Demythologizing Artificial Intelligence[1]

Reflections on the Role and Purpose of Complex Data Processing in Digital Media Transformation

Jonas Bedford-Strohm

»Digital technology begins where the world can be represented in data in order to perceive patterns and structures that cannot be noticed by the human eye and the natural mental capacity to perceive and compute,«[2] Nassehi observes. Once this world is precariously duplicated into data, this representational world develops a life of its own. This duplicate »reality of its own kind«[3] is only loosely related to what we consider our »original« life world. The duplicate reality is self-referential in that it can only communicate or relate to that which is external if it comes in its own form. Data »know the world only in their own data form and cannot escape from it. Everything that appears in it must take data form itself.«[4] Thus, data can process the world only in its own image.

---

1 Earlier results of the research for this article are published in German: Bedford-Strohm 2019.

2 Nassehi 2019: 229. The German original: »Die Digitaltechnik fängt dort an, wo sich die Welt in Daten repräsentieren lässt, um Muster und Strukturen zu erkennen, die mit bloßem Auge und den Wahrnehmungs- und Rechenkapazitäten des natürlichen Bewusstseins nicht erfasst werden können.«

3 Nassehi 2019: 114. The original: »Realität eigener Art«.

4 Nassehi 2019: 111. The original: »Daten … [sind] in besonders radikaler Weise auf sich verwiesen …, denn sie kennen die Welt eben nur in ihrer je eigenen Datenform und können daraus nicht ausbrechen. Alles, was dort wieder vorkommt, muss selbst Datenform annehmen.«

Jonas Bedford-Strohm 2022: Demythologizing Artificial Intelligence. Reflections on the Role and Purpose of Complex Data Processing in Digital Media Transformation. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 55–80.
https://doi.org/10.11588/heibooks.1106.c15193

Hence, we are faced with problems of representation because all simplistic notions of »original« and »duplicate« are rendered precarious. The digital twin in data form is never a perfect representation, rather it is a precarious duplicate that impacts what it duplicates, especially when the object of representation is a human subject. Hence, Nassehi speaks of duplication (*Verdopplung*) as an »ironic concept, since [...] what appears to be a duplication in practice turns out to mean the exact opposite: a new creation of something that only exists by being duplicated.« In this practice, he notes, »we stabilize our life world by duplicating it and pretending that it is as it appears.«[5]

Since the widely applied forms of artificial intelligence rely on the complex data processing techniques of machine learning, Nassehi's observations about data can help us identify key problems in the application of artificial intelligence. If, the reasoning goes, we are able to understand AI's most vital processing *resource* (data), we can better understand the processing *techniques* that we call »AI.« Hence, we shall explore the »stuff« that makes artificial intelligence algorithms effective first, before we zoom out to further explore the conditions under which these algorithms are deployed.

If we take Nassehi's ideas seriously, we are challenged by (at least) two sets of questions – one conceptual, one ethical: 1) What do we mean when we speak of data? How can we understand this form that is so self-referentially totalitarian that it will only accept communication with the world when it takes its own form? 2) What does the process of duplication or representation in data form look like and imply? What are its conditions? Who can trigger this process? Who can navigate and utilize the »duplicate« reality it creates? And why would they be incentivized to do so?

The former challenge takes the form of a conceptual exploration: We need to reconstruct and understand what we are speaking of, before we evaluate and analyze its uses and impacts. The latter challenge takes the form of an ethical exploration: We need to notice the uses and impacts of what we reconstructed conceptually, and critically probe the conditions for its possibility. We will therefore first define our concept of data (A) and reconstruct data processing in the form of a four-fold typology (B). Only then will we trace relevant contentions of critical data theory (C) and explore the ethics of complex data processing (D) by analyzing the necessary conditions of its practice.

---

5 Nassehi 2019: 113. The original: »Verdopplung ist gewissermaßen ein ironischer Ausdruck, weil er auf die Paradoxie aufmerksam macht, dass das, was praktisch als Verdopplung erscheint, exakt das Gegenteil bedeutet: eine Neuschöpfung von etwas, das nur dadurch existiert, dass es verdoppelt wird. ... Wir stabilisieren Lebenswelten, indem wir die Welt verdoppeln und zugleich so tun, als sei sie so, wie sie lebensweltlich erscheint.«

A Definitions of Data

In the German subculture of those who deal with data as a feature and resource of digital society, the international buzzwords Big Data, Artificial Intelligence and Deep Learning, but also German terms like *Datenschutz*, *Datenleck* und *Datenverarbeitung* are used ubiquitously. In the vast sea of thematic content, however, precise definitions of the discussed terms are surprisingly rare – despite their ubiquitous use. Although the term »data« in its various forms is employed so ubiquitously, its use, as with many terms of everyday use, remains blurred. In the German discussion, »data« usually means: *digital micro-packets of communication that can be stored electronically and, with the right methods of interpretation, become substantive information in the right context.*[6]

If we understand data in this way, the relationship between analog and digital dimensions in the source and structure of data comes into focus with the criterion of electronic storability.[7] It implies that only *electronic* data is to be understood as data in a meaningful sense. But the word »date« (*Datum*), derived from the Latin word »dare« which translates to the English »to give,« permits a diverging definition: One single »date« as the elemental building block of the multitude of »data« is simply a single instance of something »given« – the Latin participle perfect of »dare« is *datum*.

A »date« or »datum«, therefore, is an entity that appears distinct from other entities and yet forms part of an integrated, perceivable realm of specifically structured experience – the precariously »duplicated« world Armin Nassehi writes about. In this respect, »the given« can be characterized by the romantic concept of the *unity of unity and difference*. Data is comprised of a vast multitude of singular instances of given information yet forms an integrated unit – the sum is more than its parts. This is illustrated by the fact that a given »date« occurs almost exclusively in the plural »data« in colloquial discussion – we say »a piece of data,« rather than »one date,« even though it would be the usual way of applying singular and plural in the English language. Like the quantum in physics, the »date« never seems to exist in isolation, but always as part of a larger horizon of meaning. In this respect, the use of the plural is not only grammatically correct, but epistemologically meaningful.

Hence, we can recognize data as a collective multitude of communicative micro-packets that can, at least potentially, be mapped in *quantitative* structures. Taking this preliminary reflection into account, a refined version of our

6 This definition is, in fact, already more nuanced than the definition employed in most content on the subject in the popular media and everyday language, because it is based on Joseph Weizenbaum's critical theory of information. Cf. Weizenbaum 2001, especially chapter 1 on information and meaning.

7 On the »conditions of existence« for specific media forms, see Parikka 2012: 6.

colloquial terminology emerges. For our purposes here, we shall understand data as *digital and analog communicative micro-packets that can, in principle, be stored electronically due to their quantifiable structure and (when interpreted with adequate methods under the conditions of shared grammar and semantics) can become meaningful information, which can guide action and thus can impact the material or embodied practices of communicative agents*.

B Typology of Data Processing

For our purposes, we do not reduce the concept of data to digitally generated, electronically stored information, but rather define data broadly enough to take a *multimedia* perspective, including analog forms. For practical purposes, we cannot place the general concepts of rationality or communication at our theoretical center, because both would require a solid cognitive-theoretical foundation, which is impossible to adequately deliver here. We will thus limit ourselves to data processing in *media*, which appears in the daily as both task and tool. A variety of different typologies have been employed to draw out different functional dimensions of data processing. One might, for instance, consider the distinction of data storage in 1D or 2D arrays. Or one might point to the distinction of data transmission into serial and parallel transmission. Because, however, our epistemic interest is to draw out the difference in practical impact on media transformation for different types of data processing, we require a different set of types, because neither the form of storage, nor the form of transmission can suffiently expose the impact and uses that different types of data processing might have for its practical application in media. For our purposes, it will suffice to distinguish four ideal-type forms: linear, variable, spatial and explorative data processing.

I Linear

A large part of human media history is shaped by a type of data processing that we can characterize as *linear*. This linear type of data processing manifests itself in all forms of end-to-end communication between individuals or between the individual and mass media broadcasting (especially radio and television). Such end-to-end communication is more or less successful when based on shared code. This includes the often-subconscious socio-cultural coding of semantic contents in their transportation through media, but also the conscious technological encryption with cryptographic intention.

Due to the end-to-end structure of linear data processing, medial actions

of this type are difficult to scale. Linear data processing is nevertheless ubiquitous in all human social practice. Every simple form of transported messages between individuals displays the characteristic of linearity – be it the one-to-many mode of a broadcast model or the one-to-one mode of a simple communication model. A holiday postcard, for example, might be classified this way. The linear characteristic might also be used to reconstruct the more complex media form of a newspaper, which is created in the editorial room of a publishing company, produced at the print shop and then directly sent to households.

Even the basic functions of the internet are based on this type of linear data processing. The TCP/IP suite, the collection of foundational internet protocols, is based on this principle. On the internet, individual hosts send small data packets to an address via a digital network, much like the holiday postcard. Using TSL/SSL encryption, the information on these digital postcards can be encrypted, even on an open network like the internet, thus enabling the relatively secure transmission of confidential communication in linear form. Even after the invention of more complex data processing technologies, linear data processing remains the foundation and majority of digital communication. With Nassehi, we can note that precisely its *simplicity* is what explains digital technology's *ubiquity*.[8]

What used to be stored on paper in address books is now often stored electronically, but the basic structure of data processing remains the same.[9] This can be illustrated by the example of the electronic mailing list: An initial communication agent sends the general message to a more or less specified audience that *serial* linear communication is possible in the form of a newsletter, for instance by advertising it on a website. The recipient then transfers a contact address to the address book of the initial agent, for instance by typing it into a contact form on a website. In this way, the initial agent collects a multitude of linear contact addresses and begins a serial *broadcast* of linear messages.

From a data protection perspective, the decentralized nature of the data is particularly noteworthy here. In the form of a silo, the address books of the various communication agents are stored separately. The strategic use of this

8 He hones in on the quantitatively infinite possibility of recombination with the simple binary code that makes up all digital technologies at their core. Cf. Nassehi 2019.

9 Again, Nassehi picks up this thought and develops it further than we can here. His theoretical approach is to ask what societal conditions needed to be in place for digital technology to be adopted at such a swift pace and high rate. This leads him to conclude that, in fact, the very foundation of modern society is digital in structure, which in turn explains why digital technology could function as an effective problem solver in this society. Cf. Nassehi 2019. See chapters 1 and 2 especially.

data for pattern recognition therefore remains limited.[10] And even if one of the silos is attacked, the other silos remain secure. Linear data processing, therefore, implies a kind of safeguarding clause: One compromised silo does not automatically compromise other silos. In this respect, the linear use of data can be classified as fairly secure. However, the linear use of data is also impractical for many use cases, because the findings from such data remains limited.

The only strategic analysis possible is *category formation* and *individual analysis*. The example of itemized telecommunications bills can explain this. The data generated from itemized connections is only informative if the identity of both the contacting and contacted agents is known and available for investigation with a concrete epistemic interest (meaning: you know what you are looking for). If, for example, police want to check an alibi after some type of criminal offence, the registration of the linked mobile device in a mobile cell tower far away from the crime scene for the purpose of a telephone call with an unconnected third person offers strong indication that the alibi is valid. For such simple analysis with a pre-existing epistemic interest, the linear data processing of individual end-to-end data flows is sufficient. However, inquiries beyond individual analysis and category formation require more sophisticated data processing, especially when one does not know what exactly one is looking for.

II Variable

Data processing becomes more complex when a vast amount of data is available from linear data processing and a system for high level pattern analysis is developed strategically. Such methods were invented long before computer-based technologies. The indexing of analog crime scene photographs and the strategic comparison of murder weapons, murder methods and special features of a crime, for instance, can lead to more complex data processing based on the linear type.

Applied to our telecommunications example: Through *data aggregation* and stray search for extraordinary prevalence of certain types of actions, the system can *identify patterns*, thereby allowing investigators to deduce potential habits and strategies. In this case, the individual pieces of data become *metadata* in aggregated form and thus can serve as an ideal basis for more complex forms of variable data processing. In such processing, one or more vari-

10 This explains both why Facebook intends to combine user data from all its services that hitherto had remained separate – the value of the data increases manifold once it is combined into a shared silo. This also explains why it is so controversial.

ables are defined in a fixed formula, which ensures that a change in *signals* actually shows up as a change in the *result* in *real-time*. As formula-based data processing, we may consider simple algorithms that include a variety of forms and quantities in the variables and thus renders differences in incoming signals visible in the final result.

Practical examples of variable data processing are simple personalized purchase algorithms such as the book suggestion function on Amazon's online retail platform. When the potential buyer places a book in the shopping cart or even just clicks on it to read through the description, data is generated through these click signals which indicates interest in this specific book's general category. If a large amount of such individual pieces of data is available from other users, Amazon can determine which other books have also been viewed or added to the shopping list in similar purchase processes. Based on the known formula of these purchasing patterns, Amazon can develop a personalization algorithm suggesting interesting books to new customers: »If user A clicks on book X and user B has bought book X and Y at the same time, then suggest book Y to user A as well.« Although the principle is simple, it is based on the sophisticated variable inter-relation of linear end-to-end types of data processing.

III Spatial

While both linear and variable data processing is largely based on simple algorithms,[11] many industrial algorithms show the characteristics of *spatial* data processing. The application of such algorithms does not result in a 2D visualization as in Facebook's News Feed, but in a multidimensional rendering. A practical example for this form of data processing are 3D printers, which are able to bring linearly stored data structures into spatial application by means of multidimensional blueprints in a computer program. Just as in linear and variable data processing, the data is broken down into small and simple communicative micro packets. And yet the applied algorithms are able to draw a coherent spatial picture from this complex multitude of information packets in order to reconstruct them materially.

Many different branches of industry use this type of data processing on a daily basis. Stress tests of manufacturing materials and prototypes in aero-

11 Even if scaled and connected into more complex algorithmic systems, the basic operations are in the form of simple formula-based algorithms. The only deviation from this rule is the application of machine learning on top of the formula-based processing. Certain personalization algorithms (e.g. Facebook's News Feed) are increasingly applying machine learning and technologies from spatial and explorative data processing.

space engineering, for example, cannot be performed physically with sufficient replication – either due to prohibitively high cost or simply a lack of time. Therefore, complex mathematical models are used to simulate the physical effects of wear and tear in order to calculate where reinforcements have to be made and where weight can be saved to increase efficiency in fuel usage and material cost.

Another instance of spatial data processing are the calculation and visualization programs for architects, product designers, vehicle engineers, structural engineers and meteorologists. In these fields of application, operational safety is of particular importance from an ethical perspective. Since public infrastructure, product application, vehicle operation, building usage and weather calculations often impact the chances of survival in emergency situations, the inviolability of the person is of utmost importance in this form of data processing, given the foundational consensus of the modern concept of personal dignity. In applications of medical and scientific research, the ethical questions of the good life and holistically life-enhancing strategies for spatial data processing are even more evident.

IV Explorative

The category of *explorative* data processing shall summarize the technologies known as »artificial intelligence« in popular discourse. Usually, the term artificial intelligence means a more or less intelligent algorithmic system that, in most cases, is trained by humans with annotated training data and recognizes patterns in these data sets with none or little structure. The recognized patterns are then applied to new incoming data and if the domain area of this incoming data matches the domain of the training data, these patterns can produce meaningful insights for the human employing such systems. Such machine learning techniques are commonly called »artificial intelligence« because a human being could never have manually defined the patterns recognized by the system and thus required skill augmentation by human-made technological tools, in this case: artificial in the sense of »made,« »created« or »built« intelligence. When the searchable data set is so vast that human beings cannot go through it themselves, explorative data processing with machine learning is exponentially more powerful than any category formation in linear data processing could ever be.

The »artificial intelligence« system, in such cases, is nothing like the mystical all-powerful god-like general intelligence portrayed in popular culture and requires a very narrowly specified domain to function appropriately. A facial recognition system will likely produce gibberish if applied to music and

a music recognition system will likely produce gibberish if applied to images. But the technologies employed, even if limited to a narrow domain, have powerful properties that impress humans enough to inspire the title *artificial intelligence*. In our example, the machine learning system assumes the role of *detective*, processing a super-human amount of information simultaneously, as well as the role of *analyst*, interpreting the recognized patterns through *quantitative* means which can then be augmented by human *qualitative* analysis to produce an end product deemed in many cases superior to human analysis without computer assistance.

The relevant methods for explorative data processing are mainly data mining techniques employing machine learning methods, machine learning methods. In more complex tasks, these might take the form of deep learning, which attempts to mathematically map and functionally imitate the neuronal structures of the human brain.[12] If paired with high-speed computing power, deep learning can vastly outperform machine learning (for instance in machine translation of natural language). But for many simpler applications, machine learning comes close and is the more resource-efficient option. For many personalization algorithms in media platforms, for instance, machine learning methods will suffice.

The neural networks utilized in deep learning methods are designed for evolution and learn a certain intelligent behavior for a specific area of application, in some cases with the help of human trainers and always with large amounts of data. That is why the algorithms generated through these methods are categorized as *self-learning* algorithms. In contrast to the formula-based algorithms of variable data processing, artificial intelligence procedures are less of a strategic approach and more of an investigative, discovering, unstructured trial-and-error approach. This trial-and-error philosophy imitates the human learning curve marked by empirical experiences of pain and happiness.

C Contentions of Critical Data Theory

Much of the discourse on artificial intelligence has been (inadvertently or not) shaped by the product marketing initiatives of tech companies and euphoric researchers in search of funding on one side, and the fundamental critics of technology and big business on the other side, while politicians try to safely

12 Though brain scientists reject that metaphor, because computational systems require a stability that human brains never reach. To them, the attempt to imitate a dynamic, self-stabilizing system with a static, stable system (be it ever so fast in computation) is a dead end. See for instance Singer 2003. We should, therefore, consider the analogy more poetic device than scientific characterization.

tread on middle ground, reminding us of both the »risks and opportunities« in mantra-like fashion. Because the discourse is not yet broad enough to realistically mirror societal complexity, the discourse remains caught in extreme perspectives from either side of the polarized spectrum mixed with fearful repetition of vague set phrases that more often than not demonstrate a lack of technical knowledge (which further piles onto the reasons for politicians to be afraid of clear statements and initiatives for fear of ridicule). What would a critical theory of artificial intelligence look like? Could it be truly critical, in the sense of both critiquing practices and dialectically critiquing insufficient critiques of such practices? What topics would such a theory have to confront? Among them, certainly are these four: (1) autonomy, (2) transparency, (3) mythology, and (4) contextuality.

I Autonomy

No existing artificial intelligence system can rightfully be called *autonomous* if we follow the literal sense of *self-legislation*, derived from the Greek *autos* = self and *nomos* = law). Machine learning, at least, can still not do without human input, even if the human input is less than in linear or variable data processing. The utilized algorithm is not based on simple formulas with variables and signal prioritization explicated by humans. However, a framework and training data set must still be given to the system by humans in most cases. In short: AI does not just fall from the sky. To create powerful AI systems, immense amounts of human work, model training and optimization are necessary, thus begging the question: Is it really cheaper to invest in expensive AI systems for a given problem? Or is it more expensive to hire AI experts for a job that can be done by manual labor as efficiently?

An example for this is machine learning in community management on social media for publishers. In order to find patterns in comments, humans must define for the algorithm what the relevant data point is, such as the most common word in the trove of comments that is not a filler or sentence-connecting word, such as »like« or »and«. Alternatively, humans could optimize the algorithm to discover the most swift and steep increase in usage of a word, which can power trend analysis, because it could identify which increase occurred after a defined period of stagnation or regression in use. One could also search for signals occurring in pairs (to establish correlations), for parallel appearing changes (to establish interdependencies) or other forms of patterns and connections.

All such analyses, which ultimately might produce a meaningful result, are more directly related to human analytical competence than the popular dis-

course on artificial intelligence makes us believe. The result of such analyses becomes truly exciting for the analytical teams comprised of both humans and algorithms when a variety of data sets are superimposed on each other and the identified patterns can be compared with other data sets. In this way, the human-machine teams can identify correlations to specific events, weather developments, demographic changes, time of day and much more which by no means could be discovered by human beings alone. Hence, we can understand machine learning as experimental and explorative, but not truly »autonomous« from human influence and decisions.

This even goes for the algorithms applied in so-called »autonomous driving,« since the algorithmic decision-making is strictly determined by the data collected through the sensors of the self-driving car. If you change the sensors, you change the decision. If you change the training to a more aggressive driving style, you will get a more aggressive self-driving car. The car has no conscious reflection and decision-making about what to optimize its driving towards: Speed at all cost? Avoidance of injury? All those guidelines are human guidelines, external to the algorithm and trained or programmed into it. The »autonomous driving« algorithms hence are more dependent on external guidance than their names imply. If a self-driving vehicle identifies a human being in front of them, the algorithm has been trained to hit the brakes. Truly autonomous decision-making, as ascribed to humans, would imply that running the human over is a possible option in this case. The data sets that have trained the algorithms and the humans training the systems, however, never allowed for that possibility. Therefore, the algorithm is not *autonomous* in the meaningful sense of *self-legislation*.

II Transparency

The term artificial intelligence is usually used in public discourse as a collective term for all those procedures that result in a computer system performing tasks considered to be intelligent in humans. It is imprecise, but expresses a new quality of complexity in computer processes. Simple »if X, then Y« formulas develop into more complex instructions for machine learning: »If X results in result A, then assume A for the next experiment Y. But if Y results in result B, then correct A into B for case Y. And replicate this procedure n-fold to calculate probabilities for each further result by aggregating individual results and discovering patterns through similarity analysis.« Some claim that due to its so-called »autonomous« and evolutionary nature, such a computational feat should not be called an algorithm anymore, because it is unlike the formula-based algorithm of linear data-processing. But since it is still a quan-

titative process based on calculation with human involvement in the data set, computational framework and learning instructions, it is more similar to traditional algorithms than the evangelists of AI mythology would have it.

The word algorithm has its roots in the Latin word *algorismus* and used to mean the Indian art of calculation in reference to the Greek word *arithmós*, meaning »number.« The word was created from the name of the Persian-Arabic 9th century mathematician Al-Hwarizmī and is defined by the standard German dictionary as a »procedure for step-by-step transformation of number sequences« and »a process of calculation according to a certain [repetitive] scheme.«[13] Similarly, the Oxford dictionary defines an »algorithm« as »a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.«[14] Since, in the case of machine learning and even deep learning we are talking about a process of calculating probabilities (even in the case of deep learning the computational imitation of neurons in the brain is far less complex than its organic original and has well-established, stable math at its procedural core), we are talking about algorithms, which come into the world only because of human creativity. But their evolutionary nature allows these algorithms their own learning biographies as they were known only from humans and other animals.

It is difficult for the most complex of these experimental algorithms to give a meaningful account of their decision criteria, especially in the hidden layers in deep learning's neural networks. Not unlike cases involving human action, complex investigations into the decision criteria are necessary when something goes wrong, and only the most specialized machine learning experts can estimate where the root problems is. Especially when the root cause is in biased data or mistaken annotations of the training data, it takes time, focus and effort to find the source of bugs. Users of an AI system in a consumer product usually cannot identify any such bias or mistakes in the system in their own user interface. Similar to the pre-Reformation priesthood of the Church, machine learning experts are granted far-reaching competency to decide what counts as responsible development. But if it is true that such technologies will permeate every aspect of our lives in the not-too distant future, such authority must meet the highest of standards of transparency and accountability.

One of the key problems in terms of transparency and accountability in AI algorithms, has been the *black box* problem. Arthur Clarke has famously offered a poignant rule of thumb, commonly known as Clarke's third law: Any

13 Duden 2021. Author's translation. The definition in its original German wording: »Verfahren zur schrittweisen Umformung von Zeichenreihen; Rechenvorgang nach einem bestimmten [sich wiederholenden] Schema.«

14 Oxford University Press 2021.

sufficiently advanced technology is indistinguishable from magic.[15] The black box problem hits deep learning algorithms based on neural networks the hardest, because the neural network's hidden layers are precisely that: hidden. Analyst Alok Aggarwal notes that »even researchers are currently unable to develop a theoretical framework for understanding how or why they give the answers they do.«[16]

As an example, Aggarwal offers the Deep Patient experiment run by Joel Dudley and several colleagues. Deep Patient's objective was to use deep learning technologies »to predict health status, as well as to help prevent disease or disability« and »provide a machine learning framework for augmenting clinical decision systems.«[17] The project was successful and achieved improved predictions »for severe diabetes, schizophrenia, and various cancers« by using aggregated electronic health records of around 700,000 patients from their hospital's data warehouse.[18] Will Knight has reported that the Deep Patient algorithm anticipates »psychiatric disorders like schizophrenia surprisingly well.«[19] But given how difficult the prediction of schizophrenia is, the algorithm's co-inventor Joel Dudley »wondered how this was possible.«[20] But even Dudley himself has no way to find out because the algorithm »offers no clue as to how it does this.« He acknowledges that his team »can build these models … but we don't know how they work.«[21] Will Knight suggests that in order for such an algorithm to reliably help doctors, it will have to »give them the rationale for its prediction, to reassure them that it is accurate and to justify, say, a change in the drugs someone is being prescribed.«[22]

Among the open questions for algorithmic accountability studies is how to reconcile the public value of *transparency* with the public interest in *privacy*. What kind of transparency is possible if personal data must stay protected and secured from the very public eye that attempts to deliver transparency? In sensitive areas like medical application, who receives explanations from the »Explainable AI« is key. Under the traditional data privacy framework, only the patient and their medical team should have access to the data employed in the computational process. This has been eroded by the complexity required to process and store the vast troves of electronic data for medical purposes, which

15 For Clarke's third law's context, please see Clarke 1973.

16 Aggarwal 2018.

17 Miotto et al. 2016: 1.

18 Miotto et al. 2016: 1.

19 Knight 2017.

20 Knight 2017.

21 Knight 2017.

22 Knight 2017.

has led most doctors to outsource their data processing. This involves a third party in the process, adding complications to the task of transparent attribution of influences and factors in the process.[23] The complexity of AI systems further adds another layer of complexity in this attribution.

III Mythology

Developing AI systems has been a key goal in computer science and statistics for decades, not least due to the lucrative applications in medicine, biotechnology, industrial design, logistics management, quality control and many other commercial fields. Due to cost-effective availability of huge quantities of computing power, as well as the growing availability of training data (though this is still a massive hurdle for many AI projects), the goal is slowly becoming more and more realistic. Nevertheless, AI development remains difficult and error-prone even in the most successful systems and requires rare and costly talent that only the most attractive employers have available.

This is just one of the reasons why it remains doubtful whether a general AI can ever reach the much-discussed stage of *singularity*. The claim that a reliably flawless metasystem (termed general AI or strong AI) can result from the sophisticated interconnection of domain-specific subsystems (called narrow AI or weak AI) created by error-prone humans with imperfect data is logically impossible without some type of mythical leap.[24] For leapfrog innovation towards singularity to happen, some other foundationally new approach to error elimination must be found. The neuroscientist Wolf Singer has shown the flaws in the claim that AI systems can actually replicate the human brain's neural networks. Singularity theorist Kurzweil, Singer charges, »has fallen prey to a huge misunderstanding if he believes that an increase in computing speed alone will lead to a qualitative leap. The analogy of computer and brain is superficial at best. While both systems can execute logical computations, the systems architectures are radically divergent.«[25] While the human brain is both complex/non-linear and stable in its neural processing, all computing systems

23 The complexity in this process lead to blind spots in the processing chain, often leaving sensitive medical data exposed. Cf. Dangelmayer et al. 2019 and Gillum et al. 2019.

24 In Schmidhuber's contributions the leap takes story form and is presented as a logical progression: »As I grew up I kept asking myself, ›What's the maximum impact I could have?‹ … And it became clear to me that it's to build something smarter than myself, which will build something even smarter, et cetera, et cetera, and eventually colonize and transform the universe, and make it intelligent.« Cf. Markoff 2016.

25 Singer 2003: 33. The German original: »Ich denke, dass Kurzweil einem riesigen Missverständnis aufsitzt, wenn er glaubt, dass Vermehrung der Rechengeschwindigkeit allein zu einem qualitativen Umschlag führen würde. Die Analogie zwischen Computer und Gehirn ist

are either complex/non-linear and unstable or simple and stable.[26] The great riddle, Singer concludes, is how the non-linear complex processing of the brain »retains its stability and … integrates the various partial functions.«[27]

Nevertheless, some (self-proclaimed) pioneers of AI technology such as Schmidhuber[28] remain vehemently self-confident advocates of strong AI, which he propagates with mythological language as a godlike hyperintelligence and expects to emerge in the medium-term through continuous technological advancement. Schmidhuber's storytelling exploits the widespread lack of technical understanding and has significant power to frame how AI technologies are viewed. But since the average person has no way to verify or falsify grand claims, the discourse on the future of artificial intelligence has become a question of trust.

Because »everything we know about the world in which we live, we know through the media«[29], this question of trust is a question of *media* trust: Are journalists independently and critically verifying the grand claims of computer scientists and marketing directors? Are they even technologically capable of critical judgment on such specialized issues? In all critical probing of this kind, we must take note of what Luhmann adds after his famous dictum about the mediated nature of social reality: »we also know enough about the mass media to not be able to trust these sources. We resist it suspecting manipulation, but without consequence, since the knowledge we derive from the mass media, as if by itself, completes itself into a self-reinforcing framework.«[30]

---

bestenfalls eine oberflächliche. Beide Systeme können zwar logische Operationen ausführen, aber die Systemarchitekturen sind radikal verschieden.«

26 Singer's definition of »simple« includes machine learning.

27 Singer 2003: 37. The German: »Das große Rätsel ist, was die Großhirnrinde im Einzelnen macht, wie sie es macht, wie sie sich stabil hält und wie die vielen Teilfunktionen, die in ihren verschiedenen Arealen erbracht werden letztlich gebunden werden.«

28 Schmidhuber's student Hochreiter developed the Long Short Term Memory method that is used today in billions of smartphones for speech recognition, handwriting recognition, image analysis and other applications. Schmidhuber is cited as an author on the paper. Other AI researchers have questioned his credit. LeCun for instance, is not impressed: »Jürgen is manically obsessed with recognition and keeps claiming credit he doesn't deserve for many, many things … It causes him to systematically stand up at the end of every talk and claim credit for what was just presented, generally not in a justified manner.« Cf. Markoff 2016. Schmidhuber's research partners defend his claims for credit.

29 Luhmann 1996: 9. Given in the author's translation. The German original reads: »Was wir über unsere Gesellschaft, ja über die Welt, in der wir leben, wissen, wissen wir durch die Medien.«

30 Luhmann 2009: 9.

IV Contextuality

Depending on the culture in which a critical discussion of technical processes takes place, the popular assumptions about this process transported in media and everyday practice vary:

In the German-speaking world, AI is often portrayed as an *enemy* that destroys safe and secure working conditions and symbolizes impersonal coldness.[31] In the Anglo-Saxon world, AI is staged more as a *servant* or even a slave, which is also reflected in the usability dogma in the marketing of products developed by U.S. tech companies. In Chinese culture, AI is more often seen as a *partner* and colleague, which is reflected, for example, in the initiative of the Chinese news agency Xinhua to »hire« an AI-based avatar as news anchor.[32] Japanese reports repeatedly show that AI is viewed more as a *friend* there, as can be seen, for example, from the use of robots in assisted living for seniors.[33]

Even if we cannot provide methodologically rigorous evidence for these cultural differences and their socio-technical consequences here, such heuristic indications put the topic on the radar and stimulate interdisciplinary research. Reliable comparative ethnographic studies would turn this discourse into a highly productive interdisciplinary field of learning for the ethical evaluation and socio-psychological analysis of the hopes, fears and uses of AI applications.

D Ethics of Complex Data Processing

Once we clarified our definition of data and reconstructed four prevalent types of data processing in our typology, we were prepared to explore four areas of contention for critical data theory. Now we can venture into the ethics of complex data processing by exploring the conditions for the possibility of its practice. We can identify six hallmarks of ethical evaluation: I) technological *capability* of the responsible processor, II) general *availability* of data, III) *equitability* of the training data, IV) *computability* of the intended function, V) applied *methodology* and its corresponding distortions, and VI) *directionality* for the use and optimization of algorithms.

31 E.g. the dramatic headline »Die Jobfresser kommen.« Cf. Schultz 2016.

32 Cf. Kuo 2018.

33 Cf. 3sat 2018.

I Capability: Who can develop it?

As digital divide research has shown: access to digital opportunities is unequally distributed. This applies in amplified ways to artificial intelligence opportunities. Whoever can access and use large data sets, is in a good position to train machine learning algorithms, while those with little data are in a weak to impossible situation. Public institutions with strict data privacy regulation, for instance, are forced into competitive disadvantage to more liberally regulated private actors that can collect large quantities of data as long as the user has given consent.[34] Hence, an ethical evaluation of AI must include power analysis: Who is in a position to develop artificial intelligence systems in the first place?

There is, of course, an indirect limit to this type of power, because even those who can train AI systems in one domain will not necessarily be able to train systems in other domains. Even massive corporate conglomerates like Facebook with vast amounts of user data in many domains struggle to develop AI systems that can effectively take down live-streamed shooting videos from their platform before they are distributed to millions of users in real-time.[35] Improving such preventative AI systems requires domain-specific data of what such first-person shooting videos look like, which few companies have available in sufficient quantities to train a machine learning model. Facebook, for instance, has resorted to »working with American and British law enforcement authorities to obtain camera footage from their firearms training programs to help its A.I. learn what real, first-person violent events look like.«[36] Ethically, the question arises what kind of publicly funded data should be provided to privately held digital platforms for such preventative law enforcement purposes.

II Availability: What data is used?

After the stage of *power* analysis might come a stage of *data* analysis, because the type, source and structure of training data matters greatly for the ethical evaluation of a given machine learning solution. What data is available for

34 This consent remains precarious if most users unconsciously tick a box without reading privacy terms.

35 In March 2019, for instance, Facebook was used by the mass shooter in Christchurch, New Zealand to spread live video of 51 killings. And in August 2019, Facebook's platform was used in El Paso, USA to distribute the shooting plans posted on 8chan through Facebook and other social media sites.

36 Alba et al. 2019.

training could, for instance in medical research, inadvertently decide about who gets to live and who has to die. Taking genomics as a concrete example: If only European genomes are sequenced because of resources available there and few African genomes get sequenced, and groundbreaking research is thus based on European genomics, the developed treatment strategies might not work when applied in African contexts. The type, source and structure of the data, therefore, might serve to perpetuate existing power and illegitimate privilege.

III Equitability: Is there structural bias?

If an algorithm is trained with a set of data, this data will impact the results of this algorithm the application it powers. This has led to instances where racial prejudice or other forms of discriminatory patters in the training data have caused the algorithm to reproduce such prejudice in its results. A famous example is Microsoft's conversational bot Tay trained on tweets which turned it into a racist in less than a day.[37] Another example is an HR tool developed by Amazon that was intended »to review job applicants' resumes with the aim of mechanizing the search for top talent.«[38]

The problem was that »Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period« which meant that most resumes were from male candidates because of the massive gender gap in the tech industry. The system had »taught itself that male candidates were preferable«[39] because of the data underlying it. The system »penalized resumes that included the word ›women's,‹ as in ›women's chess club captain.‹ And it downgraded graduates of two all-women's colleges.«[40] At first, Amazon attempted to make the system more neutral to these specific words, but since the data set was biased, the algorithm was necessarily biased and even if singular instances could be corrected, the overall patterns could not. Amazon had no choice but to pull the plug on the project.

There is an increasingly established discourse on algorithmic bias and there are numerous attempts to develop best practices against such bias.[41] And while algorithmic bias is not easy to solve, it is still one of the easiest AI ethical problems to solve, because it is (a) evident in most cases, (b) quantifiable in

37 Vincent 2016.

38 Dastin 2018.

39 Dastin 2018.

40 Dastin 2018.

41 Cf. Lee et al. 2019.

many cases, and therefore (c) addressable through technological refinement. While it remains a challenge to test and evaluate training data ethically, the more fundamental ethical question is the limit of the quantitative paradigm.

IV Computability: What limits are inherent?

The basic fact about any computational technology is that it is just that: computational, and therefore quantitative. This is the foundationally inherent limit in any artificial intelligence system. Unless AI research comes up with a fundamentally different approach to intelligence – one that is not solely mathematical – there will be severe limits to the types of cognitive tasks artificial intelligence systems can take on. And there will even be mathematical limits, as Gödel has proven with his theorem of incompleteness,[42] which further casts doubt on euphoric anticipation of general AI. If the quantitative mathematical paradigm remains the only relevant paradigm in AI development, essential dimensions of human experience will never be captured as part of artificial intelligence systems due to the methodology's inherent limits.

Love, for instance, is one of the key human experiences and experienced as an integrated emotion with strong cognitive elements. Yet, it is impossible to quantify, which is why quantifiable proxies need to be found to even begin to approach the phenomenon. We might visualize which parts of the brain are active when we experience an intense moment of love. But love in a deeper, more philosophical sense manifests itself in so many ways and nuances that it becomes too complex to reduce to mathematical, quantified calculations. And even if we identified the most important part of the brain for the act and experience of loving, we would still not have proven anything. For like force in physics, love is impossible to prove. We can only deduce its existence from the effects we can observe. For a low standard of proof, this might suffice. But any sophisticated concept of love will include non-quantifiable dimensions and limit the functionality of AI applications in this area.

The limitations of the current methodologies do not just have philosophical implications for those who worry about the limits of a quantitative paradigm. They have implications for those who invested their capital in the commercialization of artificial intelligence systems. Analyst Alok Aggarwal notes that »several of the obstacles that led to the demise of the first AI boom phase over forty years ago remain unresolved today, and it seems that serious theoretical advances will be required to overcome them.«[43] Aggarwal concludes that »the

42 Cf. Rautenberg 2008.

43 Aggarwal 2018.

predictions … are unlikely to be met in the next fifteen years, and financiers may not receive an expected return on their recent investments in AI.«[44]

V Methodology: What might skew results?

Beyond the inadvertent impacts of the quantitative paradigm and the theoretical limitations of the current concepts in artificial intelligence research, there are other ways the process might skew the results. An example for this is what has at times been called *coding populism*. The concept means that applying machine learning to content distribution (as Facebook has increasingly done with its News Feed) will necessarily advantage populist contributions on the platform. If the distribution system is based on billions of trial and error experiments (for instance, with contextual bandit testing) designed to find the posts that trigger the highest engagement and thereby increase the time spent on the platform which translates directly into higher advertisement revenue, then nuanced contributions will be drowned out on the platform and attention-grabbing, incendiary, controversial, aggravating content will necessarily be the most-distributed content on the platform.

Applying machine learning to content distribution instead of operating on editorial philosophy and principles of human curation, might actually end up feeding our subconscious *worst* angels instead of Lincoln's proverbial »better angels of our nature.« Such machine learning applications exploit subconscious behavioral patterns, including the bias and prejudice that we try to fight *consciously* but still often act out *subconsciously*. This does not mean we endorse our subconscious hopes, fears and prejudice consciously. It means that we are imperfect human beings who might not want to engage in discriminatory behavior, but unknowingly contribute to racist, sexist, or violent structures built into an engagement-only based algorithm. Because machine learning is based on *preformed* and not *intended* behavior (as an editorial policy would expound), it does not ask users to support audacious ideals, but rather feeds into their subconscious failings. It leaves users feeling manipulated, as is observable in the low trust in social media platforms. One example: Only 14 percent of the German population trust social media networks generally, with Twitter specifically at only 10 percent.[45]

44 Aggarwal 2018.

45 Institut für Demoskopie Allensbach 2016.

VI Directionality: What is the purpose?

Despite its outspoken commitment to the high-flying ideal of »making the world more open and connected,«[46] Facebook's ultimate purpose of applying machine learning to the content distribution has been to increase advertising revenue.[47] And it has been successful with it – the revenue can now sustain operations at scale and leaves significant capital for innovation projects, acquisitions and other investments in Facebook's overall technological capabilities. While any business analyst would hail this move as responsible business practice with an impressive record of success, those who do not earn from this success as shareholders and worry about the societal impacts will question the integrity of machine learning application for this purpose. And even those who have earned a fortune from Facebook's ad technology have started to question the impact of the News Feed algorithm. Several former employees have expressed concern about »the unintended consequences of a network when it grows to a billion or 2 billion people« and »exploiting a vulnerability in human psychology« through »a social-validation feedback loop.«[48] Others have aired »tremendous guilt« for helping to create »tools that are ripping apart the social fabric«[49] – for instance Facebook's micro-targeting technology[50] and the like button.[51] Mark Zuckerberg's co-founder has even called for a breakup of the company.[52]

This vigorous debate, however, has not just been stimulated by former Facebook employees, but a great number of other individuals and organizations. One of those individuals is Tristan Harris, a former Google engineer, who has critiqued Silicon Valley's dopamine-driven product strategies as the »attention economy« and a »race to the bottom of the brain stem.«[53] Harris started the Time Well Spent movement[54] and went on to start the Center for Humane Technology. Through the work of this center, Harris wants to fight

46 Hoffmann et al. 2016: 1.

47 Given its massive user growth, venture capital alone was not enough. Facebook needed a solid revenue stream and perfected its role as ad-broker and micro-targeting to reach highly focused user groups.

48 Allen 2017.

49 Vincent 2017.

50 Cf. Garcia-Martinez 2017.

51 Cf. Morgans 2017.

52 Hughes 2019.

53 Thompson 2019.

54 The *Time Well Spent* movement has been taken up by Apple, Instagram, Facebook and Harris's former employer Google through the implementation of time-monitoring apps. Some have seen this as a move to coopt the movement, others as a legitimate response to it. Cf. Stolzoff 2018.

the »downgrading« of humanity trough technology. Tech companies have furthered what he sees as a race to the bottom »by promoting shortened attention spans, outrage-fueled dialogue, smartphone addiction, vanity, and a polarized electorate. Harris called for tech companies to enable a new ›race to the top,‹ centered on building tools to help people focus, find common ground, promote healthy childhoods, and bolster our democracy.«[55]

The intense debate on the purpose and impact of tech giants like Facebook and Google has strengthened the wider ecosystem of debate around the purpose and impact of technology more generally. In Germany, for instance, the Conscious Coders student group works towards »beneficial and sustainable use of digital technologies for the society and the environment« as well as »a profound understanding of emerging technologies throughout the whole society« and calls for »critical developers who review their work against ethical questions and are aware of their responsibility.«[56]

This mission-driven approach as garnered momentum in the field of artificial intelligence as well. The AI for Good Foundation, for instance, wants to build »lasting communities that bring the best technologies to bear on the world's most important challenges … by coordinating the AI research community, technologists, data, and infrastructure with the stakeholders on the ground, policy makers, and the broader public« in support of the United Nation's Sustainable Development Goals.[57] Another organization from within the U.S. tech community working »to ensure that artificial general intelligence … benefits all of humanity« is OpenAI.[58]

VII Summary: Ethical Use of Complex Data Processing

By asking ourselves the right questions at the right time in the process, ethical review can become an integral part to technology operations. Our probing of the conditions for the possibility of complex data processing has yielded a non-exhaustive list that can power such an operationalization of ethical review. Underlying this exploration is the assumption that all technology is a cultural product and requires a whole range of different constructively linked factors for its successful implementation. Before the popular questions about singularity become meaningful, a whole host of things can go wrong in day-today AI systems that are already in widespread use.

---

55 Newton 2019.

56 Conscious Coders 2019.

57 AI for Good Foundation 2019.

58 OpenAI 2018.

Sometimes, it seems, the storytellers of AI mythology deploy the smokescreen of singularity, to hide more day-to-day AI applications in plain sight. Instead of narrowing the ethical scope to the grand questions of imagined futures, researchers should expand the ethical analysis of existing technologies in complex data processing. Our non-exhaustive list of review questions might serve as a first step in that endeavor: Who can deploy complex data processing in the first place? What (kind of) data is available for training of intelligent algorithms? Does the training data show signs of prejudice or structural bias? What are the inherent limitations of data processing? How might the specific process skew the results? For what goal is a given technology deployed and optimized?

References

Aggarwal, Alok 2018: The Current Hype Cycle in Artificial Intelligence. https://scryanalytics.ai/the-current-hype-cycle-in-artificial-intelligence/ (accessed 9 October 2019).

AI for Good Foundation 2019: About Us. https://ai4good.org/about/ (accessed 11 October 2019).

Alba, Davey/Edmondson, Catie/Isaac, Mike 2019: Facebook Expands Definition of Terrorist Organizations to Limit Extremism. In: The New York Times, 17 September 2019. https://www.nytimes.com/2019/09/17/technology/facebook-hate-speech-extremism.html (accessed 17 October 2019).

Algorithmus, der. In: Duden 2021. https://www.duden.de/rechtschreibung/Algorithmus (accessed on 30 November 2021).

Algorithm. In: Oxford University Press 2021. https://www.lexico.com/definition/algorithm (accessed on 30 November 2021).

Allen, Mike 2017: Sean Parker unloads on Facebook: »God only knows what it's doing to our children's brains«. In: Axios, 9 November 2017. https://www.axios.com/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792-f855e7b4-4e99-4d60-8d51-2775559c2671.html (accessed 11 October 2019).

Bedford-Strohm, Jonas 2019: Mythologie, Typologie, Pathologie: Bausteine einer kritischen Theorie der Datenverarbeitung in den Verfahren der künstlichen Intelligenz. In: Görder, Björn/Zeyher-Quattlender, Julian (Hgg.): Daten als Rohstoff: Die Nutzung von Daten in Wirtschaft, Diakonie und Kirche aus ethischer Perspektive, Münster, LIT.

Clarke, Arthur C. 1973: Profiles of the Future: An inquiry into the limits of the possible. London, Macmillan.

Conscious Coders 2019: Vision. https://www.consciouscoders.io/ (accessed 11 October 2019).

Dangelmayer, Pia/Meyer-Fünffinger, Arne/Hagmann, Ulrich/Köppen, Uli/Kühne, Steffen/Nierle, Verena/Schnuck, Oliver/Streule, Josef/Tanriverdi, Hakan/Thamerus, Tatjana/Zierer, Maximilian: Millionenfach Patientendaten ungeschützt im Netz. In: BR24, 17 September 2019. https://www.br.de/nachrichten/deutschland-welt/millionenfach-patientendaten-ungeschuetzt-im-netz,RcF09BW (accessed on 20 July 2021).

Dastin, Jeffrey 2018: Amazon scraps secret AI recruiting tool that showed bias against women. In: Reuters, 10 October 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (accessed 10 October 2019).

Garcia-Martinez, Antonio 2017: I'm an ex-Facebook exec: don't believe what they tell you about ads. In: The Guardian, 2 May 2017. https://www.theguardian.com/technology/2017/may/02/facebook-executive-advertising-data-comment (accessed 11 October 2019).

Gillum, Jack/Kao, Jeff/Larson, Jeff 2019: Millions of Americans' Medical Images and Data Are Available on the Internet. Anyone Can Take a Peak. In: ProPublica, 17 September 2019. https://www.propublica.org/article/millions-of-americans-medical-images-and-data-are-available-on-the-internet (accessed on 20 July 2021).

Hoffmann, Anna Lauren/Proferes, Nicholas/Zimmer, Michael 2018: »Making the world more open and connected«: Mark Zuckerberg and the discursive construction of Facebook and its users. In: New Media & Society, 20 (1), 199–218.

Hughes, Chris 2019: It's Time to Break Up Facebook. In: The New York Times, 9 May 2019. https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html (accessed 11 October 2019).

Institut für Demoskopie Allensbach 2016: Welche dieser Informationsquellen halten Sie für vertrauenswürdig, wo kann man besonders zuverlässige Informationen über Politik und politische Ereignisse erwarten? In: Allensbacher Archiv, IfD-Umfrage 11062.

Knight, Will 2017: The Dark Secret at the Heart of AI. In: MIT Technology Review, 11 April 2017. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (accessed 14 October 2019).

Kuo, Lily 2018: World's first AI news anchor unveiled in China. In: The Guardian, 9 November 2018. https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china (accessed 13 November 2018).

Lee, Nicol Turner/Resnick, Paul/Barton, Genie 2019: Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. In: Brookings, 22 May 2019. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ (accessed 10 October 2019).

Luhmann, Niklas 1996: Die Realität der Massenmedien. Opladen, Westdeutscher Verlag.

Heidelberger Institut für theoretische Studien 2017: The Dark Side of Natural Language Processing. https://www.h-its.org/scientific-news/ethics-nlp/ (accessed on 15 October 2018).

Markoff, John 2016: When A.I. Matures, It May Call Jürgen Schmidhuber ›Dad‹. In: The New York Times, 27 November 2016. https://www.nytimes.com/2016/11/27/technology/artificial-intelligence-pioneer-jurgen-schmidhuber-overlooked.html (accessed 21 July 2021).

Miotto, Riccardo/Li, Li/Kidd, Brian A./Dudley, Joel T. 2016: Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. In: Scientific Reports, 6 (26094). https://www.nature.com/articles/srep26094 (accessed 14 October 2019).

Morgans, Julian 2017: The Inventor of the ›Like‹ Button Wants You to Stop Worrying About Likes. In: Vice, 6 July 2017. https://www.vice.com/en_uk/article/mbag3a/the-inventor-of-the-like-button-wants-you-to-stop-worrying-about-likes (accessed 11 October 2019).

Nassehi, Armin 2019: Muster: Theorie der digitalen Gesellschaft. München, Beck.

Newton, Casey 2019: The leader of the Time Well Spent movement has a new crusade. In: The Verge, 24 April 2019. https://www.theverge.com/interface/2019/4/24/18513450/tristan-harris-downgrading-center-humane-tech (accessed 11 October 2019).

OpenAI 2018: OpenAI Charter. https://openai.com/charter/ (accessed 11 October 2019).

Parikka, Jussi 2012: What is media archaeology? Cambridge, Polity.

Rautenberg, Wolfgang 2008: Unvollständigkeit und Unentscheidbarkeit. In: Einführung in die Mathematische Logik. Wiesbaden, Vieweg+Teubner, 167–208.

Schultz, Stefan 2016: Arbeitsmarkt der Zukunft. Die Jobfresser kommen. In: DER SPIEGEL, 2 August 2016. http://www.spiegel.de/wirtschaft/soziales/arbeitsmarkt-der-zukunft-die-jobfresser-kommen-a-1105032.html (accessed 14 November 2018).

Singer, Wolf 2003: Ein neues Menschenbild? Gespräche über Hirnforschung. Frankfurt/M., Suhrkamp.

Stolzoff, Simone 2018: Technology's »Time Well Spent« movement has lost its meaning. In: Quartz, 4 August 2018. https://qz.com/1347231/technologys-time-well-spent-movement-has-lost-its-meaning/ (accessed 11 October 2019).

Thompson, Nicholas 2019: Tristan Harris: Tech Is ›Downgrading Humans.‹ It's Time to Fight Back. In: WIRED, 23 March 2019. https://www.wired.com/story/tristan-harris-tech-is-downgrading-humans-time-to-fight-back/ (accessed 11 October 2019).

Vincent, James 2017: Former Facebook exec says social media is ripping apart society. In: The Verge, 11 December 2017. https://www.theverge.com/2017/12/11/16761016/former-facebook-exec-ripping-apart-society (accessed 11 October 2019).

Vincent, James 2016: Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. In: The Verge, 24 March 2016. https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist (accessed 10 October 2019).

Weizenbaum, Joseph 2001: Computermacht und Gesellschaft. Frankfurt/M., Suhrkamp.

3sat 2018: Mein elektrischer Freund. Für Japaner haben auch Roboter eine »Seele«. https://www.3sat.de/page/?source=/nano/technik/184917/index.html (accessed 13 November 2018).

ORCID

Jonas Bedford-Strohm https://orcid.org/0000-0003-4165-1881