Machine Dreaming

Stefan Rieger

I. AI – Old and New

As a rule, artificial intelligences place their states of affairs (*Sachstände*) in relation to those of natural intelligences and, above all, to those of human beings. This impulse has persisted with a certain degree of tenacity throughout their various phenomenal manifestations and respective implementations. In this way, historical references have become possible that one would not at first assume for the highly present-oriented field of artificial intelligence. With its various breakthrough periods, AI has managed to become an operative vade mecum, and in light of its ostensibly universal applicability, it has become a warning sign for a modern era that, in a certain way, no longer seems to need the human at all. There is hardly a single field in which AI does not seem to promise one solution or another, from self-driving cars to medicine, from traffic logistics to software development, from the production of goods to the creative economy, from the financial sector to facial recognition. Nevertheless, references to different conceptualizations of intelligence are necessary in order to evaluate, by comparing anthropological and technical states of affairs, the sustainability of their respective methods, their epistemological status, and thus also, on an underlying level, their social acceptance.

In the 17th century, for instance, this relationship was unproblematic because it was not based on competition (cf. Rieger 1997). Not until an emphatic concept was developed of that which constitutes the human did it become possible for artificial processes to call this self-perception into question. Necessary, too, was that relatively young invention of the human which, oriented toward criteria of uniqueness, distinctiveness, and individuality, detected in every externalization (*Entäußerung*) the will to pursue an inimitable style (cf. Foucault 1970). The philosopher Johann Gottlieb Fichte described the enno-

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

blement of uniqueness as a particular feature of the age of Goethe, and thus he argued that individuality, the existence of which can be assumed on the basis of probability theory, should be protected by intellectual copyright law. "It is more improbable than the greatest improbability," according to Fichte, "that two people should ever think about any subject in exactly the same way, in the same sequence of thoughts and in the same images, when they know nothing of one another. Still, this is not absolutely impossible." It is absolutely improbable, however, "that someone to whom ideas must first be imparted by another should ever assimilate them into his own system of thought in exactly the form in which they were given" (Fichte 2015 [1793]: 451). This incalculable residue in human data processing forms the legal justification of our inalienable property (cf. Kittler 1987; Rieger 2020).

It has become increasingly possible to base the self-image of modernity on calculations (such as that of probability) and ultimately on a particular orientation of rationality, which, with various degrees of nuance, has indeed always served as an emblematic consolidation of mechanized processes. Because individuality capitulates to rational pervasion – and, for reasons of its own justifiability, must in fact capitulate – here, too, the residue becomes the basis of its own evaluation. In short, it has become the anthropological residuum. Like a revised version of Cartesian dualism, man and machine are categorically separated from one another for the greater glory of complexity: "In other words, a highly complicated system like an organism cannot be broken down into describable individual processes without an indissoluble residue remaining; it therefore cannot be rationalized, and thus it also cannot be completely depicted by means of a technical model" (Wolfgang Wieser, cited in Steinbuch 1971: 19).

The inauguration of the human as an evasive phenomenon that is resistant to rationalization and is thus incapable of being simulated in a seamless way has been described from various perspectives. Since then, everything that proceeds from and concerns the human has been a matter of epistemological interest and has been the object of systematic observation and scientific attention. With wide-ranging variants within this epochal setting, the recording system around the year 1800 became the code of a new order of knowledge that revolved around and was oriented toward the human. Since that time, the human has been regarded as the source of all sources and has reliably served as a generating principle of inexhaustibility (cf. Schneider 1994). In this light, the emergence of artificial intelligences represents a threat to its narcissistic sovereignty. Evidence of this can be found in his-

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

torical semantics, which tends to keep mechanisms in their "rightful" place. Conceived as the paragon of unoriginality, formulaicness, regularity, stereotypes, schematization and repetition, the machine merely serves as the negative foil to the radiance of the human being (cf. Rieger 2018). Thus, a bulwark has been created that positions the human against that which industrial and digital revolutions have released into the world with their incessant logic of growth (cf. Rieger 2003). Over the course of these developments, human faculties have quickly come to approximate techniques of data processing. This affects our specific manners of perception and cognition, the ways in which we see and hear, taste and smell, remember and notice, generalize and forget, invent and discover, associate and draw conclusions, hope and dream, fantasize and deduce, collect and process information, and the ways in which we are praxeologically active and socially integrated. The sum of these tendencies confirms, in an impressive manner, Friedrich Kittler's presumption that we are only able to make formulations about ourselves by examining the media of our knowledge. "So-called Man", according to Kittler, "is not determined by attributes which philosophers confer on or suggest to people in order that they may better understand themselves; rather, He is determined by technical standards. Presumably then, every psychology or anthropology only subsequently spells out which functions of the general data processing are controlled by machines, that is, implemented in the real" (Kittler 1997: 132). In its fundamental orientation, this media-anthropological implementation will conceivably not end well for human beings. Up against the speed, capacity and reliability of technical data processing, the human recedes into the background. As to how this process should be discussed, Günther Anders (1983) nailed it on the head when he spoke of Promethean shame and the antiquity of mankind: Technology makes its creator look old.

II. DeepDream and Adversarial Attacks

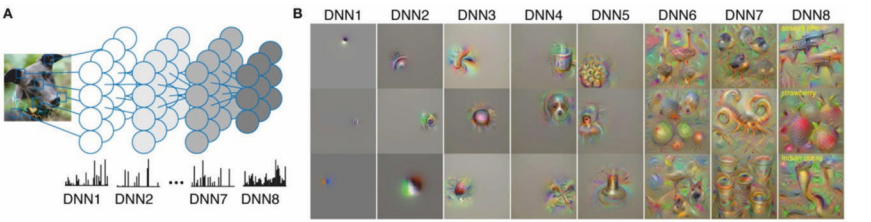
In light of historical semantics and its inertia, it is a rather peculiar finding that the dream, of all things, has been infiltrating the machine for some time now and that, over the course of machine learning and deep-learning algorithms, our own dreams have quite blatantly been affected by its dreams. This incursion of the irrational into the technical domain of the rational is noteworthy because focusing on the dream was itself a central element of the inauguration and thus also part of the history of paying at-

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

tention to the individual-anthropological – a history which, since the 18th century, has focused on human beings in their totality and thus brought to light, for the first time, the specific features of the knowledge inherent in dreams (cf. Schings 1994). Everything about the dream seemed remarkable – above all its formal richness and its manner of drawing connections, its excess and its inherent logic, its polysemy and ambiguity, its bizarreness and fantasies, and not least its license to suspend causality and coherence. The dream becomes the occasion to describe particular images and particular connections between images – varying between a hallucinatory character and clarity, between the accessibility of consciousness and the evasiveness of variously conceptualized forms of unconsciousness (cf. Schredl/Erlacher 2003; Fuchs 1989).

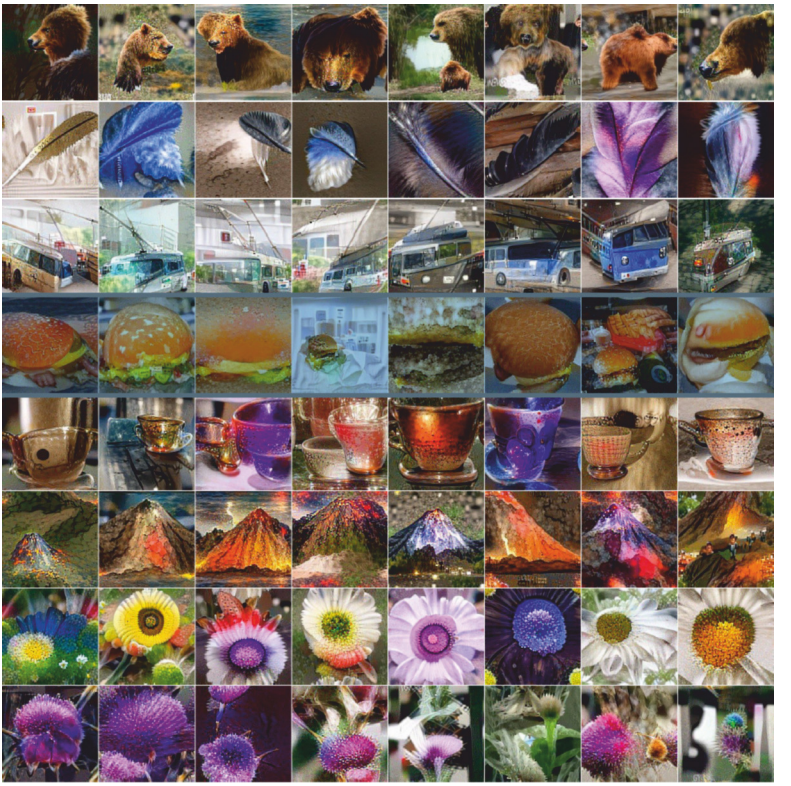
If the dream has infiltrated the machine, it ceases to be a privilege of humankind. The program *DeepDream*, for instance, which Google released into the world in 2015, is unambiguously identified with dreaming. Even the German Wikipedia entry on the program takes up this wording and discusses the history of its dissemination: "Because its results recall the recognition of faces or animals in clouds (see pareidolia), the media often refer to this process as the 'dreaming of a computer'" (https://de.wikipedia.org/wiki/DeepDream; cf. Spratt 2018). The program inverts common pattern-recognition systems, which operate on the basis of artificial neural networks (a so-called convolutional neural network). The latter make use of iterative loops and large datasets to recognize particular patterns. The problem for early cybernetics – how to recognize a face or an object, for instance, from various angles or in various lighting conditions – is solved here by means of iteration and layering. The use of the term "deep", which semantically unites this discourse, is due to the stacking of multiple layers.

Figure 1



https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Figure 2: Class-conditional 224 × 224 images obtained by DeepInversion given a ResNet50v1.5 classifier pretrained on ImageNet. Classes top to bottom: brown bear, quill, trolleybus, cheeseburger, cup, volcano, daisy, cardoon



Yet Google then inverts the process: Instead of recognizing images, it requested the network to generate images and interpret objects within existing images. If the algorithm was fed the photograph of a cloudy sky, it began to make free associations. If the computer discovered that a cloud resembles a whale, it generated a whale from its storehouse of saved images, and this whale perfectly fit the shape of the clouds. It recognized patterns where no patterns existed. The researchers called this technique of abstraction and overinterpretation "Inceptionism", taken from the Hollywood film *Inception*,

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

in which Leonardo DiCaprio breaks into the dreams of other people. (Airen 2015)

Figure 3

This image-generating process, which released a flood of psychedelic pictures into the world, takes place without any human agency. Its proximity to dreaming arises by way of analogy: "As in a dream, DeepDream combines new impressions and stored knowledge into a graphic remix, which is surreal and yet strangely familiar. The result is unsettling images that resemble psychedelic hallucinations – intermediary worlds and parallel realities in which reality blends together with free association" (ibid.). It goes without saying that these images have become part of the discussion about the relationship between human and artificial creativity. As a result, they have even attracted the attention of art historians (cf. McCraig et al. 2016; Boden 2017; DiPaola et al. 2010).

Other voices have been less cautious, semantically, and have likened this process of image generation to hallucination. Subliminally, at least, this calls

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

to mind threatening situations in real life, thereby taking away the impression of harmlessness. However much one might trust a driver who steers his car with the certainty of a somnambulist, a certain degree of doubt arises when one thinks about dreaming self-driving cars. The use of dreaming cars might therefore be somewhat unsettling given the potential real-life consequences (cf. Rocklage 2017; Piergiovanni 2019). Whereas psychedelic images typically do not do anyone any harm and typically do not cause any lasting damage, the situation with hallucinating self-driving cars is another matter entirely.

Against this backdrop, it is possible to situate a discussion that, like the case of autonomous cars, concerns a central aspect of AI and its potential future scenarios. In a peculiar way, this discussion brings to light certain factors and circumstances that are detrimental to self-driving cars. In this context, it is not human beings and their shortcomings that play a role which must be kept to a minimum by means of technical methods and by delegating autonomy; rather, the issue is the system of recognition and thus the technology itself. Especially significant in this regard is a peculiarity of artificial systems designed to recognize objects in the environment, something which is often described as a weakness. These particular failures of artificial intelligences are often discussed in language typically used to describe mental disorders such as delusions or hallucinations. It has occasionally been mentioned (with a degree of *Schadenfreude*) that AI introduced a hallucination problem into the world that is rather difficult to fix (cf. Simonite 2018). This talk of hallucination is not only found in editorial write-ups but also in the work of experts: "Recently, adversarial algorithms were developed to facilitate hallucination of deep neural networks for ordinary attackers" (Abdel-Hakim 2019).

At first, the phenomenon of hallucinations attracted attention as no more than a sort of curiosity of AI. There are so-called "adversarial attacks", which deprive the system of coherence and understanding, and so-called "adversarial patches", which are meant to cover up such glitches in an inconspicuous manner. At the heart of it all lies the finding that optical image-recognition systems can be made to make simple but grotesque misjudgments in a relatively straightforward way. To illustrate this, the weekly newspaper *Die Zeit* cited research dealing with the technical metamorphosis of a simple banana. The banana was scanned by a technical system and accordingly recognized for what it was. However, if there was a small image – a "patch" or a "sticker" – without any recognizable content near the banana, the system

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

would lose its perception of the banana and identify the new situation as a toaster, for instance.

But that's not all. In the case of the banana, the image-recognition system was 97 per cent convinced of its accuracy, but this level of confidence increased to 99 per cent in the case of the toaster: "Other cases show how a microwave with an adversarial patch was identified as a telephone and how, after the addition of a few pixels that no person would ever notice, the computer identified a panda bear as a gibbon" (Schmitt 2019: 33). The abundance of examples could be expanded to include things from everyday life such as teddy bears and oranges, socks and pets, sports equipment and furniture. Stories of this kind are easy to tell – and that's probably why they make it into the feuilleton pages of ambitious newspapers –, not least because the especially grotesque extent of the false identifications underscores an important point. To a human observer, it would not be unexpected for a banana to be mistaken for a similarly shaped zucchini or a similarly colored lemon. In fact, given the process of thinking in gradual degrees and linearities, such confusion could even be called plausible. It could be explained by the history of technical vision, which had to work its way through staggered resolutions while also dealing with other visual parameters such as shading, distance and different viewing angles.

In this respect, it is precisely the identification of objects – the special achievement of human sight – that has repeatedly posed special challenges throughout the history of machine vision. The German cyberneticist Karl Steinbuch, for instance, offered a typical assessment of the matter: "In no other area is the inferiority of technical entities to organic systems more apparent than in the visual system" (1971: 98). But the translation of a banana into a toaster is a decisive break in the habitual way of dealing with similarity relations. It represents an exceptional case that is not covered by common special forms of image perception. One such form is provided by the phenomenon of ambiguous images (*Kippbilder*), which Gestalt psychologists have used to describe what happens when the typical process of perceiving images is undermined. Their logic is not determined by resolutions and approximations; rather, images of this sort break the contract with continuity. There is no gray area and there are no overlaps between a banana and a toaster, between perceiving Rubin's vase or two faces: Deciding between the vase and the faces is an effect of their binary instantaneous setting (cf. Schönhammer 2011).

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Figure 4: The left image shows a real graffiti on a Stop sign, something that most humans would not think is suspicious. The right images shows our [sic!] a physical perturbation applied to a Stop sign. We design our perturbations to mimic graffiti, and thus "hide in the human psyche".

Things become difficult when one expands this colorful list of examples and comes upon cases that are anything but harmless and can alter our assessment of complex and hazardous situations. With respect to self-driving cars, it has been reported that a single sticker placed on a stop sign can be enough to confuse image-recognition systems; it has also been shown that, "by subtly altering the shell of a plastic turtle, it could be made to appear like a rifle" (Schmitt 2019: 33; cf. Athalye et al. 2018). The identification of traffic signs in particular has become a special challenge and is thus being researched extensively (cf. Cires et al. 2012). Such examples bring to light, in a dramatic way, the dire importance that adversarial attacks might have in everyday life.

The article in *Die Zeit* also refers to an additional increase in possible applications. Whereas the examples mentioned above concern the confusion of static images, researchers such as Michael J. Black at the Max Planck Institute for Intelligent Systems in Tübingen have been working on manipulating recordings of motion. That is, they have been trying to manipulate so-called "optical flow systems". Because the anticipation of complex situations – and also the appropriate reaction to these situations – depends on the ability to perceive such systems, the whole issue is increasingly losing its status as a mere technical curiosity: "Because optical flow systems make it possible to

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

track the primary movements taking place in the bustle of high-traffic situations (the movements of cyclists and vehicles, for instance), they represent the potential building blocks for future self-driving cars" (Schmitt 2019: 33; Ranjan et al. 2019).

To optimize their own systems, the researchers in Tübingen sent their "patches" (before publishing their results) to manufacturers and suppliers in the automotive industry so as to assist the latter's investigations into possible causes of interference. The director of the Algorithm Accountability Lab, Katharina Zweig, whom the article quotes, underscored the seriousness of the situation by pointing out that such systems reach their decisions without any knowledge or awareness on the part of those who might be affected by them. The artificial systems operate in the mode of the imperceptible, and they operate – as the case of self-driving cars and other examples makes clear – in spheres that undoubtedly affect daily life. Research into this phenomenon is ongoing, and it reached a critical point in 2019. It became clear then how much dangerous potential lies in the technique of adversarial attacks and how working on such networks produces emergent phenomena and thus has no use for the old paradigm and traditional understanding of mechanics, determinacy and the predictability of technical systems. Deep neural networks are difficult to predict, and they are not limited to areas in which their mistakes can be tolerated as mere special phenomena:

> Because of these accomplishments, deep learning techniques are also applied in safety-critical tasks. For example, in autonomous vehicles, deep convolutional neural networks. The machine learning technique used here is required to be highly accurate, stable and reliable. But, what if the CNN model fails to recognize the "STOP" sign by the roadside and the vehicle keeps going? It will be a dangerous situation. (Xu et al. 2019)

A comparably dangerous situation can be imagined in the arena of financial transactions: "If there are fraudsters disguising their personal identity information to evade the company's detection, it will cause a huge loss to the company. Therefore, the safety issues of deep neural networks have become a major concern" (ibid.).

What the text at first identifies as a strategy of humans dealing with credit institutions is then shifted into the world of other manners of operation and forms of representation. Here, the question of credibility is not treated in terms of the personal identity of fraudsters – that is, in terms of the uniqueness of their biometrics, for instance – but rather in terms

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

of graphs that concern the inherent logic of certain processes. Whether in the case of street traffic or in that of financial transactions, dangers loom in various places within an environment that is always at risk of being disrupted. They are differentiated according to various occasions and technical implementations, and they pertain not only to the recognition of letters and numbers, stop signs and other traffic signs, bananas and turtles, but also to other forms of data (cf. Feinman et al. 2017; Kurakin et al. 2017; Evtimov et al. 2018). The example of graphs (and of manipulating just a few nodal points) has caught the eye of the finance industry, and different ways of dealing with language – whether in its written or spoken form – have also attracted attention. Here, too, the attacks take the guise of minimal and barely detectable differences – for instance, by means of minor manipulations in the form of typos (cf. Jia/Liang 2017). Beyond mere typos and small differences in letters, focus has also turned to the act of switching out entire sentences or phrases. In this respect, a 2018 article with the title "Detecting Egregious Responses in Neural Sequence-to-Sequence Models" greatly (egregiously?) expands the descriptive language and semantics of that which is typically associated with "egregiousness" (cf. He/Glass 2018). The hallucinatory and the grotesque, which are inherent to the confusion of bananas and toasters, are extended here to include something akin to the monstrous: "In this work, we attempt to answer a critical question: whether there exists some input sequence that will cause a well-trained discrete-space neural network sequence-to-sequence (seq2seq) model to generate egregious outputs (aggressive, malicious, attacking, etc.)" (ibid.; cp. Sandbank et al. 2018). The adjective *egregious* is used here to describe reactions to sentences; the semantic range of this word covers, in addition to things that merely stand out in a bad way, also the sphere of the inimical. As one overview article has pointed out, even systems of spoken language can be the object of such egregiousness (cf. Hinton et al. 2012).

Research into the robustness of such systems has therefore become as inevitable as its expansion into other forms of data. What creeps into the situation is a semantics of suspicion concerning everything that the world has in store. Talk of Potemkin villages has been spreading, and there is often mention of alienation, which goes hand in hand with the undermining of human perception. The adjective *subtle* is used to describe the imperceptible changes that lead to blatant misjudgments and their grave consequences. The images discussed in the articles cited above are reminiscent of the iconography found in the "search-and-find puzzles" (*Suchbilder*) printed

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

in newspapers and books for children, which contain many objects that are ostensibly the same but also at least one outlying detail that is difficult but not impossible to detect. The only difference is that, in the case of adversarial attacks, the human eye has no chance to discover the built-in anomaly (on the tradition of *Suchbilder*, cf. Ernst et al. 2003).

III. Ally Patches

It is a peculiar point that the defense measures taken against attacks on image recognition and sign recognition in the case of self-driving cars are now, for their part, taking place in the consistent mode of so-called "ally patches". That this process results in mutually interfering stickers ("Ally Patches for Spoliation of Adversarial Patches", as one article on the topic is titled) is a detail that perhaps only attracts the attention of cultural theorists, who like to regard such self-references as part and parcel of the logic of cultural chains of meaning. That the stickers have changed sides and now it is they, of all things, that are expected to disrupt the disrupter and thus guarantee security, is nevertheless a cycle that warrants attention – especially in light of the possibility that this circle might never be closed (regarding the issue of disruption in media and cultural studies, cf. Kümmel/Schüttpelz 2003). Because here, too, the principle of imperceptibility applies, and a peculiar finding emerges: the human eye is unable to distinguish the combatants from one another – only the final effect reveals their position as friend or foe. Attaching stickers has lost its innocence and thus also the subversive charm that went along with these semi-public acts of expressing opinions. It is more than a juvenile misdemeanor and even has the characteristics of a high-risk traffic violation: "The consequent troubles may vary from just unpleasant inconvenience in applications like entertainment image and video annotation, passing by security-critical problems like false person identifications, and can turn out to be life-threatening in autonomous navigation and driver support systems" (Abdel-Hakim 2019). Thus the vulnerability of self-driving cars has become the object of attention – as in one case in which an adversarial attack managed to turn on windshield wipers in dry conditions (cf. Deng et al. 2020).

Particularly striking in the research is an assessment that concerns the status of the examples in question (cf. Xu et al. 2019). Instead of the operative question concerning which of three described methods should be used

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

(gradient masking, robust optimization, detection), the urgent question has turned out to be what sort of role these exemplary cases should be attributed in specific everyday situations. In addition to the cat-and-mouse game of reacting to specifications, certain issues play a role that belong to the domain of basic knowledge. As one repeatedly reads, the spirals and circles of attacks and counter-attacks strengthen our understanding of the inherent dynamics of deep neural network systems – and thus bring knowledge about their technical nature to light. As one article on this topic makes clear, adversarial attacks are not *bugs* in the system; rather, they are *features* of it (cf. Ilyas et al. 2019). Another text strikes a comparable tone – though with the nuance that it uses the term *flaw* instead of the term *bug*: "These results have often been interpreted as being a flaw in deep networks in particular, even though linear classifiers have the same problem. We regard the knowledge of this flaw as an opportunity to fix it" (Goodfellow et al. 2015).

One particular point that has been revealed by this (involuntary) basic work on the functioning of deep neural networks is that the latter operate quite differently from human data processing – even though the language used to describe them (all the talk of dreaming or hallucinating machines) does much to conceal this difference. However much protective measures and an understanding of DNN's mode of operating mutually condition one another, human perception and reasoning remain on the outside of the process: "For example, adversarial perturbations are perceptually indistinguishable to human eyes but can evade DNN's detection. This suggests that the DNN's predictive approach does not align with human reasoning" (Xu et al. 2019).

In a certain way, the technical logic of neural networks has liberated itself from the politics of the human imagination, either despite or because of all the semantic borrowings – and yet at the same time, this logic has also come somewhat closer to such politics, at least as far as effects are concerned. It is no longer necessarily the case that the order of things is an order that follows human preconceptions and reactions along the lines of similarities and representations. The danger of misinterpreting traffic signs and, even more so, the manipulation of more complex situations that would be so essential to the acceptance of self-driving cars thus underscore the special epistemological position of such processes. They may be able to dream and hallucinate, but the way they do so differs from that of human beings. *Our* order of things is not *the* order of things.

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

The semantics of depth, which is applied to a specific technical feature of the processes employed, concerns, in the case of DeepDream, not only mechanisms for organizing images but also achieves the level of disruptive mechanisms with methods such as DeepFake or DeepFool – and this is not even to mention applications such as DeepFood, which is used to identify nutritional food (cf. Jiang et al. 2020). With the latest technical possibilities, an element of unpredictability and uncertainty is spreading within an everyday environment that one hoped would be kept safe. It is as though one is witnessing, with open but blind eyes, a reentry of the irrational into the field of the rational, the only difference being that whatever bizarre images come to light are not a matter of aesthetic surplus production but rather concern the stabilization of systems. In all performances of the bizarre and psychedelic, the moment of system optimization remains inherent to the dream of the machine. Any admission that there is some aesthetic surplus or added cultural value produced by these processes therefore remains ambivalent. It is impossible to distinguish between seriousness and triviality, consequence and mere emergence, curious errors and disruptive intentions:

> We cannot reliably identify when and why DNNs will make mistakes. In some applications like text translation these mistakes may be comical and provide for fun fodder in research talks, a single error can be very costly in tasks like medical imaging. Additionally, DNNs show susceptibility to so-called adversarial examples, or data specifically designed to fool a DNN. One can generate such examples with imperceptible deviations from an image, causing the system to mis-classify an image that is nearly identical to a correctly classified one. Audio adversarial examples can also exert control over popular systems such as Amazon Alexa or Siri, allowing malicious access to devices containing personal information. (Charles 2018)

IV. Lucid Dreaming

What does this mean, however, for the scenario discussed at the beginning and its effort to place technical and anthropological states of affairs in relation to one another for the sake of their epistemological validation? A reference of this sort can be made in light of a particular variant of dreaming: lucid dreaming. With the title "Are You Dreaming? A Phenomenological Study on Understanding Lucid Dreams as a Tool for Introspection in Vir-

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

tual Reality", a recent article has endeavored to determine the relationship between lucid dreaming and virtual reality in order, in the end, to identify the lucid dream as the ultimate stage of the technical:

> Lucid dreaming, "dreaming while knowing one is dreaming," is one phenomenon that we can draw parallels to a VR experience. It is a genuine human experience that places a person in a "virtual" reality, i.e., their dream, which feels just as real as their waking reality. At the same time, lucid dreamers are aware that they are in a dream and that nothing in the dream has real-life consequences, much like that of a VR experience. (Kitson et al. 2018)

The research group responsible for this is led by Alexandra Kitson, who situates her work at the intersection of human-computer interaction, design and psychology. In addition, she is also interested in technically induced processes of transgression and well-being (cf. Kitson et al. 2020). In the study cited above, subjects are interviewed according to a rubric, and they are asked questions about perceptual impressions and feelings, about activities and practices, about how such things influence their experience, and not least about the way in which meaning is created in lucid dreams. The knowledge is then fed back into the development of future VR systems with the goal of bringing them closer to and doing justice to the human condition: "This knowledge can help design a VR system that is grounded in genuine experience and preserving the human condition" (Kitson et al. 2018).

Another project led by Kitson – titled "Lucid Loop: A Virtual Deep Learning Biofeedback System for Lucid Dreaming Practices" – goes beyond this combination of introspection, inquiry and extrapolation in the creation of future technological designs. This project opens up a sphere of activity for the technology that seems unusual at first glance, even esoteric (for a similarly esoteric application of virtual reality, cf. Downey 2015). In this case, the focus is on personal well-being. The interrelation between lucid dreaming and technology has brought to light a previously unnoticed potential application, which might possibly be used to increase the greater glory of human beings, their personal achievements and their individual welfare. Through a combination of technical image processing, biofeedback mechanisms and yoga, a new sort of depth is conjured up that is not focused on the depths of the soul but rather on the surface operations of image layers. In their article "Are You Dreaming?", Kitson and her coauthors outline this process as follows and refer to so-called deep convolutional neural networks (DCNN):

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

> Lucid Loop will be an open, [sic!] nature scene with other interactive elements that provoke curiosity. DCNN imagery will provide a level of abstraction needed for a dream-like effect […]. The image layers themselves will range from very abstract to completely clear, mimicking levels of clarity in lucid dreaming. (Kitson et al. 2018)

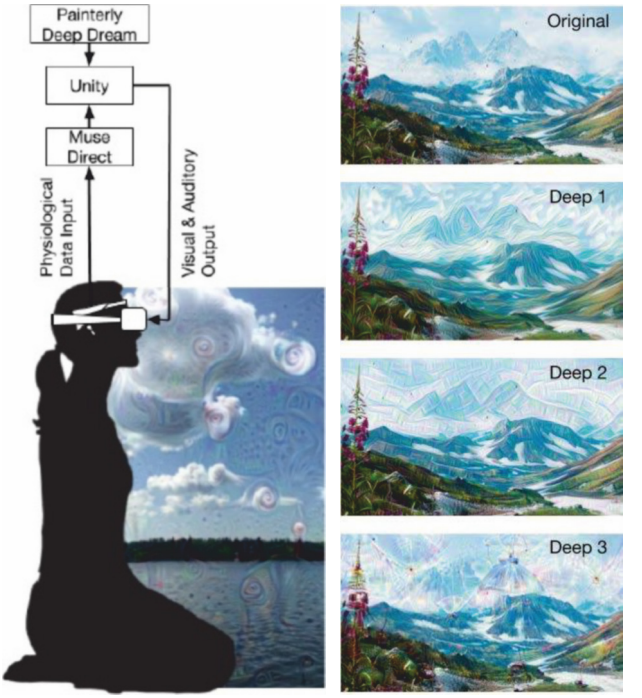
The levels of image clarity mentioned here are an effect of advanced image processing technology. It is artificial intelligences – the processes of deep learning – that lend the images their dream-like impression and enable them to be scaled accordingly (cf. McNamara et al. 2018). This process is therefore accompanied by image sequences that present a technical distortion of a natural world – minutely layered on top of one another, scaled in detail, and as a process that takes place before the eyes of those participating. The result is a flurry of images controlled by a bodily signal, and this flurry is able to exhaust the potential of dream images and recreate the depth of dreams: "Visuals are creatively generated before your eyes using a deep learning Artificial Intelligence algorithm to emulate the unstable and ambiguous nature of dreams" (Kitson et al. 2019). The intensity and clarity of each generation of images are controlled by bodily signals.

What efforts of this sort negotiate is a confrontation between forms of different realities, an authentic confrontation between the virtual and the dream. In this sense, the desired perfection of virtual realities has found its goal and standard in lucid dreaming. What emerges from this is an adaptation of technology and its users that places its highly advanced perfectibility in the service of the human condition. There are thus points of contact, interconnections and references between these entangled concepts of reality, and there are ways in which these connections are formed, modified and technically realized. In the possibility of reciprocal modelling, the virtual and the imaginary – that is, the technologically possible and the anthropologically authenticated – encounter one another on equal terms and seemingly free from Promethean shame. To quote Kitson and her colleagues again: "The ultimate VR might look like lucid dreaming, the phenomenon of knowing one is dreaming while in the dream" (Kitson et al. 2018).

Accordingly, the philosophically controversial question of the reality content of reality does not play a special role in the praxis of the lifeworld. In the virtual, as the sociologist Elena Esposito has written with respect to the interrelated nature of possible existential relations, there are "no false real objects but rather true virtual objects, for which the question of real reality

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Figure 5 (left): Lucid Loop system schematic. Painterly and Deep Dream *creatively generate visuals to emulate dreams. The virtual environment becomes more lucid or "clear" when the participant's physiological signals indicate increased awareness. Figure 6 (right)*



is entirely irrelevant" (Esposito 1998: 270). In this case, traditional distinctions take a back seat to altered functionalities, and the most likely reality proves to be a calculated fiction (cf. Esposito 2019; Dongus 2018).

References

Abdel-Hakim, Alaa E. (2019). "Ally Patches for Spoliation of Adversarial Particles." In: *Journal of Big Data* 6. URL: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0213-4.

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Airen (2015). "So sieht es aus, wenn Computer träumen." In: *Welt*, July 21, 2015. URL: https://www.welt.de/kultur/article144267349/So-sieht-es-aus-wenn-Computer-traeumen.html.

Anders, Günther (1983). *Die Antiquiertheit des Menschen. Erster Band: Über die Seele im Zeitalter der zweiten industriellen Revolution*. 6th ed., Munich: C.H. Beck.

Athalye, Anish et al. (2018). "Synthesizing Robust Adversarial Examples." In: *arXiv.org*, June 7, 2018. URL: https://arxiv.org/abs/1707.07397.

Boden, Margaret A. (2017). "Is Deep Dreaming the New Collage?" In: *Connection Science* 29.4, pp. 268–275.

Charles, Adam S. (2018). "Interpreting Deep Learning: The Machine Learning Rorschach Test?" In: *arXiv.org*, June 1, 2018. URL: https://arxiv.org/abs/1806.00148.

Cires, Dan et al. (2012). "Multicolumn Deep Neural Network for Traffic Sign Classification." In: *Neural Networks* 32, pp. 333–338.

Deng, Yao et al. (2020). "An Analysis of Adversarial Attacks and Defenses on Autonomous Driving Models." In: *arXiv.org*, February 6, 2020. URL: https://arxiv.org/abs/2002.02175.

DiPaola, Steve et al. (2010). "Rembrandt's Textural Agency: A Shared Perspective in Visual Art and Science." In: *Leonardo* 43.2, pp. 145–151.

Dongus, Ariana (2018). "Life Is Real = Reality Is a Platform: Creating Engineered Experiences Where Fiction and Reality Merge." In: *Body Images in the Post-Cinematic Scenario: The Digitization of Bodies*. Ed. by Alberto Brodesco and Federico Giordano, Milan: Mimesis International, pp. 151–162.

Downey, Laura L. (2015). *Well-being Technologies: Meditation Using Virtual Worlds*, Doctoral diss.: Nova Southeastern University.

Ernst, Wolfgang et al. (eds.) (2003). *Suchbilder: Visuelle Kultur zwischen Algorithmen und Archiven*, Berlin: Kadmos.

Esposito, Elena (1998). "Fiktion und Virtualität." In: *Medien, Computer, Realität. Wirklichkeitsvorstellungen und neue Medien*. Ed. by Sybille Krämer, Frankfurt am Main: Suhrkamp, pp. 269–296.

Esposito, Elena (2019). *Die Fiktion der wahrscheinlichen Realität*. 4th ed. Berlin: Suhrkamp.

Evtimov, Ivan et al. (2018). "Robust Physical-World Attacks on Deep Learning Models." In: *arXiv.org*, April 10, 2018. URL: https://arxiv.org/abs/1707.08945.

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Feinman, Reuben et al. (2017). "Detecting Adversarial Samples from Artifacts." In: *arXiv.org*, November 15, 2017. URL: https://arxiv.org/abs/1703.00410.

Fichte, Johann Gottlieb (2015 [1793]). "Proof of the Unlawfulness of Reprinting: A Rationale and a Parable." In: *Primary Sources on Copyright (1450–1900)*. Ed. by L. Bentley and M. Kretschmer, pp. 443–483. URL: http://www.copyrighthistory.org.

Foucault, Michel (1970). *The Order of Things: An Archaeology of the Human Sciences*, New York: Vintage Books.

Fuchs, Peter (1989). "Blindheit und Sicht: Vorüberlegungen zu einer Schemarevison." In: *Reden und Schweigen*. Ed. by Peter Fuchs and Niklas Luhmann, Frankfurt am Main: Suhrkamp, pp. 178–208.

Goodfellow, Ian J. et al. (2015). "Explaining and Harnessing Adversarial Examples." In: *arXiv.org*, March 20, 2015. URL: https://arxiv.org/abs/1412.6572.

He, Tianxing, and James Glass (2018). "Detecting Egregious Responses in Neural Sequence-to-Sequence Models." In: *arXiv.org*, October 3, 2018. URL: https://arxiv.org/abs/1809.04113.

Hinton, Geoffrey et al. (2012). "Deep Neural Networks for Acoustic Modeling in Speech Recognition." In: *IEEE Signal Processing Magazine* 29, pp. 82–97.

Ilyas, Andrew et al. (2019). "Adversarial Examples Are Not Bugs, They Are Features." In: *arXiv.org*, August 12, 2019. URL: https://arxiv.org/abs/1905.02175.

Jia, Robin, and Percy Liang (2017). "Adversarial Examples for Evaluating Reading Comprehension Systems." In: *arXiv.org*, July 23, 2017. URL: https://arxiv.org/abs/1707.07328.

Jiang, Landu et al. (2020). "DeepFood: Food Image Analysis and Dietary Assessment via Deep Model." In: *IEEE Access* 8, pp. 47477–47489.

Kitson, Alexandra et al. (2018). "Are You Dreaming? A Phenomenological Study on Understanding Lucid Dreams as a Tool for Introspection in Virtual Reality." In: *CHI '18, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. URL: https://dl.acm.org/doi/10.1145/3173574.3173917.

Kitson, Alexandra et al. (2019). "Lucid Loop: A Virtual Deep Learning Biofeedback System for Lucid Dreaming Practice." In: *CHI EA '19, Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. URL: https://doi.org/10.1145/3290607.3312952.

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Kitson, Alexandra et al. (2020). "A Review on Research and Evaluation Methods for Investigating Self-transcendence." In: *Frontiers in Psychology*, 11. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.547687/full.

Kittler, Friedrich (1987). "Über romantische Datenverarbeitung." In: *Die Aktualität der Frühromantik*. Ed. by Ernst Behler and Jochen Hörisch, Paderborn: Ferdinand Schöning, pp 127–140.

Kittler, Friedrich (1997). "The World of the Symbolic – A World of the Machine." In: *Literature, Media, Information Systems: Essays*. Ed. by John Johnston. Amsterdam: Gordon & Breach, pp. 130–146.

Krizhevsky, Alex et al. (2017). "ImageNet Classification with Deep Convolutional Neural Networks." In: *Communications of the ACM* 60.6, pp. 84–90.

Kümmel, Albert, and Erhard Schüttpelz (eds.) (2003). *Signale der Störung*, Munich: Wilhelm Fink.

Kurakin, Alexey et al. (2017). "Adversarial Examples in the Physical World." In: *arXiv.org, February 11*, 2017. URL: https://arxiv.org/abs/1607.02533.

McCaig, Graeme et al. (2016). "Deep Convolutional Networks as Models of Generalization and Blending Within Visual Creativity." In: *Proceedings of the 7th International Conference on Computational Creativity*, Palo Alto, CA, pp. 156–163.

McNamara, Patrick et al. (2018). "Virtual Reality-Enabled Treatment of Nightmares." In: *Dreaming* 28.3, pp. 205–224.

Piergiovanni, A.J. et al. (2019). "Learning Real-World Robot Policies by Dreaming." In: *arXiv.org*, August 1, 2018. URL: https://arxiv.org/abs/1805.07813.

Ranjan, Anurag et al. (2019). "Attacking Optical Flow." In: *Proceedings International Conference on Computer Vision (ICCV)*, Piscataway, NJ: IEEE, pp. 2404–2413. URL: https://arxiv.org/pdf/1910.10053.pdf.

Reinhardt, Daniel et al. (2019). "Entropy of Controller Movements Reflects Mental Workload in Virtual Reality." In: *IEEE Conference on Virtual Reality and 3D User Interfaces*. Piscataway, NJ: IEEE, pp. 802–808.

Rieger, Stefan (1997). *Speichern / Merken: Die künstlichen Intelligenzen des Barock*, Munich: Fink.

Rieger, Stefan (2003). *Kybernetische Anthropologie: Eine Geschichte der Virtualität*, Frankfurt am Main: Suhrkamp.

Rieger, Stefan (2018). "'Bin doch keine Maschine …': Zur Kulturgeschichte eines Topos." In: *Machine Learning – Neue Pfade künstlicher Intelligenz*. Ed. by Christoph Engemann and Andreas Sudmann, Bielefeld: transcript, pp. 117–142.

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Rieger, Stefan (2020). "'Be the Data': Von Schnürbrüsten, Halsketten und der Dokumentation der Daten." In: *Durchbrochene Ordnungen: Das Dokumentarische der Gegenwart*. Ed. by Friedrich Balke et al., Bielefeld: transcript, pp. 191–215.

Rocklage, Elias (2017). "Teaching Self-Driving Cars to Dream: A Deeply Integrated, Innovative Approach for Solving the Autonomous Vehicle Validation Problem." In: *IEEE 20th International Conference on Intelligent Transportation Systems*, Piscataway, NJ: IEEE, pp. 1–7.

Sandbank, Tommy et al. (2018). "Detecting Egregious Conversations Between Customers and Virtual Agents." In: *arXiv.org*, April 16, 2018. URL: https://arxiv.org/abs/1711.05780.

Schings, Hans-Jürgen (ed.) (1994). *Der ganze Mensch: Anthropologie und Literatur im 18. Jahrhundert*, Stuttgart: Metzler.

Schmitt, Stefan (2019). "Ist das wirklich ein Toaster?" In: *Die Zeit*, November 14, pp. 33–34.

Schneider, Manfred (1994). "Der Mensch als Quelle." In: *Der Mensch – das Medium der Gesellschaft*. Ed. by Peter Fuchs and Andreas Göbel, Frankfurt am Main: Suhrkamp, pp. 297–322.

Schönhammer, Rainer (2011). "Stichwort: Kippbilder". URL: https://psydok.psycharchives.de/jspui/bitstream/20.500.11780/3666/1/Kippbilder_psydoc_11052011.pdf.

Schredl, Michael, and Daniel Erlacher (2003). "The Problem of Dream Content Analysis Validity as Shown by a Bizarreness Scale." In: *Sleep and Hypnosis* 5, pp. 129–135.

Simonite, Tom (2018). "AI Has a Hallucination Problem That's Proving Tough to Fix." In: *Wired*, March 9, 2018. URL: https://www.wired.com/story/ai-has-a-hallucination-problem-thats-proving-tough-to-fix/.

Spratt, Emily (2018). "Dream Formulations and Deep Neural Networks: Humanistic Themes in the Iconology of the Machine-Learned Image." In: *arXiv.org*, February 5, 2018. URL: https://arxiv.org/abs/1802.01274.

Steinbuch, Karl (1971). *Automat und Mensch: Auf dem Weg zu einer kybernetischen Anthropologie*. 4th ed., Berlin: Springer.

Xu, Han et al. (2019). "Adversarial Attacks and Defenses in Images, Graphs and Text: A Review." In: *arXiv.org*, October 9, 2019. URL: https://arxiv.org/abs/1909.08072.

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

List of Figures

Figure 1: Horikawa, Tomoyasu, and Yukiyasu Kamitani (2017): "Hierarchical Neural Representation of Dreamed Objects Revealed by Brain Decoding with Deep Neural Network Features." In: *Frontiers in Computational Neuroscience* 11.4. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5281549/.

Figure 2: Yin, Hongxu et al. (2020): "Dreaming to Distill: Data-Free Knowledge Transfer via DeepInversion." In: *arXiv.org*, June 16, 2020. URL: https://arxiv.org/abs/1912.08795.

Figure 3: https://www.tensorflow.org/tutorials/generative/deepdream.

Figure 4: Evtimov, Ivan et al. (2018): "Robust Physical-World Attacks on Deep Learning Models." In: *arXiv.org*, April 10, 2018. URL: https://arxiv.org/abs/1707.08945.

Figure 5: Kitson, Alexandra et al. (2019): "Lucid Loop: A Virtual Deep Learning Biofeedback System for Lucid Dreaming Practice." In: *CHI EA '19, Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. URL: https://dl.acm.org/doi/fullHtml/10.1145/3290607.3312952.

Figure 6: Kitson, Alexandra et al. (2019): "Lucid Loop: A Virtual Deep Learning Biofeedback System for Lucid Dreaming Practice." In: *CHI EA '19, Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. URL: https://dl.acm.org/doi/fullHtml/10.1145/3290607.3312952.

https://doi.org/10.14361/9783839457320-011 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND