Artificial Ignorance

What AI Cannot, Would Not Want to, Should Not, and Must Not Know

An essay by Anil K. Jain

Artificial intelligence (AI) is the salvation of humanity. And it is its downfall. Both theses have passionate advocates. The AI discourse (beyond engineering) oscillates between utopia and dystopia. It is in love with the extreme, and, seemingly, there is no in-between. So, will intelligent machines carry us into the realm of freedom (and not only from A to B) because they release us from the economic imperative of alienated labour and enable us to live in a »lotusland« of unlimited consumption and permanently available services for everyone? Or will they soon not only make us completely superfluous, but disempower and enslave us – their masters (and »ladies«)? The misanthropist hopes for the latter, the optimist fears the former. Only one thing seems very unlikely: that AI will not radically change our lives. Accordingly, tame scientists in 2015 wrote a »digital manifesto« in which they would not call for a revolution against the power of AI machines, but lamented about the expected changes and demanded a »digital enlightenment«. This term was not meant not imply the reduction of enlightened thinking to 1 and 0, to yes or no, to good or bad. They simply wanted to impose rules on the »digital revolution« so that »we all benefit from the fruits of the digital revolution: economy, state and citizens alike« (Helbig et al. 2015 [own translation]). In order to achieve there, among other things, informational self-determination should be supported and transparency should be improved (see ibid.).

Unfortunately, no one seems to have told them that revolutions rarely run regulated. And that, in the context of digitisation, especially in the field of artificial intelligence, the dimension of nescience or »ignorance« is crucial – as paradoxical as it may sound. Therefore, in the following, the significance of ignorance in the field of AI (which is the »natural enemy« of transparency) will not only be examined with regard to what AI cannot known, but also to what it would not want to, should not and must not know. However, first we have to ask ourselves: Is artificial intelligence really intelligent, respectively, can it ever be intelligent? Even: is there such a thing as intelligence at all?

1. THE »CONSTRUCTS« OF AI AND ITS NECESSARY »LIMITEDNESS«

Already the term »artificial intelligence« self-evidently implies that a) there is intelligence and b) it can be modelled artificially. The model of intelligence which is relied on here and which is the measure for the evaluation of the reached level of (artificial) intelligence is generally human intelligence. However, looking at ourselves realistically we ought to admit: we are rather »limited« beings. Already dialectics, so to say the synthesis of -1 and +1, not only overcharges mathematics (no: the sublating »solution« is not 0!) and the machines relying on it, but most human »spirits«. Accordingly, how much intelligence can we really expect from systems that are modelled after such a poor »original«?

Thus, the first limitation of AI, which causes its »ignorance«, is related to its model and measure: (the hu)men. The human is, at the same time, systematically overrated and underrated in the context of AI. The overrating does not only relate to the suitability of the human as a model for (artificial) intelligence but also to his/her readiness for and interaction competence with intelligent machines. Not only do diffuse fears of contact matter here. Sometimes, humans simply *cannot* act as expected by the more or less intelligent machines. An example that probably many have themselves already suffered from is the dialogue with (increasingly »intelligent«) voice recognition systems and chat bots of hotlines. Even if the voice recognition works (i.e. the machine »understands« what we are saying) it does not at all mean that one gets the expected support – when formulating a request that exceeds the horizon of expectation of the system. The only »way out« is to ask to get connected to a »real« person – if this is even offered. However (and that is also demonstrated impressingly by the common practice): you can never be sure that the human counterpart is more »responsive« than the bot. This is especially unlikely when encountering a person who is not willing (or authorized) to depart from common routines and formal rules. Unflexibility and a lack of sympathy are thus not »exclusive« problems of artificial intelligence but are generally a side-effect of rule-based action which is so common in many areas. The algorithmic coding of AI systems tends to reproduce and even enforce this kind of »bureaucratic« behaviour – which already drove quite a few into madness.

In order to find real acceptance, AI systems would need to learn to show a level of flexibility and »empathy« that many humans are not capable of (or willing or »empowered« to show). For artificial intelligence, however, it is especially difficult to overcome rule fixation due to its technical and algorithmic fundamentals. This difficulty is sustained by a »blaming effect«. Like racists are (amongst others) transferring the misconduct of individuals to the entire stigmatized group (see, e.g., the discourse on »refugee delinquency«), it is likely that malfunctions or singular mistakes in AI systems will be attributed to the entire technology. Even if, for example, the total number of traffic fatalities could be significantly reduced by AI control, one spectacular accident resulting in deaths and caused by an autonomous vehicle could probably question the social acceptance of this technology in general. We would rather be run over by a drunken fellow human than by a misguided mobility system. Thus, artificial intelligence would need

to work almost error free and still be hyper flexible in order to be accepted by us – an obvious optimization conflict.

Our own mistakes we, however, gladly ignore. Let's be honest: What we expect from the machines, we do not meet ourselves in any way. We are quite often incompetent, unpredictable and, yes, »stupid«. In organization studies, the »Stupidity-Based Theory of Organizations« (Alveson/Spicer 2012) takes this into account. And in (application-oriented) IT the DAU – the dumbest assumable user – has long been the determining factor. In the field of AI, however, the DAU is a more difficult challenge than in, say, the design of graphical user interfaces. Just as the (seemingly) stupid are a disturbance to the (seemingly) clever – by forcing them to make the world »dumber« than it would be desirable (and necessary) from their perspective –, human stupidity systematically limits the developmental potential of artificial intelligence, because AI systems need to adapt to the stupidity of users and to imitate them in order to be accepted by them. For who always knows better and who wins each and every chess match will not be loved. Artificial intelligence will therefore have to pretend to be even more ignorant than it inevitably already is (since it is generally oriented towards a »stupid« model, the human being, and it is also created by stupid, i.e. »human« developers who do not/cannot know better).

But AI is – in current practice – not only doomed to stupidity and ignorance towards certain user needs. Foremost, it is ignorant of that which would be necessary to actually unfold its utopian potentials. Since technology – and its pro-gress – is never impartial. Social conditions (their injustices and imbalances) »solidify«, manifest in it. In this respect, it was naive of Marx to assume that the capitalist economy would necessarily unfold the productive forces needed to satisfy the demands of the communist »realm of freedom«. Also artificial intelligence will in fact rather reproduce the social power structures than create the means to transform them – at least as long as it is (by design) ignorant of the real needs of the individual and the society.

That is why I ask for a, so to speak, »Copernican turn« in the treatment of AI. Formulated loosely in reference to Kant (1929 [1787]: p. 22): We must therefore make trial whether we may not have more success in the tasks of artificial intelligence, if we suppose that ignorance (not intelligence) were its determining element.

2. THE »NATURAL« LIMITS OF KNOWLEDGE: WHAT AI CANNOT KNOW

In 1979, the French philosopher Jean-François Lyotard published a text commissioned by the »Conseil des universités du Québec«. This text (originally entitled »La condition postmoderne«) became one of the programmatic works of the discourse of postmodernism/poststructruralism. The main topic of this rather short volume is knowledge, which Lyotard (1984 (1984]) regards as a more and more contested resource. He thus notes: »It is conceivable that the nation-states will one day fight for control of information, just as they battled in the past for control over territory, and afterwards for control of access to and exploitation of raw materials and cheap labor« (p. 5). Everything that cannot be transformed into usable information, however, is left

on the side-lines (see ibid.: p. 23). At the same time, knowledge inevitably becomes more and more a question of governance (see ibid.: p. 35) – and thus an object of conflicting positions. But – as it were – also from within scientific progress more and more questions the (formerly accepted) claim of the absoluteness of knowledge (see ibid.: p. 112ff.). This primarily feeds the distrust towards the great meta-narratives (such as the emancipation of the subject). According to Lyotard (see also ibid.: p. 13ff.), this delegitimization of knowledge and the doubts about meta-narratives are characteristics of the »postmodern« present.

However, it is safe too assume that there has actually never been any indubitable knowledge. How else could one explain the desperate, »obsessive« attempts of theologists and – later – positivists and rationalists to establish certainty? Descartes' striving for absolute certainty, for example, is born out of a global uncertainty, as he himself explains in the first chapter of his »Meditationes«. His attempt was therefore doomed to fail. Historically, there have always been struggles for interpretational sovereignty – between Catholics and Protestants, Christians and Muslims, empiricists and idealists, materialists and hermeneuticists, etc. None of them were finally able to decide those in their favour.

These struggles have existed and continue to exist simply because there are apparently »natural« boundaries of knowledge. Werner Heisenberg's investigations on »The Actual Content of Quantum Theoretical Kinematics and Mechanics« (1927) are frequently and readily used here as a testament in which his famous »uncertainty principle« found its first formulation: At the same time, location and momentum of a particle cannot be exactly determined. The actual (epistemological) consequence that we should draw from these quantum-physical considerations is, however, not the »focussing« on blurriness, but that, if we follow Heisenberg's argument, there are »objective« limits to knowledge: things that one cannot know and, thus, about which there will never be complete certainty. But even if we do not follow Heisenberg (e.g. because we are not able to since we lack the quantum theoretical foundations), we must, nevertheless, suppose certain »natural« limits of knowledge: the limits of knowledge that are rooted in our perceptual apparatus and our actual capacity of understanding.

As (in this case actually) without any doubt, there are limits to our perception – and thus there are limits to what we can »meaningfully« know. Although we can try to extend these limits by technical means, we still can only experience the »meaning« of, say, microwave radiation (which we cannot »sense« visually) if we can perceive it in other ways: for example when it generates heat that we can »feel«. If it is not possible for us to make such a »sensual« connection, the knowledge about the existence of this kind of electromagnetic radiation remains abstract and »empty« – phenomenologists (and practitioners) do know this.

One could now try to expand the boundaries (of perception) further and further and to connect the resulting new experiences with meaning – in order to become »smarter«. And, in fact, the limits of knowledge and perception are crucial to intelligence. However, not so much by being limiting, but rather by forming the very basis of our intelligence. For in the field of intelligence paradoxically it is true: less (knowledge) is more. This, at first, seems to contradict our everyday experience, where we tend to put knowledge and intelligence essentially into one. Knowledge is, so to speak, the

common »objectification« of intelligence. The underlying assumption to this »objectification« is that one can only attain knowledge through intelligence. But pupils know better: diligence is the much more promising approach here. And also a book is not »smart« just because it contains a lot of information. Rather, this depends on the meaningful connections made.
But what does it mean to make meaningful connections? – It means above all selection and reduction, i.e. not the arbitrary and equally-weighted linking of everything with everything, but the concentration on certain, relevant links. And thus, even in the traditional art of thinking, classical logic, it is not the most exhaustive conclusion, which is considered particularly intelligent and »elegant«, but the simplest possible syllogism – at least if one follows the reasoning of Dignaga, who, in the 5th century, pleaded for a reduction of the five-stage syllogism that had previously been favoured in India (see Roloff 2005). This five-stage syllogism followed the form:

»1. Thesis: There is fire on the mountain.
2. Reason: Because we can see smoke on the mountain.
3. Example: Where there is smoke, there is fire, like in the kitchen, unlike at the lake.
4. Application: There is smoke on this hill.
5. Conclusion: There is fire on this hill.« (Quoted according to ibid. [own translation])

Dignaga proposed to shorten this pattern to the first three steps:

»1. Thesis: There is fire on the mountain.
2. Reason: Because we can see smoke on the mountain.
3. Example: Where there is smoke, there is fire, like in the kitchen, unlike at the lake.« (Quoted according to ibid. [own translation])

Thus the syllogism proposed by Dignaga follows a similar form as the syllogism in the *»logic«* of Aristotle (1889 [ca. 367–344 BCE]: book 3). However, there are good reasons to assume that here (as well as there) the logical circle is not completely closed and parts of the chain of argumentation (specifically the actually necessary »con-clusion«) are to be supplemented in personal contribution. The underlying assumption is obvious: The intelligent person can draw his/her conclusions independently. So, when someone explains everything to us down to the last detail, we may quite rightly assume that we are considered to be somewhat mentally limited.
In the field of perception, reduction is even more important than within logic. A five-step syllogism may seem cumbersome. With perception, though, insufficient »filtering« threatens to collapse the system. Thus, our perceptual apparatus is not only limited from the outset, for example, in that only a narrow band of electro-magnetic radiation is »visible« to us. In the following steps, too, adequat filtering is perhaps the most important task (see also Wessels 1990: p. 90ff.). Only a fraction of the incoming information is actively processed. Individuals who have deficits in the field of filtering, such as in the clinical picture of autism, have considerable problems

navigating the environment. All that remains is unstructured noise, as if we stand too close to a pointilistic painting or a megascreen and can no longer recognize anything »sensible«. Because sense is also a result of the filtering. The everything is nothing (»smart«). Something is always something specific.

Again, from this example we can see that it is not only »useful« but also necessary to address the question of (artificial) intelligence from the other side: the – unavoidable – ignorance (towards specific objects). One could also formulate: The (right) degree of ignorance determines the degree of intelligence. The »nature« of intelligence is therefore based on the »art of discrimination« into the important and the less important. All differences, however, are – according to this statement – ultimately »constructed«: produced, an *achievement* of our perceptive and cognitive apparatus. If we follow this view of intelligence, then intelligence is necessarily subjective, because what is important or unimportant and what can be meaningfully linked with something else is always dependent on the respective point of view and the respective situation. Objectively seen (i.e. detached from the subjective point of view) there is accordingly no intelligence but only the arbitrariness of discrimination.

This is why intelligence is not objectively measurable. The witty phrase »intelligence is what the tests test« (Boring 1923) aptly expresses this subjectivist dilemma of intelligence. And, therefore, we all quite rightly consider ourselves to be the smartest: because we are – from our subjective perspective. Nietzsche's (1908: ch. 2) writing »Weil ich so klug bin« (»Because I Am so Smart«) was hence perhaps less a sign of the emerging megalomania, but owed to the wise insight into the radically subjective nature of intelligence. In artificial intelligence, however, the question arises: Who or what is the subject? (Which means: for whom does it »make« sense?) Irrespective of this question, which at best can be answered politically, sociologically and psychologically (but not philosophically), the only thing that remains to be done is to note: Probably, »artificial ignorance« is the better term (if/especially when we are talking about actually intelligent machines according to the understanding of intelligence as presented above).

3. WHAT AI WOULD NOT WANT TO KNOW

Not wanting to know represents, so to speak, the very core of ignorance, for ignorance not only means a state of lack of knowledge, but, according to general understanding, also implies the will to ignore. Ignorance in the actual sense can thus be distinguished from a simple lack of knowledge by the fact that something could be known, but that it does not want to be known. Ignorance is the decision not to know.

In the field of AI machines, the mere simulation of the human limits of knowledge (e.g. through filter technologies) could already be interpreted as such a desire not to know. After all, one thus deliberately passes certain, actually »accessible« information – for the sake of »more intelligent« results and for reasons of information processing efficiency. Although ignorance

is not very popular (who would like to call themselves »ignorant«?), not wanting to know however does indeed make sense for people. Yes, ignorance is at times even necessary for survival – and sometimes it's quite »stupid« to want to know something.

This is not only true for those cases where something cannot be known at all, but one still puts a lot of energy into finding it out. Much more often knowledge is burdening – and we therefore do not wish to know. This can happen consciously or unconsciously. The conscious decision to ignorance can be called »deflection«. We try to focus on other things, try to overlook something. The unconscious will not to know is commonly referred to as »repression«. According to psychoanalysis, repression plays a central role in the mental economy (see e.g. Freud 1977 [1940]: S. 225ff.). However, psychoanalysis, too, considers it to be predominantly problematic. Repression contradicts the spirit of Enlightenment which seeks to bring all knowledge to light. And, after all, psychoanalysis represents a climax in the movement of Enlightenment as it promises the subject – by means of therapeutic reflection – self-perfection and liberation from inner constraints. Psychoanalysis is thus an expression of the rationalist desire for knowledge and control over one's own self. This control, however, has to be acquired by hard work and it requires to unveil the impulses of the »Id« (which are suppressed and repressed by the super-ego) in order to experience »healing« (through knowledge). But there is always the danger that the underlying traumas thus will break their (new) path. Accordingly, we are dealing with a dialectics of repression (and the will not to know), which I in general like to call the »dialectics of reflection and deflection« (see Jain 2000).

The reflexive impulse, which essentially is to permanently address and question each and everything (including oneself), bears an excessive character. Reflection cannot stop. This applies even more to reflexive learning processes – even in the rather »formalistic« understanding of AI approaches. For instance, the website of the »Artificial Intelligence Lab« of the »University of Michigan« warns: »the agent [in a reflexive learning system] does not consider the possible costs of learning a particular piece of knowledge. These costs hinge on the usefulness of knowledge: reflexive systems learn everything, even knowledge that does not promise to enhance the agent's behavior.«

Reflexivity therefore tends towards excess. It is not only annoying (for others), but straining (for oneself), in that it demands to progress further and further. That is why reflexivity calls for counter forces, the counter forces of deflexivity, of distraction, of displacement, of the will not to know. Deflexivity, however, is not only a protection from the excesses of reflexivity, but in some sense a dialectic expression of the forces of persistence that are directed against the change demanded by reflexive impulses. Deflexivity is thus »ignorant« towards innovation as it fears change and the unknown. On the other hand, every innovation must become concrete and real if it were to prevail in practice. And that means: at some point there has to be an end to reflection, you have to stop questioning. It is time to act and to ignore everything that could lead one astray. Stubborn and persistent. Against the (probably) better knowledge of others and their criticism. That may not always be wise, but it is necessary.

And this kind of (deflexive) ignorance, this defensive desire not to know, is therefore also an essential faculty for AI systems in respect of decision making and in order to assure agency. There must be a stop rule that prevents reflexive excesses: the infinite regress. However, it is difficult to define such rules in a purely formalistic and »numerical« way. For example, if you limit the number of passes through an algorithm, who grants that the ideal result would not have come out exactly the next (prevented) run? Humans are (mis)guided by »gut feelings« and »inherent stubbornness« in such situations. They often fail. But sometimes they have indeed success with that »method« (few times even ground-breaking one). Again, it is crucial not wanting to know the right thing. What that is can usually only be determined retrospectively. AI systems would therefore have to be able to look into the future – or develop their own kind of »stubbornness« in order to simulate this ability of not wanting to know the right thing. However, we will hardly tolerate and stand machines that are stubborn, ignorant and dismiss our »feedback« – a dilemma for artificial ignorance. Thus, AI is doomed to perpetual learning and will never know anything (complete) – and some will want to add: just like us. For in fact, not wanting to know (any more) is the actual »safeguard« of (practical) knowledge. And as one could summarize (so far): A relevant portion of the »actual« intelligence is the right kind and timing of ignorance (in the sense of »not wanting to know«).

4. WHAT AI SHOULD NOT KNOW

Besides the epistemological (not being able to know) and the »pragmatic« (not wanting to know), nescience also implies an ethical, normative dimension: certain things should not be known. A legal expression of this »obligation to ignorance« are, for example, the regulations of data protection. It is assumed that individuals have privacy and should retain control over their personal information. AI systems must, of course, comply with these data protection standards. That means, however, they either should not at all be able to obtain certain information, or, that they must »forget« this information, respectively shield it from access in such a way that »misuse« is impossible. This model of data protection seems to be based on the assumption that certain knowledge »belongs« to the subjects (of knowledge). It is »private« knowledge. But it is precisely this »private« knowledge that triggers desires – not only of state supervisory bodies. It is a highly »interesting« resource for companies to generate surplus, since such knowledge can be used to »play off« targeted, individualised advertising to the users of services. The business model of entire segments of the IT industry is based almost exclusively on this knowledge commodity, which threatens privacy. And the more intelligent the information networking systems, the greater the danger.

Thus, on the one hand, the increasing »transparency« of the individual in the course of digitisation in combination with techniques of »algorithmic intelligence« (in my opinion the more appropriate term) is perceived as a threat. On the other hand, »transparency« is a necessary condition for the acceptance of new technologies. This also applies to algorithmic or artificial intelligence.

We want to know and understand how these technologies work – and above all what happens to the generated knowledge. However, this calls for the subsequent question: Whose knowledge is it? Does it belong to the individual? The society? The companies that develop and apply the technologies to generate knowledge? (Or even the AI systems?)

We obviously give different answers to this question depending on what the concrete content of knowledge is. In the case of knowledge/information about a specific person, we tend to attribute the »ownership« of this knowledge to the respective individual. As opposed to this, knowledge which is not so strongly linked to individuals, we believe should be publicly available. This, however, is conflicting with economic desires. Patent law is an attempt to balance these diverging interests. It offers patent holders the opportunity to economically exploit knowledge for a limited period of time. But they need to publish it – and thus hand it over to the »community«. The counter-model to this is practice is the non-disclosure of knowledge. Here, too, different ethical norms obviously apply to individuals and to »corporations«. We allow the individual to keep his/her knowledge to himself/herself (especially if it is »personal« knowledge). The concealment of (useful) knowledge by private »corporations«, on the other hand, is regarded as immoral by many.

In the field of artificial intelligence, this ambiguous attitude towards what should be known and what should not be known creates a number of problems: Should artificial intelligence be regarded as an individual and therefore be allowed to keep certain pieces of information secret and use it exclusively for itself? Or is artificial intelligence only an »instrument« of its owners (and has no »right« of its own)? Then one would also tend to affirm that the knowledge generated by the AI must be available to the general public (at least after the expiry of certain »protection periods«).

This »problem of ownership« of AI-generated knowledge is boosted by the fact that certain technologies within AI (such as neural networks) technically speaking provide no disclosable knowledge at all but only »applications«. The concrete rules of action the applications are based on cannot be explicated, as is often the case with humane everyday experts. Or, to give another example in this problem area: Who is the author (entitled to royalties) of the composition of an AI-based music composition software in which the humane »composer« has only entered a few parameters? Does not the result »belong« to the AI (or its »creator«) rather than the user? At the latest when AI systems (whether intended or as a »side effect«) develop a kind of (self-)consciousness, artificial intelligence will not be able to be denied certain »subject rights« and thus be entitled to private »knowledge sovereignty«.

But these are ultimately »technical« or legal problems. Even if they could be solved, the related ethical problems of what should not be known would remain. In the area of AI, these are even more infuriating than in general. And they can hardly be solved algorithmically. For at least in this case I tend to largely agree with the »traditionalists« within ethics who refer to Aristotle (2009 [ca. 350 BCE]). He explained that an ethical good must always be in accordance with the common rules. Ethical virtues are therefore a result of habituation (see ibid.: book 2-1). Except that one should strive for moderation and the middle way (instead of the extreme),

according to Aristotle there are no general and »reasonable« rules in the area of ethics (as the latter belong to the area of »intellectual virtues«). Consequently, he equates justice with respect for the law (see ibid.: book 5-2). But ultimately, this means nothing other than to say, like the »sophist« Thrasymachos, that norms are a mere expression of power (see Plato 1992 [ca. 380 BCE]: book 1). The thought of Enlightenment could, of course, accept neither such »realism« nor the mere reference to the »custom of the fathers« as a basis of »morality«. And thus, with his »categorical imperative«, Kant (1997 [1788]: p. 28) tried to formulate a general rule of »right« action based on rational insight: »Act so that the maxim of the will can also be admitted as a principle in the giving of a universal law.«

However, if one were now to attempt to anchor Kant's categorical imperative algorithmically in AI systems, one would quickly find out: such a rule is »impracticable« because it is too abstract. Without references to certain (conventional, generally shared) values, it is impossible to determine the »grounds« of such a general »law«. The obligations of ethics cannot be founded rationally nor theoretically, but only empirically and practically. Any attempt to implement ethics algorithmically must fail because of the (spatial and temporal) contingency of the »moral sense«. At best, artificial intelligence could »learn« ethical norms (from the observation of human behaviour and with appropriate training). But in this case, we cannot predict how, for example, AI systems would deal with knowledge that should not be known – unless explicit instructions are given. These, however, can hardly be given exhaustively for all applications. (Rule) conflicts and the violation of (data) protection rights are thus »pre-programmed«.

5. WHAT AI MUST NOT KNOW

The dilemma of rule conflicts is aggravated when it comes to the question of what must not be known, respectively, the legal dimension and its implementation in AI systems. The area of privacy (and its relation to what should not be known) has already been mentioned above. However, for AI systems as well as for human beings it is true: in our modern world, which is highly juridified, it is compulsory to take legal regulations into account – if one does not want to stand outside the law. Therefore, all actions (and non-actions) of AI systems tend to be affected by legal regulations. And this inevitably leads to conflicts – not only with habitual »moral« rules (because the law is not always in accordance with morality), but also because the legal regulations (can) contradict each other and are ambiguous in their interpretation. After all, law was not conceived by logicians, and also for humans, yes, even for (highest) judges, the balancing of different legal norms is sometimes a difficult task (in which they fail). Every legal norm needs to be interpreted – and the interpretations, like law itself, is subject to (social and cultural) change.

Therefore, you cannot simply »feed« AI systems with laws and »impose« to follow them. They must be able to properly interpret these – ever and ever anew. And, in case of, in fact inevitable, norm conflicts, they must be able to find new »equities«, since there is no clear hierarchy of

legal regulations and norms. Even the application of the »laws of robotics«, as devised by Isaac Asimov (1942) his science fiction short story *»Runaround«*, which seek to assure the prevention of human damage caused by autonomous machines, is not »trivial« at all. The first and (supreme) law indeed sounds quite simple: »A robot may not injure a human being or, through inaction, allow a human being to come to harm.« But how is artificial intelligence supposed to decide if all action (and non-action) alternatives will lead to human injury? By (victim) numbers? Or should there (also) be other evaluation criteria, such as the social relevance/performance of the people in question? One can easily see that even such a seemingly clear and simple rule is in need of interpretation and there is no imperative reading.

For artificial intelligence, there is, however, still another problem: We consider people who fully conform to each and every rule to be rather stupid (or at least »unconfident«). For sophisticated people, the sovereign (self-legislative) handling of legal regulations is, so to speak, one of the core competencies of their »intelligence«. Intelligent people do not rigidly stick to rules but use the interpretative space to maximum extent or even exceed legal boundaries – if the respective rule is obsolete in a specific case, if it benefits them, if it benefits others (and, when acting in a »moral« attitude, if the damage caused is non-existent or minor in comparison to the benefit). However, we would not like to grant such sovereignty to machines. But thus, we condemn the machines to the »stupidity« of rigid rule application.

Paradoxically, even in the event of a rule violation, AI systems would not have to fear any legal consequences. They would remain »unpunished«, because they are not (yet) subjects with legal capacity. Depending on the point of view and interpretation, (legal) responsibility lies with their owners, their users, their designers, their »controllers«. Liability usually implies »free will« – although there is good (scientifically underpinned) reason to assume that human will is actually not as free as legal schoolars presume (see e.g. Chun et al. 2008). In the case of AI systems, however, a free will cannot be supposed at all because algorithmic intelligence can only want what the algorithm dictates. But despite all the problems of interpretation, this »dictate« can only be: strictly follow the rules. Anything else would, of course, be illegal (action). And that's why an AI system must even potentially know how to bypass laws. This artificial ignorance in regard of laws must (statutorily) be implemented in AI.

6. (ARTIFICIAL) IGNORANCE AS A NECESSITY AND PROBLEM

As a conclusion of this reflection on the epistemological, psychological, ethical and juridical problems in the context of AI I want to retain: There is no intelligence without ignorance – and this is even more valid for artificial intelligence. Not only can intelligence only appear in contrast of its »other« – ignorance. What is more, this »other« is an integral part of it. Ignorance is »determining« for intelligence since the core competence of any performance of intelligence is to know what not to know (and when). (Artificial) ignorance is predominantly problematic

only where individual and social needs are ignored and where there is a conflict between intelligence and (ethical and juridical) norms.

But also for a rather »trivial« reason we may admit that we would be rather interested in artificial ignorance than in artificial intelligence: we do not desire real artificial intelligence but prefer its artificial ignorance: »intellectual slaves« searching for the next rail connection or making more valid diagnoses than our family practitioner. Machines which point us to our mistakes, write better criminal stories than us and refuse to follow our stupid orders – we do not want this, for sure. And that is why artificial intelligence is cleverly modelled after the stupid: us. Thus, the circularity of the dominant AI approaches ensures (at least partially) necessary ignorance as well as social acceptance.

LITERATURE:

- Artificial Intelligence Lab/University of Michigan (1994): *Reflexive Learning*. Online Ressource: http://ai.eecs.umich.edu/cogarch0/common/prop/reflexlearn.html.
- Alvesson, Mats/Spicer, André (2012): *A Stupidity-Based Theory of Organizations*. In: *Journal of Management Studies*. Vol. 49 (2012), No. 7, pp. 1194–1220.
- Aristoteles (1889 [ca. 344–367 BCE]): *Organon*. London: George Bell & Sons.
- Aristoteles (2009 [ca. 350 BCE]): *Nicomachean Ethics*. Oxford: Oxford University Press.
- Asimov, Isaac (1942): *Runaround*. In: *Astounding Science Fiction*. March 1942, pp. 94–103.
- Boring, Edwin G. (1923): *Intelligence as the Tests Test It*. In: *New Republic*. Vol. 36 (1923): pp. 35–37.
- Descartes, René (1986 [1641]): *Meditations on First Philosophy [Meditationes]*. Cambridge: Cambridge University Press.
- Freud, Sigmund (1977 [1940]): *Vorlesungen zur Einführung in die Psychoanalyse*. Frankfurt: Fischer.
- Chun, Siong Soon et al. (2008): *Unconscious Determinants of Free Decisions in the Human Brain*. In: *Nature Neuroscience*. Vol. 11 (2008), pp. 543–545.
- Heisenberg, Werner (1927): *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*. In: *Zeitschrift für Physik*. Vol. 43 (1927), Nr. 3–4, pp. 172–198.
- Helbig, Dirk et al. (2017): *Das Digital-Manifest – Eine Strategie für das digitale Zeitalter*. In: Könneker, Carsten (Hg.) (2017): *Unsere digitale Zukunft*. Heidelberg: Spektrum der Wissenschaft Verlag. S. 23–28 sowie der hier zitierte Abschnitt zu *»Digitale Demokratie statt Datendiktatur«* online unter: https://www.spektrum.de/news/wie-algorithmen-und-big-data-unsere-zukunft-bestimmen/1375933.
- Jain, Anil K. (2000: *Politik in der (Post-)Moderne - Reflexiv-deflexive Modernisierung und die Diffusion des Politischen*. München: edition fatal.
- Kant, Immanuel (1929 [1787]): *Critique of Pure Reason*. London: MacMillan.
- Kant, Immanuel (1997 [1788]): *Critique of Practical Reason*. Cambridge: Cambridge University Press.
- Plato (1992 [ca. 380 v.u.Z.]): *The Republic*. Indianapolis: Hackett Publishing.
- Lyotard, Jean-François (1984 [1979]): *The Postmodern Condition*. Manchester: Manchester University Press.
- Nietzsche, Friedrich (1908): *Ecce homo*. Leipzig: Insel Verlag.
- Roloff, Carola (2005): *Dignaga – Vater der indischen Logik*. In: *Tibet & Buddhismus*. Vol. 75 (2005), pp. 30–33.
- Wessells, Michael G. (1990): *Kognitive Psychologie*. München/Basel: UTB.

Information Sheet

Author(s):	Anil K. Jain
Title:	Artificial Ignorance
Subtitle:	The »Phagic« Character of Capitalism
Year of Origin:	2018
Version/Last Updated:	25/04/2019
Original Download-Link:	http://power-xs.net/jain/pub/artificial_ignorance.pdf
First Print Publishing in:	—

In case that you want to cite passages of this text, it is preferred that you quote the PDF-version (indicating the version number/update date) – even if a print-version is available – since the PDF-version may be more profound and/or corrected and updated.

Find other texts of Anil K. Jain and further information at: http://www.power-xs.net/jain/
E-Mail contact: jain@power-xs.net

Feedback is welcome! (However, there is no guarantee of a reply.)

Translation (from German): Auris E. Lipinski and Anil K. Jain