



The Humanities and the Rise of AI

Implications of Cultural and Societal Engineering

*Johannes Pause, Christoph Purschke,
Isabell Eva Baumann, Till Dembeck & Georg Mein*

Eight Notes on Power and Algorithms

Matthias Kettner

The Ends of the Humanities - 3

Edited by Johannes Pause, Christoph Purschke, Isabell Baumann, Till Dembeck and
Georg Mein



Melusina Press, 2025

Published in **Melusina Press**, 2025

11, Porte des Sciences

L-4366 Esch-sur-Alzette

<https://www.melusinapress.lu>

Management: Niels-Oliver Walkowski, Johannes Pause

Copyediting: Carolyn Knaup, Niels-Oliver Walkowski

Cover and Layout: Valentin Henning, Erik Seitz

The digital version of this publication is freely available at

<https://www.melusinapress.lu>.

The PDF and the master copy are generated with the help of princeXML.

Print and Distribution is carried out by BoD - Books on Demand, Norderstedt

Bibliographic information of the National Library of Luxembourg: The National Library of Luxembourg lists this publication in the Luxembourg National Bibliography; detailed bibliographic data are available on the Internet at bnl.public.lu.

ISBN (Online)	978-2-919815-79-1
ISBN (PDF)	978-2-919815-80-7
ISBN (Print)	978-2-919815-81-4
DOI (Publication)	10.26298/1981-5791-enop
DOI (Volume)	10.26298/1981-5791
DOI (Series)	10.26298/2716-7550
ISSN (Online)	2716-7550

This work is licensed under CC BY-SA 4.0. Information about this license can be found at <https://creativecommons.org/licenses/by-sa/4.0/deed.de>. The images and resources contained in this work are subject to the same license unless licensed otherwise or taken from another source.



Eight Notes on Power and Algorithms

1. Digitalization as a cultural process

The momentous developments that we sweepingly refer to as *digitalization*, *algorithms*, and *artificial intelligence* are intertwined developments, but contingently so. There can be algorithms without digitalization, and digitalization without algorithms. Even artificial intelligence without digital computers is possible, e.g. organically implemented (but nonetheless artificial) intelligence. For half a century of techno-cultural history, however, we have had a contingent but highly synergistic combination of digitalization, algorithms, and artificial intelligence (AI), the latter currently consisting primarily of machine learning (ML).

Scientific inquiry into digital transformations needs an analytically powerful notion of change, and so does normative ethical inquiry. I propose a triple-perspective on digitalization: As information technology, it is a new basic technology (D1). From a media-theoretical perspective (D2), it is a global project of constructing connectivity within a computer-accessible data sphere. From a cultural-philosophical perspective (D3), it is an unbounded autotransformative cultural process.

In order to construct the rich concept of digitalization or “digital transformation” as a cultural process, let us start with the purely technological dimension of analysis.

D1: *Digitalization as a new basic technology*. Digitalization is not just another “key technology” (such as nanotechnology, genetic engineering, microelectronics), but rather a general basic technology that permeates virtually all areas of life. A suitable comparison in terms of the history of technology would be “electrification”, i.e. the utilization of electricity in all areas of life (in Europe since the end of the 19th century). If we want to take a narrow technical perspective on digitalization, the core of this basic technology today is the combination of three kinds of devices:

D1.1: Electronic calculating machines (computers) that can be used to automate mathematical calculations and thus perform them with unparalleled efficiency.

D1.2: Input devices to make information “machine-readable”. According to the state of today’s computer technology, they perform the reformatting of any other form of information (e.g. information in analogue form sound wave patterns, such as meaningful speech) into the form of information we now call “digital data”. Digital data are chains of discrete characters, in the simplest case chains of numbers in the binary system of numbers. Information in this form can be processed and stored by computing machinery. The materiality of the formal elements is arbitrary. It changes with advances in computer technology (punched cards, magnetized spots on tape, voltage states in semiconductors, quantum states), a progress that is eventually constrained only by the laws of physics.

D1.3: Output devices for translating data processing and its results from the machine-readable universal format (“digital data”) back into other desired forms and formats of information in order to make the results either accessible to other machines, i.e. “machine-readable”, or to make them “human-readable” again in forms and formats that are

meaningful to humans and can be practically used by humans, i.e. natural persons, or both humans and machines (“front-end”/“back-end”).

Let us now mark out a media-theoretical dimension.

D2: *Digitalization as a collective project of globalizing connectivity.* The new basic technology (D1) is spreading worldwide in the same way as other, e.g. economic, globalization processes. Seen in this light, we can understand digitalization in terms of media theory as a project pursued worldwide by countless diverse actors, mostly corporate actors, with two goals that are interrelated in complex ways:

D2.1: The construction of a globally extended, hyper-connected “datasphere” (cf. Bergé et al. 2018) with potentially unlimited capacity to mediate any and all machine-readable information flows. Note that some authors use the term “infosphere” instead. For instance, Luciano Floridi (1999; 2014) does not want to limit the extension of the term to machine-readable information. However, in my view, this over-inclusiveness makes Floridi’s concept of the infosphere less useful than that of a datasphere.

D2.2: The expansion of machine-controllable connectivity between all meaningful human activities (sociologically speaking, between all social practices). The well-functioning and reproduction of meaningful human activity depends on the linguistically mediated creation and interpretation of meaning generated in communities of communicative action. (To explain: The social-philosophical concept of *communicative* action as opposed to *instrumental* action was first developed by Jürgen Habermas (1984). According to Habermas’ distinction, people act communicatively when they coordinate their actions on the basis of beliefs that they assume are consensually shared with their interaction partners and that they could revise if necessary by articulating, comparing and re-evaluating their reasons, i.e., in “argumentative discourse”.)

The goal in D2.2 is to mediate the meaning of meaningful human activities with flows of machine-readable information generated for and circulating in the Data Sphere; and to include, i.e., link, all humans (present, past, and future) into the machine-controllable connectivity of the Data Sphere.

Who can control which connections, how, and with what intentions and goals? This is the key ethical and political question of power in relation to digitalization.

A further dimension of analysis is recommended if we want to take into account the fact that the so-called “digital transformation” is a cultural process and involves cultural change. We should therefore also approach digitalization from a cultural-philosophical perspective.

D3: *Digitalization as an auto-transformative cultural process.* Wherever information is processed by means of digitally realized computational operations, it will be super-fast and hyper-connected. This promises astonishing gains in effectiveness and/or efficiency for all purposes, whatever they may be. With the new basic technology (cf. D1), ‘we’ - or rather the most powerful driving actors - have collectively changed almost all cultural processes and practices in the course of the last three decades since the globalization of the Internet (cf. D2). We should therefore understand digitalization as a cultural process that manifests itself in changes in an enormously wide range of our social practices. (For example, in our practices of reading, writing, learning, teaching, assembling, making art and waging war, loving and living, in short, in all our practices of organizing and co-

ordinating everything and everyone.) Many of these cultural changes are gradual, but eventually have a dramatic result (“disruption”), much like the proverbial slowly boiled frog. Other cultural changes are the result of deliberate planning (e.g., powerful marketing strategies) of how to turn new practices into taken-for-granted cultural normality as quickly as possible (e.g., transferring normal practices of informal communication to social networks (Friese et al. 2020; Grunwald 2019; Grunwald 2024; Gramelsberger 2024) or disrupting the way we used to take photos by normalizing practices of using smartphone cameras (Weilenmann/Hillman 2019).

The practices that make up our ways of life are constantly changing and adapting to the influences, impacts, and challenges of a wide variety of changes in all kinds of processes (e.g., the vicissitudes of economic life, population dynamics, migration, warfare and terrorism, climate change). How then can we attribute certain changes in certain practices specifically to digitalization? The test is whether (or to what extent) we can explain this change as a consequence of the construction and use (cf. D2) of new means, namely digital technical means (cf. D1), perhaps better than by alternative explanations.

As a matter of fact, digital technology solutions are constantly being sought for all sorts of practical problems, large and small, and really better, or only purportedly better, digital technology solutions are being sought for existing solutions, and new practical problems are being invented, with certain digital technology applications then being promoted and marketed as the solution. Morozov (2013:1-16) calls this “solutionism”. Many ethically relevant problems of a “solutionist” mentality arise from the fact that solutionism is internal to a predominantly capitalist world culture.

Social practices, whatever artifacts, biofacts, technofacts, or pre-existing material they include in their inventory, necessarily involve people as their personnel, so to speak. Motivated human behavior, especially intentionally controllable behavior alias *human action*, becomes meaningful behavior within networks of normatively ordered social practices: here, such behavior becomes interpretable and governable in meaningful ways. All our practices, being the stuff of cultural processes, “work” directly with meaning. Meaning, whatever its symbolic form, can be articulated mostly linguistically. The original and irreplaceable media for us to make sense of things and communicatively share and revise meaning are, of course, our vernacular or “natural” languages. The medium of natural language is foundational. This holds regardless of the extent to which our “natural” sense-making mediated by natural languages may also involve non-linguistic media and their technical modes of mediation, e.g., mediation by technical media and technical infrastructures of interpersonal communication. In contrast, all information processing machines, e.g. computer systems, work exclusively with appropriately formatted data. With the help of this conceptual distinction between “sense and meaning” and “data”, the cultural-philosophical concept of digitalization D3 can be pointedly formulated:

D3: *Within the total cultural process, digitalization is a historically new technocultural activity, engendering change in the entire realm of our meaningful practices, i.e. in our life-worldly sphere of sense and meaning, through interpenetration with the global connectivity of the computer-based “data sphere”.*

Note that D1, D2 and D3 are methodologically and theoretically independent of each other, when taken as dimensions of analysis. Taken as concrete realities, D3 presupposes D2, and D2 presupposes D1.

What is particularly striking about the cultural process of digitalization (D1+D2+D3) in its current phase is, first, that it generates disruptive developments (not only within the economy); second, and seemingly contrary to this, there is a remarkably rapid familiarization and banalization of all digital-cultural changes. Algorithmically controlled applications are becoming normal elements of our technoculture in more and more areas of practice. Thirdly, amazing things are happening in the realm of collective fantasies and narratives. In a sober technical understanding, algorithms are automatable solution programs for machine-calculable goals. In the popular understanding, however, algorithms have become an object of fantasy. The popular topos of the ‘power of algorithms’ functions today in the digital revolution in much the same way as the topos of the ‘power of genes’ did thirty years ago in both the euphoria and doom-mongering in the wake of the molecular genetic revolution. The power of algorithms is a theme that runs through many plots in popular culture and is also attracting attention in journalism and cultural studies. A case in point:

“Algorithms are everywhere, supposedly. We are living in an ‘algorithmic culture’, to use the term coined by author and communication scholar Ted Striphas (2015). Google’s search algorithms determine how we access information. Facebook’s News Feed algorithms determine how we socialize. Netflix’s and Amazon’s collaborative filtering algorithms choose products and media for us.” (Bogost 2015)

We curious, inventive, technology-loving humans have learned to construct algorithmically controlled machines that are now able to realize *virtually*¹ an astonishingly wide range of animal and human intelligent performance. This raises a number of difficult theoretical questions and practical (political, ethical, pragmatic) problems that are not easy to tackle. If we continue to refine and augment virtually realized intellectual capacities, e.g. intelligence, through technological progress, will we at some point stop making a strict distinction between “virtual” ones (i.e. simulated in machines) and “real” ones (i.e. embodied in human persons)? And should we perhaps get used to, or have to get used to, attributing certain actor characteristics to suitably developed technocultural systems and machines not only in terms of intelligence (a construct invented by psychologists at the beginning of the last century) but also in many other respects, which we currently only attribute to ourselves and to our fellow human beings? In aspects such as empathy and emotionality, love of life and concern for the future, creativity, capacity for suffering - but also their negative inversions, such as callousness, destructiveness, maliciousness, etc.)? Should we want, or should we have to expect, that in the near future the cultural process of digitalization will give birth to technofacts whose agency makes them so capable of speaking, thinking, relating and acting that we will encounter such actors as if

1 ‘Virtual’ taken in Charles Sanders Peirce’s precise sense of the term: “ virtual X (where X is a common noun) is something, not an X, which has the efficiency (virtus) of an X. - This is the proper meaning of the word; but it has been seriously confounded with ‘potential’, which is almost its contrary. For the potential X is of the nature of X, but is without actual efficiency.” (Peirce 1902:763)

they were new kinds of persons? And conversely, that we will appear alien to these new kinds of persons?

If we want to think seriously about all these big and moving questions about the future, we have to start with basic questions. One of these basic questions is the question of power. It is a very necessary and very demanding task to develop a new dynamic understanding of power that allows for the analysis of power relations not only with respect to persons and quasi-personal corporate actors, but also with respect to a-personal software agents and other types of machine agency. Speaking of ‘agency’ is not per se speaking of the agency of human persons. ‘Agency’ is a technical term whose meaning can vary depending on the conceptual content one chooses to associate with the term, the only downward constraint being that the entity one finds appropriate to describe as an ‘agent’ must be such as to be able to do one thing or another on its own. The theoretical task to specify different concepts of agency for different kinds of agents seems to me to be urgent, but in my opinion we are not yet advanced enough philosophically to solve it satisfactorily. What are we talking about when we talk about the power of algorithms? Do we thereby conceive of algorithms or algorithmically operating machines (computers) as powerful agents of sorts? In preparation, I offer below a series of reflections that attempt to clarify common conceptions of power through philosophical analysis and synthesis.

2. Towards a real definition of power

The time-honored vocabulary of power formed in the European history of concepts and ideas is impressively ambiguous (especially *dynamis*, *energeia*, *potentia*, *potestas*, *dominium*, *auctoritas*). The German sociologist Max Weber (1864-1920), a teacher of the important American sociologist Talcott Parsons, made an attempt, which is now considered classical, to articulate the general form of clear cases of power. Weber’s proposal for a real definition of power is as follows:

“*Power* means every chance, within a social relationship, to enforce one’s own will even against resistance, whatever the basis for this chance may be. *Domination* should mean the chance of finding obedience for a determinate *command* among identifiable people; *discipline* should mean the chance of finding prompt, automatic and schematic obedience for a command among a definable number of people by virtue of their habitual attitudes. [...] The term ‘power’ is sociologically amorphous. All conceivable qualities of a person and all conceivable constellations can enable people to enforce their will in a given situation.” (Weber 1956: 38, translated by M.K.)

For at least eight reasons, Weber’s real definition seems to me to be an excellent one:

#1. The definition is tailored to action, namely social action among people, but is not strictly anthropocentric. Power relations beyond human beings are also conceivable.

#2. It avoids the confusion of power as the *capacity* of a power holder and the *exercise* of this capacity.

#3. By referring to the *will*, it incorporates central elements of human intentionality into the concept of power, without reducing the power-relevant intentionality to intentions to act. Also cognitive intentionality is integral to the concept of power.

#4. Conceived of, and cognized as, an “opportunity”, power is a way of using a given favorable opportunity, and thus has to be conceived dispositionally and contextually, since it depends on the circumstances of the respective situation which opportunity is given, how and for whom it is so given.

#5. Weber’s definition renders the question of the *means of power* – a question that plays the main role in many purported theories of power, while their concept of power itself remains vacuous (e.g. deVrio et al. 2024) – an unbound variable (“whatever the basis for this chance may be”). One of the implications of this is that the exercise of power does not necessarily have to be violent or involve violent means.

#6. With regard to power relations, the definition does not prejudge a necessary asymmetry between power and counter-power (subjugation, overpowering, etc.), nor even necessarily a counter-power, but only a resistance, a power-affected opposing part of exercised power. The opposing or resisting part can also be something resistant in the actors themselves that they have to overcome, e.g. weakness of will, stupidity, dysfunctional habits, disorganizedness.

#7. Conceiving of power along Weberian lines avoids the misconception so common in many sociological power theories of declaring the relation of *influencing* or *being influenced* to be the essence of power, thus trivializing the power relationship.²

#8. The definition does not prejudge the assessment of power phenomena under moral and other kinds of normative standards. It is important to note that the exercise of power does not inherently imply (though it does not preclude) a morally questionable restriction of freedom and certainly does not necessarily involve an unjust deprivation of freedom.

I would now like to take Weber’s basic formula a step further, with some conceptual engineering, in order to make it useful for the description, explanation and evaluation of forms of power and power relations within the autotransformative cultural process of digitalization (cf. D3).

Let me note from the outset that Weber’s definition of power is tailored to human beings acting and interacting in pursuit of their purposes or goals in the social practices that are accessible to them and make sense in their cultural world. In Weber’s basic formula, the term ‘will’ refers to the role of intentionality in the concept of power. With #1, it is clear that Weber has in mind practices of people who want to do something that does not happen by itself, with or against other people who want to do something (different), or in confrontation with something resistant (material, things). The question of whether, and to what extent, adequately organized non-human beings can also be said to have the ability to want to do something and to will accordingly, but could have wanted and willed something else instead, leads into the philosophical labyrinths of theories of

2 To equate the exercise of power with the exertion of influence means equating power with the creation of whatever social effect. Every exertion of influence by actors on one another then becomes an exercise of power, and society, at the macro level of its most important institutions as well as at the micro level of personal relationships and social interactions, is uniformly stylized as a phenomenon of power, because the social presupposes the mutual influence of individuals. However, a specific concept of power is then superfluous. The term ‘influence’ would do and would also have the advantage that it can be used as a noun as well as a verb (cf. Morriss 2002:8-13).

free will. If we could still hold fast to Kant's practical philosophy today, it would be easy to claim that power relations exist only by and between beings who possess what Kant called "rational will". We can, in my opinion, relieve the metaphysical burdens of Kant's conception somewhat by interpreting the meaning of Weber's basic formula through the concept of purpose and purpose-setting (although these, too, admittedly have their metaphysical pitfalls). Power (*sensu* Weber) then means a situational ability of purposeful actors to achieve something that is important to them by seizing opportunities for its realization.

This reformulation makes it clear that a great deal of cognition, evaluation, and motivation - in short, rationality - is invested in a modified but still broadly Weberian concept of power. Circumstances have to be *recognized as relevant* to the respective goals, and they have to be *evaluated* as something that would possibly (i.e., potentially) contribute to the realization of the goals so and so well. Likewise, the circumstances have to be *appreciated as suitable* circumstances. Moreover, opportunities, once identified, must be *seized*, i.e., the actors pursuing the respective goal must *decide to act* purposefully.

If all these capacities are required of an actor said to possess power in clear cases of our attribution of power to "full" actors (i.e., human persons in full possession of their intellectual and volitional capacities or powers), how far can the standard range of our core concept of power be extended without overstretching our reliable practice of attributing power to actors? The constraints described so far certainly allow for certain kinds of socially constituted entities beyond human persons, namely those to which we can ascribe enough identity and intentionality to consider them capable of action and to hold them responsible for the exercise of their powers, not only in a causal sense, but also in a normative (e.g. legal or moral) sense. Consider commercial enterprises (e.g. *Alpha*), which we hold morally and legally responsible for what they do and do not do though they ought to (e.g. curbing fakenews), or consider governments. Do the conceptual burdens of the concept of power allow its generous extension to (some) animals? And possibly also to (some) artifacts or technofacts, such as machine-based intelligent systems that we have constructed to pursue purposes and make decisions, although they are neither alive like animals nor normatively organized like corporate actors, e.g. companies? I will take up the question of sufficiently qualified agency again in section 5.

3. Power as steering change processes through control of forces

Although the pre-theoretical understanding of power undoubtedly includes the conceptual component of force or strength, it cannot be reduced to this. It would lack something important that we take for granted as soon as we want to define human agency: the aspect of control through intentions that a person has and wants to put into action.

Suppose I enter an unlit room at night and do not know that the air is enriched with explosive gas because of a defective pipe. I feel for the light switch, want to turn on the light, and there is a huge explosion (which I fortunately survive). With my behavior, I have indeed caused a very strong effect. But we wouldn't say I 'had the power' to blow up the room. The effect and the way it was brought about was beyond my control, although

no one other than me caused the effect by my actions. Leaving aside the possibility of malice, the explosion was an accident, so it was caused by the unfortunate confluence of certain circumstances of my action and by my action. If effective events are caused by an actor's intentional behavior, but occur outside the actor's sphere of control, then this is certainly not what we mean when we attribute power to an actor in this regard (even though we may hold the actor responsible, at least in a causal sense of responsibility, for the effects he happens to produce). It is obvious that we have to build a moment of control into a meaningful concept of power.

We have seen: Power, which somehow powerful actors have, cannot be conceptually reduced to their ability to cause events or produce effects. I now propose to construe the concept of power in such a way that it implies *a relationship of forces*. The claim is: Power and power relations can only exist where there is control, forces and relations of forces. What it means to have or exercise a certain power can best be explained as the mastery of forces, without us having to negate Weber's root formula.

If we relax Weber's point #1, we can think of the relevant point of power this way: power is control over change processes through mastery over forces within relations of power. I will now try to make this idea a little more precise. We can explain what it means for a powerful actor A to have power as A's ability to steer, in a counterfactually robust way, courses of change that are significant for A. To express this schematically in a concise formula:

The power that A has relative to the power of other actors A' is a function of the ability that A has to control certain forces (let's call them F+) with the help of certain other forces (let's call them F) that A can already control with such certainty that A rightly believes that A could still control F* and steer F+ if circumstances other than the current real circumstances, i.e. counterfactual circumstances, were present.*

Power, then, is a relationship that exists whenever there are certain forces that persons (or other kinds of actors with relevantly similar abilities for pursuing goals intelligently) are able to control intentionally within a certain range of circumstances (=intentionally self-controlled forces F*) such that if they wanted to then they really could control via their self-controlled forces (F*) other forces (F+) which are not, or not yet, self-controlled.

If we understand the fact *that an actor A has power P* as A's ability to steer (or govern or direct) trajectories of change in a counterfactually robust way through intentionally self-controlled forces then which kinds of powers actors have and to what extent will depend, as a matter of fact, on what forces F+ there are available to them for being brought under indirect control via A's intentionally self-controlled forces F*.

To control, in the sense of having something under control, means to be able to significantly influence or dominate something, to have decisive influence in a matter, or to exercise effective authority over certain people involved in certain practices. For example, an important part of the control specifically required for political governance is having the reproduction of belief in its legitimacy under control. Consider the political realm: Any political activity can count as a form of governance as soon as it exerts a formative influence on the course of actions and events by directing the behavior of significant actors (possibly including the governing actor himself), or keeps them in check, or deci-

sively influences the course of events by exercising authority or enforcing discipline. In fact, the semantics of the terms 'steering', 'control', 'governing' and 'directing' overlap to a large extent, although sometimes terminological composites indicating distinctions are used in the literature (controlling power, governing power, steering power). I choose the term 'steering' rather than 'governance' because 'steering', unlike 'governance', has such a flexible sense that it fits all the different ways in which actors' exercising powers they have can manifest itself. .

Where are we now? We have moved away from Weber's root-formula in several steps, without losing sight of it, and have arrived at a more abstract real definition of power: In relation to people and relevantly similar actor-entities, power is their ability to steer, in a counterfactually robust way, trajectories of change by controlling forces of their own.

The power that several actors have in a common situation finds its limit for each individual actor in what the forces that each actor has at his disposal in this situation could direct or change in this situation, and finds its limit for all of them together in what trajectories of change they could steer with the forces they would have at their disposal if they were to unite into a corporate or cooperative actor. Let's take an example in macro format: Suppose we consider the set of all members of a politically connected society, e.g. the citizens of Germany. How much power is there in this society? The power-theoretical answer, at the highest level of abstraction, is: The totality of all changes in behavioral trajectories that all members can steer with all the forces they could control counterfactually robustly, if they wanted to collectively bring about certain changes that each of them considers important.

Whenever the exercise of power can be definitely described in an individual case (e.g. entry into force of the law of the AI Act via EU decision on August 2, 2024), the power exercised manifests itself as the totality of those changes in practices that can be described as effects (power effects) for which the power holder is causally responsible through the exercise of his power in this case. It goes without saying that the totality of the factual effects of power can also include changes that are beyond those within the scope of control. Control can be spurious. No power is perfect. No counterfactually robust control is absolutely robust.

4. Computing power and powers of the intellect

Consider the ability of most people to perform simple calculations in their heads, e.g. the addition of two numbers in the number range up to 100. If I want to, I can increase my existing mental arithmetic ability, for example by developing a practice of training or by learning a simple algorithmic cultural technique, calculating with paper and pencil, or with an abacus, or, as is taken for granted today, calculating with electronic calculators. I can want to learn to master such techniques, e.g. because I want to be able to calculate better (more complex, faster, more reliably, in a larger number range). I can intend to do it and actually do it. We could speak here of the *power of learning*. The cultural technique of being able to calculate without having to call on other people as assistants or using technical agents must be learned and then empowers me (to a certain extent) to

calculate correctly. Once I have learned (F^*) how to control powerful computing agents and thereby to steer their vast computing power ($F+$) towards my purposes, by enlisting the computing power of the technofact I have improved my ability to achieve all kinds of purposes for which computation is required. My power has been augmented. Let us assume that persons A and A' are entrepreneurially active, involved in competitive practices. As such, they are meaningfully oriented toward each other and thus interactively situated in the social force field of market forces. Then A has more power in competition with A' than A' whenever it is important, for example, to calculate the profitability of an investment more quickly than one's competitors.

What about the "power of reason"? What we mean when we say of people, for example, that someone is "in full possession of her intellectual forces" (or conversely, that someone is "no longer in possession of his reason", that someone has "lost her mind"), can certainly also be understood as possession or loss of intellectual powers. Among all the abilities and competencies whose possession we can summarize as the "possession of intellectual powers" or as the "complex ability to orient oneself and to behave rationally" (also known as general intelligence), there is surely also the competence to orient oneself in the culturally interpreted "space of reasons" in an insightful and projective anticipatory manner, just as one thinks that others whom we take seriously as interaction partners would do. By virtue of our power of *reason*, we find ourselves in a realm of *reasons*, a realm that has always already been culturally established and that acts, so to speak, like a field of intelligible forces; and we are not only immersed in this space, but we are also able to actively change this space through the power of our reason.

Wherever we really *work* with our General Intelligence and make a practical difference - in whatever situations and practices where we are concerned with chances of success, i.e. where we orient ourselves "rationally" - we exercise powers of the intellect: By directly controlling relevant intellectual forces (F^*), we can more or less control many other kinds of forces ($F+$) to which we would otherwise only be exposed. By virtue of the powers of the intellect, we are better able to steer away from events, outcomes, consequences that we want to avoid for good reasons, and to approach goals that we want to achieve for good reasons.

5. Working forces, agents, and actors

If, as I have suggested, we think of power as a complex relation of steering via controlling *forces*, we underline an important point of power, namely that something is working; that some kind of *work* occurs. In the case of the mere potentiality of power, it is work that would take place if the powerholder in question were to actualize its power potential. Of course, it is also possible that an actor possesses a certain power in order to *prevent* something that would otherwise happen, materialize or be done. In this case too, work occurs, namely resistance or blockade work. In any case, energy (=actualizable work) must be expended.

In our pre-scientific understanding, it makes perfect sense to speak of power in terms of, for example, the force of people's will, or the force of strong valid social norms, or

the force of people trying to do something. However, let us try for a moment, with the help of the concepts ‘work’ and ‘energy’, to trace certain analogies that exist between the *naturalistic* (i.e. non-normative) power vocabulary of mechanics, which is theoretically specified and refined in physics, and the power vocabulary of our lifeworld, which is theoretically specified and refined (if at all) in the social sciences and covers *both* naturalistic and normative cases, depending on the context. There are no normative forces and normative standards in the world as conceived in physics. Rather, the analogy is based on the concept of a natural force. In physics, the concept ‘force’ is used to describe and explain the behavior of objects of a certain type (e.g. mechanical bodies) when objects act on each other through a force (e.g. gravitational force). Categorically, forces have vector form, i.e. magnitude and direction. They cause changes or changes in changes (e.g. when an opposing force blocks a change that would otherwise occur). Changes caused by classical mechanical forces (gravity, friction, compression, and tension) are typically changes in velocity (acceleration, deceleration, change in direction of motion of objects) and changes in shape (deformation of objects). Forces that only operate in cultural worlds (e.g. willpower, legal power, purchasing power, property³) cannot be reduced to mechanical forces, but only *involve* them (e.g. in physical strength, firepower of weapons). Involved in practices, people who can behave in this or another way are *meaningfully* oriented towards each other and thus interact. Fields of *culturally constituted* forces in the background of the lifeworld of encultured people manifest themselves in directional changes in the will and motivation of individuals as well as in directional changes in their networks of reasons and beliefs, with all the consequences that such directional changes may have for all the persons involved in all the practices concerned. All social practices taken for granted in the cultural environment of human actors are part of the “background” (sensu Searle 2019: 145-173) of human agency.

As I said, the exercise of power is associated with work. When actors have power and exercise it, some sort of work is being done. We need not unnecessarily limit the ontological constitution of the respective “workers” to human beings, but can treat their constitution as an ontologically free variable. We can call the values of this variable ‘agents’. In the perspective of the general theory of power I am exploring here, machines that are set to perform a certain service are also agents (e.g. steam engines, computers executing essentially unchangeable algorithms). *I think it makes sense to make a clear conceptual distinction between persons, actors and agents.* The usual English terminology is unreliable here. If we embrace methodological anthropocentrism, then persons who set and achieve purposes and goals remain the paradigmatic and, in terms of intentionality, most sophisticated value of the actor variable in the concept of power. If (as suggested in section 2) we exclude the problem of human free will and modestly replace it with the ability to achieve goals and purposes intelligently, then we can make the conceptual cut between persons and agents: Agents (let us say) are actors fixated on specific performances. Human persons, on the other hand, can set, pursue or reject a wide variety of objectives. Persons are not actors fixated on specific tasks, but actors who can determine their

3 Consider: Within the normative force field of property rights, having property gives a person the normative power to exclude others from using that property.

tasks and performances themselves. They have the *power of self-determination* to a certain extent (relative autonomy). A distinction between actors who are agents and actors who are persons would be a plausible specification of the generic concept of an actor. Likewise, one could also say that persons are “full” actors, or “actors par excellence”, and agents are only “reduced” or “specifically limited” actors.

One might object that the distinction I want to draw between persons, actors and agents seems unnecessarily complicated when considering that, speaking from a theory of action perspective, humans (and certain kinds of animals) are so far the only recognized instances of intentional, purpose-setting entities that operate to some extent independent of the imposition by other entities. In this sense, the distinction between actors and agents may seem sufficient from the standpoint of an ordinary theory of action. What I want to question, however, is precisely the sufficiency of such a theory-of-action perspective with respect to technically created artificial entities with some sort of agency.

The design of multiple-agent systems plays an important role in the interaction between humans and robots (Dahiya et al. 2023). It would be interesting to analyze such systems from a power-theoretical perspective. As mentioned earlier, certain legally constituted entities, e.g. corporations, may, under certain circumstances, be considered agents and multiple-agent systems. In line with the terminology used in computer science and AI research, certain types of technofacts can also be considered agents, e.g. computer programs that, in action, run to some extent independently of user intervention, make decisions independently within a defined scope, exchange information with other programs (‘communicate’) and react more or less adaptively (‘intelligently’) to changing conditions and inputs in virtue of their own forces (‘autonomously’).⁴

I have just referred to common uses of language in Computer Science and related fields. Here we encounter a difficult problem at the intersection of philosophy of language, theory of action and metaphysics: anthropomorphization. Why take this use of language “at its word” and uncritically ignore the fact that it is strongly anthropomorphizing? The question is this: Under what conditions does the intentionalist language of action, on which we can rely when we act as humans among humans, lose its explanatory and predictive power and turn into anthropomorphic fiction? Is it misleading to speak of “agents”, for example, when we ascribe certain capabilities to artefacts and describe them as behaving “to some extent independently of user intervention”? There is a spectrum of philosophical positions on this question. At one end of the spectrum, one can argue for the descriptive use of the language of human action without postulating particularly demanding conditions of adequacy. Who’s afraid of anthropomorphism? An argument in favour of this inflationary position is that the attribution of intentionality (“my computer did not remember this password and also forgot to remind me of an important appointment”) does not imply the attribution of genuine mental states, as we would ascribe them to persons whose behaviour we describe in an intentionalist way (“my colleague did not remember this password...”). At the other end of the spectrum, we find

4 With regard to the new philosophical research field of machine ethics, see Cervantes et al 2020; for an overview of the research literature on conversation agents in particular, which has grown immensely since the 1960s, see Schöbel et al. 2024.

positions such as psychological behaviorism, which remained relatively unsuccessful. As early as the 1940s, anthropomorphism had already begun to be eliminated in the way that the philosophically much more elaborate “eliminative materialism” in the philosophy of mind from the 1970s onward wanted to do.

Is it perhaps sufficient in terms of power theory to understand agents as ‘means’ or ‘aids’ (as aids for persons who want to use these means or aids in order to accomplish something specific that they cannot accomplish on their own, or only with difficulty)? In many cases, this rather banal categorization will indeed suffice. For example, in the case of partially autonomous systems that steer their own work activities controlled by algorithms but still rely on human input, so-called Human in the Loop (HITL) systems. Such systems are power-enhancing tools for people, or ‘means of power’, if you will. Means of power enable or enhance possessors of power but have none themselves. The instance of a rationally willing subject (person) or at least an actor pursuing goals or purposes in an intelligent way is missing here. Algorithms as code or program do not pursue goals or purposes, but we (more or less intelligently) do that with their help. Algorithms do not pursue their own purposes even when they are active in the appropriate machine environment, i.e. when they control the course of the machine – that is, so to speak, the “performance” of the algorithms – if they only do the work for which we have programmed them.

However, once we construct algorithmically controlled machines in such a way that the latter virtually realize a familiar power of humans (and also of many kinds of non-human animals), the power of learning, then *we empower* such machines in certain respects. This creates a power relationship within the relationship between persons and machines. One could perhaps characterize the empowerment of the machine as an ‘externalization’ of human power.⁵

Again, we come up against the question of the proper rational constraints on the tentative (and perhaps merely metaphorical) extension of concepts originally rooted in how we as human beings make sense of our behaviors and powers. The concept of learning is certainly an indispensable part of the human language of action, because we humans experience ourselves as learning beings in countless contexts (e.g. when we find human or even animal behavior “intelligent” – a term coined by scientific psychology 120 years ago). Is what we now describe and explain as “machine learning” real learning, or just so-called learning? I have to admit that I don’t think this is a good question. Whatever our “exact” understanding of learning, it is certain that not only humans learn, but also animals. And if animals, then why not the corresponding devices and machine systems that we have created through technological practices? Or should the fact that we do not bring animals (capable of learning) into the world with the help of technological practices, whereas we do bring machines (capable of learning) into the world with the help of technological practices, make a significant difference here? I don’t think so.

5 For lack of reciprocity (Loer 2021) from person to person, I don’t want to call it “delegating,” as it is sometimes called in the marketing jargon of advertising for so-called personal intelligent assistants. When I set the alarm on my smartphone to go off at 5 a.m., it would be quite strange to say that I “delegate” waking up to my smartphone, as if it were a hotel receptionist or my personal butler. I just set the alarm.

Machine empowerment arises in fully autonomous systems that no longer rely on human-controlled inputs: human-out-of-the-loop (HOOL) systems. (Admittedly, humans are not “out of the loop” in an absolute sense, since the very existence of the loop has been or is being produced in human technological practices. Yet the fact that y originally depended on x does not mean that y continues to depend on x . And if there is a z that originally depended on y and a y that originally depended on x , this also does not imply that z continues to depend on x .) If we think, for example, of military HOOL weapons systems, such as anti-aircraft missiles that autonomously pursue their target, then this externalization of power, once released, in some circumstances can no longer be contained, although it remains within the intended and technically established scope of the purposes for which we design this type of device. Even a more far-reaching externalization of power is conceivable with self-learning machines. The power of learning, which we externalize into this kind of technofact by technically designing a virtual instantiation of this power in the machine, possibly can empower self-learning machines to leave the desired and technically established scope of our initial design-purposes for them (Davidson 2023), or, if you want to put it ominously: it can empower them to break out.

6. The power of algorithms as propaganda

Sufficient purchasing and learning power given, if I can make suitable software agents serviceable, I can augment and enrich the power of my agency for a variety of tasks: If such things are important to me, for example, for the purposes of targeted research on the internet, for checking and prioritizing my emails, for filling out electronic forms, for synchronizing profiles in social networks and curating news feeds, and last but not least for finding good offers in online retailing. Digitalization as a cultural process is evident, not least in the increasing incorporation of algorithmically organized agents into the routines and habits of our everyday practice. The sweeping global success of language, image and sound generating AI models that set in with *Open AI*'s coup at the end of 2022 – a perhaps similarly significant historical turning point as the arrival of market-ready smartphones in 2007 – testifies to an enormously strong general interest in power-expanding technofacts of all kinds. Of course, it would be naive to think of this interest as sudden or “natural”. It is an expression of a long socialization in a culture increasingly molded by the cultural process of digitalization (D1+D2+D3).

Here we come across a new important aspect of the power of algorithms, namely the ideological power of digital cultural innovations (AI being the most recent one) manifesting itself in the vagaries of our imagination. Before AI there were Clouds and Big Data, more is in the waiting (e.g. quantum computing), and all of this is summed up in the phantasm of the algorithm. “Being digital” was the formula for a future full of hopes for progress in the 1990s, when the internet became a global reality; today, “becoming smart” is the only libidinous narrative of progress left, replacing the remnants of earlier social utopias in a world of artificial intelligence. Those who love progress welcome the algorithm as a symbol of a better future. A comparison with the alchemy that the medieval goldmakers hoped for comes to mind. The ability to stimulate fascination and

mimetic competition, to promise the blue sky and get away with it, is certainly a strong power, the power of controlling collective attention for the purpose of transfiguration, hype and affirmation, and business. It is well known that the most powerful actors in this very unequally distributed power are commercial actors, such as the tycoons of digital corporations.

It seems to me that the perspective on AI and algorithms as a glorifying ideology opens up a wide field of fruitful research questions for philosophy, cultural studies and social sciences. A formulation by David Beer (2016: 9) points in the right direction:

“The notion of the algorithm is part of a wider vocabulary, a vocabulary that we might see deployed to promote a certain rationality, a rationality based upon the virtues of calculation, competition, efficiency, objectivity and the need to be strategic. As such, the notion of the algorithm can be powerful in shaping decisions, influencing behaviour and ushering in certain approaches and ideals. The algorithm’s power may then not just be in the code, but in that way that it becomes part of a discursive understanding of desirability and efficiency in which the mention of algorithms is part of ‘a code of normalization’.”

The same applies to AI and the vocabulary associated with it, which is used inflationarily to evoke a certain rationality based on the virtues of calculation, competition, efficiency, objectivity and strategy. For the formulation and dissemination of the most suggestive slogans, rhetoric and narratives, there are many actors with very little communicative power and a few actors with enormous communicative power who are able to shape and reinforce the corresponding frames. This is obviously a highly asymmetrical power relationship. We should carefully examine the ways in which algorithms are part of broader programs of social change and development that are touted as rational and progressive.

Another worthwhile concern on the agenda for philosophical digitalization research on questions of power might be to reveal the massive unreflective trust placed in systems that are being highlighted as “algorithmic” systems and hyped as “artificially intelligent”. What strikes me as particularly interesting is the gratuitously optimistic idea that once some of the messy problems of digital cultural discrimination are technically solved, such systems will be neutral, reliable, trustworthy and much better than anything human beings can do. Those of us who see ourselves as critical humanists should carefully examine who gains and who loses from the social construction and selective distribution of this trust.

7. Algorithmic steering, normative steering, and discursivity

If social steering can take many forms, what is the difference between steering by norms and steering by algorithms? Due to their anchoring in speech acts, normative ‘programs’, from the most banal and weakest, which are individually chosen as a means for private purposes, such as a cooking recipe, to the most powerful, politically authorized, universally binding laws, have an inherently extremely valuable property: all normative forces brought into play by speech acts and all possible effects that come into play through

them can be comprehensively reflected on by (almost) every person who belongs to the relevant communication community in which the normative ‘program’ in question is running. People can be or become addressees, but also authors or modifiers of normative textures, i.e. of patterns of culturally constituted normative forces. They can question the norms as to their alleged meaningfulness, can demand and find good reasons for their worthiness of recognition, can distinguish really good from merely supposedly good reasons and reject bad reasons concerning the claims of authority and validity inherent in normative textures. The exercise of reflection on validity, the ‘force of reflection’ (F^*) can change, strengthen or weaken the available justifications and thus the normative forces ($F+$) of the norms in question.

The fundamental openness for reflection on validity – for short: discursivity – is a characteristic of normative textures that can be explained by the fact that normative textures are culturally formed and formatted and can be formulated and explicated in a natural language. Algorithmic programs, on the other hand, i.e. the compositions of instructions formulated in programming languages and designed for the purpose of steering the trajectories of change in the operations of computing machinery, do not have this property inherently. Can algorithmic control programs have this property at all? To a certain extent, yes, but only to the extent that ‘we’ purposefully and consciously build up discursivity around them, so to speak. It will not be possible to write the property of discursivity into the programs themselves, pace so-called “explainable AI” (XAI). And the more ‘we’ enable the programs to change themselves, so that after an initial phase their self-transformation becomes increasingly independent of the human power of instruction, the more we relinquish our control over the externalization of human power and empower the machine learning systems (ML aka AI). Of course, we are and remain responsible for this - this development does not happen naturally, but as a cultural process in which we, initially at least, always have a certain (political etc.) power to shape things.

The machine-learning apparatus is still in need of instruction during a training phase (this is also likely to change as soon as ‘we’ decide to construct communities of robots and “put them into the world”), but when it works as it should it will continue to learn on its own – and soon the box turns black. The exact activities of the program in its immediate operational environment then become opaque to us. ‘We’ have wanted this loss of control or at least we accept it. If ‘we’ want to reverse it or at least keep it within limits, then technical manoeuvres are required that (currently still) involve an almost absurd amount of effort (Pangutti et al. 2023). Even if XAI research makes great progress and solves the technical side of the problem of non-transparent machine learning processes (Ali et al. 2023), still the normative, ultimately political problem remains that different groups in very different positions in the overall context of the development and application practices of (semi-)autonomous ML systems generally also have very different control interests and views. ‘We’ are many, with diverse, often irreconcilable interests.

Think of students, their parents, teachers, administrators, system developers, entrepreneurs, data protection officers and education policy makers in the transformation of a traditional school into a smart school (McConvey/Guha 2024). The utopian (dystopian?) vision here is AI-controlled personalized teaching and the automated monitoring

of all educationally relevant psychosomatic parameters of all students (Dimitriadou/Lanitis 2023). What they want interpreted and what they don't, what they understand by better and worse explanations and what everyone believes they are entitled to - none of this can be brought down to a common denominator without sensible political negotiation processes. Such political negotiation processes, in turn, cannot be delegated to algorithmically controlled agents unless politically empowered actors such as parliaments and regulatory bodies irrationally wanted to renounce the creative normative powers they in fact possess.

8. Conclusion: What is the hard problem of AI?

How can we make machine-learning AI systems explicable in the sense that their outwardly opaque input-output functions become intelligible to people with natural human intelligence? The problem of discursivity – not of algorithms or programs, of course, since that is impossible, but discursivity of the practices in which ‘we’ engage with algorithmically controlled machine-learning systems – is about much more than just “explainable” or “interpretable AI”. Rather, the problem of the discursivity of AI as today’s flagship of digitalization (D1+D2+D3) raises important scientific and political questions (Hedlund/Persson 2024, Walter 2024). How *should* discursive power best be distributed?⁶ Who of ‘us’ can, who should be empowered and authorized to demand, give and evaluate justifications in case of doubt? (The problem of the objectively appropriate and ethically sound (“fair”) distribution of discourse opportunities and the scarce resource of discourse). What are the actors that *actually* fill this idealized and potential ‘we’ today?

These questions point to the hard problem of whether, in the cultural process of digitalization, the innovation, dissemination and application of products of digital technology can be sufficiently steered at all by the social forces at the disposal of the political sovereign (here: the democratic ‘we’). In light of the current disillusionment with the possibility of control of political rule in democracies, the problem of regulating the technical development and cultural normalization of devices that learn by themselves through training or even independently of human-made instruction appears to be a challenge that is so massive that it threatens to overtax political power.

The challenge of the hard problem can also be formulated abstractly in terms of two concepts that occupy a prominent position in the cultural self-understanding of Western modernity: autonomy and its opposite, heteronomy. How do ‘we’ manage, in the course of the progressive autonomization (intelligent self-control) of the apparatuses that ‘we’ want, to maintain at the same time so much heteronomy (external control) of the apparatuses that our private autonomy (individual freedom) and collective autonomy (po-

6 My term ‘discursive power’ is meant to refer to the kind of power that actors have when and to the extent that they can control the rhetorical forces (F*) of argumentative utterances and can thereby control the forces (F+) that reasons gain among persons who value good reasons (Kettner 2018, Haugaard/Kettner 2020). The normative force of a convincing reason to consider something to be true (or false), right (or wrong), beautiful (or ugly), good (or bad) are some examples of respective F+ forces.

litical self-government) – two civilizational achievements that are always precarious and endangered anyway - will not regress, erode, or disappear.

References

- Ali, S., T. Abuhmed, S. El-Sappagh et al (2023): Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence. In: *Information Fusion* 99: 101805. <https://doi.org/10.1016/j.inffus.2023.101805>.
- Beer, David (2017): The Social Power of Algorithms. In: *Information, Communication & Society* 20 (1): 1–13.
- Bergé, Jean-Sylvestre, Séphane Grumbach and Vincenzo Zeno-Zencovich (2018): The 'Datasphere,' Data Flows Beyond Control, and the Challenges for Law and Governance. In: *European Journal of Comparative Law and Governance* 5 (2): 144–178.
- Bogost, Ian (2015): The Cathedral of Computation. In: *The Atlantic*, January 15. <http://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/>.
- Cervantes, J. A., S. López, L. F. Rodríguez et al (2020): Artificial Moral Agents: A Survey of the Current Status. In: *Science and Engineering Ethics* 26: 501–532.
- Dahiya, Abinav, Alexander M. Aroyo, Kerstin Dautenhahn and Stephen L. Smith (2023): A Survey of Multi-Agent Human-Robot Interaction Systems. In: *Robotics and Autonomous Systems* 161: 104335. <https://doi.org/10.1016/j.robot.2022.104335>.
- Dimitriadou, Eleni and Andreas Lanitis (2023): A Critical Evaluation, Challenges, and Future Perspectives of Using Artificial Intelligence and Emerging Technologies in Smart Classrooms. In: *Smart Learning Environments* 12 (10): 1–26. <https://doi.org/10.1186/s40561-023-00231-3>.
- Floridi, Luciano (1999): Information Ethics. In: *Ethics and Information Technology* 1: 33–52.
- Floridi, Luciano (2014): *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press.
- Frieze, Heidrun, Marcus Nolden, Gala Rebane and Miriam Schreiter (Eds) (2020): *Handbuch Soziale Praktiken und Digitale Alltagswelten*. Wiesbaden: Springer.
- Gramelsberger, Gabriele (2024): Die Philosophie des Digitalen: Zur Struktur, Signatur und Phänomenologie des Digitalen. In: *Philosophische Digitalisierungsforschung. Verantwortung, Verständigung, Vernunft, Macht*. Edited by R. Adolphi, S. Alpsancar, S. Hahn, and M. Kettner. Bielefeld: transcript, 285–296.
- Grunwald, Armin (2019): Digitalisierung als Prozess: Ethische Herausforderungen inmitten Allmählicher Verschiebungen zwischen Mensch, Technik und Gesellschaft. In: *Zeitschrift für Wirtschafts- und Unternehmensethik (zfwu)* 20 (2): 121–145.

- Grunwald, Armin (2024): Digitalisierung als Prozess. In: Philosophische Digitalisierungsforschung. Verantwortung, Verständigung, Vernunft, Macht. Edited by R. Adolphi, S. Alpsancar, S. Hahn, and M. Kettner. Bielefeld: transcript, 415–435.
- Haugaard, Mark and Matthias Kettner (Eds) (2020): Theorising Noumenal Power: Rainer Forst and His Critics. London: Routledge.
- Habermas, Jürgen (1984): The Theory of Communicative Action. Translated by Thomas McCarthy. Vol. 1+2. Boston: Beacon Press.
- Hedlund, Maria and Eric Persson (2024): Expert Responsibility in AI Development. In: AI & Society 39: 453–64.
- Kettner, Matthias (2018): The Forstian Bargain: Overrationalizing the Power of Reasons. In: Journal of Political Power 11 (2): 139–50. <https://doi.org/10.1080/2158379X.2018.1479039>.
- Loer, Thomas (2021): Reziprozität: Annäherungen an Eine Grundlegung der Kultur- und Sozialwissenschaften. Wiesbaden: Springer.
- McConvey, Kelly and Shion Guha (2024): “This Is Not a Data Problem”: Algorithms and Power in Public Higher Education in Canada. In: CHI ’24: Proceedings of the CHI Conference on Human Factors in Computing Systems, Article No. 16. <https://doi.org/10.1145/3613904.3642451>.
- Morozov, Evgeny (2013): To Save Everything, Click Here: The Folly of Technological Solutionism. New York: PublicAffairs.
- Morriss, Peter (2002): Power: A Philosophical Analysis. 2nd ed. Manchester: Manchester University Press.
- Panigutti, C., R. Hamon, I. Hupont et al. (2023): The Role of Explainable AI in the Context of the AI Act. In: Proceedings of the 2023 ACM conference on fairness, accountability, and transparency, 1139–1150.
- Peirce, Charles Sanders (1902): Virtual. In: Dictionary of Philosophy and Psychology. Edited by James Mark Baldwin, Vol. II. London: Macmillan, 763–764.
- Schöbel, Sofia, Anuschka Schmitt, Dennis Benner, Mohammed Saqr, Andreas Jansen and Jan Marco Leimeister (2024): Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. In: Information Systems Frontiers 26: 729–54.
- Searle, John (2010): Making the Social World. Oxford: Oxford University Press.
- Striphas, Ted (2015): Algorithmic Culture. In: European Journal of Cultural Studies 18 (4–5): 395–412. <https://doi.org/10.1177/1367549415577392>.
- Walter, Yoshija (2024): Managing the Race to the Moon: Global Policy and Governance in Artificial Intelligence Regulation. A Contemporary Overview and an Analysis of Socioeconomic Consequences. In: Discover Artificial Intelligence 4 (14). <https://doi.org/10.1007/s44163-024-00109-4>.

- Weber, Max (1964): *Wirtschaft und Gesellschaft: Grundriß der Verstehenden Soziologie*. Erster Halbband. Studienausgabe. Edited by Johannes Winckelmann. Köln/Berlin: Kiepenheuer & Witsch.
- Weilenmann, Alexandra and Thomas Hillman (2019): Selfies in the Wild: Studying Selfie Photography as a Local Practice. In: *Mobile Media & Communication* 8 (1): 42–61.