Intelligence. And what computers still can't do

JOBST LANDGREBE[1]
AND
BARRY SMITH[2]

[1] University at Buffalo/Founding and Managing Director, Cognotekt Corporation, Cologne

[2] University of Buffalo

It is with pleasure that we comment on this collection of papers inspired by our book *Why Machines Will Never Rule the World*. In what follows, we summarize the arguments made by the contributors about what we say in the book, and then show where we disagree.

Intelligence is the most complex and differentiated capability of higher animals and humans. According to Max Scheler, who provides the best available definition of intelligence, a living organism

> behaves intelligently when, without engaging in trial and error, it behaves in a meaningful, appropriate way in relation to a situation that is novel (which means: typical neither for an organism of this species nor for this individual organism), and it does this suddenly and above all independently of its prior attempts to solve a problem, and using a solution that is determined instinctively (Scheler 2009, p. 29, translation adjusted).

In higher animals, behaviour of this sort is essential for survival. It is the result of billions of years of evolution. In traditional metaphysics, it is something created by God. Therefore in the traditional framework, to which Harari (2016) alludes in his book *Homo Deus*, we would indeed be like God if we could create intelligence. But, as we show in our book, creating—or better: engineering—an intelligence is something we cannot do: *Homo non deus*.

In her introduction, Janna Hastings portrays the launch of ChatGPT as a major step in the evolution of AI. But has anything really changed since our book was published, as a result, perhaps, of what we have learned from ChatGPT? Certainly not. Will it ever change? We do not think so.

In our book, we discuss sequential stochastic models (including GPT, on which ChatGPT is based) in the section on Foundational Models and GPT (pp. 177 ff.) in chapter 10 (of 'Why machines will not master human language'). Foundation Models are trained on large corpora and can compute the most likely output sequence given an input sequence. They are sequential models parametrised without outcomes to allow the generation of syntacically correct text across a wide variety of domains.

LLMs are just one type of such models, all of which fall short of any sort of intelligence because they do not understand at all what they do. They do not understand the texts they receive as prompts, and they do not know what it is that they produce as responses to these prompts; because they do not know anything. They merely compute the most likely symbol output sequence by using their parametrization, or in other words by using the huge multivariate distribution which was created by using gigantic bodies of digital text corpora in their training. They thereby model the most frequent, regular patterns encountered in this training corpus. That is all.

They also know nothing about the world in which we live. Rather, we can imagine the entities putatively described in their responses—we can call them 'ChatGPT entities'—as similar in some respects to the entities we encounter when we read a work of fiction. These are well-defined, at best, only along a small number of dimensions and then only at a very coarse level of granularity. For all their qualities are determined by the sentences the author chose to include in the work (Ingarden 1973). The difference is that ChatGPT entities can have multiple, contradictory qualities of this sort.

Our book is of course primarily about artificial general intelligence (AGI), the achievement of which is a popular goal of AI engineers, some of whom are confident that this goal will be achieved 'before long' (section 9.1). When ChatGPT appeared it was indeed greeted by some as being *already* an example of AGI, since it could respond to prompts on every subject under the sun (Arcas and Norvig 2023). We must recognize, however, that ChatGPT is an example not of general but of narrow AI. For general AI requires the ability to respond to novel situations, where ChatGPT can respond only to prompts pertaining to the contents of its training material. The repertoire defined thereby is admittedly very large as a result of the fact that it rests on an enormous (but still finite) amount of data used for training. But we can conceive the result of this training, for each successive version of any given Large Language Model, as a very long operator formulated as an equation with (nowadays) some billions of parameters. It is this operator which generates ChatGPT's responses from user prompts as inputs. These responses can appear to us (for different reasons) as surprising. Like all computable algorithms, however, these responses are generated through the mathematically limited resources of a Turing machine.

ChatGPT is in this respect very much like so-called neural machine translation (NMT, for example Google Translate, DeepL). The latter can be applied to many languages and to expressions in these languages covering many different kinds of topics. Yet NMT is nonetheless an example of narrow AI. It is a logic system, which is defined solely by the parameters derived from the data upon which it is parameterized in each successive release. This system is *complicated* (it is based, after all, on an algorithm using 100s of millions of parameters.) But it is not *complex* in the sense of thermodynamics outlined in our book. We show what it is for a system to be complex by identifying seven properties which are characteristic of such systems and which are defined at length in chapters 7 and 8 of our book and discussed further below.

These properties often—for instance in the case for organisms—go hand in hand. The possession of just one of these properties, however, implies that we are dealing with a complex system. We introduce the term 'logic system' to refer to systems which have none of these properties. Such a system has models on the basis of which we can predict its behavior using logic and extended Newtonian mathematics. A nuclear reactor, for example, is a logic system in this sense. But when it gets out of control, it becomes a complex system.

Those who believe in the possibility of AGI make two big mistakes:

1. They extrapolate the innovation pace of the last 100 years to an imagined future evolution of engineering.
2. They do not take into account the negative results of physics obtained above all in thermodynamics, which place natural limitations on what we can model mathematically in our scientific reasoning.

It is in thermodynamics, more than in any other part of physics, that we encounter such limitations. When we survey the world we discover that thermodynamic phenomena are universal. They dominate the behaviour of all natural systems, including all organisms. This means that, for complex systems, we cannot obtain the sorts of synoptic and adequate mathematical models of behaviour that would be needed to engineer a device that would have the ability to simulate such behaviour. Only engineered systems are logic systems, though some natural systems (like the solar system when viewed in terms of a gravitational interactions) can be approximated well as logic systems. The seven properties of complex systems are

1. evolutionary,
2. have element-dependent interactions,
3. show major force overlay,
4. are non-ergodic,
5. driven,
6. context dependent and
7. characterised in their observable behaviour by deterministic chaos.

Human intelligence has all of these properties, which thus rules out the possibility of our engineering a simulation of human intelligence in any sort of machine. This is because engineering a technical artifact satisfying given requirements is possible only when we have a mathematical model which allows us to predict the behaviour of the artifact in a way that conforms to these requirements.

Can we then perhaps create intelligence without mathematics? Certainly. One way might be to use narrow AI. An algorithm like AlphaGO is certainly more *capable at playing Go than human beings. But it is not more intelligent, as it cannot find novel solutions to problems it has not encountered before*, and given what we know about algorithms of this sort it seems inappropriate to use the term 'intelligence' when describing how they work. Another way is by conceiving children, which we will hopefully continue to do. In his book *Superintelligence* Bostrom (2003) suggests some other ways, which we describe at length in our chapter entitled 'Digital Immortality', a chapter which is included in our book in order to provide the reader with some light relief. With all of this in mind, let's review what our critics have to say.

BILL RAPAPORT

First, Bill Rapaport in his opening essay defends the idea that approximations to intelligence could be good enough to emulate intelligence in a way that is sufficient for AGI. He believes that we, in contrast, assume that only what calls a 'perfect emulation of intelligence' would suffice for such an emulation.

Our argument in his eyes thus takes the following form:

1. Only a perfect emulation of the human mind could realize AGI.
2. Such a perfect emulation is impossible
3. Therefore, AGI is impossible.

Rapaport challenges 1. But a proposition like 1. in fact plays no role in our argument. It seems rather to be a product of Rapaport's fascination with the idea of perfection (a term which he uses in one or other form 33 times in his essay).

Our argument is rather that our mathematical capabilities are insufficient to model any complex system, including the complex system that enables the behaviour of a foraging parrot that encounters a novel situation and masters it without any trial-and-error attempts or prior experience.

Why, then, will we never reach the point where such models could be built? Technology evolves, after all. But it does so in accordance with and not against the laws of thermodynamics. We are just unable to understand even how the parrot manifests its version of intelligence, let alone how we ourselves manage to produce intelligent verbal pronouncements. And so we have no idea how to produce the sought for emulation, at any level of exactness.

All of this follows from one of the major insights of physics—that non-ergodic, driven systems with unstable, changing state spaces elude mathematical modelling. The restriction is not technical, but is grounded in the (almost certainly biological) foundations of our mathematical skills. The latter are limited to describing and predicting regularities. Irregularities cannot be captured by mathematical models. Mathematicians and theoretical physicists know this, and experimental physicists struggling to repeat their experiments even more so. But Bill Rapaport does not.

Most importantly, in focusing on the idea that some imperfect emulation of the human neurocognitive system would surely suffice for the purposes of engineering an AGI, Rapaport dismisses the importance of turbulence. 'What,' he asks, 'does turbulence have to do with cognition?'

Well, where should we start? The blood vessels enabling the activity of the neurons and thereby contributing to the way the brain works are full of turbulence. The pattern in which electric currents spread in the brain which is a major foundation of consciousness has characteristics of turbulence at the microstate level. About all of these things we have not even the beginnings of a mathematical model.[1]

One final point relates to that part of our argument in which we show that complex systems 'cannot be modelled in a way that would yield the sorts of mathematical predictions that would enable the sorts of predictions that can be reliably used in technological applications' (p. 123). We cite in this connection a standard work on the mathematics of complex systems, namely Thurner, Klimek, and Hanel (2018, p. 5). Rapaport, however, points to an intriguing line of thought in this book, where its authors state that it is incorrect to 'say that complex systems will never be understood or that, by their very nature, they are incomprehensible. [C]omplex systems are algorithmic' (op. cit., pp. vi, 7).

On closer inspection, however, it turns out that such passages pose no challenge to our main thesis. Thurner et al. do indeed contrast those parts of physics which they call 'analytic' with complex systems, which they call 'algorithmic'. Thereby, however, they define 'algorithmic' not in the way that is customary in computer science. Rather, they use it to apply precisely to the complex system properties studied in thermodynamics. They use it, in other words, to refer not to algorithms *running in computers* but rather to those that *run in nature*. The corresponding natural processes are of course lawful. But we are unable to work out what the laws are and to formulate them in a comprehensive and synoptic way.

JONATHAN SIMON

One running theme in our book is that the human mind does not work like a computer. Simon tries to challenge our argument for this view, and the core of his challenge is that we make the 'claim that a model *in* a logic system can only be an (adequate and synoptic) model *of* a logic system' (p. 10). This, he claims, trades on a conflation of vehicle and content, and as such it 'proves too much.' (p. 11). To prove this point, Simon points to examples of logic systems—namely cellular automata and the game of life—that do seem to model non-logic systems. His thesis to the effect that these logic systems model complex systems is, however, false. Certainly their output can mimic some aspects of the output of complex systems. This is true also for Penrose's tiling, which is aperiodic even though it is constructed using a logic system.

But the fact that we can use a logic system to create output resembling to some extent the output of a complex system by using a logic system does not mean that we can create a complex system. The output of ChatGPT does, after all, resemble to some extent the output of a human intelligence. But this does not mean that ChatGPT is a complex system. The reason, again, is that it is impossible to use mathematics to build a model of a complex system. This means also that it is impossible to engineer a machine that can, for example, take care of young children in the playground, or deal successfully with a heckler by countering his interruption with a witty riposte.

Humans have the ability to cope with such situations, an idea that we use in our definition of AGI as an AI that has a level of intelligence

> that is either equivalent to or greater than that of human beings or *is able to cope with problems that arise in the world that surrounds human beings* with a degree of adequacy at least similar to that of human beings (p. xi, emphasis added).

Because humans live always in environments filled with multiple overlapping and continuously changing complex systems, we have evolved in such a way as to cope successfully with the resultant massive variability. We have evolved, in brief, to cope.

Simon now formulates what he calls the Coping Argument, which he formulates as follows:

1. There is a non-empty class C of complex systems that cannot be adequately and synoptically computationally modelled.
2. We cope.
3. If our brains are computers, we cope iff we adequately and synoptically computationally model some complex systems from C (those in our environmental niche).
4. Therefore, our brains are not computers.

We do not put forward an argument like this in our book for a number of reasons. One reason is that humans—in contrast to other animals—have no environmental niche. To see why, we recommend that Simon read Scheler or Gehlen. He would then see that humans do not have a niche because they have objectifying intelligence, which is universal in the sense that it can be deployed in relation to any subject, at any place and at any time in the past, present or future. This universal intelligence is specific to humans and obviously the capability of a complex system, not a computer, which is a logic system.

It is for this reason alone that Simon's argument is of little concern to us here. What is of (slight) concern is that he claims to find a flaw (he calls it a 'critical equivocation') in our argument. For if, on the one hand 'coping' means, *performing as well as you would if you had adequate and synoptic models of complex environmental systems*, then premise 2. is false: we can't cope, because there can be no such models. If, on the other hand 'coping' means anything less demanding than that, then premise 3. is false.

Very well. We agree. This does not mean Simon has found a way to prove that clause 4. is false. Rather, he has shown that the argument in question has a flaw. Both premises 1. and 3. are false for all the reasons presented in our book. But should we care about this? It is, after all, *Simon's argument*; and we look forward to seeing him fix it. This can be done by abandoning the world view which sees the human neuro-cognitive system as some sort of computer.

His concluding statement reads as follows:

> Even if there are abstract upper bounds on how well computational intelligence can cope with its environment, it doesn't follow that there aren't other computational intelligences that perform better than us on a wide-range of relevant power-seeking tasks (just as the chimpanzees).

This shows that Simon does not understand the pre-condition of power-seeking, which is the possession of a will. Chimpanzees have a will, they are intelligent, and of course they can perform better than we can in many tasks. So can computers. But computers can never have a will (see pp. 275 ff. in our book).

SCHULZ AND HASTINGS

Like Simon, Schulz and Hastings, too, claim that there is 'perhaps a danger that L&S are drawing too narrow a picture of what form AGI may take with their clause (II)', which identifies 'being a machine' as amounting to:

1. the outputs are known and intended by the engineers,
2. some mathematical model of the behaviour exists to predict and verify the behaviour in any given situation.

They conclude that we might be arguing with a straw man. Sure, being an A+B machine precludes having the capability we call AGI. But perhaps, they say, there are other sorts of machines for which this is not the case. First, clause II is not to be found anywhere in our book. As to A, being a machines means to be a logic system planned and created by humans for the purpose of exhibiting an intended behaviour. What Schulz and Hastings miss is that this intended behaviour does not always occur. As to B, there must exist a testing scheme for machines, but that is not necessarily mathematical at all; for example, there is no mathematical model for deceleration with friction (car or airplane tyres).

No engineered system can run complex system processes, though it can approximate some of them if it is designed to do so by a human engineer. Schulz and Hastings complain that we do not sufficiently define what is an artefact, a machine, or an evolutionary machine. They hold that because of this, our 'discourse about evolutionary and potentially intelligent machines [remains] vague'.

Yet our book clearly defines the process of creating machines, which are artefacts, which we deal with at length in chapter 9. This is engineering, which is the planning and realisation of logic systems. Machines are logic systems, and these we define at length in chapter 7. As we prove in chapter 10, there can be no evolutionary machine of the sort which the authors speculatively sketch in their thought experiments. From our point of view the latter come very close to science fiction philosophy, philosophy of a sort that has turned away from real problems.

The main argument of Schulz and Hastings against our view relies on the idea that we can 'create artefacts with evolutionary potentials'. It is precisely this that we cannot do, as our chapter 10 shows. A machine always remains a machine, it cannot have evolutionary properties because it is engineered using a model with a constant phase space (shown in chapter 8). Evolutionary systems—for example you, dear reader—do not have such phase spaces and (unlike Penrose tiles) they exhibit flexible non-ergodic behaviour.

Finally, the authors claim in their abstract that our 'argument presupposes a very contemporary vision of artificial intelligence as a model trained on data to produce an algorithm executable in a modern digital computing system'. This is of course not true. We explicitly deal (chapter 12, section 3 'Transhumanism') with claims (by Bostrom and others) to realise so called super-intelligence without computers by using humans. We show that it is impossible to do this with any type of biological intelligence. But may there not be a type of device that still waits to be invented? Even such a device would be a logical system, and therefore not able to run complex system processes. Again, a better understanding of the negative results of thermodynamics is warranted in this case, too.

EMANUELE MARTINELLI

The main avenue leading to the idea of a complex system in our book goes through the phenomenon of thermodynamics, and specifically through the thermodynamic view of systems first laid out by Prigogine in the 1950s. Martinelli, in contrast, follows Hayek's theory of complexity to elucidate the notion from a different perspective. We quote Hayek several times in our book and find much that is of value in his view of complex systems. We also see certain parallels between Prigogine and Hayek as concerns their respective treatments of limitations to prediction. But where we have focused on the limitations that apply in physics, Hayek, like his predecessor Mises, focuses rather on limitations in economics and in economic planning (illustrated most impressively in Mises' calculation argument) (Steele 2013).

Martinelli defines a complex system as a system for which 'any synoptic description ... necessarily entails reference to at least one particular entity,' where by a synoptic description, he means a complete description (roughly) of all aspects of the entity in question. In other words, every complex system is in a sense *sui generis*. Simple systems can be copied *ad libitum*.

He sees, quite correctly, that systems are not autonomously demarcated portions of reality which would exist independently of human cognitive processes. Rather systems rest on selection and demarcation of the sort which is practiced in all cognitive endeavours. We can agree with his definition. But we would nonetheless argue that it is inferior to the definition we offer, which uses essential properties of the universals and mathematical entities (such as phase spaces) of thermodynamics in its formulation. Our definition is also stronger, because it draws on multiple aspects of complex systems, where the definition of Martinelli draws only on what Hayek filtered out of Prigogine's theory for his own purposes.

It does, though, allow Martinelli to make valuable points in his treatment of the intentions of agents. It is our immediate access to the pre-intentional world—above all through perception - that gives us the flexibility that we require to grapple with the complexity and unpredictability in our environments. And as Martinelli points out, this flexibility is lacking in the case of an artificial agent, which has only 'a mediated access to the pre-intentional world' which is characteristic also of collective agents (clubs, nations, armies).

But Martinelli goes wrong when he speaks of 'intentional states' of machines. Machines have no consciousness, and therefore no intentional states. Their states are logic states of the state machine implemented in the algorithms they run. These state machines define the set of states the machine can be in and how the state transitions occur depending on the parameters given by the process step they are engaged in. Nothing of this is comparable in any way with intentions of animate systems.

Martinelli goes wrong, too, in his statement that 'many complex systems can be approximated by simple systems'. To illustrate his point he uses the example of what the physician does when 'recognizing the symptoms of the seasonal flu on the inside of your own very particular throat.' Here approximation via simple system is indeed at work. But the sorts of simplifications required in physics to describe nature are often of a radical and counterintuitive nature, and principles of mathematics have to be pushed aside to make them work. The strength of physics is its ability to derive simplified models from nature that can be used to engineer logic systems. This is what drove the radical change of our environment to create the modern technosophere, in a development we have been witnessing since the mid-17th century. But these simplified models were hard won, and could be created only for a tiny sample of the massive numbers of complex systems by which we are surrounded.

Overall, it is nice to see that Hayek is hereby revived by Martinelli, but his view does not add anything to our arguments; it is a small subset of what we describe. Hayek, it has to be said here, is well received by non-postmodern contemporary philosophers because his philosophy, like Popper's, is metaphysically essentially as flat as Humean empiricism. Scheler, Gehlen and even Bergson provide much deeper material to tackle the problems we discuss. But their insights are rarely used by contemporary philosophers.

RAGNAR FJELLAND

Fjelland's contribution makes a very interesting point, namely that 'modern science did not completely liberate itself from metaphysics'. The situation we face is indeed even worse. Modern science is a form of metaphysics. The most important reason for this is that modern science relies on the mathematical modelling of the world and on the idea that modelling real entities using mathematical entities gives us information about the former. But does it really do so? This question belongs to the realm of pure metaphysical thinking, albeit only implicitly, since most scientists do not engage in considering questions such as this; they do not reflect on what they are doing.

As Fjelland emphasises, Husserl clearly saw how science remains within the realm of metaphysics, and we agree with him that our book falls very much within the tradition of Husserl. Husserl saw above all the important role played by Galileo in opening the door to the mathematics-based methodologies adopted by later generations while at the same time helping to close the door to metaphysical reflection (Husserl 1970).

But we are facing today another equally fundamental transition between Galileo, Newton and their successors on the one hand and quantum field theoreticians on the other. For the former model universals. That is, they develop theories of *types* of entities, such as physical bodies or orbits or velocities, theories which can then be applied to instances of these types in reality. The latter, in contrast, do not have universals at their disposal. Rather, they use in their place mathematical entities such as Hilbert spaces and operators on these spaces. The latter cannot be instantiated by any entities in the world, but acquire their connection with the world only through the measurement tuples which are the outputs of experimentation (Landgrebe and Smith 2023).

Consider, for example, spin. As Fjelland puts it: 'spin is really only a property that we can measure with complicated machines and which we model using quantum projectors [...]. We cannot imagine this property: we can only think of it as a -space projector'.

Fjelland addresses another important point, which critics of technology have often raised[2], namely that the use of technology can alter or damage the human mind. He is indeed right that the mass usage of digital systems and AI algorithms to take over basic tasks can lead to a decline in human cognitive skills and thereby also to a decline in the associated fields of endeavour. One of the best examples is map-based navigation, a skill that was mastered by every car driver a generation ago but is now rapidly being lost. One simple way of making this point is to say that the use of technology can make its users stupid.

There are three reasons why we do not cover this aspect of AI in our book. First, our book is about the question whether we can create intelligence. It is not about the assessment of the value of technology, a subdiscipline of sociology. A second reason why we steer clear of questions pertaining to the sociology of technology is that all technologies change human culture, and while some aspects get lost, such as the oral tradition, folklore, religious practices, and so on, others emerge. Viewing the impact of technology as mainly a negative factor for human culture is part of cultural criticism, an activity we prefer to leave to others.

Third, most human beings in all cultures exhibit sterotypical behaviours and are prone to adopting the perception and views of their peer group. The usage of technology just shifts there stereotypical behaviour from one cultural technique to another, for instance from the rosary to the iPhone. Yet humans do not become any more stupid as a result of shifts of this sort; they just change repetitive behaviour from one cultural stereotype to another. Therefore, the idea that technology usage harms the masses needs to be seen in the light of their relative lack of cognitive interests. From that perspective, it does not matter too much how they spend their time.

ROBERT WEST

West, sadly, is too much influenced by the on-going hype on behalf of AI, and so he seriously underappreciates the points we are making in our book. To begin with, he takes the marketing term 'generative AI' seriously. 'Generative' is an attribute describing the ability to create something truly new. But LLMs are not 'generative'; they merely produce an output that they were parametrised to compute. This output is the most likely continuation sequence of symbols given the input symbols and the parametrisation of the model.[3] Also, LLM-based chatbots cannot participate in conversations. They can certainly react to user input. But they cannot maintain a conversation that meets any psychologically demanding needs of human beings.

Let's briefly look at what a conversation is. It is a sequence of speech acts consisting of utterance-response-tuples that are pragmatically interrelated, i.e. which are contextualising each other and depending on each other. These tuples form the trace of a conversation, which is the sequence of utterances and responses of the conversation participants. After very few such tuples, there are more possible traces than stars in our galaxy; after a few more, there are more traces than stars in our universe. It is thus impossible to train a multivariate model to enable such conversations, the training material cannot be obtained.

His fifth reference shows that he does not understand that quantum computing (which is unlikely to ever develop beyond existing toy models) would not change anything in the field of AI at all, as we explain in the book. While the book was in press, one of the leading experts of quantum computation theory showed that the algorithms used in AI would gain little benefit in terms of speed-up from quantum computers, even should the latter ever be built, which seems unlikely (Dyakonov 2020; Aaronson 2022).

West then makes an even worse mistake by claiming that intelligence can be attributed to a computer based on its output and that today's outputs are getting close to showing this. No, they do not, as we show in great detail in the book. They just compute what they were parametrised to compute. West, like many other AI believers, uses a definition of intelligence (as 'solving complex problems', section 'Human and computer intelligence') which enables him to make the claim that machines might become intelligent. But even this definition does not really help, because he uses it merely to support the nebulous claim that machines will become intelligent by the interaction of man and machine (’What this suggests is that AGI may be best construed as more than just a set of algorithms; it emerges from the interaction between humans and those algorithms’).

In the book, we show in detail that a scenario of this sort is unrealisable. ‘AGI’ means artificial intelligence; not the sort of intelligence exhibited by human beings acting in partnership with machines of whatever sort. Such a partnership could indeed lead to the emergence of new kinds of machines. This is already happening with successive versions of LLMs. But this is a matter of humans engineering the models in question. The idea that machines might somehow evolve in and of themselves to create something new—something like AGI—is, once again, science fiction. Certainly the process could be nothing like the evolutionary process that gave rise to intelligent organisms. For this is a process in a complex system comprising a complex system of the huge number of complex systems, including intelligent organisms, that have interacted here on earth since life began. There is no coherent scenario under which computers might spontaneously develop an intelligence through a process of this sort. And nor, of course, is there any way in which we might engineer such a process, since we have no mathematically modellable idea how evolution works.

KIRILL KRINKIN

The main point made by Krinkin concerns what he calls ‘co-evolutionary hybrid intelligence’. He holds that such a thing is possible and might lead to new sorts of capabilities. There is a sense in which we agree to this thesis. Thus it is obvious that there are many ways in which humans obtain more possibilities when using technology. As stated above, technology is a major component of human culture. Machines will never become intelligent. But the evolution of technology continuously changes the world in which we live.

As will by now be clear, however, we reject the idea of machine evolution, because as we show in chapter 9 of the book, machines, which are logic systems, cannot be used to emulate an evolutionary process. Also, machines and humans do not fuse in any way. Rather, as in the case of a car driver or a computer user, humans use machines to achieve specific purposes. This has many psychological consequences for the individual and for society as a whole which are interesting to analyse, but such an analysis is not the purpose of the investigation conducted in our book.

Humans will continue to use machines. But there will be no fusion of man and machine, as we show in chapter 12 of the book, because it is mathematically impossible to design a seamless interface of mental processes and processes inside machines.

JANA SEDLAKOVA

Jana Sedlakova’s contribution, on ‘conversational AI’ (CAI) and its applications in psychotherapy, provides a number of insights which chime well with the arguments in our book.

As she points out, when a patient engages with a CAI, then the CAI will seem to be making expressive speech acts, for example making knowledge claims about the patient’s state of mind. The consequence, as she notes, might be that the user forms a relationship with the CAI, which might involve treating it as if it were an epistemic authority.

Rightly, however, Sedlakova points out that a CAI ‘is not part of the space of reason it is not undertaking and attributing commitments and entitlements. It is not social and normative. It also does not perform expressive speech acts.’

The problem she raises, concerns the fact that ‘human interlocutors will get used to the type of conversation with an AI that first was perceived as limiting and different from conversations with human interlocutors. With time, she says, it will become natural. ‘We can get used to talking to CAI and adapt our conversation style to it.’ None of this, as she admits, will change the fact that CAI does not have human capabilities. Yet still, CAI becomes in a sense a part of our discursive practices, because human discourse requires intersubjectivity. Machines in themselves are incapable of intersubjectivity because they do not feel or think and have no personality.

She holds that our arguments cannot account for what is happening here. But we have referred already in the foregoing to the phenomenon whereby the introduction of a new technology can have the effect of making humans—in some respects—more stupid. This is perhaps not what is going on here in every case. For even an intelligent person might engage with a CAI. We insist, however, that in such cases the person in question will be aware that she is dealing not with a conversation partner in the fullest (which is to say the human) sense, but rather with a machine that she can, for the moment, trick herself into having apparent conversations with.

MARIA M. HEDBLOM

The topic of creativity repeatedly played a role in many of our conversations in advance of completion of the manuscript of *Why Machines Will Never Rule the World*. In the end, however, we found the topic too difficult to enable the sort of conclusive treatment it would require to appear in the book. We are thus grateful to Maria Hedblom for her contribution which, as she puts it, zooms in on creativity as an aspect of intelligence that we neglected.

We respond by citing from a paper by Landgrebe derived of our book on the topic of AlphaFold, an impressive AI algorithm devoted to the problem of protein folding structure prediction (Landgrebe 2022). As Landgrebe points out, the dNN design methodology used in Alphafold (as also in ChatGPT and other LLMs) 'is like alchemy—input data and architecture are chosen in a heuristic manner using different patterns until a design emerges that can tackle the problem'. AlphaFold is built in addition around design decisions which show the nature of the cognitive style of the community to which its authors belong (what we might call mathematico-biology). Thus it uses simple mathematical (triangular) relationships between amino acid residues to design the system's representations of the transformations which these residues undergo during folding processes. Other operations built into the system by its designers are even less motivated by biology. The system underperforms when it comes to novel proteins, or in other words to proteins belonging to families in which no structure has as yet been identified by traditional (and very time-consuming) crystallographic methods. The system is in this sense non-creative. When it comes to proteins homologous to known structures however, its performance is excellent, illustrating once more the ability of dNNs to help in identifying regular, recurring patterns in reality. What the dNN models reveal here is the conservation of protein folding mechanisms in evolution—the evolution, namely, of molecules that we have learned about from crystallography experiments.

As Landgrebe summarizes:

> The AlphaFold project is an achievement at the very apex of human culture. It is a very impressive interdisciplinary effort which rests on decades of research and brings together biology, chemistry, x-ray crystallography, mathematics and computer science to create a new technology of considerable usefulness.

We would now propose to Hedblom that she should include in her theoretical knapsack a parallel diagnosis when she encounters putative examples of creativity on the part of dNNs or any other AI technology. There cannot be genuine creativity in machine learning output, because humans are responsible for that output just as much as it was humans who were responsible for building Cologne Cathedral. To see this using another algorithm from the same manufacturer, for example, to expert players of the game of GO. Their play was enhanced through the creativity of the mathematicians and engineers who designed the elegant reinforcement-learning based algorithm of the automaton.

NOTES

1 To see where he is going wrong, here, we recommend that Rapaport consult the 1,696 pages of Kandel et al. (2021), which summarises some 100 years of neuroscientific research—and contains almost no mathematics.

2 For an exhaustive overview see: (Landgrebe 2021).

3 If we imagine an AI algorithm is a long equation, then the process of identifying the parameters of this equation is called parametrisation.

REFERENCES

Aaronson, Scott. 2022. How Much Structure Is Needed for Huge Quantum Speedups? *arXiv:2209.06930*.

Arcas, Blaise Agüera y and Peter Norvig. 2023. Artificial General Intelligence Is Already Here. *Noema Magazine*.

Bostrom, Nick. 2003. *Superintelligence. Paths, Dangers, Strategies*. London: Oxford University Press.

Dyakonov, Mikhail I. 2020. *Will We Ever Have a Quantum Computer?* Cham: Springer.

Fjelland, R. 2023. Computers will not acquire general intelligence, but may still rule the world. *Cosmos + Taxis* 12:5+6.

Gehlen, Arnold. 1988. *Man: His Nature and Place in the World*. New York: Columbia University Press.

Harari, Yuval Noah. 2016. *Homo Deus: A Brief History of Tomorrow*. London: Random House.

Husserl, Edmund. 1970. *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. Evanston: Northwestern University Press.

Hedblom, M. M. 2023. Every dog has its day: An in-depth analysis of the creative artistic ability of generative AI. *Cosmos + Taxis* 12:5+6.

Ingarden, Roman. 1973. *The Literary Work of Art. Investigations on the Borderlines of Ontology, Logic and the Theory of Literature*. Evanston: Northwestern University Press.

Kandel, Eric, John D. Koester, Sarah H. Mack and Steven Siegelbaum. 2021. *Principles of Neural Science*. New York: McGraw Hill.

Krinkin, K. 2023. Back to evolutionary intelligence. *Cosmos + Taxis* 12:5+6.

Landgrebe, Jobst. 2021. Technikkritik Im Zeitalter Der Technosphäre. In: *Coram Deo Versus Homo Deus: Christliche Humanität Statt Selbstvergottung*, ed. Thomas Seidel and Sebastian Kleinschmidt, pp. 209-34. Leipzig: Evangelisches Verlagshaus.

________. 2022. What AlphaFold Teaches Us about Deep Learning with Prior Knowledge. https://doi.org/10.21203/rs.3.rs-1582914/v1.

Landgrebe, Jobst and Barry Smith. 2023. Ontologies of Common Sense, Physics and Mathematics. https://arxiv.org/abs/2305.01560.

Martinelli, E. 2023. Complexity and Particularity: An Argument for the Impossibility of Artificial Intelligence. *Cosmos + Taxis* 12:5+6.

Rapaport, W. J. 2023. Is Artificial General Intelligence Impossible? *Cosmos + Taxis* 12:5+6.

Scheler, Max. 2009. *The Human Place in the Cosmos*. Evanston: Northwestern University Press.

Schulz, S. and Hastings, J. 2023. What is a machine? Exploring the meaning of 'artificial' in 'artificial intelligence'. *Cosmos + Taxis* 12:5+6.

Sedlakova, J. 2023. Conversational AI for Psychotherapy and Its Role in the Space of Reason. *Cosmos + Taxis* 12:5+6.

Simon, J. A. 2023. Is Intelligence Non-Computational Dynamical Coupling? *Cosmos + Taxis* 12:5+6.

Steele, David Ramsay. 2013. *From Marx to Mises: Post Capitalist Society and the Challenge of Economic Calculation*. LaSalle: Open Court.

Thurner, Stefan, Peter Klimek and Rudolf Hanel. 2018. *Introduction to the Theory of Complex Systems*. Oxford: Oxford University Press.

West, R. 2023. Semi-autonomous Godlike Artificial Intelligence (SAGAI) is conceivable but how far will it resemble Kali or Thor? *Cosmos + Taxis* 12:5+6.