Edward J. Alam

From Aristotle to AI and Back Again: In Praise of Responsibility and Irrationality

In an attempt to achieve some unity and continuity between this presentation and the one I had the honour of delivering here last February, titled *Friendship in a Digital Age*, I have chosen to remain within the domain of Aristotelian virtue-centred ethics, given that, when our International Institute for Ethics and Contemporary Issues launched this annual conference series in cooperation with UCU's Faculty of Social Sciences, it was within the broader framework of a theme we named *Integral Human Development in the Digital Age*, wherein 'Integral Human Development' is understood to be commensurate with what many moral philosophers today call *human flourishing*, which, in turn, is often meant to capture what Aristotle's *Nicomachean Ethics* is ultimately all about, namely, *Eudaimonia* (εὐδαιμονία). This rich Greek term, traditionally translated into English as *happiness*, is the main theme of Aristotle's virtue-centered *Ethics* and the very aim and ultimate purpose of each and every human life; everything we desire, Aristotle claims, is but a means to this final end and definitive goal. I think few would

disagree: at the end of the day, everyone wants to be happy, but not everyone agrees on what happiness is, nor on how to get it; these disagreements form the basis of the two great questions of the *Nicomachean Ethics*. Immediately in Book I, Aristotle makes it clear that although happiness entails good feelings, it is not just a feeling, but "something final and self-sufficient – the very end of *action*" and the highest "human good [as] an *activity* of the soul in accordance with virtue."[1] In this, Aristotle gives his answer both to what happiness is and to how to get it. This explains the first part of my subtitle, *In Praise of Responsibility*, because, for Aristotle, our ability to act, to choose, to respond to whatever life gives us, whether good and bad, is what determines our happiness. When our response is virtuous, happiness happens[2]; when our response is vicious, it does not.

Last year, I reflected upon contemporary notions of *Friendship* in the light of Aristotle's presentation of it as one of ten moral virtues.[3] This year, in a somewhat similar vein, I would like to critically reflect upon the contemporary phenomenon of artificial intelligence (hereafter AI), but in the light of the Aristotle's treatment of the intellectual virtues. My rather self-serving aim throughout this reflective exercise – namely, to keep my own sanity – is not entirely selfish, for if I manage to stay sane, there is a higher probability that my friends and loved ones and students will remain sane as well. And this brings me to the other part of my subtitle, *In Praise of Irrationality*. I shall never forget the force of a definition of insanity by G. K. Chesterton, who said that "a madman is not someone who has lost his reason, but someone who has lost everything but his reason." The great dignity (and danger) of being a human animal with a rational organizing principle of life, i.e., a soul, that is far superior to the organizing life principles of other living things, i.e., the souls of non-human animals and of plants, is only fully realized when we focus on how much we all share in common. The vegetative and nutritive processes in plant life are, of course, present in both non-human and human animals; and the irrational or non-rational powers of desire coupled with perception and consciousness present in non-human animals are in turn present in human animals. When we forget this, we not only tend to arrogantly and destructively lord our superiority over the animal and the plant kingdoms, but we lose sight of the very *reason* why reason is there in the first place, namely, to raise and train and tame those *irrational* powers and desires that we share with the other life forms on our planet. The trick is to discipline and elevate these desires without ever denying or ruining them. No doubt, reason can only do so much when it comes to the vegetative and nutritive powers of the soul; these are fixed by nature and cannot change. But when it comes to the irrational or non-rational powers and desires we share with non-human animals, it is entirely different. These powers can and must be trained and elevated by reason in order for us to perfect our nature, acquire virtue, and thus live a good life. The *Nicomachean Ethics* is not named after Aristotle's son, Nicomachus, for nothing. It reads at times like a parenting manual with loving parents giving advice to their children regarding how to be happy. Perhaps it could be titled something like, training without breaking: how to bring out your child's particular gifts and individuality without leading them to the precipice of false pride and individualism. And as for the main theme of happiness: what brings more genuine happiness

1 Book One, chapter 7.

2 The idea of letting happiness happen rather than aiming for it per se, first made a deep impression upon me when I read Viktor Frankl's *Preface* to his remarkable little book, *Man's Search for Meaning* (London: Rider, 2004) 14. Even the etymological similarity is striking: happin-ness/happen-s.

3 Friendship, as a moral virtue, is so important for Aristotle that he spends two entire books on it while he devotes only three books to the other nine moral virtues.

to a mother or father than seeing their children happy, and knowing that they had something to do with it?

But I want to concentrate on Aristotle's discussion of the intellectual virtues in order to see whether their inherent connection to the moral virtues sheds any light on the important topic of the ethics of AI that is before us. For Aristotle, as I have said, the human soul has two distinct capacities, the rational and the irrational, the latter being further distinguished by those powers that are fixed and those that can somehow speak and understand the language of rationality, thus allowing for change and, potentially, virtue. Aided by and in communication with theoretical wisdom, practical wisdom does the work of communicating with the irrational or non-rational parts of the soul to achieve moral virtue and good character. The rational power is also further distinguished into what Aristotle calls theoretical and practical powers, the perfection of which, in both cases, is the virtue of wisdom: theoretical/scientific wisdom involving pure thought (*Sophia*) or practical/deliberative wisdom involving rational choice (*Phronesis*). Whether theoretical or practical, the two virtues of wisdom, like the other three intellectual virtues, are ordered to truth, whereas the moral virtues are ordered towards the good. The intellectual and moral virtues all work together for the final end of the whole person, for to *know* the true good and then to *choose* it brings happiness, and such choosing of the good over and over leads eventually to finding pleasure in the good, which, for Aristotle, is the perfection of human nature.

In addition to the theoretical virtue of wisdom, Aristotle describes two other theoretical intellectual virtues, which involve the perfection of two distinct rational powers, the power of intuitively grasping first principles and of inductively understanding what is real and true; this is the intellectual faculty he refers to as *nous.* And then there is the power of demonstrative knowledge, primarily involving deduc-

tion, which he calls *episteme*. We might call the two theoretical virtues associated with these rational powers *right intuition* and *proper logical demonstration* respectively. For our purposes, it is the latter that is most relevant because it is possible to historically trace the invention of the computer all the way back to Aristotle's own development of this particular rational faculty – something I will attempt to do in order to make the claim that because Aristotle's logic laid the foundation for computer science, he is also the distant great-grandfather of one of its branches: AI – a great-grandchild, to continue the metaphor, that is in need of some serious fatherly discipline and moral advice before he destroys himself and indeed the whole world.

Although AI, as a branch of computer science, can be traced back to the now-famous brainstorming workshop at Dartmouth College in the summer of 1956, which brought together leading researchers[4] in what was then called the field of "machine intelligence", its deepest intellectual antecedents stretch back, not a mere six and a half decades, but nearly two and a half millennia – back to Aristotle himself. His logical system was so solid and influential in the history of thought that no one less than Immanuel Kant wrote that although there were great logicians after Aristotle, none of them were really able to "take a single step forward, and therefore [Logic as a discipline] seems… to be finished and complete."[5]

[4] I had the pleasure of briefly meeting and listening to one of the Dartmouth workshop participants, Mr. Raymond Solomonoff (one of the founders of AI), in February of 2008 when gave the keynote address at a conference titled "Current Trends in the Theory and Application of Computer Science" at my university in Lebanon. Mr. Solomonoff graciously stayed on for a few weeks to give a short course on his own theory of algorithmic information. On the last day of the course, many students and teachers participated in conversing and taking pictures with Mr. Solomonoff. He died one year later.

[5] This comes in Immanuel Kant's own Preface to the 2nd edition of his *Critique of Pure Reason*, tr. Kemp Smith (London: Macmillan, 1929) 17.

As it turned out, "seems" was the key word here because a significant step forward did finally take place; and it came half a century after Kant's death when a largely self-taught philosophical genius, the son of a shoemaker, published a monograph titled *An Investigation of the Laws of Thought*. The year was 1854, the name of this genius, George Boole (1815-1864). Contrary to what is too often taught, Boole never challenged or undermined the basic principles of Aristotelian logic. In fact, concurring with Kant, he acknowledged his debt to Aristotle when he wrote:

> In its ancient and scholastic form, indeed, the subject of Logic stands almost exclusively associated with the great name of Aristotle. As it was presented to ancient Greece in the partly technical, partly metaphysical disquisitions of *The Organon*, such, with scarcely any essential change, it has continued [and for good reason] to the present day.[6]

But Boole was able to do what Kant apparently never even dreamt of: to extend and expand Aristotelian logic by translating the syllogism into an algebraic calculus, thus giving it a new mathematical foundation that allowed for novel and powerful applications – somewhat analogous to the way Descartes extended Euclidian geometry by converting lines into numbers, and diagrams into formulas, thus allowing the manipulation of symbols to go beyond mere spatial intuition.[7] Boole's principle of "wholistic reference", in general, and his theories of concept and proposition formation, in particular, broke new ground in the extension of Aristotelian logic and paved the way

[6] George Boole, *Laws of Thought* (Cambridge: MacMillan & Co., 1854) 1. The original publication has been digitized and is available on line at: http://www.ccapitalia.net/descarga/docs/1847-boole-laws-of-thought.pdf

[7] See my "Descartes' *Discourse on Method*: More Discourse?" in *Budhi, A Journal of Ideas and Culture*; vol. 6, No. 2 & 3 (2002).

for novel developments in mathematical logic and for the establishment of the new field of computer science.

At the time Boole published his mature ideas, another philosophical genius, Gottlob Frege, was born in Germany, who would likewise play a major complementary role in the extension of Aristotle's *Organon* and in the development of the kind of mathematical logic that would pave the way for the new field of computer science, eventually leading, of course, to that particular branch known as AI. Without entering into the complex and somewhat controversial question regarding Boole's influence on Frege, since Frege was definitely aware of Boole's work, there is no doubt that his original genius played a crucial role in the story we are telling. Not only was he able to formalize the very notion of proof through a new analysis of quantified statements, but his entire philosophy of language allowed for a novel way of conceiving the relation of mathematics to logic. In this, both Boole and Frege were mutually inspired by Leibniz's dream (almost two centuries earlier) of a universal concept language, a formal mathematical language that could symbolically capture the truth-value of mathematical statements. Leibniz had even built a calculating machine for the purpose of manipulating these symbols in such a way as to allow for a kind of *decision* on the part of the machine regarding the truth values of these statements. Although the machine was successful at one level, it never achieved what Leibniz ultimately wanted due primarily to the shortcomings in his formal language.

Surely, Leibniz's dream of a universal concept language was conceived only in the aftermath of Descartes' attempt at a universal mathematics,[8] which, paradoxically, led to Descartes' radical turn towards subjectivity. We may be able to understand this central paradox at the core of modern philosophy if we approach it from what

[8] See footnote 7.

we might call the world's "first and second greatest ideas". Here, I am deeply grateful for the work of Linda Zagzebski (she delivered the Gifford Lectures in 2015), who argues that these "two greatest ideas" account, in one way or another, for all of the significant intellectual discoveries in history and indeed, in some sense, for civilization and history itself. She argues that the first idea (the idea that the mind can grasp existence in terms of a unity, after which the key term *uni-verse* emerges), is dominant and foundational in the ancient and medieval eras, while the second greatest idea (that the mind can grasp itself) is what establishes and characterizes modernity – embodied, in part, by this unique Cartesian turn towards subjectivity – a shift paradoxically achieved in the context of a search for universality. To see that the second idea is only possible *because* of the first sheds light on all aspects of the perennial metaphysical problem of the one and the many[9] at the heart of virtually every possible field of enquiry, including ethics – which is why my title is "from Aristotle to AI and back again." We will soon return to Aristotle, but it is first necessary to complete the story leading up to the birth of AI so that our return to his *Nicomachean Ethics* may be all the richer.

Less than a quarter of a century after Boole and Frege, the field of mathematical logic was in full swing and took a major step forward with the 1910 publication of *Principia Mathematica* by Bertrand Russell and Alfred North Whitehead. No one could have predicted then, not even Russell or Whitehead, how this field was on the verge of erupting into a computer science that would dras-

[9] Here I am indebted to, and inspired by, Hans-Georg Gadamer's lectures on the Pre-Socratics, particularly his brilliant interpretation of what he calls "Parmenides' didactic poem", wherein he claims one can detect, as Heidegger did, the "ontological difference" between *ousia* and *on* (being and beings) which is not "made" by thinking, but which is already there for us to discover. See Gadamer's *The Beginning of Philosophy* tr. Rod Coltman (New York: Continuum, 1998) 107–125.

tically change the modern world forever. To be sure, the first few decades of the 20th century witnessed the invention of mechanical devices and prototype computers that could calculate differential equations at amazing speed. One such machine, an analogue computer,[10] was developed by a certain Vannaver Bush, Professor of Electrical Engineering at MIT, who would go on to play a key role in the Manhattan Project during World War II. It was Bush's student, though, Claude Shannon, an original member of the 1956 Dartmouth summer workshop, who made the decisive breakthrough. As an undergraduate, Shannon was required to take several courses in general philosophy; in addition to being introduced to Plato and Aristotle, he was also exposed to the philosophy of George Boole, which caused a spark to go off in Shannon's young mind that would set an entire field ablaze. He began to see the connection between his own field of electrical engineering and the deepest modern philosophical foundations of symbolic mathematical logic rooted in Aristotle. Perhaps we could say that the modern field of computer science was partially conceived in this spark. The subsequent publication of his 1938 paper, "A Symbolic Analysis of Relay and Switching Circuits," turned out to be one the most important academic papers of the century. While doing the research for this paper, I found Shannon's 1938 paper and read it as one would read a detective novel. Since I am virtually illiterate in the language of Electrical Engineering, I understood precious little of it, but in struggling through the text, it became clear that his reliance upon the language and principles of modern symbolic logic was central to his work. Equipped with the insights and achievements of Boole and Frege, and encouraged by the systemic

[10] Analogue computers eventually became obsolete with the invention of digital computers; these latter machines applied the use of symbolic language in a way that allowed for much greater complexity, efficiency, and a much broader general range.

power of the *Principia Mathematica* (there is a notable reference in this paper to the work of Whitehead), Shannon combined his own field of electrical engineering with the most momentous insights of the previous century and positioned himself to become one of the key participants in the now famous 1956 Dartmouth seminar at which AI was officially born. The last point I will briefly make here before I begin my conclusion is that at about the same time Shannon was doing his work at MIT, a young Englishman in his early twenties by the name of Alan Turing was independently working on a paper titled, "On Computable Numbers, With an Application to the [Decision Problem] *Entscheidungsproblem*." The origin of this problem, as we have seen, went all the way back to Leibniz, but a modern and more advanced version of the problem had been presented in 1928 by David Hilbert and Wilhelm Ackermann, and Turing had set himself the task of solving it, which, in a way, he did – precisely by showing it could not be solved – at least not by any computational procedure.[11] Professor John McCarthy, the main

[11] At a talk given in Calgary on 24 January 2012 by Richard Zach celebrating the Alan Turing Centenary, Zach stated correctly that Turing "showed that no Turing machine, and hence no computational procedure at all, could solve the *Entscheidungsproblem*." "Many scientific questions are considered solved to the best possible degree when we have a method for computing a solution. This is especially true in mathematics and those areas of science in which phenomena can be described mathematically: one only has to think of the methods of symbolic algebra in order to solve equations, or laws of physics, which allow one to calculate unknown quantities from known measurements. The crowning achievement of mathematics would thus be a systematic way to compute the solution to any mathematical problem. The hope that this was possible was perhaps first articulated by the 18th-century mathematician-philosopher G. W. Leibniz. Advances in the foundations of mathematics in the early 20th century made it possible in the 1920s to first formulate the question of whether there is such a systematic way to find a solution to every mathematical problem. This became known as the decision problem, and it was considered a major open problem in the 1920s and 1930s. Alan Turing solved it in his first, groundbreaking paper "On computable numbers" (1936). In order to show that there cannot be a systematic computational procedure that solves every mathematical question, Turing had to provide a convincing analysis of what a computational procedure is. His abstract, mathematical model of computability is that of a Turing Machine. He showed that no Turing machine, and hence no computational procedure at all, could solve the *Entscheidungs* problem."

organizer of the Dartmouth seminar, one of the founders of AI, and the one who coined the term, noted in 2006 that because Turing might have been the first to really understand that programming computers was the main way to realize AI, he would have played an important role at the seminar; unfortunately, he had died just two years before.[12]

Before commencing my conclusion with a return to Aristotle's *Nicomachean* Ethics, I must say something about these last 64 years. My brief remarks may be surprising and I anticipate the following question: are we not gathered today to reflect upon the ethics of AI primarily in the light of the remarkable developments that have taken place since 1956, and especially in the last ten years or so? Part of my answer to this legitimate question is developed in the light of an article published by John McCarthy himself, whom, as I have just mentioned, was the main organizer of the Dartmouth seminar, one of the founders of AI, and the one who coined the term. Not too long before his death in 2011, McCarthy stated:

> My hope for a breakthrough towards human-level AI was not realized at Dartmouth, and while AI has advanced enormously in the last 50 years, I think new ideas are still required for the breakthrough… [b]esides proposals for extending logic, there are many systems that restrict logic in order to make computation more

[12] See McCarthy's article: http://www-formal.stanford.edu/jmc/slides/dartmouth/dartmouth/node1.html
John Vincent Atanasoff, too, played a key role in the invention of the digital computer, but has not really given the credit he deserves.

efficient. I'd prefer to use full logic but want systems that can reason about their own reasoning methods in order to *decide* on efficient reasoning. After all these years, I still have not been able to make specific proposals.[13]

[13] (My emphasis in italics) In an article titled, "The Dartmouth Workshop – As Planned and as it Happened," which is available on line at: http://www-formal.stanford.edu/jmc/slides/dartmouth/dartmouth/node1.html McCarthy importantly stated: "I remember well only the events at Dartmouth that intersected with my own scientific interests, so this is not a comprehensive account of what went on. Good work that I am ignoring here includes Raymond Solomonoff's work on algorithmic information and E. F. Moore's further development of his ideas on automata. What came out of Dartmouth? I think the main thing was the concept of artificial intelligence as a branch of science. Just this inspired many people to pursue AI goals in their own ways. My hope for a breakthrough towards human-level AI was not realized at Dartmouth, and while AI has advanced enormously in the last 50 years, I think new ideas are still required for the break through. What has happened since 1956? AI research split, perhaps even before 1956, into approaches based on imitating the nervous system and the engineering approach of looking at what problems the world presents to humans, animals, and machines attempting to achieve goals including survival. Neither has achieved human-level AI. Proposals that one approach should be abandoned and all resources put into the other are silly, as well as being unlikely to happen. I'll confine myself to engineering approaches. Within the engineering approach, the greatest success has been accomplished in making computer programs for particular tasks, e.g. playing chess and driving an off-the-road vehicle. None of these purport to have achieved general common sense knowledge. Thus the chess programs do not know that they are chess programs. Their ontology consists mainly of particular positions. The logical AI approach is in principle more ambitious. It requires representing facts about the world in languages of mathematical logic and solving problems by logical reasoning. It faces many difficulties, some of which have been overcome, and there are proposals for overcoming others. Nevertheless, there is still not a well-accepted plausible plan for reaching human-level AI. For some years, I have thought mathematical logic needs to be extended in order represent common sense knowledge and reasoning. That extensions are possible may seem paradoxical in the light of Gödel's 1929 completeness theorem for first order logic. (Don't confuse this with his 1931 incompleteness theorem for formalized arithmetic.) The 1929 theorem tells us that any sentence true in all models of some premises has

McCarthy gives expression here to a tension in, and a debate about, the future of AI that is still very much alive and shows no signs of going away. On one hand, he acknowledges the enormous advancement on the engineering level, but then admits his disappointment at the failure, including his own, to come up with any *new ideas* on the ontological level to really solve the one fundamental problem that has been around since the time of Leibniz: the decision problem. What he seemed to appreciate so much about Turing, was not only that Turing was the first to launch a serious scientific enquiry into the question of human level machine intelligence, but also showed, by the invention of his own Turing machine, that the fundamental obstacles to achieving human-level AI would not be solved through computational procedures. Today, the latest cutting-edge research in this field still grapples with the same

a proof from these premises. Therefore, any genuine extension of logic must allow inferring some sentences that are untrue in some models of the premises. The various systems of formalized nonmonotonic reasoning do precisely that. They allow inferring sentences true in *preferred* models of the premises. Human commonsense reasoning is often nonmonotonic, and human-level logical AI requires nonmonotonic reasoning, but how to do this in a sufficiently general way is still undiscovered. The need for nonmonotonic reasoning is well accepted in AI, although for specific domains, the human designer often decides what interpretations are preferred and relegates only monotonic reasoning to the computer. This is at the cost of generality. Besides nonmonotonic reasoning, I propose other extensions to logic to be able to do common sense reasoning. These include systems with concepts as objects, systems with contexts as objects, and admitting entities that cannot be characterized by if-and-only-if definitions. I'm sure there's lots more needed before logic fully covers common sense. My proposals are in articles published here and there but all available from my web page: http://www-formal.stanford.edu/jmc/. Besides proposals for extending logic, there are many systems that restrict logic in order to make computation more efficient. I'd prefer to use full logic but want systems that can reason about their own reasoning methods in order to decide on efficient reasoning. After all these years, I still have not been able to make specific proposals."

problems that McCarthy identified a few years ago before his death, with very little progress, while the progress at the level of engineering continues to sky-rocket. In other words, at the philosophical level, with which we are primarily concerned, there is not really that much more to report on since 1956, with the exception of the notable breakthroughs in the philosophy of mathematics relevant to AI application. But, again, these are primarily in the realm of engineering, not at the level of philosophical breakthroughs in Cognitive Science, and thus are not directly relevant to my express goal of shedding light upon the ethics of AI in the context of the links to the Aristotelian Intellectual Virtues and our general theme of *Integral Human Development* – to which, by way of conclusion, I now return.

I began this reflection by pointing out that, for Aristotle, the human soul has two distinct capacities, the rational and the irrational, with the former distinguished by theoretical and practical powers, the *perfection* of which, in both cases, is the one two-fold virtue of wisdom: theoretical/scientific wisdom involving pure contemplative thought (*Sophia*) and practical/deliberative wisdom involving rational choice (*Phronesis*). I also noted that, for Aristotle, this two-fold rational virtue of wisdom which contemplates truth and chooses the good, both complements, and is complemented by, two other theoretical capacities which involve the intuitive ability of inductively grasping first principles (*nous*) and the deductive power of logical demonstration (*episteme*). There is one more intellectual capacity that I did not mention above, but which should now be specified as I draw my conclusions; this is the power of (*techne*), the perfection of which is the practical intellectual virtue of art or know-how that works closely with the theoretical virtue of *episteme*. Although distinct, these three powers are not separate; they dynamically overlap and are designed by nature to work in

harmony with the two-fold virtue of wisdom for the final end of the whole person, which is happiness or human flourishing. Furthermore, I have described the achievement of AI as a significant historical development of *episteme* – the deductive power of logical demonstration, the foundation of which is laid in Aristotle's *Organon* – a development that has been aided too by a significant expansion over the ages of the practical intellectual capacity of *techne*. But the ultimate question is this: are these remarkable developments and achievements virtuous? This is not only a question for us who are gathered to reflect upon ethics in a digital age, but for every human being that desires happiness. It may be somewhat sufficient to say that AI in itself is morally neutral, neither good nor bad, but is and will be precisely what we make of it. This may be entirely sufficient for other related and powerful contemporary developments (like globalization for instance), but given that we are here talking about developments analogous to the workings of the human mind itself[14], and about the express desire on the part of some of the very founders of AI to achieve, what they call, human-level AI, I think our answer must be more nuanced and should elicit more questions such as "what exactly does the expression human-level AI mean?" By retaining the adjective, *artificia*l, to qualify intelligence, the phrase seems to maintain a distinction of kind between human and artificial intelligence, though one could argue it is only a distinction of degree. Other questions are: "why do we want to achieve human-level AI in the first place?" And "will such an achievement, assuming we can even come up with those 'new ideas' (McCarthy did believe that one day we should be able to reach human-level AI,[15] and until his death was working on

[14] Made possible, I suggested above, by the world's second greatest idea.

[15] John McCarthy, "What is artificial intelligence?", 171 (2007) 1174–1182.

identifying the fundamental formidable problems[16]) that McCarthy spoke about[17], contribute to human flourishing and happiness?" To be sure, some contemporary research is responsibly focusing on ways of using AI to solve some of the serious global problems the world is facing, but the dark reality is that the vast majority of AI research is driven by the very same industry in which it had its origin, the war industry. We saw above that Claude Shannon's supervisor played a key role in the Manhattan project during WWII, which perversely named its operation "Trinity" and which even more perversely ended up naming the atomic bombs dropped on Japan, "Little Boy" and "Fat Man". But even if some radical shift

[16] *The Dartmouth workshop, planned and as it happened*: "For some years, I have thought mathematical logic needs to be extended in order represent common sense knowledge and reasoning. That extensions are possible may seem paradoxical in the light of Gödel's 1929 completeness theorem for first order logic. (Don't confuse this with his 1931 incompleteness theorem for formalized arithmetic.) The 1929 theorem tells us that any sentence true in all models of some premises has a proof from these premises. Therefore, any genuine extension of logic must allow inferring some sentences that are untrue in some models of the premises. The various systems of formalized nonmonotonic reasoning do precisely that. They allow inferring sentences true in *preferred* models of the premises. Human commonsense reasoning is often nonmonotonic, and human-level logical AI requires nonmonotonic reasoning, but how to do this in a sufficiently general way is still undiscovered. The need for nonmonotonic reasoning is well accepted in AI, although for specific domains, the human designer often decides what interpretations are preferred and relegates only monotonic reasoning to the computer. This is at the cost of generality. Besides nonmonotonic reasoning, I propose other extensions to logic to be able to do common sense reasoning. These include systems with concepts as objects, systems with contexts as objects, and admitting entities that cannot be characterized by if-and-only-if definitions. I'm sure there's lots more needed before logic fully covers common sense."

[17] I highly doubt this is possible because it is impossible to computationally produce agency, as Turing showed. Of course, this also depends upon what is meant exactly by *human-level* AI since by retaining the adjective, *artificia*l, to qualify intelligence, the suggestion (or desire) may not be to mechanically produced human intelligence.

were to take place and AI research scaled down its war and weapons of mass destruction[18] impetus in favor of more benign and humane goals, I believe the ethical debate should continue; for as Werner Heisenberg once wisely said, "[o]ne has to remember that every tool carries with it the spirit by which it has been created." And the spirit of over-developing one intellectual capacity at the expense of all the others creates an unbalanced monstrous energy that undermines integral human development. In alienating the other intellectual powers of *nous* and theoretical wisdom, it prevents practical wisdom from its crucial task of communicating and negotiating with the irrational parts of the soul that we have in common with our non-human animal companions, and which are part and parcel of the praiseworthy characteristics that make us the unique creatures we are. When cut off from the wisdom dwelling in the wild woods and the wild animals, an entirely new kind of mechanistic human arrogance emerges that precludes a healthy cultivation of and relation to these other vital life forces on our planet. The myopic rational development of *episteme* in isolation from the other higher powers of rationality leads not to a responsible society of rational animals with moral character, who feel a sense of responsibility for one another and for all the life forces in our planet, but to a greedy gang of irresponsible mechanistic beasts or beastly machines – beastly, in part, because so artificial: artificial birth-control, artificial sex, artificial love, artificial friendship, artificial life, artificial death, and… artificial intelligence? Surely, there is something intellectually virtuous about this stupendous development, as I have tried to show, but the moral questions I have raised ought to be addressed sooner rather than later.

[18] And today, the "weapons of Math destruction" – a reference to mathematical algorithms that destroy human reputation, life, etc., by turning persons into commodities without even knowing it: the two senses of WMD.

Given that it is still winter here in Lviv, I will end with a little *winter* poem that marvelously captures our theme of moral responsibility, with its celebration of the centrality of human promises, and our praise of irrationality, with its allusion to how even a little horse, when next to the forest, seems to know the difference between what is right and wrong. Written in 1922 – very near the place in New Hampshire where the famous Dartmouth seminar took place, Robert Frost's words are as fresh and powerful today as they were a century ago when he wrote:

Whose woods these are I think I know.
His house is in the village though;
He will not see me stopping here
To watch his woods fill up with snow.
My little horse[19] must think it queer
To stop without a farmhouse near
Between the woods and frozen lake
The darkest evening of the year.
He gives his harness bells a shake
To ask if there is some mistake.
The only other sound's the sweep
Of easy wind and downy flake.
The woods are lovely, dark and deep,
But I have promises to keep,
And miles to go before I sleep,
And miles to go before I sleep.

[19] See Adolph Potmann's *Animals as Social Beings* (1961).

Czesław Porębski

Information Overload, Big Data, and Freedom

1. Information Overload?

"Information overload" implies a thesis that sounds like a denial of the ancient wisdom: *scio me nihil scire* – I know that I know nothing. We seem to complain that we know too much, which makes us feel ill at ease. And this feeling of uneasiness seems to be one of the symptomatic ailments of "our time".

Before we accept or reject what "information overload" implies, we should ask: who is speaking, who are we, and what sort of "information" is meant? Many would protest: we still do not know very much; and this little we know helps us only realize the immensity of our ignorance! Things like these may be expected first of all from those who, working on the first line of scientific research and serious reflection, return to fundamental questions and find in them their source of inspiration and motivation. They would point to how recent are those discoveries and theoretical findings in which our image of humankind, of different levels of reality, starting from that of quantum physics and ending at the cosmological level, is grounded.