Miklós Hoffmann

Science as a Human Vocation and the Limitations of AI-Based Scientific Discovery*

Abstract

In his essay *Science as a Vocation*, Max Weber took the essence of scientific activities to consist in specialisation and enthusiasm. His arguments, together with works by Michael Polanyi (Mihály Polányi) and others, are explored and compared with recent results and expectations of automatised, artificial-intelligence-driven scientific discovery. Our aim is to show that artificial intelligence systems (AI systems) – while they can evidently and effectively support everyday scientific activities as useful tools – are not, in themselves, able to produce genuine invention, are not suitable for breakthrough scientific discovery. And this limitation, we argue, is due to AI systems' inability for specialisation and their lack of enthusiasm. Our observation is that while selection by intrinsic interest is unavoidable and an essential part of science, this interest is unquantifiable and unmetrisable by an objective function, therefore cannot be learned by an AI system. We conclude that being a scientist full of passion and with the ability of selection remains humans' intellectual privilege.

Keywords: scientific discovery, artificial intelligence, invention, enthusiasm

I. INTRODUCTION

In February 1996 the reigning world chess champion Garry Kasparov was defeated by the IBM computer Deep Blue. Although Deep Blue controlled the white pieces, and right after that game Kasparov won the next game, and was the overall winner of the six-game chess match (Kasparov versus Deep Blue: 4–2), this date still marks an unusual technological achievement: this was the first win of an artificial-intelligence-driven system (AI system) over the highest-ranked human specialist in a specific field of expertise. One year later Deep Blue defeated Kasparov 3.5–2.5 in another six-game chess match.

* This research was supported by the grant EFOP-3.6.1-16-2016-00001 (Complex improvement of research capacities and services at Eszterházy Károly University). The author would like to thank the anonymous reviewers for their thoughtful comments on this work.

Almost exactly 20 years later, in March 2016, Lee Sedol, one of the greatest players of Go, a highly complex strategy board game popular in East-Asia, was defeated by Google's AlphaGo software. Go has far more variations than chess, and strategies are more complicated (Bouzy 2001), therefore this win is another important milestone in the development of artificial intelligence. The event was selected as one of the scientific breakthroughs of the year 2016 by Science Magazine.[1] And the momentum was unstoppable: in less than one year, AI software defeated over 100% of the best poker players in several poker tournaments. Why are these developments so interesting? While chess and Go are called games with complete information – that is, players possess full information about their opponents and their potential (straightforward or surprising) actions – poker is clearly a game with incomplete information. The possibility of bluffing makes a poker game somewhat independent of the consequences of previous steps, it liberates players from the restrictions of logic, therefore opponents need to study not only combinations and strategies – beside learning game rules, computers must also learn the behaviour and attitude of other players (Moravčík 2017).

Besides board and card games, it was a natural next step to compare humans and AI systems in other fields as well. In this paper we intend to study how the rapid development of AI may impact on human scientific activity, science as a vocation: more specifically, the potential automatisation and algorithmisation of scientific discovery.[2] Some are sceptical about such impact. Others are warning the sceptic: it may be prudent to reassess doubts given that Kasparov and Sedol were also antecedently doubtful as to whether an AI system could beat them. For example, Nobel laureate scientist Wilczek (2016) (among others) strongly believes that scientific discovery will soon be fully automatised. According to Wilczek and other scientists (see e.g. Kitano 2016) it is a realistic scenario that an AI system will be the best physicist and will be able to win the Nobel Prize in the near future.

Attempts to find scientific achievements through automated discovery have a long history during which tools and concepts have significantly evolved. In this paper the label "AI system" is taken to encompass earlier serial-computation approaches as well as more recent machine learning approaches, genetic algorithm based methods, or their fusion (for an overview of these methods and their history see Alai 2004). We will call all these AI systems without differentiating among them, because we believe that there are two significant common attributes to all of these methods: their data-driven approach and their algorithmic, procedural nature. Regardless of the method and tool, artificial intelligence requires data – a large amount of training data in which it can find typical patterns

[1] Besides gravitational waves and customised proteins, see *Science* 354. 1518–1523.

[2] Throughout the paper notions like "algorithm" and "computer" are used in the usual manner for software and hardware tools developed for determined computation executed in a finite number of steps. However, we note here that illuminating discussion about these notions is now under way in the literature, see e.g. Rapaport 2018.

and correlations. And this is done by a procedure, an algorithm, even in the case of the most sophisticated neural network methods.

One may ask whether scientific discovery or even the scientific description of the world can have a substantially different path than what we have experienced throughout the history of science. We cannot answer this question here, but the fact remains that no alternative approach has been envisioned so far: all the attempts at automatised scientific discovery follow our classical path and a potential new, uncharted path may well diverge considerably from what we now call science and knowledge. Nevertheless, our discussion remains in the classical framework: we consider scientific discovery and science as an enterprise whose results were, over many centuries, produced by human scientists.[3]

In this paper we intend to point out those substantial aspects of scientific discovery that make the personal involvement of human scientists inevitable, and consequently make the replacement of scientists by computer algorithms and artificial intelligence in the scientific process highly doubtful. Our arguments will extensively rely on Max Weber's stance, who saw the essence of scientific activities in specialisation and enthusiasm (Weber 1946). These key notions will be analysed in our study from the perspective of AI-driven scientific discovery. We aim to show that AI systems – while they can evidently and effectively support everyday scientific activities as useful tools – are not, in themselves, able to produce genuine invention, are not suitable for breakthrough scientific discovery. And this limitation, we argue, is due to AI systems' inability for specialisation and their lack of enthusiasm.[4]

One may think that specialisation cannot be an obstacle to AI in terms of automatised scientific discovery: for the computational and learning capacity of these algorithms can easily be focused on an arbitrary narrow field. However, as we will show, from a theoretical point of view, the specialisation requirement yields an insurmountable problem for artificial intelligence. Enthusiasm, as we will also point out, raises an even more difficult issue.

We put special emphasis on the enthusiasm-filled moment that anticipates scientific work. Max Weber writes about this moment as follows:

> Yet it is a fact that no amount of such enthusiasm, however sincere and profound it may be, can compel a problem to yield scientific results. Certainly enthusiasm is a prerequisite of the "inspiration" which is decisive. Nowadays in circles of youth there is a widespread notion that science has become a problem in calculation, fabricated

[3] This view also gives credibility to the thoughts of scientists from past centuries about science and scientific discovery, even if automatised scientific discovery was not an issue, or it was technically less developed in their time.

[4] From a Kuhnian perspective: artificial intelligence is able to support "normal science" through day-to-day experimental studies, but it cannot discover results forcing a paradigm shift.

> in laboratories or statistical filing systems just as "in a factory", a calculation involving only the cool intellect and not one's "heart and soul". First of all one must say that such comments lack all clarity about what goes on in a factory or in a laboratory. In both some idea has to occur to someone's mind, and it has to be a correct idea, if one is to accomplish anything worthwhile. And such intuition cannot be forced. It has nothing to do with any cold calculation. (Weber 1946. 135)

This – in our view, essential – moment, the birth of the first idea, the exciting promise of the discovery, the moment of entering the force field of the problem, I will call – applying a physical metaphor – *the gravity of invention*.

II. AI-DRIVEN SCIENTIFIC DISCOVERY – INABILITY FOR SPECIALISATION

The first research result about an AI system engaging in scientific discovery was published by Pat Langley and his colleagues (Langley et al. 1987). In this study an AI system was programmed by the research team to explore new scientific results based on a data set. In their groundbreaking study the most interesting aspect is the history-oriented approach, which, to some extent, already predisposes it towards verifying a preconceived outcome: during the training period, data fed into the AI system was selected from a certain historical period of science. Physical and chemical observations and laws known around the 17th and 18th centuries were learned by the system. Based on these data, the AI system "discovered" now well-known, but at-the-time new scientific results such as Ohm's law, Kepler's third law of planetary motion, and various chemical reactions.

However, besides these apparently successful outcomes the computer also "discovered" superseded scientific theories such as the phlogiston theory mistakenly put forth to explain oxidation. Moreover, other outcomes were true but totally uninteresting from a scientific point of view. Note here that those results, such as Kepler's law, discovered by the AI system, can be deduced (and in fact have been subsequently discovered by Kepler) by systematically tracking the available observational data over a long period of time. In other words, systematic computational work on observational data can readily lead us to this discovery. The phrase "systematic" is used here as the opposite of "heuristic", following a distinction drawn by Michael Polanyi:

> The difference between the two kinds of problem solving, the systematic and the heuristic, reappears in the fact that while a systematic operation is a wholly deliberate act, a heuristic process is a combination of active and passive stages. A deliberate heuristic activity is performed during the stage of Preparation. If this is followed by a period of Incubation, nothing is done and nothing happens on the level of consciousness during this

> time. The advent of a happy thought (whether following immediately from Preparation or only after an interval of Incubation) is the fruit of the investigator's earlier efforts, but not in itself an action on his part; it just happens to him. And again, the testing of the "happy thought" by a former process of Verification is another deliberate action of the investigator. Even so, the decisive act of discovery must have occurred before this, at the moment when the happy thought emerged. (Polanyi 1974. 134)

A scientific discovery is called systematic if the final result is reached by a series of intentional, algorithm-based steps, even if these steps are very complicated. By contrast, the discovery is heuristic if – beside the above mentioned steps – it is based on one or more unanticipated, unenforceable moments, which cannot be explained as a simple logical consequence of preceding steps. These are the moments of Weberian inspiration, the moment when a – perhaps brilliant – thought arises. For example, contrary to Kepler's law, the thought of the heliocentric system by Copernicus cannot be the outcome of a systematic discovery, since observational data available given that era's level of accuracy provided stronger support for the Ptolemaic system. Analogously, the theory of general relativity by Einstein cannot be algorithmically derived from the observational data of that age – it was experimentally proven only decades after the publication. Since every result the AI system can produce is inherently based on the analysis of available observational data, it can yield systematic scientific discovery, but we claim that brilliant heuristic moments and thoughts lie outside the repertoire of an AI system.

One may think that even if we cannot expect from AI systems groundbreaking discoveries in the natural sciences or mathematics (discoveries that require the power of a compelling paradigm change), many useful and interesting results in a specialised narrow subfield may still be gleaned by an AI system. And this leads us to the question of specialisation, whose importance was also emphasised by Weber. But specialisation certainly involves selection: scientists have to select among topics, within the given topic they have to select among related theorems, laws, data which are to be learned, improved or further developed. Moreover, one even has to select among the potentially solvable problems and among the provable theorems. Selection is unavoidable due to our limited resources, but there is an even more important aspect: the intrinsic interest of the problem. It is worth citing Michael Polanyi again on this:

> An affirmation will be acceptable as part of science, and will be the more valuable to science, the more it possesses: (1) certainty (accuracy) (2) systematic relevance (profundity) (3) intrinsic interest. (Polanyi 1974. 143)

While (1) and (2) sound natural requirements in the realm of scientific inquiry, (3) is a property that is difficult to make precise, yet it is of central importance. We clearly have no exact tools or algorithms or conditions to evaluate effectively

the level of interest of a scientific statement. No one can assess based on exact criteria what theorem or law is more interesting (or will be in the future) than another statement of physics, chemistry or mathematics. Having said that, selection by intrinsic interest looks not only unavoidable, but also essential. It is evident that our (human or articifical) intellectual capacity is restricted in terms of time and computational power, therefore it is highly beneficial to focus this capacity on problems which may yield higher "gains", and can improve our scientific knowledge in a more effective way. The higher the intrinsic interest of a problem, the stronger its gravity of invention. Stronger gravity can also affect, influence more scientists. We provide some examples for such an interest arising among mathematicians because – compared to the natural sciences – mathematics is a field where scientists can formulate new valid statements in a relatively easy way, thus in relatively large numbers.

Since mathematics is a cumulative, aggregate field of science, whenever a statement is correctly proved, it will be part of mathematics forever. The so-called Ulam's dilemma (Ulam 1976) describes the ever-more-complex situation as follows: in mathematics (and partly in theoretical physics) we have discovered so many theorems, and scientists extend this list daily by such a vast amount of valid statements, that nobody is able to overview the entire field, only some sufficiently small subfield.[5] The only solution to this dilemma is specialisation, also encouraged by Weber. Specialisation means selection: selection among theorems, among subfields, among problems. This selection, however, is not a drawback, not a restriction, not a systemic limitation, contrary to how one may view it at first glance. Selection is the essence of scientific discovery. It is worth citing here one of the greatest mathematicians of the 19th and 20th centuries, Henri Poincaré:[6]

> What, in fact, is mathematical discovery? It does not consist in making new combinations with mathematical entities that are already known. That can be done by anyone [*even a computer – M.H.*], and the combinations that could be so formed would be infinite in number, and the greater part of them would be absolutely devoid of interest. Discovery consists precisely in not constructing useless combination, but in constructing those that are useful, which are an infinitely small minority. Discovery is discerment, selection. (Poincaré 2009. 50)

[5] In his book, Stanislaw Ulam estimated the number of yearly published mathematical theorems around 200 000 – and this number evidently further increased (probably exponentially) in recent decades.

[6] In the original version: "Qu'est-ce, en effet, que l'invention mathématique? Elle ne consiste pas à faire de nouvelles combinaisons avec des êtres mathématiques déjà connus. Cela, n'importe qui pourrait le faire, mais les combinaisons que l'on pourrait former ainsi seraient en nombre infini, et le plus grand nombre serait absolument dépourvu d'intérêt. Inventer, cela consiste précisément à ne pas construire les combinaisons inutiles et à construire celles qui sont utiles et qui ne sont qu'une intime minorité. Inventer, c'est discerner, c'est choisir." (Poincaré 1912. 48)

Invention thus practically amounts to selection when done well, and well in time. But such selection cannot be algorithmisable, since it is not a mechanically scientific, but rather a meta-mathematical selection. If we start from an axiomatic system, say, the Peano-axioms of natural numbers, then human as well as artificial intelligence can prove many valid statements, for example, that there is an infinite number of primes; and can falsify many other untrue statements, such as there is no even prime. Moreover, artificial intelligence can evidently "produce" many more valid theorems and can falsify many more untrue statements in a given period of time than human scientists can. However, as Karl Popper[7] (1950) also points out, computers have no instruments or algorithms to draw a distinction between what are – in our view – interesting, thought-provoking, ingenious statements and statements which are totally uninteresting (although true). A very simple, yet convincing example of Popper's can further illuminate this problem and make it plausible: besides the statement $2 + 1 = 3$, a computer will find infinitely many statements like $2 + 1 \neq 4$; $2 + 1 \neq 5$… and further statements like $2 + 1 \neq 3 + 1$; $2 + 1 \neq 4 + 1$, all arrived at based on the same set of starting axioms. For each substantial, interesting statement an AI system systematically generates infinitely many uninteresting yet valid statements.[8] Overall, the probability of observing the few promising ideas worth further investigation among the many-many uninteresting statements by the computer is very close to zero.

III. AI-DRIVEN SCIENTIFIC DISCOVERY – LACK OF ENTHUSIASM

As we have already mentioned, besides the ability, instinct and delight of specialisation, Max Weber has seen the substance of scientific activities in enthusiasm. What does enthusiasm – or lack thereof – mean in terms of science as a vocation? When engagement with a problem is externally driven (a typical example for most of us is solving a task provided by the teacher in a mathematics class) then one can feel the sense of duty or competition, the wish to surmount the hurdles related to the problem, but the extrinsic nature of motivation deprives us of feeling passion and enthusiasm. By contrast, if the motivation for solving the problem comes from an intrinsic interest, if an unforced and unforce-

[7] "A calculator may be able, for example, to produce proofs of mathematical theorems. It may distinguish theorems from nontheorems, true statements from false statements. But it will not distinguish ingenious proofs and interesting theorems from dull and uninteresting ones. It will thus 'know' too much — far too much that is without any interest." (Popper 1950. 194)

[8] Although it is not well defined what we mean by "interesting" and "uninteresting" results, mathematicians have a surprisingly well-functioning common intuition in judging the value of propositions. Overall this leads to the question of (un)metrisability of scientific interest, which we will discuss in the last section.

able seed of idea emerged in our head, then we will engage this problem with personal commitment, passion and enthusiasm.

In his famous book *Proofs and Refutations*, Imre Lakatos (1976) studied and demonstrated through several examples how a (mathematical) problem and invention may arise, among which here we briefly refer to one typical scenario.

Suppose that – as a beginner in maths – we study divisibility of numbers and we observe that every number whose last digit is 2 (such as 12, 22, 32 etc.) is divisible by 2. Meanwhile numbers whose last digit is 3 are not always divisible by 3 (for example 63 is divisible by 3, but 13 is not divisible). We find it interesting that there are numbers analogous to 2, for example numbers whose last digit(s) is 5 (or 10 or 25) are always divisible by 5 (or 10 or 25), call these last-divisible numbers. Meanwhile we find several numbers in the other class as well: for example 24 is divisible by 4, but 14 is not. Now we are right in the middle of the field of gravity of the problem, the gravity of invention, and the data we collected (last-divisible examples are 2, 5, 10, 25, 50, 100…) make the heuristic idea clear: all the last-divisible numbers are products of powers of 2 and 5, possibly including powers with exponent 0 (note, however, that not all numbers that are such products are last-divisible, for example, 5^4 isn't while 5^3 is). This is far from a rigorous proof, but the rest is simply a mechanical computation for formulating and justifying the precise statement.[9]

In the example described above and in many other examples Lakatos has presented (in a much more detailed form in his book), he describes the athmosphere of raising a problem and finding a heuristic solution. Here we intend to focus on one important aspect of this process: assuming an underlying principle in the collected examples and counterexamples, based on which one can heuristically create a conjecture is of utmost importance. The key moment is the perception of the first couple of aspects of the pattern, the excitement of foreseeing the potential existence of some (ir)regularity. This excitement is not about the foreseen result, but about the promise of an interesting result. It is the gravity of invention, the engagement of the scientist in the field of gravity of the problem. The first perception about the number 2 is not specifically exciting, but the moment of understanding that another number (3) works differently than 2 may put our mathematical thinking in action. Anticipating the promise of success, we try to find new examples and counterexamples. Finding these data it can happen that the problem turns out to be too simple, too trivial, or uninteresting. But it can also happen that the intrinsic interest of the problem drives us into a new field and activates our heuristic problem-solving abilities.

[9] The theorem in its final form is as follows: a number n is last-divisible if and only if $n = 2^p 5^q n = 2^p 5^q$, where $p, q \geq 0 p, q \geq 0$ and $0 \leq q - p + 1 \leq 40 \leq q - p + 1 \leq 4$.

Note that the first moment, the promise of a future fruitful cogitation is, in fact, not part of the heuristic problem-solving process. It is a "preheuristic" flash, yet it is essential in terms of mathematical discovery and in general in scientific vocation.

The start of gravity of invention, the passion of thinking, the way how the scientist is getting engaged by the problem, is unexplainable, unenforceable, and, more importantly, unpredictable. It cannot be foreseen, cannot be measured (as we will see soon in some detail). And overall this foreshadows that an AI system, evidently driven by external forces (i.e. programmers) when studying chess, making stock market transactions or proving geometric theorems, cannot be programmed for this central passion, for the enthusiasm towards science. Some aspects of this discussion ultimately lead us to the most fundamental questions of artificial intelligence, notably issues having to do with what it takes for a system to have mental states, and more specifically, mental states of the sort that can underpin goals, motivation. These questions are beyond the scope of this paper, but even if we suppose that future AI systems can have mental states, these states (and the change of them) are outcomes of a causal process, externally driven (by the programmer and partly by the input data). Therefore it is entirely unclear if and how enthusiasm and passion of thinking is achievable by artifical intelligence systems.

There is no doubt that computers, without any passion or enthusiasm, can indeed find interesting results in some fields of science. For example, the AI-driven computer called "Eve" has been searching and finding effective pharmaceutical components, carrying out a vast number of trials (see King 2018). A further recent example for this type of discovery is from the field of material science, reported by Tshitoyan (2019) and Kauwe (2020). AI systems can discover new materials or new properties of old materials, but only following the typical patterns of an extremely large training data set of information. However, if we are seeking to discover something atypical, a kind of material which achieves its extraordinary properties by leveraging a new mechanism that is not common in the training data set, it will be unlikely to identify it through AI-driven discovery (c.f. Kauwe 2020). Atypical discovery always needs heuristic impulse and vision. The following sentences from Michael Polanyi clearly demarcate the barriers:

> The heuristic impulse links our appreciation of scientific value to a vision of reality, which serves as a guide to enquiry. Heuristic passion is also the mainspring of originality. (Polanyi 1974. 169)

Automatic scientific discovery without enthusiasm can only happen through a great number of trials, through following or finding typical patterns. Without heuristic impulse there is no chance of realising and evaluating the potential value of a future discovery which may come from a certain direction of research.

Even if a significant discovery is found by an AI system, it is not necessarily able to realise its importance.

The scientist is not cold and unemotional during research, not even at the beginning, at the preheuristic moment of involvement. As, following Heidegger and Gadamer, István Fehér M. (2017. 15) formulated this succintly, the scientist always has a – positively considered – prejudice:[10]

> …with regard to the type of interpretation that is directed at texts, in most cases it is illusory to refer to what "stands there" in the text as decisive evidence. For what is first and foremost "there" – provided there is any sense in speaking of "standing there" – is not so much the text itself, but rather "the self-evident, undisputed preliminary prejudice [*Vormeinung*] of the interpreter".

We can extend this approach to non-textual (visual or machine-based) sources of scientific information as well: there is no decisive, original meaning of pictures, figures, graphs, equations, data, and software output. What exists is a meaning interpreted by the scientist who studies that source, and this meaning is filtered and fertilised through the preliminary prejudice [*Vormeinung*], and positively considered preconception [*Vorurteil*] of the scientist. An AI system does not possess and cannot be equipped with such a preconception and prejudice: computers can treat only the information "standing there" technically or syntatically without any relationship to the source of information, without any preliminary opinion, because these are all beyond (or rather before) the pure binary information. It is as yet entirely unclear how to equip an AI system with more about the subject matter of investigation than the pure binary data we have provided. Let's compare this to Polanyi's words:

> To see a problem is to see something hidden that may yet be accessible. The knowledge of a problem is, therefore, like the knowing of unspecifiables, a knowing of more than you can tell. But our awareness of unspecifiable things, whether of particulars or of the coherence of particulars, is intensified here to an exciting intimation of their hidden presence. It is an engrossing possession of incipient knowledge which passionately strives to validate itself. Such is the heuristic power of a problem. (Polanyi 1961. 466)

Note that knowing the problem is not the same as knowing the solution to the problem. To know problems, or even to feel problems requires the recognition of their hidden presence, and the excitement of this recognition, the possibility of invention is gravitating us to the search for a solution to the problem and to the application of heuristics. The latter, that is to say, our attempt to solve

[10] In this paragraph the author refers to Heidegger (1962. 141).

the problem, may or may not succeed, but the gravitational attraction already mentioned will trigger the process. Mathematics uses one concise word for all of this: conjecture. The mathematical conjecture is preconceived knowledge, prejudice, similar to what is discussed in Gadamer's legal example (Gadamer 2004. 194) as a preconceived judgement: something I think about the thing before I know the thing, which can be verified or falsified by subsequent careful examination.

The AI system tries to grasp the problem without prejudice, with the question "what is that?," while we begin to engage the problem because it already means something to us, so our question (according to Nietzsche) is: "what is that for me?".[11] With the question "what is that?" the computer searches for an objective constitution, an absolute meaning in all data. When evidently expecting two computers to analyze the same data to produce the same result, we actually discover and demonstrate limitations to artificial intelligence. Let's quote Gadamer again:

> The paradox that is true of all traditionary material, namely of being one and the same and yet of being different, proves that all interpretation is, in fact, speculative. Hence hermeneutics has to see through the dogmatism of a "meaning-in-itself" in exactly the same way critical philosophy has seen through the dogmatism of experience. (Gadamer 2004. 507)

The question "what is that for me?" can be answered differently even if two of us look at the same text, same data, same picture. Moreover, here, in addition to our semantic relation, the expression "means something to me" as it is used in everyday language, carries an emotional charge, and this emotional charge is the passion. Artificial intelligence is an attempt to realise the "meaning-in-itself" in the modern age, trying to ignore personal enthusiasm and passion. But passion-free invention cannot exist, and, for now this seems to remain a lasting if not eternal barrier to artificial intelligence.

[11] See Nietzsche (1968. 301): "A 'thing-in-itself' just as perverse as a 'sense-in-itself', a 'meaning-in-itself'. There are no 'facts-in-themselves', for a sense must always be projected into them before there can be 'facts'. The question 'what is that?' is an imposition of meaning from some other viewpoint. 'Essence', the 'essential nature', is something perspectival and already presupposes a multiplicity. At the bottom of it there always lies 'what is that for me?' (for us, for all that lives, etc.)".

IV. SCIENTIFIC INTEREST IS UNQUANTIFIABLE

Finally, let us examine the reason for the apparent contradiction that while AI systems can beat top-ranked minds in the mental sports and games mentioned in the introduction, they cannot produce groundbreaking novelties in the field of scientific discovery.

In chess, various calculation methods that assign numerical values to each piece and each step or position are well known (for example, according to classical piece value calculation, 1 unit is assigned to pawn, 3 to bishop, 5 to rook, 9 to queen). The purpose of the computer is to find a step that optimises the cumulative value of the current position. It can be done by examining an easy-to-construct mathematical tool, the so-called *objective function*, and to find its optimal value. A similar objective function – or in multi-criterion decision models, functions – can be defined in other games and in very different areas of the application of artificial intelligence as well, such as automated stock trading, where the obvious objective function is the amount of profit.

The objective function (or functions) must clearly quantify which of the two situations or states is more valuable, that is, when we make a better choice, to which direction belongs more *utility*. However, in the light of the above mentioned problems, it seems that such an objective function cannot be defined in scientific discovery.

It is worth mentioning here the notion of utility measured by a given objective function (a certain type of objective function is also called a utility function), because when we apply it to science, to scientific discovery, it brings to mind Kant's classical discussion of the conflict of faculties:[12]

> …*truth* (the essential and first condition of learning in general) is the main thing, whereas the *utility*… is of secondary importance (Kant 1992. 7).

Of course, this allows that what is useful may be untrue (and vice versa); meanwhile, usefulness and utility cannot be the primary guiding principle for a theoretical researcher. As Mihály Vajda expresses in his commentary on Kant's work above (Vajda 2016):

[12] This text is especially relevant to our topic because we may well assume that Kant, had it existed at that time, would have classified the Information Technology (or simply IT) faculty as one of the higher-utility faculties, in contrast to the lower faculties such as Philosophy (and Mathematics), where the guiding principles are pure erudition, free choice of subject, and critical approach.

> I would like to hope that the university will continue to train not only smart professionals but also a group of people who are carriers of something which is *per se* unreasonable, because it is useless… What can be contrasted with utility is a kind of irrationality: a world where the useless – beauty and tranquility – (also) reigns.

Scientists may choose a direction which is (at a given moment) seemingly useless, if they find this direction interesting. Moreover, accoding to Vajda's commentary, this – perhaps unreasonable – moment showcases the freedom and beauty of pure science.

Are we able to algorithmise and measure the motivation behind the drivers of what may seem like an unreasonable, useless choice? Recent scientific experiments show that we are powerless in this matter. Let us examine what happens when someone tries to define an "objective function of intrinsic interest", that is a metric stimulating and measuring the curiosity of an AI system based on the novelty of information or the amount of information obtained per unit of time.

In a recent experimental study (Burda 2018), authors found that the most interesting series of events (that is to say: the series providing the most interesting information defined by the objective function of intrinsic interest) to the AI system as it "watches" television is the continuous, instantaneous switching of channels, or even the black-and-white noise of the television (when there is no broadcasting), since the information per pixel changes the most in these cases. If, ironically, the computer was playing computer games to arouse its interest, it was observed that the computer – after several wins – sometimes "intentionally" lost the game in order to see "GAME OVER" which was rarely shown, so it was interesting new information according to the objective function. The irrationality here also appears, but the algorithmic uselessness is a dead end.

The above anomalies also prove that we cannot as yet properly allocate value to the interest of a process, situation, or even to the interest of an unknown scientific claim. While we can measure their information content in a technical and syntactic sense, we are unable to mathematise its semantic aspects, value, and intrinsic interest. Between stacked syllogisms, we are technically unable to set up a scale or order of values that can be automatically calculated and verified. All this eventually results in the AI system being able to function as a "smart professional" in the sense of Vajda, but – in the absence of a proper objective function – the system will be incapable of decision and choice, in the sense of Poincaré. Thus it remains entirely unclear if and how fully automatised scientific discovery could be carried out. Consequently, being a scientist full of passion and with the ability of selection remains humans' intellectual privilege.

REFERENCES

Alai, Mario 2004. AI, Scientific Discovery and Realism. *Minds and Machines*. 14. 21–42.
Bouzy, Bruno – Tristan Cazenave 2001. Computer Go: An AI Oriented Survey. *Artificial Intelligence*. 132. 39–103.
Burda, Yuri – Harri Edwards – Deepak Pathak – Amos Storkey – Trevor Darrell – Alexei Efros 2018. Large-Scale Study of Curiosity-Driven Learning. *arXiv:1808.04355*
https://arxiv.org/abs/1808.04355
Fehér M., István 2017. Prejudice as a Precondition for Understanding. *Hungarian Philosophical Review*. 61. 9–28 (in Hungarian).
Gadamer, Hans-Georg 2004. *Truth and Method*. Transl. by Joel Weinsheimer and Donald Marshall. London – New York/NY, Continuum Publishing.
Heidegger, Martin 1962. *Being and Time*. Transl. by John Macquarrie and Edward Robinson. London, SCM Press.
Kant, Immanuel 1992. *The Conflict of the Faculties*. Transl. by Mary J. Gregor. Lincoln, University of Nebraska Press.
Kauwe, Steven K. – Jake Graser – Ryan Murdock – Taylor D. Sparks 2020. Can Machine Learning Find Extraordinary Materials? *Computational Materials Science*. 174. 1–7.
King, Ross D. – Vlad S. Costa – Chris Mellingwood – Larisa N. Soldatova 2018. Automating Sciences: Philosophical and Social Dimensions. *IEEE Technology and Society Magazine*. 37. 40–46.
Kitano, Hiroaki 2016. Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery. *AI Magazine*. 16. 39–49.
Lakatos, Imre 1976. *Proofs and Refutations*. Cambridge, Cambridge University Press.
Langley, Pat – Herbert A. Simon – Gary L. Bradshaw – Jan Zytkow 1987. *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge/MA, The MIT Press.
Moravčík, Matej – Martin Schmid – Neil Burch –Viliam Lisý – Dustin Morrill – Nolan Bard – Trevor Davis – Kevin Waugh – Michael Johanson – Michael Bowling 2017. Deepstack: Expert-Level Artificial Intelligence in Heads-Up No-Limit Poker. *Science*. 356. 508–513.
Nietzsche, Friedrich 1968. *The Will of Power*. Transl. by Walter Kaufmann and John R. Hollingdale. New York, Vintage Books.
Poincaré, Henri 1912. *Science et Méthode*. Paris, Flammarion.
Poincaré, Henri 2009. *Science and Method*. Transl. by Francis Maitland. New York, Cosimo.
Polanyi, Michael 1961. Knowing and Being. *Mind*. 70. 458–470.
Polanyi, Michael 1974. *Personal Knowledge*. London, Routledge.
Popper, Karl R. 1950. Indeterminism in Quantum Physics and in Classical Physics. Part II. *The British Journal for the Philosophy of Science*. 1. 173–195.
Rapaport, William J. 2018. What is a Computer? A Survey. *Minds and Machines*. 28. 385–426.
Tshitoyan, Vahe – John Dagdelen – Leigh Weston – Alexander Dunn – Ziqin Rong – Olga Kononova – Kristin Persson – Gerbrand Ceder – Anubhav Jain 2019. Unsupervised Word Embeddings Capture Latent Knowledge from Materials Science Literature. *Nature*. 571. 95–98.
Ulam, Stanislaw M. 1976. *Adventures of a Mathematician*. New York, Charles Scribner's Sons.
Vajda, Mihály 2016. The Conflict of the Faculties. *Magyar Tudomány*. 177. 410–416 (in Hungarian).
Weber, Max 1946. Science as a Vocation. In Max Weber: *Essays in Sociology*. Transl. and ed. by Hans H. Gerth and Wright C. Mills. New York, Oxford University Press. 129–156.
Wilczek, Frank 2016. *Fantastic Realities: 49 Mind Journeys and a Trip to Stockholm*. Singapore, World Scientific.