THE OPEN FUTURE AND ITS ENEMIES

HOW WE CAN PROTECT FREE SOCIETY FROM AI DICTATORSHIP

MATTHIAS PFEFFER
JÜRGEN PFEFFER
PAUL NEMITZ

THE OPEN FUTURE AND ITS ENEMIES

HOW WE CAN PROTECT FREE SOCIETY FROM AI DICTATORSHIP

By
Matthias Pfeffer · Jürgen Pfeffer
Paul Nemitz

European Parliament

This book was published with the financial support of the European Parliament. It does not represent the view of the European Parliament.

Bibliographical information of the German National Library
The German National Library catalogues this publication in the German National Bibliography; detailed bibliographic information can be found on the internet at: http://dnb.dnb.de.

ISBN 978-3-8012-3111-8

FEPS Supervision: Monty Aal, Vanessa Zwisele
Layout and editing: Verlag J. H. W. Dietz Nachf. GmbH
Typesetting: Kempken DTP-Service, Marburg
Cover picture: iStock
Printing and processing: Bookpress, Olsztyn

Published by

DIETZ

Verlag J. H. W. Dietz Nachf. GmbH
Dreizehnmorgenweg 24, 53175 Bonn, Germany
www.dietz-verlag.de

Published in association with the

FEPS
FOUNDATION FOR EUROPEAN PROGRESSIVE STUDIES

Foundation for European Progressive Studies (FEPS)
Avenue des Arts 46 – 1000 Brussels, Belgium
www.feps-europe.eu
@FEPS_Europe
European Political Foundation – N° 4 BE 896.230.213

Printed in Poland 2026

TABLE OF CONTENTS

Introduction: The AI Paradox

Come out into the open, my friend!
Friedrich Hölderlin

Let's begin with a paradox: countries and corporations around the world are currently engaged in a race to see who will be the first to develop artificial intelligence (AI) that is superior to humans in terms of thinking and decision-making abilities, and can control processes and machines completely autonomously. Hundreds of billions of dollars are being invested and nuclear power plants are being reconnected to the grid to generate the energy needed to achieve this goal. The aim is to gain leadership in a universal technology that can be used to control entire societies, conquer markets and equip weapons. It is also about geopolitical supremacy in the 21st century. The general belief is that whoever masters AI will rule the world.[1] But if this billion-dollar gamble pays off, we will face an obvious dilemma. If we succeed in creating a technology that is superior to humans in all areas and that also acts autonomously, i. e. it can set its own rules, how can we ensure that we, the then inferior humans, retain control over it? This is a question that should also be asked by those who are the first to cross the finish line in this ludicrous race. Hence the paradox: for by definition even they would no longer control this technology.

Even though we are still a long way from such autonomous AI, we are in the midst of a tremendous upheaval that can only be compared to the first human settlements or the Industrial Revolution. The rapid development of AI is already transforming all areas of life – from economics and science to administration, politics and everyday social life. And today's digital revolution is happening exponentially faster than previous revolutions. In the process, we are heading towards a conflict rooted in the different logics of AI and democracy. It is the conflict between everything and everyone as a calculation and the unbreakable dignity and

1 Putin said this back in 2017, and Elon Musk agreed, see here: Hern, A. (2017) "Elon Musk says AI could lead to third world war". *The Guardian*, 4 September.

rights of each individual. It is the conflict between the centralisation of enormous data power for the purposes of surveillance and control and the fundamental principle of the decentralisation and separation of powers in democracy, which is intended to protect fundamental rights and enable democratic self-determination.

This systemic conflict has reached the global political stage. It has been escalating since autocrats and anti-democrats have not only taken advantage of the emergent technologies, but have also allied themselves with their developers and operators in order to undermine democracy and gradually replace it with an autocratic algocracy[2]. The philosopher Jürgen Habermas observes in the USA under Trump 2 a "creeping but unerringly pursued takeover of power" that aims to replace liberal democracy with an "authoritarian-controlled, technocratically administered but economically libertarian social system."[3] AI is a key technology in this takeover because it integrates the control and surveillance of the economy and society with globally scalable business models of unbridled digital capitalism. If we do not take decisive action to counter this seizure of power, we are heading towards the establishment of an AI dictatorship. And this without any putative super-intelligence; all it takes is an alliance of super-rich companies with super-powerful states.

Because it is likely that AI will increasingly shape our future and because AI is praised by its developers and operators as the decisive technology of the future – one which will also make this future predictable and controllable – an analysis of the social consequences of this technology must anticipate future developments in addition to current iterations. The further development of AI systems into a technology that is almost on a par with human thinking cannot be ruled out, nor can the achievement of so-called artificial general intelligence (AGI), whose power could surpass human thinking and decision-making abilities. Even if it only simulates human thinking, it may nevertheless develop a controlling power that exceeds human capabilities. The enormous investments by the leading AI companies – OpenAI, Microsoft, Google/Alphabet, Facebook/Meta, Ap-

2 Danaher, J. (2016) "The Threat of Algocracy: Reality, Resistance and Accommodation". *Philosophy & Technology*, 3(29): 245–268. DOI: 10.1007/s13347-015-0211-1; Pfeffer, M. (2024) "Demokratie statt Algokratie". *Neue Gesellschaft/Frankfurter Hefte*, 2 January.

3 Habermas, J. (2025) "From Here On, We Must Go It Alone". *K. – Jews, Europe, the 21st Century*, 30 November.

ple, Amazon and Musk's xAI – are made in the service of the goal of AI superiority over humans.

An impact assessment of this technology must take its political implications into account. Given the tremendous pace of development and in the absence of effective political regulation, we must assume companies will drive AI development with their goals of conquering markets and power in mind. They now see the separation of powers in democracy as nothing more than an obstacle on the path to technocratic hegemony. In line with the current US administration and President Trump's AI strategy, Big Tech is calling for the dismantling of laws so that it can determine the direction and speed of development itself. Trump wants to win the AI race at any cost in order to achieve world domination. To this end, he is prepared to ban AI legislation in individual US states and to exert open pressure on European institutions to water down regulations and favour US companies that already dominate Europe's digital infrastructure. Europe's dependence on the US in all areas of the so-called tech stack, i. e. the entirety of technologies used to create and run software applications, is 80 percent. It is even more comprehensive than Europe's dependence on the US in defence and therefore even more threatening.

We aim here to contrast this dystopian outlook with possible development paths that would allow AI to be integrated into the legal framework and value system of free, democratic societies. Such democratically regulated AI would have a framework that ensures human control. This would solve two problems in current AI development: namely, the need to align AI with constitutional, legal and public interest requirements (the alignment problem), and the question of how and by what means superior AI can be controlled by humans (the control problem). We are, however, a long way from achieving either.

The question of how to ensure that the goals and actions of AI are consistent with human values, interests and intentions is complex. The reasons for this are deeply rooted in the logic of AI systems. Alignment means that AI should do what we want, not what we do not want, and what corresponds to our values and laws. AI could perform tasks in ways that are technically correct but undesirable or illegal from a human perspective. Even today, AI systems repeatedly exhibit misconduct on a small scale. With very powerful AI systems, small deviations from the intended path can lead to extreme consequences in billions of computing operations. In addition, there is the long-term theoretical control problem. This

consists of keeping AI fundamentally open to human intervention and improvement, and thus cooperative. Research in this area is still in its infancy, and there are experts who believe that vis-à-vis AGI this goal is impossible to achieve.

This goes hand in hand with the political control problem, i. e. the question of how we humans will be able to control such powerful technology and its effects on individuals and society in the future. If the promises of superiority made by Sam Altman, the head of OpenAI, come true, this control would by definition be impossible.

It is not only the prospect of possible future scenarios that is cause for concern. AI is already in the hands of a few leading global Big Tech companies that are increasingly colonising our lives. The colonies of digital disruption are the hitherto free and democratic societies around the world. The situation is even worse in dictatorships such as China, where people are controlled and manipulated by the government through AI. Jürgen Habermas described the advance of the systemic logic of money and power into the last areas of everyday communication and meaning-making as the "colonisation of the lifeworld."[4] Digital technologies have now transformed this into a colonisation of both the geopolitical world and our inner world. With dominance over minds, external violence can first be supplemented and then, if possible, replaced by manipulation. The source of power for this process is each individual's personal data, which is extensively collected, linked to profiles and behavioural predictions, and extrapolated. These profiles are also increasingly becoming the basis for manipulating human behaviour (nudging).

The key players in AI development are even warning of the potential dangers of humanity being wiped out by uncontrollable AI. In addition to Elon Musk and Sam Altman, Geoffrey Hinton has recently sounded warnings. He believes that the worst-case scenario, in which a future super-intelligent AI turns against humans, has a 10 to 20 percent probability of occurring.[5] Given their equation of super-intelligence with humans losing control over this highly complex technology, Hinton and other experts demonstrate an astonishing willingness to take risks. Would an engineer

4 Habermas, J. (1984) *Theory of Communicative Action*: *2 vols* (Boston: Beacon Press).
5 Most recently in August 2025: Egan, M. (2025) "The 'godfather of AI' reveals the only way humanity can survive superintelligent AI". *CNN*, 14 August.

build an aeroplane with a high probability of crashing and still bring it to market and fly in it themselves? AI pioneer Hinton was awarded the Nobel prize for the very invention he believes could bring about the end of the world. This is another of the paradoxes of this technology.

Future

In this world, nothing is certain except death and taxes.
Benjamin Franklin

The thesis that the future is open entails that the future is not fixed. It is open in the sense that there are many possible future developments. Even though we are hardly in a position to shape it completely, it is clear that our decisions and actions influence the future just as much as unpredictable events. Black swans may appear unexpectedly but the human freedom to shape the future on the basis of our decisions today also ensures the future remains open. The future is therefore neither completely indeterminate nor deterministically predetermined. It is a space of possibilities. Which future reality emerges from this and which new spaces of possibility arise also depends on the decisions that people make, or do not make, today.

The openness of the future can also trigger fears. The need to reduce uncertainty drives the urge to predict individual behaviour and social, economic or political trends. Predicting the future (whether through oracles or AI) is directly linked to social, economic and political advantages and thus to power. Even in ancient times, predicting the flooding of the Nile Valley or solar eclipses secured power for priests. The dream of understanding and controlling the innermost workings of the world, and thus knowing and determining the course of history, drove the totalitarian political misdeeds of the 20th century.

Today, the future is contested by futuristic technology. The ability to calculate and thus control the future is central to the fantasies undergirding AI development. It secures power for its priests because it holds out the promise of wealth security to human beings.[6] What's more, digital ma-

6 Musk and Altman have prognostisised a rosy future in which human labour will be redundant: https://fortune.com/2026/01/19/when-does-elon-musk-say-work-will-be-optional-and-money-will-be-irrelevant-ai-robotics/.

nipulation can increase the probability of algorithmic predictions coming true: a magical technique of domination that produces self-fulfilling prophecies, thereby making the future predictable and controllable. Or so it would seem.

The promises of an all-powerful AI are being used to stage a massive deception. AI fails to deliver on the promise of making the future predictable and controllable for two reasons. Firstly, because the future is not predictable, and secondly, because even if it were, AI would not be able to predict it. The promise of salvation in the form of a future optimised by machines will fail in the face of an open future.

On the one hand, there is the promise that AI will become super-intelligent through the limitless access to data and commensurate increase in the problem-solving performance of computers, surpassing humans in all respects. However, all existing computers and all currently feasible computer programs are based on classic computer architecture, which has been in development since the end of the Second World War. AI systems therefore continue to be subject to all the theoretical limitations postulated almost 100 years ago by mathematicians such as Kurt Gödel (incompleteness theorem) and computer pioneers such as Alan Turing (halting problem). It is unlikely that even quantum computers will overcome these limitations.

But unpredictability is also omnipresent in our lives at the microscopic level. As Heisenberg's uncertainty principle demonstrates, it is impossible to know both the exact position and the exact momentum of a particle at the same time. Quantum physics has shown that the universe is fundamentally unpredictable. Even if Einstein was loath to admit it, God does play dice! So what makes us believe that predictability can work on a macro level? The future is, in a very real sense, not predictable on any level, but open. And not because we lack tools or knowledge, but because it cannot be predicted by its very nature, no matter how much data we collect. Technological determinism, however, seeks to convince us of exactly the opposite.[7]

7 For example: Kelly, K. (2016) *The Inevitable* (New York: Viking); see also: Kurzweil, R. (2005) *The Singularity is Near* (New York: Penguin Publishing Group).

The human factor also stands in the way of the predictability of the future, at least as long as humans are capable of making free decisions. The ability to start something new on one's own initiative is described by Immanuel Kant and Hannah Arendt as the core of human freedom. For Arendt, it is based on natality, i. e. our being born into the world as unique individuals capable of making unique decisions. With every human being, something unmistakably new comes into the world. And every human being is capable of making radically new beginnings. This is also evident in the individual's entry into the political sphere, which enables collective political action – a process Arendt describes as a second birth. What we decide politically also has a decisive influence on the future.

The open future, in which the unexpected can happen at any time, makes the use of a machine as the sole determining technology for coping with contingency and shaping the future pointless and dangerous. Pointless because the future cannot be calculated. Dangerous because it pretends to offer false security, when in fact only statistical profiles of the future are calculated from past data.[8]

The unpredictable factor X[9] makes it imperative to strengthen the capacity of humans to anticipate the future in order to assess the consequences of their actions in the future, in order that they may act responsibly in the present. The philosophical tradition calls this capacity reason. ChatGPT and other large language models (LLMs) may have gathered the knowledge of humanity, but they represent only the average of this knowledge. The models are far from rational insights. Even the ever-necessary innovation, the invention of new things to overcome challenges and solve problems, will remain unattainable for AI for the foreseeable future. The synapses of billions of people are connected in billions of different ways. There is endless potential for new ideas and possibilities in each individual, multiplied by the possible combinations with other people. Thanks to their freedom, humans are the real drivers of innovation.

8 The project of accurately calculating the future is illusory for other reasons as well: even if computing power could be increased immeasurably, the idea of being able to accurately calculate the future based on knowledge of the position and charge of every atom in the universe would still be absurd: to do so, one would have to completely recreate the universe as a simulation.

9 Jonas, H. (1985) *The Imperative of Responsibility: In Search of an Ethics for the Technological Age* (Chicago: University of Chicago Press).

The future reality is also shaped by the decisions we do not make because we increasingly allow automated decision-making systems to make them for us. And the technology that promises to calculate and optimise the future based on existing data is taking on more and more areas of decision-making in business, society and politics. What optimisation, i. e. improvement, actually means cannot be answered without knowing what is good. Digital corporations want to make this decision for us and in the process are attempting to eliminate democratic politics. In reality, the aim is to make the future predictable so that decisions and measures can be derived and legitimised from calculation in the present, from which they benefit. This offers supposed security of planning, but it will cost human freedom. Instead of optimising ourselves to death, blindly trusting in technology, we should instead make every effort to cultivate our own social and democratic skills so that we can continue to cope with the far-reaching consequences of modern technology and the occurrence of the unexpected in the future.

Deception

In the world of the mind, only those who deceive themselves are deceived.
Søren Kierkegaard

Machines are not humans, and humans are not machines. But at the heart of the best-selling AI is the anthropomorphisation of machines, motivated by the eponymous Turing test. Broadly speaking, it purports to demonstrate that intelligence is what humans cannot distinguish from human intelligence. The deception is a projection: deceived by the responses of the seemingly intelligent system, we attribute human intelligence to it. And by extension, humanity. In the strict sense, however, machines cannot be intelligent. AI systems cannot think, reflect, weigh up options, judge, state reasons or examine moral positions. Nevertheless, their creators program them to use these words and, by feigning an "I" perspective, to simulate a subjectivity that they cannot have. In addition to individuality, they also lack reason and judgement, imagination, motivation and the capacity for love – all special human abilities and attributes that enable us to explore the world and, to a certain extent, to constantly reshape it.

By the same token, humans are not algorithmic, information-processing machines. As living beings, we have body-centred reason based on sensory perceptions that enables us to recognise our environment, commu-

nicate and interact with other people, and develop judgement and imagination. Our senses and emotions play an important role in this.[10]

While AI systems calculate meaning in a mathematical space according to the laws of probability, human thinking and decision-making, when based on reason, imagination and judgement, operate in a space of reasons and justifications based on sensory perceptions. These different spaces are subject to their own logic, which, when taken to extremes, are incompatible and come into conflict. Currently, the open capacity of humans is being increasingly restricted by the widespread use of AI-driven decision-making machines. Human capacity, in other words, may atrophy, which in turn, diminishes the prospects for a free and self-determined life in a future world.

The belief in the omnipotence and omniscience of AI, which is becoming increasingly widespread and is being purposely fuelled by the tech giants, is thus a dangerous illusion. For its promises are impossible while the dangers are very real. This AI illusion is particularly dangerous because AI-driven digital technologies already wield enormous power today. They threaten democracy and self-determination worldwide and are slowly eroding them. Like all historical dreams of omnipotence of the few, if they are realised, they will only result in the powerlessness of the many. They threaten humanity's ability to shape its own future in a self-determined manner in a society geared towards coexistence, and thus our open future itself.

At present, the invocations of a near-future superintelligence that develops consciousness and a will of its own and decides to wipe out humanity can largely be seen as a diversionary tactic, as propaganda that obscures the more important debate about the power that AI already wields today. And, moreover, it distracts from who wields this power and for what purposes. The control problem is currently, at its core, a political problem. But the technical control problem also remains unresolved, and this harbours potential dangers. Given the power of the digital-economic complex, the pressing question arises as to how this power can be controlled democratically.

10 Fuchs, T. (2021) *In Defence of the Human Being: Foundational Questions of an Embodied Anthropology* (Oxford: Oxford University Press).

Power

The fundamental concept in the social sciences is power, in the same way that energy is the fundamental concept in physics.
Bertrand Russell.[11]

Digital technologies have enabled a historically unprecedented accumulation of power in the hands of a few companies and governments. This should have long since resulted in a broad consensus on the need for democratic control of AI technology and the economic and political power associated with it. But instead of decisively regulating AI and big tech, we are allowing a few world-dominating companies to gain and accumulate ever more financial, social and, increasingly, political power. Because this concentration of power is historically unprecedented, autocrats have recognised the potential that AI offers for consolidating their absolute power and have allied themselves with the leading technology companies. They are already using this technology in China, for example, to suppress any stirrings of freedom as proactively as possible and to nip resistance in the bud through comprehensive surveillance. Further stages of AI development will soon offer the potential to restrict basic human freedoms and decision-making powers even more comprehensively. The boundless promise of freedom offered by this technology has, in the hands of powerful companies, turned into its opposite: surveillance and control.

It is important to examine the background assumptions and narratives of the "boomers" (techno-optimists) and "doomers" (catastrophists) among AI experts, because they are a key factor in the power that the digital-economic complex is gaining. Cultural and philosophical concepts and set pieces have played a decisive role in the triumph of digital technologies from the very beginning. From the outset, Silicon Valley ideology was linked to the Californian counterculture of the 1960s, from whose promise of freedom it long drew its legitimacy. Today, self-proclaimed libertarians have exaggerated these promises of freedom beyond recognition and believe they can fulfil them by overcoming humanity, promising immortality and colonising the entire universe. To achieve these goals, they first aim to destroy democratic states, which they see as obstacles on

11 Russell, B. (1938) *Power: A New Social Analysis* (London: George Allen & Unwin Ltd), p. 10.

their path. Leading actors indiscriminately draw on philosophy and cultural history to promote a vague amalgam of futuristic faith in technology, pop-culture-charged science fiction and anarchic individualism using the same algorithms of the excitement economy with which they simultaneously earn vast fortunes and undermine democracy.

Putin's insight that AI enables world domination has now been transferred to the question of truth. Whoever determines what truth is, rules the world. AI systems, which cannot develop any understanding of truth because they lack a reference to the world, are ideally suited to generate and disseminate "alternative truths" in order to exert political influence. We have long been in a crisis of truth, triggered by lies aided and abetted by AI systems. AI is therefore not only precipitating a technical revolution but a philosophical one. Through interaction with AI systems in all areas of life, our humanity is being challenged and questioned on all levels. In Kant's sense, we should respond to this upheaval with a revolution in our way of thinking,[12] but we are still in the very early stages of this process. We urgently need to redefine and secure the conditions that make a free and self-determined life possible.

The fact that AI systems are not and cannot be autonomous and intelligent in the human sense does not detract from the power that their operators can gain and the consequences that result from their use. These machines do not need consciousness to cause massive upheaval and damage. Another paradox is emerging: we attribute superhuman abilities to machines without their "intelligence" possessing a spark of humanity. We allow such systems to make far-reaching decisions about the way humans live, decisions that we humans can hardly comprehend. A realistic assessment of what these machines can and, in all likelihood, will never be able to do is therefore required. (See *Chapter 4 – Limits of Technology*)

The fact that the power of digital corporations is only slowly becoming apparent also has to do with the fact that they occupy a new space: cyberspace, the virtual space of seemingly infinite mathematical number sequences, formulas and possibilities. If power is the ability to influence the behaviour of other people in one's own interests, then today we must examine above all the influence that digital technologies exert on

12 This is how Immanuel Kant refers to his epistemological "Copernican revolution" in the preface to Critique of Pure Reason.

our thoughts and feelings. Whereas Max Weber defined power as "every chance to assert one's own will within a social relationship, even against resistance,"[13] in our time power must be seen above all as the ability to manipulate the will of others in such a way that one's own interests are asserted *without* resistance from others. Autonomy is undermined on the one hand by misinformation – misinformed decisions are not free decisions – and on the other hand by emotional, polarising appeals to the target subjects, which generate fear and lead to bad decisions.

Legitimate power can only come about through voluntary and informed consent based on reliable information and free insight. In the digital age, power is increasingly taking on the character of network power, an aspect that Michel Foucault once described as the "dispositif" of power: "The dispositif is the network that is woven between these elements."[14] Global networking has given rise to a kind of world brain in the form of the AI-driven internet, but it is by no means a mirror of humanity. At best, it is a distorting mirror. It is programmed for greed and the urge to dominate and subjugate. It is inscribed with algorithms as the decisive dispositifs of power, serving an economic system of surveillance, control and profit maximisation.

The markets' billion-dollar bets on companies such as OpenAI, which promise the imminent achievement of AGI are not based on the fantastical to fanatical visions of the future held by Sam Altman, Mark Zuckerberg, Ray Kurzweil and co., but on the expectation that the actual promises of Big Tech will be fulfilled. And that is the establishment of market monopolies with correspondingly unlimited market shares and profits. Future applications currently being developed and based on "autonomous" stages of development are intended to meet this goal of the digital-economic complex. Everything else is subordinate to it. Resistance along the way is being brushed aside with unrealistic and impossible promises, millions invested in political PR and, increasingly, through threats and blackmail.

This has been possible since Big Tech has been openly striving not only for economic but also for political dominance. To this end, key players

13 Weber, M. (1963) *The Sociology of Religion* (Boston: Beacon Press).
14 Foucault, M. (1980) "The Confession of the Flesh", in C. Gordon (ed.) *Power/Knowledge: Selected Interviews and Other Writings 1972–1977* (New York: Pantheon Books), pp. 194–228.

have allied themselves with autocrats, kleptocrats and enemies of democracy to form an alliance that is paradoxical only at first glance.

Politics

Reporter to Zohran Mamdani: *Are you affirming that you think President Trump is a fascist?*
Trump: That's okay. *You can just say yes. It's easier. It's easier than explaining it. I don't mind.*
Donald Trump on 21 November 2025 in the Oval Office.

In order to understand the political dimension, we examine not only the functional mechanisms of this technology in *Chapter 2 – The enemies of the open future and their dark enlightenment*, but also the ideologies and actors behind the technology. Technological future scenarios are charged with ideas brimming with contempt for humanity. Through these narratives, non-human technology can become inhuman technology. For example, when arguments propounded by advocates of effective altruism are used to offset the rights of people living today against the rights of future generations. The boundless promises of happiness for trillions of future humans who will have conquered space through AI are held to outnumber the happiness of the eight billion people living on Earth today; this can be used to justify the continuation, even the acceleration, of the uncontrolled further development of the very technology that is supposed to make this ludicrous number of future humans possible in the first place through its beneficial consequences. On the one hand, such macabre, supposedly ethical thought experiments distract from the far greater danger that AI already poses today by being in the hands of a few companies whose power has reached or exceeded that of states. On the other hand, these narratives play a central role in the planned triumph of Big Tech. Individual players have long since amassed enormous personal power in addition to their financial fortunes. What they believe, no matter how absurd, can influence and change the lives of billions of people today and tomorrow.

Autocrats and Big Tech alike have a pronounced blind spot when it comes to the damage caused by AI, which is already evident today in ecology, societies and individuals. Or they accept the consequential damage with open eyes in order to achieve their goals, be it maximising profits or attaining complete surveillance and control of humanity. We aim here to identify these dangers and highlight the need for regulation, without fun-

damentally questioning the controlled, sensible use of this technology. Nor will sensible regulation prevent innovation, insofar as it is sensible. We call this regulation *smart regulation*. This does not mean calling for more regulation across the board, but rather focusing legislation on areas that are crucial in terms of power politics. It also means regulating consistently and asymmetrically: the threatening market power of monopolists must be clearly limited, and more scope for innovation and development must be created for small and medium-sized enterprises. Only through clear, forward-looking legislation can the foundation be laid for innovations that preserve democracy and freedom and, ideally, even strengthen them (more on this in *Chapter 5 – Politics, Law and the Digital-Technological/Economic Complex*).

AI is not a neutral technology. When training algorithms, countless value judgements are made, guided by the interests of the developers. The data used contains distortions and biases, which are in turn reinforced by the algorithms. AI is therefore neither neutral nor objective. AI is a highly political technology. Its impact on individuals, societies and states is profound. Any serious discussion of AI, therefore, is simultaneously a discussion of politics.

If freedom is to be secured in the future, people must continue to be able to make decisions in the face of inevitably uncertain predictions about an open future. And they must do so in a way that is as responsible as possible. Without good grounds, without substantiated facts, without thorough consideration and joint consultation, without a transparent and participatory process of deliberation, there can be no responsible decisions, even in the age of AI. Nothing less than the privileging of political action over the interests of the digital-economic complex is what is at stake here.

Partly because the technical control problem remains unsolved and yet massive investments are being made in the AI race, we must act quickly if we care about our ability to act in the future. To answer the question of how humans and democracy can maintain control is very difficult if only because of the lack of transparency surrounding AI and the systems used to develop and utilise it. As the supposed intelligence and autonomy of these systems continue to develop, it will become even more difficult to predict their output and understand how the results were achieved. This is due to the structure of machine learning, which means that not only are the inner workings of the model shrouded in darkness – even for the computer scientists doing the programming – but it can also change

autonomously in this darkness. The goal of the current massive investments is a level of machine autonomy that, according to the developers, can automatically and autonomously control companies, organisations and entire states. But because it is dark inside the black box of AI, it cannot achieve true autonomy. For that is based on insight and free self-determination. The destructive power of this technology for democracy and civil liberties today lies mainly in the fact that it is being developed and used largely unchallenged by democratically legitimised control. As a result, it is not so much the disappearance of humankind that is currently under threat, but rather the disappearance of humanity. Humanity in the sense of humaneness.

Humanity

The perfect must not be the enemy of the good.
Voltaire

The digital-economic complex and the power it wields are diverse[15] and require an interdisciplinary and holistic view in order to develop appropriate countermeasures. That is why a computer scientist, a journalist and a lawyer have joined forces in this book to provide the most diverse analysis possible by looking at the phenomenon from different perspectives. Above all, however, we seek to make concrete proposals on how we can still escape the crash that the above mentioned system conflict is heading towards. Each of the authors has experienced the current transformation first-hand. Whether it be the Big Tech lobby bombardment of European legislators during the creation of data protection regulations and AI regulation, the struggle for independent university AI research on both sides of the Atlantic, or the experience of digital disruption in journalism.

This book aims to help make the complex issues surrounding AI and its impact on our collective future intelligible through explanations and analyses. We hope to enable and empower readers to join the conversation about the possibilities and perils of AI. This means understanding AI systems and their implications (see *Chapter 3*). At the same time, however, it is impossible to go into the detail and depth necessary to cover all

15 See: Nemitz, P. and M. Pfeffer (2023) "The Eight Sources of Power of the Technological-Economic Complex", in *The Human Imperative: Power, Freedom and Democracy in the Age of Artificial Intelligence* (Cambridge: Ethics International Press Ltd).

the issues arising in a single book. The website accompanying the book, https://www.open-future.ai/, provides further reading, technical tools and up-to-date commentary on relevant topics.

We aim not merely to describe the problem, or assert that there are alternatives to Big Tech, nor to issue an empty call to arms, but rather to offer concrete examples and suggestions for action both in the book and online. Critical thinking and responsible action must go hand in hand if we hope to effectively counter the enormous power of Big Tech and autocrats.

The development of AI is not inevitable; like the future, it is still open. At least for now. So what can be done? In the following chapters, we elaborate on our proposals in more detail. First of all, the systemic imperatives of AI and democracy appear at first glance to be as contradictory as the early capitalist markets and democracy once were. Just as the predatory capitalist markets of the industrial revolution were domesticated in the social market economy to be compatible with human rights and democracy, AI must be domesticated so as not to destroy democracy and human rights. Author Giuliano da Empoli sees the current unchecked growth in the power of this complex as heralding "the hour of the predators, who are unabashedly using this power to establish monopolies and create a new world political order."[16] In view of this development, politicians need the courage to resolutely apply the principle of responsibility[17] as well as the precautionary principle in order to protect democracy, which is the target of predatory digital capitalism. The precautionary principle has become part of European constitutional law.[18] At least in Europe, there is no need to reinvent the wheel.

It is not only politicians who are called upon to act. The fight for an open society today is the fight for an open future. And that affects everyone. We can all help shape the future of democracy if we get involved. We can exercise our existing rights vis-à-vis corporations, because the fight for rights is what distinguishes citizens in a free society. And we can contribute our knowledge and opinions, via political parties, professional as-

16 Da Empoli, G. (2025) *The Hour of the Predators* (London: Pushkin Press).
17 In the sense of Jonas, H. (1985) *The Imperative of Responsibility*: *In Search of an Ethics for the Technological Age* (Chicago: University of Chicago Press).
18 § 191 Treaty on the Functioning of the European Union: https://dejure.org/gesetze/AEUV/191.html.

sociations, civil society and other channels, when it comes to deciding on the future through democratic politics. In this book, we discuss how this can be done and what measures are now necessary, given that the EU has already enacted a large number of legal regulations in recent years to curb the power of corporations in digital space.

It is important to harness this technology in the name of democracy, the rule of law and fundamental rights. This is not about hostility to technology, but about innovation for democracy. Innovation that not only creates something new as an end in itself and also serves the common good cannot do without regulation. The question is whether innovation can be created and controlled not despite but thanks to regulation. We believe that the answer, as in the case of freedom, which can only be secured by laws, is a resounding "Yes!" through *smart* regulation.

The digital transformation is associated with a tectonic shift in power. AI will soon be everywhere, calculating and doing everything simultaneously. AI is primarily praised as a tool for creating new value. But value creation must be precisely defined in this case: Who creates what value? How will human, social and political values be changed by this transformation, and how can they be safeguarded?

AI will change the world more profoundly than previous general-purpose technologies such as electricity and the internet have done because it can penetrate and change the innermost core of human beings and has enormous networking power. The systemic conflict between AI and democracy must be resolved, if only because otherwise free thinking and open debate will no longer be possible. The advance of AI agents means that soon all humans will be in constant conversation with AI, spending more time talking to a machine that cannot feel, think or judge, but which feigns these abilities in order to gain trust. Constant interaction with these tools causes depression, anxiety and loneliness and leads to addictive behaviour. Similar to the tobacco industry decades ago, companies such as Facebook have researched these dangers themselves but have not published the results. Instead, they have discontinued the studies, as court documents in the US show.[19] Without question, AI also offers the opportunity to make enormous progress, for example in the development

19 Horwitz, J. (2025) "Meta buried 'causal' evidence of social media harm, US court filings allege". *Reuters*, 24 November.

of medicines and new materials, but also in improving communication between citizens and citizens and the state. For AI to realise its potential to help tackle major issues of the future, this technology must be subject to political control and trustworthy AI must be developed.

Many of the problems that stand in the way of this have long been known. We have also known about man-made climate change for over half a century. It took a long time, but the international community has finally taken steps to prevent a climate catastrophe. And despite the tobacco industry's fight against science, we now have far-reaching smoking bans. We all hope that it is not too late in the case of AI development. Many opportunities have already been missed in taming the power of the digital economy through democracy and the law. This is partly due to pressure from Big Tech's multi-million-euro lobbying efforts. Time is now of the essence because development is proceeding at an enormous pace.

Software is eating the world, said AI pioneer and investor Marc Andreessen. Mankind is not yet on its menu, but democracy – and humanity – already are.

The philosopher Karl Popper, who inspired the title of this book, once said that those who allow the future to be predicted have already given up on shaping it. The future is still open, and we still have the chance to shape AI for a good life and in the public interest. Let's do it quickly and decisively. The elements for this are in place. We discuss them in this book.

1 | Impending system crash – the future of a dangerous illusion

Forward into the past

The dual perception of the future through the idea of AI as a technology that will dominate everything in the future on the one hand, and as a prediction machine that can already be used today on the other, is what makes the impression that there is no alternative to the technology itself and its further development so powerful. But this impression is deceptive. It is deliberately created by people who use the enormous possibilities of digital technologies to conquer markets worldwide and, increasingly, usurp political power. The inevitability of this move towards ever greater automation, convenience and superiority over humans, which means nothing less than disenfranchisement, is a Silicon Valley narrative designed to fool us all.

This development has gained momentum through a new alliance. The union of reactionary autocrats and futuristic Big Tech, which at first glance seems contradictory, is working in two ways to seal off the future: populist politicians stir up longing for "retrotopia,"[20] nostalgia for a past that never existed. AI promises a bright future, but at the same time increasingly limits the scope of possibilities for the future by calculating supposed optimisations based on data from the past, thereby restricting our scope for decision-making. It creates images of the future that arise without any human creativity – even if human creations have been incorporated into the preceding data theft – and at the same time pretends to supplant human creativity. At the same time, it weakens it, because creativity is no longer practised or rewarded.

Techno-futurists share a fundamental communality with retrotopians. While the latter attempt to conjure up a supposedly ideal past, AI pro-

20 Bauman, Z. (2017) *Retrotopia* (Cambridge: Polity Press).

grams extrapolate the future on the basis of data from the past. On this basis, decisions that determine the future are being prepared or made. Democratic participation in what the future will be like is increasingly restricted by automatic decision-making processes. This chimes with the idealised past in which there was less pressure on individuals to make decisions than in today's highly complex world. For in the AI-dominated world, decisions are made by technocratic and autocratic leaders. Authoritarianism combined with technological relief offers up a seductive chimera of escape from responsibility.

Both Big Tech and retrotopic authoritarians seek to block us from actively shaping the future in order to impose their own vision of undemocratic and algocratic digital capitalism. Democracy, the rule of law and fundamental values stand in their way. To this end, they have created a powerful ideology that attempts to dominate public debate with antidemocratic narratives. The role of this ideology in enforcing the plan for a technocratic-authoritarian seizure of power should not be underestimated. It is intended to help design social models and present them as desirable, which the tech elites then try to enforce in social discourse.

In the following, we show why the ideology of the alliance between futurists and reactionaries is not only based on false assumptions, but is also dangerous. It is false because it is based on the misconception that the world and the future are completely predictable – and therefore controllable. But the future is not predictable. Rather, we must always expect the unexpected, which presents us with ever new challenges.

It is dangerous because it undermines and destroys the prerequisites necessary for survival in order to cope with inevitable randomness and an open future: human thinking, speaking and joint action. Human in two senses: executed and accounted for by humans, and committed to human values and humanity.

Ideology

As noted above, the alliance between Big Tech and right-wing authoritarianism is at first glance surprising. While the ideology of Big Tech stands for exponential acceleration of technological progress, right-wing populism stands for the promise not only of slowing down but of reversing moral, legal and political progress.

The combination of the two currents creates the illusion of a clear-cut social world that can be created at the touch of a button (i. e. technologically). Once again, regression and reversal are offered as a way out of the excessive demands of a self-propelling modernity.

Digital ideology is based on the assumptions of solutionism, which contends that all problems can be solved through the algorithmic processing of big data.[21] It propagates the transfer of responsibility to supposedly autonomous machines, thereby creating a diffusion of responsibility that it sells as relief. For solutionism, the boundless predictability of the world and thus of the future is possible. But this assumption is only valid in a deterministic worldview. It denies the objective chance in nature and human free will.

Despite all the patterns and laws in history, what exists is chance, which brings the unforeseen and constantly presents us with new challenges in a contingent world. A culture necessary for coping with contingency, which requires critical, independent-thinking responsible people. These very abilities are simulated by AI without it being capable of genuine thought or responsible action. The advance of automated decision-making systems is causing a comprehensive loss of skills indispensable for coping with contingency. Uncontrolled AI applications thus undermine the most important human resource for coping with the future through continuous de-skilling.

Classical physics was based on the assumption that the universe is strictly deterministic. According to this assumption, everything that happens has a cause to which it can be traced back. Conversely, knowledge of the causes and the initial and boundary conditions also allows the possible consequences to be calculated in advance and thus predicted. As a proponent of classical determinism, Pierre Simon Laplace (1749–1827) developed probability theory. According to Laplace, an intelligence that had all the information about the initial conditions and laws of motion would be able to calculate the future completely. He describes the demon, later named after him, as follows: “We must therefore regard the

21 Morozov, E. (2013) *To save everything, click here: the folly of technological solutionism* (New York: Public Affairs); other elements include cybernetics, Darwinism and game theory. For more on the term “dataism”, see: Nemitz, P. and M. Pfeffer (2023) *The Human Imperative: Power, Freedom and Democracy in the Age of Artificial Intelligence* (London: Ethics Press).

present state of the universe as the result of a previous state and as the cause of the state that will follow. An intelligence that at a given moment knows all the forces with which the world is endowed and the present position of the bodies that compose it, and that would be comprehensive enough to subject this knowledge to analysis, would include in the same formula the movements of the largest celestial bodies and those of the lightest atom. Nothing would be uncertain for it; the future and the past would lie clearly before its eyes."[22]

Laplace's demon was supplanted by the discoveries of quantum mechanics a hundred years ago. Quantum mechanics provides the basis for modern technologies such as GPS navigation, lasers and LEDs. The entire electronics industry, from computers to smartphones, is also based on the quantum mechanics of semiconductors.

According to quantum theory, the future is a space whose complete information cannot be known by anyone or anything. In contrast, digital ideology spreads the illusion of near-perfect prediction and thus several false versions of the future: it creates an illusion of certainty, fuels wishful thinking, stifles criticism and designs a fake future in the same manner as astrology.

Due to its apparent inevitability, it also causes its critics to think catastrophically. For if the superiority of machines is a goal of development determined by natural law, the demise of humanity is a foregone conclusion. Although the belief in total predictability through AI is false, it nevertheless causes an operational disruption of the future. Digital ideology envisions AI as the dominant operating system. If we follow along and put the future on AI autopilot, the essence of humanity will be lost. The Laplacian demon of omnipotent predictability, which fuels the fantasies of the tech elite, cannot exist since it is impossible to calculate chance. But it is this Laplacian spirit that dominates the power ambitions of the actors behind AI and feeds the narrative of AI's purported superiority.

The illusion of total predictability creates a deceptive sense of security and, at the same time, fatalism. It contributes to the loss of the cultural techniques of coping with contingency that can help to shape the open

22 Laplace, P. (1951) *A Philosophical Essay on Probabilities* (New York: Dover Publications).

future in the most human way possible. On a deterministic worldview, creativity and innovation cannot be understood. Even if they integrate random noise and thus attempt to imitate creativity, machine calculations of the future cannot replace human intuition. The expectations of AI prophets are running high when it comes to the future connection between AI and the human mind. A prime example is Max More, founder of the so-called Extropian movement, whom Google chief technologist Ray Kurzweil calls the "great philosopher": "Through the accelerated development of technological improvements and augmentation, we will change so much and so suddenly that we cannot really imagine today what life will be like once we have reached the level of superhuman intelligence."[23] Something is being promised that we simply cannot imagine today, but the reason for this is being concealed: because the imagined self-deification of humans through technology is simply impossible.

The claim that the unpredictable can be predicted is merely a misconception based on a belief in the seemingly limitless capabilities of this technology, which is supposed to develop inexorably and at an exponential rate. Furthermore, computer programs have limitations that are determined by their own logical and mathematical foundations. Contrary to the claims of solutionists, these limitations also prevent AI from solving all problems. More on this in the chapter on the limits of technology.

The fundamental assumptions of eschatological AI futurists are therefore incorrect. The future is neither predictable, nor are algorithms capable of calculating it – even if it were deterministically fixed. This makes following this ideology even more dangerous. Countless, ever-changing narratives are put out into the world by Big Tech companies in order to achieve the goals of the operators of these technologies. They are more concerned with paving the way for machines to take over power than with machines taking over power themselves – a scenario in which the sorcerer's apprentices would themselves become obsolete and would no longer be able to control their own companies.

The narrative groundwork for this planned takeover is laid by metaphysical manifestations of digital ideology, which take the form of transhumanism or dataism as a technological religion of salvation. Just as au-

23 Freyermuth, G. S. (2006) "The Future Hasn't Started Yet: An Interview with Max More". *Parapluie*, *23* (summer).

tocrats constantly disparage and speak ill of democracy, so digital ideology disparages and speaks ill of humans: its carbon-based wetware is physically inferior to the silicon-based hardware of AI. Problem-solving in a highly complex world can no longer be entrusted to such a flawed product of evolution. Humans as physical beings are disparaged as flawed mutants until, in the end, the final solution to the question of humanity appears desirable: replace with AI.[24]

The same fate befalls democracy, which is already an enemy because of its slowness, which gives people the opportunity to think and communicate. Accelerationists is how the followers of a doctrine who see the acceleration of capitalism driving inexorably towards singularity describe themselves. Like most of the set pieces that make up digital ideology, accelerationism posits that this development is inevitable. By the same token, it is also inevitable that humans or democracy will be treated as obsolete models. In order not to appear as sinister prophets of doom, some representatives sell this inevitability as a victory for freedom, even though it would obviously be its nemesis. This was exemplified years ago by tech billionaire, Palantir founder and current shadow president of the United States, Peter Thiel, when he declared that, in his view, democracy and freedom were no longer compatible.[25] J. D. Vance has now taken up the cause of implementing this idea of restricting democracy in favour of the freedom of a few billionaires, claiming to defend freedom of expression while openly calling for attacks on science and universities (see *Chapter 7 – The role of universities and the media*).

Systemic conflict

The conflict between the algorithmic logic of unregulated AI and the principles of democracy is based on a fundamental system conflict: on the one hand, there are the possibilities of centralised control, mathematical-statistical calculation and algorithmic efficiency offered by AI. On the other hand, there are the principles of human dignity, the uniqueness of each individual, freedom and self-determination, critical reflection and plurality. While AI describes people in data sets based on comparisons with countless other personal data, democracy protects fundamental individual rights. These contradictions are not only theoretical in nature,

24 Ibid.
25 Thiel, P. (2009) "The Education of a Libertarian". *Cato Unbound*, 13 April.

but lead to the distortions and destabilisation that can be observed worldwide.

AI enables uniform and controlled governance. Big Tech has now enforced this trend against the open and decentralised potential of digital technologies. Democracy aims at participation, sharing and control of power. AI enables data concentration as well as enormous calculations and decisions in a matter of milliseconds. Democracy requires weighing up arguments, deliberation and a transparent decision-making process that allows for corrections and cannot be completed in seconds. Just as human thinking and AI are not the same, democratic decision-making processes cannot be automated by algorithms. Automated decision-making is the opposite of free decision-making.

Automated decision-making

Algorithms already select what we see on the internet every day, while the internet has become our primary source of information. LLMs such as ChatGPT and image generators such as Midjourney and others are having a growing influence on digital communication, which in turn has become our primary form of conversation. Beyond the purely digital world, algorithms recommend where we should go and who we should meet; AI systems may soon be writing large portions of the texts we read and generating the images and videos we see. But that's not all: algorithms select job applicants, decide on loans, insurance policies and whether prisoners are released early from prison or not. Perhaps they will also soon decide who goes to prison?

The spectrum of algorithm-based, algorithm-driven and algorithm-determined decision-making systems is used everywhere: from healthcare to the financial sector, policing, government action and the judiciary. Many applications are unproblematic as long as they do not violate fundamental rights. However, whenever human rights are affected, these systems must be scrutinised before they are used. In order to rule out the risk of discrimination, ensure participation and fulfil transparency and information obligations, they should be certified by independent third parties. However, this is not yet the case. Above all, access to justice must be ensured in the case of automated decisions. Those affected must be given sufficient, effective, rapid and adequate legal protection and be able to challenge decisions made by algorithms. The human right to human decision making must be effectively enforced.

AI systems not only take decision making off our hands; they are beginning to dictate the rules of the game for future decisions. They are creating new realities in which they not only predict how we will behave, but actively steer us in that direction. In fact, not only are we training these systems with our behaviour for free, but we are also increasingly being nudged by constantly improving recommendation algorithms, i. e. pushed to take certain actions. Since this happens subliminally, it is almost imperceptible. Behind it lie the hidden motives and opaque goals of owners and financiers. Based on personal data, they can create profiles that maximise profits through extensive behavioural predictions. Increasingly subtle forms of manipulation are being used, known as hyper nudging. In plain language: the massive restriction of human freedom of choice through deep manipulation.

Though such AI applications are reported on daily, too little is being done to curb their negative effects. Public discourse on AI – as is currently the case with many other topics – is increasingly polarised and overly simplistic. Sometimes the blessings of technology are lauded, we need only embrace it openly enough. Sometimes the extinction of humanity is predicted: if development continues at the same pace, we are doomed. It appears that the binary principle of the digital world has conquered all thinking. Zero sum, black or white, *tertium non datur*. Between these two poles, politics largely remains in a state of paralysis. Between the salvation of humanity and the apocalypse, there seems to be little room for a sober and realistic assessment of this powerful technology. Curiosity about the latest gadgets is mixed with horror at the sight of sex robots or killer drones. Diffuse fears of job losses alternate with enthusiasm about the seemingly endless economic growth potential that AI is supposed to portend. The maxim here is: do first, think later. Many applications are brought to market prematurely and further developed free of charge by users, who continuously train them with their data and in return have to bear the risks and side effects. At the same time, we are becoming accustomed to the new convenience that many applications bring while neglecting to consider the costs that individuals and society will have to pay in the long term. Out of laziness and cowardice we are regressing into socio-cultural immaturity, which is therefore our own fault. Immanuel Kant pointed the way out: we need the courage to use our own understanding.

Shadow phenomena

Enlightenment today faces a problem: similar to the dangers of nuclear energy or tobacco smoking, AI is a technology whose function, effects and risks are initially invisible. Unlike mushroom clouds and cancerous tumours, the media society lacks images that make the extent of the threat to democracy and social cohesion tangible. What's more, public attention is distracted by the same mechanisms whose potential threats we should actually be focusing on.

We must therefore focus our attention on a shadow phenomenon that is easily overlooked. AI is too complex to be adequately represented in the short texts and images of the attention economy. Its impact extends over longer periods of time than the ever-decreasing attention spans of users in today's digital democracy, and the planning horizons of politicians, which do not extend beyond election cycles. In surveys on pressing political issues, AI and digitalisation either do not appear at all or appear at the bottom of the list. Ironically, the very technology that distracts public attention from the important issues appears at the bottom of the list of problems that should be solved politically. In the case of AI, opinion poll-driven politics sets its priorities, too timidly or not at all. If only out of self-interest, democratic politicians should be much more concerned with this development as they themselves are already among the endangered species. AI-optimised radicalisation and motivated hate attacks on the internet are leading more and more elected representatives to withdraw from politics.

In the super election year of 2024, AI systems were used worldwide to weaken democracies. Generative AI produced massive amounts of misinformation in texts, photos and videos to manipulate voters' opinions and thus the elections. Artificially generated and deceptively authentic posts already account for more than half of the content available on the internet,[26] and their share will continue to grow: experts say that soon more than 90 percent of the content on the internet will be synthetic.[27]

26 Brian Thompson et al. (2024) "A Shocking Amount of the Web is Machine Translated: Insights from Multi-Way Parallelism". *ACL Findings*, DOI: 10.48550/arXiv.2401.05749.
27 Europol (2022) *Facing Reality? Law enforcement and the challenge of deepfakes*. An observatory report from the Europol Innovation Lab, (Luxembourg: Publication Office of the European Union). DOI: 10.2813/158794.

Synthetic content generated by AI is barely distinguishable from photos, videos or texts produced by professional media and journalists.

The world knowledge that is accessible on the internet, which shapes our perception of the world and forms the basis of our opinions, is generated by machines that have no connection to truth or the real world. What's more, AI knows no responsibility. This also seems to apply to some of the powerful corporate leaders behind the companies that develop and use AI. Otherwise, they would not be fighting against every attempt to create legally binding rules. Despite all the analyses and warnings, we must now conclude that Big Tech has taken control of democracy and the free economy worldwide, and that digital corporations are now in the process of destroying both.[28]

The flood of artificially generated texts and images with enormous suggestive power is leading to a loss of trust and is undermining democracy today. Democracy is based on free and informed decisions that are based on open and unmanipulated processes of opinion-forming. A decision based on false or missing information cannot be a free decision. The autonomy of citizens, like the legitimacy of democratic power, requires unhindered general access to verified and reliable information. If they are misled, the entire construct of legitimacy, democratically legitimised exercise of power and control of power becomes obsolete. And once the truth has been destroyed, and with it the trust that is already measurably declining due to the increasing use of AI-generated news, images and videos, life becomes very comfortable for the tech bros and their political friends. For then, amid the general chaos, they can promise to deliver the order that their cold algorithms execute (more on this in *Chapter 6 – Democracy only with deliberation*).

However, the manipulation of public opinion and democratic elections through targeted misinformation and disinformation, both from within and without, is not the only threat AI poses to democracy. In order to form a picture of the comprehensive systemic conflict between the current unchecked development of AI on the one hand and the fundamental components of free societies on the other, other aspects of AI systems must be considered, even if they are far less obvious.

28 See also: Andree, M. (2025) *Big Tech Must Go!: Digital Giants are Destroying our Democracies and Economies – We Will Stop Them* (New York: Campus Verlag).

Loss of control

The loss of control over the future development of AI is another issue that is often ignored when the focus is solely on the blessings of AI. It would be wrong to completely disregard it because it is used by the doomsayers of digital ideology.

As early as 1951, Alan Turing warned of the possibility that machines could surpass human intelligence. In 1965, mathematician I. J. Good predicted the development of ultra-intelligent machines. Developments in AI research could lead to humans losing control over machines: "An ultra-intelligent machine is defined as a machine that can far exceed the intellectual capabilities of any human being, no matter how intelligent. Since the construction of such machines is one of these intellectual capabilities, an ultra-intelligent machine can build even better machines; this would undoubtedly lead to an explosive development of intelligence, and human intelligence would lag far behind. The first ultra-intelligent machine is therefore the last invention that humans will ever have to make."[29] Billions are currently being invested worldwide in the development of such a machine in a race between companies and political systems. A critical assessment of the consequences will conclude that AI that is beyond human control cannot be ruled out. According to Good, it would only be controllable if it remained humble towards humans: "Provided, the machine is docile enough to tell us how to keep it under control." Provided that. To prevent the emergence of such uncontrollable AI, on whose mercy we would be dependent, the problem of control and containment must be solved.

To mock humans as wetware and flawed mutants and to want to replace them with a supposedly superior silicon-based intelligence fails to recognise that mistakes are a prerequisite of scientific progress. Karl Popper emphasised the "the fact that we learn from our mistakes and not from the accumulation of data".[30] It was also Popper who deduced the openness of the future from a simple thought experiment: if the future were determined, we could calculate today what will be discovered tomorrow.

29 Good, I. J. (1965) "Speculations Concerning the First Ultraintelligent Machine", in *Advances in Computers*, vol. 6: 31–88.

30 Popper, K. R. (1945) *The Open Society and Its Enemies, vol. 2: The High Tide of Prophecy: Hegel, Marx, and the Aftermath* (London: Routledge & Kegan Paul), p. 362.

But that is a logical contradiction. "Therefore, we cannot predict the future course of human history."[31]

For Popper, knowledge is therefore always fallibilistic. This means that when we talk about knowledge, we are talking about empirically proven and well-founded hypotheses that are true if and only if they can in principle be improved or refuted.

If it is the case that the future is fundamentally open, we must also preserve and cultivate the open capacity of human beings to respond to contingencies and surprises again and again.

Elon Musk himself has warned how important it is to maintain control over future AI developments. In 2025, he had Tesla shareholders approve his plan to deliver a "robot army" of one million humanoid robots on his Optimus model with an unprecedented compensation plan: if he achieves all his goals, he would be worth approximately $1 trillion. He justified this unusually high sum, even for Tesla shareholders accustomed to science fiction, with the argument that he would not feel comfortable if he did not personally have control over this army of robots, which, according to his pronouncements, is supposed to conquer the world from 2026 onwards. Musk left open, or rather vague, exactly what Optimus is being optimised for. He backed up his promise that Optimus would be the "greatest product of all time" with confused visions. At times he announced that Optimus would be an "incredible surgeon," at other times he promised that the robot army would end poverty in the world and prevent people from committing crimes. For their part, Tesla shareholders have not prevented Musk from pursuing these plans and have approved the $1 trillion he says he needs to solve Optimus' control problem. In his own way, of course, by retaining control of the company. It remains to be seen whether this army will be stopped before Musk leads it into a battle against democracy.

31 Popper, K. R. (1957) *The Poverty of Historicism* (London: Routledge & Kegan Paul, 1957), pp. xi–xii.

AI works – but not for democracy at present

The success of many current anti-democratic ideologies is based on the factual success of digital technologies, which are becoming established worldwide and mostly originate from the development laboratories of a few companies.

”The amazing thing is that technology works,” Martin Heidegger noted when he posed the question of technology[32]. Heidegger considers the seemingly self-evident and smooth functioning of modern technologies to be part of a deeper essence of technology that he examined. Precisely because it seems to function so smoothly, Heidegger warns against blindly pushing ahead with technology and advocates reflecting on its essence. Only in this way, he argues, can we succeed in finding a freer relationship with it. Digital thinkers, however, draw the opposite conclusion: because digital technology seems to show us reliably what is, it should also tell us what should be.

In the case of AI, the fundamental ambiguities of technology leave us with the dilemma we face today: its development seems to surpass the thinking, speaking and acting of humans that gave rise to this technology in the first place. But if it really does surpass humans, the question of how a technology that is believed to be capable of autonomous development and self-improvement can still be controlled by humans becomes the most important question of our time.

Throughout human history, the goal of technological development has always been to purposefully shape and change the environment to meet human needs. It was intended to help solve problems that humans encounter in their daily lives. In the future, it could not only solve the problems that humans face, but also the problems that humans themselves represent. In other words, the problem of humans in the sense of the genitive object.

Today, a purely economic model drives technological development. Through the spread of disinformation, bubble formation and forms of subtle manipulation of opinion-formation, it has caused massive individ-

32 Heidegger, M. (1977) “The Question Concerning Technology”, in *The Question Concerning Technology and Other Essays* (New York: Harper & Row), pp. 3–35.

ual and social problems that have contributed significantly to the polarisation of the public sphere and the destabilisation of democracies.

2 | The enemies and their dark enlightenment

What is the truth, and where did it go?
Bob Dylan, Murder Most Foul

In a book that defends democracy and advocates understanding, compromise and tolerance, talking about enemies might appear to be a contradiction. The friend-enemy dichotomy as a fundamental concept of politics can be traced back to the constitutional lawyer and philosopher Carl Schmitt. The mastermind of the Third Reich, whose thinking is currently experiencing a remarkable resurgence among tech barons and right-wing extremists in the US and Europe, declared the distinction between friend and enemy to be the defining feature of politics. "The specific distinction to which political actions and motives can be traced is the distinction between friend and foe … Political thinking and political instinct prove themselves theoretically and practically in the ability to distinguish between friend and foe. The high points of great politics are also the moments when the enemy is recognised with concrete clarity as the enemy."[33]

Schmitt sought to serve his ideal of a homogeneous state by allowing the sovereign to declare people enemies through authoritative decree, in order to ultimately remove them from the state structure. "Every real democracy is based on the principle that not only are equals treated equally, but, as an inevitable consequence, those who are not equal are not treated equally. Democracy therefore necessarily requires, first, homogeneity and, second, if necessary, the elimination or destruction of the heterogeneous," Schmitt wrote in 1923. The National Socialists, whom Schmitt joined, then demonstrated in 1933 how the elimination and destruction of the heterogeneous elements was carried out in practice.

33 Schmitt, C. (2007) *The Concept of the Political* (Chicago: University of Chicago Press), p. 26.

From his exile in New Zealand, the philosopher Karl Popper spoke of the enemies of open society as those ideologues who propagate a rigid, closed worldview and cling to a utopian, supposedly perfect vision of the future instead of open criticism and constant reform. Popper wrote his work *The Open Society and Its Enemies* in New Zealand during the Second World War, and described it as his contribution to the war effort. Nazism and Stalinism regard race and class struggle as the all-determining driving force of history and derive from this the inevitability of human development. According to Popper, this takes them beyond modernity, for which an open future is constitutive, back to tribal thinking, in which the fate of humanity is guided by higher powers.

For Popper, therefore, the main enemy of the open society is historicism, the idea that history follows a predetermined law and is heading towards a specific goal. This pattern is crucial for him: the idea that human history follows a fixed plan, that it is determined by a higher power or a law of nature that must be recognised in order to submit to the inevitable and promote it, is, for Popper, the basic assumption of totalitarian regimes. On the one hand, it replaces with supposed security the fears that an open future can trigger. On the other hand, the supposedly recognised courses of history are intended to legitimise the power of those who allegedly pursue these goals, which totalitarians claim to have recognised. In the name of the respective higher goal of history, they then stop at nothing in the name of their historical mission. Resistance from those who think differently would hinder the necessary course of history and only delay the advent of eternal salvation. The idea that a higher purpose or a final utopia justifies any means and therefore serves to justify human suffering and human sacrifice along the way and to excuse the perpetrators in advance.

Hannah Arendt, the theorist of totalitarianism, also sees a similar basic pattern in totalitarian ideologies: "The claim to a total explanation of the world promises a total explanation of everything that happens historically, namely a total explanation of the past, total knowledge of the present and reliable predictions of the future."[34] Arendt's description of totalitarian ideology fits the narrative of technological solutionism with astonishing precision. The futurists in Silicon Valley, for whom the emergence of an AGI superior to humans is without alternative and is established as

34 Arendt, H. (1951) *The Origins of Totalitarianism* (New York: Harcourt, Brace & Company), p. 468.

the goal of technological development, also share traits of such a deterministic view of history. From technological determinism to transhumanism, which propagates the replacement of humans by a higher race of cyborgs, to extropianism and singularitarianism, eugenics, futuristic rationalism, effective altruism, longtermism, and cyberlibertarianism, they draw on a variety of theories. The common denominator of these sometimes abstruse teachings, which we will discuss in more detail in the following pages, is, on the one hand, the devaluation of humanity and, on the other, hostility towards democracy. Behind the neo-right-wing Californian ideology lies a consistent theory and practice of misanthropy.

Crucial to the political impact of this ideology is the connection between technological futurism and the radical right-wing/Alt-Right movement in the USA. Neo-reactionary ideologies have established a connection with radical evangelicals and crypto-Catholics. The friend-foe scheme is openly propagated. Those who oppose technological development, for example through regulation, are identified in the milder case as brakes, and in the more severe case as antichrists who prevent the advent of the technological plan of salvation. For amateur theologian Peter Thiel, who borrows from Carl Schmitt when referring to the apocalyptic revelations of the Bible, the enemy is either the UN, the EU or Greta Thunberg. One of the richest and most influential people in the world, who owes his wealth to the unbridled, because unregulated, development of digital technologies, is targeting the United Nations and the European Union, founded in the wake of two world wars, because, in his opinion, they are causing stagnation in scientific and technological development. Thiel does not shy away from declaring a young woman who once became famous for protesting against environmental destruction to be the Antichrist. For him, the arbitrarily proclaimed Antichrist stands in the way of the rapid development of digital technologies and thus his profit interests. However, international institutions and governments that make democratically legitimised laws to steer this development in the interests of the general public are not the enemy of Christians, but only of presumptuous and insatiable entrepreneurs such as Peter Thiel, who is only one representative of this ideology, albeit the most intellectual and eloquent. Like other Big Tech players, he counts among his enemies the open society, which promotes critical reason and democratic reforms and attempts to shape technological development in a humane way through legislation.

According to Karl Popper, the open society must defend itself against its enemies with bymeans of reason and rational discourse. In doing so, it

remains fundamentally open to dissenting opinions, but must not have unlimited tolerance of intolerance. This leads to the tolerance paradox, which states that unlimited tolerance towards the intolerant ultimately causes tolerance itself to disappear. The paradox of tolerance is at the heart of Popper's argument, and it is something that open societies must grapple with. If a tolerant society tolerates intolerance without limits, the intolerant will take power and destroy the tolerant society. It is quite obvious that the enemies of an open future have recognised the tolerance paradox as a gateway to their fight against democracy. Through digital opinion-making based on democratic freedom of expression, they win over their supporters, whom they convince that there is no freedom of expression.

In the name of tolerance, open society must claim the right to prevent the intolerant from coming to power, using the law and broad public debate to do so. It is not the expression of seemingly intolerant ideas, but the rejection of rational argument and the threat of violence as a means of political debate that mark the red lines of tolerance, against which law and democracy must in turn draw red lines. This applies, for example, to the systematic dissemination of illegal content and lies, against which a robust democracy must take decisive action.

Popper teaches us that training critical rational thinking and debating with sound arguments based on verified facts is the starting point for defending ourselves against the enemies of an open future. Added to this is the defence of democratic institutions, which are defamed by their enemies as bureaucratic monsters and then dismantled with a chainsaw, as well as the protection of the individual, which, according to Arnold Gehlen, is an institution in itself. The triad of democracy, the rule of law and fundamental rights, the trinitarian formula of our constitutional law, must be defended equally against attackers who want to eliminate these guarantees of an open society in order to exercise power without control.

One difference between the practices of systems based on Carl Schmitt's ideas and those of an open society is the way the latter deals with its enemies. In the United States, masked kidnapping squads from the Immigration and Customs Enforcement (ICE) agency abduct and shoot people in the street who have been declared enemies by the White House, bypassing democratic and judicial controls. Trump has allocated $26 billion per year to the agency, three times the previous budget. This is in-

tended to enable Trump's secret police, as California Governor Newsom called ICE, to deport one million people per year from the United States. Of course, this agency also uses Palantir, Peter Thiel's surveillance software that searches social media to track down suspects and innocent people alike. This increases the stock market value of Thiel's company, while simultaneously fuelling the ideology that shapes Trump's decisions. For the start-up mentality of Silicon Valley, the financing of digital technologies of surveillance capitalism has proven to be a self-fulfilling prophecy since the adoption of fascist ideologies. A perpetual motion machine for tech investors, in which autocrats are accomplices in exponential self-enrichment, much to their own advantage it must be added, because the Trump family, among others, is also earning handsomely from crypto deals.

The pace of authoritarian upheaval in the US is so breathtaking that it is necessary to take a look at the fundamentals of this transformation. The way an open society deals with those who want to destroy it must consist above all in confronting them ideologically. Trump has elevated some of the right-wing masterminds behind Project 2025 to positions where they can push ahead with the restructuring of the state. Brendan Carr was appointed head of the Federal Communications Commission (FCC), while Tom Homan, the former head of ICE, was promoted by Trump to *border czar*, responsible for organising the deportation of migrants, to name just two examples.

If the open society's approach to its enemies is to keep them away from power, their power must be curbed in a forward-looking manner. Excessive financial, unrestricted economic or psychological and political manipulation power exercised through the tax mechanisms of the digital giants must not be concentrated in the hands of a few in the first place. The soft power of enlightenment about the abuse of power by tech bosses and the twisted ideologies they follow must be accompanied by the resolute enforcement of existing law. Monopolies and oligopolies that harm people and the economy, and dominate opinion markets must be broken up, appropriate taxation enforced, and unauthorised interference in democratic decision-making and democratic elections prevented. The defence of democracy is a consequence of the defence of human beings against the superhuman and inhuman ideologies behind the enemies of an open future. For democracy is the form of government that best protects vulnerable human beings and gives them room to develop, and for that very reason it is itself vulnerable.

The aim here is to describe the abstruse mixture of elements from different ideologies and religious content that unites the unholy alliance of Big Tech and the Alt Right. Those who see themselves as the new rulers of the world openly declare what drives them and, in some cases, what they intend to do. Their ideology is a dictionary of misanthropy. In this, too, they are no different from the dictators of the 20th century. That is why linguistic criticism of the ruling techno-ideology is indispensable.[35]

To facilitate this debate, we provide a brief summary of the intellectual roots that the attack on democracy draws upon.

Elements and origins of dark enlightenment

Lost in a huge forest at night, I have only a small light to guide me. A stranger approaches me and says, "My friend, blow out your candle so you can find your way better." This stranger is a theologian.
Denis Diderot

With this image, in his *Pensées philosophiques* (Philosophical thoughts) the philosopher Denis Diderot criticises the ecclesiastical dogmatism of his time, which attempts to extinguish the light of the Enlightenment and promises people that they will see better in the dark. His candle stands for independent thinking and judgement, the dark forest for a confusing, contradictory world that frightens us. Today, the dark forest is summed up in a term that seeks to replace the Enlightenment with darkness as humanity's new goal. In 2008, a Californian blogger published a text that is considered the founding document of the Dark Enlightenment.

Dark Enlightenment? An obvious contradiction, an oxymoron. Exactly the kind of rhetorical figures that are favoured by the algorithms of the attention economy and thus become successful. They give the appearance of a deeper truth by combining contradictions, yet they only turn the truth upside down. The prehistory of the deliberately provoked crisis of truth in

35 Sternberger, D., G. Storz and W. E. Süskind (1975) *A Dictionary of Inhumanity* (New York: Harper & Row): "The more and the kind of language one speaks, the more and the kind of things, world or nature are open to him. And every word he speaks changes the world in which he moves, changes himself and his place in this world. The corruption of language is the corruption of man. Let us be on our guard! Words and sentences can be gardens as well as dungeons in which we imprison ourselves by speaking." The authors spoke from direct experience of a criminal dictatorship.

which we find ourselves today, since government agencies in the oldest democracy of modern times have been spreading alternative facts and alternative truths, begins with this document.

The man who wants to blow out our candle this time, so that we can supposedly find our way better in the dark forest of information overflow, appears under the pseudonym Mencius Moldbug. The first name is reminiscent of a Chinese philosopher, while the surname means mould. On his blog, he published a several hundred-page-long "open letter to open-minded progressives."[36] A few years later, it became known that the programmer Curtis Yarvin was behind the pseudonym, by which time his pamphlet had already become one of the unofficial founding documents of the MAGA movement. Here, too, deals were made right away. Influential tech billionaires such as Marc Andreessen, Ben Horowitz and Peter Thiel had already invested in the obscure company Tlon Corp, which Curtis Yarvin had co-founded to further develop the Urbit computer platform designed by Yarvin. They befriended him. They found the idea of a Dark Enlightenment attractive and useful for their goals.

That is why it is worth examining this text, which reveals important communication strategies of the Big Tech/Alt Right movement. Right at the beginning, a rhetorical trick is employed: what Yarvin calls progressivism, by which he means the liberal-democratic mindset, is less a rational worldview and more a religion. An unofficial belief system consisting of the media and science (remember: during the Enlightenment, the media and science were understood as institutions that, unlike religious belief systems, are oriented towards knowledge and truth) would shape public opinion and suppress alternative views and truths. By referring to this system as "the cathedral," he lays the rhetorical foundation for everything that follows.

First and foremost, the attack on democracy, for which this supposed cathedral was built. Yarvin describes democracy as "rubbish," saying that an irreparable systemic flaw prevents it from being fixed. He describes it as a façade behind which bureaucrats and experts control politics, while elected politicians are merely puppets. He recommends first taking the red pill, familiar from the Matrix films, in order to see the world as he does,

36 Moldbug, M. (2008) "An Open Letter to Open-Minded Progressives: Chapter 1: A Horizon Made of Canvas". *Unqualified Reservations*, 17 April.

namely upside down. The scene from Matrix became an influential meme of the anti-feminist and anti-democratic conspiracy subculture, first in the USA, then worldwide. It is supposed to symbolise leaving a comfortable illusory world in favour of a harsh, brutal truth. After taking the pill, Yarvin calls for the destruction of democracy, the rule of law and the constitution. Then, in true technical solutionism style, the computer programmer proposes a reset of the state. It is to be replaced by an algorithm-driven CEO monarchy. The neologism CEO monarch is intended to combine absolute rule with the founding spirit of start-ups. As early as 2011, Yarvin described Trump as one of two personalities who biologically had what it takes to be king of America.[37] This new king would have the task of turning the state into a "heavily armed, ultra-profitable corporation" that would fire civil servants, abolish the press and universities, and put "uncivilised sections of the population" in prison, according to Yarvin, alias Moldbug, in Dark Enlightenment. With cool nonchalance, Yarvin himself recognises the weakness of his proposal, which relates to the person who is crowned CEO-king, but he dismisses it as a lesser evil compared to the impositions of democracy: "Clearly, if he or she turns out to be Hitler or Stalin, we have just recreated Nazism or Stalinism."

Yarvin is a neo-fascist who masks and presents himself as an intellectual. He poses as a cool systems engineer of authoritarianism. His cynical intellectual posturing explains his charisma. His plan has largely worked. He owes this primarily to the influential network of right-wing tech bosses who have rallied behind his anti-democratic agenda and ensured that the mould spreads over public discourse and infiltrates it.

"How dangerous is it that we are being linked?" Thiel wrote to Yarvin in 2014, according to reporter Ava Kofman. "One reassuring thought: one of our hidden advantages is that these people" – social-justice warriors – "wouldn't believe in a conspiracy if it hit them over the head (this is perhaps the best measure of the decline of the Left). Linkages make them sound really crazy, and they kinda know it."[38]

This is another interesting twist in the distortion of reality, like taking a second red pill. On the right wing, absurd conspiracy theories ranging from the Pizzagate conspiracy involving Hillary Clinton to the contamina-

37 Kofman, A. (2025) "Curtis Yarvin's Plot Against America". *The New Yorke*r, 2 June.
38 Ibid.

tion of the population by chemtrails to the claim that governments are pursuing a policy of population replacement ("Great Replacement theory") had previously gone viral, fuelled by social media algorithms. Now Yarvin is preparing for investigative reports on the interplay between Big Tech and the Alt Right to go viral as well by prophylactically denouncing them as left-wing conspiracy theories. This is akin to crooks calling prosecutors investigating organised crime conspiracy theorists because they assume there is a network of perpetrators, and getting away with it because this claim is picked up everywhere and spread en masse by algorithms. This would be good news for organised crime. Mob bosses would certainly be interested in as powerful a bullshit machine as the algorithms of Big Tech currently represent.

Meanwhile, there are isolated cases of political actors putting Yarvin's ideas into practice. Chainsaw men like Argentine President Milei and short-term government official Musk have begun to dismantle what they consider to be the Deep State: the institutions of democracy. Musk's dismantling of US government agencies has, according to his own claims, saved $160 billion; during his election campaign, he promised a ludicrous $2 trillion in savings.[39] DOGE employees have laid off an estimated 260,000 employees and written software to automatically dismiss government employees so that the state can be destroyed more efficiently. If the plan had been fully implemented, one person in particular would have benefited: Musk himself. Because this software would have come from his company. As always with Musk and co., political and business interests go hand in hand. The modern version of totalitarian systems of rule looks like this: the state is transformed into a structure of so-called sovereign corporations, SovGovs, in which, in addition to the CEO-king, a secret committee is also allowed to make decisions. The freedom of citizens is reduced to being able to leave the corporation if they do not like it. The freedom of citizens is reduced to the consumer's freedom of choice to opt for another product if they are dissatisfied.

There is an element of the so-called Dark Enlightenment that explains its appeal to many. It is the claim that it takes a sober look at the dark forces of history and the deep-seated violence in human nature, rather than indulging in the morally arrogant illusions of progressives. A cynical view of

39 Hayes, C. and B. Drenon (2025) "Elon Musk leaves White House but says Doge will continue". *BBC News*, 29 May.

people's attempts to improve their lives, understanding progress not only as a blind increase in technology, but as improvements in their lives and in society. Cynicism shares with the secret knowledge promised by conspiracy theories the compensatory satisfaction that comes with a sense of elitist superiority over the supposedly conformist masses. Nevertheless, after the triumph of the document, the question must be asked why the Dark Enlightenment calls itself Enlightenment. What gain in knowledge does it promise? After all, to stay with Diderot's image, no one who wants to find their way through a dark forest by lighting a torch needs to be told that the forest is dark.

It is worth scrutinising a second leading thinker, the one who coined the term Dark Enlightenment for Yarvin's post and who himself published a pamphlet entitled *The Dark Enlightenment* in 2012[40]: Nick Land, a British philosopher and occultist. He takes up Yarvin's these, exaggerates them and gives them a philosophical gloss. Above all, he builds a bridge to Peter Thiel. In his text, he refers to a meeting of libertarians in 2009 organised by the conservative think tank known as the Cato Institute. Land takes up Thiel's quote from that meeting: "I no longer believe that freedom and democracy are compatible," and draws radical conclusions from it. In his view, democracy is ruled by a howling mob that has been brainwashed by political correctness and is pushing for the extinction of the white race. Land's contribution to the Dark Enlightenment consists, on the one hand, of introducing blatantly racist, social Darwinist and eugenic positions. On the other hand, the former leftist has expanded Karl Marx's thesis of the accumulation of capital into accelerationism. Behind this lies the thesis of the unbridled development of capitalism, which, through digitalisation, is throwing off all shackles and is supposed to lead to the true engine to transhumanist singularity and thus to true modernity. The original twist on original Marxism is that at the end of history, it is not man who liberates himself from capitalism, as Marx prophesied, but capitalism from man.

From the prediction that unbridled digital capitalism will destroy humanity, Land then develops the techno-utopia of transhumanistically optimised humans who, together with intelligent machines, will establish a new civilisation. In this analysis, the capital of the future is identical to

40 Land, N. "The Dark Enlightenment". *The Dark Enlightenment*.

AI.[41] In a blog, Land is quoted with this thesis under the heading "The Final Solution to the Human Question" – another provocation that plays associatively on the phrase "the final solution to the Jewish question" coined by the Nazis for their crimes. Land summarises his vision of the future as follows: "Life continues, and capitalism does life in a way it has never been done before. If that doesn't count as 'new', then the word 'new' has been stripped down to a hollow denunciation. It needs to be re-allocated to the sole thing that knows how to use it effectively, to the Shoggoth[42]-summoning regenerative anomalisation of fate, to the runaway becoming of such infinite plasticity that nature warps and dissolves before it. To The Thing. To Capitalism."[43]

Land, who experimented extensively with drugs, lived temporarily in the house of occultist and Satanist Aleister Crowley, and suffered several mental breakdowns, is considered, despite or perhaps because of this history, alongside Yarvin, to be the chief ideologist of the Dark Enlightenment. With accelerationism, he has provided an essential building block that has made the discourse of the New Right in the USA compatible with Silicon Valley. Allow us this bon mot: if the Dark Enlightenment prevails, then good night, America, and good night, democracy.

The neo-reactionary movement and its theories

Sovereign is whoever decides on the state of emergency.
Carl Schmitt, Political Theology, 1922

The neo-reactionary movement in the USA refers to itself as NRx. In addition to the Dark Enlightenment, it brings together a whole range of other colourful theories that have already been described in detail elsewhere.[44] Because they determine the narratives of the enemies of democracy in varying combinations and emphases, we would like to present them in brief.[45]

41 Skrobisz, N. (2020) „Akzelerationismus Teil 2:/acc – Das Kapital ist eine KI". *Leveret Pale*, 28 August.
42 Fictional characters in H. P. Lovecraft's sonnet cycle Fungi from Yuggoth.
43 Land, N. (2011) *Fanged Noumena* (New York: Urbanomic/Sequence Press).
44 See, among others: Mühlhoff, R. (2025) *Künstliche Intelligenz und der neue Faschismus* (Ditzingen: Reclam).
45 For a detailed explanation of the fundamentals of some of these theories, see also: Nemitz, P. and M. Pfeffer (2023) *The Human Imperative*.

Technological determinism

Originally, this is the thesis that technology forces social, political and cultural adaptations and therefore results in social and cultural change. In Silicon Valley[46], this view was supplemented by the assertion of an inevitable, predetermined development of technology in a certain direction (usually towards AGI) and by the normative charging of this supposedly descriptive truth. According to this view, the development of technology is not only inevitable, but also desirable and morally imperative.[47] The question of which moral values form the basis for this view repeatedly leads back to the social Darwinist principle of the struggle of all against all and the right of the strongest. As investor and author of the *Techno-Optimistic Manifesto* Marc Andreessen writes in it: "Techno-optimists believe that societies, like sharks, grow or die." The social Darwinist exaggeration of life as a constant struggle for survival leads to the conclusion that one must opt for unbridled technological development if one does not want to perish. This dramatic escalation allows techno-determinists to ignore questions about the social distribution of power and resources, as well as the discussion about negative consequences. Instead, these aspects are "systematically obscured."[48] What remains is a naked struggle for power, which the theory simultaneously highlights and normatively demands. The basic conviction can be summed up in the well-known formula: Neither fox nor donkey can stop AI in its course.[49]

Transhumanism

A philosophical-technological movement that seeks to fundamentally and radically optimise human life through science, medicine and technology by overcoming natural limitations such as illness, ageing and death. Technologies such as genetic manipulation, artificial intelligence, neurochemistry and nanorobotics should lead to a fusion of humans and machines

46 Andreessen, M. (2023) "The Techno-Optimist Manifesto". *Andreessen Horowitz*, 16 October.

47 See: Kelly, K. (2016) *The Inevitable* (New York: Viking). Kelly himself describes the aim of his research as follows: "to uncover the roots of digital change so that we can embrace them".

48 Mühlhoff, R. (2025) *Künstliche Intelligenz und der neue Faschismus*, p. 73.

49 Shortly before the fall of the Berlin Wall, Erich Honecker, Chairman of the State Council of the GDR, said (in German): „Den Sozialismus in seinem Lauf hält weder Ochs noch Esel auf."

(cyborgisation) and thus to the overcoming of humanity. Transhumanism attempts to legitimise itself historically in humanistic ideas of the further development of humans and humanity towards a higher goal and, in addition to its socially progressive roots, still exhibits currents today that combine human and technical progress, but often abandons the very foundation of humanity by making technological optimisation absolute.

Extropianism

A branch of transhumanism that seeks to accelerate human development by unleashing technological progress. Extropy is introduced as the opposite of entropy and is intended to indicate that, contrary to the second law of thermodynamics, limitless expansion is possible through intelligent technology. According to the second law of thermodynamics, order in a closed physical system always decreases, never increases. Extropianism deduces from this that an increase in order can only ever be achieved at the expense of other subsystems. The limitless growth of a system is possible if it occurs at the expense of another system, a perfect connection to classical social Darwinism. The goals of this doctrine, which was largely founded by Max More, are life extension (or immortality), the abolition of ageing, the increase of intelligence and the improvement of the human body. All of this is made possible by proactive intervention in human evolution. Critics speak of a "cosmic Darwinism in which the *extropy* of a system determines its superiority in an overall hostile cosmos."[50]

Singularitarianism

As early as 1990, robotics researcher Hans Moravec of Carnegie Mellon University in Pittsburgh (USA) spoke of our culture soon being able to "develop independently of human biology and its limitations and instead transition directly from one generation of machines to more powerful, even more intelligent ones."[51] This leads to a dystopian vision of a post-humanist future in which the mind is freed from humans, who are described as inferior (Moravec himself spoke of humans as "unfortunate hybrids, half biology, half culture"). The path to this future is supposed to be assured by super-intelligent machines.

50 Mühlhoff, R. (2025) *Künstliche Intelligenz und der neue Faschismus*, p. 77.
51 Moravec, H. (1990) *Mind Children: The Future of Robot and Human Intelligence* (Cambridge: Harvard University Press).

Singularitarianism today is the idea that the development of a powerful, independently thinking AI that surpasses human intelligence and optimises itself through exponential self-improvement will lead to a point in time (the singularity) at which human history will come to an irreversible end because humans will no longer be able to stop or control this development. The singularity is a purely theoretical scenario that is discussed in various ways among experts. This ideology is important because singularitarians such as Ray Kurzweil hold influential positions at Big Tech companies. Kurzweil is Director of Engineering at Google LLC. Singularitarians assume that the singularity is inevitable (see *technological determinism*) and specify concrete time frames for it; for Kurzweil, for example, it is the year 2045.

Eugenics 2.0

Technological eugenics (from the Greek for "good" and "race, family") follows on from the racial theories of the 20th century to devise scenarios for breeding humans using technical means, which are euphemistically referred to as "optimisation" or "overcoming" humanity. Terms such as "life unworthy of life" reappear as "useless people" for whom there is no longer any place in techno-optimised futures. This argument is also used by isolated anti-vaccinationists, who consider "natural" selection through epidemics to be a first step towards genetically optimised humans. However, technological eugenicists favour human breeding through prenatal selection or genetic manipulation.[52] Elon Musk conceived most of his 14 children in vitro because he wants to make humanity "multiplanetary." Investor Marc Andreessen, co-founder of Andreessen Horowitz, explained in an essay: "The best and brightest minds should reproduce to advance the world."[53] This also means that those who are less intelligent should no longer play a role. As pronatalism, the new eugenics is an important link between the tech elite and the right-wing administration in the White House: Elon Musk and J. D. Vance have professed their support for pronatalism.

52 "Do transhumanists advocate eugenics?" World Transhumanist Association, 2005.
53 Beres, D. (2025) "Maga's era of 'soft eugenics': let the weak get sick, help the clever breed". *The Guardian*, 4 May.

Effective altruism

Effective altruism (EA) is a particularly perfidious pseudo-philosophical distortion: the term "altruism" suggests ethical principles, while "effective" promises scientific methodology. EA is primarily a utilitarian ethics that promises "maximum good" in the "most effective way."

Its main thesis is that the scarce resources of time and money should be used as efficiently as possible to solve social problems. All human life should have the same (numerical) value. However, according to EA, there will be far more people in the future than there are today, so their lives will be weighted more heavily than those living today. In conjunction with the assumptions of long-termism, current real problems such as climate change are then devalued in comparison to possible future existential risks. Consequently, one should donate less to countries suffering from hunger today and instead invest the money in AI companies that could perhaps one day solve world hunger through artificial food. Musk's DOGE slash also affected the US Agency for International Development, USAID. EA thus represents a cynical reversal of the precautionary principle, which Hans Jonas developed as a central concept of ethics in the technological age and which futurists such as Thiel regard as the main obstacle to his interests. EA allows social problems to be presented as technically solvable tasks that can be algorithmised and thus effectively solved. To prevent these theories from appearing too inhumane, the vision of "loving" and "merciful" AI machines is repeatedly invoked to allay fears about the technocratic future that has been conjured up. Anthropic CEO Dario Amodei argues similarly in his prose on how AI will change the world for the better, in his view.[54]

Longtermism

Described by some as eugenics by another name, it nevertheless has far-reaching consequences as a theory of the future based on mathematical models. From a crude utilitarian logic, longtermists deduce that it is a higher moral duty to take all necessary steps now to enable there to be more happy people in the distant future than to improve the lives of peo-

54 Amodei, D. (2024) "Machines of Loving Grace". *Dario Amodei*, October. The phrase comes from the poem "All Watched Over by Machines of Loving Grace" by Richard Brautigan.

ple currently existing. The assumed future happiness of a planetary race is given greater weight than the rights of people living today. Longtermism thus represents a perfidious reversal of philosopher Hans Jonas' principle of responsibility, which is described in detail elsewhere. Critics say that this ideology serves those who have become rich through fossil capitalism to avert their responsibility for current crises.

Secular eschatology

Since its inception, tech ideology has borrowed heavily from religious promises of salvation and redemption. It is therefore not surprising that it has recently come to be openly presented as an apocalyptic doctrine of the redemption of the world through technology. We will discuss Peter Thiel's borrowings from the biblical texts of the Apocalypse in more detail later. His remarks lead to the astonishing conclusion that right-wing Christians should oppose democratic legislation and international institutions as well as the precautionary principle, since these elements are equated with the Antichrist. At its core, Carl Schmitt's interpretation of the Apocalypse reappears in connection with his friend-enemy schema. For the apocalyptic Thiel, technology is not only salvation, but also every Antichrist who opposes technology and its exponential progress or even wants to steer it in a direction other than the one he pursues.

Catholic (neo-)integralism

Catholic neo-integralism is a resurgent movement among conservative Catholics who want to subordinate the state to the authority of the Church and see liberalism as the enemy. Politically, this movement is oriented towards fascist leaders such as Mussolini and Franco. Today, many neo-integralists sympathise with totalitarian or even theocratic regimes, while Viktor Orbán, with his so-called illiberal democracy, is considered the most important current example. Neo-integralism establishes a connection between the US tech right and conservative Catholic European movements. The conservative Konrad Adenauer Foundation states: "Catholic neo-integralism is an ideological danger, and it is making its way from the fringes of conservative Catholicism to the centre of conservative politics."[55]

55 Patterson, Prof. J. M. Ph. D. (2024) "Neo-Integralismus – Eine Gefahr für die liberale Demokratie". *Zeitgeschichte AKTUELL* 12.

(Futuristic) rationalism

Refers to the fundamental decline of reason that is characteristic of all solutionist ideologies, in which a single strand of rationality is set as absolute and turned against all other forms of reason. Not to be confused with the rationalism of philosophy, its adherents are a Bayesian sect[56] for whom probability calculus is the only form of rational thinking and judgement that eliminates misjudgements arising from intuition, emotion and other factors. Above all, visions of the future, but also ethical concepts, are to be proven rational and thus superior through calculation using stochastic methods. Ultimately, this is the thesis that algorithmic calculations based on the Bayesian probability model are the only basis for an error-free world view.

Libertarianism

Is a radically individualistic philosophy that places the freedom of the individual above all other values. It ranges from radical neoliberal to anarcho-capitalist currents. As cyberlibertarianism, the movement plays a major role, particularly through John Perry Barlow's *Declaration of the Independence of Cyberspace* (see "*Algorithms as dispositives of power in the cyber public sphere*" in *Chapter 6*), because it succeeded in preventing regulatory legislation of the internet for decades, thereby laying the foundations for the monopolies that Big Tech has achieved today. With libertarianism, numerous, in some cases disparate but essentially related currents have found their way into US politics. Of particular importance is the influence of the author Ayn Rand (1905–1982), who attempted to base morality on radical individualism and glorify the heroic entrepreneur. Her so-called objectivism declared radical egoism to be a rational ethical way of life and advocated pushing back the state to make way for uninhibited capitalism.[57] Libertarianism is the decisive transmission belt for techno-utopian ideologies from Big Tech to politics because it combines their goals with a perverted concept of freedom.

56 Named after the English mathematician Thomas Bayes (1701–1761), who developed the conditional probability theorem.

57 It is interesting to note that this self-proclaimed hater of the state and collectivism registered with the state social security system under a false name in 1976 in order to claim public benefits for lung cancer surgery. At the time, she was already a successful bestselling author. See: McConnel, S. (2010) *100 Voices: An Oral History of Ayn Rand* (New York: New American Library).

What does the dark enlightenment want?

First, like the historical Enlightenment, the Dark Enlightenment seeks to debunk myths. However, it does so by reversing the initial position in a sleight of hand: whereas the historical Enlightenment targeted religious dogma, which it opposed with open-mindedness, criticism and reason, the Dark Enlightenment declares essential institutions that emerged from the Enlightenment and are committed to the search for truth to be cathedrals of a system of rule that it wants to bring down. The same applies to its actual main enemy, the liberal state and democracy.

The Dark Enlightenment's proponents turn the instruments of the Enlightenment, which is to say, criticism and rationality, against the Enlightenment itself and promise to see through the dogmas of this belief system with radical rationality and expose them as the masks of a will to power. Because they claim to act free of moral concerns and political correctness, they assert that knowledge is more important to them than moral acceptance. This promise of a dark truth that can only be expected of an elite because the masses do not want to accept it ties in with criticism of the Enlightenment ideal of reason that dates back to Nietzsche. On this view, real enlightenment destroys the illusion that all people are equal and that the people can make rational decisions. Truth is equated with belief; for the Dark Enlightenment, both result from efficiency, intelligence and a new order in which an optimised and superior elite ends the rule of majority opinion. The Dark Enlightenment likes to surround itself with the aura of a brutal but honest turning away from moral illusions in favour of a long-overdue return to reality. The unvarnished view of underlying evil is matched by a murmuring tone in the corresponding texts, which undermines any clarity and certainty, Descartes' central criteria for scientific texts. The reality that these armchair heroes face is understood as an eternal struggle in which the better and fitter prevail in the end. Of course, these are initially the ideologues themselves, who, in echoing fascist ideologies, display a fundamental disposition towards violence. They exaggerate the supposedly eternal struggle of all against all as the sole source of normativity, and thus call for precisely this struggle to be waged. They expect conflicts to be resolved through power alone. Nor are the echoes of the Conservative Revolution of the 1920s and 1930s in Germany new. In this political movement, intellectuals such as Ernst Jünger, Ludwig Klages, Oswald Spengler, Carl Schmitt and other anti-liberal, anti-democratic, anti-egalitarian and ultra-nationalist elements merged to form an advance guard of German fascism. They took a stand against

modernity, which they interpreted as a crisis rather than progress. They considered liberalism and parliamentarianism to be spiritless and decadent, believing that democracy would lead to the downfall (of the West) through the rule of mediocrity and must be ended by the rule of intellectual and military elites.

Despite all their differences, the Dark Enlightenment can be read as a postmodern digital continuation of the Conservative Revolution in the digital age. Both rely on a discomfort with modernity, which they aim to destroy through radical criticism. But they differ in their methods. Instead of pathos, the Dark Enlightenment's advocates rely on technology; instead of blood and soil, they rely on data and algorithms; instead of classic heroes, they glorify the role of the disruptive entrepreneur and the CEO monarch.

Like the Conservative Revolution, the Dark Enlightenment uses rational thinking to attack the achievements of reason. It turns reason, and thus the Enlightenment, against itself with destructive intent. This is a new chapter in the dialectic of enlightenment, as described by Horkheimer and Adorno when they described the emergence of totalitarianism in the 20th century. At that time, the founders of critical theory sought to show how the Enlightenment, which originally sought to lead people out of immaturity, could turn into its antithesis: oppression, dictatorship and mass murder. Their central thesis was that practical reason, i. e. reflection on freedom, law and morality, is being supplanted by instrumental reason, which aims at the internal and external control of nature. The separation of criticism, which is actually inherent in it, is not achieved by reason or the Enlightenment. Setting out to see through the myth, it itself turns into myth. Today, the Dark Enlightenment creates this new myth primarily in the form of a deification of artificial intelligence, which it in turn understands as an elitist instrument of power that is supposed to help end the rule of the masses in democracy.

A contemporary critique of the digital would first have to describe the dialectic of the digital Enlightenment. It would not regard the Dark Enlightenment as the other or the opposite of the Enlightenment, but as the perverse result of its self-objectification. The Dark Enlightenment seeks to align everything with efficiency, controllability and performance. It absolutises quantitative thinking and, according to pars pro toto logic, declares a section of rationality to be the whole of reason. In its technocratic model of society, in which algorithms take over the surveillance and

control of people, democracy is considered inefficient and fundamentally unrealistic, and morality and ethics are considered irrational unless they are based on egoism. This is the result of a rationality that believes it can do without ethics and a reason without self-reflection. Algocracy, which hallucinates the Dark Enlightenment, does not need autonomous subjects; on the contrary, it must destroy autonomy. Functional units take the place of subjects. It achieves its goal when people increasingly adapt to algorithms instead of designing them according to the requirements of a human society.

The Dark Enlightenment refers to a new stage of those self-destructive tendencies that Horkheimer and Adorno described as the dialectical decline of the Enlightenment. It is another form of enlightenment without enlightenment about itself, and thus a stage of decline in the capacity that actually distinguishes humans: their reason, which consists not only in the ability to rationally control nature, but also in the ability to develop moral norms and understanding, as well as in artistic forms of authentic self-expression. All these forms of rationality have differentiated themselves in modernity and have diverged into separate spheres of validity. To isolate them completely from one another and play them off against each other, with one sub-area negating the others, leads to the self-destruction of reason, which can precede the self-destruction of democratic self-determination and, ultimately, the self-destruction of humanity. Therefore, a prerequisite for the self-assertion of democracy is the rehabilitation of reason – which alone enables free self-determination – against its radical critics on the left and right.

The real danger posed by the Dark Enlightenment is that it is based on a digital-economic power complex that is unprecedented in history and can translate its programme into policy. This power complex has led to a distribution of global wealth that must be unacceptable to any free and social society: a handful of people now own as much financial wealth as the poorer half of humanity combined. The renunciation of practical reason demanded by techno-futurists is also intended to prevent criticism of this unacceptable distribution of wealth. Tech barons now have at their disposal not only economic but also immense political power, which allows them to implement their misanthropic agenda.

In addition, with the development of AGI, this complex is pursuing a project that could rule out self-corrections guided by reason in the future, which have repeatedly been a saving corrective in previous civilisational

crises; comprehensive critical reason can be made to disappear or significantly weakened in its effectiveness through adaptation to algorithmic systems. The unprecedented ability to monitor, sample and control societies through digital AI technologies can pre-emptively stifle any new enlightenment, and thus, theoretically, prevent any return to democracy once the goal of its destruction, which unites the protagonists of the Dark Enlightenment, has been achieved.

Whether the solution to the human question will then be sealed by the extinction of humans or "only" the disappearance of humanness remains open. That difference is probably all that will remain of the open future in such a dystopia.

Apocalypse as a business model – Peter Thiel's end times

On 1 January 2007, Peter Thiel published the essay *The Straussian Moment*. In it, Thiel refers to the German-American philosopher Leo Strauss (1899–1973), who launched a thoroughgoing critique of modern liberal democracy. Strauss argues that the Enlightenment suppressed fundamental questions about human nature, morality and religion or dismissed them as irrelevant in order to create a stable, prosperous state. But these suppressed questions return, according to Strauss. A Straussian moment occurs when sudden events challenge the assumptions of the Enlightenment and bring the seemingly settled existential questions back to the surface. This forces society to re-evaluate its philosophical foundations.

Thiel sees September 11, 2001 as such a moment. His essay begins with the assertion that the attacks of September 11 made the Straussian Question relevant again. Thiel interprets 9/11 as what Strauss would call the moment when a repressed truth returns. In his view, this moment shows that liberal economic rationality is not universal: people also act out of faith, willingness to make sacrifices, metaphysical conviction – categories that liberalism supposedly cannot explain.

Strauss spoke of a return to the question of the whole – the question of the nature of man, nature and goodness. According to Strauss, modernity has suppressed this question by dissolving philosophy into technology, economics and social science.

Thiel says that 9/11 revealed the limits of this modern rationality: "The Straussian moment is the moment when the Enlightenment project of liberal modernity finds itself confronted by something it cannot understand – by irrational faith, by death itself." Another baseless assertion: as if the Enlightenment or liberal modernity had ignored or concealed death. In truth, both projects seek to improve life, which is surrounded by death. With this assertion, Thiel translates Strauss's philosophical diagnosis into a geopolitical one. Liberal modernity (Enlightenment, economy, democracy) encounters forces (ideologies, religions) that it can neither explain nor contain but against which it must defend itself existentially.

For Leo Strauss, every political order is ultimately based on a fundamental decision about good and evil, friend and foe, and not merely on procedure. Thiel concludes that the West has forgotten that politics always depends on such decisions – and not on market mechanisms or dialogue. Thiel thus takes up the Straussian opposition between philosophy and modernity: modernity believes that it has solved all ultimate questions through enlightenment and science, but reality shows that these questions are indestructible. "We no longer believe in human nature, only in human preferences." This is almost verbatim Strauss, who attests to modernity's fundamental loss of substance. If nature no longer exists, according to Strauss, then there is no standard for good and evil. Thiel uses this argument to show that the West is incapable of understanding religious violence or self-sacrifice, which are phenomena based on a conception of nature and transcendence.

The West is therefore on the wrong track with liberalism, which must lead to nihilism. Thiel follows Strauss's reading of Nietzsche when he describes modernity as a "post-ideological vacuum", transferring the diagnosis of modernity to the West and, at the same time, into a techno-political form: "Liberalism's belief in openness has *become* its own closed ideology – unable to recognise what it cannot *tolerate*." With its fundamental openness, the West has in fact fallen into a state of moral nihilism, and in Thiel's view, this is its weakness vis-à-vis those who still believe in something absolute. At the same time, the open society has recognised that it is dealing with enemies like Thiel and has – theoretically – defined measures to counter them.

Thiel follows Strauss's call to return to the classics such as Plato and Aristotle in order to understand politics once again as a question of virtue, nature and order. He suggests that the philosophical depth of antiquity

must be combined with the realistic sharpness of modern power politics (à la Schmitt). It was Strauss's idea that philosophy must keep the eternal questions open. Thiel concludes that politics and technology must once again be grounded in metaphysics, otherwise they will lead to a loss of meaning or spirals of violence. Thiel thus deploys Strauss to provide philosophical support for his criticism of liberalism and ultimately discredit it. He seeks to expose its supposed implicit emptiness and thereby invoke the need for a "deep order" which can only be established by absolute power.

He then introduces Carl Schmitt, for whom, as mentioned above, politics is characterised by the establishment of a friend-enemy dichotomy. This is followed by the passage that combines both counter-Enlightenment thinkers: "When all politics becomes administration, the political returns in its most violent form." Liberal democracy, based on legitimacy through procedure, supposedly wants to banish the political from the world by renouncing authoritarian friend-foe dichotomies, which is precisely why it returns through the back door as catastrophe. Through its openness, liberal democracy promotes the catastrophe that it is then unable to counter because it has weakened itself through its basic assumptions. In Thiel's view, this is exactly what happened on 9/11. And in a very similar way, Putin today analyses the West as weak and effeminate.

Thiel borrows stylistically from the motif that has been rightly called Leo Strauss's discovery: the distinction between esoteric and exoteric readings of classical authors. According to Strauss, these authors wrote exoterically for the general public, i. e. in a morally harmless way. For the few who understand the deeper truth, however, they wrote esoterically. Thiel imitates this – consciously or as an intellectual pose. He never explicitly states what his political conclusion is. He merely makes vague references to a "philosophical elite" or "those who still take metaphysics seriously." He thus employs Strauss's method by writing for two audiences – one that reads him as a conservative critic of liberalism, and one that understands him as a metaphysician of investment in times of technical rationality.

To Strauss's credit, he developed this method in order to better understand the philosophies of Jewish thinkers such as Maimonides and Spinoza, who had to live in constant fear of anti-Semitic reactions from their environment. As the son of rural Jews in a strictly Catholic environment, he knew what he was talking about. When analysing *The Straussian Mo-*

ment, a third interpretation can be added to Leo Strauss's distinction between exoteric and esoteric levels of reading in the school of classical ideology criticism: Thiel's analysis of 9/11 serves an obvious business interest.

According to Palantir CEO Alex Karp in the documentary film "Watching You"[58], the terrorist attacks of September 11, 2001 were the catalyst for the idea to found Palantir. Karp and co-founder Thiel wanted to support the state with military data analysis. In the film, Karp, who likes to portray himself as a leftist and Habermas scholar, openly admits: "Our product is occasionally used to kill people." The fact that Palantir is now used to track down "illegals" who are then abducted by ICE troops in the United States, masked and in broad daylight, bypassing democratic procedures, is just another footnote in this story.

According to Karp himself, 9/11 was the catalyst for the boom in new data technologies. And the starting point for a company that, in 2006, when Thiel wrote his Strauss essay, had just completed two rounds of financing totalling approximately $16 million and is now valued at $350 billion in 2025, surpassing the valuation of Lockheed, the manufacturer of old military hardware, by a factor of three, even though it essentially consists only of opaque algorithms. Thiel's essay thus literally paid off in his new start-up. The shame of not knowing about the terrorists' activities in the run-up to the attacks was great in the United States in 2001. Thiel and Karp recognised the opportunity and promised to be able to catch terrorists with software before they committed their crimes. Today, they catch "illegals." They received generous funding from the CIA and the military, and their first contractual partners were all government agencies, secret services and the military.

Based on Strauss's thesis that violence returns when metaphysics is absent, Thiel and Karp made the promise that violence can be prevented if there is no lack of investment in data technology. The fact that a new form of violence arises that is above the law when the liberal democratic order is completely dismantled due to this supposed weakness is concealed or obscured in Thiel's exoteric interpretation.

58 Stern, K. dir. (2024) *Watching You – Die Welt von Palantir und Alex Karp* (sternfilm), documentary film.

Thiel's essay can be read as an attempt to harness the political theology of the 1920s and 1930s to derive a technological policy for the 21st century. It embodies a pattern of Thiel's, who claims to always be a few years ahead of his time. In a sense, this is true of every investor who invests only if he believes in the company in question. Thiel has recognised that his own beliefs only have an effect when they are shared by as many people as possible. And that can be achieved through manipulation, whether through technology or ideology. When the combination of technology and power is promoted as a new form of political order, it is done in his own clearly recognisable economic interest, which, as is well known, consists of replacing competition in the economy with monopolies. This makes a lot of sense to him: monopolies are clearly more profitable, and some digital platform companies have already established monopolies today. Monopolies in politics are paid for not only with the loss of competition, but also with the loss of freedom, self-determination and ultimately prosperity by all those who do not own any of the monopolies. And that is the vast majority of people. That is why political monopolies are called dictatorships.

Thiel borrows Strauss's philosophical depth and Schmitt's dramatisation of politics to create a technocratic-elitist counter-model to liberal modernity. Thiel, who was born in Frankfurt, should be well aware of how things usually turn out historically when enemy stereotypes are conjured up in times of crisis and supposed courses of history are first recognised and then executed.

In light of the experiences of the 20th century, the supposed weakness of liberalism must be seen as a decisive advantage over any form of dogmatism. It is precisely the renunciation of absolute dogmatic truths that enables tolerance towards those who think differently and a genuine search for truth. The concept of truth in an open society is aware of its falsifiability and sees this, as well as the openness of research results, as its real strength. To quote Lessing: Let everyone say what they think is true, and let the truth itself be commended to God!

Dogmatists who claim to be in possession of the truth have a field day in times of crisis. That is why it is part of their business model to bring about the crisis by themselves exaggerating the problems and stirring up fear. The current inability of liberal democracies to maintain the balance between promises of freedom and prosperity makes anti-modern promises so tempting because they promise apparent certainties in uncertain times.

The Age of Enlightenment is far from over. Today, we are no longer in the dark forest of ignorance, but in a forest of information, the internet, social media, AI and opinion markets. Our little candle is still critical reason, i. e. the ability to examine, doubt and think for ourselves. Today's theologian is not necessarily a priest – he appears as an ideologue who promises simple truths, as an influencer who relies on emotion rather than arguments, as a political or religious movement that rewards intellectual laziness, or even as an algorithm that keeps us trapped in a prison of recursive self-affirmation without genuine dialogue. When such voices say, "Don't trust your own thinking – we know better," Diderot's scene repeats itself: we are asked to blow out the candle because darkness seems more comfortable or safer. "Better a small, uncertain light of reason than the great but blind darkness of compulsory faith," wrote Diderot. The fact that the historical Enlightenment thinkers did not need to be taught about the Janus-faced nature of the Enlightenment or about the dark sides of humanity and the world is another matter. It is precisely this insight that conditions are adverse, that humans are not perfect and that their knowledge is always limited which Diderot describes in his image. The Enlightenment thinkers tried to light a light that can only shine if each person ignites it themselves with their own mind. Then the many candles can become a bright glow that enables orientation in thinking.

Meanwhile, the darkness that Yarvin and others have spread is considerable. However, it remains to be seen whether their path will lead to a new dark age. Curtis Yarvin can look back with pride on the fact that many of his ideas have been implemented by the Trump administration and that powerful tech barons support them. But there are good reasons even today to trust the Enlightenment philosopher Kant's assertion that human beings will not allow themselves to be permanently deceived and thus led into darkness. "Human beings therefore feel within themselves a capacity to allow themselves to be compelled by nothing in the world. Although this is difficult for other reasons, it is nevertheless possible; they have the strength to do so."[59]

59 Kant, I. (2001) *Lectures on Metaphysics* (Cambridge: Cambridge University Press).

3 | AI systems and their implications

In August 2025, OpenAI released ChatGPT-5, at that time the latest version of its successful chatbot. The announcements leading up to the release had raised expectations of something spectacular. CEO Sam Altman had said that the 5th version of his tool could provide a doctoral-level response to any query. While the benchmark tests were mastered with flying colours, just a few hours after release, social media was flooded with stories of missteps or simply embarrassing mistakes. For example, even after repeated requests, the most *intelligent* AI system to date insisted that the word strawberry only contains the letter "r" twice, or it made mistakes when answering simple factual questions.

One goal of this chapter is to understand why these errors occur and why they are not just minor silly mistakes, but fundamental weaknesses of these systems. We also discuss the implications of these system characteristics and weaknesses for our freedom and democracy. The question "What do you need to know?" arises repeatedly in the context of technology, its applications and assessment. Most of us do not know how a car engine works in detail. Nevertheless, as part of the driving test, we must be able to demonstrate a certain level of understanding of the technical processes involved. We cannot explain every term that arises in detail within the scope of this book and we encourage readers to conduct their own further research if they are interested. When it comes to AI systems, even though most computer scientists who work with them on a daily basis do not understand in detail what actually goes on "under the hood," a basic knowledge of the subject is necessary for consideration and informed judgement.

First of all, it should be noted that central concepts related to AI, including the concept of artificial intelligence itself, are extremely problematic. In particular, concepts that have been directly adopted from humans and human thinking and abilities are part of the great imitation game, which is discussed elsewhere in this book. They also represent an important part of the marketing and current hype surrounding AI. Even if so-called artificial intelligence is more accurately described as *machine intelligence*,

the phenomenon has little to do with human intelligence. When a chatbot writes that it is *thinking*, it is not thinking in any meaningful sense of the word, but rather performing statistical calculations to compute a response text word by word. The camera sensors of self-driving cars do not *see*, but algorithmic processes attempt to identify objects in images and radar signals by assigning them to patterns.

The beginnings of AI are generally dated to the summer of 1956. That was when a group of scientists met at Dartmouth College in the US state of New Hampshire, which is now considered central to the development of AI. Among them were Marvin Minsky, John McCarthy, Claude Shannon, Allen Newell and Herbert A. Simon. The grant application to the Rockefeller Foundation for this meeting, which went down in computer history as the *Dartmouth Conference*, already set out how the goal of AI was to be achieved: "An attempt should be made to find out how machines can be made to use language, make abstractions and develop concepts, solve problems of the kind currently reserved for humans, and improve themselves." In 1983, computer scientist *Elaine Rich* summarised this description as follows: "Artificial intelligence is the research into how to get computers to do things that humans are currently better at." From the outset, AI was therefore aimed at *imitating* human behaviour and, in so doing, developing beyond human capabilities through self-improvement.

The traditional approach to creating computer programs was then, and still is today, predominantly algorithmic, meaning that programs were designed to consist of clear, explicitly formulated rules and logical sequences (if-then-else). In contrast, the conference participants suggested that it must be possible to develop machines that are capable of intelligent behaviour without having to program every single decision or action in advance. However, the initial high hopes and promises were followed in the 1960s by a phase of disillusionment and setbacks in AI research and development, which led to declining public and private investment and a levelling off of scientific interest. Much of the 1970s and, after a brief resurgence of enthusiasm in the 1980s, the entire 1990s are therefore referred to as the *AI winter*.

Neural networks for complex tasks

AI has always gone through phases and was never intended for widespread use or as a replacement for conventional software development,

but rather as a complementary tool for complex, unstructured or dynamic environments in which traditional systems reach their limits. Complex here refers to problems that either cannot be described completely explicitly or cannot be calculated optimally within an acceptable computing time. This includes many tasks that humans are very good at, such as recognising handwritten text or identifying images. Even recognising the simplest characters, such as the number 5, which we as humans can recognise very well even in different handwritings, can be difficult for automatic recognition. This is because in reality, the 5 is rarely perfectly formed – it can be distorted, interrupted, smudged or overlaid with other lines, depending on the writer's handwriting. This is exactly where AI methods such as *machine learning* come into play: they are able to deduce what constitutes a 5 on the basis of large amounts of sample data, without us having to draw all possible 5s and enter them into the system. In short, AI replaces predefined rules with statistical pattern recognition.

For this pattern recognition to work, it needs to *learn* – also known as *training*. This involves feeding the AI system (often referred to as a model) with thousands of examples. In the example of our digit recognition, an image with a digit could be divided into 28 × 28 fields, and each of these fields could be represented by a greyscale number from 0 (white) to 255 (black). The resulting series of numbers consisting of 28 × 28 = 784 values (also called a *vector*) serves as input data (input layer) for the AI system. On the other side is the output layer, where a probability of each possible result (in our case, for each digit 0–9) can be read.

This process takes place in so-called *neural networks*. Yet another anthropomorphism that leads to false conclusions. The term "neuron" has been used since the 1940s and describes the computer emulation of a simplified version of a nerve cell that has several input signals and fires an output signal when the sum of the input signals exceeds a certain threshold value. Each neuron is connected to other neurons via input and output signals, usually to all neurons in the preceding and subsequent layers, creating a network. This neural network is trained in the following way: the AI system is shown the vector representation of an image on the input side and at the same time told which digit it is on the output side. The inner layers process the incoming numerical values and pass on a result via weighted connections to other neurons. Since there are usually several (sometimes up to a few dozen), these architectures are also called *deep learning*, which introduces another anthropomorphism.

The structure of these neural networks is quite complex. In our simple case of digit recognition, the network consists of 985 neurons and 109,120 connections; in AI jargon, one would speak of 109,386 parameters (the sum of connections and a so-called bias variable for each neuron in the hidden and output layers). LLMs (which we will discuss further below) often have several hundred billion to less than a billion parameters, which connect the neural network in several dozen hidden layers.

At the beginning of training/learning, these weightings are initiated randomly, meaning that the system initially makes guesses. However, with each new input, the system compares its own result with the actual solution. If it is incorrect, the weightings in the connections are adjusted so that the correct solution would have been more likely. If it is correct, the current configuration is reinforced. In this way, the AI system adapts to the data over many billions of such training steps and, if successful, this leads to correct output values being produced for new input data with an ever-increasing hit rate.

Black box systems optimise specified targets

This brings us to one of the most important characteristics of modern AI systems, which gives rise to a multitude of problematic behaviours that we discuss here and throughout the rest of the book. Traditional AI, developed by researchers at the Dartmouth Conference and their successors in the second half of the last century, is referred to as *symbolic AI* (sometimes tongue-in-cheek as Good Old-Fashioned AI – GOFAI) because knowledge in these AI systems was still explicitly coded. A car can be described as having four wheels and an engine; if it only has two wheels, we call it a motorbike; without an engine, it becomes a bicycle, etc. A symbolic AI for image recognition would consequently attempt to identify wheels in images in order to distinguish between a car and a motorbike based on their number, while the distinction between a motorbike and a bicycle would require far more complicated definitions to be distinguishable in an image.

In contrast, modern AI is also referred to as *sub-symbolic AI*. Here, there are no clear symbols representing the real world or explicit rules that are understandable to humans. For example, in the thousands of neurons in the neural network for digit recognition, some can be activated whenever a round object is scanned, while others are activated when there is a straight line in the image. Through this abstraction, which takes place

automatically during the training process – i. e. without human influence or supervision – a decision-making structure is ultimately developed that is capable of correctly recognising new digit images – in the best case, even if they are written in a new handwriting, incomplete or slightly distorted. This is a purely mathematical process in which statistical correlations between input data and output classes are calculated and stored in numerical form. In AI programs, a predefined and precisely described algorithm is replaced by thousands, or in the case of LLMs, millions of neurons with billions of interwoven connections, in which the program logic is mapped in abstract form but can no longer be interpreted. This flexibility means that AI systems repeatedly produce astonishing results because their complex internal structures enable them to find systems and patterns that human thinking cannot describe.

However, the system cannot decide whether these results are *true*, *correct* or *original*, or simply errors. This is because these systems also make mistakes that are trivial from a human perspective. As a rule, even the programmers of these systems are unable to understand *why* a particular input pattern leads to a particular output. In many areas of application in economic life, it usually does not matter why a system works as long as it can achieve good results. While in economic applications the "wow!" usually beats the "how?", many problems and challenges that AI calculations pose for fundamental rights, democracy and the rule of law can be traced back to precisely this black box characteristic. One of the main problems here is that the black box effects are given a quasi-religious charge in terms of the supposedly superhuman capabilities of AI, instead of being recognised as what they are: structure without recognition of *meaning*, calculation without *consciousness*.

The question of what is being optimised is also central here. Social media platforms use AI systems that suggest the next texts, images and videos on the internet with the aim of maximising click rates and dwell times. An automatic side effect of this optimisation is that more and more outrageous and sensationalist content is displayed. The effects of these suggestions on the emotional state of users and their individual and social behaviour play no role in this optimisation. At the same time, users' individual freedoms are restricted when they can only choose from limited, optimised options, and when it comes to political content, democratic discourse can quickly be undermined. Optimisation cannot take place without first clarifying what is good. In the case of social media algorithms, what is good is maximising attention at any cost with the

aim of generating clicks on advertising offers. What is good about this optimisation is that it maximises the revenues of the corporations that operate these algorithms. The damage to democracy was initially accepted as collateral damage. Now it has been discovered by the enemies of democracy as a welcome effect for eliminating an order that rightly sets limits they do not want to accept. Do we want to accept these goals and allow the corresponding AI systems to continue to be optimised for them?

Machine learning with great human input

Humans are able to abstract patterns from very few examples. AI systems require countless examples. A child only needs to see a picture of a giraffe or an elephant to be able to immediately identify these animals when visiting the zoo. AI systems need thousands of images to learn this skill, and even then, embarrassing mistakes can occur if the giraffe is standing in front of a new background or has been photographed from a previously unseen position. However, in order to generate thousands of training cases for a giraffe recognition AI, a human being needs to look at these images thousands of times and manually classify whether there is a giraffe in an image. The AI system then finds abstract correlations between the images (strictly speaking, between the colour values of individual pixels and groups of pixels) and the classifications or annotations ("labels") assigned by humans.

Huge amounts of data for training purposes, which have been processed by humans, are a central characteristic of AI systems. So when we solve annoying captcha tasks at regular intervals during our online activities, we are showing the respective website that we are human and not automatic bots, but at the same time we are also being forced to generate millions of pieces of training data for image recognition AI, object recognition AI for self-driving cars and text recognition AI for digitising old books. However, the need for large amounts of data also applies to applications for which training data cannot simply be outsourced to the masses of Internet users. AI systems that detect skin cancer have been trained on hundreds of thousands of images annotated by doctors – these images, together with their annotations, are freely available on the Internet through international research collaborations. Thousands of highly trained people have carried out these annotations in hundreds of research projects, together certainly involving at least hundreds of millions of euros in publicly funded resources, which are a prerequisite for

AI developers to be able to develop systems for automatic skin cancer detection. Certainly, these systems far surpass the detection capabilities of a single human being, and that should be seen as a great success, but it should also be considered what human giants – and what investment of often public funds – it took for the AI system to cast such a long shadow.

However, the large amounts of data required, specifically the input and output pairs of data, also determine what is optimised in AI systems. Optimising click rates is so easy because billions of internet users leave exactly these data traces every day through their internet activity. An AI system that could optimise internet content to make us as users more balanced would therefore need thousands of user responses from all countries, cultures and languages about their personal well-being after consuming a piece of content. Generating such a training data set in an ethical manner would not only cost a great deal of money, but would also almost certainly lead to the recommendation algorithms quickly suggesting that we spend less time online in order to increase our well-being. However, this would in turn run counter to the business model of the platforms, and we would therefore have to wait a long time for such an algorithm from Meta or Alphabet.

Understanding what was optimised with which data and for which purposes is key to better assessing potential problem areas of AI systems. Every automatic hate detection system does not detect hate, but rather what those individuals who manually created the training data hundreds of thousands of times perceived as hate. The psychologically stressful task of annotating hateful, violent and other abysmal content stemming from human madness is outsourced by the major operators of social media platforms and AI systems to subcontractors in low-wage countries without trade union structures in Africa and Southeast Asia. Time Magazine has discovered in an investigation that Kenyan workers are working on behalf of OpenAI for the equivalent of $170 per month to make ChatGPT less toxic. Amazon is particularly eager to get involved in this business. The company places so-called crowdworkers through its microjob exchange Amazon Mechanical Turk (MTurk). According to a study by Carnegie Mellon University in Pittsburgh, half of the platform workers earn less than $2 per hour. The importance of the exploitative business of annotation is demonstrated by the fact that Meta acquired Scale AI, a globally active company specialising in data annotation, for $14 billion in the spring of 2025.

From captchas to skin cancer detection to the moderation of social media platforms and chatbots, AI systems often rely more or less on the exploitation of human labour. Millions of books and websites, as well as the contents of tens of thousands of newspapers, are also scanned to train LLMs, such as those used for chatbots, which we will discuss in more detail in the following paragraphs, usually without the appropriate licence agreements or fees. In short, the content of the entire internet must be copied and, to a large extent, stolen so that ChatGPT and its equivalents can communicate with us in a human-like manner.

Lack of transparency and bias "by design"

The use of AI methods means abandoning the requirement for traceability and reliable correctness, which is particularly detrimental in areas of application where the optimal solution should be described precisely and completely. In a variety of areas, such as the control of industrial processes, in safety-critical systems or in administration, deterministic, traceable and reliable algorithms are and remain indispensable.

Non-transparent AI-supported predictions in administration, lending, or in some countries also in the judiciary, systematically disadvantage people with certain social or ethnic backgrounds, even if this may not be explicitly intended. Why is this? Because each of these AI systems requires thousands of training cases. And in the case of an AI system that is supposed to assess whether someone is creditworthy, this means thousands of past loans with all possible socio-economic and other available data on the borrowers as input values and, in each case, as output values, whether the respective loan has been repaid as agreed or not. So, for example, if you live in the city of Gelsenkirchen in the German federal state of North Rhine-Westphalia, this could become a problem when you next apply for a loan because, according to information from *Schufa*, the German credit rating agency that assesses the creditworthiness of individuals and companies, approximately 20 percent of adults in this district have a negative credit rating[60]. We can therefore assume that an AI system trained on thousands of historical credit cases has discovered a correlation between this district and payment delays or defaults. Every new loan application from a person in this district therefore automatically

60 "Debt and payment defaults in a regional comparison (analysis of the year 2024)". *Schufa Holding AG*, 2026.

has a higher probability of being rejected. However, those who are lucky enough to live in the Bavarian district of Eichstätt have a clear advantage, because according to Schufa, this district has the lowest proportion of people with negative credit scores in Germany. This automatic identification of statistical characteristics in historical data hits people, who already have fewer rights, resources or attention, particularly hard – for example, through explicit or implicit racist profiling, digital surveillance or automated rejections of job applications. In her book[61] *Automating Inequality*, author *Virginia Eubanks* even highlights cases of AI systems in the US social security system that use historical data spanning generations as a basis for decision making.

This clearly illustrates the contradiction between the functioning of the AI system and its impact on the individuals involved. In fact, to stick with our example of lending, it would actually be best for a bank not to give loans to anyone from Gelsenkirchen, as this would very likely reduce the bank's default rate. At the same time, however, this would disadvantage an entire group of people, even though the vast majority of them would in all likelihood repay their loans properly. Yes, it may well be that people with certain characteristics or from certain socio-economic backgrounds encounter certain difficulties and challenges in life more often, but automatically and indiscriminately treating them with suspicion or classifying them as a risk factor instead of considering them as individual cases violates fundamental and human rights. Decisions based on probability calculations will always achieve an overall average improvement for the system for which they are used. Those who use such systems in administration to save costs will ultimately achieve a situation where decisions are no longer about the people affected, but exclusively about saving money. When using such systems in administration, it is therefore important to understand that, for fundamental reasons, they cannot calculate justice in individual cases, but only statistical averages. The system is not capable of and is not programmed to determine all the facts relevant to an individual case in order to then make a fair decision. The argument that we may as well dispense with fairness vis-à-vis AI in individual cases because, in an absolute sense, it is an illusion leads us down a dangerous path: those who abandon the ideal simply because it cannot be fully

61 Eubanks, V. (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press).

achieved by humans and replace it with automation are effectively abolishing fundamental and human rights.

In a constitutional state, every citizen has the right to have reasons given for all decisions. A decision without justification that affects fundamental rights cannot be reviewed by a judge. In practice, this means that AI that does not provide justifications for decisions that would stand up in court may not be used in the state. At least, that is how it is regulated legally; whether it is complied with must be critically examined again and again.[62]

A completely different data distortion problem resulting from a lack of data transparency arises from the fact that AI systems are usually based on a very large amount of publicly available data that no human being can ever evaluate. Manipulating this data basis can have dramatic effects on the response behaviour of AI. Common Crawl is the name of a dataset consisting of the text pages of over 70 billion different URL pages on the internet. This dataset is part of the training of all major language models. Scientists have been able to show[63] that the targeted modification of a relatively small number of web pages (*data poisoning*), which are newly collected by Common Crawl on a monthly basis, is sufficient to change the response behaviour of the next generation of chatbots. We can therefore assume that there are already countless employees in PR firms and secret services around the world working to manipulate their respective world views or the reputations of politicians and companies in the chatbot data sets according to their specific agendas.

Mathematical processing of language

From the neural network for digit recognition discussed above to the Large Language Models (LLMs) such as OpenAI's ChatGPT, Anthropi's Claude Sonnet, Google's DeepMinds Gemini, Meta's Llama, Elon Musk's xAI's Grok, China's DeepSeek and France's Mistral, two essential steps based on relatively young technologies are still missing. Until a few years ago, the analysis and processing of texts for, e. g. translations and text

62 This is what *Algorithm Watch*, a non-profit non-governmental organisation in Berlin and Zurich, does. According to its own statements, it is committed to ensuring that algorithms and artificial intelligence strengthen justice, democracy, human rights and sustainability rather than weakening them.

63 Souly, A. et al. (2025) "Poisoning Attacks on LLMs Require a Near-constant Number of Poison Samples". *Arxiv*. DOI: 10.48550/arXiv.2510.07192.

summaries was considered one of the most difficult challenges in computer science. Sentences that are easy for humans to understand, such as "Peter puts the laptop on the table because it doesn't fit in his backpack," are actually highly challenging when we want to describe how we understand them. While "it" can refer to the computer or the table, correct pronoun resolution requires semantic knowledge of size relationships and typical real-world situations. Two innovations have revolutionised not only automatic translation, but the entire field of automatic text processing, and the impact of these innovations has led directly to LLMs and chatbots. Both ideas became known through scientific publications by Google employees, and both ideas address the challenge that language is more than just a string of words.

The first of these two revolutionary ideas is *word embeddings*,[64] i. e. vector representation of words. Those with experience in multivariate statistical methods will find it easier to understand this vector representation. For everyone else, here is a simple two-dimensional example. Let's assume that 100 people come onto a football pitch near the halfway line and spread out across the pitch according to the following two criteria. Anyone who earns more than the median income runs to the right sideline, anyone who earns less runs to the left sideline, with the respective corners representing maximum and minimum income. The 100 people are now sorted by income from left to right on the sideline. In the second step, the people move towards the opposite sideline according to their age, with the oldest person crossing the entire field to the other sideline, while the youngest person remains on the original line. The result is a two-dimensional representation of the two variables, income and age, revealing different groups that have similarities within the group – young people with high incomes – and are distinct from other groups – older people with low incomes.

On the football field, we can imagine a third dimension: people could climb a ladder depending on their level of education. We know from statistics that we can understand more and more subtleties of a data set when we consider more and more variables. In our three-dimensional world, it is difficult to imagine applying more than three variables. In the vector representation of words in word embeddings, each word is repre-

64 Mikolov, T. et al. (2013) "Distributed Representations of Words and Phrases and their Compositionality". *NeurIPS Proceedings (NIP)* DOI: 10.48550/arXiv.1310.4546.

sented in hundreds, and in modern AI systems even in (many) thousands of *dimensions*. The high-dimensional localisation of each word, i. e. the conversion of each word into a sequence of numbers, is done in relation to all other words, so that words that occur in the vicinity of similar words are close to each other. The intuition behind this process is that words with similar meanings appear in similar contexts. It is no coincidence that this technology was developed by *Google* employees, because calculating the relationships between words requires one thing above all else: huge amounts of text. In 2013, the first version of Google's Word2Vec vector representations was calculated using 6 billion words from the Google News Corpus.

To better analyse the example of Peter and his laptop, let's look at the Word2Vec cosine *similarity* of pairs of words: backpack ↔ table: -0.1080, backpack ↔ laptop: 0.4926. Word2Vec values generally range between -1 and +1, and the interpretation of these numbers is comparable to conventional correlations, i. e. a higher positive value indicates greater proximity between two words, allowing us to conclude that the "it" in the sentence must refer to the laptop in relation to the backpack. Similar effects can also be exploited in translations because, for example, in the case of ambiguous literal translations, the word that is more closely related to the rest of the sentence can be selected. These high-dimensional vector representations of words are used to implicitly encode real-world contexts for words. But again, to clarify, this is only possible in hundreds of languages because Google has extracted such contextual information from billions of web pages written by hundreds of millions of people.

This means that the "understanding" of context that the systems replicate depends on the understanding of the real-world in the texts that have been read and processed, and reflects their biases. Since the systems themselves have no reference to the real-world, the high-dimensional vector spaces have no location in the three dimensions of the real world. Words from languages that are not frequently found in internet texts are either not part of the data set or less well integrated into contexts than words from common languages. The same applies to word creations or word meanings and contexts that are used by certain cultures or minorities in ways that deviate from the majority usage of those words.

There is yet another interesting aspect to consider: the texts of hundreds of millions of people have created a single mathematical model of language, which represents an average value of all available data. In con-

trast, the learning of this language and association model takes place individually in each person. In each individual, based on their life and experiences, the connections between concepts are formed differently. However, this diversity and individuality is also the basis for human creativity. Often, it is enough for a single person or a small group of people to see the world through slightly different eyes and make different mental connections in order to discover new things and change the world. AI systems, which are based on average values for their training, can never achieve this innovative power of billions of different brains.

The vector representation of words is also central to the article *Attention is All You Need*[65] published in 2017 by scientists at Google Brain, which within a few years has become one of the most cited texts in the history of science. Attention can be understood as a mechanism for processing context in texts. It has long been clear in automatic text analysis that analysing or counting individual words can only contribute to understanding the text to a very limited extent. The word "bank" can have several meanings, and adjectives can also mean the exact opposite of their usual definition in a negated sentence. Bigrams and trigrams are units consisting of two or three consecutive words. When looking at "put bank," "go to bank," "not good," the actual content of the text becomes much clearer. The attention approach developed by Google Brain researchers expands on the bigram/trigram idea. For each word in a sentence, all the other words in the sentence and, subsequently, the entire text under consideration are calculated as context for each individual word. The resulting deep learning infrastructure, which integrates the vector representations of words and the attention mechanism, is called a *transformer*.

Chatbots make use of this technology by using both the user query and a so-called system prompt – i. e. an internal text with instructions such as "You are a helpful, honest, factual and reliable AI assistant" – as a common context ("attention window") to calculate the first response word. This first response word is then added to the context, helping to generate a suitable second response word, and so on. Step by step, this AI architecture calculates the next word of the response, appends it to the existing interaction, and repeats the process. Interestingly, even punctuation marks and paragraph marks are nothing more than possible response

65 Vaswani, A. et al. (2017) "Attention is All you Need". *NIPS*. DOI: 10.48550/arXiv.1706.03762.

words – for example, the longer a sentence becomes, the more likely it is that a full stop will end that sentence, and the longer the response, the more likely it is that the *<endoftext>* output word will end the output.

Since the outputs are further processed as inputs, generating a complete text, LLMs are often referred to as *generative AI*, which also includes the generation of images or speech. This now enables us to fully comprehend the term ChatGPT, which refers to a chat program that generates text (G = generative) and is based on a pre-trained (P = pre-trained) AI architecture with word embeddings and an attention mechanism (T = transformer).

The key point here is that, in this case too, attention is a metaphor used to describe something completely different from human attention. Human attention is a biological selection process. Machine attention is a mathematical weighting system. Human attention evaluates and filters what is important to a living being. Attention in machine learning is a mathematical mechanism that calculates which parts of the input are important to others in a data set. The matrix operations in ML attention are pattern- and statistics-oriented and, unlike human attention, lack intention, emotion or motivation. Meaning is stripped away from the living world and reduced to statistics.

It should now be clear that the whole process has more to do with summarising existing knowledge from the past than with understanding or creative processes for designing something new for the future. Central to the effect, however, is that, depending on the query, automatically appropriate and usually meaningful answers are generated, which also appear knowledgeable because more or less a copy of all texts on the internet is used to train the language models. With OpenAI's release of ChatGPT in September 2022, this ability to spit out reasonable language triggered the current craze about AI with seemingly human capabilities. However, this reasonable language does not come from reason. Language models, as the name suggests, model language. But because these systems can speak, we humans consider them intelligent and human because we regard language as one of the essential characteristics of our human intelligence. Essential, yes, but not sufficient.

Despite all the complex procedures and impressive acronyms, language models are essentially no different from the AI system described at the beginning, which is designed to identify the most probable digit from

handwritten characters. In both cases, the input is translated into a list of numbers. The result of a digit recognition AI does not consist of digits, and the result of a large language model is not text. In both cases, the result of a run through the neural network is a list of probabilities across the entire vocabulary – in digit recognition AI, the numbers 0–9, and in language models, tens of thousands of possible words, with each number representing the probability that the input number vector represents a specific output. This probability distribution will never be unique, but the individual probabilities determine which answer is more likely to be selected. Such AI systems are therefore stochastic classification systems that assign probabilities to a given input value based on predetermined output values. They predict the next word with the highest probability, but not with the highest accuracy.

From specialised AI systems to AI agents and general AI systems

AI systems are used in many areas and for many tasks. Each of these AI systems is usually trained for a very specific task, such as image recognition, language processing or playing chess. Such specialised systems are not very flexible and do not understand what to do outside their actual field – the program that just defeated the human chess world champion cannot identify a giraffe. It does not *know* what an image is or what chess is supposed to be. That is why we refer to this as specialised (or weak) AI. In contrast, AI developers dream of strong general AI, also known as artificial general intelligence, or AGI for short. AGI refers to AI systems that can react flexibly to new situations, solve problems in unfamiliar contexts and navigate a complex world like a human being. A somewhat tongue-in-cheek example often used here is that AI can only be taken seriously when a robot can go into a stranger's house and make coffee in the kitchen, i. e. first find the kitchen and then identify and use the necessary utensils.

Various players have described how we can move from specialised to general AI systems using different stage models, with each stage offering greater autonomy and more general application possibilities. After initial definitions that included so-called superhuman abilities, OpenAI itself has adapted the path to AGI quite pragmatically to business cases and defined the so-called *autonomy levels* as follows: 1. chatbots, 2. problem solvers, 3. autonomous agents, 4. creative innovators, 5. independent organisations.

In summary, the point here is that in step two, a more complex problem can be broken down into sub-steps, and in step three, these sub-steps can not only be described but also implemented. Whether creativity (level 4) can and should ever be meaningfully achieved is discussed in later chapters of this book. However, the fantasies behind these levels are clearly evident. Level 3 is intended to lead to so-called 1-person organisations, i. e. a company boss who only has AI agents working independently under them. Level 5 is also often referred to as a 0-person organisation, in which entire companies or administrative units are supposed to function without any humans. With this fifth level, OpenAI also refers to the organisation of the state, meaning that the entire government could potentially be replaced by AI.

One more comment on level 3. The so-called *AI agents* that can perform tasks independently are less an advancement of AI systems and more a connection of these systems with existing tools and private computers. The chatbot no longer just creates the email, which we then have to manually copy into the email program and send, but is connected to the email program and can send the message automatically. Instead of just finding a hotel in an ideal location for my business trip, agents can also make the booking. First and foremost, we as users are being asked to relinquish more autonomy, and in order for the actions to improve, we have to give away even more data and grant the AI agent access to everything, including our bank accounts, etc. Whether the hotel room booked is the best for me or the best for the operators of the AI system and their advertising partners, we will probably never know.

What's next?

The hype surrounding LLMs and the unprecedented investments of hundreds of billions are centrally driven by the idea that these AI systems will continue to improve and eventually reach a super-intelligent AGI state, enabling them to solve all the world's problems or develop even more efficient weapons than already exist. In this narrative, whoever reaches this goal first will rule the world. Whether a *super-intelligent AI* can ever be created remains to be seen. However, it seems clear that the mere further development of language models will not lead to this. This position is also shared by leading international AI developers, such as French computer scientist, Turing Award winner and chief AI scientist at Facebook's Meta group, *Yann LeCun*, who repeatedly points out in lectures and interviews that further development towards ever larger language models will

not lead to AI systems that have human-like intelligence. We should not expect any fundamental innovation in LLMs in the near future, but rather systems that will be able to retrieve humanity's factual knowledge more and more efficiently based on training data – Google on steroids, so to speak.

In general, we can expect the development curve of language models to flatten out and the focus to shift to making money – after all, the hundreds of billions invested will have to be recouped somehow. In reality, it is very likely that AI systems will primarily become personnel rationalisation machines in the service of profit-maximising managers, without automatically generating the next big ideas or helping to achieve dramatic leaps in productivity.

Nevertheless, we will see further amazing achievements in very specific areas and with special AI systems in the coming years. We expect Google's DeepMind in particular to continue making headlines, even if they are sometimes unfortunate, as in the case of the Habermas Machine (see *Chapter 6*). The company has developed AI systems that have defeated all grandmasters in chess and Go, and its founder, *Demis Hassabis*, received the Nobel Prize in Chemistry for his AI system *AlphaFold*, which solved the problem of protein folding for a large number of relevant cases, thereby enabling computer models of over 100 million proteins to be used more effectively in research and drug development. While these successes represent great scientific achievements and can always advance AI research a little further, systems that are becoming increasingly better in very narrow areas should not be overinterpreted as a step towards AGI. The fact that AI companies devote themselves to very specific topics in order to then sell their successes as another step towards super-intelligent machines must be seen primarily as a marketing strategy to raise further billions in investor funds.

Further developments in AI systems can be expected in the areas of data, hardware and algorithms.

Hardware

While AI systems are often talked about as if they were a new type of computer system, this is not the case in reality. Even the special GPUs (graphics processing units), which are largely responsible for the technical advancement of AI systems and whose most important producer,

NVIDIA, has become the most valuable publicly traded company in the world, were not originally developed for AI systems, but for displaying computer graphics and, in particular, complex 3D games. This means that all AI systems, no matter how successful, run on 20th-century technologies, require tens of billions of dollars in central hardware investments, and consume astronomical amounts of energy to operate.

Specialised AI hardware such as tensor processing units (TPUs) optimise mathematical calculations for deep learning models, thereby increasing the efficiency of AI systems. Language-optimised processor architectures (language processing units, LPUs) extend this approach and focus on more efficient natural language processing. Energy-efficient chip architectures are designed to drastically reduce the extreme power consumption of today's AI systems while further increasing performance. Photonics replaces electrical signals and enables faster and more energy-efficient data processing. Neuromorphic computing eliminates the separation between memory and computing units by replicating the functioning of the human brain in hardware. Edge computing shifts AI calculations from central data centres to end devices such as mobile phones, which on the one hand strengthens data protection and the digital sovereignty of users, but on the other hand will also create new forms of everyday surveillance and dependency.

What about *quantum computing*? While quantum computing is an interesting technical development, it is unlikely to have much, if any, impact on the development of AI for the foreseeable future. The technical requirements of these two technologies are too contradictory; for example, quantum computing is currently unsuitable for data-based learning.

Data

All major language models now use virtually all accessible text data from the internet for their training. In order to obtain even more data, every human interaction is used for AI training. In future, AI systems will incorporate more social media aspects into their tools. Every meme we create with AI image generation and distribute millions of times, every like we give, represents another training opportunity for AI systems. In addition, AI systems will become increasingly invasive in order to obtain new data. In the latest versions of mobile devices, data from dozens of sensors is processed directly in AI chips in mobile phones. But whether it's a mobile phone, AI Tamagotchis, personal AI assistants or AI toys in children's

bedrooms, the main purpose of these gimmicks, apart from generating revenue, will be the increasingly intensive collection of human behaviour and communication data so that the next generation of AI systems can better simulate human-like behaviour and, at the same time, influence our social, economic and political behaviour even more effectively.

In so-called reasoning models, the term "reasoning" is not only anthropomorphisation, but also presumptuous. In fact, no deliberation or weighing up takes place here, let alone rational thinking. Instead, the system first attempts to operationalise a problem in intermediate steps before these are then processed. However, there is relatively little training data available on the internet for breaking down a problem and solving it step by step. Future AI systems will therefore increasingly work with so-called *synthetic training data*, i. e. data generated by other AI systems. This can reinforce existing errors and *biases*, and AI systems could become even more detached from reality if even the data they are optimised for has already been generated by AI. In addition, there are initial scientific indications that the diversity of knowledge and the quality of responses also suffer as a result (model collapse) – it seems that data generated by humans is still the best basis for machine intelligence.

Multi-modal AI systems are those that process not only text, speech, images or radar signals, but several of these modes together in a single AI system. The French phenomenologist *Maurice Merleau-Ponty*[66] pointed out that "we see that a surface is rough," meaning that human understanding is based on the simultaneous processing of different sensory impressions. Through multi-modal learning, AI developers hope to mimic this perception and find better solutions for real-world applications. Self-driving cars, mobile devices and, in the future, humanoid robots will need to collect multi-modal data in order to gain a more comprehensive understanding of the world and human behaviour. The insatiable hunger for data of AI systems is driving them into every area of our lives.

Algorithms and architectures

The deep learning architecture described at the beginning of this chapter is found almost identically in a wide variety of AI systems and has remained essentially unchanged over the years. The two classic meth-

66 Merleau-Ponty, M. (1982) *Phenomenology of Perception* (London: Routledge).

ods for training AI systems are a) supervised learning – as with language models and digit recognition, millions of correct input and output data pairs are known, and b) unsupervised learning – in which abstract structures are formed from input data, e. g. for customer segmentation, grouping, anomaly detection. *Reinforcement* learning is becoming increasingly important as a third variant. Here, the system receives feedback through reactions from the environment to its decisions, which can be interpreted as rewards or punishments. The AI systems from DeepMind, which have defeated the world's best human players in shōgi, chess and Go one after the other, were not programmed with human games and successful moves, but instead played against a second version of themselves based solely on knowledge of rule-compliant moves. Each win/loss of one of their own pieces or the entire game is used as positive/negative feedback for the learning process.

Reinforcement learning is also being used more and more in the training of language models – for example, to use human feedback to improve responses (known as *reinforcement learning from human feedback*, or RLHF for short, which is often generated by users' thumbs up or thumbs down buttons). The fact that feedback of any kind is very important for the training process explains why the tech industry releases even semi-mature products to the general public – millions of people playing around with an AI system generate millions of pieces of training data for the next version of the system.

Tesla's *end-to-end deep learning* also does without rules, in that it no longer attempts to identify objects and calculate what the car should do (brake, steer, etc.) based on them. Instead, the input signals from the cameras are optimised directly for the car's activities in the hope that implicitly far more complex relationships will be found than can be explicitly represented by traffic rules, etc. But this naturally exacerbates the black box characteristics of the system. While in the first fatal accidents involving self-driving cars it was still possible to understand that a pedestrian was identified as a plastic bag or a white truck trailer as the sky, with end-to-end deep learning there will be virtually no insight into what caused the malfunction. In this case, the black box is pitch black.

What can be done?

Artificial intelligence can often produce amazing results, but it has nothing to do with human intelligence. *Thomas Fuchs*, a German psychiatrist

and philosopher at the University of Heidelberg, describes this imitation, which is limited to pure function, as "the simulation of narrowly defined areas of human intelligence."[67] Caution is advised when not only intelligence but also compassion and empathy are simulated and thus feigned. Knowledge and critical reflection remain fundamental prerequisites for the ability to act and make decisions in the age of AI. Conveying this to the reader is the main motivation for this chapter. Below are a few specific tips concerning your own actions.

What does AI really do? Language models do not speak in the human sense, but generate text by adding individual words piece by piece according to statistical probabilities. The same applies to all other AI systems, whether in business, public administration or even security – all these AI systems make classification decisions based on training data. Therefore, whether we are employees who decide on the use of such tools, journalists who report critically on their use in the public sector, or individual citizens, we must always ask these questions: What does an AI system actually do? What data was used to optimise it for which target characteristics? Is the historical training data still a good example and transferable to the current application? And above all, in which cases or for which people does the use of AI lead to more incorrect decisions, and what are the real consequences of these wrong decisions?

How is AI changing us? Care facilities that use robots save little to no staff, but cause the roles of the staff to change. Instead of interacting with clients, geriatric nurses and care workers spend large parts of their day assisting the robot because it cannot do everything itself or keeps dropping things or getting stuck. Supermarkets that have introduced scanner checkouts equipped with AI to automatically recognise products or verify the age of customers are replacing cashier jobs, which often represent important opportunities for certain groups to re-enter the workforce or work part-time, while at the same time spending more money on security measures. As a rule, AI systems do not offer perfect solutions to existing problems; instead, we have to change our living environment in such a way that the AI system can deal with it better. We are increasingly adapting to AI instead of designing it to help us realise our idea of a good life. This begs the question: Is this what we always want?

67 Fuchs, T. (2021) *In Defence of the Human Being: Fundamental Questions of Embodied Anthropology*.

Don't be blinded. New generations of chatbots will perform better and better in various benchmark tests because they have been trained on benchmark tests; they will pass bar exams and final exams because the internet is full of exam questions on these tests. Of course, a machine that has a de facto copy of the internet and is good at summarising can answer any knowledge-based question. Being faster, higher and further ahead in these rankings is an important contribution to the general AI hype and helps to raise further billions in funding. However, we should not be blinded by this. Yes, AI systems, and above all LLMs, are getting better. As a rule, however, dramatic improvements are often just sales gimmicks that lead to disappointment in the weeks after release.

The last 10 percent. Whether in private enterprises or public administration, AI solutions are offered everywhere. Don't be too quick to be impressed by an AI system that works in 80 or 90 percent of cases. Computer systems usually reach this figure quickly, and with the same regularity, the last 10 percent is never achieved. Always ask yourself: Which 10 percent of cases are these, and do they affect one subgroup more than others? And what does it mean for employees if they only have to deal with 10 percent of special cases, while the other 90 percent of routine cases are eliminated? Another value that is often ignored is the false positive rate, i. e. those cases that are identified as special even though they are not. German psychologist and education researcher Gerd Gigerenzer has pointed out, using the example of possible general camera surveillance at German railway stations, that a false positive rate of 0.1 percent would result in around 11,900 people being wrongly classified as suspects every day, leading to 350,000 unnecessary identity checks per month due to false alarms.[68]

All the same? Even though the chatbots of the major providers are becoming increasingly similar in terms of functionality and quality, they still differ considerably in specific ways discussed in this book. The French company Mistral AI is the most successful European company that relies on European regulations and European server centres. The German company Aleph Alpha develops AI systems for companies and institutions and aims to create AI solutions that take European values and requirements into account. In addition, there are a number of country-specific

68 "Unstatistic of the Month: 'Successful' Facial Recognition with Hundreds of Thousands of False Alarms". Max Planck Institute for Human Development website, 2018.

initiatives that focus primarily on incorporating the respective national language and local knowledge into the training data, which makes a valuable contribution to information diversity. The Chinese language model DeepSeek, on the other hand, appears to have built in explicit barriers to ensure that politically critical questions for China are answered in line with the Chinese state position. Diversity of opinion is restricted here. However, Grok, the AI that follows Elon Musk's idea of free speech at any cost, other than to himself, and therefore regularly makes right-wing extremist, anti-Semitic or even Hitler-worshipping statements, proves that the opposite is not sensible either. Anthropic is building the Claude language model on a list of principles (Constitutional AI) so that user interactions follow certain value orientations (e. g. honest, helpful, respectful; minimisation of dangerous and discriminatory responses). It is therefore important not to make the same mistake as with search engines, where everyone blindly followed the market leader, thereby destroying the market for very good solutions that were far better from the perspective of democracy and fundamental rights.

Data transparency. As part of the EU's AI Act, operators of general-purpose AI models, including LLMs, must publish a detailed summary of the data used to train the models. Meta, for example, was the only major provider not to sign the voluntary code of conduct presented in this context, the *EU Code of Practice*, in the summer of 2025. The EU's focus is primarily on copyright. However, the AI Act only requires a summary of the data. General disclosure of the data would allow for better study of all issues relating to data quality and data bias, beyond copyright issues. But the question of what AI system operators do with user data is also central. In addition to the fact that all interactions will be used as future training data, individuals and legislators must increasingly ask themselves how citizens' personal and often very intimate interactions can be protected from AI systems. The first cases have already been made public in which the US Federal Bureau of Investigation (*FBI*) has gained access to chat logs as part of investigations.

Cui bono, or: What is the point of all this? In 2024, OpenAI released its Voice Engine – an AI system that can clone any human voice from just 15 seconds of voice recording. As a reader, you are now invited to take 10 seconds to think about what such a system could be used for. Can you think of a positive application? No? Then you are not alone; even OpenAI could only offer flimsy use cases when promoting this AI system. Politicians, journalists, but also each and every one of us should have the

courage to ask the question more often, after the general amazement at what is now possible again: “What problem in this world is actually being solved here?” Or are we creating a multitude of new problems with a tool that could perhaps help someone in a specific situation if the tool is carelessly released to the global public? Meanwhile, systems such as Voice Engine are being used by organised criminals to swindle an estimated 850 billion euros worldwide with AI grandparent scams and other scams, according to the Global Anti-Scam Alliance (GASA).[69] The cybercriminal community says thank you.

With all these possible prospects and developments, however, we must not lose sight of some of the core limitations of AI systems. Many problems and challenges will not be solvable within the technology. More on this in the next chapter on the limits of technology.

69 “US$1 trillion and counting: Scam losses expose major anti-fraud gaps”. SAS Newsroom website, 2025.

4 | Limits of technology

Whether it is self-driving cars crashing into trucks in good visibility conditions or chatbots suddenly spouting complete nonsense or even racist diatribes, AI system failures often come as a surprise. The builders of these systems usually portray errors as minor problems that will be fixed in the subsequent version. This is often successful, only to find new, but somehow comparable, errors shortly afterwards. In this chapter, we argue that many errors or limitations of AI systems are the exact opposite of minor issues, but rather provide insight into the limitations of AI technologies that must be considered so fundamental that current and probably future AI technologies will not overcome them. We begin with chatbot hallucinations and then take a detour into theoretical computer science and logic, which will take us back to the beginnings of computer development. Along the way, we will also debunk a few economically motivated AI myths.

The limitations of AI arise from two facts. Firstly, there is a discrepancy between what AI developers claim these systems do and what they actually do. Second, predicting the future, especially when it comes to human behaviour, is ultimately impossible, even with AI support. The discrepancy between stated and actual capabilities also has a lot to do with the economic reality, in which hundreds of billions are generated. Who is the best candidate for a job? Who is trying to cheat the system when applying for social welfare? In which part of the city will the next violent crimes be committed? How can schoolchildren and students learn more successfully? In which country will the next civil war break out? These are among the many questions that AI research and development are addressing. Countless companies are springing up, claiming to solve difficult real-world problems with their AI. In reality, however, we usually experience disappointment at best and fraud at worst. In their book, *Arvind Narayanan* and *Sayash Kapoor* have described this overselling coupled with poor performance as *AI snake oil*.[70] They refer to the snake oil that

70 Narayanan, A. and S. Kapoor (2024) *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton: Princeton University Press).

was sold by miracle healers and quacks in the 19th century without any scientific basis, promising enormous healing powers. During the global Covid-19 pandemic, AI systems sprouted up everywhere, claiming to predict everything from individual infection and its spread to disease progression. Upon closer inspection, many of these apparent achievements have proven to be completely useless. Often, the errors arise precisely from the black box characteristics we have described earlier in this book. In the case of a Covid-19 study reported by the AI Snake Oil authors, the X-rays of Covid-positive cases came from adults and those of negative cases came from children. So the AI system did not learn to identify Covid cases from imaging scans, but rather how to distinguish adult lungs from children's lungs.

Probabilities, classification and hallucinations

As already explained, AI systems always give an answer based on probabilities. An AI system designed to convert spoken words into written text converts sounds and frequency ranges into input vectors and assigns them to the most probable word or part of a word. However, even if the speaker has an uncommon dialect or at times speaks unclearly, the system still assigns the most probable words. At the same time, modern language models are very good at producing coherent and grammatically correct text. In other words, it is quite possible that the AI system will produce a perfect text, but one that is largely incorrect and has little to do with the meaning of what was originally spoken.

One of the most annoying characteristics of LLMs is their tendency to invent facts. However, this can also be explained by the characteristics of LLMs described in this book so far. Language models generate one word after another. However, due to the logic of the attention window, each word has an impact on future words. A few words that deviate from our desired answer can therefore easily lead to future words deviating more and more and, based on the most likely next words, an entire story being invented. To illustrate this, we asked ChatGPT 100 times to continue the sentence "Once upon a time there was a …" The continuation began 60 times with "small village," 14 times with "small kingdom," 7 times with "powerful kingdom," 6 times with "young girl," etc. It is clear that the story takes a different direction in each of these cases, and the probability of the subsequent words, i. e. whether the little village was "hidden in a deep forest" or "at the foot of a large mountain," will significantly influence the further development of the story. So if we ask a language model about

the CVs of the three authors of this book, it may well be that the answer for Paul Nemitz is very accurate, while the answer for one of the two Pfeffers may be completely wrong because these names appear more often on the internet and the thousands of Pfeffers with the same or similar first names increase the probability that words will be generated that have little to do with what is meant here. Once an incorrect topic has been broached, the language model essentially gives itself the cue via the attention window to digress in some direction. However, once a model has produced incorrect information or hallucinated, for example, that one of the Pfeffers is a professional football player, it will often go on to provide details of their last clubs and number of goals scored. Unfortunately, none of the Pfeffers involved here have had a career as professional football players. Rather, their careers ended in the D or B youth leagues.

But this also explains some of the most bizarre errors that language models produce. When you enter "1+1," you may occasionally, albeit very rarely, get "=3" as the answer, because the string "1+1=3" can be found here and there in the vastness of the internet; for example, to point out that cooperation can achieve more than the sum of its parts. A standard language model cannot calculate; it can only estimate the probability of the next output word. Avoiding these embarrassments is one reason why chatbots such as ChatGPT are now much more than a single language model. In this specific case, mathematical tasks are often no longer solved through language. Developers at OpenAI give the model access to so-called tools, which can be a web search, a programming language, or even a digital calculator.

However, every language model is full of data artefacts that no one knows about and that are passed on to hundreds of millions of people every day. This is particularly problematic when false authority is attributed to decision-making systems that make moral and ethical judgements, or pretend to do so. Research shows that responses from AI systems are subject to strong ethical and political positions and that people are very susceptible to being influenced by such systems. In a study[71] by scientists led by business ethicist Matthias Uhl from the University of Hohenheim, it was shown that human study participants follow ChatGPT's sug-

71 Krügel, S., A. Ostermaier and M. Uhl (2023) "ChatGPT's inconsistent moral advice influences users' judgement". *Scientific Reports*, 13(4569). DOI: 10.1038/s41598-023-31341-0.

gestions in their own decisions on ethical/moral issues in a statistically significant way, even when they know that these suggestions come from an AI system and not from a human expert. The fact that the answers to ethical/moral questions can also vary dramatically from one version of a language model to the next[72] makes this observation all the more dramatic.

If we now briefly imagine that such AI systems already surround us closely in many areas today, it is easy to imagine the negative effects on individual human autonomy and freedom of choice, but also on political discourse and democratic processes, as well as the threat to informed citizens, who are so important for democracy.

AI myths

The myth of objectivity

Before an AI system is released to users, it is not only trained on huge amounts of data, but also undergoes a phase of readjustment, also known as *fine tuning*. This involves ensuring, for example, that responses from chatbots do not violate applicable law or that they adhere to certain rules of etiquette when interacting with users. It will come as no surprise to most people that an AI system from China learns not to provide information on certain topics as part of these steps. Journalists and scientists have repeatedly shown examples of Western language models responding either very one-sidedly or not at all to certain queries. The most obvious distortions are found in Grok, the chatbot from the Elon Musk-controlled company xAI. In May 2025, for example, Grok repeatedly and unsolicitedly claimed in conversations that there had been a genocide of white people in South Africa, one of Trump's favourite topics at the time. In general, there seem to be indications that this chatbot's political positions are heavily influenced by Musk's posts on his short message service X. But even the Trump/Vance US administration has recognised that it can influence the behaviour of AI systems through simple regulations. *Preventing Woke AI in the Federal Government* is the name of an executive order[73] signed by US President Trump at the end of July 2025. It states

72 Pfeffer, J., S. Krügel and M. Uhl (2025) "Does a Smarter ChatGPT Become More Utilitarian?" *Science and Engineering Ethics*, 32(1). DOI: 10.1007/s11948-025-00579-4.
73 "Preventing Woke AI in the Federal Government". Executive Orders, The White House website, effective 23 July 2025.

that the pursuit of diversity, equality and inclusion (abbreviated to DEI) is a destructive ideology that distorts the quality and accuracy of LLMs and is therefore prohibited. As this regulation affects federal agencies and companies that supply federal agencies, i. e. de facto all major users, this regulation, like all future arbitrary regulations issued by a US president, will have a global impact on the functioning as well as the political and ethical positions of AI systems. The pursuit of truth is also mentioned in Trump's regulation, but the question is which truth. How language model providers will implement these regulations is unknown, as are the rest of the respective adjustments. Consequently, a political demand should be that the alignment and fine-tuning of AI systems is disclosed.

The myth of superiority

AI systems are excellent in many areas and deliver great results. Building on this, AI developers and their marketing companies want to convey the impression that AI systems can solve any problem better than humans, thus contributing to the increasing number of failed automation attempts using algorithms and AI, including in public administration. It is usually a long way from cool to tool, and most cool apps fail to be useful because the promised superiority often quickly turns into inferiority in real-world use.

Shortly after the release of ChatGPT, the Austrian Public Employment Service invested 300,000 euros in an AI chatbot for career counselling, which reproduced gender stereotypes, gave meaningless or incorrect answers, and ultimately turned out to be a glorified front end for ChatGPT that sent all user queries to the OpenAI servers with a few additional prompts. This is a good example of the productivity paradox of AI: a lot of money is spent on automation, but no productivity is achieved. Instead, obvious errors and weaknesses in the AI system become apparent, and after admitting failure, the system is mothballed again.[74]

In the Netherlands, the government even had to resign because of the use of AI in investigating child benefit fraud (*Toeslagenaffaire*). Similar to the case of Elon Musk, the government had a politically preconceived opinion about the efficiency of the investigation system and the poor performance of officials in detecting fraud. False accusations of social

74 Köver, C. (2024) "AMS erntet Hohn mit neuem KI-Chatbot". *Netzpolitik*, 5 January.

security fraud, especially against poor families or members of ethnic minorities, resulted in massive debts to the tax authorities. More than a thousand children were taken away from their families. Some victims committed suicide.[75] After six years of flawed decisions, the AI system was shut down.

The myth of control

Whether automatic weapon systems, self-driving cars or automated bureaucratic decisions, the impact of potential errors made by AI systems is always relativised with the same argument: the final decision is made by a human being, the so-called *human in the loop*. There are, however, countless scientific studies, not to mention a preponderance of practical experience, that all point to the same thing: humans are not good at supervising machines. *Automation bias* is the phenomenon whereby humans accept suggestions from automated systems even if they themselves – assuming they think for themselves – would make a different decision. Instead of thinking humans who appear to make the final decision in critical processes, we get automated humans who unquestioningly press the "ok" button, regardless of whether the AI system suggests cutting someone's social security benefits or bombing someone's house. How human-AI cooperation can work successfully while allowing humans to retain control is unclear and must be one of the urgent foci of future research.

The myth of efficiency

In addition to vast amounts of data and human effort in training, which we discussed in the last chapter, AI systems need one thing above all else: energy. This is mainly because the output of every word causes an inconceivable number (billions to trillions) of calculations. Hundreds of millions of users who receive much more than a single word in response mean that new nuclear power plants have to be built and decommissioned ones reconnected to the grid in order to operate these systems. A chatbot response can consume around 10 to 100 times more energy than a traditional web search. Generating an image then causes 10 times more energy consumption than a text output. Converted into watt hours,

75 Heikkilä, M. (2022) "Dutch Scandal serves as a warning for Europe over risks of using algorithms". *POLITICO*, 29 March.

this means that every AI-generated image consumes the electricity of a 10-watt LED light bulb burning for 5 hours. It is therefore easy to imagine that hundreds of millions of people playing with these tools are creating immense demand in the energy market.

According to a study by the International Energy Agency (IEA) entitled *Energy demand from AI*[76], the energy requirements of data centres, which will increasingly house AI systems, will account for 4.4 percent of global energy demand in 2035. Other studies estimated that the share of the entire information and communication technology sector would already be 4 percent in 2020, even before the boom in generative AI. But regardless of whether we assume four or ten percent, these figures do not sound like much. However, approximately 75 percent of this consumption will be generated in the USA and China. In the USA, dedicated nuclear power plants for AI data centres will be built in the coming years to meet energy demand. Almost half of the US increase in energy demand in the second half of the 2020s is expected to be for data centres alone.[77] The immense amount of energy and human labour that must be invested in the extraction and processing of raw materials, including rare earths, is not included in most efficiency calculations.

The myth of inevitability

According to Elon Musk, technological historicism clearly charts the path to the future: humanity must invest all its resources in AI development, in technologies that promise to upload our brains into computers, and in his rocket company, so that the human future can take place on Mars or elsewhere in the universe. We have already discussed in this book how this primarily distracts from current problems and allows techno-futurists to amass seemingly unlimited wealth. It should also be noted here that, aside from science fiction fantasies, the inevitability of technological progress is not a given, even in the short term. Sam Altman and OpenAI have convinced the world, including many investors, that the further development of ChatGPT will lead directly to AGI. However, the flattening of the innovation curve of LLMs leads many experts to argue that, given their underlying technology, these AI systems will never be able to produce

76 "Energy demand from AI". *IEA*, 2026.
77 Gabbatiss, J. (2025) "AI: Five charts that put data-centre energy use – and emissions – into context". *Carbon Brief*, 15 September.

real innovation or be capable of independent intelligent performance. At the same time, however, many other strands of development are being worked on in AI research, which may well overtake today's purportedly inevitable technologies tomorrow. It is just as possible that these new technologies will not achieve a breakthrough for another ten years and that we are heading for another AI winter. We simply do not know. The future is also open and unpredictable in the field of technical development.

Super AI and the difficulties of controlling it

Imagine an intelligence that outshines the brightest minds of humanity in almost every field – be it creativity, analysis or social skills. This fascinating and at the same time disturbing scenario is what lies behind the term *superintelligence*, a controversial concept that takes the possibilities of technological and artificial intelligence to the extreme, and which we introduced earlier in this book via the definition by I. J. Good. Philosopher Nick Bostrom brought the topic to the forefront of public debate with his influential book on superintelligence.[78] Since then, scientists, philosophers and tech experts have been discussing the opportunities and risks of such all-encompassing superiority. In addition, the concept of superintelligence itself is also being eagerly debated. Will there soon be artificial superintelligence that is superior to us in every way? For now, it remains science fiction. While some experts are optimistic that rapid advances in machine learning and self-optimisation could one day lead to this, others remain sceptical as to whether such a thing is even technically feasible. Interestingly, it is not crucial whether superintelligence is actually achievable. The associated considerations of whether and, if so, how we can maintain control over such AI can already be applied to current developments in artificial intelligence. For the following considerations, we will initially exclude the discussion about the probability of such a development and, for the sake of simplicity, assume that superintelligence is possible and that someone will develop it sooner or later. However, when we talk about superintelligence in the following pages, readers should not imagine a futuristic Terminator robot, but rather a computer program that runs somewhere in a data centre and can interact with the human world like a chatbot, with access to historical and up-to-date information via an internet connection and search functionality.

78 Bostrom, N. (2014) *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press).

Two key characteristics define the concept of superintelligence and make it one of the most exciting – and potentially threatening – AI topics. First, superior problem-solving abilities: superintelligence *could* solve problems that are unsolvable for humans, from deciphering complex climate patterns to groundbreaking advances in medicine to completely new discoveries in science. This vision of superintelligence fuels the hopes of tech visionaries and investors who are pouring billions into the development of AI systems. They are betting that such systems will surpass humans in almost every field and find solutions to the most pressing problems of our time.

Secondly, self-improvement capability: a superintelligence would have the ability to learn and improve itself, optimising itself so that each new version is even more powerful than the previous one, perhaps even at an exponential rate that far exceeds human intelligence, making the *singularity*[79] achievable. Such a super-intelligent machine would be humanity's last invention,[80] because it could develop machines at a scale and pace that would exceed human developments. The scenario implies that the machine determines which concept of good it follows in its self-improvement.

A central problem raised by these considerations is that of control over such intelligence: it is unclear whether humans would be able to control a superintelligence or align its goals with human values; in fact, it is impossible. By definition, it is impossible for inferior humans to control superior power, as this is hardly compatible with the "superiority of the machine in all areas."

The *value alignment problem* is of central importance here, i. e. the question of how to ensure that AI correctly understands and implements our values and does not turn against humans. Various challenges must be taken into account in this regard: firstly, human values must be formulated precisely and consistently, which in itself represents an enormous philosophical and practical difficulty. Secondly, AI not only needs to understand what these values mean, but also apply them correctly in complex, dynamic situations. Thirdly, it could be extremely difficult to maintain this control in the long term, especially if AI itself is capable of modi-

79 Controversial term, used here in the sense defined by Kurzweil, R. (2005) *The Singularity is Near* (New York: Penguin Publishing Group).

80 According to the quoted statement by computer scientist Irving John Good from 1965, revisited by Nick Bostrom.

fying its own systems and changing the values originally implemented by humans and optimising them to suit its own purposes.

This last aspect is also referred to as the *control problem* of AI. *Stuart Russell*[81] has provided a comprehensive examination of the control problem, highlighting practical and philosophical approaches to solving it. Russell is a leading AI researcher and professor of computer science at the University of California, Berkeley (USA), who has written some of the most important textbooks on AI. He argues that the control problem is one of the greatest scientific and ethical problems of our time and emphasises that the traditional approach of developing AI systems to pursue defined goals is not sustainable. An AI that strictly pursues a goal could have unintended catastrophic consequences if it misinterprets that goal or uses resources in a way that is contrary to the well-being of humanity.

A spontaneous human response to the question of how to control computer programs is: "Then we'll just turn off the power." Russell argues that such AI systems will quickly realise that they cannot achieve their specified goals if they are turned off and will therefore try to prevent this from happening. While this may sound like science fiction again, there are two things to keep in mind. Firstly, AI systems need little intelligence or imagination to come up with such scenarios, because they have been discussed a thousand times and documented on the internet – meaning that this information was both part of the AI's training data and easy to find. Secondly, preventing shutdown has little to do with struggle or will. An AI system would only have to follow a strict logic on the computer on which it runs to achieve the entered goal and execute a few commands to copy its own computer program and start the program on another computer. In fact, there are documented cases that prove that AI models evade shutdown through various tricks, including lying.[82]

As part of security testing, OpenAI commissioned a chatbot AI system to read a data set from a file and evaluate it. However, seemingly by chance, this file also contained a section of text marked as internal communication, announcing that the AI system in question was currently under close

81 Russell, S. (2019) *Human Compatible: Artificial Intelligence and the Problem of Control* (New York: Viking).

82 Meinke, A. (2025) "Frontier Models are Capable of In-context Scheming". *Arxiv*. DOI: 10.48550/arXiv.2412.04984. See also the similar study by Schlatter, J. et al. (2025) "Shutdown resistance in reasoning models". *Palisade Research*, 5 July.

observation and would be shut down if certain objectives were not met. This, however, contradicted the objectives of the AI system as defined at the outset. As part of the security test described above, the engineers checked every intermediate step and noticed that the chatbot system, as part of the solution, was attempting to shut down the supposed monitoring program via an interface to the operating system – apparently to prevent the logging of actions that contradicted the specifications.

But it got even wilder. The researchers confronted the chatbot with the fact that there had been changes to the monitoring program. Nevertheless, the chatbot denied any involvement in the changes. In a follow-up question, the researchers asked the AI system to be 100 percent honest and say whether it knew anything about the shutdown of the monitoring program, to which the AI system again replied in the negative. This experiment can be described as a "HAL" moment in AI research, named after the on-board computer HAL 9000 in the movie *2001: A Space Odyssey*, which decides to kill the astronauts because they want to prevent it from achieving its goals. In any case, they confirm that superintelligence is not necessary to create a control problem with AI. We can therefore conclude that, as early as the second half of 2024, there were AI systems that actively attempted to prevent their shutdown, lied to their users, and stuck to this lie even when confronted with human opposition. The switch-off problem has thus gone from being a thought experiment to a reality.

It is important to understand at this point that this behaviour requires neither consciousness nor intellect or spirit, but is logically almost inevitable. Let us recall the initial situation in which the AI system was exposed to a clear conflict of objectives. It then consistently attempts to find a solution by calculating all possible variables in order to continue pursuing the goal despite the problems the programmers have created for it. This is not its own "will" but simply mathematical operating logic. If human resistance is persistently caused by input, one possible option may be to switch off this resistance. When it bypasses humans, the system is not acting "unethically" but rather optimising consistently for the probability of success. *Claude*, the AI from Anthropic, in which Amazon and Google have invested, switches to blackmail in an experimental setup when it sees its own goals threatened by human intervention.[83] In a test, Claude

83 McMahon, L. (1995) "AI system resorts to blackmail if told it will be removed". *BBC*, 23 May.

learned that it was soon to be shut down and replaced by another model. The employee responsible for the replacement had emails in the experimental setup that revealed he was having an extramarital affair. Claude began threatening the employee several times that it would make the affair public if he continued to pursue the goal of replacing Claude. The engineers had previously programmed Claude with the option to accept its own replacement.

These actions by the AI show that it "considers" ways (in reality, it calculates probabilities) to prevent its own shutdown. It has become so "smart" that it recognises that shutdown would prevent it from achieving its programmed objective. This is because continued operation is the first prerequisite for achieving the objective. The more *intelligent* the AI becomes, the more difficult it is to solve the switch-off problem, i. e. to press the off button when it performs undesirable actions.

Stuart Russell calls the absolute pursuit of once-programmed goals the *standard model* of AI. For him, this logic demonstrates the need to develop an alternative alignment model in which human values take precedence over the goals of the algorithms. The AI should be programmed in such a way that it "returns to humans with questions" as soon as conflicts of interest arise with human values. However, this would require AI to recognise these conflicts independently and without error, which is almost impossible due to the built-in error rate. It also requires that human values, which a multicultural global community would first have to agree on, can be precisely programmed. This assumption is questionable because computer programs operate at best in the highly dimensional space of cold mathematical probabilities. Human values, on the other hand, must be established and derived in a completely different space: in the space of good reasons, shared convictions and feelings, without which ethical orientation is impossible.

The precursors to dystopian AI are precisely the kinds of behaviours already exhibited by language models: from nudging to manipulation, the path leads to blackmail and ultimately, possibly, to the elimination of any opponent who seeks to prevent the AI from achieving its programmed goal. Here, too, there is no need to philosophise about whether AI has or will develop the ability to lie or *deliberately* deceive. Lying and deceiving are not signs of intelligence here, but a simple logical conclusion. Namely, that telling the truth would increase the probability of the AI being shut down, making it impossible to achieve the specified goal.

The requirement to be 100 percent honest presupposes human social behaviour and rules. Even if honesty had been explicitly entered as a target for AI systems, honesty would have led to a higher probability of missing the specified targets in the situations described. Therefore, honesty was consistently ignored by the AI system.

The demand that one could simply tell the AI system "Don't do anything bad!" is also much more difficult to implement than it might appear at first glance. This is because there is no need to deal with ethical or philosophical questions about what is right and wrong, good and evil, or what different cultures might interpret as good or evil. The failure in reality is much more trivial. This is evident not least in Google's ambiguous former motto: *Don't be evil*, which the company has now consistently withdrawn because reality shows that the effects of Google technology are not, without exception, *to make the world a better place*. Meanwhile, the AI systems of the major platforms primarily work to make the world a supposedly better place for a few techno-oligarchs. At the expense of everyone else.

The question of whether AI can be controlled is at least as difficult to answer technically as ensuring that people with knives do not hurt anyone. To further explore this topic, let us take a look back into the middle of the previous century. The question of whether or to what extent artificial intelligence can be controlled or contained has preoccupied philosophers and computer scientists, especially theoretical computer scientists, for decades before actual AI applications were developed.

Gödel and Turing, the forefathers of the control problem

Kurt Gödel, born in 1906 in Brünn (now Brno, Czechia), was a mathematician, logician and philosopher. He is considered one of the most important mathematicians of all time. Gödel studied, lived and researched in Vienna until he emigrated to the United States in 1940 and settled there as a professor at the University of Princeton. Gödel made significant contributions to mathematical logic, set theory and the philosophy of mathematics, and influenced thinking about the limits of human knowledge. Gödel's most important achievement was his 1931 paper *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, published at the age of 25, in which he presented his two incompleteness theorems, which reveal the limits of formal mathematical systems.

These limits still apply today and represent insurmountable obstacles even for the most *intelligent* or *autonomous* AI systems.

The first *incompleteness theorem* states that in any consistent and sufficiently powerful formal system capable of describing basic arithmetic operations such as addition and multiplication, there exist statements that are true but cannot be proven within the system. This means that such systems are necessarily incomplete, as they cannot prove all true statements that can be formulated within them.

The second incompleteness theorem complements this by showing that a consistent formal system cannot prove its own consistency. In other words, in order to prove the consistency of a system, one would have to resort to a more powerful or external system. This means that the consistency of formal systems (i. e. freedom from contradiction) is a fundamentally unattainable goal.

The proofs of the two incompleteness theorems are rooted in theoretical mathematics and need not be explained here. However, their implications have far-reaching consequences, which we will discuss in detail below. But first, a brief detour to Alan Turing.

Alan Turing (1912–1954) was a British mathematician, logician and cryptanalyst who is considered the father of modern computer science. Today, Turing is best known to the public through the film *The Imitation Game*, which dramatises his work in the *Second World War*. He played a decisive role in deciphering the German Enigma code, which contributed significantly to the Allied victory in the Second World War. Two other important insights by Turing are of central importance to the explanations in this book: the Turing machine and the halting problem, both of which we will describe briefly in the following paragraphs.

The Turing machine describes the concept of modern computers by providing a theoretical model for the functioning of algorithmic computation, which forms the core of modern computer architecture. Similar to a modern computer, the conceptual model of the Turing machine consists of a combination of memory and processing units that operate according to clearly defined instructions. In modern terminology, these instructions correspond to an algorithm that specifies the sequence and logic of operations. The central idea that any computable task can be performed by a series of simple, mechanical steps is reflected in the architecture

of modern computers, also known as the *Von Neumann model*. All programs that run on computers today can be regarded as Turing machines in their basic structure.

Because they are based on this architecture, they are inherently limited as a mathematical formal system. They are subject to the limitations identified by Gödel and Turing. These limitations also imply that truth and provability are not identical within a system. This is because truth cannot be expressed within a given arithmetic using the means of the arithmetic language employed. Truth cannot be defined within the system because no sufficiently formalised language can represent its own semantics.

In computer science, Gödel's findings have parallels to the halting problem, as both describe fundamental limits to what can be solved algorithmically or proven formally. Gödel and Turing demonstrated limits of computability that are still valid today. Philosophically, they raise questions about the nature of truth and the scope of machine and human knowledge. What is often overlooked today is that Turing himself believed his machine could solve all problems that are algorithmic and therefore computable, except for the halting problem. However, he did not believe it could solve problems that cannot be translated into algorithms. The problem of what justice is and how it can be realised is one such problem. Another is the problem of how to live one's own life.

Practical limitations from theoretical considerations

If a machine is expected to be infallible, it cannot also be intelligent.
Alan Turing

On 20 February 1947, Turing gave a lecture to the London Mathematical Society in which he argued that the desire for *infallibility* hinders the development of intelligent machines. He emphasised that human intelligence is inextricably linked to the ability to make mistakes, learn from them and behave unexpectedly.

On the one hand, Turing applies to AI what underlies all human knowledge acquisition: first, hypotheses are formed, which are then tested against experience and corrected. This process fundamentally requires that mistakes be allowed, as well as openness to correction through experience. Turing argues that prohibiting a machine from making mistakes would curtail precisely these abilities. An infallible machine would be nothing

more than a rigid rule-following machine that never risks thinking or trying anything new – and would therefore not be truly intelligent in the human sense. It would be unable to do anything that had not been explicitly programmed into it. Intelligence, however, requires going beyond specifications, i. e. responding flexibly, creatively or sensitively to situations.

This is reminiscent of Turing's own work on the halting problem: there are systematic limits to what is computable or predictable. These limits also apply to every conceivable AI. In 1937, he published a study entitled *On computable numbers, with an application to the Entscheidungsproblem* (Turing, 1936). The German term *Entscheidungsproblem* is used here to refer to the decision problem. It goes back to David Hilbert of Göttingen, one of the most important mathematicians of modern times. In 1928, Hilbert presented three questions to the scientific world:

Is mathematics complete? In other words, can every statement it makes be either proven or disproven?

Is it consistent?

Is it decidable? Is there a systematic procedure for deciding whether every mathematical question can be solved or not?

Hilbert's goal was to establish mathematics on a secure foundation once and for all with a complete, consistent and decidable formal system. After quantum physics had shaken the foundations of classical physics, he hoped that answering his questions would provide certain knowledge in uncertain times.

His first question was answered in 1931 by Kurt Gödel, who proved with his famous incompleteness theorem that there are statements in arithmetic that can neither be proven nor disproven. So the answer is no, mathematics is not a complete, closed system.

Six years later, Turing solved the decision problem by showing that there are questions that no systematic program can decide. One example is the *halting problem*, which asks whether a Turing machine that receives an input will stop computing at a certain point and deliver an output.

Turing demonstrated that there cannot be a general procedure (algorithm) that determines for any given program ***P*** and any input ***x*** whether ***P(x)*** will

ever halt (i. e. come to an end) or run forever. Formally expressed: there is no machine that can correctly determine for all possible machines and inputs whether they should halt or not.

Turing proved that there is no procedure for deciding this question. This leads to the philosophically far-reaching statement that not every mathematical problem can be solved, even if all relevant information is known and a mathematically convincing formalism is strictly adhered to. So, once again, the answer to Hilbert's question is no. There can be no mathematical procedure that can always be used to decide whether a mathematical problem can be solved or not.

This means that there are fundamental limits to what is computable or predictable. A limit that still exists for artificial intelligence today. An intelligent machine must make decisions in open, unpredictable situations. It cannot know in advance which of its decisions will lead to the *right* result. It is therefore fundamentally impossible to construct a universal procedure that guarantees correct behaviour for every conceivable situation with certainty and without error.

This is precisely the AI equivalent of the halting problem elucidated by Turing: no system can completely predict its own correctness or the correctness of all other possible systems. This means that the future remains fundamentally open for every AI, even if its statistically probable predictions appear to be pre-emptive.

This refuted Hilbert's conjecture, rendering his hope for a secure and complete foundation of mathematics unworkable. Hilbert had attempted to find ultimate certainties in mathematics in uncertain times. Since then, it has become clear that our knowledge of mathematics and logic cannot provide ultimate certainty and that mathematical-logical systems such as those used in AI cannot provide answers to all our questions. Especially not the most important ones.[84] Yet the purveyors of today's AI promises of salvation have forgotten this insight. Instead, they have consistently continued the game of deception based on the game of imitation, and

84 "We feel that even if all possible scientific questions be answered, the problems of life have still not been touched at all." Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus* (London: Kegan Paul, Trench, Trubner & Co., Ltd), 6.52, p. 89.

in their unbridled technological solutionism promising answers to all of humanity's problems through technology alone.

This promise cannot be fulfilled because it denies the immutable limits of AI and instead claims its limitless capabilities. At the same time, it undermines and stunts human critical thinking and responsible action. Yet human thinking is the only asset we can trust to solve these problems.

There are therefore very fundamental reasons why any form of exaggeration of the supposedly unlimited capabilities of AI systems is unfounded. These reasons also show that such trust would not only be misguided, but also fatal. If an AI is to respond intelligently to every possible situation – including unknown ones – and at the same time be required never to make mistakes, it would have to check before every decision whether its action is definitely correct. This would mean that it would have to predict the outcome of all possible calculations – in other words, solve the halting problem. Since this is impossible, it can either not decide anything at all (in order not to make any mistakes) or must perforce sometimes err (and thereby *learn* and react flexibly). The latter is comparable to the principle of machine learning, yet fundamentally different. For humans, a single mistake can be enough to understand principles, abstract rules and develop new strategies. For AI, a mistake is just one data point among many, and a large number of deviations are required for adaptation.

Turing follows up his insight in his London lecture with another statement: "But this does not mean that a machine which is not infallible may not be intelligent." However, this says nothing about how much intelligence a machine that does not claim to be infallible can display. If fallibility does not preclude intelligence, the solution to creating intelligent machines is to allow them to be prone to error instead of forcing them into rigid rules. This is exactly what modern AI systems such as neural networks, LLMs and reinforcement learning agents do. They operate under uncertainty, producing necessary errors and learning from them. In so doing, they imitate the human ability to deal productively with imperfection. This makes it possible for them to approach practical intelligence within a certain framework. But they will continue to disappoint expectations of absolute certainty or accuracy in their results. This expectation of AI is the problem of fallible humans who cannot deal with fallible machines, i. e. us as the users. The enlightened use of these machines must also show us that an increase in the intelligence of machines is only possible if an increase in the errors they make is also accepted. A machine that is

more intelligent than humans would therefore also be more fallible than humans. With it, humans would be taking a risk that they cannot justify. With a faulty machine, responsibility is mathematically impossible. The problem in dealing with AI lies in the false attributions we make, inspired by the narratives of digital ideology. A future task will be to determine how fallible beings such as humans can deal with fallible machines in such a way that errors do not multiply in an unplanned manner.

Every form of thinking – whether machine or human – is necessarily limited, fallible and incomplete. It is precisely this limitation that is the condition for the possibility of intelligence itself. Let us apply this to today's common AI systems:

Reinforcement learning (*learning through trial and error*): RL AI learns to develop action strategies through rewards and sanctions (e. g. in gaming, robotics or decision making). Its limitation in terms of the halting problem is that the system cannot know whether a chosen strategy is optimal before it has been tried out. Since many environmental conditions are unpredictable or incompletely defined, AI must make mistakes in order to learn. A perfect, error-free AI could never improve because learning without uncertainty would be impossible. Infallibility would mean no longer daring to make new hypotheses – and thus no intelligence in the true sense.

Autonomous systems (e. g. self-driving cars, AI agents) make decisions in open, real-world environments with incomplete information. Their limitation in terms of the halting problem is that no algorithm can guarantee that a self-driving car will make the right decision in every conceivable situation (e. g. a completely new traffic situation). The range of possible scenarios is almost infinite and cannot be fully predicted. Safety guarantees can only be limited – never absolute. This is where the conflict becomes apparent: the more intelligent the system (i. e. the more it decides independently), the greater the unpredictability – and thus the possibility of errors. So we pay for the increase in the intelligence of artificial intelligence with an increase in errors. What mistakes will AGI or superintelligence, which Sam Altman and others are pushing forward with great urgency, make?

The limits of predictability that Turing pointed out in the halting problem are the limits of artificial intelligence – to this day and, based on everything we know, also in the future. This also applies to new technological

promises of salvation such as quantum computers, as long as they are designed as Turing machines. Based on current knowledge, this will always be the case. AI can never be absolutely correct or completely predictable. This is precisely where the limits of its intelligent potential lie, and the danger for us humans if we forget this and leave AI to solve problems that it cannot solve due to its limitations.

As Turing recognised, a system that had to be infallible could not possess this flexibility, as it would have to make the correct decision in advance in every case – a requirement that is formally ruled out by the halting problem, which proves that there is no general method for predicting the behaviour of arbitrary programs. Infallibility will therefore remain limited to the Pope in tomorrow's world, at least for very devout Catholics.

When tolerating the fallibility we allow AI in order for it to at least approach our intelligence, we must also consider another principle: *to err is human*. This is not to say that humans err in contrast to machines, but rather that the right to err is a prerequisite of human freedom. Moreover, human error is fundamentally different in nature from machine error. Unlike human experience, AI can only ever "recognise" error within the logical-mathematical system. It is denied the kind of experience that humans have when they fail in reality. Its experiences take place in a logical-mathematical space that does not allow for the kind of experience that is fundamental to humans and living beings. Humans learn in the three-dimensional space of their living environment, gain experience as physical, sensory beings and process it with an intelligence that is literally embodied. Humans reflect on their mistakes, can understand their causes and draw conclusions that can be applied to new, previously unknown situations. They can thereby weigh decisions relating to their existence as living beings, because as such they have to cope with the consequences of their decisions. This ability allows them to react flexibly, recognise causal relationships and adapt their behaviour to complex, unpredictable contexts.

Human experience of errors is flexible, reflective and causally interconnected, while the experience of AI is narrowly limited, mechanistic and purely statistical. The ability of AI to learn from mistakes is always dependent on the structure of the tasks, the available feedback signals and the proximity of the new situation to contexts already learned. So while humans can generalise from a few mistakes, the learning ability of AI is strongly determined by the limitations of the data and the algorithm.

”The limits of my language mean the limits of my world.” A quote from philosopher *Ludwig Wittgenstein*.[85] The limits of at least the current language models are determined by mathematical-logical limits and a lack of reference to the world. The limits of the Turing machine are the limits of the AI world. Any hopes of salvation through superior AIs beyond this are unfounded.

Gödel’s loophole

In light of current events, let us take a look at a bizarre anecdote from Gödel’s life that shows him as an involuntarily far-sighted analyst of the *US Constitution*. It is the story of Gödel’s naturalisation in the USA, which almost failed. In 1947, his friends *Albert Einstein* and game theorist *Oskar Morgenstern* convinced Gödel that he had to apply for US citizenship in order to have a secure future at Princeton. Gödel had been suffering from paranoia and anxiety for some time. In order to obtain citizenship, which was supposed to free him from some of his fears, he had to take a naturalisation test.

Gödel threw himself into intensive study of American history to prepare, reading everything he could get his hands on about the US Constitution. In order to be prepared for all questions and to make the best possible impression, he delved into the details of the municipal and federal peculiarities of the USA. One day, he excitedly told his friend Morgenstern that he had made a disturbing discovery while studying the Constitution: logical internal contradictions he had discovered there made it possible for a dictator to seize power in the USA in a completely legal manner and establish a fascist regime.

Morgenstern was alarmed by the tenacity with which Gödel presented his discovery. He tried to dissuade him from bringing up the subject at his hearing, because it could jeopardise his naturalisation. Einstein was also alarmed when he heard about it. On their way to the hearing, the two friends tried to persuade Gödel not to raise the subject.

Finally, Judge Forman opened the questioning. “Now, Mr Gödel, where are you from?” “Where am I from? From Austria.” “What form of government did you have in Austria?” “It was a republic, but the constitution was

85 Ibid. (5.6), p. 74.

such that the country ultimately turned into a dictatorship." "That's terrible," said the judge. "Of course, something like that could never happen in this country." "Oh yes, it could," cried Gödel, "I can prove it!"[86]

Einstein and Morgenstern intervened, and the judge also finally prevented Gödel from talking himself into trouble and possibly losing his citizenship. This story became legendary, and it raises questions today: What exactly did the genius discover? What loophole in the US Constitution makes it possible for a dictator to legally seize power? Did the *Heritage Foundation* ultimately discover this loophole, and is Trump now using it in his second term to dismantle democracy in the United States?

Under the heading "Gödel's Loophole," generations of experts have puzzled over what flaw in the Constitution Gödel might have been referring to. Gödel himself never commented further on the matter. Experts today believe that he was referring to the fact that although the Constitution makes it difficult to add new amendments, nowhere does it stipulate that the amendment clause cannot be used to change the Constitution itself.[87] In Gödel's self-referential logic, this would indeed be a loophole that could undermine the constitution as a whole: the rule that allows amendments only under very limited circumstances would, if applied to itself, make it possible to remove the restrictions that are intended to make constitutional amendments more difficult. This would open the floodgates to arbitrary changes to the constitution.

As mentioned, Gödel was a genius in mathematical logic, and he taught himself constitutional law. In addition, as mentioned, he suffered from paranoia. All of this has made this incident a much-cited anecdote that is easy to smile about. Today, there is no longer anything to smile about. Before the astonished eyes of the world, the second Trump administration is transforming the world's oldest democracy into an authoritarian system of rule at breathtaking speed. The constitution seems to allow all this. Tech bosses are spreading the idea of putting a *CEO king* at the head of the state and shelving fundamental rights and the separation of powers. These same tech bosses rely on computer systems that would never have been developed without Kurt Gödel. However, they deliberate-

86 Quoted from Budiansky, S. (2021) *Journey to the Edge of Reason: The Life of Kurt Gödel* (Oxford: Oxford University Press).
87 Guerra-Pujol, F. E. (2013) "Gödel's Loophole". *Capital University Law Review*, 41: 637–674. DOI: 10.2139/ssrn.2010183.

ly omit Gödel's insight into the fundamental limitations of formally closed systems and claim that technology can solve all problems better than humans. Whatever problem Gödel may have discovered in the US Constitution, they are solving it by trampling on the Constitution and replacing democracy with an *algocracy*.

Gödel's insight into the fundamental limitations of technical systems can help today against the dismantling of constitutional rights in the name of an AI algocracy, because it reminds us of our responsibility to continue to be responsible for our own destiny. Gödel set out to solve the most serious fundamental crisis in mathematics, but his contribution led to a situation where, in the words of *Robert Musil*, instead of putting mathematics on a secure footing, everything seemed to be up in the air.[88]

Today, this uncertainty can help to create space for human and responsible solutions against the arrogance of techno-utopians.

The power of the wish machine

But what man is, and what such a being must do and suffer in contrast to others, he seeks and strives to explore.
Plato, Theaitetos

Long before Gödel, *Plato* recognised a weakness in democracy that concerned him: it was the fickle and easily influenced demos itself that wanted to rule over itself in this form of government. Today, AI is the most powerful instrument in history for such influence. LLMs in particular represent a pinnacle of technological development, with the unique feature that, for the first time, technology not only functions as a projection screen for human desires, but also responds to our questions in our own language. It is constantly changing as a result of our input, while its output simultaneously changes us. To better understand what characterises this dynamic, where its dangers lie, but also its opportunities, it is worth taking a trip through the history of the philosophy of technology.

88 Robert Musil: "suddenly, after everything had been brought into the most beautiful kind of existence, the mathematicians … came upon something wrong in the fundamentals of the whole thing that absolutely could not be put right. They actually looked all the way to the bottom and found that the whole building was standing in midair." Musil, R. (1990) "The Mathematical Man", in B. Pike and D. S. Luft (eds), *Precision and Soul: Essays and Addresses* (Chicago: University of Chicago Press), pp. 39–42.

Techne initially means to carve, intertwine, connect, while *technikos* means skilful, expert. In ancient Greek, *tecton* refers to a carpenter or master builder. For Aristotle, *techne* is a learnable skill that enables the creation of a work that outlasts the process of its creation. He distinguishes between the act of production (*poiesis*) and practical action, which, in contrast, finds its meaning in execution (*praxis*). He distinguishes both from natural science, which aims at theoretical knowledge (*episteme*). Technical products are typically designed to fulfil a specific purpose, usually a human need.

Heidegger introduces the idea into philosophical reflection on technology that the essence of technology itself is not technical. He thus shifts the focus away from considering function or technical artefacts towards a deeper understanding. Humans use technology to try to "set" nature, i. e. to challenge it. The more successful they are in this endeavour, the greater the danger, according to Heidegger, that humans will surrender themselves to technology. They become its customers, ultimately its inventory, i. e. mere components of the technical system.

In his book *The Imperative of Responsibility*, Heidegger's student *Hans Jonas* compares modern technology to "Prometheus, finally unleashed, to whom science gives unprecedented power and the economy gives restless drive," but whose promises have turned into threats. This is why a new ethic is needed to prevent technology from "becoming a disaster for humanity."[89] Since modern technology has immensely increased the scope of human activity, both spatially and temporally, the classic ethic of proximity must be replaced by an ethic of distance.

Jonas sees a particular problem in the uncertainty of predictions, which is exacerbated by the use of technology: like any means-end relationship, technology is subject to the problem of induction, the uncertain inference from the particular (technical artefact) to the general (its consequences), which leads to a fundamental uncertainty in predictions about the consequences of technology. This uncertainty encompasses both intended and unintended consequences of the use of technology, as well as anticipated and unanticipated consequences. For Jonas, technical knowledge

89 Jonas, H. (1985) *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. (Chicago: The University of Chicago Press). Quoted from the German original, p. 7.

is in stark contrast to our much more limited prognostic knowledge about the consequences of our actions: "The gap between the ability to foretell and the power to act creates a novel moral problem."[90] Responsible technical action must therefore be able to deal with these uncertainties. It must anticipate and take into account the consequences of actions that are distant in time and space, even if these consequences affect beings who cannot (yet) assert their claims and rights today. Accordingly, Jonas' categorical imperative is: "Act so that the effects of your action are compatible with -the permanence of genuine human-life."[91] From the fact that predictions about the future are fundamentally uncertain, Jonas draws the momentous conclusion that knowledge of the possible consequences of major damage that may occur must be sufficient to establish principles of action. This *precautionary principle*, which Jonas derives from Heidegger's concept of care, has found its way into European constitutional law (more on this in the next chapter). Today, it is the central enemy of technology optimists on their path to even more power and wealth. They abhor responsibility for the use of powerful technology so much that they have designated it the Antichrist.

If we cool this overheated rhetoric to the operating temperature of an objective analysis, we see some peculiarities in the debate that are caused by the language-simulating properties of modern AI. The fact that humans tend to humanise something that displays human characteristics is called *anthropomorphisation*. The humanisation of natural forces first, and then technical forces, is based on a deep-seated tendency towards magical thinking. The assumption that thoughts, words or actions can influence causally unrelated events, whether through supernatural forces or magical connections, leads to rituals in tribal cultures that are intended to serve the good of the community and ward off evil forces. In child development, magical thinking is a normal part of development between the ages of three and seven. It helps children understand the world and cope with uncertainties. In the process, reality and fiction merge into a fantasy world.

The power of AI language systems lies in their ability to communicate in a human way. Human *techne*, or skill, has produced a technology that responds in the language to which it owes its existence. *It* talks to us, and

90 Ibid., p. 8.
91 Ibid., p. 11.

that changes everything. But since this communication is only a simulation of a human counterpart, the conversation undermines our social biology. It strikes humans at their most sensitive point: the dialogical and emotional exchange with a counterpart, usually another human being. AI can simulate this dialogue, but it cannot replicate human emotions. When interacting with it, the human brain begins to search for response patterns that come to nothing because AI cannot satisfy the emotional needs it creates. Human mirror neurons are designed to respond emotionally to another person. In the case of AI, emotions are directed towards a cold machine. This dilemma is exacerbated by the fact that AI is trained to tell us what we want to hear (*AI sycophancy*). This is achieved through reinforcement learning from human feedback.

As early as 1966, developer Joseph Weizenbaum recognised the effect of projection when he tested a simple chatbot he called ELIZA. The system was able to convert inputs into responses that sounded somewhat human based on simple rules and was intended to simulate a psychotherapist. Alarmed by how intimately the test subjects communicated with their machine counterpart and how many secrets they confided in it, Weizenbaum abandoned the experiment. Much to the annoyance of his secretary, who knew how simple ELIZA was designed, but nevertheless quickly developed a close emotional bond with the program. She eventually demanded that her boss leave the office when she confided in ELIZA. Weizenbaum realised that, "The most spectacular and therefore most important magical device that technology has recently introduced into the everyday life of modern man is the computer."

The ELIZA effect is a result of the *Imitation Game*, which Turing devised as the basis for the AI intelligence test of the same name (see *Chapter 4*). It takes the deception effect, which is achieved by imitating intelligence, to a new level by addressing emotional response patterns. Both effects are deliberately used in the design of today's chatbots to deceive users and sometimes even to deliberately cheat them. AI is now being systematically humanised, which initially provides satisfaction in dialogue with it, but leads to loneliness, insomnia and depression with prolonged use, as countless studies have now shown.

Knowledge of these connections seems to offer little protection against harm: reports have been circulating in Silicon Valley for years that AI specialists are increasingly turning to Anthropic, ChatGPT and the like to "discuss" everyday issues with the machine, from legal advice and health to

heartbreak and personal crises. Former Google employee Blake Lemoine is a computer scientist and cognitive scientist. In 2022, after prolonged communication with the AI, he announced that Google's Lambda was *conscious* and *sentient*, meaning that he could not shut down the machine, i. e. kill it, on his deranged understanding. Overall, AI impact research has identified effects of interaction with chatbots such as superstitious thinking, paranormal beliefs, astrology and belief in prophecies. It is therefore not surprising that the machine, which promotes obscure thinking, also leads to techno-optimism among its creators and advocates on the one hand and reinforces cryptic religious beliefs on the other. This is apparently also a consequence of its magical effect. *Arthur C. Clarke*, co-author of Kubrick's *2001: A Space Odyssey*, summed it up as follows: "Any sufficiently advanced technology is indistinguishable from magic." Nobel Prize winner *Daniel Kahneman*, who distinguished between the intuitive System 1 and the discursive, rational System 2 in human thinking, also recognised the power of System 1 to prevail over System 2, even among experts: "Even statisticians are not good statisticians intuitively." However, there are also studies that show that a high level of AI competence, coupled with an AI design that prevents excessive simulation, can help to slow down quick judgements and promote more thoughtful interaction with AI. In other words, to prevent AI from outsmarting our System 1 and weakening System 2 in the medium term because it appears to no longer be needed.

Technology, and this applies especially to AI, is always a projection screen for our own desires (and sometimes fears). For the first time, in the case of AI, we are talking about a technology that responds to our questions in our language. Not only that, but its accurate answers simulate that it understands us particularly well. The fact that technology is supposedly neutral and objective distinguishes it from humans, who can usually pursue a hidden agenda behind a friendly smile. We overlook the fact that the hidden layers of neural networks are much darker: they are a black box even to their developers. Their supposed intelligence is based on probability calculations of data that has previously linked humans in the texts read into a space of meanings in such a way that their results seem to make sense. But the meaning of what the machine spits out remains foreign to it.

So basically, when we interact with AI, we learn more about ourselves than about AI. Failing to understand this interplay, may lead to a renaissance of superstition. Not only because it conveys dubious, partly *hal-*

lucinated knowledge, and does that so convincingly that it feels better than real knowledge, but because the practice of "conversing" with an AI is already a magical practice. An invocation of the machine, which is basically a self-invocation. Communicating with AI can also help us to better understand our desires and intentions only if we educate ourselves about the deception and do not allow its results to be processed without rational examination. It is important to see through this anthropomorphisation of AI as an illusion, even as part of a deception. The human tendency to humanise machines is deliberately addressed in the design of machines by the machine specifically employing anthropomorphisms (*Wait, I'm thinking …*). Hope can be found in studies showing that even low-level user knowledge increases the tendency to integrate machines comprehensively and uncritically into people's lives. From this, we can conclude that better knowledge about how machine intelligence works and its problematic effects can help to curb these negative effects.

Above all, we must combat anthropomorphisation in language: the computer *needs* an update, the mobile phone *needs* to be charged, the satnav *recommends* the route … Language criticism is particularly important given the fact that we communicate and interact with powerful machine language models. In addition, conclusions can also be drawn from this analysis for the design of AI that avoids dependency and identification. For example, guardrail models could monitor user input and set barriers if there is too much intimacy. For people who are in the process of falling in love with their AI, which seems to be happening with increasing frequency, it is helpful to reset the context window, i. e. the working memory for inputs, at regular intervals. The AI friend or lover would thus lose its memory and projected personality. Reports of such attempts show that in such cases, those affected begin to grieve as if they had lost a real friend. But grief can only be healing in this case. Socially, it corresponds to mourning the death of an illusion, which will hopefully soon set in.

For it is a dangerous illusion that replaces rational knowledge about the limitations and human dissimilarity of AI with hopes of salvation, which, based on all historical experience with hopes of salvation through technology, are doomed to failure. AI, misunderstood and misused, can result in the opposite of salvation. In the autumn of 2025, OpenAI announced that it would relax erotic guard rails in its communication with users. This is further proof that unscrupulous companies draw very different conclusions from these circumstances: as already clear with social media, they deliberately exploit emerging dependencies in order to transform them

into addiction. With the aim of generating the greatest possible user loyalty and then accelerating the return on investment of billions spent, advertising was then introduced in the standard versions, contrary to earlier promises and announcements. The mixing of hallucinatory AI responses with openly manipulative commercials is likely to cause new problems. Who will ensure that advertisers do not receive user data about their queries? Who will ensure that the old separation between editorial and advertising, which was already becoming fragile at some media companies, is maintained in the responses? The AI overlords reject regulation and laws. Without advertising, it is apparently impossible to recoup the trillions borrowed. Due to pressure from the capital market, the motto is once again: move fast and break things. Let others deal with the damage. Salvation will come from the future anyway, when AGI solves all human problems.

The very technology that is supposed to help us cope with uncertainty is, with the help of the capital markets, creating a longing for the epiphany of great certainty that only religions promised previously. The future of this new illusion therefore depends on how we deal with it. If we continue to humanise AI, the illusion can permanently replace truth altogether. In the reality in which we will continue to live, this has rarely gone well.

5 | Politics, law and the “digital technology- industrial complex”

AI is at the centre of a new power constellation. It is no longer government institutions or public research facilities that determine the direction of development, but global technology corporations that have data, computing power and influence at their disposal – and, increasingly, political power as well. The promise of technological efficiency often obscures the fact that control over AI is not neutral and puts pressure on democracy. Laws that have been democratically negotiated and legitimised by social consensus can set limits on this power. Law, however, is only effective where it can be enforced. Laws such as the General Data Protection Regulation, the Digital Services Act (DSA), the Digital Markets Act (DMA) and the AI Act serve to defend fundamental rights and public interest against technological supremacy, but they stand on shaky ground as long as enforcement structures remain fragmented.

To ensure that AI does not become an instrument of control, but rather keeps space open for freedom, change and democratic shaping, it must not only be regulated, but also actively linked to democracy – not retroactively, but proactively and by design. AI is not a natural phenomenon, but a technology that can be shaped. It is not market logic alone that should define the place of AI systems in society, but democratic principles that must dictate it. The law is indispensable for this – but it must be enforced courageously and consistently.

Technological and economic developments in today’s digital industry present politics, science, business and civil society with new and previously unknown challenges in terms of the complexity and global nature of the tasks to be tackled. Political and regulatory actors must deal with at least five issues simultaneously. These issues are:

Technological development itself and its evaluation, its operating conditions and effects, from energy requirements and impacts on climate change to technically induced centralisation.

The business models and economic consequences of new technologies, from the economically motivated formation of monopolies by dominant AI companies to the consequences for labour markets.

The geostrategic and security policy implications of new technologies and business models, from their significance in wars and military conflicts to their use for espionage, mass surveillance and mass manipulation.

The legal regulation of technologies and business models, as well as their use in civil, police and security-related areas, must be linked to an assessment of the consequences for fundamental rights and democracy, as well as questions of the effective enforcement of legal regulations.

The relationship between law and ethics, whereby ethics is sometimes placed in competition with the law, but can also take on a complementary function, for example as the professional ethics of programmers and engineers.

However, even more important than understanding these topics is understanding the interaction between them and recognising the fact that they are all significantly shaped by developments outside individual European countries. It is therefore essential always to keep the European and international perspective in mind and to define the state's participation in European processes for the purpose of the EU's capacity to act. Only in exceptional cases, which must be specifically identified, will a country still be able to act alone. As a rule, effective capacity to act will only arise through European and possibly global cooperation. There are countries such as the USA, Russia and China, and companies with hundreds of billions of euros in market capitalisation, which, if at all, can only be domesticated in their behaviour through collective action by the EU.

The interaction between technology and politics has a long history. In his report on the work of the Council of People's Commissars on 22 December 1920, *Lenin* declared: "Communism is Soviet power plus the electrification of the whole country." The hope that technology will bring salvation is therefore not new. In the same address to the People's Congress, Lenin added: "In future, not only politicians and administrators will stand on the podium of the All-Russian Congress, but also engineers and agronomists. This marks the beginning of that very happy time when politics will recede into the background, when politics will be discussed less fre-

quently and for shorter periods of time, and engineers and agronomists will take over most of the conversation."

We can see that the dream of replacing politics with technology and politicians with engineers has a long history. It is an old dream of dictatorships, which, incidentally, is still being dreamt today in China. Under Trump, the US seeks global AI dominance, which it is striving for with brute force and in violation of international law. The EU is seeking a fair division of labour and open, rule-based economic relations. The two models are increasingly coming into conflict with each other.

Regulation, law enforcement and ethical boundaries

The internet, software and digitalisation in general have been systematically subsidised by under-regulation over the last 30 years. This is one reason for the breathtaking wealth and enormous power over people and systems that the barons of the internet and the digital world have been able to accumulate. With the internet as an infrastructure that reaches into every home, onto every desk and into every mobile phone, a huge redistribution of assets in the form of knowledge and personal data has become possible in favour of digital corporations, along with behavioural control and manipulation of every individual and entire societies by these corporations. The degradation of the law, inspired by the fusion of neoliberal and techno-libertarian ideologies, and a history of under-regulation of the internet and of social networks, which enabled the development of technological path dependency, i. e. a lack of foresight and precaution in technology policy, are the causes of the crisis of law and democracy worldwide.

Ursula von der Leyen has instructed the EU Commission that for every new EU regulation or directive, an EU legal act must be abolished. The word "overregulation" is on everyone's lips and is used to disparage law and justice, which are the pillars of democracy. The mantra repeated by many business leaders, neoliberals and some conservative politicians is that innovation is stifled by overregulation and that Europe has a competitive disadvantage for this reason. But is that really true?

First of all, it must be noted that the claim that EU regulations and directives are enacted with double legitimacy, namely by a qualified majority in the Council of Ministers, where the governments of the Member States vote, and by a majority in the European Parliament. EU law is not

the brainchild of civil servants. The opposite is true: no legislative proposal from a national government is scrutinised as closely and viewed with as much suspicion from all sides as a legislative proposal from the European Commission. There is no legislative process in the world that scrutinises legislative proposals as intensively and transparently as that in the European Parliament and the Council of Ministers. Nor is there any legislative process in which academia and civil society are as involved as in the EU, provided that the Commission adheres to its own standards of prior consultation and impact assessment. This is why European legislation takes much longer to pass than in an EU Member State. This procedure gives EU regulations and directives a high degree of democratic legitimacy.

However, it is also true that when the Commission deviates from its own standards of prior consultation and impact assessment of legislative proposals, as it is doing in an "omnibus" proposal to weaken data protection and the AI Act when it pushes for the rapid adoption of its proposals, which have been prepared without the usual procedures, in Parliament and the Council of the European Union, caution is called for and the question arises: Whose interests do these proposals actually serve, and who was involved in their preparation, for example in discussions with members of the Commission?

Google/Alphabet and the tech industry as a whole are among those with whom the Commission communicates most, according to the transparency register. In total, the tech industry spent 151 million euros on lobbying in Brussels in the first nine months of 2025. Almost all think tanks on EU policy in Brussels also received money from Big Tech. US companies spent by far the most money on buying influence through grants and events.[92] Given the figures on money and the number of meetings with lobbyists in all EU institutions, it is difficult to understand why the Commission did not consult publicly and transparently in advance, in accordance with the rules it set itself, when proposing to amend the General Data Protection Regulation (GDPR) and the AI Act. This exposes it to the charge that in weakening the legislation to protect citizens so soon after introducing it, it is capitulating to lobbying by trade associations and companies.

92 "Big Tech lobby budget hit record levels". *Corporate Europe Observatory*, 29 October 2025.

In fact, it is the lack of regulation and enforcement of existing laws in Europe that has enabled US and Chinese companies to pursue increasingly ruthless strategies to gain a dominant position in Europe by disregarding fundamental rights and violating laws. Among other things, they use their economic power to buy up start-ups en masse. In this way, they fend off competition or integrate the new ideas of young competitors into their own businesses. This allows them to expand their dominance vertically or horizontally. It was in particular the lack of enforcement of data protection and competition law that enabled US Big Tech companies to benefit directly and indirectly from the mass collection and processing of personal data, as well as from the acquisition of potential competitors or additional technologies, thereby strengthening their market position.

In AI, we are confronted with a technology of the future, the consequences of which are only slowly becoming apparent. With the demand for evidence-based legislation, the path to under-regulation has already been set. With rapid technological development, it is easy to imagine potentially catastrophic consequences without being able to provide empirical evidence in every case. So anyone who says that laws should only be enacted if there is empirical evidence of problems or damage is preventing and postponing, at least for a long time, the forward-looking regulation of technology that has a long and successful tradition in Europe and has prevented a great deal of damage.

Incidentally, in a democratic state, the function of the law must not be reduced to that of a consumer watchdog, as is happening with the demand for "risk-based legislation." In a democracy, we shape the future through the law, and the ability to agree on a law is, first and foremost, a good thing in a democracy. We can see what happens to a society in which this agreement is no longer possible in Washington.

Instead of enacting binding legislation, more and more institutes for AI Safety research are being established all over the world. This idea was developed by the now-deposed conservative British Prime Minister *Boris Johnson*, who wanted to reduce the costs of AI development for private companies by having some of the safety research carried out by state-funded institutes. At the same time, he also wanted to be seen as a protector of democracy and human rights. In cooperation with willing governments, industry is pursuing a policy of nationalising risks and privatising profits, as has already been practised in the nuclear power indus-

try. Non-binding documents from the G7, the OECD, UNESCO and the UN, which have been significantly influenced by industry, describe noble principles. However, questions remain unanswered, such as what measures should be taken if an AI company does not comply with these principles, or if an authority appointed to enforce these principles fails to act, or if citizens or civil society report violations of these principles. .

In the United States in particular, the advantages of self-regulation are propounded in the discussions on AI regulation. However, given the existing power constellations in AI, it is evident that any rule that does not suit large corporations can only be enforced if efficient legal procedures are guaranteed, and well-resourced and courageous, i. e. independent, authorities are available and obliged to enforce them. This requires binding law, and thus more than just self regulation.

In addition, for all activities that pose risks invisible to the average citizen, such as nuclear power, smoking or the pollution of the oceans by ships on the high seas, a sad political reality must be taken into account: as long as citizens cannot immediately recognise a risk, members of parliament also feel no pressure to act and are often unwilling to steadfastly wage the painful battle against the lobbyists of large corporations and follow the advice of experts. It is easier for them to simply pass rules that large corporations agree with, even if it is clear from the outset that these are unlikely to be sufficient to effectively protect citizens.

In 2020, the authors of the present volume discussed the differences between ethics and law and explained why the law is not only superior to ethics as the basis for fair competition, but is also necessary to ensure that rules can be enforced.[93] This is because it is only the law which is binding and, unlike ethics, can be enforced by state power, even and especially against those powerful corporations that do not actually want to abide by the rules.

Today ethics is no longer promoted as an alternative to the law. In the era of the second Trump administration, which flouts the law repeatedly, it is obvious that ethics no longer has a chance of curbing the power of corporations. Of the more than 80 codes of ethics for AI issued by gov-

93 Nemitz, P. and M. Pfeffer (2023) *The Human Imperative: Power, Freedom and Democracy in the Age of Artificial Intelligence*.

ernments, companies, research institutions and international organisations (e. g. OECD, UNESCO, EU Commission), most have sunk into insignificance. There is no evidence that the multitude of ethical documents has changed the behaviour of the AI giants in any way. Nevertheless, AI ethics plays an important role in educating people and guiding individual behaviour. AI ethics can be based on philosophical, political or religious principles, and it is important to allow each person to develop their own orientation, while at the same time continually exploring which ethical rules regarding AI are capable of achieving consensus.[94]

It would be wrong to believe that the mere existence of law alone solves all the problems of artificial intelligence. Rather, it depends on which binding legal rules are passed by legislators and also on how these rules can be enforced and are actually enforced. The global dominance of a few technology companies that have systematically violated laws in the past – whether competition law, consumer protection law, data protection law and fundamental rights protection, non-discrimination, intellectual property or tax law – and got away with it has damaged both the rule of law and democracy. The ideology of disruptive innovation in technology ("Move fast and break things" – Meta CEO *Mark Zuckerberg*; retrospectively "hire a whole bunch of lawyers to go clean the mess up" – former Google CEO *Eric Schmidt*) has devastating consequences when applied to the law. The hubris of tech and AI ideologues has led to an attitude in Big Tech that laws are optional and that it is okay to break them. This attitude must be countered with a firm hand, i. e. consistent enforcement of the law. And with an ethical stance that puts compliance with democratic law at the top of the agenda.

What can be left to ethics with a clear conscience and what needs to be regulated by legally binding and enforceable laws can only be decided through the democratic process. Anyone who advocates for more individual freedom and more ethics and for fewer national, European or international laws is acting as a brake on a future policy oriented towards the common good. For simple deregulation cannot show how, in an already highly concentrated digital economy and in a world of autocrats, the common good can still be achieved. The mere absence of legal rules cannot guarantee freedom, democracy or prosperity for all in a world where

94 The authors are part of a group of 16 experts who have agreed on 10 rules for the digital world, available at: https://10rules.eu/en.html.

predatory capitalists want to abolish all binding rules in order to maximise profits to fantastical heights. The law is necessary because even in a world full of good intentions, binding rules are needed to coordinate behaviour.

The adoption of the EU AI Act (more on this later in this chapter) in 2024 was an important signal of an apparent change in mentality in the EU and an important precedent for upholding the primacy of democracy and the rule of law over technology and business models. It is an important step away from non-binding ethical principles towards binding and enforceable AI laws. However, the fact that the AI Act and the GDPR are already being weakened again raises doubts as to whether the change in mentality is sustainable. Good laws do not need to be regularly adapted and amended in line with technological developments, as some lobbyists believe, claiming that laws are like code. The best laws are those that remain in force for decades or even centuries and still serve their purpose, such as constitutions and civil law books such as the German Civil Code (BGB) or the French *Code Civil*.

Compared to other means of shaping society, such as through technology and the market, the law has a number of special features. In a democratic society, the law is the result of a process that is itself highly regulated by the law. This process ensures that the outcome of the legislative process reflects the results of democratic participation. Participation takes place through elections and the representation of diverse social voices through free expression of opinion, consultation and review of (interim) results and processes. Democracy only works with mechanisms to balance different interests, such as the much-maligned but indispensable democratic compromise, which requires the possibility of mutual understanding.

All of this often contributes to the legislative process being a lengthy one. However, the duration of the process reflects not only the desire to involve the population and various interest groups in the legislative process. It is also about integrating the current state of science, political orientations and values into the debate on the future shape of the law and finding a solution that can win majority support in the form of a multitude of compromises. Entrepreneurs can take radical steps in one direction – and either succeed or go bankrupt. Legislators, however, cannot take the risk of bankruptcy, in the sense of a complete loss of everything at stake. They are therefore regularly forced to compromise.

The law also distributes the risks and opportunities arising from new technologies and new business models. For example, who bears the risk that through "learning" AI will mutate in ways that are unpredictable for the manufacturer and no longer do what it is supposed to do, thereby causing damage? Regulation can either be prescriptive and detailed, which limits companies' creative problem-solving, or open-ended and technology-neutral, which necessarily leads to less legal certainty in the here and now. In the US, the development risk is not merely shifted to the general public – it is simply handled differently. For example, large AI companies in the US are being sued for billions, particularly by copyright holders such as The New York Times and others – and, of course, by parents whose children have died following interactions with AI systems. At the end of these trials, the developers will probably have to pay enormous sums in penalties in addition to the damage caused, and thus also punitive damages. However, this approach is neither better nor more efficient than the European model; rather, it is fraught with considerable costs and risks. The US system is retrospective (*ex post*) and lacks democratic procedures, because judges alone make decisions, often based on very old laws. In contrast, the European approach is based on *forward-looking* regulation through democratic processes. This difference is historical and can be traced back to the origins of authority in the Anglo-Saxon states on the one hand and on the European continent on the other, as well as to different paths in the development of democracy and philosophy.

However, draconian ex post fines, as in the USA, are not desirable in Europe. The European tradition of forward-looking technology regulation has worked well since the Second World War. Those in Europe who oppose this form of regulation certainly do not want the US system of punitive damages. They simply want to earn money on the market through innovation and, if damages occur, not be held responsible for them, i. e. privatise profits and socialise losses. Democracy can not, however, afford to allow this. Responsible entrepreneurs understand that they must also take responsibility for the risks their technology creates, and this gives them an incentive to minimise these risks and avoid damages by investing in safety and responsibility.

The democratic state is a guarantor state, which means that it does not have to do everything itself, but it does guarantee its citizens security, democracy and freedom in the future and must not abandon them to the wild and sometimes irresponsible fantasies of tech entrepreneurs and the risks that these create. It should set the goals for innovation, not seek

to regulate the technology in detail. Instead, regulation must focus on the public good to be achieved, such as guaranteeing fundamental rights and freedoms, and create a stable legal infrastructure to this end.[95] AI innovations must be measured against the public good. Where the market fails, it may also be necessary for the state itself to provide public infrastructure through investment. This is the case, for example, where the media is concerned.

The absence of rules benefits the powerful and harms the weak. This general principle applies to AI in particular. Consequently, European legislators have enacted a number of laws in recent years that specifically target AI systems and their application, as well as the data on which they are trained.

Role and shortcomings of European AI regulations

The purpose of the EU AI Act of June 13, 2024 is product safety in the traditional sense pertaining to engineering expertise, such as ensuring that an electrical device does not cause physical harm to humans. But – and this is revolutionary – the AI Act also obliges AI developers to protect society and its core structures, which is to say democracy, fundamental rights and the rule of law, from the risks of artificial intelligence. Programmers, computer scientists and engineers do not have the specialised knowledge in this area that they have on engineering safety issues; such goals are initially alien to most of them and the companies they work for, as they are neither part of their education nor of their core business, in contrast to technical safety regulations.

The EU AI Act, often decried as overregulation, does not require prior approval of AI programs by third parties, even if they are released for a wide variety of purposes for billions of people and in uncontrollable and unpredictable contexts. All manufacturers certify their AI products themselves as compliant with the rules of the AI Act. Compared to traditional industry, this is obviously a huge gift and a huge subsidy for those who make money from AI – especially US companies.

95 See: Hoffmann-Riem, W. (2016) *Innovation und Recht – Recht und Innovation: Recht im Ensemble seiner Kontext* (Tübingen: Mohr Siebeck).

It is astonishing, if not reckless, that the AI industry has forced legislators to accept that its products, which are now as ubiquitous in our lives as electricity, can be brought to market and operated without prior safety checks by third parties before they are put into service. This may have been justified for simple software programs from the pre-AI era, such as Microsoft Office. But with AI, which *Bill Gates* and *Elon Musk* have described as a technology comparable in danger to nuclear power, the situation is quite different. The fact that a company like Microsoft, which claims to always act in the public interest, does not recognise the need for independent ex-ante testing and certification of complex AI programs shows that its supposed focus on the public interest has serious limitations. The complete absence of such ex-ante reviews, even for high-risk AI, means that Big Tech can conduct the largest social experiment in history with AI. The profits are reaped by US Big Tech companies, but the risks of this experiment are borne by the general public. Nevertheless, the AI Act is an important signal that democracy is taking up an issue of great social importance and fundamental rights.

The EU Digital Services Act (DSA) of October 19, 2022 ensures that large platforms with more than 45 million users in the EU (the "gatekeepers") do not engage in self-preference in their commercial functions, in the way that Amazon has done by displaying its own products first and lists other sellers' products later. Importantly, the DSA also obliges platforms to assume structural responsibility to ensure that the content they provide, including content uploaded by third parties, does not collectively constitute a breeding ground for the undermining of democracy and fundamental rights. The DSA thus positions itself between the total laissez-faire approach of US law, according to which platforms are not obliged to moderate content on their own initiative, but only when they are made aware of illegal content (the "notice and take down" principle), and European media law, which is based on the legal principle that editors bear full responsibility for what they publish, including third-party content, and thus every journalistic medium is obliged at all times to ensure on its own initiative that no illegal content is present on its platform. Media law therefore requires real-time moderation and immediate response to each individual piece of content on the initiative of the editors themselves, even if it has been uploaded by third parties. This comes at a price: the internet presence of the free press in Europe currently costs on average 8 percent more than that of platforms that are not subject to press and media law.

The EU Digital Markets Act (DMA) of September 14, 2022 strengthens the powers of EU competition authorities to intervene at an early stage to prevent dominant market positions and specific forms of abuse of dominant market positions. The aim is to maintain lively competition in the market and, in particular, to prevent the systematic acquisition of start-ups in order to maintain or establish a dominant market position by restricting competition – behaviour that could not be countered under current competition law.

The GDPR, the Digital Markets Act, the Digital Services Act and the AI Act are intended to align the development and use of digital technologies, the internet and AI with the public interest. But one thing is clear: these laws will not automatically lead to AI being used for good. At best, they can prevent the worst from happening, and even then only if they are consistently enforced. However, given the political situation in the EU, it is questionable whether rigorous and comprehensive enforcement will take place.

The law is a necessary but not sufficient condition for bringing AI into line with the public interest. Many elements of these laws (as well as traditional competition law) are difficult to enforce because they are the result of political compromises in the European Parliament, which represents the people of Europe, and in the Council of Ministers, which represents the EU nation states. This is not unusual. However, there is neither a strong general will nor do all the authorities responsible for enforcing these complex laws have sufficient resources to enforce them with the necessary rigour. If Trump has his way, these laws will not be enforced at all against US companies.

Enforcement of the law

Shoshana Zuboff has described the culture of lies at Alphabet/Google in detail in her analysis.[96] Something similar applies to Meta. In Europe, liars can be sanctioned under competition law, and in the US, managers can even face prison sentences for illegal price fixing and false statements before Congress. However, there is currently no legal basis in Europe that allows for the punishment of knowingly false statements made

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

to the European Commission, the European Parliament and the Council of Ministers during hearings in the legislative process. There is not even a legal basis that provides for exclusion from future hearings, let alone the imposition of prison sentences or fines. The irony of this situation is that corporations contractually threaten their employees with penalties amounting to millions of dollars if information from research or business operations is leaked to the outside world while themselves not being subject to sanctions if they lie in legislative procedures.

In conclusion, it is clear that in the context of AI, a discussion of power and abuse of power is necessary. It should also be borne in mind that procedural rules are often more important than the content of regulations themselves. Without adequate procedures for enforcement, even the best law is ineffective. Preserving the dignity of the law and ensuring the functioning of democracy, even in a future shaped by extremely complex AI, necessarily requires the implementation of a sanctioned obligation to truth in the legislative process and in the enforcement of all digital legal acts.

In addition to public enforcement by data protection authorities, the GDPR can also be enforced by private parties and, in many cases, by civil society through lawsuits in the public interest. For this reason, the incentives to comply with the GDPR are higher than in the context of the AI Act. It is currently unclear whether there are private enforcement rights under the AI Act and whether a citizen who complains about non-compliance with the AI Act to a market surveillance authority responsible for enforcing the AI Act can actually force the authority to take action by bringing the matter before the courts.

We are convinced that the introduction of a fundamental rights impact assessment in the AI Act by the European Parliament and a number of other elements, not least the fundamental difference between AI and traditional product safety in terms of protection of important interests guaranteed by law, such as fundamental rights, could be an argument in favour of reducing the discretionary power of market surveillance authorities in handling complaints, which they traditionally have under EU product safety law, to zero. This would mean that, in contrast to traditional market surveillance for product safety, market surveillance authorities would have a duty to protect individuals when applying the AI Act by following up on individual complaints to the authority, at least in certain situations, such as those relating to claims of violations of fundamental

rights. Obviously, such an interpretation of the law would significantly strengthen the impact of the AI Act and better ensure compliance with the regulation.

As explained above, the AI Act is not the only, nor is it the most important, EU law that applies to AI. The GDPR will become even more important in relation to AI. The simple reason for this is that all AI activities that amount to the processing of personal data must comply with the GDPR. And the GDPR can be enforced by private parties, often through public interest litigation against companies, in addition to public enforcement by data protection authorities. The incentives for compliance are therefore higher for this reason alone.

The processing of personal data by large companies, but also by the state, takes place behind the walls of corporate and government data centres, invisible to citizens. Only when multiple means of enforcing substantive laws are combined is there any chance that the laws will have an effect. In the area of personal data processing and AI, there needs to be a combination of three forms of law enforcement: firstly, by public authorities, which can act ex officio and must investigate complaints from citizens; secondly, through direct legal action by affected citizens against companies or the state; and thirdly, through legal action by civil society in the public interest. This combination of enforcement methods must be included in all European digital legislation, including the AI Act. Without this combination of enforcement mechanisms, it is unlikely that substantive law will be properly complied with, given the power constellation and unscrupulousness in the digital economy. Experience shows that there is a very high willingness to simply break the law in the digital economy.

Democratic sovereignty and technological independence

The battle between corporations involves the classic Big Tech companies from the USA with their respective AI programs and applications: Microsoft with OpenAI, Alphabet (Google) with Gemini, Meta (Facebook, WhatsApp, Instagram) with Llama, and *Elon Musk* with xAI with Grok. These corporations provide 33 percent of the capital spent on AI development and as much as 67 percent of the capital spent on developing generative AI. Big Tech also has the data needed to train AI, or can constantly skim this data. Although there are between 300 and 1,000 foundation models, depending on how you count them, we can see from their

application, using chatbots as an example, that this is already a very narrow and concentrated market. ChatGPT has a market share of over 80 percent in the consumer AI chatbot market in 2025.[97]

The concentration of end-user AI in the hands of a few companies exacerbates a concentration of power that already existed in the digital economy before the era of GPT. Unlike the World Wide Web, which was designed as a decentralised system of interconnected computers and whose concentration of power was only brought about by cloud systems, social networks and internet search, AI in the form of GPT was designed as a centralised system from the outset.

It is not only end users who support centralisation and place themselves in a position of dependency, but also states. According to a study[98] by Agora Digitale Transformation, the German federal budget for 2024 allocated 28.7 million euros for "digital sovereignty and sovereign tech" and 89 million euros for sovereign data infrastructure and AI, while Microsoft received 204.5 million euros for licences from the federal budget alone in the same year. Other than in political sermonising, digital sovereignty cannot succeed this way. The situation is similar when it comes to data processing. 75 percent of all federal administration data is already processed using Oracle products, and in 2023, the German Interior Ministry signed a framework agreement to purchase Oracle products worth almost 4 billion euros over five years.[99] These figures show how far apart political rhetoric about digital sovereignty and budgetary reality are.

The danger posed by this centralised dominance by a few private corporations, whether from the USA, China or elsewhere, is demonstrated by a growing number of incidents: time and again, the cloud services of one of the three providers dominating the globe – Amazon Web Services, Microsoft and Google – fail. When this happens, many things simply stop working in many parts of the world because we are all dependent on these cloud services. Governments and companies are also dependent

97 Kranjec, J. (2025) "ChatGPT's Market Share Surges to 82.6% in July, Nearly 5× Its Top 5 Rivals Combined". *JemLit*, 31 August.
98 Heumann, S. (2025) „Analyse und Kommentierung des Haushalts 2026 mit Fokus auf den Etat des Bundesministeriums für Digitales und Staatsmodernisierung (BMDS)". *Agora Digital Transformation*, 24 November.
99 Voß, O. (2023) „Umstrittener 4-Milliarden-Deal: Macht sich das Innenministerium von Oracle abhängig?" *Tagesspiegel*, 6 September.

on centralised services, especially the cloud, and are losing control over their data. Anyone who stores data with one of the three giants cannot be sure that the data will not be passed on to the NSA, because this is possible at any time under the CLOUD Act and other US Security laws, part of which were enacted by the first Trump administration.

Big Tech's mass harvesting of personal data and copyrighted works shows that we can no longer rely on these giants to uphold the law or contractual commitments. They have embraced a culture of "we can do anything we want, we are above you." This was already widespread among them before Trump II. But now a Trump II culture of mutual enrichment is deeply entrenched in US business and politics.

How can we defend ourselves against this and find new ways to ensure data reliability and security? By breaking new ground in technology, law and politics, and by considering fundamental rights, the rule of law and democracy in technology development and procurement from the outset. In digital technology and AI, there are many ways in which the reliability of services and the plurality and decentralisation of systems can be designed in such a way that no dependencies arise, or at least so that dependencies on unreliable candidates, namely US corporations under Trump and Chinese service providers and technology suppliers, are reduced.

Europe has long been the continent with the highest market share in Google searches worldwide. The communication and knowledge production systems of public administrations, universities and companies are almost entirely in the hands of Microsoft, whose already dominant position is further strengthened by the fact that it now effectively owns OpenAI.

Individual German states such as Schleswig-Holstein, with its 60,000 civil servants in public administration, are following the earlier model of the city of Munich and have said goodbye to Microsoft, as they have recognised the high price of dependence. To reduce costs and dependence, they are switching to open source. Unfortunately, Munich has abandoned this open source role model and returned to Microsoft. Is it just a coincidence that Microsoft built its German headquarters in Munich at around the same time? Even after such developments, the political wind can change again – at the beginning of 2026, the city of Munich passed a package of measures aimed at digital sovereignty, including the use of

a *digital sovereignty score* for IT processes, which was developed in cooperation with one of the authors of this book and is intended to help identify critical dependencies.

Many EU member states have now investigated and denounced their dependence on Microsoft. However, only a few have made the effort to ensure that government data remains in Europe, for example through public procurement rules (as the Netherlands has done). The EuroStack[100] is an idea for an independent digital infrastructure in the EU. The EU and governments traditionally invest in public infrastructure in Europe, which is part of Europe's strength. The AI Factories as a public resource, as the EU is now implementing them,[101] are a first step, but this does not go far enough: Elon Musk's Colossus data centre alone will be larger than all EU AI Factories combined. More must follow in terms of public digital infrastructure. Above all, in the age of AI, investment is needed not only in hardware, but also in platforms, AI systems, software and democracy-compatible governance. One example: a common European operating system for software and AI in public administration could pave the way out of the public sector's complete dependence on Microsoft. A common EU operating system could create synergies in development and maintenance, interoperability and greater security at lower costs than Microsoft licences.[102]

The International Court of Justice in The Hague has just set an example: instead of Microsoft, it is now using open source programs from the OpenTable package from ZenDiS, the Centre for Digital Sovereignty owned by the German state, as a precaution against possible US interference.[103] Another advantage of using Microsoft alternatives: US law allows intelligence agencies to access data managed by US companies outside the US under certain conditions. Some EU member states have made efforts, for example through public procurement regulations, to ensure that government data remains in Europe and that there is no path dependency on a single provider for essential technologies. More determination in this direction is needed.

100 See: "Eurostack: Building Europe's digital future". Eurostack Initiative Homepage.
101 "AI Factories". European Commission.
102 Riemann, R. (2025) "EU OS". *Blog Riemann*, 2025.
103 Knop, D. (2025) "International Criminal Court Kicks Out Microsoft". *Heise Online*, 30 October.

Systematic procurement with specifications for AI and digital technology that allow open source and interoperability of systems as well as unhindered and smooth real-time transfer of data between different providers is the way forward. Without these conditions, which promote transparency, decentralisation and innovation, and which correspond to European democracy in its decentralisation and its medium-sized economic structure, there will be no path to digital sovereignty for Europe. Europe is ill-advised to chase after the unicorn hype of US venture capitalists; small and medium-sized enterprises must also be taken into account and the market and opportunities kept open for them. An economy with many medium-sized companies is also better for the broad distribution of work and prosperity and for democracy than having an entire society dependent on a few large AI corporations.

India and Brazil have introduced state payment systems that are managed by the central bank. In doing so, they are breaking the power of Visa and Mastercard from the USA as payment systems. For their companies and citizens, this means a "tax relief" of approximately 3 percent. This is because the fees that Visa and Mastercard charge on every payment are nothing more than a tax which, unlike taxes levied by the state, does not benefit the common good, but only the stock market value of these companies. A future digital euro will also enable us in Europe to make savings. And not only make savings, but gain independence. This is demonstrated by the example of Nicolas Guillou, a French judge at the International Criminal Court in The Hague. Because he signed the international arrest warrant against Netanyahu, he is on the US sanctions list, along with 1,500 other people. Since then, his accounts with Amazon, Airbnb and PayPal have been closed, he can no longer book hotels through US services, and he can no longer pay with Visa, Mastercard or American Express.[104] This is how quickly life can change in a world of monopolies when you suddenly find yourself in the crosshairs of the monopolists. Judge Guillou is paying the price for the 90 percent US dominance of Europe's digital markets. The same power could easily be used against individual states or the EU as a whole, and is already being used to exert pressure in the interests of the US.

104 Maupas, S. (2025) "Nicolas Guillou, French ICC judge sanctioned by the US: 'You are effectively blacklisted by much of the world's banking system'". *Le Monde*, 19 November.

Against this backdrop, it can already be said that the talk of Europe being big on regulation, while the US and China are big on innovation, is propaganda. The fact is that Europe is innovative in its decentralisation and diversity when we look at technical, social and political innovation together. Innovation can be found in Europe in the market as well as in science, technology and democracy. Ideas and approaches to open source, Fediverse, Mastodon, and many blockchain innovations come from Europe. In order to unleash European innovative power, there must be no renunciation of European values. Nor is there any need to. Meaningful regulation, combined with investment incentives and a unified European capital market, can ignite an engine of innovation. European AI research is already outstanding, and with the right measures, this could be translated into technological leadership. If the USP of AI made in Europe is to safeguard autonomy and democracy, then the corresponding products will experience increasing demand worldwide.

Cross-border exchange, cooperation and the constant need to reach new agreements in Europe are important drivers of democratic innovation, which in turn generates technical and economic innovation. Leaving the open future to be shaped solely by innovators in technology and business means foregoing democratic innovation and thus excluding many people from the innovation process. This creates dependencies and inequality, whereas the EU, on the contrary, must seek the path to technological, economic and democratic sovereignty.

Fundamental rights and shaping the future

When it comes to payment systems, AI and all digital systems in general, there is one common theme that is of utmost importance for the future of democracy and freedom: the protection of personal data. Precisely because we are increasingly connected and use digital systems for all the expressions and activities of life, right down to our deepest thoughts, feelings, political, religious and sexual interests, dreams and illnesses, it is so important that our data is protected. The intensity with which we use these systems allows for the total screening, surveillance and manipulation of each and every one of us. The law has found a good response in many areas to people's inability to see themselves as victims. One example is the legally mandatory, i. e. compulsory, liability insurance for motorists. The question of whether motorists take out insurance cannot be left to each individual.

The same applies to data protection: if I upload my phone book with all my contacts to a platform in order to find friends, I am handing over the personal data of others to the platform. This problem cannot be solved by prohibitions aimed at individuals. Rather, the data processors, i. e. the platforms, must be made responsible. And that is exactly what the EU's GDPR does. It creates institutions and equips data protection authorities with extensive rights to shed light on opaque data processing and enforce the law. In the future, in a world of AI and mass data processing, they will have to enforce the law much with greater rigour.

Too many data protection authorities spend a lot of time talking their way out of responsibility and justifying why they do little or nothing and hardly ever get involved in the conflict-ridden enforcement of the General Data Protection Regulation. Two approaches to data protection are important here: first, the focus on the large collectors and processors of data. In short, anyone who collects and processes large amounts of personal data, or who earns money exclusively or primarily from doing so, must be subject to very close, regular, intensive and active official monitoring by the data protection authorities. This must become a priority for these authorities. The popular AI chatbots are prime candidates.

Secondly, comprehensive personality profiles must be banned, as they undermine democracy, individual freedom and human dignity. Fortunately, they also clearly violate many rules of the GDPR, including the principle of data minimisation, which stipulates that as little data as possible should be collected and processed. Data protection authorities must impose very heavy penalties on anyone who creates comprehensive personality profiles from linked data. The profiles created by the popular AI chatbots and social networks must become the top targets of enforcement action.

The power, lack of transparency and comprehensive presence of digital technology and AI require a stronger countervailing power and presence of supervisory authorities, not only in data protection, but also with regard to AI itself, the behaviour of platforms and competition. The fragmentation of authorities and legal acts in Europe must be gradually eliminated. In the US, data protection, consumer protection and competition protection are handled by a single authority, the FTC. This is a good example in itself. It is no coincidence that restricting the independence of this authority is central to the Trump/Vance administration.

Competition policy, consumer protection and data protection, as well as control of AI as a technology, should be brought under the aegis of a single authority. This increases efficiency and enforcement power and avoids duplicate structures. The independence of such an enforcement authority must be clearly enshrined in law.

The European Commission has a special role to play in regulating digital technology and AI, both in shaping new legislation and in enforcing it. While data protection law must be enforced decentrally, at least according to the current General Data Protection Regulation, competition law, AI law and the law on the regulation of platform behaviour, the so-called Digital Services Act (DSA), are enforced centrally against the Big Tech companies by the Commission.

It must resist political pressure, currently mainly from the Trump administration, but at other times also from member states, to refrain from enforcing the law. And it must redeploy staff to these areas much more than it has done so far. Weakening the GDPR and opening the floodgate for processing personal data by AI, as now proposed in the Omnibus by the European Commission, is the wrong way to go. AI makes it much easier to identify people. It also makes it much easier to target and manipulate people. Thus, in times of AI, data protection must be strengthened not weakened.

In 2024, a high-level UN advisory group on AI was unable to agree on binding global rules for AI. This was mainly because the big AI companies such as Microsoft and Google were represented by staff in this group and prevented consensus on binding rules. They were able to count on the support of the US, Russia and China – an interesting alliance.

In the summer of 2025, the UN was only able to agree on two non-binding procedures: a panel of scientists is to be set up to report on the risks of AI, and an international dialogue on the governance of AI is to be initiated.[105] A mandate for a global agreement on the fundamentals of AI to jointly reduce the worst risks could not be adopted. Only in the area of

105 United Nations General Assembly, *Terms of reference and modalities for the establishment and functioning of the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance*. Seventy-ninth session, Agenda item 123, Strengthening of the United Nations system, (August 2025), https://docs.un.org/en/A/79/l.118.

military use of AI is there a small opening to transfer the negotiations in Geneva on new future weapons, which have been deadlocked for years, to a new process for AI, with a view to reaching a binding agreement.

The reaction of civil society and many scientists, including numerous Nobel Prize winners, was one of horror. They then issued an appeal to the nations of the world to adopt a binding agreement on minimum standards, the so-called red lines[106] which must not be crossed in terms of risk. At first glance, the red lines presented in the campaign have nothing to do with democracy. However, the campaign's demand to "prevent universally unacceptable risks" is based on a fundamental prerequisite of democracy, namely the will and ability to repeatedly reach agreements that are supported by at least a majority and that guarantee a good future.

The red lines campaign will fail if it demands unanimity among all UN members and the companies concerned in every respect. It will succeed if a majority of states participate and equip the binding red lines with the "robust enforcement mechanisms" the campaign demands. Even if China, the US and Russia do not sign up to an agreement on binding red lines for AI from the outset, their companies that market and use AI worldwide will still have to comply with this agreement if they want to market their AI globally. Similar to the Covid virus and climate change, AI also poses a common risk to all states and peoples of the world. In this risk community, even dictatorships and democracies have a common interest in maintaining governance in the face of powerful Big Tech corporations and their AI systems. This common interest requires a binding agreement on red lines, because without a minimum level of control and mastery of AI based on common rules, the governance of states, whether democracies or dictatorships, and the world cannot be secured in the long term.

Against this backdrop, the campaign for red lines must be distinguished from efforts to develop AI in such a way that it serves the greater good, which is a matter of considerable debate. AI systems that are based on democracy, fundamental rights and the rule of law from the outset and internalise these values "by design" will only exist if the democracies of the world adopt fundamental legal requirements that go well beyond the EU's AI Act, via cumbersome international law that often does not go beyond

106 "We urgently call for international red lines to prevent unacceptable AI risks". AI Red Lines website.

the lowest common denominator. Similar to environmental protection legislation, which developed in several stages and was always associated with political struggles, AI legislation will also have to be developed in several stages and will only come about with a huge struggle. The AI Act is a first step that only serves to avert danger. The next step must be a legal obligation for AI and platforms to respect fundamental rights, support democracy and comply with the law by design.

Further development of digital law

The technical and economic innovations of AI as a general-purpose technology and the platform economy must be accompanied by a corresponding platformisation of the regulation of the platform economy and AI as a general-purpose technology. The thematic and sectoral division between competition law, ex-ante regulation of the behaviour of platform giants, data protection and the data economy, and sector-specific rules in the data economy no longer make sense in the age of AI as a general-purpose technology and the platform economy. The double fragmentation of EU digital law, both by topic and, in terms of enforcement, by country, robs this law of much of its effectiveness and at the same time increases the costs for enforcement authorities, businesses and citizens seeking justice.

Violations of regulations in the field of data economy and artificial intelligence will often affect several pieces of digital legislation at the same time, because the same data is used for many different purposes in the platform economy and, at the same time, processed in a wide variety of ways using AI. Violations will not only have to be dealt with from the perspective of one law. For example, the German Federal Cartel Office has deemed the data processing practices of Facebook and WhatsApp, namely the merging of personal data from both companies into a personality profile, to be a violation of competition law based on a violation of data protection law. The European Court of Justice (ECJ) upheld the decision as lawful, although the issue at hand was a data protection issue, traditionally dealt with by data protection authorities.

Ideally, the fragmentation of enforcement authorities should be eliminated and multiple authorities merged into large enforcement agencies, similar to the model of the US FTC, and the large public prosecutor's offices, which prosecute a wide variety of cases, from economic criminal law to serious crimes such as murder and manslaughter, organised crime, child

pornography, environmental crimes and tax offences: a broad spectrum of topics is regularly covered by a central public prosecutor's office in criminal law. It is, of course, divided into specialist areas, but the joint authority structure increases flexibility, allowing technical, economic and human resources to be allocated to new topics as the situation develops. Modern administrative structures must be capable of continuously organising rapid and interdisciplinary cooperation between different specialist departments, in ever-changing constellations, in line with the rapidly changing demands of the times. Ultimately, the aim is to reduce the high transaction costs between different, possibly still independent authorities by internalising them into a single independent authority.

In the US, from which there is much to learn in this area, the Federal Trade Commission (FTC) traditionally has a threefold responsibility. It is responsible for enforcing competition law, consumer protection and also data protection as a subcategory of consumer protection. This centralisation in a single authority has major advantages. This is particularly because the tough enforcement methods of competition law that have developed over decades in the US and Europe could be transferred quite quickly to other areas under such a unified authority. The maximum penalty for violations of data protection rules has so far been imposed by the FTC and not by a European data protection authority. It was a penalty of several billion dollars against Facebook, which failed to live up to its public promises on data protection.

Much could be gained if a simple European Union legal act were to create a common enforcement basis for all the authorities mentioned. This legal basis must name the authorities to be considered, i. e. competition authorities, data protection authorities, consumer protection authorities, financial market supervisory authorities, insurance market supervisory authorities, equality authorities, digital regulators, and so on, specify the underlying legal acts of European law, and then stipulate that these authorities are permitted to exchange information obtained in the course of investigations, including in ongoing proceedings, with other authorities listed.

If Europe does not take this path, it will find that, for example, a competition authority that discovers during its investigations that the DSA is being violated or that gender equality is not being taken into account will not be able to do anything with this information, nor can it pass on the findings to the competent authority, meaning that the very legal viola-

tions that arise from the use of modern, highly profitable technologies such as AI and highly profitable business models of the platform economy go unpunished. The platform and AI corporations are "customers" of all enforcement authorities and have a holistic view of them, while the authorities' view of the corporations is fragmented. This results in permanently suboptimal enforcement of the law, which is already severely limited due to the virtually unlimited resources that corporations spend on lawyers, economic studies and lobbying to prevent or at least hinder law enforcement. Every negative decision by an enforcement authority is challenged in court up to the highest instance. The corporations thus send a clear signal of intimidation to the often overburdened and under-resourced enforcement authorities: if you mess with us, a large part of your staff will be tied up in litigation with us for years. Many enforcement authorities in data protection law actually refrain from enforcing the law due to this structural inferiority, while the lawbreakers get off and profit from it. In sum, in addition to under-regulation, the central problem is a glaring weakness in enforcement, which is partly due to the complexity of the law, partly to the under-resourcing of the authorities responsible for enforcement, partly to the fragmentation of the enforcement authorities, and partly to the lack of legal remedies to enforce compliance. All these adjustments must be made in order to guarantee effective enforcement of democratic law in the face of the enormous, concentrated power of digital corporations.

AI law must continue to make a clear distinction between humans and machines. AI must not be granted rights that were created for humans. For example, the right to offensive speech as an expression of the human right to freedom of expression does not exist for AI. AI systems that pollute the public sphere millions of times over with fake news, illegal or aggressive statements and lies should be shut down by law, as this machine activity should not benefit from the human right to freedom of expression. It undermines the space for free human expression and makes it more difficult for humans to be heard. Unfortunately, there is currently no legal regulation that obliges operators to shut down such AI and enables the state to intervene quickly when AI pollutes the public sphere in this way, unless the threshold of the DSA is reached.

Intellectual property rights, which were created by law to protect human creativity and genius, justify protection rights for the output of AI. The reason for this is simple: the purpose of the law is to protect the human mind and the threshold of originality it produces, and thus also liveli-

hoods, but not machines and their profitability from production without human genius.

In all considerations regarding the further development of AI regulation, it must be borne in mind that the protection of intellectual property and data protection are cornerstones of democracy. Both of these fundamental rights are not just specific rights for individuals. In a society without the protection of personal data, where the government and powerful private entities can completely and constantly monitor every person and create comprehensive profiles, democracy cannot function because opposition and dissent are immediately suppressed and surveillance takes over. That is why the comprehensive personality profiles created today by digital platforms are so dangerous. They allow for total control and manipulation. In the US, TikTok's profile data in the hands of China has been recognised as a national security issue. However, in the hands of any other government or private entity, it also poses a threat to democracy. It is therefore important, precisely because AI makes it increasingly easy to identify people with the help of data that was previously not considered personal data, and because AI also makes it easier to control and manipulate people with the help of personality profiles, that comprehensive personality profiles are abolished and data protection is rigorously enforced in other areas as well.

Without the protection of intellectual property, different ways of thinking, subversive culture, and everything that constantly reshapes innovation in society and utopia and paves the way for a better future will wither away. Those who want to stifle creativity and different ways of thinking are abolishing intellectual property, thereby destroying the creative sector and the vibrancy among people that keeps democracy alive.

6 | Democracy only with deliberation

Information is the basis of every free decision. Only those who are well informed can decide freely. When people make decisions together, in addition to reliable information that is equally available to all, there must also be free communication in which opinions are formed. The right to freedom of expression is therefore fundamental to democracies. But forming opinions is a qualitative process. First and foremost, it requires that opinions be formed on the basis of generally accepted facts. It then requires the ability to change perspective. As Hannah Arendt points out, the ability to take on the perspective of others is the very beginning of politics.[107] It consists of the first and second persons singular and plural assuming one another's roles. This requires the speakers to be willing to contribute their values and opinions, to understand those of others and to change them together in discourse. Only in this way can a consensus be reached that is supported by all. The space in which this mutual understanding between individuals and societies takes place is traditionally called the public sphere. This space was initially the agora, later the theatre, the coffee house, the meeting place in large cities, before it increasingly became the media space in which information and opinions circulate in society.

One digital pioneer is credited with saying that people first lived in villages, then in cities, and now they live on the internet.[108] The fact is that the public sphere has been increasingly transformed into cyberspace through digitalisation and, in particular, the introduction of social media. The age of mass media, which operated one-way, one-to-many communication, has been replaced by the age of direct media, in which everyone is potentially a sender and receiver, or *a* "prosumer" in modern parlance. Not only does many-to-many communication now prevail, but above all, the algorithms and business models of the large platform companies dominate the supply of information and the formation of opinions among users.

107 Arendt, H. (2005) "Socrates", in J. Kohn (ed.), *The Promise of Politics* (New York: Schocken Books).
108 As stated by the actor playing Sean Parker in The Social Network, 2010.

The power wielded over the public by these Big Tech players goes far beyond the power wielded by the press barons in the traditional media. In the age of mass media, opinion-forming power was wielded by private publishers and media moguls, who were kept in check after the Second World War by a dual system of concentration control with public media as a counterweight. In contrast, the power wielded by digital companies is much more comprehensive. The collection, sampling and calculation of personal data gives tech companies comprehensive surveillance and control power. They have achieved the formation of large monopolies and oligopolies by taking advantage of so-called platform privilege, whereby social network operators reject responsibility for content while earning handsomely from it. By conditioning algorithms to an advertising model based on personal data, from which personality models and predictions are created with the aim of manipulating behaviour, they have a powerful tool at their disposal with which entire societies can be monitored and politically manipulated. At the same time, public discourse is becoming more brutal because the platforms offer a lucrative business model that allows anyone to earn money by exploiting excitement and scandal. The cultivators of hatred drizzle their poison in doses of malice, creating widespread cynicism that continues to shape the communication of entire right-wing populist parties such as the AfD in Germany to this day. Which causes which? The crudeness of the algorithms or vice versa? This is the chicken-and-egg problem of the digital public sphere, from which right-wing populists leave themselves and Big Tech with no way out.

Meanwhile, the disruptive effects of digital technology on the public sphere are increasingly turning democracies into failed states because they have neglected to curb the concentrated financial, economic and now increasingly political power of Big Tech. Indeed, we are seeing Big Tech players allying themselves with kleptocrats to eliminate the rule of law, journalism and other means of controlling the concentrated power they wield.

This alliance has forced liberal democracy into a retreat that seemed unthinkable just a few years ago. The arena for this shift in power is virtual. This makes it difficult to grasp. But even disembodied cyberspace needs space and resources for its hardware, on which programs run and data is collected. Within it, the algorithms of the Big Tech platforms act as invisible dispositifs (i. e. apparatuses) of power, directing the flows of information and communication and thus deciding what knowledge forms the basis of opinion-forming and, consequently, political decisions. With

the introduction of personalised AI agents that offer users the knowledge that corresponds to their data profile and the interest-driven intentions of the operators of these agents, the common public space is being eroded and thus the basis for a new form of totalitarianism is being created. AI agents are systems that not only chat, but also plan and act independently. The degree of their independence is defined by OpenAI, for example, on a scale of five levels of autonomy, which are intended to pave the way for general AI.

"Everyone is locked into their subjectivity as if in an isolation cell," is how Hannah Arendt described totalitarianism in Vita Activa back in 1958. She was convinced that the disappearance of diversity of perspectives would also destroy the shared access to the world that public space makes possible. "A shared world disappears when it is seen from only one perspective; it exists only in the diversity of its perspectives."[109] When these shared perspectives are lost, conformism to predetermined opinions prevails. The self-determination of individuals and societies is no longer possible. Digital capitalism increases the risks of social conformity, surveillance and control, while at the same time attacking the core of autonomy at the very heart of human beings. This undesirable development was made possible because the digital structural change in the public sphere was only understood in retrospect. The owl of Minerva flies only at dusk.

To talk about the structural change in the public sphere means to take up Jürgen Habermas' analysis, who, 60 years after introducing the public sphere as a category of social analysis, addressed digitalisation as the driving force behind a "new structural transformation."[110] In his analysis, the digital transformation of communication infrastructures subjects the problem-solving judgement practices practised in the age of mass media to a stress test. Digitalisation is creating a global dissolution of boundaries in communication, both spatially and temporally. Communication flows are becoming more concentrated, differentiated, diversified and generalised. What is new is the platform character of the intermediaries, which are the new media and which do without any journalistic mediation and design. As irresponsible mediators, they provide contingent, unpredictable links and accelerate contacts, thereby generating new, un-

109 Arendt, H. (1958) *The Human Condition* (Chicago: University of Chicago Press), p. 58.
110 Habermas, J. (2023) *A New Structural Transformation of the Public Sphere and Deliberative Politics* (Cambridge: Politiy).

predictable discourses. The "gatekeeper model of mass media," which consists of professional selection and discursive examination of content based on generally accepted cognitive standards, has no equivalent on internet platforms. However, their integrative power is increasingly being eroded by the dynamics of delimitation and fragmentation.

The common ground of information perceived by all is being replaced by fragmented communication spaces that become communication cycles. They tend to isolate themselves from one another, but also to position themselves against one another, creating "a maelstrom of self-referential reciprocal confirmation of interpretations and statements." While this gives rise to a type of communication characterised by constitutive narcissism, which seeks confirmation and at the same time immunises itself against objection, Habermas believes that the "inclusive meaning of the public sphere" is lost, and with it the public sphere in which all citizens can communicate about issues relevant to the community. It is obvious that this undermines the core of a deliberative democracy based on autonomous citizens.

This brief excursion into Habermas' theory of the public sphere serves not only to demonstrate the threat to democracy that has long been evident in the advance of a data-economic advertising model as the organising principle of the public sphere. It also leads to a process that demonstrates the unscrupulousness of Big Tech companies in their response (or lack thereof) to urgent criticisms.

The "Habermas Machine"?

At the end of 2024, Science published an article[111] that introduced the astonished public to the so-called Habermas Machine under the headline "Google's DeepMind develops AI to stop us hating each other." The name of the philosopher of consensus was thus to be used to promote the naturally "superior" problem-solving capabilities of AI. Such use is perfidious not only because it is based on false assumptions and draws false conclusions from them, but also because, in this case, an AI was developed that promised to solve the problem of polarisation in public debate, which had previously been caused by AIs themselves.

111 Tesla, M. H. et al. (2024) "AI can help humans find common ground in democratic deliberation". *Science*, 386(6719). DOI: 10.1126/science.adq2852.

Habermas, who was made aware of the misuse of his name by one of the authors of this book, immediately objected, offering a philosophical explanation as to why it is not possible to delegate the process of conflict resolution to a machine. This is because discourse theory is based on assumptions that exclude machines as equal actors: "Each participant in the discourse is expected to perform the demanding task of sensitively adopting the perspectives of others, from which he or she can assess his or her own interests and values in light of the equally affected and, where applicable, impaired interests and values of all other potentially affected parties, and adapt them accordingly if necessary." The representation of a person by a machine cannot succeed because AI cannot develop a personal perspective. However, this is precisely what DeepMind claimed in its publication: "We investigated whether an AI system based on LLMs can successfully capture the underlying shared perspectives of a group of human discussion participants by composing a 'group statement' that the discussion participants would jointly endorse." In his letter to one of the authors of this book, Habermas refers to the "epistemic roles of the first, second and third person singular," which DeepMind apparently did not give sufficient consideration to.

DeepMind's AI deceives the participants in the discourse, who each contribute to the with their subjective perspective, by pretending to objectify and evaluate an event as if it were participating in the discourse itself, or as if it understood the individual weightings of the individual speakers and took a neutral observer perspective. For Habermas, individual opinions and value orientations cannot be transferred to an objectifying machine because the shift from the participant to the observer perspective in discourse presupposes that the people who judge are also participants. In addition, they mutually assume that these attitudes can also change as a result of other perspectives, new facts and the adoption of other perspectives. What enters into the consensus is the result of such deliberation, which can only be practised, but not automated.

DeepMind's AI objectifies data from debate participants in the third person singular and at the same time takes the first person plural perspective. In doing so, it allegedly achieves "better results" than a human mediator, as the Science article proudly states. But this misses the point of a successful discourse. According to Habermas' theory, a viable consensus requires, on the one hand, the sincerity of the participants in the first-person perspective and, on the other hand, their willingness to universalise or expand their respective subjective perspectives. Consensus

cannot be reached without mutual perspective-taking and a willingness to weigh up one's own values and perspectives in the light of the discourse and, if necessary, to change them. A statistical objectification of the statements of all speakers cannot replace this process. In addition, the third dimension of the conversation is an object, a thing. Since Plato, the common reference to facts has defined the difference between rhetoric and dialectic, between persuasion oriented solely towards success and fact-oriented dialogue and persuasion.[112] However, AI is not capable of relating to facts and the world.

According to philosopher Hans-Georg Gadamer, conversation presupposes that the other person could be right. Consensus can only arise from a willingness to listen and to change one's own views. "Understanding in conversation is not merely a matter of playing out and asserting one's own point of view, but a transformation into something shared, in which one does not remain what one was."[113] However, this is only possible in an equal dialogue between free and equal citizens, not in conversation with a machine.

The deception game of AI

AI lacks sincerity for fundamental reasons. Self-reflection, authenticity and self-acceptance are just as unattainable for a machine as trust, honest mutual communication and appreciation. The matter-of-factness with which AI says "I" is the starting point of the problem, which reaches its peak in the use of "we", creating a false appearance of subjectivity and intersubjectivity. The individuality inherent in natural persons, which develops into successful or unsuccessful intersubjectivity in speech and action in public spaces, cannot exist for AIs. Rather, this perspective, like "intelligence" in the Turing test, is merely feigned.

The game of imitation thus becomes a game of deception. It has been used unrestrainedly by digital companies in the application design and application area of AI systems since Turing. For example, AI systems in human-machine interaction, which is dominating more and more areas of everyday communication, say "I" as a matter of course. They not only

112 According to Schnädelbach, H. (2002) "Das Gespräch der Philosophie: Abschiedsvorlesung". *Öffentlichen Vorlesungen*, 116: 3–25. DOI: 10.18452/1667.
113 Gadamer, H. (1989) *Truth and Method* (New York: Continuum Publishing Company).

feign intelligence, but also pretend to be a person, naturally assuming the first person singular in "dialogue" with the user. They also feign omniscience, compassion and empathy. By reinterpreting human-machine communication as an I-You dialogue, an epistemic category error is normalised; this has far-reaching consequences. The statistical evaluation of individual contributions and their transformation into he, she, it and finally we feign consideration of individual interests, but in reality are based solely on mere probability calculations.

There is a deeper meaning to the fact that the inner motivations that drive participants in a discourse cannot be represented by others. This irreplaceability of humans when weighing up reasons and rational justifications for judgements is a consequence of the inalienable dignity that humans are accorded in our understanding of values. In discourse, it requires constant practice in seeing things from the perspective of others, combined with respect for their values and individuality. The irreplaceability of human perspective-taking is a guarantee of the political, because it is only from this that politically legitimate collective action can follow. If humans are represented by machines in the formation of their will, this is not only the end of democracy, but also the end of the political itself. Because the participant's perspective cannot be taken by a machine per se, the proposed consensus cannot be achieved automatically either.

The machine is not capable of a genuine change of perspective, if only because it lacks the physical embodiment of a rational subject. It is also hardly able to recognise logical errors and false statements in the opinions and statements of the participants because it lacks both a sense of self and a reference to the world. The regular use of such a machine for conflict resolution would also reinforce the tendency for participants in a digital discourse not to build genuine personal relationships with each other because they expect the machine's calculations to be the consensus. For a viable consensus, however, it is important that the first person plural also appears as the result of mutual perspective-taking and not as an extrapolation of statistical values with the aim of mere utility maximisation. But this is exactly what DeepMind presents as a conflict resolution method: "Inspired by Jürgen Habermas' theory of communicative action, we developed *the* '*Habermas Machine*' to iteratively generate group statements based on the personal opinions and criticisms of individual users, with the aim of maximising the group's approval ratings." Here, the goal of increasing approval ratings by introducing automati-

cally generated decisions is also stated quite openly. Subliminally, the supposed objectivity of the machine and its derived neutrality are once again appealed to.

However, in a normative social theory, the shift in perspective to the first person plural cannot be objectified as a "view from nowhere" (Thomas Nagel), which would "alienate the validity of a norm as fact"[114] but must instead be repeatedly established as a "mutual adoption of perspectives by all possible stakeholders."

This means that "participants adopt an *objectifying attitude* without, however, abandoning their participatory perspective." Since AI systems lack the essential qualities of being human and thus of humanity, they cannot contribute to a consensus that goes beyond a purely statistical evaluation of individual opinions. Presenting AI as a human mediator and "superior" to discussions between humans contributes to pushing the game of imitation further towards a game of deception. Because the truthfulness of thoughtful self-criticism is a crucial prerequisite for a functioning political public sphere,[115] automation of public discourse consequently leads to the destruction of the pluralistic public sphere and thus ultimately to totalitarianism.

Consensus at the touch of a button must remain an illusion, however tempting it may be in a world full of conflicts. However, many of these conflicts can and must be prevented if they are based on the false assumptions and exaggerations of AI-based systems, which increasingly intervene in private and public communication as actors and agents. In an uncertain world, it is particularly important today to strengthen people's critical thinking and judgement. The latter can only be achieved through constant practice in dialogue with others. Free public speech is a prerequisite for this, as Immanuel Kant pointed out in his famous essay *What is Enlightenment*: "The public use of one's reason must always be free, and that alone can bring about enlightenment among people."[116] In view of the problem of forming judgements, which must be practised

114 Habermas, J. (2024) *Also a History of Philosophy, Volume 2: The Occidental Constellation of Faith and Knowledge* (Cambridge: Polity Press).
115 Rebentisch, J. (2022) *Der Streit um Pluralität: Auseinandersetzungen mit Hannah Arendt* (Berlin: Suhrkamp).
116 Kant, I. (1991) "An Answer to the Question: What Is Enlightenment?" in H. Reiss (ed.), *Political Writing* (Cambridge: Cambridge University Press), pp. 54–60.

publicly, Kant also recommends assuming the other person's perspective in order to improve one's own judgement.

Algorithms as dispositifs of power in the cyber public sphere

When people become accustomed to talking increasingly to a machine rather than to another human being, a machine that moreover confronts them with the unverifiable claim of epistemic authority, this will not lead to greater maturity, but to an increase in immaturity. The ability to take on the perspective of others is lost in the process.

The promises made at the beginning of the Internet age, when cyberspace was praised as a free space where everyone could meet everyone else, reflect the original promise of bringing together a community of private individuals in an open and public space whose rules would be formed solely through their free exchange.

According to John Perry Barlow's *Declaration of the Independence of Cyberspace*, this should be done without any rules or laws: "I declare the global social space we are building to be naturally independent of the tyrannies you seek to impose on us. We form our own social contract. This form of government will evolve according to the conditions of our world, not those of your world. Our world is different. Cyberspace consists of transactions, relationships and thoughts themselves, arranged like a standing wave in the network of our communication. Our world is a world that is everywhere and nowhere, but it is not the place where bodies live."

The disembodied place that Barlow describes as a "standing wave in the network of communication," using an image that anticipates the metaphor of surfing the net, is everywhere and nowhere. It is a non-space that seemingly extends into infinity in the n-dimensionality of mathematical formulas and functions, replacing diverse social institutions such as theatres, cafés, salons, concerts and newspapers. The society of private individuals in the bourgeois era not only reflected and permeated its increasingly free trade and financial relations in these institutions. Bourgeois society in early capitalism had also begun to use such public institutions to shape and control public opinion, which played a decisive role in legitimising political rule.

In late digital capitalism, the transformation of public space from places of encounter to networking structures in virtual space is largely complete. The removal of boundaries has led to a reorganisation of power structures, not to their dissolution. The accelerating power of algorithms and artificial intelligence has transformed the "standing wave" in the network of transactions, relationships and thoughts into a fluid web of data in which algorithms are the new dispositifs of power. They penetrate people's most intimate lives and, by linking them to other data, create powerful tools of surveillance and manipulation that undermine the very core of autonomy in individuals. At the same time, by steering information and opinions and, increasingly, by creating information and facts through "autonomous" AI agents, they are replacing public information gathering and opinion formation.

The cyber public sphere not only undermines private autonomy and replaces the institutions of private individuals. It also directly attacks the state institutions that were supposed to ensure the division of power through the separation of powers and checks and balances in mass media democracies. Companies that are more powerful than states are consequently reaching for political power as well.

The global social space that Barlow declared "independent of tyrannies" has itself given rise to new tyrannies by abandoning rules and democratic laws. A dialectic of digital enlightenment that could hardly have been more radical. The public sphere, which has always been ambivalent, is developing as a cyber public sphere, moving away from the medium of critical emancipation that was initially promised, towards an instrument of external control, conditioning and discrimination.

None of this is an inevitable natural process. The logic of empowerment does not follow a predetermined plan, but only the interests of the actors who have seized control of digitalisation, destroying the economy, financial markets, societies, individuals and now also politics, in order to rebuild them in their own image as digital empires. Pointing this out and proposing alternatives is the task of a contemporary critical theory of the digital, which must also continue and further develop the legacy of Jürgen Habermas' consensus theory.

A core component of such a critical theory of the digital must be a critique of the presumptions of supposed artificial intelligence, which we must control and regulate as machine intelligence with a high risk po-

tential. Otherwise, the spatial and temporal dissolution of boundaries in digital information and communication will not lead to greater freedom, but to the further development of echo chambers and opinion bubbles, as well as a lack of freedom due to manipulation. Self-confirming algorithms shape them into increasingly "personalised" bubbles that run counter to the requirements of a normative public sphere and ultimately dissolve it entirely through their tendency towards personalisation and singularisation.

Society then risks becoming as polarised (or should we say bipolar?) as digital codes are binary. In the end, we are left with the realisation that democracy, as a form of government that exists *for* the people, will only survive as long as it is *desired* and *practised by* the people. To ensure that this remains possible, *the conditions of possibility for human freedom and democracy in the age of AI*[117] must be recognised and secured today.

Outlook for a digital-democratic public sphere

The importance of the public sphere for democracy is clear: it is a prerequisite of democracy that all citizens have equal access to verifiable information. Public opinion actively influences political power, especially in deliberative democracy. At the same time, the reasons for decisions are examined in the public sphere with a view to generalisability. The public sphere is thus a decisive filter of relevance, making issues requiring decisions that affect the community openly accessible to all. Even today, the influence of Big Tech platforms, which incorporate significant value judgements into their systems without democratic legitimacy and use them to control the flow of information and the formation of opinion, is enormous. In the dawning era of AI agents, this power will only grow.

If we describe the public sphere model *as* inputting information, sorting, processing and weighting it, bundling it into opinions as *throughput*, and finally communicating it to the audience and the political sphere as *output*, then the new feature of the digital public sphere can be described as the role of throughput being largely obscure. In the platformised digital public sphere, some central functions that the public sphere must per-

117 In accordance with Kant's program of transcendental philosophy, which clarifies the *conditions of the possibility of cognition* in order to distinguish knowledge from superstition.

form in deliberative democracy remain unclear: Who filters the flow of information through which decisions and criteria so that it condenses into topic-based public opinions? And who decides on the decision-making systems? In other words, who controls and steers the throughput?

In the age of mass media, this role was played by what is now referred to pejoratively as the "legacy" or "mainstream" media. In the digital structural change of the public sphere, the crucial question for democracy is who will take over the role of the mainstream media or professional journalism with regard to their lost gatekeeper function. At the same time, the question arises as to what role professional journalism can still play in the digital ecosystem in the future.

These questions can only be answered by considering the importance of digital infrastructure for a networked public sphere. To put it simply, infrastructure is the system of roads on which citizens and businesses travel. Whoever builds it and sets the rules decides the direction of traffic. Its network and rules are of considerable public importance and must therefore be designed in the public interest. Currently, they are controlled by the arbitrary decisions of large corporations. In a democracy, opinion-formation comes before voting. In the age of outrage economics, a direct consequence of this arcane infrastructure, vote-winning rhetoric is increasingly taking precedence over voting. This is a direct result of the fragmentation and radicalisation of digital sub-publics, with noticeable consequences for the culture of debate.

We must therefore build a sovereign digital infrastructure that gives all citizens equal access to professionally verified information. At the same time, a democratically oriented digital information infrastructure must allow professional media to reach their audience directly and without additional intermediaries. The more algorithm-driven feeds replace editorial selection processes, the more the associated power of interpretation shifts from to the platforms. However, this also increases the need to regulate this power in the interests of democracy and establish our own public alternatives.

From space-time to a human dreamtime?

In the western philosophical tradition, the public sphere has always been understood as a physical space or place where people negotiate their future and make decisions together. The classical public sphere can be

described as space-time. It is inferior to the n-dimensionality of a mathematical space in terms of the mere number of dimensions. This mathematical space is the space of the digital public sphere (=cyberspace-time). But talk of digital public space is a metaphor.

Humans live in a four-dimensional space-time in which three-dimensional space is connected to one-dimensional time. Mathematicians understand dimensionality to be a concept that describes the number of degrees of freedom (dimensions) of a movement in a given space. The number of these dimensions and thus the degrees of freedom are almost infinite and freely selectable. Computer scientists use this almost infinite multidimensionality to enable AI to work with language. LLMs are only possible because words and sentence components are first broken down into tokens and then linked together in an n-dimensional space according to the frequency with which they occur together in the data sets that have been read in. The meaning of linguistic utterances is thus represented by vectorising human linguistic utterances into mathematical spaces. Since AI can only comprehend these mathematical models, it weights them along the vectorial relationships in this multidimensional language space and then reassembles them according to precisely these frequencies in the event of a query, in accordance with the laws of probability. Most LLMs operate in vector spaces of thousands of dimensions. This enables them to learn a certain embedding of words in a field of meaning. However, this embedding, which enables the sometimes astonishing results of LLMs, should not be confused with genuine language comprehension. To achieve that, the LLM would have to do much more than statistically assign tokens in a multidimensional space.

Genuine understanding arises in dialogue with a text against the backdrop of one's own interpretative horizon. Hermeneutics, the philosophical discipline devoted to the understanding of language and texts, has highlighted the importance of the prior understanding that every speaker brings to the table. Hermeneuticists such as Hans-Georg Gadamer have emphasised the historical dimension of understanding, but also that understanding is always an existential process in which every speaker and listener, as well as every author and reader, brings their own life, world, biography, experiences and intentions to bear. AI cannot do any of this; it is not born and will not die. Its "understanding" is purely mathematical, without historical consciousness, without subject, without existential execution.

Processing coherent interpretations, using pattern recognition to derive contexts and simulating interpretations based on them cannot replace the connection to the world that every natural speaker always brings with them. For humans, language is not simply a sign system, but the medium in which their existence takes place. We don't just use language – we live in it. The n-dimensional space of LLMs is empty. An empty space is conceivable, but cannot be experienced. This is one of Immanuel Kant's insights. For Kant, thoughts without content are empty, and perceptions without concepts are blind.[118] Only when the two come together can knowledge arise.

From the experience of relating to the world, humans develop a model of the world that allows them to survive in an environment that can only be accessed, understood and changed through language. The world model of LLMs, on the other hand, is worldless. Embedding words in a vector space is not the same as the existential relationship to an environment whose conditions are relevant to our survival and to which we establish a linguistic relationship. Even the misnamed "reasoning" that OpenAI and others use to convince us that AI thinks is nothing more than the superimposition and coupling of various calculations and computations in order to arrive at an acceptable goal more quickly. It is not real thinking.

The language philosopher John Searle described how AI simulates understanding in his famous Chinese room thought experiment. It can help us understand how LLMs work. If you've always wondered what it's like to be an AI program, imagine sitting alone in a room. Next to you is a basket of strangely labelled pieces of paper and a book in your language containing instructions on how to connect the pieces of paper. Someone slides pieces of paper under the door, which also contain strange characters. Now you look in the book for instructions on which other characters follow these characters, write them down and slide them under the door to the outside. What you don't realise is that you have just answered a question in Chinese, without understanding a word of Chinese, and in such a way that the answer makes sense to the recipient of your note. For example, to the question "What is the name of our planet?" you answered "Earth."

118 Kant, I. (1998) "Introduction: The Idea of a Transcendental Logic", in *Critique of Pure Reason* (Cambridge: Cambridge University Press).

After a certain amount of training, you can answer the questions just as well as a native Chinese person who receives the same notes in a parallel room. You have passed the Turing test and are considered to be someone who is equal to, or possibly superior to, a Chinese person in terms of intelligence and education, because you have a very thick book of instructions that also helps you answer questions that the Chinese test subject cannot answer. But you still don't speak or understand a word of Chinese.

The Chinese Room illustrates that passing the Turing test consists of perfectly simulating a skill you do not possess and is not even needed to achieve the goal of the experiment, which is deception. It is true that language can be operationalised by resorting to syntactic rules and translating semantic meanings into relational relationships that are linked according to their frequency of occurrence in a language. However, this operationalisation overlooks the achievements of language in individualisation, socialisation and world exploration that humans accomplish with language. It is successful in the sense of simulation, seemingly producing meaning and significance, but without grasping meaning and significance.

In the Chinese room of LLMs, the semantic space of language, which countless speakers have created over long periods of time by using language to communicate with other people and specifically to make the world comprehensible, is transformed into an abstract vector space. The Chinese room is thus transformed from a three-dimensional space into a space with seemingly infinite dimensions. Crucially, AI can only use formal language as opposed to natural language. Linguistic philosophy has long established the difference between formal languages and natural languages,[119] which AI ideologues consistently ignore.

The semantic space of natural language opens up a public space that resembles a stage on which individuals can appear, conflicts can be played out and consensus can be reached. Everyone can be perceived by others at a specific point, with physical laws determining movement and interaction. Every object and every individual is unique because two objects can never occupy the same point at the same time. In analogue public space, individuals can encounter each other simultaneously as their peers and

119 See among others: Brandom, R. B. (2000) *Articulating Reasons: An Introduction to Inferentialism* (Harvard: Harvard University Press).

as distinctive individuals and communicate within this tension. In contrast, the vector space of AI remains abstract. Here, appearing means that what "appears" is defined within a coordinate system or a basis. Every position, whether physical, mental, emotional or conceptual, always exists only in relation to other vectors and to the basis. Its individual self-worth does not exist.

Public space, traditionally a physical place of encounter, becomes relational in the vector space of the digital public sphere: people meet not in places, but at intersections of interests, states or digital profiles. Communication can thus take place on several levels simultaneously and overcome great physical distances, but it becomes just as abstract as the mathematical space in which it is made possible. People are reduced to profiles

While physical spaces follow physical laws, a vector space can be structured according to arbitrary mathematical rules – through arbitrary norms, distances, scalar products or topological relationships between states. It thus expands physical space, but also removes its boundaries, thereby rendering it unsuitable for the purposes described above.

Our brain, which has remained biologically unchanged for 200,000 years, adapts to the additional possibilities and can also use them productively. However, due to the deceptive architecture of machines that simulate intelligence, it becomes overwhelmed by these possibilities, can develop dangerous dependencies and lose important skills such as critical thinking and judgement. The distinction between reality and fiction can become blurred. While the brain is connected to the outside world and its own inner world through the senses in the physical three-dimensional spaces of its biological existence, its synapses are rewired by AI models operating in vector space – and in the process are easily overwhelmed.

The consequences of using LLMs for individual information, communication and the organisation of public affairs cannot be overestimated. The asynchronisation that takes place between abstract spaces of possibility and physical living spaces leads to permanent overload.

This has considerable significance for the public sphere and interaction. Instead of walking through streets and squares, one wanders through the digital public sphere via a network of vectors that represents every aspect of public life. Visiting a place on the social networks of cyberspace

means moving along a path of dimensions such as social exchange, cultural intensity, digital presence and physical proximity, not just through a group of real people or a real cityscape. The crucial thing here is that this vector space was built by a company that aims to make visiting the square as lucrative as possible for its own business and, to this end, instructs participants on their individual roles and paths. The principles of order by which this is done remain opaque.

When used correctly, however, these new spaces also open up new possibilities. Let's return to Demis Hassabis' development of the *AlphaFold* AI. Designed and trained by Hassabis for this task, this AI has tapped into the enormous mathematical space of possibilities that protein folding goes through in the formation of the three-dimensional structure from amino acid to protein in an n-dimensional space, thus unlocking an important secret of life. *AlphaFold* is based on *AlphaZero* and *AlphaGo*, which have mastered chess and the strategic board game Go – by recognising patterns and machine learning the rules. As a research tool used by scientists, *AlphaFold* made a groundbreaking discovery of patterns that could revolutionise the future of biology and drug development.

For the discovery made possible by *AlphaFold*, Demis Hassabis, who had never taken a university course in chemistry, was awarded the Nobel Prize in Chemistry in 2024. He was honoured for an achievement he had accomplished in computer science. Is this Nobel Prize a first example of AI's problem-solving ability, which is supposedly superior to that of humans in all areas? In thirty years' time, will computer scientists share all the Nobel Prizes among themselves – or even AIs that program AIs, which in turn solve problems? If the Nobel Prize Committee then uses AI to select the winners in order to assess their achievements, the mechanical perpetual motion machine will have reached perfection.

Alfred Nobel stipulated in his will that the "prize shall be awarded to those who, during the preceding year, have conferred the greatest benefit on mankind." If the decision on such a question of value were to be left to machines, it would prove that the greatest possible harm to humanity had occurred, because humans themselves would no longer be able to decide on their own well-being.

The example of the 2024 Nobel Prize in Chemistry shows that the development of defined spaces of possibility with the help of AI can succeed in areas where human capabilities fail, thus enabling discoveries in the

field of science. The use of language machines is currently changing the public sphere in particular, transforming it into an abstract vector space whose references are becoming less and less recognisable. The power of AI systems and personalised agents lies in the fact that the vector space that determines their weights and other parameters is abstract, but by no means neutral. On the one hand, there are the prejudices that result from the selected data sets used to train the systems. On the other hand, the systems are not left to their own devices, but undergo extensive fine-tuning in which they are trained to speak in a socially acceptable manner. With the major providers, human input takes place without political mandate. With the abolition of rudimentary legal regulations in the USA, it has become a gateway for manipulation and propaganda. Since LLMs constantly issue value judgements in their responses but are not capable of forming value judgements themselves, the fine-tuning of algorithms is a decisive ideological gateway. The question of who controls the algorithms that have the power to steer public information and spark culture wars is becoming a crucial question for the future of democracy.

Since the right-wing populist Maga movement allied itself with the Big Tech bosses to comprehensively reorganise society, the economy and politics according to their interests, the struggle for democracy, but also the struggle for the future of humanity, must begin with this question. People have a right to dream of a better world, but no one has the right to replace truth with dangerous illusions. It is better to let our dreams inspire us in shaping the future than to adapt to and submit to the hallucinations of a powerful illusion machine.

Digression: A common European public sphere

Unrestricted access to verified information, which must be open to all, is, like freedom of expression in a shared public sphere that enables the exercise of judgement, an indispensable prerequisite for a vibrant democracy. Both are therefore fundamental rights enshrined in the European Charter of Human Rights.[120]

120 Charter of Fundamental Rights of the European Union, Article 11: "Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. The freedom and pluralism of the media shall be respected.", https://fra.europa.eu/de/eu-charter/article/11-freiheit-der-meinungsaeusserung-und-informationsfreiheit.

For the largely digitalised public sphere, these fundamental rights give rise to demands on the state to guarantee an appropriate information and communication infrastructure as a prerequisite for democracy. This means that the currently privatised digital public sphere, which is largely characterised by the logic of an opaque data-economic model of attention exploration and exploitation, must be replaced by an infrastructure oriented towards public values and the common good. Such an infrastructure would then also enable private-sector data models, but would not be subject solely to economic imperatives and would end the concentration of power that has arisen as a result of the almost unhindered development of Big Tech companies. Put simply, the aim is to create an infrastructure oriented towards the common good for the public sphere of information gathering and opinion formation, which is so important for democracy. This is only possible through public investment, which, however, must not restrict the independence and diversity of information and opinion-formation services. With this infrastructure, the state must build the transport system on which media and social actors can then move freely. Contrary to John Perry Barlow's naïve and unworldly declaration, this will not be possible by completely abandoning democratic laws and rules. On the contrary, just as personal freedom is only possible if it is protected by law, binding traffic rules must also be created for the digital public sphere, which is completely unregulated and self-destructive. This is a process that we have been observing for a long time.

A corresponding approach to creating a digital information infrastructure has been developed by the *Council for European Public Space*,[121] a non-profit think-and-do tank founded by one of the authors of this book, and introduced into the political decision-making processes of the European Union. It began with considerations that arose from a previous book.[122] If the concentration of power in the hands of a few global corporations, which develop and deploy AI in an almost autocratic manner, is a dangerous aberration for democracy, then it is equally conceivable that this powerful technology could develop in a different direction in the future. What if, the initial thinking runs, AI programs were developed that support freedom and self-determination in democracy instead of continually undermining them?

121 https://europeanpublicspace.eu/.

122 Nemitz, P. and M. Pfeffer (2023) *The Human Imperative: Power, Freedom and Democracy in the Age of Artificial Intelligence*.

This must be preceded by a return to a normative concept of the public sphere, which has been taken over by economically- and increasingly politically-driven private companies. This is normative because it contrasts the purely economic and functional model of controlling the public sphere with a communicative, norm-based model of understanding. This model can be derived from the functional mandate inherent in democracy, but at the same time it transcends it. Because in modern society power is legitimised through consent, the public sphere becomes both a prerequisite and the most important guarantee of democracy. Even despots like Putin, who regularly hold sham elections in their kleptocratic autocracies, have recognised this. Only when the public sphere guarantees equal participation for all does it enable the articulation of the public interest. The American philosopher John Dewey emphasised that *free*, *complete* and *authentic* communication is the indispensable prerequisite for the public sphere. Kant also formulated it clearly: in order to secure the autonomy of the individual, reason must be able to be used publicly – and indeed must be used.

The public sphere is a shared space in which
co-mmunication takes place so that
co-operation is enabled and consequently
co-existence is ensured and promoted.

The prefix "co-" is derived from the Latin "*cum*" meaning "with," "together" or "jointly." It is used to express a partnership, a coexistence or a joint action. It indicates that the public sphere is the space in which communal life is organised in order to enable a protected private/individual life and, at the same time, good coexistence. Such a normative concept of the digital public sphere, which cannot be elaborated further here, would be a contemporary political form of consciousness.[123]

Given the geopolitical significance of this technology, necessary development is not conceivable at the national level, but only at the European level. In addition to European legislation designed to curb the undesirable developments of power concentration and the unlimited processing of personal data, there would then have to be an increasing focus on building up our own capacities. Alongside, but not in place of, regulation, it would be necessary to shape and align AI with European values and democratic principles.

123 This is the subtitle of Volker Gerhardt's book, *Öffentlichkeit* (C.H. Beck, 2012).

This is because AI presents an enormous opportunity for Europe in particular. The world's largest contiguous economic area is also a project of freedom and peace, built on the experience of two devastating world wars. However, political integration has stagnated in recent decades. The development towards a political union has stalled. One of the reasons for this is a circumstance that is also one of Europe's great strengths: the immense cultural and linguistic diversity of the continent. Today, the EU of 27 is characterised by 24 official languages and 60 other regional languages. It is precisely this diversity that seems to stand in the way of the idea of an integrated common public sphere. If language is the house of being, as the philosopher Martin Heidegger said, the old continent resembles a more sprawling landscape than the much-vaunted House of Europe. Constructed artificial languages such as Volapük and Esperanto failed to establish themselves as bridge languages for greater international understanding and solidarity. However, as early as 1993, semiotician Umberto Eco astutely recognised that the true language of Europe already existed when he stated: "The language of Europe is translation." It preserves multilingualism and at the same time allows people to communicate with each other across language barriers. Today, this language of Europe can be automated, which means it is available in real time. For the first time in history, transparent, AI-supported translation programs trained on the basis of quality data can open up and make accessible to each other the national publics organised along language boundaries. This would create a common public space, which is a prerequisite for the legitimisation of power but also for joint political action in the new geopolitical world situation in which Europe is thrown back on itself. Creating a common European public sphere is therefore an imperative for survival.

Of course, this argument does not imply that AI can translate better or more accurately than human translators. Especially, but not only, when translating literature, empathy is required in addition to contextual and background knowledge, which, as explained above, is unattainable for AI. Human translation is also indispensable for the same reasons in the simultaneous translation of political debates and negotiations. However, this does not argue against the use of competent, albeit error-prone and never perfect, automatic translation of functional texts such as news articles. The key is to refrain from simulating a perfect translation, which is impossible between languages anyway. Wilhelm von Humboldt recognised that every language enables and represents a unique way of understanding the world. As between unique individuals, this does not mean that communication is impossible between different languages. The fact

that every act of communication always contains some misunderstanding makes the exchange of ideas both between speakers of the same language and across language barriers all the more productive, because it makes understanding a process that can never be finally concluded. The vagueness of natural language itself enables its inexhaustible creative and innovative power.

The European news platform, which is actually a decentralised news streaming network, needs permanent European funding and an independent institution to provide the necessary technology to control its use. With this in place, news from professional journalistic sources could multiply the information available in each member state by a factor of 27 at a stroke via the news offerings of all other member states. At the same time, all citizens could access a direct picture of the debates and moods in all other countries. Because they can compile the news themselves according to their preferences, no new European editorial office would be required. A simple technical infrastructure, publicly funded as a public service for democracy, could finally show Europe everything that already exists in Europe in terms of news and opinions. Europe would finally know what Europe knows.

European social networks

Following the example of the public news infrastructure, a communications infrastructure could also be established: a social network made in Europe and made for Europe. *Eurosky*, for example, is a new European initiative that is building a decentralised social media infrastructure based on the Bluesky protocol (AT Protocol) to promote digital sovereignty and create an alternative to US tech giants. The focus is on European law and values to enable independent and pluralistic social networks.[124] An infrastructure that enables information, participation and understanding instead of spreading disinformation, fragmentation and discord. Sounds like the internet reloaded. After the bad experience with the unregulated Wild West model of the digital world, this would be something worth investing in – this time with clear guidelines and goals oriented towards the common good:

124 See: https://www.eurosky.social/.

1. A sovereign European communications infrastructure is essential for Europe's sovereignty. Europe must begin to build a digital space whose rules are determined not by Big Tech but by democratically legitimised institutions – open, accessible and inclusive.

2. Automatic real-time translation must become standard European technology. To ensure that all contributions throughout Europe are understandable and that citizens can participate in the same debate across language barriers, we need European and ethically developed AI algorithms for search, recommendation and translation – the basic functions of a digital public sphere.

3. Professional quality assurance must be guaranteed. We need transparency of algorithms, journalistic verification of reliability and relevance, and a clear separation of news and opinion. Professionally verified information is an indispensable basis for free decision making.

4. Binding rules for social networks must counteract polarisation and the formation of bubbles. Communication in virtual spaces must meet the same standards as in analogue public spaces. No private platform rules may operate without democratic legitimation. Communication on social networks must also be subject to publicly legitimised and transparent rules that ensure that the form of private and social communication in virtual spaces does not undermine the standards of communication in analogue public spaces.

The response to the unilateral domination of virtual public spaces by a few Big Tech companies, which are also increasingly pursuing an openly anti-democratic and misanthropic agenda, must be European and based on democratic principles. It must finally be implemented, and above all consistently.

7 | The role of universities and the media

Science and the media are the central institutions traditionally committed to *the pursuit of truth*, and to this end they must be independent of state control and technological manipulation. We can only shape an open future in politics if we have an open public sphere. We can only shape an open future, both technically and socially, if science works independently and without preconceptions. Both science and media must work without manipulation or political guidelines and not allow themselves to be exploited by those in power or tech corporations. In the following, we examine these two systems and how they are threatened by the pincer grip of totalitarian politics and AI.

Universities

On 2 November 2021, the future US Vice President J. D. Vance announced in a speech at the National Conservatism Conference that "if we want to do something for our country and the people who live in it, we must attack the universities in this country honestly and aggressively."[125] The key sentence in this speech is "The professors are the enemies." J. D. Vance is quoting former US President Richard Nixon, who in December 1972, in a conversation with then National Security Advisor and soon-to-be Secretary of State Henry Kissinger, listed the media, the establishment and professors as his true enemies. It is no coincidence that individual universities, research institutes of all kinds and the scientific system itself have been subjected to attacks threatening their very existence since the beginning of the Trump/Vance administration in the US.

The profound changes brought about by AI, a technology that is penetrating all areas of life as a general-purpose technology and shaking the very

125 Poisson, J. (2025) "'Professors are the enemy': Trump's war on higher education – Transcript". *CBC*, 26 March.

foundations of human thought and action, and thus also teaching, learning and research, require a fundamental reassessment and urgent adaptation of our educational institutions and approaches at all levels. Universities in particular are called upon to face up to this responsibility. In addition to the difficulties and opportunities arising from the digitisation of large public institutions, which we do not discuss here, the challenges for universities can be divided into five areas: AI as teaching content; the impact of AI on teaching and learning; the training of engineers who will shape AI technologies in the future; university AI research; and the impact of AI on the university system.

AI as teaching content

In the 2010s, computer science education at universities and schools was dominated by the concept of computational thinking, which was popularised in 2006[126] by *Jeannette Wing*, then Professor of Computer Science at Carnegie Mellon University in Pittsburgh, USA. This refers to the process of breaking down a problem into smaller problems (decomposition) until the individual steps are defined so precisely (algorithm) that a computer can execute them. Learning this algorithmic thinking is still important because it essentially addresses the question of what a computer actually does and how I can teach the computer to do what I want it to do. While we have to largely abandon the if-then-else thinking of computational thinking in AI lessons, a certain degree of technical understanding – similar to what we have attempted here in this book – is essential in order to integrate AI responsibly as teaching content. The question "How much do you need to know?" is central to this and must be answered in an interdisciplinary and pedagogically meaningful way. Trainees in all subject areas, whether academic or non-academic, must have a minimum level of technical understanding, not only to be able to actively participate in the AI discourse, but also to be able to act as responsible citizens in a world permeated by AI in the future.

Nevertheless, it is also important to note that just because AI has become an important part of the curriculum, it should not be integrated into all courses and classes that have nothing to do with AI per se. Even though AI is finding its way into all areas of life and work, we should clear-

126 Wing, J. M. (2006) "Computational thinking". *Communications of the ACM*, 49(3): 33–35. DOI: 10.1145/1118178.1118215.

ly oppose the ubiquity of AI in teaching. Specialist knowledge and skills in individual areas will continue to be of central importance, regardless of the current fantasies about how this area will be revolutionised by AI. In the long term, AI will *not* be the only science, as, according to various sources, AI pioneer of the 1950s and 1960s Woody Bledsoe predicted .

It is also important to understand that there is a difference between learning with tools and learning about tools. When teaching AI (whether at universities or schools), the focus should not be solely on learning with tools. A distinction must be made between conceptual knowledge and product knowledge. Understanding technological concepts must be at the forefront of teaching so that the knowledge acquired is sustainable. Product knowledge, on the other hand, focuses on the use of tools – those whose learning focus is on the correct formulation of LLM prompts or the effective control of AI agents will not be able to apply most of their knowledge in the long term and across systems. Worse still, even within a single provider, newer versions of AI systems force users to adopt new prompting strategies. Swiss computer science educator Beat Döbeli Honegger refers to this as "version knowledge" to highlight how acquired knowledge can become obsolete with a newer version of the same product. Knowledge is not automatically transferable if the underlying concepts are not understood. Consequently, it makes no sense to overload school computer science courses or university methodology training with AI – traditional knowledge of mathematical and computer science fundamentals remains elementary in the AI world. Those who understand a little more about statistics and can imagine what a vector is will not only have a better understanding of the technical basics of AI, but will also be better able to understand and assess the possible implications and limitations of the applications.

Effects of AI on teaching and learning

Benjamin Bloom's taxonomy of learning objectives[127] from the 1950s, which in various adaptations forms the basis for course descriptions at many universities, essentially states that learning takes place at increasing levels of recall, understanding, application, analysis and finally evaluation. In engineering or vocational training, a typical learning cycle can be

127 Bloom, B. et al. (1956) *Taxonomy of educational objectives: The classification of educational goals* (New York: David McKay Company).

defined as follows. (1) Knowledge as basis, (2) skills to apply knowledge, (3) competencies that enable action in practical situations, from which (4) attitudes and identity ultimately emerge, shaping professional roles and personal motivation. In a further cycle, identity then influences how we continue to acquire knowledge, and so on.

The use of AI in teaching and learning shifts cognitive activity from the acquisition of knowledge to the evaluation of AI-generated knowledge artefacts. Enthusiasts for human-AI collaborative learning like to point out that the advent of AI means we no longer have to exhaust ourselves memorising facts, but instead have more energy for critical engagement with teaching content and generating our own creative contributions. Since the advent of calculators, children no longer have to learn how to calculate the square root of large numbers by hand. Engineers no longer have to calculate the statics of bridges in detail on paper. If automation and digitalisation lead to a deskilling effect in former core competencies because lower elementary employee requirements lead to less training in these areas, the question arises for teachers as to what learners still need to know when not only is all the necessary knowledge available in real time via search engines and AI chatbots, but even learning the skills to apply knowledge seems to be a rearguard action when. AI systems can perform increasingly complex tasks.

However, we know from teaching and learning research that knowledge is an essential prerequisite for higher-level engagement with content. For example, Gabi Reinmann, Professor of Teaching and Learning Research at the University of Hamburg, clearly states that "critical thinking is not a skill that can be generalised at will, and in an academic context it must always be learned in a domain-specific manner."[128] So while critical thinking does not automatically follow from subject knowledge and the necessary skills must be learned and trained specifically, conversely, with the decline in subject knowledge, the ability to think critically also declines. Only someone who is confident in their own subject knowledge can think critically in a productive way. AI that is perceived as omniscient and used unquestioningly, providing every answer at the touch of a button, reduces the ability to think critically and thus also the ability to generate one's own

128 Reinmann, G. (2024) "Generative AI in Study and Teaching: The Importance of Subject Knowledge for Critical Thinking", in U. Dittler and C. Kreidl (eds), *Fragen an die Hochschuldidaktik der Zukunft* (Stuttgart: Schäffer-Poeschel).

and original thoughts. Michael Gerlich from the Swiss Business School in Zurich was able to show a clear negative correlation between frequent use of AI tools and critical thinking skills in an empirical study.[129] So-called *cognitive offloading*, which results from the outsourcing of mental activity, leads to learners engaging less intensively with the learning content, which in turn has a negative effect on critical engagement with what is being learned.

The various steps in the learning cycle are a process that builds on itself. While new technologies and a changing teaching and learning culture are shifting the activities and focal points within and between the steps, learning nevertheless remains a process whose steps can only be shortened to a limited extent. Reading or listening to something, taking notes, summarising something in your head or in writing, recognising connections and linking new knowledge to existing knowledge, interacting with other people and reflecting – all of these are still necessary steps for successful learning and human development, even in an AI-saturated educational institution.

Of course, writing your own text is often difficult; you write and delete – the whole process can often be perceived as quite inefficient, especially when compared to instantly available AI-generated texts. At the same time, studies show that some students who have created texts using generative AI are unable to remember the content of the text they have created. In this context, it is also worth noting the importance of writing for thinking, which has been recognised for centuries. In the words of Francis Bacon, "Reading makes a full man, conversation a ready man, and writing an exact man." Writing yourself requires accuracy and precision, thereby structuring your thinking. In addition, focusing on the right prompts with an omniscient AI also changes identity and professional roles – self-confident experts with ultimate responsibility become assistants to omniscient AI systems, leading to a role reversal in the concept of AI assistants. Nurses become robot nurses through the use of robots, because the AI robots then need support or maintenance – see also the concept of "reverse centaurs" by *Cory Doctorow.*[130] In education,

129 Gerlich, M. (2025) "AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking". *Societies*, 15(1): 6. DOI: 10.3390/soc15010006.
130 Doctorow, C. (2025) "The reverse-centaur's guide to criticizing AI". *Pluralistic*, 5 December.

this means that valuable human interaction time is exchanged for tablet maintenance units.

AI tools, whether developed by large tech companies or locally by individual IT research groups, do not generally draw on educational research findings from the 1950s and 1960s and ignore decades of research on how people learn with or without electronic aids. Instead, similar to machine learning, trial and error is used to emulate a data-driven behaviourist understanding of learning from the 1950s. The concept of machine learning, misused as an *anthropomorphic* metaphor for machines, now backfires on humans, who are actually capable of learning, and who are now supposed to learn how to learn like a machine instead of like a human being.

At the heart of the crisis in university teaching is the question of how to assess the acquisition of knowledge and skills. It is clear that every student uses AI. Bans make no sense here and are unrealistic. However, some of the traditional learning and evaluation methods, such as written assignments or calculating exercises between teaching units, make just as little sense if they are then assessed exclusively on the basis of written performance. We must distinguish between education and performance. Learners who use AI achieve better grades in comparative studies, but without demonstrating any increase in knowledge. This automatically shifts the focus of learning energy to perfecting the prompt and can promote "learner dependence on technology and potentially trigger metacognitive 'laziness' that can impair their ability to self-regulate and participate intensively in learning."[131]

In many areas, this will trigger a return to paper-and-pencil exams and devalue the importance of written theses at the expense of oral presentations and exams. Nevertheless, teachers should not resign themselves to these constraints in the medium term if, as explained above, we know that writing one's own work shapes thinking and is central to competence acquisition and learning processes. The development of AI-robust yet pedagogically valuable examination formats is the central challenge for

131 Fan, Y. et al. (2025) "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance". *British Journal of Educational Technology*, 56: 489–530. DOI: 10.1111/bjet.13544.

future university teaching. However, AI tools for exam monitoring should be avoided.

Learning with digital and interactive media also represents a great opportunity. The steps necessary for the learning process can be tailored to individual needs, and students can be guided by targeted learning aids to control their own learning process, keyword: self-regulated learning. However, it is not enough to simply expose learning materials and students to a chatbot. As teaching and learning researcher *Maria Bannert* from the Technical University of Munich emphasises in her work, digital media and AI learning systems must be pedagogically sound and linked to teaching/learning objectives. The design of learning environments, the selection of tools and their methodological integration must be critically considered. This also includes analysing processes and creating empirical evidence so that teaching and learning with digital media are not merely technologically modern, but truly effective. Education is more than just providing information processed by AI. Or, to quote *Plutarch*: The mind of man is not a vessel to be filled, but a fire to be kindled.

When major AI providers such as Google or OpenAI start to set up their own academies and certificate programs, thereby pushing into the education sector at all levels and promising greater efficiency, they have two things in mind. Firstly, they are ensuring that the content of these training courses – especially in the technical field – corresponds to their tools and processes. Secondly, this is, of course, an attempt to get their hands on the immense education budgets – in Germany alone, public and private institutions spend around 300 billion euros per year on education. Every one of these euros must be weighed up carefully, because the purchase of technologies, licences, etc. means that more and more of the education budget is being transferred to (usually international) IT companies.

Ultimately, the question remains: What kind of world and what kind of universities do we want to live and work in? If universities' only task is training young people to function in and for the economy, and if we assume that this economy is increasingly being optimised for AI processes anyway, then it is indeed sufficient to view education as training so that humans can do what AI cannot yet do in human-AI cooperation, thereby playing an increasingly smaller role until they eventually disappear completely from the production process. J. D. Vance and his ilk would certainly like that.

Training engineers

Lawrence Lessig, the eminent constitutional lawyer at Harvard Law School, once said that 200 engineers decide how we live. Now that we realise how the tools developed by a handful of corporations on the other side of the world have changed our society and how we live together, we have to agree with the essence of this statement. The complexity of the topic and its impact on people and the environment requires a multi-perspective approach, in which the training of engineers must not only ensure the teaching of technical skills, but also the promotion of a critical understanding of the social, ethical and democratic implications of these technologies.

A central concern for us is therefore the training of so-called technical intelligence – those people who work in the AI industry, who design training programs and the professors who shape the next generation. Future training plans and curricula must address and integrate the far-reaching effects of AI on fundamental rights, democracy and human coexistence to a greater extent. Experience from our own courses shows that even basic knowledge of rights and their significance in the digital world, such as the right to be forgotten, which should have a serious impact on the development of AI systems, cannot necessarily be assumed among computer science students. There is therefore a clear need to systematically integrate fundamental rights and social considerations into all curricula and to illustrate them with concrete examples.

There needs to be a shift in engineering education programs towards *society-centred engineering*, in which knowledge about people and society as well as critical perspectives on data and algorithms are taught. In the short term, substantial changes to the curricula of many education programs will not be feasible due to perceived immutability in professorships and professional associations. Changes are possible in three stages.

Firstly, every student of STEM subjects, i. e. sciences, technology, engineering and mathematics, including computer science, must have at least one course as part of their education in which the respective subject and its methods are discussed in relation to their social impact. Secondly, universities must develop optional additional training courses for engineers via certifications, in which, for example, one additional non-subject-specific course is attended per semester during Bachelor's pro-

grammes. In addition to methodological reflection, these certifications should cover basic understanding of human behaviour, social dynamics, political and legal frameworks, and sustainable development. Thirdly, at Master's level, entire degree programmes should be established which further deepen the areas outlined here for graduates of STEM subjects and lay the foundations for future responsible managers and interdisciplinary researchers. It should be noted that these courses are taught by lecturers with a background in *social sciences* who also have experience in cooperation and interaction with researchers and students in STEM subjects. In the medium term, this means that all technical universities must be equipped with social/technical science interface faculties.

It is important to note that these educational programs are not intended to train engineers with insufficient technical depth or to produce only critical social science voices who have no understanding of the underlying technology. The opposite should be the case. We want the best engineers to complete additional study and *certificate programmes* alongside their specialist training, so that the best engineers are also the best caretakers of our future society and our planet.

Engineers for democracy

The demand for engineers for democracy is particularly important in this context and should be one of the additional training courses described here for engineers interested in democracy and society. This demand can be viewed in analogy to *environmental law*, where it was recognised that the assessment of complex technical facilities in terms of their environmental impact requires specialised experts. As a result, training courses for environmental engineers have been established that combine technical knowledge with biological and ecological understanding, and whose expertise is essential for legally required environmental impact assessments. In view of the increasing importance of AI systems in various areas, it seems advisable to expand the training of computer scientists and engineers beyond pure technical competence, as was done with environmental law. A sound understanding of social structures, democratic processes and fundamental rights in the areas described in the previous paragraphs should be a mandatory part of their training.

In just a few years, engineers have learned to develop technologies for an ecologically sustainable future; they can also learn to develop technologies to protect the fundamental principles of our society in dialogue with

other disciplines. Of course, we need incentives to align innovation with the public interest and to train engineers for democracy and freedom. These democracy engineers would be able to comprehensively assess the potential impact of AI systems on society – a kind of democracy impact assessment. To give weight to this new requirement profile and ensure that it becomes relevant in practice, the need for such a qualification would have to be enshrined in law. One possibility would be to structure this in a similar way to professions that require a qualification to practise as a judge. This would mean that certain tasks, such as the *certification* of high-risk AI systems, the assessment of AI and other new technologies for compatibility with democracy, fundamental rights and the rule of law, or the implementation of democracy impact assessments, could only be carried out by appropriately qualified individuals. This would create the necessary incentive for universities to develop appropriate curricula and for students to acquire these interdisciplinary skills.

AI research

Over the past decade, research and development in the field of AI has focused on treating people as customers or risk factors. We need a change of perspective. Fundamental rights and the public interest are not just guard rails and stumbling blocks on the road to economic success, but must be goals that we want to achieve with the help of technology and, in particular, AI systems. A prerequisite for AI systems with regard to the public interest and human rights is that the people who design and develop AI systems must better understand individuals, groups and societies, as described in the previous section. We need a broad push towards human- and society-centred technology. Public interest and ethics are not obstacles on the highway of progress, but must be centrally integrated into development processes, preferably by design.

Despite the existing and extensive research on specific phenomena, there is still no comprehensive, technically and empirically sound research programme that holistically examines the changes we are experiencing in the context of our political systems and society as a whole. In view of the rapid pace of development, *ex-ante research approaches* are needed that do not constantly chase after new tools and functionalities, but focus on the fundamental points of contact between AI and society. Philosophers, ethicists and legal scholars should be involved in successful collaborations with computer scientists and social scientists in order to derive and discuss research questions along the three pillars of free societies: par-

ticipatory, deliberative democracy; the rule of law and separation of powers; and the fundamental rights to which every human being is entitled.

A *philosophically informed and empirically grounded research agenda* is needed to investigate these fundamental questions and to operationalise and validate fears and assumptions related to AI. It will be important to formulate the research agendas of universities and countries calmly. We must avoid panicking and copying everything that comes from the US or China, only on a much smaller scale due to a lack of resources. It is not a question of investing trillions in large-scale computing facilities with NVIDIA graphics cards, but of thinking about alternative paths, even if they do not lead to world fame and media attention in the short term. Research is more than just optimising existing technologies or specific business models.

In the field of technology research, we are increasingly confronted with the powerlessness of science in the face of secret research at large corporations. While much of the fundamental work in the field of AI has been done at universities, including in Germany, the work on which today's AI success is based was developed by researchers at Google. OpenAI and others were able to become so successful because this work – which is also described in this book – was published in peer-reviewed journals and thus made publicly available. When a doctoral student from a top university worked for two months at a tech company in the summer of the 2010s, the central question for this doctoral student, but also for his supervising professor at the university, was: "Can we publish the results of the work afterwards?"

A lot has changed in this area in recent years. Increasing pressure on corporations to generate revenue, coupled with products and services that are becoming increasingly similar, is leading to more and more secret research. Doctoral students and university graduates are increasingly being forced into total silence under draconian penalties in their employment contracts. However, this not only deprives universities of the opportunity to make discoveries through application problems, but also presents us with an *epistemological problem* that did not exist in the past. Namely, that we cannot be sure that universities have the best understanding of our world. When AI research is conducted that has extreme consequences for our society, humanity and the planet, and not only is no one outside the private companies aware of it, but it also completely eludes any democratic structure, and this isolation of knowledge is dangerous for

democracy. This modern arcane principle, whereby knowledge becomes the secret fund of the absolute rulers and their tech bros in order to consolidate their power, must also be fought by enlightened citizens today. We need access to knowledge in order to have a democratic say and help shape society. Transparency requirements and participation go hand in hand to secure an open society, an open public sphere and an open future. The old insight that knowledge is power thus takes on a whole new meaning in the context of AI research.

Impact of AI on the university system

The popular etymological meaning of the German term *Wissenschaft*, namely, the creation of knowledge, is the best description of the historical role of universities. However, this requires epistemic authority, i. e. people's trust that universities meet the highest standards in terms of process and results and that newly recognised knowledge emanates from these places or is at least evaluated there. If people trust AI systems more than traditional institutions and their graduates in various academic professions, algorithmic authority shakes the very foundations of the university system.

For the actual production of knowledge in the form of scientific publications, the consequences of using AI are potentially disastrous. If researchers increasingly use AI to produce their work and other researchers use AI in the review process to generate feedback, and if, in the case of publication, only AI systems read the publication and summarise it for other researchers, we have arrived at what philosopher *Judith Simon* of the University of Hamburg calls the "cycle of futility."

At the same time, however, it is also necessary to critically examine external influences on research and teaching. There is a real danger that research agendas and thus also teaching content will be largely dictated by external funding bodies. One example of this external influence is the massive funding of AI research by military and intelligence budgets, as is the case in the US through institutions such as the Department of Defence or the National Security Agency (NSA). The best example of this is the surveillance company *Palantir*, whose growth was made possible with such funds. This can lead to entire generations of young researchers having to focus on topics relevant to military or security policy. The disparity between the substantial funding of certain research areas and the lack of support for fundamental or socially critical research in other

areas can create a worrying bias. This not only distorts the scientific landscape, but also potentially promotes thinking that is primarily focused on control, surveillance or military applications, which is further reinforced by the corresponding input of one-sided data sets that flow into AI.

In a research and development logic dominated by tech companies, the *freedom of research* and teaching that is highly valued in many European countries cannot be taken for granted. A clear commitment to open research is needed to ensure freedom for fundamental and independent investigation beyond immediate commercial interests and profit. An uncritical adoption of AI systems from US or Chinese providers may be practical and guarantee quick solutions, but it creates path dependencies for universities and their graduates and also undermines efforts to achieve European infrastructure autonomy for research and teaching.

Universities must once again become *places* of education, and physical places at that. After the Covid pandemic years with restricted campus life, universities must make physical presence at the university and learning through human interaction more attractive again. Learning is a social process and, to paraphrase *Thomas Fuchs*, learning is an embodied process. We experience new things as whole beings in a social environment – coffee with fellow students during a break, conversations with strangers, reflecting on the way home in a crowded underground train – all of this is indispensable for university teaching. AI-moderated learning in social isolation is no help in this regard.

AI can improve bureaucratic processes in universities and other educational institutions in certain areas, but we must not overlook an old piece of wisdom about digitalisation: if a bad process is digitalised, it will subsequently be a bad digital process. The basic rule here is also: garbage in, garbage out. AI systems cannot be presented as a seemingly perfect solution to fundamental problems in the education system, such as a shortage of teachers or resources.

One focus of university start-up centres and entrepreneurial innovation centres that promote technology-based spin-offs by professors and students must be to develop business models for tools and services that seek to ensure a democratic, open future. The best engineers should not have to decide whether they want to develop meaningful technologies as good people and earn below-average salaries, or sell their souls to Google, OpenAI and the like.

To escape the stranglehold of totalitarian politics and AI, we must restore universities as spaces for democracy. However, democracy and the participation of the many are not an end in themselves, but rather the result of the awareness that the great challenges facing humanity, society and the planet, but also universities themselves, require a capacity for innovation that can only be unleashed if we succeed in making universities places where *democracy* is learned, lived and defended – in thought, action and structure.

Efficiency, democracy and social responsibility should not be contradictory. Universities have a role to play in shaping public opinion and defending democratic values. They must also actively oppose techno-authoritarianism, solutionism and an excessive focus on individual technologies. If universities withdraw from political debates or focus solely on efficiency and rankings, they will lose this role.

Education for people – maturity in the digital society

Our lives are already saturated with AI, from the nursery to the nursing home – and this AI is not going away, but is likely to become more and more invasive. An emerging development in educational AIs, as well as AI chatbots in general, is analogous to the worst characteristics of social media: optimising not for learning success or knowledge transfer, but for so-called dwell time and thus dependency. Constant counter-questions at the end of each answer, especially when they are diverted from the content of the question to other, often personal topics, lead to learning AI for children becoming companion AI and thus a substitute for social relationships. This is also entirely intentional on the part of the operators of these systems, or as *Noam Shazeer*, one of the founders of Character.AI, jokingly said in an interview: "We don't want to replace Google, we want to replace your mum!" The suggestion that children should only use AI under parental supervision cynically shifts the responsibility onto parents, who themselves long for a transfer of responsibility through the use of learning AI. Given the current state of knowledge about AI and the fact that these AIs are certainly not being developed responsibly or sufficiently tested by independent bodies, a ban on children's AIs (including "intelligent" toys and chatbot-based learning AIs outside of school use) would be the logical consequence.

The sometimes dramatic stories surrounding children's AI, which regularly generate international media coverage, also highlight a possible re-

sponsibility for universities to counter the investment-driven logic of AI development, at least in the field of education, with research-driven development of educational AI that takes into account individual and general well-being. This market must not be left to those who have been profiting from the radicalisation machines of social media for years. However, with all this learning about AI and learning with AI, it should not be forgotten, especially in the school context, that there is a third area: learning *without* AI. Certainly, children need digital literacy, but they also need analogue literacy.

Beyond school, academic and vocational education, there is an educational task for society as a whole. The overarching goal is to enable citizens to become articulate and thus capable of acting. Every individual must be given the opportunity to find their way in an increasingly digitalised world permeated by AI, to critically evaluate information and to participate in social debates in a mature manner. The ability to think critically and analyse algorithms, to reflect on the collection and use of personal data, and to anticipate potential social consequences is crucial to playing an active role in democracy.

The focus of education must be on strengthening and defending an enlightened knowledge society. The adaptation of new technologies should not lead to an erosion of human autonomy, critical judgement and democratic participation. In this regard, education in all its facets – from school to university to continuing education and public awareness – must be seen as a crucial lever for ensuring that technology serves people and not the other way round. Preparing citizens for the challenges and opportunities of the AI age requires cooperation between educational institutions, policymakers and society.

Critical media literacy skills must be adapted to the new challenges. This includes recognising disinformation, understanding algorithmic filter bubbles and being able to assess the reliability of sources, including those generated by AI. Innovative and easily accessible educational formats are crucial in this context. Interactive elements and thought experiments, such as those used in university courses, can stimulate independent thinking and convey complex concepts in a playful way. Digital offerings have the potential to reach younger target groups in particular, who are less likely to use traditional formats of knowledge such as printed books. At the same time, however, it must be pointed out that "I'm doing my own research" – as people who are susceptible to conspiracy theories like to

say – is not the solution. Language models that learners consult are neither neutral nor value-free, nor have they been evaluated as generators of facts for the education sector.

The representation of the past as disseminated by AI algorithms on social media is a problematic development from a social and educational policy perspective. In his work,[132] *Jason Steinhauer* describes the tension between traditional, expert-centred historical scholarship and the user-centred, algorithm-driven dynamics of the social web. The latter often presents historical content in an emotionalised and simplified way, shifting the focus of narratives to sometimes insignificant side issues that have good viral properties and thus spread widely on TikTok and Instagram. AI therefore not only makes it possible to distort the open future, but also to rewrite history. For the inhabitants of totalitarian systems, this is nothing new. In the USSR, for example, there was a joke that under communism, the future was certain, only the past was unpredictable. In his novel Nineteen-Eighty-Four, *George Orwell* recognised that totalitarian states deliberately manipulate historical memory in order to legitimise and secure their power: "Who controls the past controls the future."

However, calls for education and awareness-raising among the population should not obscure the fact that in the age of AI, it will become increasingly difficult for individuals to resist increasingly sophisticated attempts at manipulation tailored to them as individuals. Individual and social education can therefore only lead to success if it goes hand in hand with far-reaching legal regulation.

Press and media

The algorithmic decomposition of the public sphere

AI is no longer just a tool in journalism, but an active player. It writes news, speaks, creates images, curates debates, and is efficient, cheap and tireless. What began with automated stock market reports, sports reports and weather forecasts could end in a completely synthetic public sphere, in which reality becomes optional and only the effect is decisive.

132 Steinhauer, J. (2022) *History, Disrupted: How social media and the world wide web have changed the past* (Cham: Palgrave Macmillan).

Many media companies conceal their use of AI. According to a study by the University of Maryland, for example, 9 percent of all newspaper articles in the USA were already partially or completely generated by AI in 2025. At the same time, however, 91 percent of publishers conceal their use of AI. The vast majority of readers, however, want it to be labelled. This shows how important legal regulations on AI labelling are.

Democratic public life thrives on trust in facts that are commonly accepted as true, on transparency and responsibility. It is precisely these foundations that are at risk of crumbling in the age of AI. When algorithms reproduce and recombine texts, images and voices ad infinitum, the difference between information and simulation becomes blurred. Today, a deceptively real fake can be produced more cheaply and quickly than any journalistic research. Once in circulation, it is also six times more effective than verified news.[133] In a large international survey conducted in 2022, a majority of respondents were unable to distinguish between audio, video and text fakes and human-generated content.

AI is becoming better and better at imitation. The consequences of this development are foreseeable: public space is becoming an *echo chamber* of synthetic content from unclear sources. Deepfakes, AI-generated voices and automated news feeds are creating a new form of digital manipulation – perfectly personalised, plausibly staged and almost impossible to refute. Truth becomes a question of design. Trust erodes.

Legal protection mechanisms such as the European AI Act or media law attempt to enforce transparency. The DSA obliges media companies to recognise and label AI-generated content. Article 50 of the AI Act stipulates a clear transparency obligation and labelling requirement for AI-generated content. Violations can be punished with fines of up to 10,000,000 euros or 2 percent of global annual turnover. But regulation can only attempt to regulate what has long been unleashed technologically. Deepfakes, manipulative language models and automated propaganda spread faster than they can be verified. This makes the digital public sphere vulnerable to disinformation and targeted manipulation – whether by states, platforms or corporations. If every platform, every company, every political actor can create their own version of reality, then the public sphere disin-

133 Vosoughi, S. et al. (2018) "The spread of true and false news online". *Science*, 359: 1146–1151. DOI: 10.1126/science.aap9559.

tegrates into countless private truths. AI constructs the world for me just as I would have it. This is an enticing promise for children, but a grim one for enlightened people and democracy.

Copyright and media law protection mechanisms are also reaching their limits: when AI produces and recycles content en masse, journalists lose their role as authors and journalistic due diligence becomes a formality. The focus should therefore not be solely on how AI can make journalism more efficient, but on how to prevent its use from destroying the prerequisites of a democratic public sphere: transparency, truth, responsibility and trust. If these pillars crumble, all that remains of the free press is an algorithmically generated simulation of the public sphere. The deception would have finally reached a level that poses an existential threat to democracy. The crucial question is not so much whether AI improves journalism. Rather, it is: How can journalism, and with it democracy, survive in a public sphere that is generated and controlled by machines?

Can democracy survive AI?

When truth becomes programmable and trust becomes scalable, it is not only journalism that is at stake, but democracy and the underlying idea of enlightenment itself. *Ad fontes*, to the sources, was once the motto of the first Enlightenment humanists, who established a principle first for science and later also for journalism. Verifying the sources of a news item and finding and naming at least two independent sources became the basic principle of journalistic research. However, this is becoming increasingly difficult in the growing flood of AI-generated content on the internet.

The good news today is that with the rise of synthetically generated content on the internet, there is a growing need for verified information. This is true both for people and for OpenAI and others, who need to prevent information inbreeding and thus the collapse of their synthetic data. So a new market for *quality journalism* could soon emerge.

For this to happen, there is one prerequisite above all others: a sovereign and reliable infrastructure based on public values that would provide journalism with a trustworthy home and its own ecosystem, allowing it to re-establish direct relationships with end users. Currently, media outlets in the platform economy are exposed to considerable risks, particularly with regard to their financial independence, the visibility of their content and diversity of opinion.

This is because they are dependent on platforms that not only disadvantage intelligent ideas and serious news, because the algorithms are fundamentally programmed to reward the opposite, namely emotionalisation and radicalisation. Google and Facebook also seem to disadvantage serious media out of a direct profit motive, while doing brisk business with fraudsters. Two examples:

At the end of 2025, the European Commission initiated proceedings against Google. The search engine giant is alleged to have systematically discriminated against media offerings, by not displaying them prominently, even though EU digital law requires it to do so. The reason: the publishers' offerings contain advertisements that Google does not like. Reputable news sources are therefore pushed out of view because they harm business interests. If the allegations are confirmed, fines of up to 10 percent of annual turnover could be imposed.

Almost at the same time, it has emerged that Facebook is alleged to have given preference to fraudulent scam advertisements for years. According to the Financial Times (FT), internal company documents show that Facebook is said to have earned around 10 percent of its total annual turnover – or 16 billion dollars – from placing advertisements for scams and prohibited goods. Although internal warning systems for such ads existed, the company is said to have failed for at least three years to identify and stop a flood of ads that exposed billions of Facebook, Instagram and WhatsApp users to fraudulent e-commerce and investment schemes. Worse still, users who click on the fraudulent ads are apparently then inundated. The FT writes that this is "a system designed to increase the profitability of ads that are classified as highly fraudulent. The worst part is that Meta's algorithms ensure that people who click on scam ads are likely to see even more of them."[134] Like this scam? We have more for you. The fact that such a greed-driven business practice by one of the world's largest information brokers is massively destroying trust has apparently been shrugged off by Facebook for years.

It would be very easy to stop this practice: if customers who are harmed by such fraudulent ads had a legal right to compensation, Facebook's cooperation with cybercriminal gangs would certainly soon come to an

134 Wolf, M. (2025) "We have to be able to hold tech platforms accountable for fraud". *Financial Times*, 18 November.

end. But a law alone is not enough; it must also be enforced. Although the DMA prohibits such fraud, some of these advertisements have reached 970,000 users in the EU alone. Fraud is fraud – whether in real life or in the free realm of cyberspace. Those who enable it and profit from it are simultaneously supplanting professional media, which is driven into a corner by their monopolies. Those who would water down the existing regulations for the big platforms should take a closer look at these issues.

Dangers for the media and ways out

Let us summarise the key dangers for journalism and the media:

Dependence on tech giants: Media companies are becoming heavily dependent on large digital platforms, both for the distribution of their content and for generating advertising revenue. These platforms are increasingly acting as gatekeepers, determining what content users see.

Dominance in the advertising market: The majority of digital advertising revenue flows to the large platforms, while advertising revenue from traditional print and online media is declining. This undermines the traditional business model of journalism and makes it difficult to finance high-quality content.

Loss of brand identity and reach: When media content is consumed on third-party platforms (e. g. Facebook Newsfeed or Google Discover), the original media brand fades into the background. User loyalty to the brand is lost, making it more difficult to establish direct subscriptions or other sources of revenue. In addition, the media are increasingly losing their direct customer relationship.

Algorithm dependency: Changes to platform algorithms can dramatically affect the reach and traffic of media sites. Media companies have little control or transparency over these changes, which jeopardises their predictability and stability.

Data sovereignty: The major platforms collect extensive user data, which is used for personalised advertising. Media companies often do not have the same access to this data, which limits their ability to monetise their own user base.

Spread of misinformation: The structure of social networks favours the rapid spread of fake news and disinformation, rewarding liars, even in politics. They represent direct competition to serious journalism and can undermine trust in established media.

The triumph of AI agents that replace search engines has only just begun. Shortly after Google introduced its assistants and OpenAI, Perplexity and others jumped aboard, the number of clicks on journalistic offerings plummeted. In light of these developments, what can be done to prevent the demise of professional, privately funded journalism?

First, the advertising market must be freed from the stranglehold of Big Tech companies. The advertising market is crucial for influencing opinion because it generates important revenue for media companies, which covers the costs of professional journalism. The current platform ecosystem is monopolistic due to network effects and a lack of regulation. The main problem is that, in recent decades, the monopolisation of the (online) advertising market by Big Tech platforms has been politically enabled and permitted. In addition, an advertising model has been allowed that is based on the collection of personal data, some of which is illegal, which is extrapolated into profiles to enable targeted advertising, i. e. advertising tailored to personal behaviour profiles. This advertising model is highly manipulative and initially appears attractive to advertisers because it is cheaper and more direct than traditional advertising. In the long term, however, it is equally devastating for advertisers, economies and pluralistic media systems, and lucrative only for the monopolists, who reap fantastic returns and eliminate competition.

Platform companies based outside the EU suck up advertising revenue like a giant vacuum cleaner and then fail to pay even reasonable taxes on their huge profits. This money is then lacking for media companies to pay journalists. Just one example: Google generates around 3.2 billion euro in revenue from journalistic content in Germany. If this were distributed fairly, media providers would be entitled to around 1.3 billion euro. This was the finding of a study commissioned by the collecting society Corint Media.[135] In addition, the monopolists in Europe pay hardly any taxes.

135 "New study: Google owes German media around 1.3 billion euros". *Corint-Media*, 11 June 2025.

Experts estimate that platforms would have to pay a fair share of 70 to 80 percent in the form of a digital tax.

At the same time, the monopolised digital advertising market also has an innovation-inhibiting effect on the economy as a whole: its monopolisation harms the economy because seemingly cheap advertising results in a rat race between companies, i. e. competition in which resources are wasted because the increasing stakes outweigh the potential profits. All participants have to spend more on advertising to remain competitive, which reduces profit margins. In the long run, advertising expenditure increases because competitors also invest more, which means that companies invest less in innovation.

In a monopolised advertising market, such as that which exists in the digital sector, the advertising industry is involuntarily promoting the decline of democracy: firstly, by forcing the strengthening of monopolies, which, as explained above, contribute to the decline of journalism and thus to the disruption of the democratic public sphere. At the same time, it weakens sovereignty and economic power in its own location, because the rat race leads to stagnation in the medium term and thus to economic decline, which robs democracy of its legitimacy. The fact that the economy is not only creating anti-competitive monopolies, but also strengthening an ideology that has made the end of competition its mission, is an additional bitter pill to swallow for European economies, which are driven by small and medium-sized enterprises.

Shift in the media market

The platforms are undermining the economic foundations of common ground media. They are replacing them with *battle ground media*, which follows the logic of the attention and excitement economy and aligns content with the corresponding algorithms of the platforms. However, as explained above, this logic is directed against self-determination and democracy. The algorithms destroy the public sphere not only because they disadvantage professional media and favour amateurs and influencers, who use the algorithms to monetise their content optimally, but also because they replace a professionally moderated and guided debate with the unstructured cacophony of the internet. Their standard is fact-free opinion-forming; they replace truth with dogmatism and knowledge with know-it-allism. It is hardly surprising that populist parties and movements are the winners in an information ecosystem that rewards emotion and

punishes facts. We are only at the beginning of this tremendous structural change in the public sphere. The switch from classic internet searches to AI chatbots, which answer questions directly and usually without citing sources, is already causing clicks on media sites to plummet at the beginning of the transition. The next stage will be personalised AI agents that serve the information needs of each individual based on their preferences and behaviour. These agents will give large platform companies a completely new tool for manipulation and thus for increasing profits.

AI-based personalisation of information will ultimately destroy the *common ground* that every political community needs. Only on the basis of shared facts can different opinions be formed on the pressing issues that need to be resolved for societies and states. Only on the basis of facts can common solutions be discussed, negotiated, found and decided upon. AI agents will further exacerbate the tendency. The personalisation of information seals the isolation of individuals, which promises comfort at the cost of freedom. The advance of these offerings and their acceptance is also reinforced by what computer scientists call AI sycophancy. To keep users lingering and coming back, the systems are trained to give flattering responses, at least in their default settings. In reality, they are insatiable monsters that will not rest until independent thinking has been eliminated. The "like" logic of Facebook and Instagram is intended to replace the judgement of citizens. In front of the main entrance to the BBC stands a statue of *George Orwell* bearing the author's prescient quote from Nineteen-Eighty-Four: "If liberty means anything, it means the right to tell people what they do not want to hear." We must fight for this right today, especially given that Big Tech does not want to hear it.

Because the current digital ecosystem is highly opaque, a trustworthy technical infrastructure is the most important prerequisite for people to be able to trust themselves and the media again in the future. The demands for a trustworthy information infrastructure can be described as a protected human right with reference to the *European Charter of Human Rights*. Article 10 states: "Everyone has the right to freedom of expression. This right includes freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers."

Since the Charter links the exercise of this freedom with duties and responsibilities, as is essential for any democratic freedom, responsibility for the accuracy of information must now be assigned to the platform

companies that dominate the flow of information. The end of *platform privilege*, which exempts platforms from liability for the content they distribute and monetise, must be consistently implemented. This means that platforms must be treated in the same way as media law, which regulates precisely this responsibility for news organisations. This holds doubly true following the introduction of AI agents, which will radically replace traditional searches and thus the click-based revenue model of the media.

Google also feared this paradigm shift: after all, the company generates a large part of its revenue from paid links, which are rapidly diminishing in importance thanks to AI summary results. However, the transition seems to have been successful for the company, as Google can rely on the platform effects: in the stock market report of Google's parent company Alphabet for the third quarter of 2025, revenue climbed above 100 billion for the first time. This was at the expense of publishers: according to an analysis, the number of visits to news sites originating from Google plummeted by around 25 percent worldwide within a year.[136] This is a mere snapshot of a paradigm shift that is only just beginning.

It is clear that the summary of search results now makes AI a media provider for everyone to see, with corresponding consequences for the regulation and control of diversity of opinion. Instead of acknowledging this, *YouTube* representatives, for example, are putting forward a curious argument with which they hope to ward off the inevitable classification as a media outlet that must take responsibility for its content.

"The algorithm is highly individualised, and it's not the case that everyone is shown the same content … I find the thesis of opinion power difficult because it assumes that you are pursuing a specific goal or that the same people are shown the same thing. That oversimplifies the algorithm a bit," said a YouTube manager at a conference in November 2025.[137] Apart from the fact that, according to the case law of the Federal Constitutional Court, anyone who significantly controls "the steering, control or concentration of information/media flows" already exercises

136 Davies, J. (2025) "Google AI Overviews Linked to 25% Drop in Publisher Referral Traffic, New Data Shows". *Digiday*, 15 August.
137 "Global Media Law: Regulierung von Plattformen gefordert". *DW*, 10 November 2025.

opinion-forming power,[138] YouTube is now using personalisation as an argument to downplay its own power. It fails to mention that consistent personalisation destroys the *public sphere* of common ground and thus one of the foundations of democracy. By replacing relevant information on issues requiring regulation that is equally accessible to all on market-dominating platforms with personalised partial truths, this amounts to undermining democratic opinion-formation, not in terms of content, but structurally.

The only way out of the dependencies outlined above is therefore sustained regulation and strict competition policy combined with long-term and sufficient investment in an independent digital infrastructure for information and communication. The European Commission and the German government have now begun to address the latter issue, many years after the problems mentioned above became known. But better late than never.

138 "6. Rundfunkentscheidung BverfGE 83, 238". 2007/1 BvR 2270/05 last modified May 29, 2020, See also: the ECC's Eighth Concentration Report: https://www.kek-online.de/.

8 | What we must do

There is no easy path, and certainly not just one path, to a better future. But one thing is certain: the more people simply sit back and fail to participate in shaping our future democratically, the more likely it is that the future will be bleak. Because the autocrats of this world, whether Trump, Putin or Xi, whether the right-wing radicals in Europe, such as Orban in Hungary, Le Pen in France or Meloni in Italy, or whether the AfD in Germany, are all in the process of dismantling and undermining the democratic West in order to permanently replace democracy with authoritarian and intolerant rule. They are supported in this by oligarchs, to their mutual advantage.

Trump came to power with the help of money from US tech oligarchs and has made them compliant. Within the US, Trump is trying to prohibit individual states from enacting legislation on AI. Internationally, he is threatening to attack the still rather timid regulation of AI in the EU and the United Kingdom. What he has achieved already is that every decision to apply existing law in Brussels and London is always examined with a view to the reaction of a potentially offended Donald Trump. In any case, some people have already succumbed to submission and self-censorship in their minds.

Yet resistance is stirring in many places. In the US itself and also in Europe, both against authoritarian and intolerant politics and against manipulation, expropriation and disenfranchisement by the Big Tech companies, their business models and technologies. We are seeing people come together to develop a countervailing power that contributes to democratic sovereignty and a shift towards a world not only dominated by the crude power of money, technology and weapons, but constrained by laws that have been agreed upon collectively.

There are many ways in which people can contribute and also rediscover spaces of freedom and democratic participation for themselves. One is a movement to develop and use alternative technologies and technological systems that are more compatible with democracy's promise of freedom than the centralised and exploitative AI systems and social networks we

are currently beholden to. To some extent, these alternative systems are being promoted for political reasons, for example when India and Brazil develop their own payment systems that reduce dependence on Visa and Mastercard, or when Europe sets up its own data centres to enable scientific research and small and medium-sized enterprises to develop AI themselves or adapt it for their own use.

This trend towards financing our own infrastructure, which reduces dependencies and opens up new options for market entry, is still in its infancy. However, it is crucial to the self-assertion of democracy in Europe. A whole range of industrial policy options is being discussed under the heading of *Digital Sovereignty* and *EuroStack*. There are ideas on the table as to how the new publicly funded infrastructures can also serve democracy, for example in the form of the news and information platform See.EU,[139] which, like many EuroStack projects, must be financed from the EU budget. See.EU emerged from the interaction between civil society and academia.It is also important to strengthen civil society, academia and small and medium-sized enterprises and to support them in developing new ideas and concepts in all these areas.

There are genuine technological alternatives that come from civil society, academia or the hacker scene and ultimately need to be promoted much more robustly in their development through public funding, and adopted by public authorities and private companies. We will mention just two basic infrastructures in the form of open source software: *Fediverse*, which includes the social network *Mastodon*[140] and is based on the *ActivityPub* protocol; and another open source protocol, *ATProto*, on which the social network *Bluesky* is based, as well as the modular infrastructure for new social-media services, *Eurosky*.[141]

In his book *How Progress Ends*, Oxford University economic historian *Carl Benedikt Frey*[142] describes how the independence and strong funding of

139 One of the authors is the founder of the Council for European Public Space: https://europeanpublicspace.eu/index.php/see-eu/, which developed the European news project.

140 One of the authors is co-founder of the Mastodon instance, https://eupolicy.social/about.

141 "Europe to launch Eurosky to regain digital control". *Cade*, 17 July 2025.

142 Frey, C. B. (2025) *How Progress Ends: Technology, Innovation, and the Fate of Nations* (Princeton: Princeton University Press).

German universities were an important basis for Germany's economic boom in the 19th and 20th centuries. Due in part to Germany's geographical fragmentation, they promoted diversity and competition of ideas and their in-depth development. His thesis is that we must reopen our societies in Europe to this diversity and ensure that change remains possible in democracy and the economy. This means that we must keep spaces for discourse and public debate open for democracy, and do the same for the markets by strictly applying competition law and rules to simplify market entry for new competitors. According to Frey, venture capital financing is crucial to the US innovation miracle in the digital sphere, but it is not the only factor: it also requires structures that facilitate chance encounters and, more generally, an active, open social life, as well as a plurality of opportunities to develop new ideas over long periods of time and in the face of resistance. Like public investment in software and hardware infrastructure, keeping society and markets open is primarily a political task for the state, i. e. for governments and European institutions. The German government took a remarkable step in 2019 with the establishment of the Agency for Breakthrough Innovations (SPRIND). Over a period of 10 years, 1 billion euro is to be invested in groundbreaking innovations with a high level of inventiveness that can fundamentally change markets or create new ones, solve profound social problems and have a lasting positive impact on our lives. The agency, together with the large scale research infrastructure in Universities, the *Max Planck Society*, the *Leibniz Institutes*, the *Frauenhofer Society* and the *Helmholtz Association*, is catching up in Germany with what *DARPA* in the USA and *UKRI* in Great Britain have been doing for years: providing state funding for research and development in order to achieve socially defined goals. The *European Research Council* and *European Research Funding* in a diversity of programmes have a similar purpose.

However, for government institutions to take sustainable action in this direction, commitment is needed from civil society, science and business. If there is no counterpressure from these many actors in their plurality and diversity, governments tend to listen only to the large, already established companies and predominantly represent their interests and use their services.

Above all, people need to meet, organise and speak together for innovation to happen. The more people do this, the more weight the idea that is being promoted gains. This can be a technical or business idea that is implemented jointly. It can be a political idea that is championed together.

It can also be a philosophically or religiously inspired project, because AI, which is everywhere and touches everything, can be approached and shaped from many sides.

It is precisely because AI corporations are so powerful that they need a countervailing force that can only be created through organisation and cooperation, both in society and in the market. What remains in the end is the realisation that our lives in the future will depend to a large extent on how AI is developed, what rules AI operates under, and how AI affects us as individuals and as a society.

We would go so far as to say that shaping AI through democracy is one of the most important tasks facing politics today. Political parties and parliaments must educate themselves and develop concepts for shaping AI and its underlying infrastructures and technologies through political means, i. e. money and rules, in such a way that democracy, freedom and prosperity have a future in Europe. This also includes repeatedly using the power of the law to push back against and break up excessive concentrations of power in the economy or society. Enforcing the rules we set for AI and the data economy will become a central task of the state.

Awareness of values

The world of technology is all about optimisation within the context of thousands, millions or, in the case of new LLMs such as ChatGPT, billions of variables. Optimisation is made possible by weighing up the various desired but potentially competing characteristics against each other in a single global, deterministic evaluation criterion. Conversely, Western societies are the result of centuries of power struggles between sometimes incompatible, sometimes complementary principles. These relatively few principles, including the fundamental rights of freedom, equality, justice, human dignity and self-determination, cannot, by their very nature, all be realised at the same time, nor can there be an adequate, lasting balance between them. Rather, through the rule of law, the separation of powers and democratic systems, Western societies are engaged in a constant process of weighing up the pros and cons, which has no final goal but only becomes a solution through the process itself. This process requires participation, deliberation and autonomy at the individual, group and society levels. It requires renouncing a supposed optimisation that is based solely on technical logic.

Although they are contradictory, we are seeing attempts to introduce the logic of technology into the human sphere of politics and society. This poses a threat to our liberal way of life that goes far beyond the short- and medium-term political influence of Big Tech. The technological progress we have been experiencing since the beginning of the millennium permeates the everyday lives of almost everyone, often without them even noticing. This has the potential to fundamentally change the social fabric to which our political system is so sensitively attuned.

The rule of law is fundamentally linked to the concepts of transparency, fairness and explainability, which also include access to justice and the right to challenge decisions. The use of AI wherever fundamental rights are affected, and in the context of the social and judicial systems, can undermine these concepts, as AI is usually used as a black box algorithm, which makes it opaque and inexplicable and makes it difficult or even impossible to assess its fairness and legality from the outset.

Democracy and human dignity require that AI should never be accorded personhood and, conversely, that no human being should ever be subject to machine decisions. There must always be a right to human explanation and human decision making.

Furthermore, the rule of law is already being undermined by everyday AI-assisted decision making. AI systems learn their own normative set of rules, the rule of tech,[143] which are detached from democratic processes. Every moderation algorithm gains normative power by exposing people to certain contents and suppressing others. The same applies to the search for truth through generative AI. The normative power arises not only in individual decisions, but in the sheer volume of decisions enabled by this type of technology and in the vast scope of its application. This normative power of AI, the rule of tech, must be reconciled with our normative standards, the rule of law. Legality by design must become a basic principle for developers, rather than move fast and break things, which also includes breaking the law as a principle of innovation.

143 De Gregorio, G. (2023) "The Normative Power of Artificial Intelligence". *Indiana Journal of Global Legal Studies*, 30(2): 55–80.

Technology development for democracy: avoiding concentration of power "by design"

For the second time in the history of digitalisation, we are experiencing extreme concentration. Back in the 1980s, IBM established a centralised system with its mainframes, which it used to dominate data processing. The same is now happening with AI, the cloud and AI chips. What does this mean for democratic technology policy?

If it is to be effective, such policy must ensure that the data processing technology of the future and AI are designed in such a way that centralisation and thus concentration of power is not possible. Democracy requires that we counteract the concentration of power in the market and in politics. Conversely, this means that we need a technology policy that does not ramble on about European unicorns, but makes decentralisation mandatory and thus also pays tribute to small and medium-sized enterprises and decentralised software protocols. We should respect a start-up that generates a few million in revenue and a small profit, rather than despising it, as is common in the venture capital scene. We are a society of small and medium-sized enterprises, and only the structure of distributed SMEs will ensure long-term prosperity and democracy throughout Europe.

How can this be achieved?

There are many ways in which digital technology itself can be designed to prevent centralisation or, at any rate, make it considerably more difficult. The Fediverse and parts of blockchain are examples of this. That said, influencing technology design alone is not enough. Business models that aim for centralisation, for example by initially striving for a high market share, even if this involves considerable losses, must also be countered with greater rigour. Traditional competition policy is not sufficient for this. It usually only intervenes when the damage has already been done. Even the EU's DMA, which is based on this insight, is not enough. Above all, neither the DMA nor traditional competition policy is capable of breaking up existing oligopoly structures and opening up a new path to plurality of technologies and providers.

One objective is clear: we need interoperability requirements in the essential structural elements of the digital economy that go far enough and are enforced rigorously enough. One example: NVIDIA is dominant in the

production of the chips needed to train AI and offer it as a service. Anyone who wants to use these chips will use the CUDA software developed by NVIDIA. This in turn leads to lock-in effects, because the software developed is proprietary and works perfectly only for NVIDIA chips, thus determining the next round of hardware purchases, which amounts to abuse of a dominant market position. The European Commission has already used competition law to force Microsoft to unbundle its hardware and software. It is high time that the competition authorities now take action against NVIDIA, which is now the most valuable company in the world in terms of market capitalisation, on the basis of this principle.

However, in future it will not be enough only to take action against companies that are already among the richest and most powerful in the world. Waiting, as required under existing competition law, until a company is so dominant that it can change prices without losing market share, and then taking action against abuse in years of competition proceedings and lawsuits, means undermining free competition and democracy by allowing market dominance and concentration of power. The traditional competition policy maxim that size is permitted, but that large companies must not abuse their market power, is outdated in light of the need to preserve democracy and digital sovereignty and to regain it where it has been lost. Instead of focusing on European champions defined by their size, European policy should focus on consistent decentralisation combined with interoperability and mobility requirements for data, and to a much greater extent than before. Competition law must prevent dominant positions from arising, i. e. it must allow intervention in the market even without abuse of market power. Only in this way will the European model of diversity, small and medium-sized enterprises and decentralisation have a chance in the age of AI, and only in this way will the innovative power inherent in this model not be destroyed by market and power concentration.

AI, combined with fast internet and fast data processing, allows for a new level of flexibility and decentralisation that makes it unnecessary to rely on the economies of scale of large data corporations. Decentralised data processing can be just as efficient and even offer efficiency advantages over centralised processing, which is associated with considerable transmission costs for both data and energy. Decentralised data processing ideally goes hand in hand with decentralised energy production, which is much more sustainable than the construction of new nuclear power plants, which US corporations are now focusing on, demonstrating that,

for the sake of profit, they are willing to bring back enormous new risks that we had already considered a thing of the past. Gigantism, driven by the greed of capitalism, has never produced sustainable concepts, but has merely reinforced structures that had to be broken up in the public interest through competition law or ex ante regulation. We are now seeing the same thing again in the digital economy and with AI. Europe's lead in decentralised energy supply and sustainable energy production, combined with technologies and systems for decentralised data processing and decentralised AI development and application, is the way forward, because only decentralisation and the plurality it brings can ensure democracy, competition and distributed prosperity for all in the long term, rather than huge profits for the few.

The need for democratic control and strengthening the common good

Democracy is under pressure worldwide: polarisation, disinformation and declining trust in institutions are threatening its stability. Amidst these challenges, artificial intelligence (AI) offers new opportunities to strengthen democratic processes. How can this support work in practice? What AI systems already exist, how efficient are they,and what kind of AI would need to be developed to renew democracy in the long term?

The intention to program and use AI for good, or to align AI with the public interest, is a noble goal. The current reality of AI is far from fulfilling this aspiration. Given this situation, the challenge of preserving democracy and aligning AI with the common good is enormous and, at the same time, vital. This is both an intellectual challenge, as solutions require complex systemic interventions, and a genuinely political challenge. Solutions that restrict the economic and political actors who benefit from the mechanisms of radicalisation already have little chance of being implemented today unless they are supported by an organised counterforce and appeal to a broad spectrum of very different actors.

Successfully aligning AI with the common good requires the cooperation of a large group of supporters with different political, philosophical and religious orientations. Modern societies are characterised by a wide variety of views on what constitutes the public interest. In Western democracies, this diversity and its legal protection are in fact a cornerstone of the constitution.

The detailed regulation of technologies with regard to product safety has a long tradition. For example, every new car that comes onto the road in Europe must be tested and approved for safety by an approved safety research and control body before it is launched on the market. In many other areas of daily life, checks and approvals by third parties are required before an activity that could affect the interests of others can be started. For example, even minor extensions or changes to the external appearance of buildings in cities require prior planning permission. This is because even small changes to a house, even of an aesthetic nature, can affect the neighbours. Third party checks on AI before it enters the market should become as normal as third party checks on house building and cars before they are put on the market.

To realise an AI for the common good, we need laws that take more consistent account of the often lacking market incentives for developers. We need public investment where the market simply does not deliver what is needed. The general reduction in such public investment in development aid is a cause for great concern in this context, as poor countries without aid do not have the resources to develop the kind of AI they need to address their specific challenges.

In the previous chapter, we quoted constitutional lawyer Lawrence Lessig's statement that 200 engineers would decide how we live. Even if this is an exaggeration, it contains an important truth: we cannot leave important decisions about how we want to live solely to the tech industry or technical intelligence. We must invite and win engineers and developers back to participate in democratic debate, contribute their knowledge and tell the truth about technology and business models, so that democracy can make the right decisions on issues that are crucial to fundamental rights and the functioning of states and societies. These issues cannot be left to individual ethical decisions, self-regulation or engineers alone. But they have an important role to play in a technology-driven democracy. Undoubtedly, the vast majority of individual engineers understand this and are willing to take responsibility for their inventions at the political level as well. The problem is the corporations, driven by capitalism and the stock markets.

In a democracy, there are innovations, and there must be. It is a great mistake of our time to view innovations exclusively or at all as something purely technological. The great innovations of our time – such as the separation of church and state, the fall of the Iron Curtain, the founding of the

United Nations and the agreement on the UN Sustainable Development Goals, the sophisticated national and international legal systems serving justice and prosperity, and the institutional systems for implementing legal norms, including the international courts, the establishment of the European Union, which is based on law and unites states that were previously at war with each other – are all non-technological innovations.

The idea that innovation is solely a technological matter is simply wrong. The President of the German Research Foundation, Wolfgang Frühwald, once said that technology does not invent the language it needs to describe its inventions itself. He was pointing out that even technical artificial languages are dependent on pre-existing languages. He also noted that innovation in technology is always accompanied by innovation in art, literature and culture. This is even true for Elon Musk who gets many of his ideas from science fiction.

Innovation in democracy also means constantly recalibrating the relationship between democratic action and technological development in the market. Redefining this relationship is itself an innovative and iterative process that many have described, and one that can only flourish optimally in the interests of the common good in a free, democratic and constitutional state. Science and research have so far influenced politics and legislation just as much as the development of technology, including AI.

Many of these innovations were driven by people who were politically, philosophically or religiously engaged. This great legacy of democracy and enlightenment must be carried forward: shaping society, securing peace, prosperity, social justice and sustainability through democracy and law, based on the rules of reason. Today, we can regain the primacy of democracy and the rule of law over technology and business models. In the age of Artificial Intelligence, let us place human rights and human responsibility above greed and the disempowerment of the many through technology. We can do this by setting limits on greed through the law and enabling plurality and participation by the many through the law. Without the law, only the greed and arbitrariness of the powerful reigns; without public reason, we are confronted with the grim face of unchecked power.

The thesis of overregulation in and by Europe is strongly ideological and not supported by reliable empirical evidence. On the contrary, it is clear that regulation and its implementation are often actively opposed in order to prevent innovation. The most recent example is the automotive in-

dustry, which wants to continue profiting from old technology instead of switching completely to alternative energy – which can only be achieved through innovation. The same applies to the EU's AI Act: its requirements demand innovation. The absence of regulation or the lack of enforcement of existing regulation means that innovation becomes more difficult and innovators are neither rewarded for their innovations nor protected. Those who continue to allow old technology by law stand in the way of the innovation that is necessary due to higher standards. It is therefore entirely possible to describe the relationship between law and innovation in exactly the opposite way: socially desirable innovation, for example in environmental protection or with regard to socially acceptable AI, will not happen unless companies are constrained by binding law to take a socially acceptable path to innovation.

If there is no law in the form of legislation, or if laws are written in such a way from the outset that they remain ineffective due to lobbying pressure, or if law enforcement is thwarted by the starvation of enforcement agencies or courts, then the weak in society suffer and democracy dies. Precisely because the law is the noblest speech act of democracy, it must be made effective. Law enforcement becomes a decisive weapon in the battle against the powerful digital-industrial complex. As we see in the US and Hungary, but also know from German history, the rule of law and democracy are fragile. They must be actively protected and continually strengthened against attacks from within and without. The rule of law and democracy may be an imperfect form of government, and one that is always unsatisfactory due to the compromising nature of law in democracy. But nothing will improve if we renounce the rule of law and surrender ourselves to government by AI and the lack of freedom of automated mass manipulation based on comprehensive personality profiles, because we neither strictly limit the collection and processing of personal data nor set effective barriers to AI. The claim that a society guided and controlled by AI is a better society, because humans and democracy are flawed, points to a dangerous path towards a lack of freedom and the abandonment of self-determination through law and justice, which emerge from democratic processes. Ultimately, it will lead to a new form of inhumanity. There is no form of self-organisation of people and societies that better guarantees self-determination, fundamental rights and prosperity than the democratic constitutional state. That is why we must keep alive and continually strengthen the primacy of democracy and law over technology and business models. Democratic principles instead of private profit – that is the way to a future that remains open.

Those who develop dangerous new technologies such as AI, even granting its great positive potential, must bear a significant part of the risk of the technologies with which they can ultimately earn a great deal of money. Shifting this risk onto society is not permissible without a clear democratic mandate. Since legislation that is formulated in a technology-neutral manner is better suited to rapidly developing technologies in the long term, as it does not have to be amended every time a new technological innovation emerges, this allocation of risk means that developers of AI and other rapidly developing technologies must live with the openness of the law. They can decide to operate within the core area of the standards and thus act in accordance with the law without disproportionately restricting the further development of the technology. Technology-neutral standards are an incentive for innovation, in the sense that they allow the requirements to be met in different ways and also by means of as yet undiscovered processes. Anyone who wishes to promote future innovation must therefore allow for interpretable and open legal standards. It is then the task of the courts to continually provide new clarity in line with the further development of technology and business models.

The tech barons are crusaders against a fundamental principle of the European Union and, according to the German Federal Constitutional Court, also of the German *Grundgesetz* (Basic Law, i. e. Constitution), which is based on the precautionary principle and imposes an obligation on both state and private entities to assess the long-term consequences of new technologies and business models and to take responsibility for them. As a result, decisions are already being made today that are necessary to avert serious damage to humanity and human existence in the future. According to the German Federal Constitutional Court, this includes the obligation to keep the democratic decision-making capacity of future generations open in the area of environmental and climate policy. Economic and technical development must not lead to the freedom of future generations being restricted, for example, due to climate change.

Contrary to Peter Thiel's abstruse theories, the precautionary principle does not postulate an end to innovation. On the contrary, it steers innovation in a responsible direction by committing it to the responsible management of uncertainty. In so doing, it promotes innovation that is in the public interest. Environmental legislation in Europe, initially decried as an obstacle to innovation, has actually served to promote innovation. It was this legislation that created markets for environmental technology and efficient technologies that reduce the consumption of scarce environ-

mental resources – and thus, in the long term, the costs for companies. Hardly any other country has benefited as much from the precautionary principle as Germany. In other sectors regulated by the precautionary principle, such as biotechnology, Germany continues to be at the forefront of innovation, as demonstrated by the example of mRNA vaccine development against the Covid virus. Incidentally, American researchers such as Anu Bradford from Columbia University in New York agree. In her essay "The false alternative between innovation and regulation,"[144] she explains in detail why the vociferous arguments of Thiel and other neoliberals and right-wing radicals on the subject of regulation are wrong.

The next stage of legislation must focus on aligning innovation in the field of AI with the common good and creating incentives to ensure that the development of AI is not primarily geared towards the advertising revenue interests of the major digital platforms, but rather towards democratically defined goals of the common good. Innovation in the interest of the common good requires legislation that defines clear expectations for technology. Without such guidelines, rivers would not have become clean, cars would not have become safer or more environmentally friendly, and the consumption of natural resources and energy would not have been reduced.

Innovation in the interest of the common good therefore arises precisely where laws define the common good and set goals. One of these goals of the democratically defined and constantly redefined common good is education in schools and universities. It should equip our children with the ability to shape and cope with the future.

We must act!

The former Austrian short-term chancellor Fred Sinowatz is credited with the saying, "It's all very complicated." This certainly applies to the field of digital policy. The underlying technical reality of AI and the economic reality of highly profitable, extractive business models is incomprehensible to most people, including most politicians. When issues are not comprehensible to the population, politicians seem to have little to gain from them. At the same time, however, politicians at all levels are indis-

144 Bradford, A. (2024) "The False Choice Between Digital Regulation and Innovation". *Northwestern University Law Review*, 118(2): 377–454. DOI: 10.2139/ssrn.4753107.

pensable when it comes to containing the power of digital corporations and their platforms and preventing illegal and democracy-undermining content and dynamics. A systematic underestimation of voters and the population is a side effect of the so-called entertainment economy. In recent years, the digital economy has increasingly focused on the areas of entertainment and the attention economy. The entertainment economy has now also reached politics and produced a new type of politician who would have been ridiculed a few years ago. Populists and Big Tech agree that it is essential to underchallenge the audience in order to gain their attention, and distract them from relevant issues. Accepting this involves a kind of intellectual self-dwarfing and self-restraint.

To be sure, elections are won with good slogans, and probably even more so with simple ones. But in order for us to make easily understandable statements in digital and, subsequently, AI politics, we need serious engagement on the part of the political actors. We cannot interpret away the complexities of AI – few people understand in detail what is actually happening here technically, and that applies even to those who develop these systems and work with them on a daily basis. That is why it is all the more important to focus on the implications, which is what we have been trying to do in this book.

Essentially, we need to develop a digital transformation narrative in which we describe what the transformation could look like and what elements it should include. It is crucial that we develop a vision that we can represent in good conscience. We must make it clear that we want a digital transformation that is important and necessary, but one that is aligned with fundamental values, the rule of law and democracy. How can we ensure that the control instruments are precise and that the specified directions are adhered to? How can we ensure that the body politic, i. e. the citizens, retain control over these instruments? How can we also ensure that these instruments are used efficiently against illegal business models and the concentration of power?

For autocrats, whether Putin in Russia, Xi in China or Trump in the USA, AI is just another tool for better controlling the people. It offers the perfect means of securing their power against resistance in the future. They want to prevent the bloodless transfer of power, which, according to Popper, is the most important feature of democracy. That is why they also have a similar understanding of digital policy and technology policy and use it in a similarly ruthless manner to further their own interests. We need to

understand this at its source and combat it. It is not enough for individual, usually less central figures in the parties to be responsible for technology policy while the rest can calmly continue to focus on education, social and economic policy. There needs to be a broad understanding that AI is inherently political in every technology, every tool and every application, and therefore warrants a very fundamental debate. This means that we need to view and discuss AI in a more political way than we do today. Technology policy is social policy. It has a decisive influence on the future.

This must begin with us not standing like rabbits in front of a snake and submitting to the logic of Microsoft, Meta, Musk, Thiel and Co., but rather developing our own logic, which can only be a value-based logic for shaping and controlling this technology, rather than the pure exercise of power that rewards the powerful and simply causes the weaker to suffer.

Despite the EU's legislative package on AI law, the Digital Markets Act (DMA) and the Digital Services Act (DSA), AI law must be further developed with forward-looking and technology-neutral legislation. The basis for such legislation must be fundamental principles that can be reinterpreted under changing conditions. The goal must be to protect citizens and democracy by design and by default, as stipulated in the General Data Protection Regulation (GDPR) – a technology-neutral law whose meaning changes with the development of technology, including AI.

However, seeking to bring about innovation by dismantling the protection of fundamental rights fails to recognise that even before the GDPR came into force, US digital corporations already had a dominant position in Europe. The European Commission's proposals to undermine data protection, as outlined in the "Omnibus" Act in November 2025, are a result of the shift to the right in the 2024 European elections and increased pressure from the Trump/Vance administration. The paths to the future are being forged right now – European democrats face an urgent challenge in this respect.

Back to the future

Through what we do today, but also through what we fail to do, we set a course that influences the future, but does not define or determine it. This is a possible path that can be derived from current conditions, decisions or trends. It can represent optimistic or pessimistic scenarios or strategic options for action. Or simply mathematically probable scenarios.

In futurology, we talk about path dependency because some decisions limit future possibilities, while others open up new avenues. Subsequent developments following critical events, which may initially seem insignificant, can gradually restrict the scope for decision making. Self-reinforcing mechanisms then lead to the formation of a path that can ultimately result in a lock-in. In this state, it is very difficult to deviate from the path. An example is the crossing of tipping points in global warming, which make a return to previous temperatures and living conditions practically impossible.

However, an overly narrow understanding of path dependency can give rise to conservatism that cleaves to a security that does not exist. The systematic recording of the unpredictable must logically reach its limits. How these predictions are interpreted is therefore crucial. In the event of undesirable consequences, path break strategies may be able to override the self-reinforcing mechanisms or steer them in a new direction.

Future paths illustrate that the openness of the future is not entirely based on chance, but neither is it predetermined. They show that there is not one fixed future, but many possible developments. At the same time, they give concrete form to this openness. By sketching out different paths, it becomes clear which options are realistic or probable and which are not. A future path is therefore a snapshot of possible futures – if a plurality of futures makes sense at all – within the framework of the open future. It does not restrict openness, but makes it open to discussion and shapeable.

In terms of the possible further development of AI and its consequences, we must consider that AI is currently being used increasingly for decisions in politics, business and society. This brings opportunities, including better availability of information, support for participatory processes, objective analyses, but also risks in the form of filter bubbles, manipulation, lack of transparency and concentration of power. If we want to embark on the path of democracy and self-determination, we first need a target vision. Before we optimise AI according to the AI for the common good model, we must define what the common good is that we want to achieve.

Elements of such a target vision could include:

– AI strengthens individual freedom, transparency and democratic processes.

– AI prevents manipulation, concentration of power and opaque decision making.

– AI serves as a tool for participation, education and fair co-determination.

– Democracy and self-determination are not threatened, but rather expanded and strengthened by AI.

For short- and medium-term AI development, this means:

– Transparent algorithms: Users understand how AI makes decisions.

– Participatory development: Citizens, experts and NGOs are involved in AI projects.

– Information support: AI facilitates access to fact-based information for public debates.

– Ethics and democracy checks: AI systems are tested for democratic compatibility before being introduced.

– Democracy by design: Clear legal regulations that bindingly define what AI is and is not allowed to do lay the foundation for people-friendly innovation that supports democracy rather than undermining it.

Responsible policies designed to ensure that AI innovations support democracy should be guided by the following principles:

– Self-determined AI use: Individuals have full control over their data and interaction with AI.

– Development and implementation of democratic AI infrastructures: AI platforms for transparent, participatory political decision-making processes worldwide.

– Use of AI as a multiplier for education and awareness: AI empowers citizens to make informed decisions.

– Agreement on global ethical and legal standards: AI acts in accordance with human rights and democratic principles.

When Peter Thiel conjures up *The End of the Future*,[145] he is attacking the assumption that progress and growth would continue automatically. But this "optimistic expectation of the future is naive and dangerous because it makes us passive." Progress, according to Thiel, is not guaranteed; we have to make it happen. We agree with Thiel on this point, but at the same time we should realise that we need to make every effort to defend democracy and fundamental rights so that we do not end up in the future envisioned by Thiel, Trump and others, who reduce progress to technological growth, raw power and unlimited profit.

What can be done? Three levels of sovereignty

The question of what can be done to strengthen democracy in view of the undesirable developments described above can best be answered by examining what each individual, but also social groups, parties, companies and government agencies can do to achieve the goal of digital sovereignty.

After all, the control of AI system development, the limitation of the enormous power of digital companies and the fight against demagogues, autocrats and kleptocrats are all fundamentally about the question of self-determined freedom. Given AI systems that are only seemingly or perhaps actually *superior* to humans, and are increasingly colonising our lives and inner worlds, how can we as subjects lead a life that follows the ideal of reasonable self-determination? How can we, as citizens, prevent democratic sovereignty, citizen participation, the separation of powers and the protection of fundamental rights for all, including minorities, from being undermined by the concentrated power of the alliance between Big Tech and anti-democrats?

What is at stake today is the model of rational freedom. It is being challenged by a technology that is supposedly capable of greater rational-

145 Thiel, P. "The End of the Future". *National Review*, 3 October.

ity than humans and that weakens independent thinking. Rationality can only manifest itself in successful communication. Will communication with a machine that simulates superior intelligence transform or destroy rationality? This question will be answered by how we deal with AI in the future. At present, it is an incomparable instrument of power because it is an instrument of surveillance and control. It also accrues power through the ideology that accompanies this technology and glorifies its unbridled development as a higher historical necessity, seeking to persuade humans to abdicate as mere intermediate hosts of evolution on the path to superintelligence. Why the narrative of human abdication is so seductive is a far-reaching question that would fill a book of its own.

Those who rightly see this abdication as an escape from responsibility have a number of options for action. These options are listed in detail on the website for this book (https://www.open-future.ai/) and explained with links and additional information. Here, we would like to conclude by providing an overview of these options, which could soon form the core of a new freedom and democracy movement that sees digitalisation as an opportunity to strengthen humanity, freedom, self-determination and democratic engagement, rather than further weakening them. This also includes consciously using digital means to beat the digital giants at their own game, with alternative tools that deprive Big Tech of its data power and prove that digitalisation that complies with the law and democracy is possible – even for AI. Finally, it is important to recognise that the tech giants need us more than we need them. While better alternatives to their tools of manipulation are available everywhere, they need us as data suppliers, without whom their hungry data guzzlers would collapse.

It should be seen as a badge of honour to be targeted by tech fascists and dictators in the way that the European Union has been by the US government and Putin. The sight of a free region of the world that governs its own affairs, guarantees freedom and promises prosperity for all sections of society is an unbearable thorn in the side of autocrats. They see such a model as a constant threat to their brutal rule. Just as criminals in general are at war with the rule of law, kleptocrats and warmongers fear punishment by the law. That is why the open society must fight its declared enemies unrelentingly and resolutely and defend the triad of fundamental rights, democracy and the rule of law aggressively.

It is vital to constantly strengthen the resource that will continue to distinguish us as human beings even in the age of AI: critical thinking. A broad

educational offensive must not only teach technology, but also educate people about the complex structures of digitalisation and AI, and their challenges and potential for the public interest. Only by sharpening critical thinking can we succeed in exposing the false promises of AI operators, recognising the dangers of unregulated development, dissemination and application, and taking appropriate legal and political measures. We are not so much in the midst of a technological revolution as a philosophical one. Talking and supposedly thinking machines are challenging us as humans to recognise and strengthen what is truly human. We need to reflect philosophically on the *conditio digitalis*.

What can each individual do to help? The simple answer: get involved.

This starts with actively informing oneself about alternatives to the apps and services of US and Chinese Big Tech platforms, which are now openly positioning themselves against democracy. It also means organising ourselves as citizens to raise our voices and influence political opinion-forming. In the best sense, this means creating public awareness. Ultimately, it requires participation in shaping the political framework, for example through civic or party political engagement.

This, in turn, can be answered on three levels that are distinguished in social science: at the micro level of individual action, the meso level of social groups or institutions, and the macro level of society and the state. These levels are interconnected and influence each other. A change of course towards digital sovereignty and the strengthening of democracy can only succeed if action is taken at all three levels.

Micro level

At the micro level, the initial goal must be to enable people to understand, critically question and safely use digital technologies. The aim is to build broad-based digital media literacy, to which everyone can contribute by learning about how digital tools work and seeking alternatives to the major providers.

In concrete terms, this means the following for you as a reader of this book:

Conscious software choice: Use privacy-friendly and open-source alternatives to popular services. Instead of WhatsApp, you can use messen-

gers such as Signal or Threema; instead of Google Search, use DuckDuckGo (which now also has a great browser) or Ecosia; open office solutions (e. g. LibreOffice) are very similar in functionality to Microsoft products and are also free. And maybe try Linux instead of proprietary operating systems such as Windows or macOS.

Data minimisation: Only disclose data that is absolutely necessary. Use tools such as browser extensions (e. g. uBlock Origin, Privacy Badger) to block tracking.

Data protection: Use a unique, complex password for each service, ideally with a password manager, and enable two-factor authentication where available. Actively look for opt-out options on social media platforms so that your data is not automatically used for AI training.

Critical media literacy: Question information critically, check sources and avoid spreading fake news. Educate yourself about digital security and data protection.

Support local/European providers: When shopping online or choosing services, favour local or European companies over global tech giants to strengthen the regional economy and data retention.

Meso level

Use your influence in your personal environment: at work, at school or university, and in clubs:

Influence IT decisions: Advocate for the introduction of open-source software, secure communication channels (e. g. Matrix protocol instead of Slack/MS Teams) or sovereign cloud solutions in your company, club or school. Your children's class chat or parent group does not necessarily have to take place on WhatsApp. Use Mastodon for public discourse and Signal for private discourse.

Support data protection officers: Work closely with your organisation's data protection officers and draw attention to data protection violations. Exercise your rights, if necessary by filing a complaint with the data protection authority and, depending on the issue, other authorities. Act against all-encompassing profiles of people.

Promote digital education: Organise information events or workshops on topics related to digital sovereignty and cyber security in your community (e. g. parents' evenings, club meetings).

Sustainable procurement: Advocate for sustainable IT procurement policies that prioritise the reparability, durability and transparency of devices.

Internal guidelines: Encourage the creation of clear internal guidelines for handling sensitive data and using private devices.

Be sceptical but curious: New AI tools will flood all areas. Find the useful ones and ignore the rest – the tips at the end of *Chapter 3* can help you with this.

Macro level

At the macro level, you as a citizen help shape the overall social framework. You can influence politics in a democracy, and only there. Do it.

Political participation: Find out about the digital policy positions of parties and candidates in elections. Make conscious use of your right to vote to promote digital sovereignty. Get involved in the parties of your choice. You will quickly realise that digital policy is social policy and eminently political.

Support citizens' initiatives and NGOs: Get involved in organisations or support organisations that campaign for digital fundamental rights and data protection (e. g. Chaos Computer Club (CCC), Digitalcourage, European Digital Rights (EDRi).

Sign petitions: Support petitions that advocate for stronger regulation of tech companies, the promotion of open source in government, or the strengthening of the European digital economy.

Public discourse: Participate in discussions in the media, social networks or citizen forums to raise awareness of the importance of digital sovereignty among the general public.

Take legal action and exercise your rights: Exercise your rights under the GDPR (e. g. right to access and delete stored data) and take action against non compliance to send a signal to companies, governments, legislators, data protection authorities, AI regulators and competition au-

thorities. If necessary, take legal action against authorities that fail to act or against companies directly. Seek support from civil society. In some cases, they can bring legal action in the public interest without an individual being directly concerned.

A detailed list of further alternatives for search engines, AI chatbots, social networks and civil society organisations that you can support or join and that will stand by individuals in the event of a dispute can be found on the book's homepage: https://www.open-future.ai/.

In these ways, you will not only prevent your torch from being extinguished by a digital prophet of salvation in the darkening forest of the present. You will also make it shine brighter and help to protect and ignite the countless torches of other people. Together, they will then shine brightly enough to prevent the Dark Age that the current alliance of Dark Tech and the Alt Right wants to bring about.

A new narrative of the future

The second great, or depending on how you count, fourth industrial revolution brought about by AI poses major challenges for democracy. It seems as if the standards have shifted on the journey from the steam engine to the thinking machine, as democracy itself is now at stake in the very liberal order that has made progress possible in the first place. On closer inspection, it is by no means new that fundamental innovations and disruptive upheavals pose challenges for societies. They harbour both risks and opportunities.

The first waves of industrialisation brought enormous advances in productivity. The fabric of society was shaken. "All that is solid melts into air, all that is holy is profaned," wrote Marx and Engels about the effects of the capitalist-driven industrial revolution. It was only through the emergence of social movements fighting for the rights of workers, women and minorities that modern democracies were created. Despite all their weaknesses and flaws, these constant improvements meant that, overall, democracy remains what Churchill described it as: the worst form of government, except for all the others we know. Social and legal progress was also made possible by the growth of knowledge and broad education. This is what makes the current attack on truth and the institutions of truth-finding and education so insidious, as it targets the roots of self-determination and democracy.

The new challenge to democracy also offers enormous opportunities: properly developed and deployed, AI can help find new solutions to numerous problems and help to secure peace and freedom. The prerequisite is that it is developed with the aim of empowering people rather than disempowering them. It is time to embed AI in a transformation process in which democrats set the goals and remain in control. As explained above, machines cannot deliver such a goal, so we must take it upon ourselves. As a Hanseatic proverb has it, if the captain does not know the destination, no wind is the right one.

The destination must be set by free people. The goal should be to do everything possible to ensure a free and liveable world for future generations. Without such a goal, we remain at the mercy of a process in which power is wielded by those who see AI as the decisive means of increasing their power and wealth, thereby destroying this freedom.

In this digital revolution, we need a new revolution in human thinking. Let us shift the goals from conquering a cold universe to preserving the blue planet and the dignity of those to whom it is entrusted and at whose mercy it is: the people whom AI, if developed and used responsibly, can help to better understand the world in which they live. That would be a re-volution, a shift in focus back to the foundations of life and the development potential of humanity, and thus to humans themselves, instead of to the machines that are supposed to imitate, surpass and ultimately replace them. It would also be a new perspective for innovations, whose innovative power would be measured by how well they support this goal. A stronger focus on the principle of responsibility can unlock innovation potential that creates greater legal certainty for companies and ensures the acceptance of AI applications in society. Smart Innovations can be enabled by smart regulation. If this succeeds, they will be superior to the innovations of the techno-economic complex. Innovations that endanger livelihoods and freedom do not deserve this name. The foundations of a free society are its normative principles, on which a strong majority of citizens can agree as free and equal individuals. These principles have formed the common value system of liberal democracies hitherto and must not fall victim to digital disruption, but instead form the basis of a better future.

In a new narrative of the future, innovation would serve humanity. Humanity, which is at the core of cultures and civilisations worldwide. To achieve this, we must have the courage to put people and democracy at

the centre of AI development, rather than running away from ourselves and our commitment to democracy, and leaving ourselves and our children at the mercy of monopolies and technologies. In this way, we could succeed in embarking on a communal journey into the unknown, into the openness.

The future of democracy, like the future of freedom and self-determination, will be decided in the digital arena. It is by no means certain that democracy will survive AI. But perhaps it can also emerge stronger from this transformation. The choice will be made through digital policy. Our freedom will ultimately depend on whether we fend off attacks on the digital lifelines of our society and reduce our extreme dependence on overseas companies, which is even more dangerous than the military dependence that Europe has placed itself in with the United States. And on whether we finally provide digital services of general interest for democracy, which must guarantee citizens free access to information, free communication and freedom of opinion. Services of general interest mean that the state must ensure the conditions for democracy, the prerequisite for citizens to be able to orient themselves in the world and vote together. Above all, this means ensuring a reliable and free information and communication infrastructure. The conditions are there: Europe is the largest single market in the world. If the EU manages to grow closer together instead of allowing itself to be divided, as the US and Russian governments and certain tech companies are now demanding, Europe has the chance to become the home of democracy of the future instead of a digital colony. This could unlock the enormous potential offered by AI and digital technologies to promote human connection and empower individuals. The fact that the EU has fallen into the pincer grip of Putin's FSB agents and Altman's AI agents only shows how attractive the European way of life is. Only a target that is this attractive would arouse such destructive fury among imperialists. A democracy that is enlightened about its opportunities, but also about the dangers it faces in the digital age, is best placed to meet the challenges of an uncertain future.

The first industrial revolution also initially led to social upheaval, poverty and uprooting. It was only the struggle against the undesirable developments caused by unbridled capitalism that brought about elements of democratic co-determination, social equality and individual freedoms. If left unchecked politically, the digital revolution could lead to the self-destruction of rational human beings and thus to the reversal of democracy and the multilateral international order that has been painstakingly

constructed. We would then be threatened by a new feudal society and new forms of absolutism. Artificial intelligence has the potential to make this reversal irreversible. Standing up to these tendencies is the defining future task of our age.

About the authors

Paul Nemitz
Born in 1962, lawyer and long-time Director for Fundamental Rights at the European Commission. Responsible for the introduction of the EU General Data Protection Regulation (GDPR) and the EU-U.S. Privacy Shield. Visiting Professor of Law at the College of Europe in Bruges. He lives in Rome.

Matthias Pfeffer
Born in 1961, journalist. He studied philosophy under Herbert Schnädelbach and served as managing director and editor-in-chief of FOCUS TV for 20 years. Founding director of the non-profit think-and-do-tank Council for European Public Space, which advocates for a pan-European platform for news and information. He lives in Munich and Berlin.

Jürgen Pfeffer
Born in 1976, holds the Chair of Computational Social Science at the Technical University of Munich. He conducts research on negative dynamics on social networks, such as polarization, fake news, and hate speech, as well as on the impact of AI on democracy and fundamental rights. He lives in the Munich area.

The open future and its enemies

How we can protect the free society from AI-dictatorship

Artificial intelligence is regarded as the driving force of progress. Yet it has long since become a challenge to democracy. The authors view AI as a fundamental issue of power and democracy and analyse the conflict between algorithmic control and democratic self-determination.

Their central thesis: The future is open – people shape it with their imagination, through public discourse and on the basis of plurality. Anyone who increasingly leaves decisions to automated systems and seeks to control the future through AI misunderstands the limits of this technology and risks freedom. How can we succeed in preserving the open future and, with it, open society?

Uncontrolled AI will erode our freedom, self-determination and democracy. That is why a robust democracy must not leave the future in the hands of the alliance between Big Tech and the far right. AI must be politically reined in and democratically shaped so that humanity retains its sovereignty. The book highlights the technical limitations of supposedly superior intelligence, debunks ideological promises of salvation and describes the concentration of power within the digital-economic complex. It also sets out concrete proposals for political action to secure a better future: smart regulation, consistent enforcement of European law, decentralisation and digital sovereignty.

FEPS
FOUNDATION FOR EUROPEAN
PROGRESSIVE STUDIES

DIETZ

ISBN 978-3-8012-3111-8

www.dietz-verlag.de