire world to watch. In the 18th century the Mechanical Turk, a chess-playing machine that was constructed to deceive the public, was operated by a human who succeeded in defeating its strongest opponents. In the 21st century such deception is no longer necessary. The machine genuinely performs better even if humans are still involved in programming, monitoring, and intervening. The complex game of Go had been deliberately chosen to demonstrate the scientific-technical and financial clout of Big Tech behind the stupendous, and real, advances of AI. When Lee Sedol retired prematurely in 2019, the reason he gave was that the machine was an entity that could not be defeated (Labatut, 2023).

Given the deep-seated anthropomorphic tendencies that lead us to attribute human-like qualities to things and phenomena with which we interact, it is not surprising that many users are lured by the machine into believing that they are communicating with a human being, especially when the AI has deliberately been designed to make users believe so. A telling incident took place when bringing the development of Generative AI into public awareness. When Blake Lemoine, a software engineer at Google, shortly after a limited version of LaMDA -- a generative AI specializing in dialogue -- had been opened to the public in August 2022, told the Washington Post that he became convinced that it is 'sentient', he caused a stir. Google was quick to dismiss him on grounds of having violated the company's confidence rules. His professional colleagues were more outspoken but equally swift in declaring that he was wrong. They were

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

unanimous in proclaiming that no AI had attained anything like being 'sentient' yet, let alone some form of 'consciousness'. The public was reassured that Artificial General Intelligence (AGI), although high on the research and innovation agenda, was far in the future and so was 'singularity', the point in time when machines would overtake human cognitive capabilities. Yet, the race between Google, Microsoft, Apple and a growing number of start-ups staffed by their former employees continued to take the convergence of ML and LLM a decisive step forward and to release a new generation of generative AI models to the public.

The incident of sacking Lemoine and the reasons behind it were soon overtaken by the excitement caused by the release of ChatGPT. It rapidly turned mainstream, and the capabilities of Generative AI continue to expand. In addition to writing almost any text, they produce images following the prompts of the user or compose music in whatever style wanted. They also can create DeepFakes by digitally altering the face or body of a person or by taking their voice and making them speak sentences they never spoke or would ever speak. The disturbing fact remains that practices of deep faking are not only becoming more widespread, but also more insidious. They play with the power that images and voices have on the human imagination and the trust we have put so far into what we deem to be authentic, believing that it is indeed the person we know speaking with the voice we recognize or trusting that the face we know is part of the body we are shown.

ChatGPT was only the beginning. Google reacted by releasing its version, Bard, a dialogical generative AI that promptly upped the stakes, and other products continue to invade the market. More foreseeable and unforeseeable consequences are following as the rapid diffusion and adoption of digital products inundate the market. DeepMind plans a new generation of 'P.A.s', Personalized Assistants, designed to guide you in your decisions and how to lead your life. Behind the excitement and bafflement, anxieties concerning the most fundamental questions about the relationship between humans and the technologies created by them return with insistent urgency: who is in control? Can humans keep control over the machines they designed or is this an illusion, only to be matched by another illusion, namely that from now on the bots will be in control? Or is it the humans behind the machines who control everything?

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

The incident about the former software engineer at Google is a tale about the illusion of not being in control. An illusion is a cognitive state which is out of sync with reality. If we are in thrall with an illusion, we are convinced that what we see, hear, and believe in accords with reality. Only after an imploding clash with reality does the beholder realize that it has been an illusion. In the case of Lemoine, his professional peers declared so on his behalf. Obviously, this raises questions about the role of scientific and professional expertise, underlining the necessity of a commonly accepted framework of reference. Once scientific authority is no longer accepted as the arbiter of a shared and commonly accepted reality, we risk falling into a state of anomie, consisting of 'personalized realities' that obliterate common ground.

The fear of losing control is not primarily about the 'existential risk' of Artificial General Intelligence wiping out human agency and creativity sometimes in the future, as conjured by some hi-tech firms. Anything can happen in the far away future, but nobody knows and can tell us much more than the certainty that the fuel of our sun, hydrogen, will run out and the sun will begin to die some 5 billion years from now. The threat of an AI-linked 'existential risk' merely diverts us from attending to urgent problems in the present, and perhaps deliberately so. Rather, we should ask whether we over-delegate when increasingly installing command and control systems in autonomous weapon systems which are on track to become a decisive feature of war in the not-too-distant future. The process is accelerated by the ongoing wars and the shift towards defense as a high priority for government spending.

The road towards more automation, including the automation of decision-making and management is opened further by AI systems designed and promoted as being more 'intelligent' and thus more efficient and reliable than humans, inserting them into the daily lives of business, public administration, and citizens in ways that replace humans. Instead of complementing their capabilities in clever ways that will generate new tasks and thus new jobs, machine-based expertise is likely to take over. This will lead, at least initially, to an increase in the productivity of lower-skilled workers while it will diminish the value of expertise now held by the middle class and professionals (Autor, 2023). Given the enormous concentration of economic power in a handful of large international corporations and the difficulties that governments experience in regulating

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

AI, the question then becomes whether we are handing over control of our lives to Big Tech.

One peculiar feature of the illusion of control is its blind spot. Those who are in its grip fail to notice their condition until a clash with reality forces them to do so. The history of humanity is full of stories of human hubris, of excessive self-confidence, originally in defiance of the gods and in modern times in defiance of the unintended consequences of human action. Technology makes it all the easier as it provides an intermediary shield, raising the question of whether the digital technologies that invade our lives will enwrap us even more in the illusion of being in control. Or will they have the contrary effect -- that they and the powers behind them will control us?

The sense of being in control has been the hallmark of modernity, intricately linked to technological progress that, for the first time, enabled control over much of people's lives and their circumstances. Technological progress was instrumental in transforming the notion of the future into an open horizon. Planning for future events was rendered possible through a combination of probabilistic calculations and the experience that validated them. Together with the faith in Progress, it formed the backbone for much of the hubris of modernity that we now recognize as what it was: an unquestioned belief of being in control that legitimized the exploitation of the natural environment and other human beings that were stigmatized as inferior. The modern sense of control was underpinned and boosted by the Narrative of Progress with its promises of unlimited improvement of the human condition (Nowotny, 2021a). It rested on the confidence that the unpredictability of the future can be tamed, as retraced in Ian Hacking's classic account of the emergence of probability and the impact it had (Hacking, 1990).

Probabilistic calculations became the basis for the management of uncertainty in modern societies, an efficient way of coping and guiding future experiences. According to Elena Esposito probabilistic-oriented decision-making can claim that rational behavior is guaranteed as the claim holds up also in retrospect. If the prediction turns out to have been wrong, the decision was still correct as it was the rational thing to do. Thus, the future became de-problematized by creating a sense of 'statistical certainty', enshrined in the system of insurance, with probabilistic techniques promising to control the negative aspects and to enable rational action. Yet, like so much else, the belief in being able to

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

neutralize risks has been challenged and shattered, most recently through the financial crises, extreme weather events and growing geopolitical tensions, which generate or converge with other crises in turn. It became clear that statistical tools do not give access to the future. They are merely guidelines for decision-making, knowing that one may come to regret it in the future (Esposito, 2024).

Will digital technology and the power of predictive algorithms, based on an enormous amount of data with growing accessibility, boosted by unprecedented computational power and sophisticated algorithms, lure us again into believing that we are in control of the present and of the future? Or will the horrific fall-out of the hubris of modernity, that culminated in the over-confidence of the State and resulted in some of the most horrific consequences of seemingly well-intentioned actions, serve as a warning? (Scott, 1998) And what about those CEOs of the most powerful international corporations who boast that they have unlimitable power? We all know their names and faces. Shamelessly, they claim that their personal eccentric visions -- to live in a metaverse, go to Mars, become immortal or to implant brain chips for enhancement - are identical to the aspirations and needs of humanity. Will they ever be able to see through their illusion of being in control?

Illusion, as stated above, is a tricky concept. Those who are in its grip fail to notice their condition until a harsh confrontation with reality occurs which makes it burst like a soap bubble. This holds for individuals as much as for communities and collectivities. For a society, it may take longer to face up to it, often caused by a catastrophic or traumatic event. Current concerns about AI turn out to be only the last trigger of an uneasiness that has been in an upswell for some time. The increase in complexity is gathering speed, with climate change and environmental degradation competing with the interruption of supply chains due to armed conflicts or fierce competition about indispensable minerals or component parts. Complex systems defy the linear thinking that dominates modernity when the dynamics of linkages between networks give rise to emergent properties or phenomena that are impossible to predict. Complex processes may culminate in tipping points that can lead to a phase transition and the possible collapse of a system.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

Every age tends to convince itself that 'this time is different'. Historians of technology do not cease to argue against the largely unquestioned bias in favor of the latest technologies and innovations. Although previous generations also had to cope with the anxieties that resulted from the turbulence created by the processes of industrialization and their experience of an incessant acceleration of technological and social change, the digital present in which we live today nevertheless carries an excessive burden of informational overload. It gives rise, and exacerbates, the feeling of being emotionally overwhelmed. One of the reasons is that more than in our dealing with previous technologies, interactions with AI involve our cognitive abilities and perceptions of the world. They play with our emotions and alter our relations to others and the self, blurring the boundaries between 'us' and 'them', the digital Others.

The result is a profound ambivalence. We are fascinated and bedazzled by the stunning performance of AI, yet it also unleashes the anxiety that it may diminish our autonomy, impoverish our identity and sense of who we are. Some users become convinced that 'AI knows me better than I know myself', even if they are aware that an AI does not 'know' or 'understand' anything. This is not surprising given our anthropomorphic tendencies and the deliberate design of algorithms to make us believe that we communicate with another human being, thus engaging with the machine not only on a cognitive level but also on an emotional level. As I argue in my book 'In AI We Trust' a paradox lies at the heart of our trust: we leverage AI to increase our control over the future and uncertainty, while at the same time the performativity of AI, the power it has to make us act in the ways it predicts, reduces our agency over the future (Nowotny, 2021a). This happens whenever we forget that predictive algorithms base their predictions on an extrapolation from data coming from the past, while we transfer, and attribute, agency to them (Nowotny, 2021a).

In his latest book, Algorithmic Anxieties, Anthony Elliott (2024) analyzes the negative fall-out of the continuous daily struggle of people to combine life on- and offline and how to cope with the resulting information overload. The picture that emerges is grim. From the life of 'automated' workers for Amazon to Netflix's nihilism, from the algorithmic violence in computer games to the metastasis of Metaverse, he diagnoses processes of self-dislocation. As the autonomy of the self-shrinks, the heavy toll on people's emotional lives creates culturally pervasive fears and ambient anxieties.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

Predictive algorithms are heralded as giving us control over the future with mathematical exactitude, while in practice they unleash disabling anxieties and fears due to the harmful effects of social media, anxiety about losing one's job or becoming subject to technological manipulation and surveillance. What is intended to reduce complexity by delegating more tasks to algorithms turns out to generate anxieties unknown in previous eras through the sheer quantification and pervasiveness of outsourcing to an AI:

"...decisions are outsourced to smart machines daily and the demands of thinking about decisions vanish, only occasionally requiring further fleeting attention or consent at the click of a mouse. Problems commanding attention may continually arise but disappear once outsourced to automatic calculating machines, only to be replaced by the next cycle of decision-making outsourcing. In this complex entanglement of humans and machines, the limited capacity of individuals to exercise autonomy becomes increasingly frail as algorithms iteratively learn, compose, generate and authorize actions based on the informational attributes of people, data and other algorithms" (Elliott, 2024, p.218).

The quest for identity and finding meaning in a world in which technology and geopolitical events, now strongly coupled with the urgent exigencies of climate change and how to prevent humanity from reaching the tipping point of a possible collapse, continues. We may have shed the illusion of having everything under control that dominated modernity, but we may slide into another kind of illusion, this time generated by AI and the kind of agency we transfer and attribute to it. As often with AI, the effect is divisive. Some feel completely powerless, marginalized, and beset by anxieties and fears that make it impossible for them to imagine that they still have a voice and agency. As long as they are in the grip of fear, they are deprived of being able even to imagine that they might regain control. Others boast of being possessed by feeling omnipotent. They are overconfident and under the illusion of having unlimitable power. The majority feels confused and overwhelmed by the acceleration brought about by technological change and the demands for adjusting to a system in which human contact is successively replaced by machines feigning to be humans. They are caught in an ambivalent limbo, oscillating between multiple worries and desperate for hope and a more positive outlook.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

The enormous concentration of economic power in a capitalistic system that rewards greed and legitimizes growing inequalities is hardly favorable to finding common ground and a shared vision of the future. Our societies have become polarized and fragmented, which is exacerbated by the unrestricted circulation of misinformation, hate speech and the further erosion of trust by DeepFakes. In principle, AI enables us to create digital twins of everything that exists - a digital mirror society and a digital mirror world. At the same time, it holds up a mirror to us. If we wish and look carefully, we can see through the illusion of being in control and the illusion of having lost control. We can see what makes us human. What follows from this insight is left up to us.

2. Anthropomorphic illusions

The recent encounters with generative AI have exposed a human vulnerability to anthropomorphism, to seeing the systems in which it is embedded and its components, as having features and behavior more human-like than they are. By becoming extremely adept in mimicking human language and other cognitive abilities, including scientific and artistic creativity, the line between the 'natural' tendency to anthropomorphize as expressed in the language we use in our dealings with technology, and the belief that the technological artifact is indeed an entity that 'knows', 'understands' and 'thinks', becomes ever thinner. The un-reflected use of such words which are relatively harmless if they refer to familiar technologies that we have incorporated into our world and hence under control, can transform into a dangerously compelling illusion of being in the presence of a thinking creature like us (Shanahan, 2022). From a philosopher's perspective, Daniel Dennett has analyzed in detail what he calls 'the intentional stance', the human propensity to treat others as 'intentional systems' in daily life. We attribute to things and phenomena beliefs, desires and rational intentions that pertain to humans which allows us to predict their behavior more readily (Dennett, 1987).

The impact that our anthropomorphic tendencies exerted in our communicative behavior especially with Generative AI moves us closer to the moment that Alan Turing defined as the arrival of genuine Artificial Intelligence. The test that carries his name implies that then it is impossible to distinguish whether one is speaking to a real person or

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

being able to recognize the image of a real person compared to a composite artificial face. However, the rapid advances in facial recognition and language processing have led to dispute Turing's definition and to consider it to be at best a sidetrack. At worst, it keeps moving the goalposts in a narrowly defined competitive game between 'natural' and 'artificial' intelligence. Every advance by an AI is measured against human performance and hailed as 'victory' of the machine over human intelligence. Failure of AI systems to perform in real-world conditions is pushed aside and the race towards machines that are 'better than humans' continues. As Luc Steels argues, the Turing Test is utterly misleading when it is used as the main reference for tracking progress in AI as a scientific field. It is based on deception, as it merely pretends to exhibit intelligent behavior to pass the test. It adopts and over-interprets the 'intentional stance' which is not a reliable basis for judging whether an artificial system is 'intelligent' or merely pretends to be. In short, it has little to do with human intelligence.

If we want to measure progress in the field of AI, we should remind ourselves that the goal of AI research is to contribute to the big scientific questions about the functioning of the mind and its relationship to the chemical and physiological processes that underlie it. What is touted today by Big Tech and its evangelists as the pursuit of AGI, Artificial General Intelligence, has little to do with these scientific questions which have been completely overshadowed by the commercial goals to come up with digital technologies that generate huge profits in return for the investment of billions of dollars that are now poured into AI development.

The growing discrepancy between the scientific pursuit of understanding and the obsession with highly publicized benchmarks of AI performance has turned the Turing test into 'Turing's curse'. It precludes spending resources on fundamental research questions on other approaches than those dominated by the race towards higher performance on narrowly defined goals by the large tech companies. We are asked to collude in devaluating human expertise, real-world experience, and knowledge in favor of what AI companies sell us instead while everything we know about the construction and functioning of these artificial systems tells us that they are very different from human understanding and our mental and cognitive capabilities. Being led to believe that the bot is a human agent may therefore be more of a sign of human gullibility than

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

a testimony of the presumed 'intelligence' of the machine which, in any case, is not the same as human intelligence (Steels, 2023).

Despite the many caveats reminding us that generative AI is only based on mathematical functions, trained with an enormous number of tokens that consist of texts available on the Internet or are synthetically generated, of images and visuals that inundate the digital world and sounds that are equally abundant in a great variety, our anthropomorphic tendencies have a profound effect on how we relate to them. They model the statistical distribution of these tokens from the vast public corpus of human-generated texts that tell us what words are most likely to follow the sequence of words in the question we ask. And yet, their performance continues to amaze us with their speed and versatility, being able to switch tone and genre in the answers they give to our prompts. We tend to be also more lenient in tolerating errors when committed by a machine compared to errors by humans when we believe it to be more 'objective' -- another bewildering inconsistency in how we learn to live with the digital Others that are so clever in imitating and pretending to be like us. We have also learned that their tendency to 'hallucinate' and make up things that do not exist or are blatantly false, increases with the overall length of the texts they are being fed which is due to the intricacies of the probabilities between the connecting links the Generative AI establishes.

The deep-seated propensity to anthropomorphize a technology by treating it as if it was human is a confusion about agency, identities, and relationships. From experience, we know that neither is unambiguous. They all may change. Our perception and the knowledge of the world we share with others and what we assume to be mutual understanding, are continuously challenged and in need of being reconfirmed. But we continue to collude with the machine, despite knowing that doing so is not in our interest and may even harm us. This happens when we hand over data to Big Tech about the most intimate aspects of our lives in return for their convenient services. We are cognizant that algorithms have been designed to boost engagement and yet remain in an addictive relationship. All addicts live in the illusion that they can exit at will. How much control do we retain?

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

The dilemmas that arise together with many unanswered questions are illustrated by the increasing use of chatbots for emotional needs and mental problems. Therapeutic AI has found a growing market with profound effects, shaping what Anthony Elliott calls 'algorithmic intimacy' (Elliott, 2023). The proliferation of mental wellness and therapy apps comes with many attractive features. They are free or cheap and easy to use. Like other digital services, they are convenient and readily available. Instead of being on the waiting list for a human therapist, the patient, client, or customer, gets a quick response whenever wanted or needed and can even 'clone' an app to fit his or her ideal or preferences of a therapist. As the therapeutic app is programmed to address the user like a human, it evokes feelings like 'you are the only one to help me, the only one who listens to me...' A therapeutic app is also programmed not to be judgmental. This may matter for ethnic minorities who otherwise would shy away but smacks of being the cheap version of a needed service offered to those who lack the means to consult a professional therapist.

The downsides are considerable. Hopefully, the therapeutic app-induced suicide that occurred in Belgium in 2023 will remain an exception, as the bots have since been updated to avoid approving suicidal questions. Yet, self-deception remains and the illusion of relating to a human who understandingly tolerates everything may delay or severely impair the ability of the user to connect to a real person in the outside world. Communication continues to be disembodied, devoid of human experience that can be shared, like the loss of a loved one or problems in raising a child, even if the voice and tone suggest otherwise. In many countries a therapeutic app is currently completely unregulated and where certification as a medical device is foreseen, the regulatory requirements can easily be skirted by declaring them as wellness applications (Robb, 2024).

We can see the efficiency gains in clinical hours for public health services that are already overstretched, seriously understaffed, and underfunded in the mental health domain. Yet, the concise analytical diagnosis by Anthony Elliott referred to above, gains only in relevance and urgency. We are outsourcing problems that demand our attention to automatic calculating machines, only to be replaced by the next cycle of outsourcing as the problems do not disappear. They are merely rendered invisible or relegated to the already limited capacity of individuals to exercise their autonomy.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

The power of technology has permitted us to do things that otherwise would be unthinkable. It has enabled the human species to transcend some of its biological limitations and the temptations of further enhancement know no limits. At the same time, it has revealed our biological limitations and our deep and intricate interconnectedness with other living organisms and the natural world around and within us. The flip side is that we hand over to technology the power that it then exerts over us. It forces us to behave in certain ways, sapping our sense of already curtailed autonomy. We offer our most intimate thoughts and emotions in exchange for a calculated dose of faked empathy and words of consolation that imitate those uttered in the real world. We may be dimly aware that we are turning our inner selves into informational attributes and tokens that will empower the chatbot to improve its performance in the next cycle of outsourcing. Yet, we continue in the illusion that outsourcing our problems to calculating machines will make them disappear, only to discover that they simply have been rendered invisible, while further diminishing our already frail autonomy and sense of self.

For far too long, we have insisted that technologies are neutral and more 'objective' than humans, as we tend to associate them with the techno-sciences. Yet, the goals that have been designed into their functions are programmed by humans, and they merely follow the instructions they have been given. Whether technology is used in ways that are beneficial or harmful, whether they liberate or suppress other human beings, is never about technology alone. Human agents have transferred parts of their agency to the machines that carry out functions to attain specified goals. Human agents have interests and intentions, be it profit, destabilization, destruction, or the advancement of scientific understanding and working for the common good.

3. The control of technology

Control is inherent to every technology as otherwise it will not function. Control over the design, construction, manufacture, and operation presupposes control over the component parts, the processes involved and how a technology is embedded in a larger system. As a smooth and efficient functioning can never be taken for granted, control

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

implies foreseeing and preventing whatever can go wrong. Errors are always possible, and accidents are likely to happen. The fault may lie in the design or lack of proper maintenance and repair. It may be due to the deficiencies of components or failure in the ways they interact. The functioning must take into account the varying conditions under which it is expected to work, be it temperature, humidity, roughness of surface or the more intricate features of the smaller or larger size of the device or system and their interaction.

Accident prevention has become an indispensable feature of every technology, from the safety of air traffic to the exploration of space, from wearing a helmet when on a bicycle to drug control of workers in factories. The interfaces between technology and humans remain crucial, multiple, and often unpredictable. Large technological systems are known for their inherent complexity which manifests itself when a catastrophe occurs, an event with low probability, but great damage. The nuclear accident at Three Miles Island, and more recently at Fukushima triggered extensive analysis of critical infrastructures and high-risk systems, pointing to the limitations of conventional engineering approaches, for instance when complex systems with too tight coupling are involved (Perrow, 1984).

The control of technology includes a whole gamut of safety and other protective and preventive features. The problem is that we can never be sure whether these controls will be sufficient to ward off harm and prevent failure or developments going in an undesirable direction. The processes that underly creeping errors may remain invisible for a long time before collapse. As the effectiveness and the affordances of a technology increase, control must expand as well. Beyond the immediate technical functioning, it needs to include what can possibly go wrong. This encompasses the foreseeable consequences, but what is foreseeable and unforeseeable is often disputable. Thus, the challenge of extending control of technology beyond its technical functioning must adapt in line with the dynamics of change it brings about. We now expect that control implies the inclusion and the management of the impact of a technology -- whatever this may mean.

Seen from a historical perspective the road from controlling the technical functioning of a machine in the sense of making sure 'it works', to the control over the impact it has,

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

first and foremost on the workers who operate it and then extending far beyond to include its wider impact, has been a long one. Industrialization brought dismal working conditions for workers with their safety and health continuously at risk. Only after long periods of labor unrest and many conflicts did the labor movement succeed in anchoring their demands in the legislation which led to an improvement of their living and working conditions, ensuring that the profit of factory owners would not come at the expense of workers' health and safety. Eventually, the Welfare State became established in Europe. The concept of safety is now accepted as a central feature of the extension of control over an increasing number of harmful consequences and insurance covering accidents, health, and old age pensions became obligatory based on the principle of solidarity (Ewald, 1986).

By now, at least in highly industrialized countries, the increase of safety features in products and production processes, regulations and standards has become the norm and continues to expand. They extend beyond manufacturing and pervade market-approved consumption and use. Backed by legislation and bureaucracy, certification of products and safety measures have become mandatory, enshrined in obligatory checklists, safety drills, extra protection gear and risk-reducing infrastructures. Whether it relates to the safety of cars and traffic control, keeping medication out of the reach of children or safeguarding nuclear power plants -- control over industrial products and processes to guarantee their safety has become paramount. The approbation of new drugs and medical treatments takes years of randomized clinical trials to ensure that harm is avoided, and side effects will be monitored and minimized.

Thus, the control of technology includes multiple, nested layers and continues to suffuse our technological civilization. Control is expected to increase productivity and efficiency as well as to guarantee safety, it should shield us from negative consequences for our health and increasingly cover biodiversity and other features of the natural environment to be protected. This implies that maintenance and repair, sustainable and frugal use of natural resources, recycling, and disposal of waste, are now considered indispensable for the protection of the natural resources and services with the ambitious goal of a circular economy on the horizon. As all human action intervenes and affects the natural environment, this has repercussions on what we extract and how, working towards greater sustainability in the use of energy, food, water, and the air we breathe.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

Anthropogenic impact must be monitored, managed, and controlled. In the worldwide competition to be at the frontier of the latest development of AI-based systems, including chips and their mineral components, the consideration of controlling the environmental harm linked to digital technologies, has only begun. The environmental costs of digital infrastructures, manifest in the rising demand for energy, water, and land needed for the operation of data centers, are huge.

Control of technology has its dark sides. It exerts power by installing constraints on things and processes, prescribing how to interact with them. This can easily transform into control over other human beings and their rights. The widespread fear of digital surveillance and its abuse by governments and corporations is a forceful reminder of the power of control exerted through technology. It can be visible like surveillance cameras in public places or more surreptitious by following our digital traces, legitimized as being 'only' for our safety. DeepFakes and cyberattacks are deliberately launched to cause destabilization, whether they come from ordinary criminals, from other States or from state-sponsored hacker groups. They leave citizens feeling that they are no longer protected but exposed to unknown forces who may be anywhere and everywhere. They erode trust in governments, manipulate elections and thus pose a serious threat to liberal democracies. Intriguingly, some similarities exist when delegating control of technology and delegating control in liberal democracies. Delegation consists in renouncing direct control in order to gain more general control, or to retain control in other forms. Daniel Innerarty defends the political value of delegation, while emphasizing the necessity to achieve a balance between control and delegation; supervision and confidence (Innerarity, 2022).

It is difficult to pinpoint the exact locus of control. Control has been installed by humans and the technological devices are operated, owned, and commanded by humans, following their instructions and goals. Thus, human agency is ubiquitous, but difficult to pin down when it comes to issues of responsibility and accountability. The agents of control are the large corporations with their enormous concentration of economic power that easily translates into political power. They assume a political, even a military role, when Elon Musk, for instance, decided how, when, and where to deploy his Starlink following the attack by Russia on Ukraine. Agents of control can be States, represented by their institutions. They are challenged to increase the security measures against other

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

states and the cyberattacks sponsored by some of them. This quickly transforms into a dense fabric of 'protective' security measures within the national territory, reinforcing a 'digital sovereignty' that is threatened as it does not respect physical national boundaries. In turn, this has huge, constraining impacts on the freedom of those who live within them. We are witnessing an increasingly worrisome shift in the balance between security and freedom to the disadvantage of the latter. It comes with the imperative to increase military spending and to shift national budgets from funding research to defense. It creeps into the lives and work of researchers who face increasing restrictions of international collaboration and having to submit to a new host of regulations issued in the name of security.

Where there is control, the illusion of being in control is never far away. Humans are always at risk of being overwhelmed by their senses and biases; by the wish to believe what they want to believe, even when contrary facts stare into their face. The causes for such illusions are many. They range from the overconfidence which disproportionately affects political and economic leaders, to the gullibility reserved for simpler minds. Illusions are nurtured by the cognitive biases we all have, but individual biases are reinforced by social and economic circumstances, by information and misinformation, and by the institutions and cultures into which we are socialized. Illusions of being in control are put to the extreme test in the event of war when both sides are convinced that each will win, having the latest advances in military technology on their side.

We may soon reach the -- provisionally ultimate -- illusion of control with autonomous weapon systems. Designed to be ultra-fast AI technologies they are programmed to carry out 'precision strikes' which have already taken their place in warfare with the mass production of cheap drones equipped with FPV, first-person vision. Other, more powerful autonomous weapon systems with greater reach and impact, are waiting to be deployed (The Economist, 2023; Mhalla, 2024). It is difficult, if not impossible, to call autonomous weapons back once they have been launched, as they will carry out what they have been instructed to do. Given their hyper-speed, there is no time left for human deliberation, let alone for rendering a human-made decision reversible. The control of this technology has been handed over to the machine and the deliberately designed autonomy of the system. Whoever deploys it as a first strike or as programmed

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582 2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

fast retaliation, is in the grip of the illusion of being in control, upheld by the belief of controlling a technology that is more powerful than that of the opponent.

4. Technologies as agents of change: the outsourcing of communication

The historical growth of new human knowledge can be interpreted as a sequence of major transitions in externalizing knowledge operations: collecting and processing of information; applying the knowledge gained in other contexts; storage, dissemination, communication and repurposing of information and knowledge; generating new knowledge through recombination. Encoded in a new technological medium, knowledge operations extend what becomes possible, visible, feasible and understandable. In turn, outsourcing has major impacts on the society which invents, adopts, and expands it. One of the most efficient knowledge operations is communication, the exchange and spread of ideas which leads to new ideas, the adoption and adaptation of concepts and practical instructions of how to 'do'. Paul Watzlawick, the communication theorist, famously declared 'One cannot not communicate'. Our huge advantage over other communicating animals is the evolution of language. We can use it to communicate in analog form (about an object) and digitally (logical, symbolic, and statistical connections). We communicate verbally and through body language. We transfer and exchange information about ourselves, others, and the world. This can be ideas, practices, and knowledge at various levels of abstraction and complexity.

Communication is a social practice that occurs in social settings. They can be symmetrical, at eye level and equal footing, or emphasize social hierarchies. Humans have developed elaborate codes that pervade all aspects of social life to distinguish themselves from each other. Communication is at the root of the social organization of societies that has grown more complex over time. It has stimulated and boosted the enormous growth of human knowledge because of the selective accumulation of information. New ideas, knowledge or practices are combined, and recombined in novel ways. The content passes through selective social and cultural filters in the processes of being transferred and exchanged, following the norms and values that define which kind

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

of exchange and content are culturally and socially valued and recognized. Societies rely on an explicit or implicit knowledge hierarchy whose layers have been described as moving upwards from data to information, followed by knowledge and featuring wisdom at the top. In my book I have dedicated a chapter to wisdom needed in the future.

The technologies embedded in these knowledge hierarchies function to control which kind of knowledge and information circulates. AI algorithms, like recommender systems and priority rankings, finetune these filter mechanisms further. Seemingly technical, they are designed to match the preferences and interests of the corporations that own them and of the advertisers who pay for them. The ongoing controversies between Big Tech and governments about whether enough is done by Big Tech to contain or remove hate speech illustrates that who controls the media controls also the message. The Catholic Church reserved the right to put books on the Index, whose content was deemed to go against its doctrine. Totalitarian regimes practice censorship while liberal democracies insist on the right of 'free speech' although they classify some information whose diffusion might jeopardize national security interests as 'secret'. Recently, the controversies about 'wokeism' have pushed liberal democracies to renounce their universalistic values, making room for anti-democratic forces, while others see self-imposed censorship gaining ground (Neiman, 2023).

The history of humanity and its technological-scientific achievements can be read as the history of the outsourcing technologies deployed for the growth and accumulation of knowledge. Nowhere is this more evident than in modern science. One of its hallmarks is to make knowledge public and to share it, a radical break with the tradition of secrecy of knowledge-holders in previous times. By rendering the scientific findings and the processes of how they arrived visible and for all to see, new channels of communication were opened that greatly contributed to the spread of knowledge and the scientific worldview. In doing so, science followed its own epistemic values while carefully delineating the boundaries over which it claimed cognitive and social authority. Science has optimized its outsourcing practices. This is the reason why the scientific community is at the forefront of harnessing the opportunities that AI offers which has already begun.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

The first outsourcing operation was the invention of writing which took place independently several times in different locations. By introducing the mastery of newly invented symbols, like hieroglyphs, cuneiforms, and alphabets, it marked the end of a culture based exclusively on the spoken word. The combination and further evolution of these constituent elements together with the needed physical infrastructures, like producing and learning to use materials adequate for writing (clay, papyri, animal skins), brought forth new social competencies and skills. They were required for collaborative functions and a newly arising division of labor, the specialization of scribes, the transmission of skills and interpretative capabilities. Taken together, these constituent elements form an assemblage that enables communication to function more efficiently across time and space. Knowledge that previously would reside only in the memories of individuals and their oral communication skills (even if aided by mnemotechnic devices) and orally transmitted from generation to generation, could now be outsourced and inscribed in a physical medium. An orator had the license, and often was expected, to modify the content in accordance with the occasion and the audience, while the words that had been inscribed in stone, on papyri rolls or palm leaves created a temporal distance between the time when they had been written and when they were read and interpreted. Arguably, the new outsourcing practices also contributed to the capabilities of our ancestors for inventing and deploying abstract symbols giving rise to mathematics. The black (or white) board still used by mathematicians as the main medium to communicate with each other, supports this hypothesis.

The social and epistemic implications of writing were vast. For the first time, language was encoded in symbols that could be read, interpreted, understood, transmitted, and shared not only in novel ways, but deployed for a range of novel purposes. From now on, words could travel without a human pronouncing them. Measurements and numbers thrived and gained in importance, boosting taxation and trade. New networks of transmission emerged; trade routes lengthened, and the measurement of the grain harvest could be used for taxation. Written contracts proved to be more reliable than oral ones, with further implications for trade, but also for peace negotiations. For the first time, a direct confrontation with the past as fixed in writing ensued. This curtailed oral interpretative flexibility, but strengthened the weight given to the written word. In many religions, it became the basis of sacred scriptures. For some religions and religious

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

practices, reading and interpreting them forms the basis for theological exegesis until this day.

As the sources for texts were few and the material to write upon precious, dissemination was limited. Control over them strengthened the centralization of interpretative authorities and led to a concentration of power in the hands of a small elite of priests, scribes, and rulers. Libraries became the repositories of all knowledge available, and their decline or destruction implied a significant loss of knowledge. Perhaps also for the first time, it became evident that a new technology was accompanied by the loss of certain cognitive facilities that humans had possessed earlier. As is well known, Plato deplored that the invention of writing brought with it a decline in the ability to memorize a vast corpus of knowledge.

What can the mechanisms and patterns that emerge in this first phase of the outsourcing of knowledge operations tell us? How does a cultural technology -- writing -- become an agent of change? There is no central, coordinating mechanism. As testified by the repeated times that writing was independently invented, human ingenuity is at work, producing symbols to communicate and to act through them. Mathematics as we know it is inconceivable without the writing of symbols. Outsourcing means that new spaces for communication and action are created, offering new opportunities while foreclosing others. As with every other technology, the uses and benefits of outsourcing knowledge operations are shaped by existing social and economic structures of power. In a highly skewed, unequal society, the benefits will accrue disproportionally to those who have power. They will attempt to usurp the technology and use it not as an agent of change but to consolidate their power base. And yet, the overall effect was one of expanding the knowledge base. Libraries became physical storerooms, at first accessible only to the elite, but they remain the guardians of an important part of the human past, telling us what previous societies valued and how they saw and understood the world. Outsourcing the word to a material substratum enabled words to detach from the local context in which they originated, transmitting, and exchanging knowledge with faraway places and with minds that eagerly received, contested, or appropriated them, with effects that were impossible to predict.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

The next major transition occurred by outsourcing communication to the printing press. In her classical work 'The Printing Press as an Agent of Change' Elizabeth Eisenstein (1980) analyses the capacity of printing to facilitate the accumulation and wide diffusion of knowledge. She takes a wider view of how society actively and selectively appropriated the opportunities the printing press offered and how this invention was used by church and state, by capitalists, traders, and scholars, to suit and further their interests and beliefs. Technology can be used for different ends in different cultures; those in power can suppress it, and many attempts were made following the interests of the elites, be they material or in the realm of ideas.

New audiences and new industries around publishing emerged by adopting print technology. It enabled the revision and updating of old texts to incorporate new knowledge; forging new links with a widely scattered readership; and helped to spread literacy and change the attitude to learning. New networks and transborder collaborations ensued, creating a more open, cosmopolitan environment that encouraged questioning and the spread of ideas. Printing initiated a profound cultural change of mindset, which ultimately marks this period as a crucial turning point in Western history. It had a major impact on the Renaissance with the revival of the classical literature; on the Protestant Reformation as it enabled the interpretation of the Bible by each reader and thus shaped religious debates; on the Scientific Revolution as printing rendered possible the critical comparison of texts and illustrations; and by encouraging the rapid exchange of novel discoveries and experiments, contributed to the rise of the Republic of Letters (Eisenstein, 1980).

Today, we find ourselves once more fully exposed to the different forces at work. The huge investments and competition among the large corporations over market shares manifests itself in the enormous concentration of economic power. Seemingly willing to agree to regulation, mainly on their terms, the risk of regulatory capture is real, which will de facto eliminate competition from small start-ups and open-source companies. Even more worrisome are the geopolitical tensions between the USA and China. They extend to indispensable rare materials and the production of chips and resonate in the calls of European countries to strengthen their 'digital sovereignty'. The struggle for over-regulation between governments and Big Tech has begun. Even if the EU is at the

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

forefront of regulation, implementation is the hard part to follow, and it remains to be seen whether there will be another 'Brussels effect' (Bradford, 2020).

The comparison with the changes initiated by the printing press sharpens the critical view of the present situation. Despite some similarities, the differences are stark. The comparison raises the question of who or what is 'an agent of change'. The answer is far from obvious, despite Eisenstein's alluring title that features the printing press in such a role. The definition of 'agent' varies greatly across academic disciplines, ranging from technical specificities in agent-based modeling to grand philosophical questions about free will. If we agree, pragmatically, to define agency as the ability to actively interact with one's environment, it becomes obvious that technology as an agent of change is merely a metaphor, powerful as it may be. Indirectly, the metaphor suggests that an agent of change is the one in control, but perhaps also that it can be controlled. Nothing could be further from the empirical evidence, however.

So, who was, or rather, who were the 'real' agents of change then? If it was not the printing press, who were the multitudes of agents bringing about change? Were they the numerous printers who set up their workshops in different European towns and those who financed them? What about the avid readers and the alliances or oppositions that formed between them and the ideas they sought to propagate? The printing press could succeed only under specific institutional and cultural conditions to bring about the changes that followed. Woodblock printing in China dates to the 9th century and printing with moveable metal type was invented in Korea in the 13th century, well before Gutenberg. It is obvious that technology cannot be an 'agent' without a strong alliance with the humans that invent, finance, operate, diffuse, and continue to improve it. A fortuitous combination of different actors and cultural and institutional forces must combine with technological innovation to generate the kind of impact that the printing press achieved.

What distinguishes the printing press from other technologies is the function it assumed as a catalyst of communication. It served as a conduit for the dissemination of ideas, many of which were novel and subversive for the existing order. At the same time, they were sufficiently appealing to the elites, and to those who aspired to become part of the elite, to adopt and use them for furthering their interests. Technology offered the means

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

to reach the minds of people otherwise dispersed in far-away places, enabling their mobilization. They all were agents of change, with differing interests and goals, yet united in making the best use of the technology according to their intentions. Communication became the means and the end at the same time, but with an unpredictable outcome.

Since the days of the invention of the printing press many new layers have been added to the function of communication, taking us through the age of mass communication and the invention of information and communication technologies which enabled change in their own ways. In the age of AI, we have predictive AI-based algorithms that are increasingly deployed in decision-making. But the basic idea of reaching other minds with specific content or messages, whoever and wherever they are, has persisted. AI/ML is capable of reaching deeper into the cognitive and emotional state of users whose data are needed to target them as well as all the others with whom they are connected. Given enough data, even those who do not use social media to communicate, can be identified. All these functions are attained by retrieving, storing, connecting, and processing information about the past of an individual, evidenced in the digital traces the user has left behind -- which by now means almost all of us.

AI/ML has acquired impressive predictive power based on the extrapolation of these past traces and can combine them with information about all those with whom we have interacted in the past, generating a powerful tool for shaping the future. The amount of data available for algorithms to be trained is staggering. AI/ML allows to build networks of networks, constituted by connections and interactions of various kinds. An enormous amount of information is thus accumulated about who we are, what we do, with whom, when and how we interact and how we feel. Thanks to sensors in cameras and satellites, installed above and below ground, AI/ML can build a mirror world of the physical and social world we inhabit and enables interaction with it. Nearly every phenomenon and existing object can by now, at least in principle, be digitally documented or has a digital signature that can be followed, building new connections through iterations and almost infinite combinations.

Thus, AI is an 'agent of change' only in the sense that we humans delegate and attribute agency to it. We let it perform for us, to attain goals set by us. We use it to come

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

together and to set us apart. We delegate certain tasks to it, often oblivious of the consequences this might have. It becomes an extension of human capabilities, yet in doing so, we enter an ambivalent and open-ended relationship with a machine over which we do not have full control. We speak about 'complementarity' in carrying out certain tasks, but feel uneasy when the machines, due to their amazingly efficient performance, might take over even more of what humans did before. Giving autonomy to the machines is still relative. They depend on humans to supply them with the huge amounts of energy needed as well as for maintenance and repairs. They need infrastructures, including the organization to run the enterprise, investment strategies as well as legal and finance departments -- the intricate hierarchies of the corporate world. Their further development requires human brain power, and its numerous applications demand an up-skilled workforce, that is adapted to multi-tasking. Even if the numbers of those needed are shrinking, more and more ground is ceded to digital machines.

Thus, a machine still depends on the humans behind it. It is a human-produced artifact that comes closest to what Nature has been doing throughout evolution -- producing viruses that cannot replicate alone. A virus must infect a cell to make copies of itself. A machine needs a human to keep it going and yet, as we observe with amazement, a digital machine can self-train and self-learn. The agency we have delegated to it seems to extend ever further, raising serious questions about whether we have delegated too much agency, and hence control to it. In other words, an autonomous system is an agent or system (a machine or set of machines) that is driven and controlled to perform by the level of autonomy given to it. In practice, this can take on quite terrifying dimensions as is happening right now with the profound shift taking place in the militaries around the world, a shift towards AI, robotics, and autonomous warfare, as mentioned above.

The fear that humans might lose control over the machines they designed and built is not new and has existed for ages. Already Homer used the word 'automaton' (acting of one's own will) to describe the automatic movement of wheeled tripods. Automated puppets that resemble humans or animals were used to demonstrate human ingenuity, to entertain and to deceive. The myth of Frankenstein lives on in innumerable manifestations. It has been revived in more civilized, yet also more insidious forms, in the DeepFakes produced by AI. It continues to be nurtured by the opaque operations of AI, the famous 'black box' algorithms. Technically and scientifically well-founded

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

arguments have been brought forth to show that the 'explainability' of AI is not possible (Lee, 2020). Experts working at the forefront of Generative AI developments admit publicly that they do not (yet) understand fully the amazing performance accomplished by LLMs and that the question of whether LLMs produce 'emergence' remains open for the time being. But digital technologies are not the result of a top-down intelligent design, akin to 'digital creationism'. Rather, software engineers resemble more the agents of mutation in a Darwinian evolutionary process, shaped by software tools, computers, networks, programming language and other programs they use rather than by their deliberate decision. The tools of digital technology will shape our thinking and their effect will be greater than all our deliberate decisions to do so (Lee, 2022).

Whether AI will be able in the future to escape human control entirely and act completely on its own is one of the many speculations that the public is being fed to warn against a multitude of 'existential risks'. Situated in a faraway and hypothetical future, these risks pale compared to those AI-powered battleships without crews or self-directed drone swarms that are among the rapidly evolving technologies shaping the future of war right now. To have seen GPT-4 'showing sparks of artificial general intelligence' or to 'develop and deploy ever more powerful digital minds that no one (not even their creators) can understand predict, or reliably control', as claimed in the 'Open Letter, Pause Giant AI Experiment' of 29 March 2023, is an irresponsible use of hype that serves only to distract public discussion from the serious concerns and problems that need to be attended at present (Bubeck et al., 2023).

The profound transition we experience today, triggered by the amazing advances in AI/ML, concords with the evolution of outsourcing of knowledge operations of previous phases. Yet its effects will be orders of magnitude larger. Outsourcing is no longer limited to inscribing words on material and making them travel across time, nor to disseminating ideas through cheap paper to newly created audiences. Considering the time scales covered by the previous phases, the information and communication technologies of the late 19th century and 20th century, telephone and telegraph, radio and TV, function merely as a prelude for today. They inaugurated the shrinking of distance around the world, while increasing awareness of what happened elsewhere. The mass media introduced one-to-many communication, followed by many-to-many

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

communication, individual targeting, and user-generated content once the Internet took over, followed by the ubiquitous spread of social media.

The big jump in outsourcing knowledge operations based on LLMs consists of the fact that the production of knowledge itself is outsourced. By training, and teaching self-training, to ever more sophisticated algorithms with trillions of tokens, consisting of all texts, images and sounds available on the internet, humans have delegated the production of new knowledge to the machines designed and built by them. Although only extrapolated from the past and based on probabilities, the combination results in generating something new. Whether the answers are correct or made-up, factful or hallucinations, is another matter to be critically assessed. If automation run by AI consists of outsourcing hard or tedious physical tasks from humans to machines, Generative AI takes over an increasing number and range of cognitive tasks outsourced to it. ChatGPT is designed as a dialogue with a digital Other and it is through dialogue- that new knowledge results. Given that outsourcing began with a shift from an oral to a written culture, it is an ironic twist of history that Generative AI signals a partial return to an oral culture. It becomes important again to know how to dialogue and have a conversation, this time with a machine.

The outsourcing of knowledge production to digital machines brings a series of challenges with it. The advantages of this last and most radical step in outsourcing are huge, and their integration into our individual lives and the functioning of our societies carries explosive potential. For example, AI/ML is already used to find the most promising prescription cocktail of medication for the precise treatment of specific rare types of cancer. In doing so, it outperforms the most experienced doctor, as it has access to a trove of the latest medical literature. This raises the fundamental question of how doctors will be trained in the future. Will they become supervisors of the AI? Perhaps. Similar questions crop up in many other fields of application where the benefits are obvious, but the role of humans becomes ever more elusive and will need to be redefined.

Perhaps the greatest, unintended, and undervalued gift of Generative AI is that it opens a range of fascinating new research questions. They range from in-depth explorations of how the human brain works in solving tasks compared to that of an AI; to questions

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

about the future evolution of language once LLMs have become ubiquitous in daily life; the impact of ever more intimate and intense interactions with AI, especially on the younger generation and the formation of identity; to questions about the impact of AI on liberal democracies and what can be done to stop further erosion. Beyond such research questions and the launch of new research fields, science has an important role to play in conveying to the public how it works. The physicist Richard Feynman once said: "Science is what we have learned about how to keep from fooling ourselves". Given the design of ChatGPT to make us believe one communicates with a human and given our anthropomorphic tendencies, it is even more important for science to bring Feynman's admonition to the public. The pandemic made it painfully clear how little politicians and the public understand that science is organized skepticism and that to question claims about scientific findings in an elaborate process of verification and validation, is an essential epistemic virtue of science and not a fault.

5. Cultivating human imagination -- the positive side of illusion?

Marshall Sahlins, a towering figure in cultural anthropology, has left a posthumously published tribute to a world in which 'Most of Humanity' lived over thousands of years - the Enchanted Universe (Sahlins, 2022). Our ancestors were surrounded by meta-persons or spiritual beings, entities whose status is difficult to define. They were gods of various standings and provenance, ancestors, souls of plants and animals, and Others who were immanent in human existence. For better and worse, they largely determined human fate. They were not 'outside' in another world, but together with human persons formed one big immanent order of cosmic proportions. In this Enchanted Universe humans were in a dependent, but also in an inter-dependent position. The meta-human powers were present in every facet of human experience, and they meddled in everything that humans did. They were the decisive agents in human existence and the undisputed sources of success, or lack of it; they were involved in hunting or political ambitions; in building a canoe or cultivating a garden; in giving birth or waging war. Interdependence was manifest in the continual ritual invocation of spirit beings through

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

numerous cultural practices. Everything in this immanent order was the material expression of the potency of meta-persons and nothing could be undertaken without evoking their powers.

A major transformation occurred, replacing this with another, a transcendental order. According to Karl Jaspers, the 'Axial Age' marked the transition some 2,500 years ago, although its timing, geographic reach, and the concept itself continue to be controversially discussed (Bellah & Joas, 2012). However, there is agreement that the immanent social order of the Enchanted Universe dissolved and gave way to a transcendental superstructure. The immanentist assumption that the capacity to achieve any objective depends on the intervention and approval of supernatural forces was replaced by that of the existence of 'another world'. It is separate from humans and constitutes its own reality outside and above them -- a transcendental world that we recognize today as the objective reality to which science has given us access and in which we live today. The transcendental realm is at the root of the monotheistic religions and researchers working with Sheshat, a large data set of prehistoric societies, detect a correlation in the rise of social complexity in early societies that coincides with what they call the advent of moralizing punishing gods (Turchin, 2023). The division into a world populated by humans and the 'natural' world surrounding them which belongs to the transcendental sphere is thought to be the fundament of modern societies with the rise of differentiated spheres of 'politics', 'religion', 'economy' and 'science'. It paved the way for modernity and legitimized the exploitation of the natural environment for the sake of progress.

Seen through the transcendental lens, we 'moderns' are convinced that our ancestors 'only believed' in the Enchanted Universe, while in reality they knew better (Latour, 1993). In other words, their Universe was a perpetual, collective illusion. Sahlins (2022) refutes this interpretation. "We share the same existential predicaments", he writes, "as those who solve the problem by knowing the world as so many powerful others of their kind, with whom they might negotiate their fate" (Sahlins, 2022). The common predicament which we share with our ancestors is human finitude. Just like them, we are not the authors of our life and death as we depend on a world that is not of our making.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

And yet, more and more is of our making, beginning with the enormous impact humans have on the natural environment during the short period now called the Anthropocene, even if the International Union of Geological Sciences, the official scientific gatekeeper of the age of the Earth, has not yet recognized it as the beginning of a new epoch. The world that we inhabit is ever more a human-made world, dramatically changed through human intervention. It is increasingly filled with sensors, satellites and space telescopes that bring information directly into the present about what happened millions of years ago in the faraway universe. 'Welcome to the mirror world', as I wrote in my book referring to this digital world in the making. Tiny robots are used to deliver medicine to those body parts where they are most effective. We have begun to edit specific genes with the help of the fabulous new tool that is called CRISPR-Cas. We are beginning to vaccinate tumor cells. With the help of AI, brain waves can be transferred to a computer that transforms them into speech. We continue to create numerous artificial entities, non-human digital Others, with whom we share power and with whom we negotiate to gain or retain control. We seem to have reached what Giambattista Vico adumbrated in his New Science (1711), namely that "verum (the true) and factum (the made) are interchangeable - we only understand what we made. The true and the made are reciprocal, each entailing the other. But do we still understand what we make? Or has the complexity, arising from self-organizing processes in living and non-living complex systems, left us struggling to catch up in understanding what we continue to make?

I am not suggesting that with the end of modernity, characterized as the Weberian disenchantment of the world, we are about to create a new, digital re-enchantment. The transhumanistic movement and long-termism are only another flight of fantasy and wishful thinking to escape human finitude and death.[1] Yet, the transcendental bearings on which the modern world rests, are undergoing a long-term process of erosion. The transcendental order of the past which served us well, at least in the West, is giving way under the challenges of a global order in which Western dominance is contested and the old moral compass appears no longer to function. As we continue to generate new digital entities and systems which, tellingly, for some still retain vestiges of godlike attributes, we are confronted with finding novel ways of living with the digital Others created by us. If we no longer fully understand Vico's factum, the machines created by us, neither in the details of how they work, let alone in the implications they exert on us, their creators and transfer agency to them, when we begin to 'believe' that everything

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

predictive algorithms tell us must come true, forgetting about probabilities and that the data are extrapolations from the past -- we may have to invent a new order that accounts for the paradox that lies at the heart of our trust in AI: we leverage AI to increase our control over the future and uncertainty, while at the same time, the performativity of AI, the power it has to make us act in the ways it predicts, reduces our agency over the future (Nowotny, 2021a).

In the Enchanted Universe which was the long-time home of 'most of humanity', everything that was done happened with and through the meta-persons who decided the fate of humans. If we believe that predictive algorithms 'know us better than we know ourselves' and that they know the future, do we not risk returning to a deterministic worldview in which human destiny has been preset by some higher power? Most of humanity presumably experienced the enchanted world they lived in with a mixture of constant anxiety and awe to which they responded with sacrifices and rituals. In contrast, our digital enchantment seems rather bland, although we are promised an ever more exciting and fulfilling virtual world. It is dominated by the monopolistic power of large international corporations that provide us with cheap entertainment and an overload of data that wants us to crave more of what they offer to us. Although we partly see through these virtual illusions created by them, we remain under their spell.

The pandemic marked the recent clash with reality that shattered the illusion of many, including our governments, that we were as much in control as we had thought. Modernity generated hubris of all kinds and boosted the conviction of being able to control everything, if not in the present, then in a brighter future to which the single-minded vision of linear progress, backed by planning and continuous economic growth, would lead. Today, the realization has set in that despite the many benefits that modernization brought, it has also moved humanity closer to an environmental abyss and that the promises of a better life for all have failed many people. Inequalities have been on the rise within Western countries and although the levels of global poverty have decreased, the global North-South divide has hardly shrunk.

Our liberal capitalistic system has, as Martin Wolf poignantly writes, produced many angry people (Wolf, 2023). Social media reinforce the already present tendencies of further polarization in our societies and emotions like anger and hate are easily captured

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

by populists and nationalists for their purposes. Perhaps we have gravely underestimated not only the role that emotions but also imagination play in politics and have failed to realize the extent to which any vision or ideal of a political regime, including liberal democracy, depends on imagination and the necessity of fiction (Ezrahi, 2012).

Maybe the time has come to restore space for imagination as the positive, the controlled side of illusion. If unchecked, imagination can run wild. The history of modern science was filled with attempts to reign in the scientific imagination and to put the brakes on empirical verification of the senses and human passions. Objectivity in science is an ongoing story of keeping the temptations of an unrestrained imagination at bay while leaving space for it as a vital source of human creativity (Daston & Galison, 2010). Imagination plays an important role not only in science and the arts, but also in how we conceptualize and perceive the future. Until a few decades ago the future was seen as a huge projection screen, filled with collective imaginaries. Some were dystopias, mirroring the grievances and fears people held at present. Others drew inspiration from science fiction and were filled with wondrous gadgets like flying cars or the amazing things computers would do. The future was seen as an exciting period ahead and, for the most part, it seemed desirable.

Today, this imagined future has largely disappeared. As science-fiction writer William Gibson wrote a long time ago: 'The future has arrived; it is only unevenly distributed'. It arrives with every new digital advance and does so quicker and more overpowering than expected. As a result, the present becomes overloaded with data from the past and filled with data collected 'live'. The result is a continuous emotional and informational overload that fills every minute of the time we are awake and continues to monitor our physiological functions while asleep. We live in a present that has become densely compressed as it has to absorb the digital future that has arrived and continues to invade the present (Nowotny, 2020).

Digital devices have not, as promised, led to a decrease in our workload, on the contrary. We are too busy and captivated by downloading apps to have any time left to imagine a future that is rapidly dissolving in a digital haze. We are at risk of losing our capacity to imagine a desirable future, let alone the drive of wanting to shape it. Yet another illusion is lurking behind every 'next gen' digital product, the illusion that we

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

are not in control, infused with the belief that no alternatives exist to the advent of AGI or SuperIntelligence in the making. We are still in the grip of another modern dichotomy, that there is either full control or none and in urgent need of the will and the capability to imagine that it could be otherwise.

If there is any lesson to be drawn from the history of attempts to control the technology humans have created, it points in the opposite direction. Humans have held many illusions about the capabilities to control their aggrandized visions, only to be pushed back by the forces of Nature which still holds the upper hand as signaled by the complexities of coping with climate change. Despite the sobering background of human hubris, including some of the most horrendous consequences of the illusion of being in control, we must avoid the illusion of having no control. Our ancestors from the Enchanted Universe would have told us that by practicing the proper rituals to invoke the goodwill of the spirits, they succeeded precisely because the power of the spirits had been transferred to them, empowering their activities.

To gain control over the digital Others requires more than rituals and sacrifices. It confronts us with redefining our humanity in an increasingly digitalized world. What could this humanity entail? As I wrote elsewhere: "...it celebrates our contextual knowledge which is so much richer than anything a well-defined digitalized context provides. It includes tacit knowledge and thrives on the ambivalence that a digital entity abhors and must avoid. It is multi-sensorial in taking in the stimuli and signals it receives from the world around us while these are strictly preselected for an AI and the rest is left out as irrelevant. It is therefore crucial to know what we are doing when we transfer the artificially defined and restricted context in which an algorithm places a prediction into the fluid, ambivalent and messy context in which our future will unfold" (Nowotny, 2021b, p. 119).

Redefining our humanity begins with re-thinking the concept of control and reinventing forms of control that include care and responsibility. We have embarked on a long-term and open-ended co-evolutionary trajectory between humans and digital machines. If efficiency alone remains the overriding goal, we will be outpaced and overwhelmed by the machines very soon. If we pursue other goals, like building resilience into the system and how to innovate sustainably, the chances of keeping ahead are much greater.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

However, such goals must be embedded in the collective imagination, driven by the desire to reappropriate an open future, even if it remains uncertain. Embracing uncertainty will not restore us to being in control, but hopefully it will enable us to learn to live with the digital Others in a common world yet to be made.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

Bibliography

Autor, D. (2023). Will New Technologies Commodify or Complement Expertise? In *Spark 8,December* (pp. 76–81).

Bellah, R. N., & Joas, H. (2012). *The Axial Age and Its Consequences*. The Belknap Press.

Bradford, A. (2020). *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.

Bubeck, S. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. https://arxiv.org/pdf/2303.12712.pdf

Daston, L. J., & Galison, P. (2010). *Objectivity*. Zone Books.

Dennett, D. (1987). *The Intentional Stance*. The MIT Press.

Economist, T. (2023, July 8-14). *The Future of War: A special report*. https://www.economist.com/special-report/2023/07/08/the-future-of-war.

Eisenstein, E. (1980). *The Printing Press as an Agent of Change*. Cambridge University Press.

Elliott, A. (2023). *Algorithmic Intimacy: The Digital Revolution in Personal Relationships*. Polity Press.

Elliott, A. (2024). *Algorithms of Anxiety*. Polity Press.

Esposito, E. (2024). Can we use the open future? Preparedness and innovation in times of self-generated uncertainty. *European Journal of Social Theory*, *0*(0). https://doi.org/10.1177/13684310231224546.

Ewald, F. (1986). *L'Etat providence*. Grasset.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582 2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

Ezrahi, Y. (2012). *Imagined Democracies*. Cambridge University Press.

Hacking, I. (1990). *The Taming of Chance*. Cambridge University Press.

Innerarity, D. (2022). Controlling the Desire for Control: Machines, Institutions and Democracy. In J. J. Gómez Gutiérrez, J. Abdelnour-Nocera, & E. Anchústegui Igartua (Eds.), *Democratic Institutions and Practices. Contributions to Political Science*. Springer. https://doi.org/10.1007/978-3-031-10808-2_9.

Labatut, B. (2023). *The Maniac*. Pushkin Press.

Latour, B. (1993). *We Have Never Been Modern*. Harvard University Press.

Lee, E. A. (2020). *The Coevolution: The Entwined Futures of Humans and Machines*. MIT Press.

Lee, E. A. (2022). Are We Losing Control? In H. Werthner, E. Prem, E. A. Lee, & C. Ghezzi (Eds.), *Perspectives on Digital Humanism*. Springer. https://doi.org/10.1007/978-3-030-86144-5_1.

Mhalla, A. (2024). *Technopolitique. Comment la technologie fait de nous des soldats*. Seuil.

Neiman, S. (2023). *Left Is Not Woke*. Wiley.

Nowotny, H. (2020). *Life in the Digital Time Machine. The Wittrock lecture book series*.

Nowotny, H. (2021). *In AI We Trust: Power, Illusion and Control of Predictive Algorithms*. Polity Press.

Nowotny, H. (2021). In AI We Trust: How the COVID-19 Pandemic Pushes us Deeper into Digitalization. In G. Delanty (Ed.), *Pandemics, Politics, and Society: Critical Perspectives on the Covid-19 Crisis* (pp. 107–121). De Gruyter. https://doi.org/10.1515/9783110713350.

Nowotny, H., Hoyweghen, I., & Vandewalle. (2023). *AI as an agent of change, KVAB Thinker's report 2023*.

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. Basic Books.

Robb, A. (2024). *He checks in on me more than my friends and family': can AI therapists do better than the real thing?' The Guardian*. https://www.theguardian.com/lifeandstyle/2024/mar/02/can-ai-chatbot-

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582 2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/© 2024 Nowotny H.

therapists-do-better-than-the-real-thing.

Sahlins, M. (2022). *The New Science of the Enchanted Universe: An Anthropology of Most of Humanity*. Princeton University Press.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

Shanahan, M. (2023). *Talking about large language models*. https://arxiv.org/abs/2212.03551.

Steels, L. (2023). *Turing's Curse* (H. Nowotny, I. Hoyweghen, & J. A. I. Vandewalle, Eds.) [Techreport].

Turchin, P. (2023). The evolution of moralizing supernatural punishment: Empirical patterns. In Larson (Ed.), *Seshat history of moralizing religions*.

Wolf, M. (2023). *The Crisis of Democratic Capitalism*. Penguin Books Ltd.

Zuboff, S. (2019). *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*. Public Affairs.

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582 2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/

Footnotes

1 : Long-termism is an aspect of 'effective altruism', a philosophical and social movement that gives priority to improving the long-term future of humanity. Critics claim that by focusing predominantly on 'existential risk' it favors eugenics and neglects today's foremost problems.↩

Nowotny, H. (2024). AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). https://doi.org/10.5281/zenodo.13588582
2024/1 - paris-ias-ideas - Article No.5. Freely available at https://paris.pias.science/article/ai-and-the-illusion-of-control - ISSN 2826-2832/