Secret Agents

A Psychoanalytic Critique of Artificial Intelligence and Machine Learning

Clemens Apprich

Abstract

"Good Old-Fashioned Artificial Intelligence" (GOFAI), which was based on a symbolic information-processing model of the mind, has been superseded by neural-network models to describe and create intelligence. Rather than a symbolic representation of the world, the idea is to mimic the structure of the brain in electronic form, whereby artificial neurons draw their own connections during a self-learning process. Critiquing such a brain physiological model, the following article takes up the idea of a "psychoanalysis of things" and applies it to artificial intelligence and machine learning. This approach may help to reveal some of the hidden layers within the current A.I. debate and hints towards a central mechanism in the psycho-economy of our socio-technological world: The question of "Who speaks?", central for the analysis of paranoia, becomes paramount at a time, when algorithms, in the form of artificial neural networks, operate more and more as secret agents.

Following Hubert Dreyfus' *What Computers Still Can't Do* (1992) one can assert that *good old-fashioned artificial intelligence*,[1] which was based on a symbolic information-processing model of the mind, has been replaced by neural network models to describe and create intelligence. Rather than the symbolic representation of the world, the brain's learning power is by now simulated on digital computers, in order to "automate the procedures by which a network of neurons learns to discriminate patterns and respond appropriately" (Dreyfus 1992, xiv). What Dreyfus mentions here is an alternative line of thinking within the history of artificial intelligence (A.I.), which goes back to the early 1940s, when Warren McCulloch and Walter Pitts introduced the idea of a neural network (cf. McCulloch/Pitts 1943).[2] McCulloch and Pitts proposed the idea of simulating the human brain as a network of neurons in order to acquire the brain's ability for

1 The term was coined by the philosopher John Haugeland in his book *Artificial Intelligence: The Very Idea* (Haugeland 1985).

2 Hence, this alternative approach was already there even before John McCarthy, Marvin Minsky, and others started to use the term "artificial intelligence" during

DOI 10.14361/dcs-2018-0104

pattern recognition.[3] This approach differs from the classic understanding of A. I. research, insofar as it does not depend on a rationalistic representation of the world, but thinks of the mind as a gigantic network, composed of myriads of parallel processing "mental agents" (Minsky 1985: 17), which allow the network to react to arbitrarily complex problems. Hence, the belief that the world is structured according to fixed, rational rules has come to an end within the artificial intelligence discourse.

The new A. I.-paradigm is connectionist, since neural networks are modelled on the somatic nerve system of animals. Each neuron or agent connects with another neuron through its activation, thus enabling the network to grow exponentially. However, for a long time, connectionism was identified with Frank Rosenblatt's Perceptron (1958), a neural network of only one layer, that is a layer of neurons between the input- and the output-side. The problem with this simple model was that it could not be trained to recognise more than one class of patterns at a time, because single layer perceptrons are only capable of linear learning (cf. Minsky/Papert 1969).[4] The inability to solve very simple logical operations put the neural network approach back for many years in favor of the aforementioned symbolic A. I. and it was only until multi-layered perceptrons were developed, that so-called feedforward (later also recurrent) neural networks with two or more layers revived the connectionist idea in the 1980s. From then on "parallel distributed processing" was theoretically possible and with it, a boundless neural network architecture (cf. Rumelhart et al. 1986). But it was only until recently that these complex neural networks could actually be built, thanks to the enormous increase in data, greater computer power, and better funding. After an initial identification of a pattern, for example a letter, each successive layer looks for further patterns, such as a sentence, thereby drawing on the findings of the previous layers. Hence, the processing of data isn't based on deductive reasoning, but on inductive probability, since only the input and the output data are known, not the rules that led to the conclusion. For connectionists this constitutes the simple basis for learning and pattern recognition, both in the case of humans and machines.

One could say that the classic A. I. approach is creationist, in the sense that it presumes a world of already existing (divine or rationalistic) rules, which only need to be formalised, in order to make sense to a machine (or an analytical philosopher for this reason). In contrast, the new paradigm of neural networks

the *Dartmouth Summer Research Project on Artificial Intelligence* in 1956; a six to eight weeks workshop, which today is considered to be the crucial spark in A. I.-research.

3 The simplest definition of a neural network is that of a machine that makes predictions based on its ability to discover patterns in data.

4 In their book *Perceptrons* (1969) Marvin Minsky and Seymour Papert proved that it is impossible for one-layer perceptrons to learn an XOR function, a very basic principle in mathematical logic. However, they do not claim that the same is true for multi-layer perceptrons, which indeed are able to produce a XOR function.

is evolutionary, as it is not interested in a pre-existing, exact representation of the world, but settles for an ever-closer approximation to the world as it is, or, more accurately, of how the world appears to be. It is, therefore, probabilistic by nature. Take the example of machine-based translation, such as *Google Translate*. In the classical approach, the strategy had been to specify the entirety of words of at least two natural languages and then to program all grammatical rules necessary to translate from one language to another. The problem of such a static approach is that language cannot be reduced to its dictionary definition, which is the reason why – until recently – *Google*'s translations sounded very clumsy and became the subject of countless Internet-jokes. However, at the end of 2016 *Google Translate* secretly rolled out their new system, based on five years of extensive research on artificial neural networks (cf. Lewis-Kraus 2016). The result is a major improvement in the quality of machine translations, as has been acknowledged by leading translators as well as day-to-day users. What is more, the progress in machine learning made by *Google Translate* has generated excitement about what is known as Artificial General Intelligence (A. G. I.). Following its basic assumption, logical reasoning is not some divine afflatus, but can emerge on the basis of trial-and-error over a certain period of time. Hence, deep neural networks, which are capable of learning by themselves, have the potential to create a general intelligence, or, in other words, they become general purpose.

It is not very difficult to link this idea to an empiricist understanding of the world. Consequently, a lot of innovation in A. G. I. is currently coming from robotics, in particular self-driving cars, which rely on sensory experiences. As one researcher puts it in an interview: "Self-driving cars are a data-driven problem."[5] The machine gets smarter the more it experiences in the world, which makes A. G. I. a problem that is going to be solved with data. Hence, the emergence of a new A. I. paradigm, which, in fact, goes back to the 1940s, can be explained by the concurrence of at least three mutually independent areas in the last years: Deep Learning, Network Analytics, and Big Data. Even though the latter has attracted a lot of attention in recent years, it does not explain much without the other two. In order to teach networks how to recognise patterns in a deluge of data, A. I. researchers draw on findings in neuroscience. Neural networks are modelled upon the human brain, respectively the neurons in the brain. The idea is to mimic the structure of the brain in electronic form, whereby artificial neurons draw their own connections during a self-learning process. Considering that an average human brain has about 100 billion neurons and each neuron may be connected to up to 10.000 other neurons, which means that the brain contains as many as 1.000 trillion synaptic connections, shows the complexity of this endeavour. And indeed, for a long time the computer power just wasn't enough to even try to replicate such a complex structure. However, in recent years the exponential growth of computa-

5 From the documentary *Road to A. I.*, produced by Red Hat Films documentary; URL: https://vimeo.com/219731087.

tional power combined with the massive accumulation of data has given boost to the connectionist idea within the artificial intelligence community.

What is striking with regard to connectionism is the fact that a critique of such a brain physiological materialism is almost entirely missing in neuroinformatics (cf. Dreyfus 1992: 159–162)[6]; despite the fact that the principles of connectionism can be traced back to the early days of scientific psychology, and have been subject to profound criticism ever since. In particular, psychoanalysis tried to distance itself from a biological reductionism that was dominant in the late 19th century. To reduce intelligence to a biological level and explain it in terms of neurons and its synapsis is counter-intuitive to a theory which links the human mind to cultural processes. Hence, in the following I want to take up the idea of a "psychoanalysis of things" (Sartre 1992: 765) and apply it to artificial intelligence and machine learning. I think that this approach can help us to better understand some of the hidden layers within the A.I. debate and to reveal a central mechanism in the psycho-economy of our socio-technological world. The question of "Who speaks?", central for the analysis of paranoia (cf. Lacan 1993: 54), becomes paramount at a time, when algorithms, in the form of artificial neural networks, operate more and more as secret agents.

Recurrence of the Symbolic

Reducing intellectual capacities to the material structure of the brain, as is the case in current A.I. debates, is deeply problematic, not least because it leads to a revival of biologist thinking. Brain physiology goes back to the very beginning of modern psychiatry in the late 19th century, a time, when mental illnesses were considered to be diseases of the brain. It was none other than Sigmund Freud, who, in 1895, drafted a "psychology for neurologists" (cf. Freud 1960), in which he described, for the first time, his idea of a psychic apparatus. Contrary to the dominant discourse of his time, Freud was less interested in the material structure of the brain, but rather in the functioning of neurons. He depicted psychic processes as neural transmissions, thus similar to the current understanding of artificial neural networks. However, it is important to note that Freud's apparatus, in particular in his later elaborations (Freud 1969; 1978), is by no means a merely endopsychic phenomenon, but always also an exogenous one. The apparatus consists of three parts (id, ego, super-ego), which are intertwined and yield the psychodynamic foundation of the human mind. This is the reason why psychoanalysis positioned itself against all forms of brain-physiologist materialism and similar attempts to depict the psyche by simply looking into the brain. Psychoanalysis

6 To be fair, no one in neuroinformatics seriously claims to recreate the human brain. However, the brain provides a certain model, upon which artificial neural networks are built (cf. Bruder 2017).

does not constitute the psychological as something interior, but theorises it from a relational perspective, which looks at the interplay between the inside and the outside; and takes into account not only the relationships between psychic human beings, but also their relationship to other technical apparatuses, without denying the resulting alterity, strangeness and non-integrability.

For psychoanalysis, the relation between discursive functions, psychical relations and technical media has always been central. One has only to consider Freud's recurring references to technology, be it electricity, optical media, or the *Wunderblock*, a child's toy comprising a wax tablet and a sheet of cellophane (cf. Freud 1961). These technical references are not simply metaphors, but constitute the theoretical foundation for the unconscious (cf. Vogl 2004). With the emergence of electronic media around 1900, the last remnants of the classical "language system" were transferred to an automated "recording system" (cf Kittler 1990: 229–238). It is therefore no coincidence that psychoanalysis, which promised to record unconscious phenomena such as dreams, delusions or sexual desires, emerged at the same time audio-visual recording devices were invented. Psychoanalysis' focus on language, not in terms of an ideal space of bourgeois enunciation, but rather as a messy realm of insinuations, can be seen as an attempt to inquire into technology as a necessary precondition to connect us to other humans, but also non-human beings. To make this point clear, let me briefly remind us of Gilbert Simondon's thesis, which goes to the very heart of the human-machine relationship, currently renegotiated under the term artificial intelligence. For Simondon, the psychic individual is always also a social being – we individuate ourselves at once *psychically* and *collectively* (cf. Simondon 2005). In order to do so, technology is needed, because technology refers to the pre-individual that is collective experience such as language. This is an important realisation of psychological thinking, because it explains – at least to some degree – the complex relationship between psychic, technical and social individuals. In Bernard Stiegler's words, the "formation of the psychic apparatus makes the individual pass through the circuits of transindividuation that weave and metastabilise collective individuation" (Stiegler 2012: 9 f.). In this sense, language, as the material basis of collective individuation, constitutes an outside that constantly inscribes itself upon and within the psychic individuals.

Jacques Lacan's structural psychoanalysis avows for such a language-based approach. In opposition to mere talking cures, Lacan propagates a return to Freud, by radicalising and systematising Freudian theory. The encounter between Lacanian psychoanalysis and French structuralism, in particular structural linguistics, has led to a reorientation of psychological theory after the Second World War. What is crucial here is that Lacan takes the Saussurean concept (cf. Saussure 1983) and turns it upside down: it is not the signified that holds control over the signifier within the notion of the linguistic sign, but the signifier determines the signified. Language, for Lacan, finds its realisation in articulation, rather than in a symbolic representation of the world. Only from an endless chain of signifiers, that is a network of articulated signs, do concepts or ideas draw their coherence. Hence, Freud's conception of psychoanalysis as a treatment through speech gets radicalised by Lacan, insofar as he conceives the unconscious to be structured as a kind of language (cf. Lacan 1998: 48). The speaking subject is subordinated to signifiers and the structure of language, which is mirrored in the unconscious of the subject. This differs fundamentally from a mere empiricist understanding, which indeed shares the critique of rationalistic world-representation, but goes against the idea of the symbolic, as we have seen in relation to artificial neural networks. Consequently, structural psychoanalysis may provide a way to address the symbolic without falling back to good old-fashioned artificial intelligence. As Lacan famously noted: "The symbolic world, that is the world of the machine" (Lacan 1991: 47). The psychic apparatus, in this perspective, is seen as an information-processing machine, which, by using signs and language, allows for the communication between subjects.

By taking a psychoanalytical understanding of language into account, one could argue that the symbolic actually never left the A. I. discourse. Despite current attempts to go beyond a symbolic information-processing model of the mind, it is language, or more specifically code that shapes and forms neuroinformatics. In this sense, the practice of information processing as well as the engineering of information systems is inextricably intertwined with the advent of the digital computer, a machine designed to manipulate signs. Computers make it possible to put numbers on the real, insofar as "[s]traightforward encoding transfers unlimited chance (the real) into a syntax with requirements and exemptions, that is, with laws" (Kittler 2012: 141). From this point on, computers are able to simulate reality, simply because they deal with probability.[7] Alan Turing's "Imitation Game" (cf. Turing 1950) can therefore be seen as a founding principle of computer-based intelligence, not least because it links the ability of human cognition back to the formulation of a universal discrete machine (cf. Turing 1936). For Turing, the central question is not so much whether computers can *think*, but whether machines can *do* what we do, that is to generate (or at least mimic) cognitive capacities. Turing's little test does not want to resolve the question of what intelligence or thinking are, but whether a "thinking machine" can be built or not. If a machine acts as intelligently as a human being, then it has to be considered as intelligent as a human being. Consequently, in present computer science, artificial intelligence is defined in terms of intelligent agents that perceive their environment and act to maximise their chances of success (cf. Russell/Norvig 2003: 4–5). Such an agent-based model, as can be seen in computational simulations of neural networks, is indeed rule-based, but also bottom-up, in the sense that it approaches intelligence without specific terms of reference (cf. Marchi/Page 2014). However, the question

7 Lacan, in this context, speaks of "conjectural sciences" instead of human sciences (Lacan 1998a: 43). Compared to "exact sciences" they do not only deal with facts, but also with speculation.

remains how and to what extent the notion of (human) intelligence is itself shaped by today's media technological condition.

Most recently, David Berry proposed a media theory of machine learning (cf. Berry 2017). For him machine learning systems "have an additional agenda which is the ability to create new algorithms, [...] that is, that they can construct a model of a 'world' of data and functions to transform them" (ibid.: 74). If this is actually leading to a new, self-aware form of computation, is yet to be seen, but we are, without a doubt, facing a new technological condition, within which more and more processes, formerly associated with human intelligence, get automated by computing machines. Here, Freud's psychic apparatus may shed some light on the issue, because it is, as was shown, conceptualised as an information-processing machine, which makes it possible to take the process of transfer from the conscious to the unconscious into account. The apparatus automates external influences by making them part of the internal functioning. Hence, psychoanalysis' claim that the conscious mind is only the imaginary inner view of exterior conditions. These conditions, to quote Friedrich Kittler, are ultimately media standards: "Consciousness is tied to the contingent presence of eyes or ears, to analog media; from the encoding of the real, on the other hand, the location of the other necessarily emerges – a combinatorial matrix of strategies" (Kittler 2012: 142). The location of the other, that is the world of machines, yields the unconscious as its very own discourse. Psychoanalysis, in this regard, becomes a discourse analysis of technical media and implies that information machines constitute the symbolic order, which in turn forms the psychic apparatus as a medium of storage, transmission and processing.

Hence, the biggest illusion of the human subject may be that it has created and masters the symbolic order. Instead, by participating in the realm of language, signs and symbols, the subject gets split into a subject of enunciation and a subject of statement, that is a subject that is no longer constituted by an inner consciousness (cogito), but rather by an exterior that constantly writes and speaks it. Every time the subject thinks, that is when it seeks to articulate the symbolic order, it gets trapped within. The subject is not only dominated by the symbolic, but also constituted by it. Conversely, it is the symbolic that constitutes the world of the machine. The contraposition between symbolic A. I. and artificial neural networks thus poses a false opposition, because both of them, like the human subject, are subjected to the same media-technological condition. Even the most intuitive A. I. is still bound to the rules imposed by digital computers; and every symbolic task has to be performed under real conditions. We should therefore concentrate on the interrelationship between the inner and the outer, the psychic and collective individuation, in order to decipher some of the enigmas within the current A. I. discourse.

Learning Machines

One of the most fiercely debated issues in A. I. is the question, whether machines can learn or not. One could state that machine learning is a sub-category of artificial intelligence, and, at least for the moment, its most promising one (cf. Berry 2017). Despite the fact that the implementations of machine learning algorithms are as diverse as the field of computer science, the fundamental goal of machine learning lies in the inference of general rules from specific examples, with which the algorithms were initially trained: "That is, machine learning is essentially an inductive process based on the original empirical training data fed into the network inputs and carefully reinforced so that the network pattern matching achieves the desired aims" (ibid.: p. 79). In this sense, every algorithm has an input and an output. In classic computation the data comes in and the algorithm calculates the output. Now machine learning turns this process around: input and output-data generate the algorithm that turns one into the other.[8] Hence, the big expectations from these new techniques, as these are algorithms that promise to create other algorithms, or, put differently, programs that write their own programs. The self-learning capacity of machines is currently fuelling hopes of a general-purpose A. I., as it was mentioned in the beginning of this article. The next great leap in artificial intelligence is expected to come from so-called back propagation neural networks (BPNNs), which are supposed to self-learn the ability to identify correlations within vast amounts of data sets (cf. LeCun/Bengio/Hinton 2015).[9] These deep learning algorithms have shown significant advances in image recognition and speech recognition, two fields, which for a long time have caused some of the major problems within the artificial intelligence community.

The capacity of computers to learn without being specifically programmed lies at the very heart of these problems. In fact, the whole issue can be traced back to the aforementioned Turing text on the question of computational intelligence (cf. Turing 1950). In it, one of the main objections to Alan Turing's self-imposed question whether machines can think or not, or, to be more precise, whether they could play the *Imitation Game* successfully or not, is known as the "Lady Lovelace Objection" (cf. ibid.: 450–451): Ada Lovelace, as is well known, was sceptical about self-learning capacities of machines, in particular about their ability to self-write their own code. The Analytical Engine, a mechanical general-purpose computer

8 Based on Charle's Peirce's distinction between deduction, induction, and abduction, one could argue that current forms of machine learning, albeit being inductive processes, implement a 'weak abduction', in the sense that they allow for new hypothesis to emerge from the data set; however, they do not perform a 'strong abduction', because they cannot invent new rules (cf. Pasquinelli 2017: 12).

9 Back propagation, which is considered to be the driving force behind deep learning, is a recursive technique and dates back to the 1980s (cf. Rumelhart/Hinton/Williams 1986).

designed by English mathematician Charles Babbage in 1837, was, in Ada Lovelace's account, built to follow instructions written by the programmer rather than to create anything itself. Hence, the machine has no capacity to anticipate, let alone originate rules or principles. According to Turing, this argument can be boiled down to the assertion that machines "can never 'take us by surprise'" (ibid.: 450). However, for Turing, the problem is mainly one of programming. By comparing a computer with the human brain, he points out that it may be more appropriate to consider a child's mind instead of an adult one when trying to produce a programme to simulate human intelligence. Starting with the initial state of the mind allows us to reflect on the whole process of education, as well as other experiences, to which it has been subjected to, until it reaches its later state. The problem is therefore broken down into two parts: the initial programming of a childlike mind and its following education.[10] The latter can be seen as learning, or teaching process that includes "punishments and rewards to obey orders given in some language, e. g. a symbolic language" (ibid: 457).

Turing's vision eventually sparked a quest for finding artificial intelligence, which has gone all the way to today's neural networks. Like his child-machines, these networks are subjected to symbolic language and therefore the features within it. Now features, exactly because they are part of the symbolic realm, can be designed, as is the case in almost all machine-learning procedures. This is an important fact, because it points to the problem of meta modelling in artificial intelligence. As any reasonable machine teacher will tell you, artificial neural networks, like kids, learn from the features of the world around them (cf. Domingos 2012). And if this world is a specific set of data, then everything that goes into this training set will most likely influence the development process of the network. With self-learning techniques, such as back propagation, machines do the learning, but there remains a strong human element in the sense that human engineers design the training sets which are fed to the machines: "Even if neural networks show correlations unforeseen to the human mind, they operate within the implicit grid of (human) postulates and categories that are in the training dataset and, in this sense, they cannot make the necessary leap for the invention of radically new categories" (Pasquinelli 2017: 11). Matteo Pasquinelli therefore concludes: "For sure, they do not show signs of 'autonomous intelligence' or consciousness" (ibid.: 14 f.). But then again, the question of consciousness may steer the discussion in the wrong direction. As was shown before, the conscious mind, that is the psychic apparatus is not a pre-given thing, but evolves

10 Interestingly enough, Turing states that teaching such a machine cannot be exactly the same as teaching a child, because the machine is lacking legs and eyes to make its own experiences. Now with the new developments in robotics this has become possible, which is also the reason why a lot of hopes are put into the conflation of robotics and artificial intelligence, as can be seen in the example of Hanson's most recent robot Sophia.

through the interaction between the inner and the outer world. Thus, if a child has to develop internal mental capabilities such as problem solving, memory, and language, in order to progress from dependency to increasing autonomy, then there is no good reason why a machine shouldn't.

What if neural networks, by articulating the symbolic order, develop a psychic apparatus on their own? This is not as science fiction as it may seem, considering that machine learning algorithms are now being augmented with a long short-term memory (LSTM),[11] in order to remember inputs for a longer period of time: "Beyond simple memorization, neural Turing machines and memory networks are being used for tasks that would normally require reasoning and symbol manipulation" (LeCun/Bengio/Hinton 2015: 442).[12] As development psychology shows, cognitive capacities are fundamentally linked to an increasing processing efficiency as well as a functioning working memory. The latter is a cognitive system with a limited capacity to hold information that is currently being processed. Very similar to a computer's random-access memory (RAM), this short-term memory temporarily stores information for immediate manipulation. In contrast, long-term memory, the computer's hard drive or read only memory (ROM), permanently keeps the information, in order to be available for later use. The split into a primary and secondary memory, between RAM and ROM solves, according to Kittler, one of the main issues in Freud's theory, namely to think of an apparatus capable of storing and transmitting data at the same time (cf. Kittler 2012: 133). The computer, by introducing a central processing unit (CPU), unravels the mystery of doing both storing and forgetting, of mediating between the conscious and the unconscious mind. Now simulated within artificial neural networks, such a long short-term memory is used to *remember* values over arbitrary time intervals, in order to deal with memory loss when training networks with back propagation through time.

These recurrent neural networks (RNNs), which have internal memory to process arbitrary sequences of data, are mainly used in speech recognition as well machine translation. Google Translate, for example, is using this technique, with the effect that the machine is now able to capture context and *understand* – in the Turing sense – natural languages. Yet, if we want to break with such a cognitivist or connectionist approach, within which the understanding of language is simply a question of network size, we have to consider the technological unconscious. As was mentioned before, neural networks are very sensitive on the data that they are trained on. This is a well-known effect in machine learning, called over-fitting: "Everyone in machine learning knows about overfitting [...]. One way to understand [it] is by decomposing generalization error into *bias* and *variance*" (Domingos 2012: 80). While the first refers to the problem that a network, once

11 LSTM was first proposed by Sepp Hochreiter and Jürgen Schmidhuber (1997).

12 Neural Turing machines, because the neural network is augmented by a "tape-like" memory, similar to the one Turing machines read from or write to.

trained with an error, will tend to constantly learn the same wrong thing, the second indicates the tendency to learn random things independent of the training set. Both can be seen as unconscious effects of networks, trying to make sense of the world around them. Now human language in particular is full of ambiguity and prejudices, which imply that machine learning algorithms are constantly fed with stereotyped biases that mirror our everyday culture (cf. Caliskan/Bryson/Narayanan 2017). What we teach these algorithms ultimately reflects back on us and it is therefore no wonder, when artificial neural networks start to discriminate on the basis of race, class and gender bias (cf. Apprich 2018).

The problem of biased data refers back to the issue of feature engineering: "Often, the raw data is not in a form that is amenable to learning, but you can construct features from it that are. This is typically where most of the effort in a machine learning project goes" (Domingos 2012: 82). The learning process, in other words, implies knowledge and assumptions beyond the data.[13] Hence, the determination of the desired output is ultimately a prediction based on the input given by the engineers. This has two major implications: The first relates to a macroeconomic level, as machine-learning technologies are now applied to many different aspects of value aggregation in digital cultures, from content filtering in social media and web search, to recommendation systems on Internet platforms and exchange sites, to automation systems in transport and logistics. They do not only identify objects in images, transcribe speech into text, or select news, but also drive cars, make medical diagnosis, and are foreseen to take care of the elderly. As private corporations build most of these technologies, the decisions made by their owners, employees and shareholders will have huge effects on the way we will live together in the future. The second implication is on the microeconomic level and concerns the new division of labour, or, to be more precise, the concealment of it. In fact, the labeling of data may be the most costly part and yet the most unrecognised aspect in machine learning. It is important to bring to mind how time-consuming it is to gather, clean and normalise data, before machines even process it. Hence, one of the holy grails in machine learning is actually to automate this process. But, as Pedro Domingos reminds us, "machine learning is not magic; it can't get something from nothing" (ibid.: 80). For him, machine learning is more like farming, which begs the question how *natural* this process actually is?

13 This fact is, for example, reflected by Bayesian Neural Networks (BNNs), which draw on prior knowledge in order to minimize errors in deep learning approaches (cf. Mullachery/Khera/Husain 2018).

A. I. Fetishism

Today's major tech-companies, such as Google, Facebook, Apple, Amazon, Microsoft, Baidu and IBM, are setting the standards for artificial intelligence systems. These new research clusters, compared to classic academic institutions, are ultimately driven by the need to generate revenue. It is therefore no surprise that, at the end of the day, the profits from and power over artificial intelligence are in private hands; despite the constant talk about the benefit for society as a whole.[14] Making use of ever-growing amounts of data and turning them into products and services, is the new lifeline of techno-capitalism. Think about Apple's Siri, Facebook's M, or Amazon's Echo; these are commercial applications of machine learning algorithms. In order to extract information from data and answer user's requests, these assistants rely on an immense infrastructure that has to be built and maintained.[15] What is more, this infrastructure has to be hidden in order to sustain the illusion of a truly – that is fully automated – artificial intelligence. Take the example of the new iPhone X with facial recognition and other computer vision features. Behind it lies Apple's new neural engine, which relies on machine-learning algorithms, in particular convolutional neural networks (CNNs). These networks have shown tremendous progress in visual recognition tasks over the last years, not least by validating their outputs against ImageNet, the academic but also industry standard for the classification of images.[16] What is important to note in this context is the fact that in order to create such a high-quality image dataset, a lot of labour is needed, because every image of the currently more than 14 million images on ImageNet has to be hand-annotated to indicate what objects are pictured. Hence, ImageNet was at its peak one of the biggest employers on Amazon's Mechanical Turk (AMT), a platform to outsource tasks that computers are unable to do. The name of the platform comes from an 18th-century chess-playing automaton, which was designed by Wolfgang von Kempelen to impress the Austrian Empress, Maria Theresia. However, the machine was not an automaton at all, but a human chess player hiding in the machines' architecture. Likewise, a good part of intelligent agents in our smart devices may not be perceptrons of neural networks, nor the engineers in the laboratories, but badly paid workers,

14 See, for example, the "Partnership on AI to benefit people and society", founded by Amazon, Apple, Google, Facebook, IBM, and Microsoft.

15 In the case of Facebook's M, the assistant has, for a long time, been based on actual human employees, who pretended to be chat bots. Following Amazon's "artificial intelligence", a term coined to describe the outsourcing processes of a computer program to humans; one could speak of "smart assistants", given the fact that there is no limit to stupidity when it comes to profiteering.

16 Since 2010 the ImageNet Large-Scale Visual Recognition Challenge has been issues to evaluate image recognition algorithms. It has been considered the 'Nobel Prize' in image recognition. 2017 marked the final year of the competition.

hidden within the labour intensive infrastructures of digital cultures (cf. Irani/Silberman 2013).

Tech-companies have now started their own internal micro work platforms, in order to clean up messy databases for training machine learning systems.[17] As artificial intelligence becomes ubiquitous, the stakes in the battle over infrastructures are getting higher. Google has recently developed its new custom chips called Tensor Processing Units (TPUs) to boost the performance of machine-learning algorithms. Microsoft, Intel, Nvidia, and a bunch of start-ups are working on similar projects. The idea is not only to control the narrative of artificial intelligence, but also to secure the supplies for the extremely data-hungry machine learning algorithms. This is the reason why social media platforms have still a say in the matter. In fact, the whole industry is very much dependent on "free labour" (Terranova 2000) produced by everyday users. So behind the appearance of an absolute autonomy of artificial intelligence the actual symbolic order comes to light: techno-capitalism and the denial of its dependency on labour. The recent developments in artificial intelligence and machine learning turn out to be the last push for a "frictionless capitalism" (Bill Gates), that is the idea that with media technologies the intrinsic problems of capitalism, caused by the antagonism between capital and labour, might get solved (cf. Schröter 2012). Hence, the continuous failure to acknowledge the labour that goes into the production of digital cultures. Instead its applications are fetishized, in the same sense Marx described the fetishisation of capital: "Because it is value, it has acquired the occult quality of being able to add value to itself. It brings forth living offspring, or, at the least, lays golden eggs" (Marx 1996: 165). The fantasy of a self-generating A. I. is closely linked to commodity fetishism, since neither the producer nor the consumer of smart products has a necessary or full relation with the other. The fetishisation of artificial intelligence forecloses a collective experience, which is always mediated by technology.[18]

It is not quite clear whether Marx inspired Sigmund Freud's use of the term fetish or not. But fact is that both Marx and Freud shared the same epistemological horizon, namely the proliferation of modern scientific knowledge in the course of the 19th century. Psychoanalysis' interest in the relation between the subjective inside and the social outside can therefore be read as a reflection on Marx' occupation with the proletarian subject. Lacan's "second return" to Freud then also goes through an engagement with Marx' critique of political economy (cf. Tomšič 2015). The formalisation of the structure of capitalism leads Lacan to a theory of discursive production, which tries to move on from structural linguistic. What is important in this reading is that capitalism positions itself as a self-engendering entity, thereby echoing Marx' critique of fetishism. In his book about the relation between Lacan and Marx, Samo Tomšič writes: "It [capitalism] inverts the position

17 See for example, Microsoft's Universal Human Relevance System (UHRS).

18 In keeping with Simondon's position, we can actually say that there is no society without technology and no technology without society.

of truth and agent, which makes the subject appear as an autonomous agent and the initiator of an infinite circulation, from which there is no breakout" (ibid.: 220). Capital positions itself as a vital power, one which contains the potential for a self-engendering system. And the same is true for the appearance of artificial intelligence, which, not coincidentally, is modelled upon the human brain, thereby evoking the idea of an autonomous entity. In the light of earlier attempts to conceptualise our techno-social world in terms of biology (cf. Kelly 1994), we have to ask why the discourse around artificial intelligence is exploding at a moment when the capitalist system has plunged into a crisis? May it be that the fetish of neural networks is ultimately obscuring the simple insight that a true general intellect has always been a collective endeavour? In keeping with a psychoanalytical perspective, the key to the understanding of artificial intelligence thus lies in the recognition and acceptance of the symbolic world as a world of machines. Because machines, understood as the material basis of trans individuation, go beyond the individualistic conception of the mind and allow for new relations between human and non-human actors. In the end, we all are the secret agents of artificial intelligence – we just have to accept it.

References

Apprich, Clemens (ed.): *Pattern Discrimination*, Lüneburg/Minneapolis: meson/University of Minnesota Press, forthcoming.

Berry, David M. (2017): "Prolegomenon to a Media Theory of Machine Learning: Compute-Computing and Compute-Computed." In: *Media Theory*, Vol. 1 (1), pp. 74–87.

Bruder, Johannes (2017): "Infrastructural Intelligence: Contemporary Entanglements between Neuroscience and AI." In: *Progress in Brain Research*, Vol. 233, pp. 101–128.

Caliskan, Aylin/Bryson, Joanna J./Narayanan, Arvind (2017): "Semantics derived automatically from language corpora contain human-like biases." In: *Science*, Vol. 356, pp. 183–186.

Domingos, Pedro (2012): "A Few Useful Things to Know about Machine Learning." In: *Communications of the ACM*, Vol. 55 (10), pp. 78–87.

Dreyfus, Hubert L. (1992): *What Computers Still Can't Do. A Critique of Artificial Reason*, Cambridge, Mass.: MIT Press.

Freud, Sigmund (1960): "Project for a Scientific Psychology (unfinished manuscript)." In: *The Origins of Psychoanalysis: Letters to Wilhelm Fliess, Drafts, and Notes*, 1887–1902, edited by Anna Freud and Ernst Kris, New York: Basic Books, pp. 345–445.

Freud, Sigmund (1961): "A Note Upon the 'Mystic Writing-Pad'." In: *Sigmund Freud, The Ego and the Id and Other Works*, translated by James Strachey, London: Vintage, pp. 227–234.

Freud, Sigmund (1969): *An Outline of Psychoanalysis*, translated by James Strachey, New York/London: W. W. Norton & Company.

Freud Sigmund (1978): *The Question of Lay Analysis*, translated by James Strachey, New York/London: W. W. Norton & Company.

Haugeland, John (1985): *Artificial Intelligence: The Very Idea*, Cambridge, Mass.: MIT Press.

Hochreiter, Sepp/Schmidhuber, Jürgen (1997): "Long short-term memory." In: *Neural Computation*, Vol. 9 (8), pp. 1735–1780.

Irani, Lilly/Silberman, Six M. (2013): "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk." In: *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, pp. 611–620.

Kelly, Kevin (1994): *Out of Control. The New Biology of Machines, Social Systems, and the Economic World*, New York: Basic Books.

Kittler, Friedrich (1990): *Discourse Networks 1800/1900*, translated by Michael Metteer, Stanford: Stanford University Press.

Kittler, Friedrich (2012): "The World of the Symbolic – A World of the Machine." In: Friedrich Kittler, *Literature, Media, Information Systems*, New York: Routledge, pp. 130–146.

Lacan, Jacques (1991): The Ego in Freud's Theory and in the Techniques of Psychoanalysis, Seminar II, translated by Sylvana Tomaselli, New York/London: W. W. Norton & Company.

Lacan, Jacques (1993): *The Psychoses, Seminar III, translated by Russell Grigg*, New York/London: W. W. Norton & Company.

Lacan, Jacques (1998a): *The Four Fundamental Concepts of Psychoanalysis, Seminar XI*, translated by Alan Sheridan, New York/London: W. W. Norton & Company.

Lacan, Jacques (1998b): *On Feminine Sexuality: The Limits of Love and Knowledge, Seminar XX*, translated by Bruce Fink, New York/London: W. W. Norton & Company.

LeCun, Yann/Bengio, Yoshua/Hinton, Geoffrey E. (2015): "Deep Learning." In: *Nature*, Vol. 521, pp. 436–444.

Lewis-Kraus, Gideon (2016): "The Great A. I. Awakening." In: *New York Times*, December 14, 2016, https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

Marchi, Scott de/Page, Scott E. (2014): "Agent-Based Models." In: *Annual Review of Political Science*, Vol. 17, pp. 1–20.

Marx, Karl (1996): *Capital. A Critique of Political Economy*, Volume 1, London: Lawrence and Wishart.

McCulloch, Warren S./Pitts, Walter H. (1943): "A Logical Calculus of the Ideas Immanent in Nervous Activity." In: *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115–133.

Minsky, Marvin (1985): *The Society of Mind*, New York/London/Toronto: A Touchstone Book.

Minsky, Marvin/Papert, Seymour A. (1969): *Perceptrons. An Introduction to Computational Geometry*, Cambridge, Mass.: MIT Press.

Möbius, Paul J. (1903): Über den physiologischen Schwachsinn des Weibes, Halle a. d. S.: Verlag von Carl Marhold.
Mullachery, Vikram/Khera, Aniruddh/Husain, Amir (2018): "Bayesian Neural Networks." In: *arXiv*, https://arxiv.org/abs/1801.07710.
Pasquinelli, Matteo (2017): "Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference." In: *Glass Bead Journal*, Site 1, http://www.glass-bead.org/wp-content/uploads/GB_Site-1_Matteo-Pasquinelli_Eng.pdf.
Rosenblatt, Frank (1958): "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." In: *Psychological Review*, Vol. 65 (6), pp. 386–408.
Rumelhart, David E./McClelland, James L./PDP Research Group (1986): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge: MIT Press. 1986
Rumelhart, David E./Hinton, Geoffrey E./Williams, Ronald J. (1986): "Learning representations by back-propagation errors." In: *Nature*, Issue 323, pp. 533–536.
Russell, Stuart J./Norvig, Peter (2003): *Artificial Intelligence: A Modern Approach*, New Jersey Prentice Hall.
Sartre, Jean-Paul (1992): *Being and Nothingness. A Phenomenological Essay on Ontology*, translated by Hazel E. Barnes, New York: Washington Square Press.
Saussure, Ferdinand de (1983): *Course in General Linguistics*, translated by Roy Harris, London: Duckworth.
Schröter, Jens (2012): "The Internet and 'Frictionless Capitalism'." In: *tripleC*, Vol. 10 (2), pp. 302–312.
Simondon, Gilbert (2005): *L'Individuation à la lumière des notions de forme et d'information*, Paris: Jérôme Millon.
Stiegler, Bernard (2012): "Relational Ecology and the Digital Pharmakon." In: *Culture Machine*, Vol. 13, https://www.culturemachine.net/index.php/cm/article/view/464/501.
Terranova, Tiziana (2000): "Producing Culture for the Digital Economy." In: *Social Text*, Vol. 18 (2), pp. 33–58.
Tomšič, Samo (2015): *The Capitalist Unconscious: Marx and Lacan*, London/New York: Verso.
Turing, Alan (1936): "On Computable Numbers, with an Application to the Entscheidungsproblem." In: *Proceedings of the London Mathematical Society*, Vol. 42, pp. 230–265.
Turing, Alan (1950): "Computing Machinery and Intelligence." In: *Mind*, Vol. 49, pp. 433–460.
Vogl, Joseph (2004): "Formationen des Wissens. Zur Einführung." In: Claus Pias/Joseph Vogl/Lorenz Engel/Oliver Fahle/Britta Neitzel (eds.), *Kursbuch Medienkultur. Die maßgeblichen Theorien von Brecht bis Baudrillard*, Stuttgart: DVA, pp. 485–488.