Original paper UDC: 17.025.2:004.8(045)
doi: 10.21464/sp39101
Received: 17 August 2023

Bruno Gransche
Karlsruhe Institute for Technology, Douglasstraße 24, DE–76133 Karlsruhe
mail@brunogransche.de

Normgiving Technology?

Metaphorization of Autonomy and What It Teaches Us

Abstract

Human autonomy is not technical "autonomy". That we refer to both phenomena equivocally reveals a label-transfer from human to technical contexts, i.e., metaphorization. Autonomous humans choose to submit to self-given laws (1). Machines, regardless of the metaphor, cannot shed their heteronomy (2). What is literally meant by "technical autonomy" as opposed to or interchangeable with "automation" needs to be explored in detail (3). To differentiate, multi-layered approaches to technical autonomy have been proposed (4) still revealing the unavoidable metaphorization within the description and analysis of autonomous systems. This metaphorization needs to be analyzed in terms of the similarity relation at work in the conceptual transference, which teaches us that an entity (e.g., an autonomous person) is, is not, *and* is like *another (e.g., an autonomous robot) at the same time.*

Keywords

artificial intelligence, automation, autonomous systems, autonomy, heteronomy, metaphor, self-legislation

Introduction

Humans' autonomy is completely different from the "autonomy" of technical systems. Autonomy means in essence self-determination or self-legislation, in contrast to heteronomy, external determination, or being determined by the laws of others. Autonomous humans are thought of as being rational and in possession of free will; they are autonomous when they are both lawgiver and subject to the self-given law. 'Self-given' in this context does not necessarily imply self-invented or conjured ex nihilo; the laws, goals or maxims can be found in society and be imposed by tradition, yet human autonomy implies to choose and acknowledge some of those as binding for oneself. Technical systems of any degree of automation are determined by the specifications and rules of their system designers instead; their being constructed means their heteronomy concerning their designers. Any talk of *technical autonomy* or autonomous technical systems referring to the rational human lawgiver-law follower sense is, therefore, a fiction or metaphorical talk. *Autonomy* concerning technology and humans is word-same but content-different – a mere equivocation. Yet, this equivocation invites widespread bleeding over aspects of meaning from one to the other, which has to be understood with a closer look at the underlying metaphorization at work.

The use of *autonomous/autonomy* concerning technical systems is, similar to configurations like *artificial intelligence*, today a widespread linguistic usage in research and development, industry, media, and the public as well as in politics; this usage is partly interest-driven, partly merely adopted without

reflection. It is not a matter of trying to correct a supposedly wrong use of the term (such as "autonomous robots") with reference to a supposedly right one (such as the "autonomy of pure will", Kant). After all, any supposedly correct usage could in turn be "corrected" with reference to an even earlier one (e.g. political autonomy in Herodotus). Rather, it is important to clearly differentiate area- and context-specific levels of meaning and to prevent problematic conclusions as well as questionable orientations of action through confusion of such levels.

The configurations of *technical autonomy*, *artificial intelligence*, *machine learning*, etc. all take a concept from a customarily used human sphere like *autonomy*, *intelligence*, or *learning* and transfer it to a new sphere like *technology*, *artifacts*, or *machines*. Thus, they indicate and even create a similarity between both spheres that was not seen or did not exist before these configurations were coined. This linguistic maneuver to *transfer* a concept from one sphere *over* to another is – based on the Greek expression for *over-transfer*, namely *meta-pherein* – known as metaphor. Most scientific definitions resort to metaphoric usage of concepts known from human beings to convey technical characteristics or functionalities; or: they create a similarity in the first place, where none has been before. The use of metaphor is present across sectors, it is common in media and public discourse, in politics and business talk and advertising, and it is deeply inscribed in scientific language as well. Especially that is the case with digital technology, artificial intelligence, and technical autonomy. This current omnipresent usage of autonomy and intelligence metaphors needs close examination and explicit analysis because they are here to stay and thus, they transform the meanings and our understanding of both hitherto unbridged spheres that are configured in human-technology metaphors. This article provides clarification of this connection as a first step to further much-needed examination that, eventually, might lead to a metaphor literacy of digital technology and thus contribute to adequate transformational guidance for responsible digitization of our lifeworld.

Firstly, a basic overview of autonomy (1) as used in philosophy mainly for the personal autonomy of rational beings since Kant is provided, not without pointing to its origin as metaphorization from a political state context. Secondly, the transfer of this human autonomy to technical systems is addressed (2) which then calls for closer investigation of current predominant usage in the technical sphere, contrasting autonomy to automation (3) and considering multi-layered approaches of autonomy and/or automation (4).[1] This provides the basis and many examples to address the metaphorization processes that produce the label transfers at hand (5).

Autonomy – Lawgiving Law Subjects

Autonomy since antiquity means self-legislation in the literal sense of *auto/self* and *nomos/law*.[2] The ancient origin of the word meant the right of states to make their own laws, which thus meant political-legal freedom to be independent of foreign domination on the outside and independent of tyranny on the inside. From the 19th century onward, this political-legal autonomy is understood, for example, as the "power of certain bodies within the state to set legal norms".[3] Citizens would not be autonomous in this sense, since they can decide and act freely within the framework of the legal norms that apply to them (abiding by or violating those norms in personal autonomy) but cannot

set their own legal norms (political autonomy); similarly, mere individuals cannot be politically autonomous, since recognition by others is necessary to set rules. In any case, autonomy is a relational concept, a right to choose, recognize, or reject norms, or freedom from external normative pressure and to legislate independently; this freedom is always conditioned by likewise autonomous others and wider contexts, (historical) situations, etc. Autonomy is not an omnipotent freedom without conditionality, but a relational property of and between people. The application of the concept of autonomy to persons and not to corporations or states was formatively introduced by Immanuel Kant. With Kant, autonomy means the possibility of human beings as rational beings – or rather as beings with the potential to rationality (*animale rationabile*) – to determine themselves. Kant writes:

"All philosophy is 1. autognosy [i.e., self-knowledge, BG] 2. autonomy [i.e. self-determination, BG], science and wisdom."[4]

For Kant specifically, this transfer of the notion of self-legislation from states to individuals is closer than it might seem, due to the harmony of the will and universal practical reason:

"…the third practical principle of the will, as the supreme condition of its harmony with universal practical reason, the idea *of the will of every rational being as a universally legislating will.* According to this principle, all maxims are rejected that are not consistent with the will's own universal legislation. Thus the will is not just subject to the law, but subject in such a way that it must also be viewed *as self-legislating,* and just on account of this as subject to the law (of which it can consider itself the author) in the first place."[5]

Ideally speaking, only maxims would be pursued that can be accepted as universal laws, therefore any law of states would be an institutionalized (halfway universalized, if you will) maxim, and only such state laws could be accepted that could be universalized for all humanity. Human beings with a rational will ultimately are both universal lawgivers (authors of laws) and subjects to universal law; both are decisive for humanity. The means to recognize maxims that are fit for being universal laws (for the rational will's own universal legislation) are mentioned in Kant's characterization of philosophy in addition to autognosy and autonomy above: science and wisdom. However, whether the metaphorically bridged spheres are close or far apart, they are never entirely the same and therefore the understanding of a concept in one context never is the same as in the other: individuals just do not give laws to a plethora of inner citizens like states do, "autonomous" systems just do not give, recognize

1
Several passages of sections 1 to 4, refer in essential parts to my contribution "Technische Autonomie" that will be published in: Mathias Gutmann, Klaus Wiegerling, Benjamin Rathgeber (eds.), *Handbuch Technikphilosophie*, Metzler, Heidelberg 2024.

2
On autonomy, see: Rosemarie Pohlmann, "Autonomie", in: Joachim Ritter, Karlfried Gründer, Gottfried Gabriel (eds.), *Historisches Wörterbuch der Philosophie. HWPh*, vol. 1, Schwabe Verlag, Basel 2010, pp. 2407–2463 [701–719]; Michael Kühler, Nadja Jelinek (eds.), *Autonomy and the Self*, Springer Netherlands, Dordrecht 2013.

3
R. Pohlmann, "Autonomie", p. 2419.

4
Kant, AA XXI, Erstes Convolut, 106, my translation.

5
Immanuel Kant, *Groundwork of the Metaphysics of Morals*, German – English edition, transl. Mary Gregor, Cambridge University Press, Cambridge 2011, p. 91.

or reject laws, rules or maxims like rational individuals with free will do. However, transferring these concepts via metaphorized statements has performative effects on all involved meanings, which are addressed below.

Technical Autonomy – Self-Legislating Machines?

As established, to say that humans have personal autonomy does not imply that they were not somehow socially, historically, culturally, psychologically, biologically, etc. predisposed, or conditioned; it just means that they are not entirely determined externally and that they have to be thought of as being able to choose, to acknowledge or to reject laws or maxims as applicable for themselves or not. They do not need to be doing so all the time nor most of the time to be attributed personal autonomy. Yet, autonomous human beings are in principle able to reject any heteronomous socio-psycho-biological imperative as not binding and choose to act against it. They can even acknowledge a maxim as applicable and just suspend it only for themselves for a specific situation or period. Whereas no robot could judge: "This is a valid rule and all robots including me should follow it, but just now I will not." To echo a notorious quote on political sovereignty from Carl Schmitt: *Autonomous is he who decides on the exception*.[6] Because:

"In the exception the power of real life breaks through the crust of a mechanism that has become torpid by repetition."[7]

Since sovereignty is being transferred from the context of states to humans more recently[8] just like autonomy has been after Kant and since both concepts are colloquially widely used interchangeably, Schmitt's remark on the torpid mechanicality of rule-following behavior is instructive for the differentiation of human and technical autonomy: Humans can behave mechanically, reacting widely conditioned or even determined, following given rules in a robot-like (or sometimes zombie-like) manner, but opposed to technical mechanisms or robots, they do so by continuously refraining from making themselves the exception. That people might *normally* behave mechanically rule-abiding (in a technicist, quite pessimist view of humans), "proves nothing; the exception proves everything";[9] people can and often spectacularly do make themselves the exception. Personal autonomy consists in acknowledging or rejecting given goals or rules as one's own; the fact that this acknowledgement is conditioned by multiple factors like tradition or avoidance of social punishment, does not diminish the autonomy and especially not approximate human autonomy to a technical autonomy.

The characterization of autonomy according to Kant as "self-legislation by reason" or more generally as freedom of will already shows fundamental transference difficulties of this conceptual understanding to technical systems since these cannot be qualified as rational beings and no will can be imputed to them. At least, this is the case for all current systems and those under development; the question of whether technical systems as post-biotic "consciousnesses" might eventually develop an artificial equivalent to human reason or will is a speculative one and will not be addressed here.[10] This question is also not crucial insofar as the majority of those who speak of technical autonomy today do not have "self-determination as technical rational beings" in mind anyway,[11] but mostly something like "functional without user monitoring and

interruption" (discourses around hypothetical strong AI are an exception and nonetheless fictional today).

Following a distinction by Erich Fromm (authoritarian versus humanistic ethics), man could be understood as an authority to the (thus heteronomous) technology, for which applies:

"… an authority states what is good [...] and lays down the laws and norms of conduct."[12]

While for autonomous technology in the full sense would also have to apply what Fromm formulates in the sense of his humanistic ethics for man:

"… man himself is both the normgiver and the subject of the norms, their formal source or regulative agency and their subject matter."[13]

To be autonomous, to be self-legislating, means to be the normgiver and subject of the norm at the same time. To apply this to the widespread talk of autonomous driving, this would mean: the autonomous car would not only have to be able to choose (derive or "decide") between the fastest, shortest, and, according to other criteria, preferable options (as well as between, in every available respect, equivalent alternatives) of routes, distances, or concrete control options, but would also have to be able to independently change, reject, and question the goal, the order, or the criteria mentioned. An autonomous car in this sense would have to be able to make statements like: "I don't want to go to Munich today. I'd rather rev to 4000/min (that vibrates so nicely). I'm not starting the engine now, I'm taking a break today." It is obvious that in this sense, autonomous systems with a will, norm-setting competence, and refusal authority are nonsensical and no longer useful – except perhaps as strong AI boundary explorations.

Technical autonomy refers to a variety of capabilities or functionalities of technical systems. These include, for example, the abilities to a) function autonomously without continuous supply of external energy or other resources, b) move independently of humans and execute orders without further intervention, c) select goal-oriented strategies, paths and means for an order once it has been received (but not reject or change the orders themselves!), d) adaptively orient this choice to varying contexts in different situations, e) execute the order in spite of environmental or other factors, d) to adaptively orient this choice to varying contexts in different situations, e) to align the fulfillment

6
Carl Schmitt, *Political Theology. Four Chapters on the Concept of Sovereignty*, transl. George Schwab, The University of Chicago Press, Chicago – London – Bristol, 2005, p. 5.

7
Ibid., p. 15.

8
For instance, in the context of digital sovereignty: "[T]he ultimate form of control is *individual sovereignty*, understood as self-ownership, especially over one's own body, choices, and data." – Luciano Floridi, "The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU", *Philosophy and Technology* 33 (2020), pp. 369–378, here p. 371, doi: https://doi.org/10.1007/s13347-020-00423-6.

9
C. Schmitt, *Political Theology*, p. 15.

10
Nathalie Nevejans, *European Civil Law Rules in Robotics*, Directorate-General for Internal Policies, European Parliament, 2016, p. 15, doi: https://doi.org/10.2861/946158.

11
Ibid., p. 9.

12
Erich Fromm, *Man for Himself. An Enquiry into the Psychology of Ethics*, Routledge & Kegan Paul, London 1949, p. 8.

13
Ibid., p. 9.

of the order with internal states and specifications in spite of environmental and context sensitivity, f) to successively change these internal states, i.e. to innovatively align the choice of means and ways, the interpretation of the context, etc. with the failure and success of past attempts, i.e. to learn from initial specifications and factory settings.[14]

Automatic vs. Autonomous

In the technical context, *autonomous* should be distinguished from *automatic*. To avoid misunderstandings, many of the capabilities that are currently described as autonomous would be more accurately described as automatic or gradually (partially/fully) automated.[15] *Automatic* can mean a) to function independently without human control or to execute processes from start to finish, b) to fire continuously (in the case of weapons), c) without conscious deliberation, d) to occur spontaneously, e) to follow a fixed rule. An *automatic* or *automated system* can be defined as a system that has some predetermined alternative processes programmed into it, which the system executes in a rule-based manner in response to some sensor data. Its output is predictable (given conditions assumed to be constant) if the predetermined rule set is known. Accordingly, automated systems always operate under predetermined rules which makes them heteronomous by definition, e.g. of Kant:

"If it is *in anything other* than the fitness of its maxims for its own universal legislation, […] the outcome is always *heteronomy*."[16]

Automated processes differ significantly from autonomous actions in their predictability, as well as in the fact that an automatic (spontaneous, reflexive, rule-based) reaction to input and environmental variance precisely involves no conscious decision-making, no room for redecision-making instructed by learning, no intelligence performance, and certainly no recognition performance vis-à-vis rule specifications or normative refusal freedom vis-à-vis orders. In this context, another metaphor plays a complicating role: machine "decision-making".

A commonly used definition of automation is Level 10 on the Sheridan-Verblank scale (see below), which reads, "computer does whole job if it *decides* it should be done, and so tells human if it *decides* he should be told".[17] Automated and autonomous are often used interchangeably or get confused with each other. For example, the Sheridan Scale is cited as the "Sheridan Scale for Autonomy" in the context of unmanned vehicles,[18] although this scale structures automation (and not autonomy), hence the title "Levels of Automation in Man-Computer Decision-Making".[19] This and similar confusions are probably due to the aspect of "decision-making" that can be considered the core of current technical definitions of autonomy. Sheridan correctly speaks of automation, however somewhat misleadingly metaphorically of decision making:

"In the fullest contemporary sense, the term *automation* refers to a. the mechanization and integration of the sensing of environmental variables (by artificial sensors), b. data processing and decision making (by computers); c. mechanical action (by motors or devices that apply forces on the environment), and/or d. "information action" by communication of processed information to people."[20]

A contrasting example, in which machine decision-making functions precisely as a demarcation of automation and autonomy, is provided by Stenger *et*

al.,[21] according to which an automatic system is capable of pre-programmed task execution and, in doing so, precisely cannot take any environmental context into account and cannot *decide* between different options.

"An autonomous system on the other hand has the capability to select among multiple possible action sequences in order to achieve its goals."[22]

This 'decision' of alternative choice is based on knowledge of internal and external states as well as internally defined criteria and rules.

"The challenge is to find not just any solution to a problem, but a good or ideally the best one."[23]

Deciding here represents an anthropomorphism. Human and technical decision-making differ, as do human and technical autonomy. Accordingly, the attempt to explain "autonomous systems are systems that can decide between different options" risks trying to explain something unclear by something equally unclear (*obscurum per obscurius*). The determination of technical autonomy via the capability of machine decision-making (central to artificial intelligence) reproduces the anthropomorphism of the concept of technical autonomy on the explanatory level of technical decision-making capability: human decision-making involves the normatively justified choice of one option over less wanted alternatives and thus depends on the normative positionality and will of the decision-maker. Already the concept of an alternative (being a viable option) implies normative judgments because one is never running out of options, it is always possible to just blink, sit down or hold one's breath, but those actions seldom are considered viable alternatives in concrete acting situations that aim at pursuing a goal or solving a problem; options or alternatives are already normatively judged as relevant, useful,

14
Niels Gottschalk-Mazouz, "'Autonomie' und die Autonomie 'autonomer technischer Systeme'", *XXI. Deutscher Kongress für Philosophie. Lebenswelt und Wissenschaft*, 2008. See: http://www.dgphil2008.de/fileadmin/download/Sektionsbeitraege/07_Gottschalk-Mazouz.pdf (accessed on 1 August 2024).

15
Bruno Gransche, "Handling Things that Handle Us: Things Get to Know Who We Are and Tie Us Down to Who We Were", in: Heather Wiltse (ed.), *Relating to Things. Design, Technology and the Artificial*, Bloomsbury Visual Arts, London 2020, pp. 61–80, here p. 64.

16
I. Kant, *Groundwork of the Metaphysics of Morals*, p. 111.

17
Thomas B. Sheridan, William L. Verplank, *Human and Computer Control of Undersea Teleoperators* (MIT Man-Machines-Systems Laboratory Report), MIT, Cambridge (MA) 1978, pp. 8–19, my italics.

18
Somaiyeh MahmoudZadeh, David M. W. Powers, Reza Bairam Zadeh, *Autonomy and Unmanned Vehicles. Augmented Reactive Mission and Motion Planning Architecture*, Springer Singapore, Singapore 2019, p. 11. https://doi.org/10.1007/978-981-13-2245-7

19
Th. B. Sheridan, W. L. Verplank, *Human and Computer Control of Undersea Teleoperators*, my italics.

20
Thomas B. Sheridan, Raja Parasuraman, "Human-Automation Interaction", *Reviews of Human Factors and Ergonomics* 1 (2005) 1, pp. 89–129, doi: https://doi.org/10.1518/155723405783703082.

21
A. Stenger, B. Fernando, M. Heni, "Autonomous Mission Planning for UAVs: A Cognitive Approach", *Deutscher Luft- und Raumfahrtkongress 2012*, Document ID 281398, pp. 1–10, here p. 1. Available at: https://www.dglr.de/publikationen/2013/281398.pdf (accessed on 1 August 2024).

22
Ibid.

23
Ibid.

viable possibilities, and not just doing anything. Just like alternatives, goals, and problems again both imply normative judgment, etc. Since machines do not have *problems* (they do not prefer one event or state of the world over another), they cannot judge viability in options, desirability in goals, or problematicity by self-chosen standards; they just adjust given options according to given criteria and rules and select them for given problems. Genuine (autonomous) value preferences do not come to technical systems just as little as an own will.[24] Systemic "decision-making" is, therefore, a – possibly unpredictable not ex-post explicable – application of given normative decision criteria (often these are effectiveness, efficiency, and freedom from interference).

The Air Force Research Laboratory shows an example in which technical autonomy is to be distinguished from automation and for this purpose autonomous technology is attributed a – philosophically preconditioned – free will:

"Automatic means that a system will do exactly as programmed, it has no choice. Autonomous means that a system has a choice to make free of outside influence, i.e., an autonomous system has free will."[25]

Here, either the technical property of option selection is referred to metaphorically as "free will" or automation has been "accidentally" distinguished from human rather than technical autonomy (and then "system" was used metaphorically or in a biomorphism, as in humans as biological systems), which does little to illuminate technical autonomy. Applying decision-making, free choice, and free will to technology represents – like autonomy, intelligence, etc. – "semantic trickery".[26] In this context, the route selection of an autopilot, for example, is precisely not an autonomous decision, but an option selection according to external rules and criteria, i.e. according to laws that are heteronomous for the system. Even if the system can change these rules and criteria "itself", i.e. the criteria of the selections, by machine learning, it can only do so according to external rules and criteria for machine learning. Autonomous systems differ from automatic systems in that they can pursue heteronomous goals with independent selections of means and strategy. They differ from autonomous humans in not being able to make autonomous goal decisions. Therefore, the *goals* ("its goals"[27] are not the goals that the system would have set for itself, but the ones that have been pre-set or input to it by a normative authority. Accordingly, the *internally defined criteria and rules* that, together with the *detected situation*, ground the system selections[28] are not self-selected but predetermined criteria and rules. It follows that the choice of a problem solution as *good* or *best*[29] can only be decided by the normgiver human being, but is only executed by the system under comparison of default and situation detection.

In the field of *unmanned systems* (UMS), technical autonomy is defined as:

"A UMS's own ability of integrated sensing, perceiving, analyzing, communicating, planning, decision-making, and acting/executing [...], to achieve its goals as assigned by its human operator(s) through designed Human-Robot Interface (HRI) or by another system that the UMS communicates with."[30]

Here, at least, it is clearly emphasized that "its goals" can only mean "externally assigned goals", which avoids the misunderstanding of technical self-legislation. This is a misunderstanding that, for example, Floridi and Sanders however are under when they try to counter the objection that agents must have goals for action by defining 'having goals' as "including goal-oriented behavior".[31] Showing mere 'goal-oriented behavior' does not suffice as a

criterion of autonomy, for agents could be oriented to given goals (however selected or predetermined), which would still be heteronomy and not count as a criterion for autonomy. Autonomous agents have goals that they have chosen as their own and recognize them as chosen by themselves, they subject themselves to self-given laws; in this full sense this does not apply to artificial agents and the criterion remains valid that 'having their own goals' or being normgiver and not just norm follower is decisive for autonomous agents.

An essential difference here is that systems can *detect* goals, criteria, and rules and align their processing accordingly, but they cannot self-reflexively *acknowledge themselves as subjects of recognition* on a higher level, i.e., they cannot decide on their recognition themselves. Therefore, systems – in contrast to humans – can neither refuse the recognition of a goal as binding for themselves, nor change the goal, as a result of such a refusal of recognition, and thus autonomously pursue other, self-set/self-chosen goals recognized as their own. According to Christoph Hubig, technical systems "may well have a representation of rules (possibly also as a self-formed representation)" – "possibly also a representation of themselves as bearers of the representation [...], but not a self-representation as the subject of the *recognition* or *rejection* of the representations".[32]

Multi-Layered Views on Autonomous Systems

To avoid conceptual confusion and the resulting misunderstandings, it has been suggested – that it is better not to apply terms coined in relation to

24
The ascription of "free will" and "genuine autonomous value preferences" to people does not imply that people can or will chose completely unconditioned, as if absolutely nothing influences their preferences and choices. But accepting that people are biologically, socially, psychologically etc. influenced in what they chose, does not imply that they are determined to only ever chose according to those influences. This is especially apparent since the plethora of influences come with contradictions; deciding socially appropriate might quite often imply deciding against biological imperatives. The exception proves everything (see Carl Schmitt above).

25
Bruce T. Clough, "Metrics, Schmetrics! How The Heck Do You Determine A UAV's Autonomy Anyway?", *Proceedings of the 2002 Performance Metrics for Intelligent Systems Workshop*, Air Force Research Laboratory, 2002, p. 1.

26
Klaus Kornwachs, "Der Herr der Dinge oder warum wir unsere Geschöpfe an die Hand nehmen sollten", in: Welf Schröter (ed.), *Autonomie des Menschen – Autonomie der Systeme. Humanisierungspotenziale und Grenzen moderner Technologien*, Talheimer, Mössingen-Talheim 2017, pp. 15–65, here p. 15.

27
A. Stenger, B. Fernando, M. Heni, "Autonomous Mission Planning for UAVs", p. 1.

28
Ibid.

29
Ibid.

30
Ad Hoc Autonomy Levels for Unmanned Systems Working Group Participants, Hui-Min Huang (ed.), *Autonomy Levels for Unmanned Systems (ALFUS) Framework. Volume I, Terminology (Version 2.1)*, NIST Special Publication 1011-I-2.0, National Institute of Standards and Technology 2008, p. 15. Available at: https://www.nist.gov/document/nistsp1011-i-2-0pdf (accessed on 1 August 2024).

31
Luciano Floridi, J. W. Sanders, "On the Morality of Artificial Agents", *Minds and Machines* 14 (2004) 3, pp. 349–379, here p. 365, doi: https://doi.org/10.1023/b:mind.0000035461.63578.9d.

32
Christoph Hubig, *Die Kunst des Möglichen II. Grundlinien einer dialektischen Philosophie der Technik. Macht der Technik*, transcript Verlag, Bielefeld 2015, p. 131.

humans, such as autonomy, decision-making, intelligence, consciousness, learning, etc., to technical systems[33] or to mark such an application as a metaphoric transfer or non-literal use either by quotation marks or in combination with *quasi-*. Given the enormous proliferation of unmarked anthropomorphisms – artificial intelligence, machine learning, autonomous robots, etc. – claims of avoidance or marking seem unpromising. To differentiate different types, properties, or capabilities of automated/ autonomous technology, various multilevel or scaled proposals in particular have gained acceptance.

A kind of classic is the above-mentioned "Levels of Automation in Man Computer decision-making" by Sheridan and Verblank from 1978 differentiating 10 levels:

"1. human does the whole job up to the point of turning it over to the computer to implement. 2. computer helps by determining the options [...] 10. computer does whole job if it decides it should be done, and if so tells human if it decides he should be told."[34]

Often (although inconsistently so)[35] the highest level or levels of automation are again referred to as autonomous, using highly or fully automated as a synonym for technically autonomous.

Examples of multidimensional grading attempts of autonomy/ automation would be the "Levels of Autonomy Assessment Scale" by Proud *et al*. or the "Autonomy Control Level Framework" by Clough, which respectively link eight and eleven degrees of automation with the four dimensions of Boyd's OODA loop (Observe, Orient, Decide, Act). or the "Levels of Automation applicable to dynamic-cognitive and psychomotor control task performance" by Endsely and Kaber, which assign ten levels from *1. manual control* to *10. full automation* to the four dimensions of *monitoring, generating, selecting,* and *implementing*.[36] This already indicates that it is problematic to characterize a system as a whole as autonomous or automated, since different functionalities or the OODA dimensions can be automated to varying degrees in the same system, depending on the mission phase.

Finally, approaches should be mentioned that move further away from the attribution of autonomy[37] as a system property and include more context-related variables of system use in a relational determination. Exemplary could be the ALFUS (Autonomy Levels for Unmanned Systems) model,[38] which spans the dimensions of *environmental difficulty*, *mission complexity*, and *human interface* as axes of an assessment space. According to ALFUS, systems are autonomous if they can handle complex tasks in difficult situations without human intervention. The fact that even the most 'autonomous' systems do not choose those tasks themselves and do not recognize them as worth pursuing, but that those tasks are given to them heteronomously by human authority, still distinguishes the so-called "fully autonomous" level from human autonomy in the full sense. To handle complex tasks in difficult situations without human intervention does not mean choosing what task to tackle in the first place.

The multi-level model of autonomy and control in human-technology interactions or relations AMTIR by Gransche *et al*. locates many of the aforementioned level models entirely in two of its three levels and demarcates from them a third level for a normative type of autonomy.[39] The three types of autonomy of this model are combined with three types of control and are divided into normative (1), strategic (2), and operational (3) autonomy and control: Autonomy (1) means the freedom of intentionality and of choosing, recognizing, rejecting, or setting purposes (or in terms of technology: tasks,

goals, missions). This is the normative autonomy in the narrower sense of self-legislation of recognized rules of lawgiving-law subjects.[40] The associated level of normative control asks what is morally commanded, permissible, or prohibited and orients and controls decisions at level 2 accordingly. Autonomy (2) means the freedom to decide on and guarantee strategies (including the power of control) to fulfill level 1 purposes. The choice of appropriate ways (2) here regulates the choice of means (3) and is itself oriented to the given goal (1); it can be delegated to technical systems, in contrast to autonomy (1).

"Such delegations often take place in the context of a division of labor between systems and the subjects dealing with them; here, margins are defined within which 'autonomous' [...] decisions can be made."[41]

This correlates with notions of semi- or shared autonomy, or with the understanding of autonomy that underlies many technical definitions, where "deciding", as far as the latitudes are granted to technical systems, is to be understood metaphorically. The associated control (2) constitutes an elementary system performance itself or a performance of system designers mediated through systems.

"It is referred to as control (containment, disturbance feedforward control, feedback mechanisms)."[42]

33
Patrik Stensson, Anders Jansson, "Autonomous technology – sources of confusion: A model for explanation and prediction of conceptual shifts", *Ergonomics* 57 (2014) 3, pp. 455–470, here p. 456, doi: https://doi.org/10.1080/00140139.2013.858777.

34
Th. B. Sheridan, W. L. Verplank, *Human and Computer Control of Undersea Teleoperators*.

35
The following approaches demonstrate the inconsistent use of *autonomous* and *automated*: U.S. Navy Office of Naval Research scale, for example, includes 6 levels: *1. Human operated, 2. Human assisted, 3. Human delegated, 4. Human supervised, 5. Mixed initiative*, and *6. Fully autonomous*. An Air Force scale provides for 10 levels from *0. Remotely piloted vehicle, 1. Execute preplanned mission* to again *10. Fully autonomous*. See for these and others: Andrew P. Williams, "Defining Autonomy in Systems: Challenges and Solutions", in: Andrew P. Williams, Paul D. Scharre (eds.), *Autonomous Systems. Issues for Defence Policymakers*, Innovation in capability development, vol. 2, Norfolk, Va. 2016, pp. 27–63, here p. 41–44.

36
Mica R. Endsley, David B. Kaber, "Level of automation effects on performance, situation awareness and workload in a dynamic control task", *Ergonomics* 42 (1999) 3, pp. 462–492, here p. 464, doi: https://doi.org/10.1080/001401399185595.

37
See: Karsten Weber, "Autonomie und Moralität als Zuschreibung. Über die begriffliche und inhaltliche Sinnlosigkeit einer Maschinenethik", in: Matthias Rath, Friedrich Krotz, Matthias Karmasin (eds.), *Maschinenethik. Normative Grenzen autonomer Systeme*, Springer VS, Wiesbaden 2019, pp. 193–208.

38
Ad Hoc Autonomy Levels for Unmanned Systems Working Group Participants, Hui-Min Huang *et al.* (eds.), *Autonomy Levels for Unmanned Systems (ALFUS) Framework. An Update*, 2005 SPIE Defense and Security Symposium, Orlando 2005.

39
Cf. Ch. Hubig, *Die Kunst des Möglichen III*, pp. 131–135.

40
Bruno Gransche *et al*., *Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen. Grundsatzfragen autonomieorientierter Mensch-Technik-Verhältnisse*, Frauenhofer Verlag, Stuttgart 2014, p. 49.

41
Ibid., p. 50.

42
Ibid.

The level of autonomy (3) is about freedom(s) of action in the sense of choosing, setting, and using appropriate means, and "control (3) as operational control refers to the control itself, which can be corrected, strengthened, weakened, or prevented if the choice or the way of using a means does not appear to be purposeful".[43] Depending on the level of automation and system functionality, operational freedom (e.g., choice of exact braking force, turning/spinning speed, etc.) can be delegated to machines, which – possibly equipped with sensors and capable of learning – may show operational autonomy when, for example, adapting the exact choice of means to environmental conditions or optimize the concrete suitability of certain options to given strategies from previous results. For example, an autopilot may be operationally or strategically autonomous, depending on whether it (only) follows a predefined route and coordinates all means necessary to do so, or whether it can select (not decide!) an 'optimal route' (according to level 1) depending on the environmental situation at hand and then follow it. Again, the option selection can only be called 'deciding' in a metaphorical sense.

This multi-level model of autonomy and control ensures – similar to[44] – the philosophically unique position of human autonomy (1), but proposes as a proposed solution for autonomy problems on the technology side a type differentiation of autonomy. In the literal sense, only type 1 means *autonomy*, while types 2 and 3 are already to be understood metaphorically; precisely for this reason, however, they can also be delegated to technology and thus thought of in terms of human-machine interaction, hybrid autonomy, etc. The decision to call levels 2 and 3 *autonomy* nevertheless – under the risk of anthropomorphization and thus misunderstanding – is to be understood as a step towards interdisciplinary connectivity with the tendencies of technical autonomy described above. Similarly, the AMTIR authors propose an anthropomorphization- and metaphor-aware differentiation of levels along with a set of proposals for problem solutions in the AMTIR heuristic.[45]

This overview of the concept of autonomy both in the exemplary moral philosophy of Kant[46] and in technical contexts reveals the differences between human and technical autonomy up to a point that unmasks their mere equivocation and their metaphorical use as label transference. However, the label "autonomy" is transferred from humans to artificial systems, and no new *terminus technicus* was specifically coined to refer to whatever is meant exactly when technical systems are being called *autonomous*. The current debate would surely look rather different if respective IT systems would have been (more fittingly) labeled *artificially rational* instead of *intelligent*, or if no one ever transferred the word autonomous to technical systems and only referred to those systems as *highly automated* or *remote/cable/pilot free*, *unmanned*, etc. If that were the case, there were no autonomous robots and intelligent chatbots today; primarily not because we referred to those systems differently but partly also not – given the influence of metaphors, narratives, concepts, etc. for innovation, invention, and research processes – because these alternative denotations probably would have caused different research agendas resulting in different artifacts (but that would be a fictional thought experiment).

The actual current metaphorization has consequences, especially as the origin of the label still orients the understanding of the recipients who are then charged with the difficult task of discerning which property belongs to and should stay with the old or hitherto standard meaning and which property can be transferred and reused in the new meaning. This is not only the case

with 'autonomy' itself but also with most of the explanatory concepts used to explain technical autonomy like deciding, learning, intelligence, knowing, training, evolving, etc. All these transferences of human-specific concepts to artificial systems are anthropomorphisms or biomorphisms, meaning that they describe technologies in terms of living or even rational beings, or they describe mere subjects of norms in terms that are standardly used for normgivers, thus fueling misconceptions of literal rationality or free will in technology. The 'semantic trickery' is performed mostly by way of metaphorization.

Metaphorization – to Be *and* Not to Be

The metaphorical configuration excludes absolute sameness, yet it also excludes absolute differentness. The foundation of metaphors[47] is similarity, which is a double relation of simultaneous sameness and differentness. Two entities are similar, only if they share identical aspects and if they differ in other aspects. Theoretically, for probably any pair of entities there can be assumed identical aspects, be it only in their spaciality, timeliness, or change dynamic, be it insofar they share cultural situatedness, social relevance, symbolic proximity, etc. This is why Donald Davidson called this similarity aspect trivial:

"It is trivial because everything is like everything, and in endless ways."[48]

On which identical aspects in particular and in which regard metaphors focus is not trivial though and depends on their purpose, the intended effect, and the poetic as well as philosophical mastery of the locutor.

"'In philosophy also,' adds [Aristotle in, BG] the *Rhetoric*, 'an acute mind will perceive resemblances in things far apart. […] you might say that an anchor and an overhead hook were the same, since both are in a way the same, only the one secures things from below and the other

43
Ibid.

44
Michael Funk, Mark Coeckelbergh, "(Technical) Autonomy as Concept in Robot Ethics", in: José Luis Pons (ed.), *Inclusive Robotics for a Better Society. Selected Papers from INBOTS Conference 2018*, *16-18 October, 2018, Pisa, Italy*, Springer International Publishing, Wiesbaden 2019, pp. 59–65.

45
B. Gransche *et al*., *Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen*.

46
Even if one is not willing to follow Kant's strong stance on personal autonomy or freedom of will (with good reasons), this position serves here as a contrasting point to technical autonomy. Many factors can of course be considered that condition human will or autonomy or that ground human reason in a mammal body or human *Leib* with nonrational animalistic reflexes or mechanistic repetitive behavior etc. (many of which Kant tackles), yet as established above, this does not approximate human autonomy to technical autonomy, but rather distinguishes it further from it. Mechanically rule-following people do so by omission (voluntarily, intentionally, or not) of making themselves the exception, not by inability to do so. It is autonomy-wise a huge difference, if you refrain from deviating from a given rule out of *fear* of following repercussions, out of *lack of imagination* of alternatives, or simply because rejecting a rule is not an option by *design*.

47
"Metaphors" is interchangeably used with "metaphorical statements" here.

48
Donald Davidson, "What Metaphors Mean", *Critical Inquiry* 5 (1978) 1, pp. 31–47, here p. 38.

from above' (1412 a 10–15). To apprehend or perceive, to contemplate, to see similarity – such is metaphor's genius-stroke, which marks the poet, naturally enough, but also the philosopher."[49]

According to Ricœur a metaphor such as "nature is a temple" shows this double relation of sameness and differentness in a tension of being and not being in the bridge-indicating verb "is".

"The same tension rests in the verb to be in the metaphorical statements. The 'is' is at the same time a literal 'is not' and a metaphorical 'is like' ... Nature, taken literally, is not a temple; but metaphor teaches us to see it like a temple."[50]

The "is like" actually makes a simile, whereas the metaphor drops even the "like" and identifies the configured entities. Aristotle's "A simile is also a metaphor. For the difference is small." is explicated by Ricœur:

"We shall say that simile explicitly displays the moment of resemblance that operates implicitly in metaphor."[51]

So, nature metaphorically "is" a temple regarding their shared identical aspects and at the same time nature literally "is not" a temple regarding their differences. The metaphor now, having the identity perspective as its characteristic form, emphasizes the identical in spite of the differences.

"In metaphor, 'the same' operates *in spite of* 'the different'."[52]

Yet, it does not negate the differences, which would destroy its working condition of similarity (turning it into mere sameness). Therefore, in terms of metaphor, criticizing that there are differences does not criticize the metaphor, rather stating that there are none would. By being a similarity configuration and an identity emphasis in spite of preserved differences, metaphorization *creates* different-sameness or same-differentness, that is similarity, "it is more enlightening to say that the metaphor creates the similarity than that the metaphor gives verbal form to some pre-existent similarity".[53]

"'Bringing close' what was 'far'; this is the work of similarity. In this sense, Aristotle was right when he said that 'to bring about metaphors is to see the sameness'; but this seeing is simultaneously a making. Good metaphors are those that create a similarity more than they trace it."[54]

In this regard metaphors are dialectical because they can be seen as a chain of label transfers:

"The metaphor is nothing else than the assignment of a known label with a certain past on a new object, which first resists this transfer, then gives in."[55]

This transfer (at first a *miss assignment* then a *re-assignment*[56] takes a certain label like "autonomous" and sticks it to a new entity e.g. from a) states to b) persons to c) artificial agents. Each time this transference transforms in the triple Hegelian sense (of *Aufheben*) the meaning of the label. Firstly, it *negates* the previous meaning (autonomous persons *are not* legislating citizens; autonomous systems *are not* rational lawgivers). Secondly, it *preserves or conserves* the previous meaning (autonomous people are also state-like lawgivers; autonomous systems are also self-determined or not-remote-controlled, etc.). Thirdly, it *elevates* the previous meaning to a new configuration with the word "like" as a compound similarity marker (autonomous people give themselves laws *like* states and follow them *like* citizens in one non-collective instance; autonomous systems alter their processes and some rules or thresholds *like* people but do so without a free will *like* other things). The metaphor combines these elements of *Aufheben* as a) "is not", b) "is", and c) "is like".

Metaphorized Autonomy

To grasp *autonomy* or *intelligence* in its various spheres as a metaphor allows us to focus on their similarity. This helps to understand its function in different discourses, be it public, political, or scientific. As seen, a metaphor has the identifying bridging verb "is" at its core, yet artificial intelligence and autonomous systems do not. However, it is easy enough to rephrase those configurations in a form that allows it to more explicitly see the metaphorization at work. *Autonomous robots* can serve as an instance of autonomous technology, which then can be rephrased as: "This robot is an autonomous entity." Explicating the metaphor as a simile one could add: "like other autonomous beings such as humans". Many voices in today's debate tend to use and read such a phrase as mere identifications and not as metaphorical emphases on identification (that preserves the differences). This leads to misjudgments of the autonomy of robots in an anthropomorphic way that transfers not only aspects like limited independence in operating and including sensor information even without human intervention for a certain time, but aspects that were part of the previous label standard of 'autonomy' like free will, rationality, etc. E.g.

"…an autonomous system has free will."[57]

This is indeed apparent in most of the strong AI imaginary as in debates about electronic personhood or person rights,[58] robotic citizenship,[59] themes like malevolence, world domination, the enslavement of humanity,[60] etc. The case

49
Paul Ricœur, *The Rule of Metaphor. The Creation of Meaning in Language*, transl. Robert Czerny, Routledge, London 2003, p. 30.

50
Paul Ricœur, Eberhard Jüngel, *Metapher. Zur Hermeneutik religiöser Sprache*, Kaiser, München 1974, p. 48.

51
Ibid., p. 30. Dropping the "like" in metaphors emphazises sameness as opposed to similes yet does actually means sameness. See also Davidon on the metaphor-simile difference: "The most obvious semantic difference between simile and metaphor is that all similes are true and most metaphors are false. The earth is like a floor, the Assyrian did come down like a wolf on the fold, because everything is like everything. But turn these sentences into metaphors, and you turn them false; the earth is like a floor, but it is not a floor […]. We use a simile ordinarily only when we know the corresponding metaphor to be false. We say Mr. S. is like a pig because we know he isn't one. If we had used a metaphor and said he was a pig, this would not be because we changed our mind about the facts but because we chose to get the idea across a different way." – D. Davidson, "What Metaphors Mean", p. 41.

52
P. Ricœur, *The Rule of Metaphor*, p. 232.

53
Ibid., p. 100.

54
P. Ricœur, E. Jüngel, *Metapher*, p. 48.

55
Ibid., pp. 52–53.

56
Ibid., p. 53.

57
B. T. Clough, *Metrics, Schmetrics!*, p. 1.

58
Susanne Beck *et al*., "Brauchen wir eine Roboterethik? Eine Kontroverse", in: Alexander Friedrich *et al.* (eds.), *Steuern und Regeln, Jahrbuch Technikphilosophie*, Nomos, Baden-Baden 2019, pp. 231–270.

59
FAZ, "Roboterfrau bekommt saudische Staatsbürgerschaft", *Frankfurter Allgemeine Zeitung* (27 October 2017). Available at: https://www.faz.net/aktuell/wirtschaft/digitec/roboter-sophia-bekommt-saudi-arabiens-staatsbuergerschaft-15265867.html(accessed on 1 August 2024).

60
As a meanwhile classical voice consider Moravec's *Mind Children*: "What awaits is →

of *artificial intelligence* is very similar to that of *autonomous technology*, for autonomous systems need artificial intelligence functionalities or need to be built using techniques and concepts that are employed in AI as well. A more explicit metaphorization would be: "Those systems (we refer to as AI systems) are intelligent entities (like humans or other intelligent beings)." Again, mere identity would be a misunderstanding, for all the aspects of e.g. intelligent humans that should not be transferred to intelligent systems (their differences specific for *humans*) like emotionality, sociality, embodiment (*Leiblichkeit*), culturality, etc. would be transferred nonetheless. In fact, in current discourses they often are transferred and even more often they are strategically triggered or invited to be transferred (e.g., anthropomorphic technology design, technoanimism,[61] parasocial relations with robots, etc.).

Metaphor, Ricœur said, teaches us to see nature as a temple, or transferred to this topic, teaches us to see robots as self-legislating or normgivers, it teaches us to see IT systems as intelligent. The instructive power of metaphorization was already emphasized by Aristotle:

"Metaphor, by contrast, most of all produces learning."[62]

Although Davidson insisted "Seeing as is not seeing that.",[63] this point does not undermine the teaching power of metaphors, for *seeing two entities as* similar, emphasizing their identical aspects, does not mean that it can no longer be *seen that* they are different. In metaphorization, both identity and difference are dialectically combined and elevated (*aufgehoben*). The metaphor "Such artificial systems are autonomous entities." conveys that these systems *are not autonomous* (as rational normgivers) as well that these systems *are autonomous* (as in more self-determined than 20th-century cars). Seeing that autonomous robots and autonomous humans are different does not hinder but complement seeing them as similar. This means that autonomous robots (aR) *are not* autonomous humans (aH) *and* that aR *are* aH, which in turn means – identity is shared on both sides obviously – that aH *are like* aR.

Following Hans Blumenberg, *technical autonomy* could be seen as a *concept* (*Begriff*); concepts are not directed towards present or past things or events but especially towards future ones, they are "instruments of possibility and prevention".[64] In this view, technical systems are not being labelled autonomous to deal with their presence, but rather to prepare for, advance or prevent their possible advent. Then "autonomous systems" could refer a) in a reduced sense to existing technical systems like drones and at the same time b) to possible future systems with "fuller human-like yet technical" autonomy. The advent of such systems in a hypothetical sense of b) is for various reasons (partly established above) highly unlikely or outright impossible, yet these future imaginaries of "fully autonomous systems" are influential as *present futures* – even if they never become a *future present*.[65] As concepts (*Begriffe*) – according to Blumenberg – introduce possibility into our consciousness (e.g. the possibility of personal-autonomous robots), the difference between possible and real can only be grasped by probing the possibilities (offered by the concepts) and negating what is not real:[66] So, if you think you met an autonomous robot (in the full philosophical sense established), probe its possibility, put it to the test until you either unveil the simulation and negate that it is being real (which is until today entirely the case) or until you determine that it is being real (as mentioned: unlikely). There remains another option following Blumenberg: technical personal autonomy could be considered an *idea* (a

mere concept of reason in the Kantian sense) – more precisely a subtype of the idea of freedom, which Kant sees as a necessary precondition for reason – that has no direct empirical or perceptual presence. *Technical autonomy* can be seen as metaphoric with Blumenberg because the word autonomy is being used in a different context (technical systems rather than people) where it is a semantic anomaly and used against the rules of its possible usage (figuratively) and thus gains another context-dependent meaning. Metaphors (being semantic anomalies) resist their transformation in expressive means that perfectly fit their context, they resist their re-transference in literal expression. Blumenberg defines an "absolute metaphor" as

> "… the transportation of the reflection on one object of intuition to another, quite different concept, to which perhaps no intuition can ever directly correspond.' […] That these metaphors are called 'absolute' means only that they prove resistant to terminological claims and cannot be dissolved into conceptuality."[67]

It could be argued that freedom (being an idea) and with it autonomy never had a pre-transfer literal context as an object of perception or that its origin has long been forgotten in the long chain of context migrations.[68] Then autonomy in the context of technical systems would not be an ordinary metaphor that we could (partly) retransfer in literal definitions to somehow understand its meaning adequately or orient our expectations when talking about associated phenomena (like drones or robots), not even about possible future ones (like postbiotic super-AIs), but it would be an idea, a concept of reason. Seen as absolute metaphor "autonomy" in the context of technology then serves as

not oblivion but rather a future which, from our present vantage point, is best described by the words 'postbiological' or even 'supernatural.' It is a world in which the human race has been swept away by the tide of cultural change, usurped by its own artificial progeny." – Hans Moravec, *Mind Children. The Future of Robot and Human Intelligence*, Harvard University Press, Harvard 1990, p. 1. As a recent statement consider the Open letter on AI from the Future of Life Institute from 2023: "Contemporary AI systems are now becoming human-competitive at general tasks, and we must ask ourselves: […] Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization?" – Future of Life Institute, "Pause Giant AI Experiments: An Open Letter". Available at: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (accessed on 1 August 2024).

61

Bruno Gransche, "Technogene Unheimlichkeit", in: Alexander Friedrich *et al*. (eds.), *Autonomie und Unheimlichkeit. Jahrbuch Technikphilosophie*, Nomos, Baden-Baden 2020, pp. 33–51.

62

Aristotle, *Rhetoric*, transl. C. D. C. Reeve, Hackett Publishing Company Incorporated, Indianapolis 2018, 1410b 12–14.

63

D. Davidson, "What Metaphors Mean", p. 47.

64

Hans Blumenberg, *Paradigms for a Metaphorology*, transl. Robert Savage, Cornell University Press, Ithaca 2016, p. 12, 17.

65

Bruno Gransche, *Vorausschauendes Denken. Philosophie und Zukunftsforschung jenseits von Statistik und Kalkül*, transcript Verlag, Bielefeld 2015, pp. 92–97.

66

H. Blumenberg, *Paradigms for a Metaphorology*, p. 75.

67

Ibid., pp. 4–5.

68

"… metaphors that have become worn out and have been drained of sensuous force, coins which have lost their embossing and are now considered as metal and no longer as coins." – Friedrich Nietzsche, *Philosophy and Truth. Selections from Nietzsche's Notebook of the early 1870's*, transl. Daniel Breazeale, Humanities Press International, New Jersey 1992, p. 84.

a proxy for something basically non-conceptual, not otherwise perceivable, or thinkable. Absolute metaphors lack context determination or context-specific properties that could hinder its context-migration and resist its proxy role. Therefore, especially concepts of reason (ideas) and absolute metaphors condition our understanding of phenomena without being able to be dissolved in conceptuality or retransferred into literality. Calling for such a dissolution or re-transference, then, is bound to fail; which does not make the explicit reflection on its metaphorization futile, but rather indispensable.

To teach seeing human autonomy as similar to robot autonomy by way of the widespread usage of the "technical autonomy" metaphorization equals to teach seeing human autonomy as technical autonomy in turn and thus to gradually teach a technicist understanding of human autonomy with its implications of rule-following, process optimization, training data reference, calculative rule evaluation, etc. even in humans. That is vastly different from a strictly human autonomy as (widely conditioned yet) self-recognized maxim choices, moral reflection on possible universalizability, individual suspension of rule application for oneself i.e. temporary rule-suspension or long-term rule transformation, creativity, empathy, etc. The possible convergence of human and technical autonomy due to the massive usage of identity-teaching metaphors leads to a humanization of artificial systems and a technization of human beings – the latter being by far the more problematic effect. This convergence can mean a) that we actually have "fully autonomous" technical systems (see above mentioned scaled approaches) because the technical performance reached that level (see arguments against this diagnosis above), but also b) that we actually have downgraded the meaning of human autonomy to a techno-compatible reduction that allows to non-metaphorically call robots "autonomous" in that sense. If b) is the case, then to convey human-specific autonomy properties we need new denotations to specifically convey these properties (like being normgivers).

Artificial autonomy might in language become the new normality, which is why we can drop the "artificial" and just talk about autonomous robots instead of artificially or technically autonomous robots. If Ricœur is right that the metaphor not only traces a preexisting but creates the similarity, then the omnipresence of metaphorizations like AI or autonomous robots creates, emphasizes and sediments the identical of the similarity between man and machine. Or: the more we talk about AI or technical autonomy, the more artificial and technicist the understanding of natural/human intelligence or norm-giving autonomy might become. If metaphors teach us to see technology as autonomous or intelligent, in turn, to see intelligence or autonomy in technicist terms and as unmarked standard meanings, then it teaches us to unsee previous aspects of meaning that we might want to actively preserve from such a veiling. The taught technicist view comes at the cost of possibly forgotten humanist views, of which important aspects like embodiment (*Leiblichkeit*), emotionality, empathy, intent, motivation, preference, desire, will, sociality, culturality, practical reason, etc. might fade. Now that is something that should be critically examined, explicitly and publicly decided on by the majority of moral lawgivers and not by the minority of technomoral rulers like tech giant CEOs or totalitarian governments.

Conclusion

Whoever understands, even without (in doubt not or not uniformly used) markings and indices, that *autonomous systems* can never mean the normative autonomy of self-legislation and self-reflection as a subject recognizing these laws, does not demand for autonomous robots nonsensical things like citizenship[69] or personal rights[70] or possible new criminal offenses like robot insult or robot dignity violation. Systems themselves cannot be granted any rights, since without normative autonomy they are not able to recognize or reject them in a self-determined way; moreover, by becoming part of a community of law (subjects of norms), systems would have to be granted a right to participate in deliberating the design of these rights, such as a right to have a say or a right to vote (to be or at least mandate normgivers). Technical autonomy is revealed as a metaphor for high levels of automation. The love marriage among robots or strike rights of technical systems is thus not the problem. But rather: The consequences of the interest-driven penetration of everyday life with increasingly "autonomous" systems for the personal and political autonomy of humans must be understood and intelligibly shaped as *normgiving norm subjects*.

If the systems that currently pervade our lifeworld are only metaphorically autonomous then they are actually heteronomous, meaning they follow someone else's laws. They treat us according to those laws, which are not our laws. We should not allow the metaphorization regarding technical autonomy to veil their actual heteronomy because if unveiled as heteronomous the keen critical eye examines the legitimacy of the lawgivers in the background and does not allow them to stage these systems as self-legislative and therefore themselves responsible for their normative, moral choices and laws they follow and orient their behavior on. The metaphorization of digital technology is not just a way of emphasizing similarities or giving vacuum cleaning robots cute names, it is a strategy to deflect responsibility by the actual normgivers behind those systems – since those lawgivers are not democratically legitimized the deflecting ruse should be critically called out and their lawgiving should be brought to the public court of moral reason. In a variation of Nietzsche's "The more perfect the machine, the more morality it makes necessary."[71] it holds that the more autonomous the technology, the more authority it makes necessary. However, the more metaphorized technology discourses are, the more awareness of necessary and essential differences needs to complement the teaching of sameness.

69
FAZ, "Roboterfrau bekommt saudische Staatsbürgerschaft".

70
S. Beck, "Brauchen wir eine Roboterethik?"; N. Nevejans, "European civil law rules in robotics".

71
Friedrich Nietzsche, *Nietzsche Source – Digitale Kritische Gesamtausgabe Werke und Briefe (eKGWB). Nachgelassene Fragmente 1879 – 40 (1879)*, p. 40,14. Available at: http://www.nietzschesource.org/#eKGWB/NF-1879,40 (accessed on 1 August 2024).

Bruno Gransche

Normirajuća tehnika?

Metaforizacija autonomije i čemu nas to uči

Sažetak

Ljudska autonomija nije »tehnička« autonomija. Da se na oba fenomena podjednako referiramo otkriva prenošenje obilježavanja s čovječne na tehničke kontekste tj. metaforizaciju. Autonomni ljudi odabiru podčiniti se samodanim zakonima (1). Strojevi, neovisno o metafori, ne mogu odbaciti svoju heteronomiju (2). Što je doslovno mišljeno s »tehnička autonomija« u suprotnosti ili izmjenično s »automacija« treba se detaljnije istražiti (3). Višeslojni su pristupi tehničkoj autonomiji predloženi da bi se izvelo razlikovanje, (4) no i dalje se otkriva neizbježna metaforizacija unutar opisa i analize autonomnih sustava. Ovu metaforizaciju potrebno je analizirati u pogledu relacijske sličnosti u konceptualnom prijenosu, koje nas uči da entitet (npr. autonomna osoba) istovremeno jest, nije *i* jest nalik *drugome (npr. autonomni robot).*

Ključne riječi

umjetna inteligencija, automacija, autonomni sustav, autonomija, heteronomija, metafora, samozakonodavstvo

Bruno Gransche

Normgebende Technologie?

Metaphorisierung der Autonomie und was sie uns lehrt

Zusammenfassung

Menschliche Autonomie ist nicht technische „Autonomie“. Dass wir beide Phänomene zweideutig bezeichnen, offenbart eine Übertragung von Bezeichnungen vom Menschen auf technische Kontexte, d. h. eine Metaphorisierung. Autonome Menschen unterwerfen sich selbst gegebenen Gesetzen (1). Maschinen können ihre Heteronomie, unabhängig von der Metapher, nicht ablegen (2). Was wörtlich mit „technischer Autonomie“ im Gegensatz zu „Automatisierung“ gemeint ist oder mit dieser austauschbar ist, muss im Detail untersucht werden (3). Zur Unterscheidung wurden mehrschichtige Ansätze zur technischen Autonomie vorgeschlagen (4), die jedoch die unvermeidliche Metaphorisierung bei der Beschreibung und Analyse autonomer Systeme offenbaren. Diese Metaphorisierung muss im Hinblick auf die Ähnlichkeitsbeziehung analysiert werden, die bei der konzeptuellen Übertragung am Werk ist, die uns lehrt, dass eine Entität (z. B. eine autonome Person) gleichzeitig eine andere ist, nicht ist *und einer anderen* ähnelt *(z. B. einem autonomen Roboter).*

Schlüsselwörter

künstliche Intelligenz, Automatisierung, autonome Systeme, Autonomie, Heteronomie, Metapher, Selbstgesetzgebung

Bruno Gransche

Technologie normative ?

La métaphorisation de l’autonomie et ce qu’elle nous enseigne

Résumé

L’autonomie humaine n’est pas l’autonomie technique. Que l’on se réfère aux deux phénomènes de manière équivoque révèle la présence d’un transfert de propriétés appartenant à un contexte humain à celui de la technique, ce qui a ainsi trait à de la métaphorisation. L’homme autonome choisit de se soumettre à la loi qu’il se donne lui-même (1). Les machines, quelle que soit la

*métaphore, ne peuvent se débarrasser de leur hétéronomie (2). Il est nécessaire d'explorer en détail ce que l'on entend littéralement par autonomie technique, par opposition à ou en tant que synonyme d'automation (3). Pour faire la distinction, des approches multi-niveaux de l'autonomie technique ont été proposées (4), révélant néanmoins l'inévitable métaphorisation dans la description et l'analyse des systèmes autonomes. Cette métaphorisation doit être analysée sous l'angle de la relation de similarité qui est à l'œuvre dans le transfert conceptuel, et qui nous apprend qu'une entité (p. ex. une personne autonome) e*st, n'est pas, *et* est semblable *à une autre (p. ex. un robot autonome) en même temps.*

Mots-clés

intelligence artificielle, automation, systèmes autonomes, autonomie, hétéronomie, métaphore, auto-législation