Article

new media & society
1–17

DOI: 10.1177/1461444819885334
journals.sagepub.com/home/nms
SAGE

Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning

Rainer Mühlhoff
Technical University of Berlin, Germany

Abstract

Today, artificial intelligence (AI), especially machine learning, is structurally dependent on human participation. Technologies such as deep learning (DL) leverage networked media infrastructures and human-machine interaction designs to harness users to provide training and verification data. The emergence of DL is therefore based on a fundamental socio-technological transformation of the relationship between humans and machines. Rather than simulating human intelligence, DL-based AIs capture human cognitive abilities, so they are hybrid human-machine apparatuses. From a perspective of media philosophy and social-theoretical critique, I differentiate five types of "media technologies of capture" in AI apparatuses and analyze them as forms of power relations between humans and machines. Finally, I argue that the current hype about AI implies a relational and distributed understanding of (human/artificial) intelligence, which I categorize under the term "cybernetic AI." This form of AI manifests in socio-technological apparatuses that involve new modes of subjectivation, social control, and digital labor.

Keywords

Artificial intelligence, audience labor, commercial content moderation, cybernetics, deep learning, human computation, human-computer interaction, social media, tracking, training data, user experience design

Corresponding author:
Rainer Mühlhoff, Excellence Cluster *Science of Intelligence*, Technical University of Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany.
Emails: mail@rmuehlhoff.de; muehlhoff@tu-berlin.de

Introduction: a new era of AI?

In recent years, there has been a renewed hype about artificial intelligence (AI). AI technology is attracting immense public attention as more and more real and tangible applications are emerging in industry, consumer worlds, politics, and policy. At the technological level, this trend is largely due to deep learning (DL) as *one* particular approach within the heterogeneous field of AI research. The DL is a method based on simulated artificial neural networks (ANNs) in the field of machine learning (ML) (Bengio, 2009; Goodfellow et al., 2016; LeCun et al., 2015). Various hitherto difficult computational problems such as object recognition in images, natural language processing, and identification of patterns in large data sets can now be automated with DL.

While the breakthrough of DL is often seen as a "revolution," the debate in media studies shows that this is only a momentary—and above all economic—supremacy of one of several AI paradigms that have long been running parallel (Sudmann, 2018). DL is a "bottom-up" statistical approach based on the aggregation of empirical knowledge. Since Alan Turing, learning-based AI has been contrasted with the paradigm of symbolic AI, or "Good Old-Fashioned AI" (GOFAI) (Haugeland, 1985), which essentially understands intelligence as the ability to manipulate symbols. GOFAI is modeled around problems such as automated chess play or mathematical theorem proving (Haugeland, 1981; Newell and Simon, 1976; see Brooks, 1991 for a historical overview). The current dominance of the ML paradigm over GOFAI is often explained by strong developments in computing technology toward high-performance parallel computing on graphical processing units (GPUs) during the last 10 years. That is, the current progress of DL is attributed to a new generation of hardware architectures that is better suited for the computational tasks related to ANNs that require processors different from the classical von Neumann architectures (Bolz et al., 1994; Sudmann, 2018).

In this article, I would like to add another approach to explaining the success story of DL: the diagnosis of an underlying *socio-technological* revolution. I will argue that DL's "breakthrough" required not only the development of high-performance parallel computing techniques, but a fundamental structural change in media culture and human-computer interaction (HCI) at societal scale. I start from the observation that most industrial DL implementations come with extensive media technological infrastructure for *capturing humans in distributed, human-machine computing networks, which as a whole perform the intelligence capacity that is commonly attributed to the computer system as "artificial intelligence."* Today, the scarce resource on which the success of a DL project depends is neither algorithms nor computing power but rather the availability of training and verification data, which is ultimately obtained through human participation. The importance of this resource led to the emergence of new forms of exploitation and implicit labor in the digital that build on existing socio-economic divides. Seen from the angle of this article, DL is a form of distributed orchestration of human cognition through networked media technology. The question of generating training data is so essential to DL projects that at the core of any such project today lies a characteristic problem of *human-computer interaction* (cf. Mühlhoff, 2019b): How does one design an interface, a platform, or a medial environment that can serve as an infrastructure for obtaining data through free and implitic human participation?

Historical context

For many decades in the 20th century, the symbolic paradigm of AI (GOFAI) was deemed more fruitful and received more research resources than ML approaches. This affected not only AI research but also the conception of human intelligence itself which was articulated in related fields such as cognitive science and psychology. The concept of intelligence was at any time closely tied to current techniques of computation (Brooks, 1991). The concept of the universal Turing machine (cf. Turing, 1937) and its realization in von Neumann processor architectures was not only better suited to the symbolic paradigm than to ML but also influenced the general understanding of "intelligence" and "cognition" of the time to focus on symbol manipulation and problem-solving. Despite the fact that alternative paradigms both in AI and cognitive science, such as embodiment and situatedness (Brooks, 1991), or distributed (Hutchins, 2001; Rumelhart and McClelland, 1986) and connectionist (cf. Sun, 2014) approaches, have always been pursued, it was not until the 2010s that ML based on ANNs made significant developments that eventually lead to the current dominance of the learning paradigm over GOFAI. It is common to explain this development by the discovery of the backpropagation training algorithm (Rumelhart et al., 1986), which became effective only much later by the development of high-performance parallel computing on GPUs.

Hence, the current boom of DL is largely seen as the product of a "hardware revolution"—a claim that is also maintained in media studies (e.g. Bolz et al., 1994; Sudmann, 2018). What is underrepresented in this description, however, is the fundamental shift of the relation between humans and machines that materializes in everyday human-machine interaction designs (Mühlhoff, 2018) in the wake of "web 2.0" (O'Reilly, 2005) and "ubiquitous computing" (Weiser, 1991). As I maintain in this article, the media cultural transformations of modern user experience (UX) design are not only a prerequisite for the success of DL but also instigated a shift of the conception of intelligence itself, which is densely related to the media-technological relation of humans and machines. In the DL paradigm, human cognitive skills are not simulated by a machine anymore, but embedded in machine networks. DL is less about replacing human cognitive labor by an intelligent machine but about embedding and harvesting human cognition in computing networks through new forms of labor and machinized power relations.

The perspective outlined in this article will stress this socio-technological dimension of DL. I will proceed in three steps: In section "Introduction: a new era of AI?" I will use two research contributions from 2006 and 2017 as examples to illustrate fundamental transformations in consumer media that are a prerequisite to the success of DL. In section "Hybrid processors: human-machine computing networks and AI," I will differentiate five forms of capturing human collaboration in hybrid human-machine AIs and point to the different forms of power, subjectivation and labor engendered by these modes of capture. In section "Conclusion: a cybernetic notion of AI," I will debate the shift of the understanding of intelligence that is implicit in DL, arguing that in order to accommodate recent developments, a "simulation-based" understanding must be differentiated from a "cybernetic understanding of AI."

Hybrid processors: human-machine computing networks and AI

The current, third era of AI technology is characterized by a new form of networked technology that implements intelligent devices by incorporating humans as cognitive agents. To make this historical thesis plausible, I will look at two exemplary research contributions from 2006 and 2017 that illustrate this development. Both are lectures of relevant scientists, which are available as videos.

Vignette 1: "games with a purpose"

In 2006, the computer scientist Luis von Ahn, a pioneer of "crowdsourcing" and founder of the company reCAPTCHA ([Onl.1]), gave a Google Tech Talk under the title of "Human Computation" ([Vid.1]). He says that his project started from the idea that the human brain is actually "a pretty advanced processing unit . . . that can solve problems that computers cannot yet solve" ([Vid.1]: 6 minutes 40 seconds), such as recognizing objects in images or understanding spoken language. To this, he adds the sociological observation that there is an immense number of "wasted human cycles"[1] every day in the world, evident, for instance, in "the 9 billion human-hours of Solitair [that were] played in 2003" ([Vid.1]: 7 minutes). Humans are not only good computing units, but their computing power is also available in abundance. From these two premises, von Ahn put together the goal of his research: "Running a computation in peoples' brains instead of silicon processors" ([Vid.1]: 25 minutes). To this end, "we are going to consider all of humanity as an extremely advanced, large-scale distributed processing unit that can solve large-scale problems that computers cannot yet solve." ([Vid.1]: 8 minutes; see also von Ahn, 2005)

One project of von Ahn and Laura Dabbish (2004) was the so-called "ESP game"—it was later acquired by Google and became known as Google Image Labeler. Its purpose was to obtain labels that describe images through the free participation of people on the Internet. The ESP game is a two-person online game in which play partners are randomly assigned to each other for the duration of a session and have no means of communication. In a game cycle, both players see the same image on their screens and are prompted to enter keywords describing the image. They cannot see what the other is typing, but if both enter the same keyword fast enough ("match"), they get points. In effect, these keywords can be used as accurate labels for the image.

The ESP game has gained significant popularity after its launch in 2003. Over 1.3 million labels for approx. 290,000 pictures were generated within four months (von Ahn and Dabbish, 2004). The database of Google image search at that time contained about 425 million images, and von Ahn and Dabbish (2004) estimated that their game could completely index this stock in only 6 months by the free work of the players ([Vid.2]: 15 minutes 20 seconds). The labels could then be used to improve Google's image search. Notably, this came at a time when leading image search technology relied on file names, HTML captions, and the surrounding text on the websites to associate images with search keywords.

Luis von Ahn (2006) proposed the game-theoretical term "Games With a Purpose" (GWAP) for games like this. He thus established what is commonly referred to as

"gamification" and "human computation" (von Ahn, 2005) within HCI research. Remarkably, Amazon's "Mechanical Turk" service was introduced at roughly the same time. While Mechanical Turk allows repetitive but simple tasks to be outsourced to paid clickworkers, von Ahn's vision was to turn an "extremely tedious task into a game that's fun" ([Vid.1]: 32 minutes 40 seconds). Following this principle, von Ahn et al. (2006) have developed several other online games that outsource computing problems to the free labor of humans, for example, "Peek-a-Boom" for the spacial location of objects in images or "verbosity" for the generation of a large knowledge base of common sense facts.

All these games are based on the idea of harnessing "human computing power" in hybrid human-machine networks to perform a computational task that a silicon-based computer cannot easily solve. The ultimate and best known application of this principle is "reCAPTCHA"—a company founded by Luis von Ahn and later acquired by Google ([Onl.1]). reCAPTCHA combines the idea of CAPTCHA (von Ahn et al., 2003) with that of "human computation." A CAPTCHA is a small challenge that can be built into the human-machine interaction here and there on the Internet to verify that the user is actually a "human user." For this purpose, the user is asked to solve a small task such as image recognition or text recognition, which is a low barrier for a human, but a high one for a computer bot. reCAPTCHA extends on this principle by re-using the responses of human users as training data for industrial Deep-Learning projects (von Ahn et al., 2008).

Vignette 2: the "eternal spring" of AI

A good 10 years after von Ahn's Google Tech Talk, we are in the midst of the industrial euphoria about learning based AI. In an exemplary form, this euphoria is visible in a talk given by Andrew Ng in 2017 at the Stanford Graduate School of Business ([Vid.2]). Andrew Ng is a leading AI expert, a Stanford professor, and former head of AI departments first at Google and then at Baidu. In his talk under the title "AI is the New Electricity," he explains that after the two "AI winters" in the late 1960s and 1980s, AI technology is now in a phase of "eternal spring" ([Vid.2]: 1 hour 0 minutes). Today, he says, AI has become a key technological component and transformative agent of our civilization, similar to the indispensable role of silicon-based semiconductors or electricity ([Vid.2]: 1 hour 0 minutes).

When Ng speaks of AI, he explicitly refers to the narrower category of DL in the variant of supervised learning, "because the massive economic value" of the industrial application of AI is currently (in the future this could change) almost exclusively driven by DL ([Vid.2]: 7 minutes 45 seconds). He also highlights that DL has become successful in the last 10 years because of two independent factors. (1) The development of high performance computing (HPC) on GPUs increased computing speed, and (2) DL requires an enormous amount of training data, but sufficient data sets have only become available in the last 10 years ([Vid.2], 21 minutes). This dependence on training data is because supervised learning trains an ANN using a large set of known input and output pairs until its internal parameters are calibrated so well that previously unseen input data are likely to be connected to the correct output. For example, if an ANN is to recognize objects in images, the input is an image and the output is a list of labels that designate the objects in the image. A training data set would then be a database of labeled images. According

to Ng, world-leading face recognition AIs are trained on more than 200 million facial images; speech recognition AIs are build from more than 100,000 hours of transcribed audio ([Vid.2]: 33 minutes).

Interestingly, from his business and application-oriented perspective Ng points out that today only the *second* factor, the availability of training data, is genuinely a scarce resource. This is to be seen in a context where computing power has been available as a service on an industrial scale for several years now. Services such as Google's "Cloud AI" or IBM's "Watson Machine Learning" allow any small company to bring their data and train complex DL models "in the cloud" without having to maintain their own computing infrastructure ([Onl.2]). Open source libraries such as TensorFlow ([Onl.3]) or Keras ([Onl.4]) make algorithms for DL accessible via high-level application programming interfaces (APIs), so industrial users often do not need to develop their own implementation of DL algorithms.

In a constellation where algorithms are public and computing power is for sale, the core economic asset of "each defensible AI business" is training data ([Vid.2]: 30 minutes ff.). This is a fact that determines business strategies. Ng says, "I frequently launch products where my motivation is not revenue, but is actually data; and we monetize the data through a different product" ([Vid.2]: 33 minutes 40 seconds). The AI product cycles are subject to a feedback loop that Ng calls the "virtuous circle of AI" ([Vid.2]: 35 minutes ff.): more users of an AI product typically generate more data using it; more data make the AI and thus the product better; a better product in turn attracts more users. Strategies for the introduction of new AI products on the market explicitly build on this principle. In fact, it is not unheard of that in the early stages, human clickworkers instead of intelligent computers sit "at the backend" of a new AI product. In this way, the virtuous circle of AI can still be activated, even if no training data are available yet ([Onl.5]). This trick allows to reverse the order of training and inference phases of an ANN.

Ng's talk also mentions some limitations of DL that are useful to inform the critical perspective I take in this article. First, Ng proposes a "rule of thumb" regarding which types of problems he thinks could be expected to be automated by DL. "Anything that a typical human can do in at most one second of thought, we can probably now or soon automate with AI," he says ([Vid.2]: 14 minutes). This statement includes image recognition and speech recognition tasks, but excludes, for example, the prediction of stock market prices ([Vid.2]: 16 minutes). Second, Ng mentions the learning curve of DL AIs, which is a graph that shows the performance as a function of the number of trained input and output pairs. This curve rises steeply at the beginning, that is, with an increasing amount of training data, DL makes strong progress in the accuracy of predictions; however, roughly at the point of "human-level performance," this curve typically flattens ([Vid.2]: 18 minutes). Therefore, when an accuracy roughly equal to that of human cognition is reached, additional training data have only minor effects and learning progress slows down, according to Ng. Both observations suggest that the potentials of DL are inherently tied to the cognitive skills of human beings.

Current commercial AIs do not replace human intelligence, they capture it

Remarkably, with his "rule of thumb," Ng restricts the range of problems that can be addressed by DL to exactly the same range that Luis von Ahn had envisaged 15 years

earlier with his idea of "exploiting human brain cycles." I argue that this correlation is no coincidence. In the past 10 years, ML performed well precisely on the kind of tasks for which there is now a comprehensive *media infrastructure* that involves human beings in hybrid human-machine computing networks to obtain training data. In a development that leads away from GOFAI, DL-based AI today is a product of harvesting human labor and cognition in computing networks at large scale. ML is more than algorithms and HPC: it is a media-cultural constellation involving human-machine interfaces and media technology that makes people implicitly generate data that can be used as training data. As we shall see in the next chapter, Luis von Ahn's GWAPs are only *one* of several contemporary forms such a *media technology of capture* might take.

This shows that the emergence of DL is inherently tied to recent trends in HCI and UX design (Mühlhoff, 2018). Any viable DL problem today is translated into a corresponding problem in HCI. This problem is: How can a use case and a UX world be constructed so that the data that is needed as training data can be obtained as behavioral data from the "free labor" of a general audience of users (cf. Terranova, 2000; Fisher and Fuchs, 2015; Fuchs, 2010)? The technology that solves this concurrent HCI problem must be seen as an integral part of the technical apparatus that implements the AI. This makes building an AI partly a problem of *social* engineering and interface design. The acquisition of training data goes hand in hand with the creation of digital media infrastructures that take the form of hybrid human-machine networks, which must themselves, as a whole, be described as an entity in which the AI in question is to be located.

From a broader historical point of view, the commercial breakthrough of AI is therefore closely related to key developments of the "ubiquitous computing" paradigm (Weiser, 1991) and chiefly facilitated by the rise of the interactive "Web 2.0" and social media. It was not until the end of 2006 that Facebook opened its service to the general public. The idea of "Web 2.0," which brought "design patterns and business models for the next generation of software" (O'Reilly, 2005), was popularized only in 2004. This creates an idea of how remote the concept of harnessing human cognitive resources in distributed computing networks must have appeared in 2003–2006 and earlier. Since then, however, various infrastructures for capturing human cognitive resources in networked platforms have *de facto* become a *media cultural standard* due to the penetration of the social world by networked computers and graphical user interfaces. Today, a general convergence of training data and everyday behavioral data can be observed. It has become relatively easy to collect training data if this data are a by-product of everyday usage flows.

Media technologies of capture: five types of power relations

To show how this socio-technological analysis of DL spells out in relation to real applications today, I will now distinguish five different forms of capturing human cognitive capacities in human-computer interfaces that feed into AI products. I will specifically point out how the five forms differ in terms of human-computer power relations, subjectivation of users, and new forms of labor in digital apparatuses.

The *first form* of capture has already been described above using the example of the ESP game: It can be summarized under the term "gamification." Gamification is a method to engaging users in a playful interactive world in which they knowingly or unknowingly perform tasks that originate from, and feed back into, a non-game context

(Deterding et al., 2011). In this case, the form of power that shapes the relation of users and computing machinery builds on playfulness and fun. Falling into the category of "gamification-from-above" (Woodcock and Johnson, 2018), these examples show a hierarchical extrication of "audience labor" (Fisher, 2015). By creating a subjective experience of pleasure and harmlessness, this fact (which is *known* to many users) does not dominate the user experience in a negative way.

A *second form* of harnessing human cognitive resources in computer networks can be described as "trapping and tracking." Its prototype is reCAPTCHA, fittingly described as "Human-Based Character Recognition via Web Security Measures" by its inventors (von Ahn et al., 2008). Through "trapping and tracking," a (computing) task that is to be outsourced to a human user is integrated into an interaction process so that it *must* be completed in order for the user to achieve something else they *want* to achieve. A more complex but less obvious example of this method of harnessing human cognition is provided by the Google search engine. A list of Google search results is not only the product of a calculation using AI, but it has embedded scripts that turn each user into a data provider for further calibration and re-training of this AI. This is facilitated by a click-tracking mechanism on the search engine result pages (SERPs) that records every click on that page and reports it back to a Google server (Mühlhoff, 2019a). This infrastructure allows Google to register, among other things, which search results users select and whether they return to the SERP after viewing one result (e.g. using the back button) to click another one. Thus, by simply using Google Search, users involuntarily generate a wealth of data providing information about the perceived relevance of the results and enable detailed analyses of clicking behavior (which website elements are more likely to be noticed; how far down users scroll; what bias exists between ads and organic search results, etc.). If users are logged in to a Google account, this data are linked to their personal user IDs and can be correlated with their e-mail contents, YouTube activities, calendar dates, and so on (Mühlhoff, 2019a). While I cannot go into the serious data protection issues arising from these tracking techniques (Noble, 2018; O'Neil, 2016), in the present context my point is that the real-time stream of usage data serves to continuously train and further calibrate the AI that is responsible for generating the search results.

This example shows that in practice there is often no strong separation between the training and inference phase of an ANN model. The collection of training data for continuous verification and recalibration of the Google search AI never stops. Through the participation of users in Google's search engine, a feedback loop is implemented, linking the predictions back to reality. This feedback loop is a fixed infrastructural component of Google Search necessary to make the search engine adapt to a dynamic world in which it is regularly confronted with new pages, content, cultural, and political relevance constellations, and so on. As a machine that is built to determine the relative relevance of content with respect to search keywords, a search AI never finishes training. As a dynamic process, its intelligence capacity lies in the immanence of a hybrid human-computer information processing network. The involuntary involvement of humans as data generators in Google Search creates a mediatized swarm principle, making the AI of that search engine a performative product of implicit "audience labor" (Fisher, 2015; Fuchs, 2010) in a networked infrastructure of human-machine interaction.

The kind of power relation that is at work in the "trapping and tracking" class of examples builds on a combination of two facts: first, most users are not aware that using the respective service they contribute to a distributed computing network. Although data collection is explicitly stated in Google's terms of service, it is completely invisible at the level of user interfaces; data collection happens in the background and as part of merely consuming search results (see Fisher, 2015 who highlights that this is still a form of labor that creates immediate value). Second, the strategy of "trapping and tracking" builds on the fact that these services are perceived as indispensable by a majority of users. It is not a realistic threat to those companies that users might abstain from using Google Search or from solving a reCAPTCHA. Much unlike online games, neither service is used as an end in itself, but rather is instrumental for the users to achieve another goal that they *want* to reach, and in the case of reCAPTCHA, there is by definition no way of getting past it without solving it.

A *third form* of harnessing human cognitive resources in AI systems is given by social networking platforms such as Facebook. This form relies on the extrication of *social motivations*, making the user unknowingly participate in a computing network *by acting socially*. Labeling photos on Facebook is a good example for this kind of socially motivated "free labor" in the digital (Terranova, 2000). Tagging someone on an uploaded image is part of everyday social interaction on Facebook; in fact, Facebook as a medium has *created* a UX world in which this is *made* an *essential* aspect of social communication.[2] In this way, Facebook is aggregating a database of labeled facial images that could be used to train a face recognition AI. Facebook has been building its face recognition AI since 2010, and by 2017 it was pretty accurate [Onl.6]. In that year, Facebook began to notify users when their face was automatically recognized on an uploaded photo [Onl.7]. The user could then select whether they want a label with their name added to the image, whether they prefer to stay invisible, or whether it is not even them in the photo. Facebook presents this "new feature" as a measure for better control of privacy, yet it obviously serves another purpose. This is a good trick in the field of HCI design to obtain a constant stream of *verification data* from free human labor to improve the predictions of the face recognition AI.

A lack of built-in verification mechanisms for AI-based predictions is generally one of the main sources of error and distortion in the real social use of predictive ML applications (O'Neil, 2016). A good AI needs feedback loops that help to align its predictions with reality, otherwise false positives (in other circumstances, false negatives) will not be discovered and controlled for by re-calibration of the AI. With the new "feature," Facebook set up such a feedback loop using UX design and taking advantage of a growing privacy sensitivity to capture human collaboration. The stream of training/verification data generated by this infrastructure is an integral part of the apparatus, which, *as a whole*, is referred to as Facebook's face recognition AI. Similar to the example of Google Search, this case shows that there is often no strict separation of the training and inference phases of an ANN model. DL models are often continuously re-calibrated in real time using human-generated verification data; the training phase overlaps with the interference phase and training data often take the form of verification data.

In this socially motivated form of capture, the power relation between user and machine can best be described as a social "exploit" (cf. Galloway and Thacker, 2007) in

the rich sense of the term that includes its meaning in hacker culture: an exploit is a way of taking advantage of a system through a loophole, by hijacking and subtly modulating its functions. In this sense, Facebook is "hacking" itself into the social communication habits of users to capture their cognitive capacities as free labor in a human-aided AI apparatus for face recognition. This form of power operates in part through the production of subjectivity insofar as Facebook created a social space in which such an unusual activity as tagging faces is made an integral part of everyday interaction.

A *fourth form* of capturing human collaboration in hybrid computing networks is given by information mining strategies that build on nudges and economic incentives. An example is when a health insurance service offers discount to customers who use a physical activity tracker, or a nutrition tracking app, to record step counts, movements, dietary habits, and so on. Similarly, some auto insurances offer discounts for installing a Global Positioning System (GPS) tracking and accelerometer device in ones car, tracking not only the individual location history but also the user's "driving style" (O'Neil, 2016: 168–173). Insurance companies use this kind of data to correlate it with the personal medical record of that person (health insurance) or with the rate of damages and incidents of the driver (car insurance). The idea is to use data analytics to predict diseases or addictions, or respectively, to identify driving styles and routes that correlate with higher risk of incidents. In both cases, behavioral data are used to train an AI that classifies individual users in terms of (economic) risk categories, which then is used for individual insurance pricing (O'Neil, 2016).

In order for this to qualify as an example of free human labor in a hybrid AI network, one needs to point out in what way humans, by wearing activity trackers or equipping their cars with GPS trackers, are providing a piece of *computation* to the machine network. In fact, by providing their data, each user becomes part of a distributed routine by means of which *any* other user can be classified as high-risk or low-risk. Slightly simplified, providing ones data amounts to enabling one more comparison between anyone and oneself; it means one more computational operation that refines the outcome of the prediction. This may seem an indirect way of contributing to a computational network, yet it is significant because the AI in question does not have a built-in mechanism on its own to distinguish safe from risky driving styles or healthy from unhealthy fitness habits. It has to learn this from user data and each user, by providing their data, does a little bit of the *work* of training the predictive system. At the same time, the negative consequences resulting from high-risk classifications are visible only to some users as they are often asymmetrically distributed to the disadvantage of the poor (O'Neil, 2016).

The power relation involved in this form of capture (and now I am referring only to the moment of capture, not to the negative consequences one might suffer from being classified as high-risk) is a soft one, often described as "nudging" that pushes the user in a certain direction, for instance, by economic incentives.[3] In these cases, the nudge is further enabled by the fact that many users do not see the *collective* damage of providing their data, but stick to their individual perspective in which it seems to them that they "have nothing to hide."

A *fifth form* of harnessing human cognitive resources in distributed computing networks is crowdsourcing on platforms such as Amazon Mechanical Turk ("MTurk"). This platform for small, low-paid, on-screen tasks was publicly launched in 2005, around the

same time Luis von Ahn and his team were developing their ideas to extract such work for free through gamification. MTurk was originally developed for Amazon's own purposes, as an infrastructure to outsource a number of repetitive tasks related to maintaining their product catalog, such as updating product information and identifying duplicates. In the jargon of the platform, small tasks that can be processed by humans in a few seconds for a few cents are called "HITs"—"Human Intelligence Tasks" ([Onl.8]). On MTurk, there is always a worldwide community of casual workers available to process HITs that are submitted by large companies or research institutions through an API. This community of workers is mostly located in the Global South and often economically precarious ([Onl.9–10]). Their deployment through MTurk is often cheaper than developing a full automation of the tasks; if automation is desired, these workers can be used to create training or verification data.[4]

As the computer scientist Jaron Lanier (2014) puts it, MTurk really "allows you to think of the people as software components." Through an API that is available for all major programming languages, processing HITs on a "human processor" can be integrated smoothly into classical programming code (Figure 1). Such a programming code is indeed partially executed on silicone-based processors and human brains. The access to human workers through an API largely conceals the social dimension and social consequences of this form of capture. This is particularly evident in commercial content moderation (CCM), which is the outsourcing of moderation tasks from social media platforms to service companies that rely on clickwork for manual reviews of

```
ideas = []
for ( i = 0; i < 50; i++ ) {
    // generate code for human processor:
    question = "What's fun to see in Berlin?
        Ideas so far: " + ideas.join(", ")

    // create Human Intelligence Task (HIT)
    // via MTurk API:
    hitId = mturk.createHIT( ..., question, ...)
    result = mturk.getHITResult( hitId )

    ideas.push( result )
}

ideas.sort( function( a, b ) {
    // generate code for human processor:
    question = "Which is better:"
        + a + " or " + b + "?"

    // creat HIT with MTurk API:
    hitId = mturk.createHIT( ..., question, ...)
    result = mturk.getHITResult( hitId )

    return( result == a ? -1 : 1 )
})
```

Figure 1. Sample algorithm for creating a sorted list of 50 tourist attractions in Berlin using human cognitive resources via *MTurk* API-Call.
Source: The author / adapted from Little et al. (2010).

user-generated content (UGC) (Roberts, 2016a). The army of CCM workers deployed by Facebook, Twitter, Tinder, and so on to review UGC is conservatively estimated at more than 100,000 people worldwide, more than double the number of Google employees and 14 times that of Facebook; many of them are located in low-wage areas and in the Global South [Onl.11]. The task of CCM workers is to check UGC for compliance with laws and platform guidelines. To do this, they review such content item by item, day by day, to sort it into different risk categories. Most platforms do not send all images or posts uploaded by users through such a manual review as this would be very expensive. Often UGC goes live immediately and only when another user reports it as inappropriate (which is also a form of capturing human collaboration) is it sent to CCM. In this way, as Sarah Roberts (2016b) points out, CCM workers do not see the entire spectrum of uploaded material, but a pre-selected list, which

> "often focuses on content that is highly sexual or pornographic, depicts the abuse of adults, the abuse of children (physical and/or sexual), the abuse and torture of animals, content coming from war zones and other areas besieged by violent conflict, and any material that is designed to be shocking, prurient or offensive by nature."

Investigations by journalists and researchers point to psychological damage such as post-traumatic stress disorder caused by this work. This is a form of social cost that adds to the exploitative financial working conditions in the gig economy and is not covered in the balance sheet of companies that use these kind of services ([Onl.11; Onl.12]). Hence, clickwork shows how economic power relations shaped by precarious work conditions and global economic disparities can be directly transformed into computing power. Indeed, a clickwork platform is a machine that converts economic power differentials into computing power.

In times when politics effectively force platform companies to install upload filters, AI methods for the automatic classification of content are being developed [Onl.13]. At present, these are not mature enough to allow a computer system alone to identify abusive content with great accuracy [Onl.14]. A partial automation is still conceivable; by combining silicon-based AI techniques with the selective use of clickwork, a human decision is only necessary when the ML model delivers an uncertain result. This hybrid form of automation is more efficient and cost-effective, for then it forms a hybrid human-machine computing network that implements, as a whole, a human-aided AI for content filtering.

Conclusion: a cybernetic notion of AI

I refer to the five forms of harnessing human cognitive, affective and social capacities in hybrid human-machine computing networks, together with the various (commercial) products and services that are built upon them, as the media-sociological dispositive of "Human-Aided AI." This concept aims to place the nexus of media technologies, social interaction, the molding of end-user subjectivity, and new forms of labor in everyday machinated power relations at the center of a discussion of AI. While I do not deny the relevance of developments in computing technology for the success of DL, my point is

to stress that today most commercially relevant AIs are emergent phenomena in hybrid human-machine networks that rely on specific media-cultural prerequisites.

The term "Human-Aided AI" also aims at questioning the classical notion of "intelligence" as an autonomous and sovereign rational capacity located within a physically delineated apparatus or living being. Human-aided AI is an emergent and distributed intelligence capacity of hybrid human-machine assemblages. To make this contrast clear, I will refer to the classical (autonomous and confined) understanding of intelligence, if it is applied to AI, as *simulative* understanding of AI. Simulative AI is strongly tied to the idea of an intelligent system as a black box that passes the Turing Test (cf. Copeland, 2000; Turing, 1950). In this logic, intelligence is ascribed to, and located within, a system if it can simulate human cognitive performance in its *external* interactions. (On the semantics of "simulation" see Turing, 1996 [1951].) The symbolic paradigm of AI, or GOFAI, is an example of a simulative conception of AI. It conceives of AI as a problem-solving, language processing, or chess playing capability of a system that manifests in its external relations and within the constraints of the mediality of its interactive channels to the outside world. For instance, Joseph Weizenbaum's (1966) ELIZA "chat bot" interacted through a typewriter; a chess automaton interacts via a chess board, be it physically present or visualized on a screen. By introducing the qualifier "simulative" to describe this connotation of AI, I seek for an interface and media theoretical (rather than algorithmic) characterization of this type of AI. Regardless of its concrete algorithmic implementation, simulative AI assumes that intelligence is located within an apparatus and is evident in its external interaction that resembles the intelligent behavior of humans, as it can be tested by some variant of the Turing test.

I maintain that this principle of simulation is abandoned in the switch to DL. This is because the media-theoretical logic of interaction between the intelligent device and humans has changed: from simulation to immersion of human skills, from the machine "growing into" human cognitive capacities to exploiting human cognitive capacities, from the machine substituting human labor to the power strategy of capturing human labor within distributed higher-order apparatuses. I refer to this as *cybernetic understanding* of AI—which is meant as an oppositional concept to simulative AI. I call it cybernetic because the structural form of its relation to humans is that of feed-back loop control (Rosenblueth et al., 1943): as we saw in the examples of Facebook and Google Search, human action within the apparatus generates training and verification data that feed into AI predictions; however, there is actually a double feedback effect as cybernetic AIs also back-feed on the people who use them. By communicating through Facebook, searching with Google, or providing data to one's health insurance, the human-machine network (aka "AI") modulates the user's movements, knowledge, well-being, and affects. This double feedback effect of DL-based AI apparatuses subjugates users to a mechanism of control. Control is a subtly modulating form of power that is central to the sociotechnical mechanisms Norbert Wiener and others described under the title of "cybernetics" (Ashby, 1957; Wiener, 1954).[5] Using the term cybernetic AI stresses that the AI apparatus is not just run by unilateral exploitation of free labor, but rather facilitates an emergent cognitive capacity of the apparatus that is regularly consulted by users themselves. This leads to a reciprocal co-dependence of users and AI that is at the heart of specific forms of mechanized power and control in the dispositive of human-aided AI.

While this form of power is generally weak and non-repressive, it can still manifest in strong forms of subordination that have recently been debated as "algorithmic discrimination," "automated inequality," or big data–based social selection (cf. Eubanks, 2018; Noble, 2018; O'Neil, 2016).

The conceptual difference between simulative and cybernetic AI concerns the form of mediated relations between machines and humans: In simulative AI, intelligence manifests in a relation of *comparison* or *resemblance of skills* across external boundaries of humans and machines. In cybernetic AI, intelligence is an emergent and distributed capacity of the hybrid human-machine assemblages *as a whole*, while the single relations between humans and machine are power relations that make the human a functional part of that machine. Simulative AI reproduces human skills, while cybernetic AI embeds them. This shows that recent developments in commercial applications of AI come with a significant shift in the implicit conception of intelligence itself. In our specific media-cultural context, this shift is related to concrete design principles and developments in the field of HCI. The founding father of UX design, Donald Norman (1988), speaks of design as a "psychology of everyday things." Seen from this angle, interaction design is the business of colonizing the cross-section of sociality and technology by a creative "will to power." In the media-cultural dispositive of human-aided AI, people are made to habitually attach to digital interfaces, which enable harnessing them as data servants and free labor force. In consequence, human-aided AI is not just one technology among many, but a historical formation. It is based on socio-economic conditions, technological standards, political discourses, and specific habits, subjectivities and embodiments in the digital world that are themselves a product of everyday interaction with digital media (Mühlhoff, 2018). In 2006, a specific online game had to be set up to gain training data for a specific AI problem. With the emergence of the dispositive of human-aided AI in the years since this relationship has turned upside down. Data are constantly generated and collected, and its availability even tend to precede the concrete use for an AI problem.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research of this article has in part been supportet by the Collaborative Research Center SFB1171 Affective Societies, project B05, at Freie Universität Berlin, funded by the Deutsche Forschungsgemeinschaft (DFG), 2015–2019.

ORCID iD


Rainer Mühlhoff https://orcid.org/0000-0002-3936-9919

Notes

1. "Human cycles" allude to the term "processor cycles" in computer science, thus referring to a fictitious unit of information processing power of the human brain.
2. Scholarship in the post-Marxist theoretical tradition compared Facebook to a "digital assembly line," where millions of free workers generate the economic value of the company (Scholz, 2013). See also Fisher and Fuchs (2015), Fuchs (2010), and Terranova (2000). These approaches start from extending the concept of *work* to the digital sphere in order to subject the phenomenon to a (post-)Marxist strategy of economic critique.

3. The term "nudging" originates from behavioral economics (see Thaler and Sunstein, 2008). For a critical discussion in the context of interface design, see Mühlhoff (2018).
4. "Mechanical Turk" alludes to the (fake) chess computer of the Austro-Hungarian Baron von Kempelen, who became known as the "chess Turk" in the 18th century, in whose generous wooden housing a man was hiding, covertly playing the game (Levitt, 2000).
5. As a precursor to what we see in human-computer networks today, the notions of feedback and control have been translated by the sociocybernetics movement into the sociological framework of systems theory (Geyer, 1995).

References

Ashby WR (1957) *An Introduction to Cybernetics*. London: Chapman & Hall.

Bengio Y (2009) Learning deep architectures for AI. *Foundations and Trends in Machine Learning* 2(1): 1–127.

Bolz N, Kittler F and Tholen CG (1994) *Computer als Medium*. Munich: Fink.

Brooks R (1991) *Intelligence Without Reason* (A.I. Memo). Cambridge, MA: MIT Press.

Copeland J (2000) The Turing test. *Minds and Machines* 10(4): 519–539.

Deterding S, Dixon D, Khaled R, et al. (2011) From game design elements to gamefulness: defining gamification. In: *Proceedings of the 15th international academic Mindtrek conference: Envisioning future media environments*, Tampere, 28–30 September, pp. 9–15. New York: ACM.

Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Fisher E (2015) Audience labour on social media: learning from sponsored stories. In: Fisher E and Fuchs C (eds) *Reconsidering Value and Labour in the Digital Age*. New York: Palgrave MacMillan, pp. 115–132.

Fisher E and Fuchs C (eds) (2015) *Reconsidering Value and Labour in the Digital Age*. New York: Palgrave MacMillan.

Fuchs C (2010) Labor in informational capitalism and on the Internet. *The Information Society* 26(3): 179–196.

Galloway A and Thacker E (2007) *The Exploit: A Theory of Networks*. Minneapolis, MN: University of Minnesota Press.

Geyer F (1995) The challenge of sociocybernetics. *Kybernetes* 24(4): 6–32.

Goodfellow I, Bengio Y and Courville A (2016) *Deep Learning*. Cambridge, MA: MIT Press.

Haugeland J (1981) Semantic engines: an introduction to mind design. In: Haugeland J (ed.) *Mind Design*. Cambridge, MA: MIT Press, pp. 34–50.

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

Hutchins E (2001) Distributed cognition. In: Smelser N and Baltes P (eds) *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Pergamon, pp. 2068–2072.

Lanier J (2014) *Who Owns the Future?* New York: Simon & Schuster.

LeCun Y, Bengio Y and Hinton G (2015) Deep learning. *Nature* 521: 436–444.

Levitt G (2000) *The Turk, Chess Automaton*. Jefferson: McFarland.

Little G, Chilton LB, Goldman M, et al. (2010) Turkit: human computation algorithms on mechanical turk. In: *Proceedings of the 23nd annual ACM symposium on user interface software and technology*, New York, 3–6 October, pp. 57–66. New York: ACM.

Mühlhoff R (2018) Digitale Entmündigung und "User Experience Design." Wie digitale Geräte uns nudgen, tracken und zur Unwissenheit erziehen. *Leviathan—Journal of Social Sciences* 46(4), pp. 551–574.

Mühlhoff R (2019a) Big data is watching you. Digitale Entmündigung am Beispiel von Facebook und Google. In: Mühlhoff R, Breljak A and Slaby J (eds) *Affekt Macht Netz: Auf dem Weg zu einer Sozialtheorie der Digitalen Gesellschaft*. Bielefeld: Transcript, pp. 81–107.

Mühlhoff R (2019b) Menschengestützte Künstliche Intelligenz: Über die soziotechnischen Voraussetzungen von "Deep Learning". *ZfM – Zeitschrift für Medienwissenschaft* 11 (2/2019): 56–64.

Newell A and Simon H (1976) Computer science as empirical enquiry: symbols and search. *Communications of the ACM* 19: 113–126.

Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Norman D (1988) *The Psychology of Everyday Things*. New York: Basic Books.

O'Neil C (2016) *Weapons of Math Destruction*. London: Penguin.

O'Reilly T (2005) What is web 2.0. design patterns and business models for the next generation of software. Availabe at: http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html (accessed 7 March 2015).

Roberts ST (2016a) Commercial content moderation: digital laborers' dirty work. *Media Studies Publications* 12. Available at: https://ir.lib.uwo.ca/commpub/12

Roberts ST (2016b) Digital refuse: Canadian garbage, commercial content moderation and the global circulation of social media's waste. *Wi: Journal of Mobile Media* 10(1). Available at: http://wi.mobilities.ca/digitalrefuse/

Rosenblueth A, Wiener N and Bigelow J (1943) Behavior, purpose and teleology. *Philosophy of Science* 10(1): 18–24.

Rumelhart DE and McClelland JL (1986) *Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Rumelhart DE, Hinton GE and Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323: 533–536.

Scholz T (ed.) (2013) *Digital Labor: The Internet as Playground and Factory*. London: Routledge.

Sudmann A (2018) Zur Einführung. In: Engemann C and Sudmann A (eds) *Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz*. Bielefeld: Transcript, pp. 9–23.

Sun R (2014) Connectionism and neural networks. In: Frankish K and Ramsey W (eds) *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press, pp. 108–127.

Terranova T (2000) Free labor: producing culture for the digital economy. *Social Text* 18(2): 33–58.

Thaler RH and Sunstein CR (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Turing A (1937) On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 2(1): 230–265.

Turing A (1950) Computing machinery and intelligence. *Mind* 59(236): 433–460.

Turing A (1996 [1951]) Intelligent machinery, a heretical theory. *Philosophia Mathematica* 4(3): 256–260.

von Ahn L (2005) *Human computation*. Doctoral Dissertation, School of Computer Science, Carnegie Mellon University. Retrieved from http://reports-archive.adm.cs.cmu.edu/anon/anon/usr/ftp/home/ftp/2005/CMU-CS-05-193.pdf

von Ahn L (2006) Games with a purpose. *Computer* 39(6): 92–94.

von Ahn L and Dabbish L (2004) Labeling images with a computer game. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Vienna, 24–29 April, pp. 319–326. New York: ACM.

von Ahn L, Blum M, Hopper NJ, et al. (2003) Captcha: using hard AI problems for security. In: Biham E (ed.) *International Conference on the Theory and Applicationsof Cryptographic Techniques*. London: Springer, pp. 294–311.

von Ahn L, Liu R and Blum M (2006) Peekaboom: a game for locating objects in images. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Montreal, QC, Canada, 22–28 April, pp. 55–64. New York: ACM.

von Ahn L, Maurer B, McMillen C, et al. (2008) reCAPTCHA: human-based character recognition via web security measures. *Science* 321(5895): 1465–1468.

Weiser M (1991) The computer for the 21st century. *ACM SIGMOBILE Mobile Computing and Communications Review* 3: 3–11.

Weizenbaum J (1966) Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9(1): 36–45.

Wiener N (1954) *The Human Use of Human Beings*. Boston, MA: Houghton Mifflin Harcourt.

Woodcock J and Johnson M (2018) Gamification: what it is, and how to fight it. *The Sociological Review* 66(3): 542–558.

Online Sources

[Onl.1] https://www.google.com/recaptcha/

[Onl.2] https://cloud.google.com/products/ai/ and https://www.ibm.com/cloud/machine-learning

[Onl.3] https://www.tensorflow.org/

[Onl.4] https://keras.io/

[Onl.5] https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies

[Onl.6] https://www.npr.org/sections/alltechconsidered/2013/10/28/228181778/a-look-into-facebooks-potential-to-recognize-anybodys-face

[Onl.7] https://www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged/

[Onl.8] Amazon Mechanical Turk. API Reference. API Version 2017-01-17. Online (PDF): https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMturkAPI/amt-API.pdf

[Onl.9] http://techlist.com/mturk/global-mturk-worker-map.php

[Onl.10] https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/

[Onl.11] https://www.wired.com/2014/10/content-moderation/

[Onl.12] https://derstandard.at/2000035900517/

[Onl.13] https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how/

[Onl.14] https://www.vice.com/en_au/article/wj7mv5/instagram-is-using-ai-to-filter-out-toxic-comments

Videos

[Vid.1] von Ahn L (2006) Human computation. *Google Tech Talk*, 26 July. Available at: https://www.youtube.com/watch?v=tx082gDwGcM

[Vid.2] Ng AY (2017) Artificial Intelligence is the new electricity. *Talk at Stanford Graduate School of Business*, 25 January. Available at: https://www.youtube.com/watch?v=21EiKfQYZXc

Author biography

Rainer Mühlhoff is a postdoctoral research fellow in philosophy at the Cluster *Science of Intelligence* at Technical University Berlin, where he works on "ethics of design in AI and robotics". Rainer's research areas are social philosophy and critical theory of the digital society. Rainer studied mathematics, philosophy, and gender studies in Heidelberg, Leipzig, and Belin (http://rainermuehlhoff.de).