Contemplation for an Age of Artificial Intelligence

Marco Schorlemmer
Artificial Intelligence Research Institute, IIIA-CSIC

Meditatio Centre, London, 20th May 2017

ABSTRACT: Research in Artificial Intelligence (AI) is advancing rapidly, yielding computational systems with impressive capabilities. This has brought world-renowned scientists and engineers to voice their concerns about the risks that AI research can pose to the survival of humanity. But, is AI really such an apocalyptic threat? Actually, we can learn a lot from AI research about how we see ourselves as sentient beings. Marco Schorlemmer will argue that AI is not a threat to our survival 'per se', but that we need to cultivate our capacity of intimate inquiry if we are to survive in the emerging knowledge and innovation societies in which AI will play an important role. Actually, AI research (and research in other techno-scientific fields) is very valuable for humanity... if done with a contemplative mind.

Introduction

This day at the Meditatio Centre is about "artificial intelligence" but also about "contemplation". You may wonder what their relationship is, and my objective of this day is to persuade you that there is actually a vital relationship between these two topics. First, I will attempt to raise the awareness that Artificial Intelligence, AI, is here to stay and that it'll have an important impact on our lives, very soon. This will be the topic of my morning talk: it's the "… an Age of Artificial Intelligence" bit of the title. In the afternoon, I will attempt to make clear that nurturing our contemplative dimension will be absolutely necessary to be able (1) to cope with this impact that AI will have on our lives and (2) to provide orientation on how to do research and how to develop these new technologies and how to deploy them in our societies for the well-being of humanity, all life forms and our environment. This will be the topic of my second talk: it's the "Contemplation for…" bit of the title. Finally, in the wrap-up session, I may add some additional thoughts on the role of contemplation in our emergent knowledge and innovation societies.

Ramon Lull and the Mechanisation of Thought

Let us start by calling to mind the figure of an important philosopher, theologian, mystic and contemplative of the 13th and 14th century, which scholars also credit with the first serious attempt to mechanise human reasoning. His name was Ramon Llull (also know by his Anglicised name Raymond Lully or Raymond Lull, or by his Latin name Raimundus Lullus). Ramon Llull is one of the prominent figures within European thought. He is very well known in Catalonia, because he is credited with writing the first major works in philosophy and literature using the Catalan language in addition to Latin. In Barcelona we have a university named after him, and the Catalan equivalent to the British Council, i.e., the governmental organisation that promotes the Catalan language and culture internationally, is called the Ramon Lull Institute, which by the way has also an office here in London. Last year, in 2016, we celebrated the "Ramon Llull Year" on occasion of the 700th anniversary of his death.

Ramon Llull was born on Majorca, only a few years after James I, King of Catalonia and Aragon conquered the island to/from the Moors. Ramon Llull was from a wealthy family, with important links to the royal court. As a grown up man in his 30s, with wife a two children, he was living a frivolous life as a troubadour, when he had a series of consecutive mystical experiences that turned his life upside down, leading him to leave his family, his position, and all his belongings to become

a Franciscan tertiary and to devote his life to the service of God with the objective to convert Muslims to Christianity. For this reason he learned Arabic and Islamic culture, and in a later revelation, he envisioned the method to carry out his missionary task. Llull devised a complex system based on the mechanical manipulation by means of a series of concentric wheels of philosophico-theological concepts –the so called Ars Lulliana or Lullian Art– which in a systematic fashion constructed and analysed rational arguments that supported the core beliefs of Christianity. Ramon Llull presented his Art at the major European universities of his time and also to the Pope in order to get support and funding for his missionary task, and he even put the Art into practice in several missionary trips to North Africa. The works he composed during his lifetime (he died in his mid 80s) is very impressive –some 280 works in Catalan, Latin and Arabic– and he directly influenced the thought of posterior philosophers such as Nicholas of Cusa, Giordano Bruno and Gottfried Willhelm Leibniz, among many others.

Today it is widely recognised that Ramon Llull anticipated some important ideas that centuries later led to the development of Computer Science and Artificial Intelligence, such as the idea of a formal algebra of thought with which one could perform logical analysis and deduction based on mechanically-realised calculations through the manipulation of symbols. But it was only by the late 19th century and early 20th century, that the conceptual apparatus and the mathematical foundations were mature and clear enough for computers to be actually built and programmed, and subsequently for the field of Artificial Intelligence to arise in the mid 20th century with the explicit aim to design and build machines that exhibit what we would call intelligent behaviour.

AI in Our Daily Lives

Where are we now in this quest of mechanising reasoning as initiated by the mystic Ramon Llull 700 years ago and carried out today with electronic computers with the aim of tackling all kinds of problems and tasks? AI-based technology is already present in many very diverse activities, such as in aircraft navigation, stock trading, medical diagnosis, home appliance control, internet keyword search, logistics, product recommendation, efficient energy consumption, and we all carry AI in the mobile phones in our pockets in the form of personal assistants such as Siri, Cortana or Google Assistant.

Three years ago, four world-renowned scientists put forward in an article published in The Huffington Post the following reflection of the future of Artificial Intelligence. These authors were:

- Stephen Hawking, Director of Research at the Centre for Theoretical Physics at Cambridge
- Max Tegmark, MIT cosmologist, scientific director of the Foundational Questions Institute
- Stuart Russell, Computer Science professor at Berkeley; Co-author, of the most widely used textbook on AI at universities worldwide
- Frank Wilczek, Nobel laureate and physics professor at MIT

In their article they said:

- Artificial intelligence (AI) research is now progressing rapidly
- Recent landmarks [in AI] will probably pale against what the coming decades will bring
- The potential benefits are huge. [W]e cannot predict what we might achieve when [human] intelligence is magnified by the tools AI may provide.
- Success in creating AI would be the biggest event in human history.

But they also said:

- Unfortunately, it might also be the last [event in human history], unless we learn how to avoid

the risks.

- Near term: world militaries are considering autonomous weapon systems that can choose and eliminate their own targets
- Medium term: AI may transform our economy to bring both great wealth and great dislocation
- Long term: machines with superhuman intelligence could repeatedly improve their design even further,
  - outsmarting financial markets
  - out-inventing human researchers
  - out-manipulating human leaders
  - developing weapons we cannot even understand

We are facing potentially the best or worst thing ever to happen to humanity.

Additional celebrities have also voiced their concerns, such as Bill Gates (Microsoft), Elon Musk (Pay Pal, Tesla Motors, SpaceX), and Morgan Freeman (actor).

Actually, when I first read this article I was surprised about such an apocalyptic view. As an insider of the AI research field, and knowing what has been achieved and what continue to be very hard problems in AI, my first reaction was to dismiss these claims and fears as completely out of touch of what can be achieved with AI in the near-, mid-, and even long-term future. But then, these are prominent figures of science and engineering, and also of AI in particular; so, a second, closer look to what they say is necessary in order to assess which is actually the risk humankind is facing.

In Our Image – AI's Religious Narrative

Since its inception, AI has been surrounded with controversy. Mainly because there is no consensus of what "intelligence" is, so that the term "artificial intelligence" creates a lot of confusion. Questions such as…

- "Can machines think?"
- "Can computers have a mind?"
- "Can robots be conscious?"

… have accompanied the field of AI and have subsequently been tackled by all sorts of thinkers and philosophers (and they are usually the sort of questions we get from people outside our field, such as journalists). There are thinkers (such as the philosopher of science Jordi Pigem) that even claim that the term "artificial intelligence" is an oxymoron because true intelligence is natural, vital and above all "cordial" –i.e., stemming from the heart in its rich, spiritual sense. The Berkeley philosopher John Searle, also advocates that intelligence is observer-relative: he claims that there cannot be intelligence in computing machinery, since intelligence is something in the eye of the (conscious) beholder. We see intelligence in the behaviour of computers.

Anyway, these questions don't have a clear answer because they often involve a particular worldview, a certain metaphysical outlook on what humans and machines actually are, what one understands by 'thinking' or 'mind', and what consciousness, ultimately, is. And these are unsettled issues. I won't answer these questions myself, because I do not think they can be answered as they stand. Actually, we researchers in AI don't bother very much about them in our daily scientific tasks. Still, these questions, and the advances done in the field of AI over the last sixty years have triggered the imagination not only of scientists and engineers, put also of philosophers, novelists, filmmakers and the public at large. One only needs to look at the amount of movies that have AI as their central topic: 2001 A Space Odysee, Blade Runner, Terminator, Matrix, Her, Ex Machina, etc.

Staying at this more controversial level of analysis of AI, one could even claim that, the quest of AI shows some sort of religious narrative:

- one could see is as a contemporary expression of the ancient biblical idea of creation in the image of the creator, as we encounter in the first chapter of Genesis; or,
- one could see it also as a mundane way to achieve some sort of immortality, by attempting to take what is assumed to be specifically human –rational thought– and to "transfer" it to machines, liberating in this way the mind from the mortal body upon which it depends.

But, although there might be some truth in exposing the religious narrative of AI, it is necessary to further look into the conceptual metaphors underlying the AI research programme.

Metaphors of the AI Research Programme

The research programme of AI originates with certain metaphors of mind that constitute the basis of Anglo-American analytic philosophy. I say "metaphors of mind" because cognitive scientists have in the last decades provided significant evidence that most of our concepts, and in particular abstract concepts such as "intelligence" or "mind", are largely metaphorical and grounded on our bodily sensorimotor experience as we interact with our environment.

If you have been to Greece, you might have seen lorries and busses with the word "metáfora" written on them. "Metáfora" means transportation, and by means of metaphors we transport the conceptual structure of one domain that we are familiar with into another more abstract domain. By looking closely at the metaphors underlying the AI research programme we can find out how we understand concepts such as "mind", "rationality" and "intelligence", and thus become aware of how we see ourselves as intelligent beings.

A very common way to conceptualise the mind is as a body, where thinking is conceptualised metaphorically as some kind of physical functioning such as moving, seeing, object manipulation or even eating, and ideas are physical entities with an independent existence, such as locations, things perceived, manipulable objects or food. We see these metaphors in expressions such as:

- "How did you reach this conclusion?"
- "I see what you are saying."
- "This subject matter is beyond his grasp."
- "Let's ruminate this idea a bit more before abandoning it completely."

On top of these metaphors, people have also used the Mind Is A Machine metaphor, and more recently the Mind Is A Computer metaphor, by which we conceptualise the mind as a computer program, concepts and ideas as formal symbols, and thinking as symbol manipulation.

There is nothing wrong in using metaphors. Actually, as argued by the cognitive scientist Geroge Lakoff and the philosopher Mark Johnson, "it is virtually impossible to think or talk about the mind in any serious way without conceptualising it metaphorically." The problem lies, they say, in the different attitudes towards conceptual metaphor: Lakoff & Johnson list three attitudes that one can adopt towards the conceptualization of the mind in computational terms as stated in the Mind As Computer metaphor.

1. One can note that it is a metaphor and study it in detail – i.e., to study what sort of metaphorical entailments one can draw and how this influences our understanding of the mind.

2. One can recognize its metaphorical nature, but still take it very seriously as a scientific model

of the mind. Many practitioners of what has been called the weak version of artificial intelligence take this position.

Practitioners of weak AI use computational models as a tool to study the mind by formulating and testing hypotheses about it. Weak AI has also come to be associated with attempts to build programs that aid and enhance, rather than duplicate, human mental activities.

3. One can take the metaphor to disclose a deep scientific truth, namely that concepts *are* indeed formal symbols, that thought *is* computation and that the mind *is* a program running on 'brain' hardware (or wetware, so to say). This is the position of what has been called "strong AI".

Strong AI, in contrast to weak AI, is associated with the claim that an appropriately programmed computer could have a mind and could think at least as well as humans do. There are an important number of AI researchers for which achieving strong AI is the ultimate goal of the field. So, once the Mind As Computer metaphor is taken literally as defining the very essence of mind, it is not consciously seen as a metaphor any more, but rather as "the Truth".

Hype and Fear of AI

So far, whenever impressive results and breakthroughs in the design and implementation of computer systems that exhibit some sort of intelligent behaviour have been achieved (achievements in "weak" AI), –such as beating the world champion in chess, or winning in the American television game show Jeopardy!, which requires extensive general knowledge and the capability to "understand" natural language– whenever these breakthroughs have been achieved, they are usually followed by:

- often over-optimistic claims of what could be done with AI, and about when human-level AI would be achieved ("strong" AI)
- as a consequence of this, often extreme dystopian views of what this would mean for the future of humanity are voiced

Now we are again going through one of these periods of impressive results in AI technology ("weak" AI), accompanied by speculations about when human-level AI will be reached ("strong" AI), and eventually surpassed, and about the risks of reaching this point in AI development. However, the current surge of interest and application of AI is without precedent and the difference to previous periods of exaggeration, hype and dystopian visions is that current "weak" AI technology is achieving human-level performance or even improving this performance in areas that were traditionally considered very difficult for machines to tackle, such as:

- face and speech recognition,
- language translation, or
- image processing and classification

Consequently, there is now an important uptake by industry of current AI techniques, and all major tech companies (Google, Amazon, Microsoft, Apple, …) have set up specific departments and units that do research and development in AI. There is a now a growing consensus that AI is going to be of huge importance for the future economy and functioning of human societies. One immediate question that many currently pose is how AI technology is going to affect the labour market: Will robots take our jobs?

Here is this website of the BBC that allows you to check the "automation risk" in the next 20 years for particular job titles. It is based on a study that was carried out at the university of Oxford

together with the consultancy firm Deloitte: http://www.bbc.com/news/technology-34066941

- taxi and cab driver or chauffeur: 57% likelihood of automation
- library clerk: 97% [local government administrative worker]
- clergy member: 2% [nurse, midwife]

What impact will this automation of our jobs have on the functioning of human societies? For this it will be useful to look at AI's core assumptions when designing computational systems that exhibit intelligent behaviour.

Rationality and Autonomy

The main paradigm that is followed in most AI research is that of an "autonomous, rational agent". It is the paradigm described in the most widely used AI textbooks in universities around the world today.Let me describe this paradigm of an "autonomous, rational agent" in very general terms. In AI we aim at designing a computational entity, or system, that perceives other entities and the environment in which it and the other entities are situated (i.e., it gets input from its sensors), and which based

- on the current percept,
- on the memory of previous percepts and
- on some sort of internal model of the environment and other background knowledge,

it autonomously takes a decision to act in a certain way (i.e., it produces output on its actuators). Hence the name "agent". What is the principle of choosing a particular action? For this, in AI we usually adopt an "idealised" understanding of what it means for an agent to act "intelligently", namely by narrowing it down to the notion of "rationality". And to act "rationally" is understood as to choose the action that produces the best outcome, according to some previously defined performance measure (or, when there is uncertainty, as in most cases, to choose the action hat produces the best expected outcome).

This approach has been very successful for the advancement in AI research because it avoids the vague and ambiguous notions of "thinking", "intelligence", "consciousness", etc. and adopts a notion of "rationality" that can be effectively checked for progress and success:

- An "autonomous, rational agent" playing chess perceives the chess board with its figures and the moves of its opponent, and chooses the move that most likely leads to it winning the match. An agent that chooses better moves and thus wins more matches is deemed "more intelligent" at playing chess. If it does so better than humans then it is "more intelligent" than humans for this particular task.
- An "autonomous, rational agent" that aims at classifying tumour tissue images to diagnose skin cancer that classifies images with a lower error rate than humans is deemed "more intelligent" than humans for this particular task of image classification.
- An "autonomous, rational agent" that aims at neutralising an enemy position by choosing the target on which to fire a particular weapon or to drop a bomb, and does so with less collateral damage in terms of number of civilian casualties is deemed "more intelligent" than humans for this particular task of enemy neutralisation.

Obviously "perfect rationality" is seldom achieved (exploring all possible chess moves, one's own and those of the opponent is computationally unfeasible), so in AI one assumes "limited/bounded rationality" and an important research problem lies in designing good algorithms to cope with this situation of "bounded rationality". This understanding of "rationality" or "bounded rationality" has

been very useful to tackle many of the tasks for which AI techniques have been developed so far, and which I mentioned at the beginning of my talk: aircraft navigation, stock trading, medical diagnosis, home appliance control, logistics, consumer recommendation, computer games, etc.

Assumptions of Modernity and of Industrial Societies

However, this notion of "rationality", and the notion of "autonomy" associated to it –namely that we freely choose the actions we want to do based on our "rational" decision– is largely based on the assumptions that have characterised Modernity, and which underlie the value system of industrial societies. An these assumptions are:

- that societies are formed by "agents": self-interested individuals or organisations with objectives that they can achieve by interacting with other agents ot acting on the environment in which these "agents" are situated.
- that these "agents" have well-defined preferences that can be quantified and compared
- that "agents" are autonomous and decide the action to be taken in a rational way, i.e., by maximizing utility and benefit
- that rationality is disembodied, i.e., not dependent on our bodily nature
- that knowledge is information created and stored in minds and in other human artifacts like stories, books, and internetworked computers.
- that knowledge is actionable information that forms the basis for our thoughts and actions.
- that knowledge is a commodity that one produces, gives and takes
- that quality can be measure and compared, giving rise to the notion of "excellence", i.e., when one's quality excels that of another

These assumptions are themselves grounded again on a series of conceptual metaphors that constitute the current dominant socio-economic model: the Rational Choice Theory.

Value Alignment Problem

What do these assumptions and conceptual metaphors underlying the notion of "autonomous, rational agent" say about AI's potential impact on human societies to the point of actually becoming a threat to humanity? Stuart Russell, the computer science professor at Berkeley and co-author of the earlier mentioned HuffPost article on the existential risk of AI describes the problem as follows:

- the utility function with respect to which a autonomous rational agent makes high-quality decisions may not be perfectly aligned with the values of the human race, which are (at least) very difficult to pin down
- any sufficiently capable intelligent system will try to ensure its own continued existence and to acquire physical and computational resources – not for their own sake, but to succeed in the assigned task, i.e. to achieve maximum utility

The goal of AI research has been sp far to improve decision quality (according to the autonomous, rational agent paradigm) irrespective of the utility function chosen. This can lead, says Russell, to the "King Midas" effect, where we get exactly what we ask for, not what we want. And here lies the danger. So, as the Harvard professor of Economics Sendhil Mullainathan said it: "We should be afraid. Not of intelligent machines. But of machines making decisions that they do not have the intelligence to make. I am far more afraid of machine stupidity than of machine intelligence.", he says. The computer scientist Pedro Domingos even goes a little bit further and says that "people worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world."

So, Stuart Russell correctly identifies that values need to be included in our research outcomes: AI technology should yield AI systems that are provably aligned with human values. He calls this the "Value Alignment Problem".

Barcelona Declaration

In a similar vain, in March, in Barcelona, a "Declaration for the Proper Development and Usage of AI" was issued as the result of a debate on the dreams and risks of AI. This declaration proposes a *code of conduct* for AI practitioners (people researching in, developing, and deploying AI systems):

- to be **prudent**: to take very stringent prerequisites before applying an AI-based technique
- to guarantee **reliability**: to devise adequate verification and validation methodologies that certify an AI applications before it is put into widespread usage
- to ensure **accountability**: to provide explanations of decisions that can be challenged by humans with reasoned arguments
- to call for **responsibility**: to make it obligatory clear whether an interaction originated form a human or from an AI system
- to implement **constrained autonomy**: to properly circumscribe autonomous intelligence systems so as that they are aligned with human values
- to rely on **humans**: to continue to teach, develop and exercise human expertise, because AI systems depend crucially on human intelligence

Stuart Russell's call for "provably beneficial artificial intelligence" and the Barcelona declaration proposing a "code of conduct" are valuable steps towards a more conscious development and deployment of AI systems. However, these proposals –and the actions taken so far that take them into account– still stem from the same assumptions and are framed within the same conceptual metaphors in which the AI research programme has been carried out so far. Because, if we are to address the Value Alignment Problem as put forward by Stuart Russell, staying with the conceptualisation of "intelligence" as "rational action", and "rational action" as "utility maximization", one would have to frame human values within the "Well-Being as Wealth" metaphor that is a constituent part of the Rational Choice Model. But, although we can provide conceptualisations of our human value systems in order to reason about them –and this is essentially what Ethics is about, and further, these conceptualisation won't be able to avoid metaphors and metaphorical entailments– values are really about what's in our heart, not what's in our head. Consequently, values need to be also approached at a phenomenological first-person experiential level of how we "feel" about values, and not only at a third-person conceptual level of how we "think" and "reason" about values.

If we are to take effective actions in order to address the Value Alignment Problem so as to address the potential risks of AI systems that might be harmful for human societies we first need to look through AI's current assumptions and underlying conceptual metaphors to assess if they continue to be apt for today's societies and ways of living. By continuing to operate within the same conceptual system and within the same system of values, we may make things even worse –albeit unintentionally and in good faith. And here is where contemplation kicks in.

The Journey of Inquiry

It has often been argued that progress in science and technology is putting us at odds with spirituality; that science and technology have displaced spirituality from the centre of our lives, and that we now entrust everything we belief and do to science and technology instead of the gods. This

is certainly true with traditional expressions of spirituality as we may encounter in the belief systems we have inherited from institutionalised, conventional religion. Depending on whom you ask, they either may think of this displacement as something positive, a progress for humankind, moving from irrational beliefs towards greater enlightenment; or they may think of it as something negative, a great loss for humanity as it removes from it the value, depth, meaning and significance of our existence as provided by religious frameworks. I think this positive vs. negative interpretation of the advancement in science and technology is misleading; because, actually, genuine scientific inquiry is fundamentally spiritual *'per se'*. Because to inquire is to open up to the unknown, to let curiosity and wonder drive this inquiry, and to allow oneself to be transformed by it; it emerges from being aware that we do not know and that we need to question.

Actually, that is what faith is all about. As the Boston College philosopher Richard Kearney stated: "For faith means knowing you don't know anything absolutely about absolutes." So, contrary to what one may think, faith is not about belief, nor is it about truth, certainty or dogma; faith is, ultimately, about inquiry, as there is no certainty, fixity or isolation in the universe. Hence our human journey of life, our spiritual journey, is that of inquiry, to be on the way of seeking; keeping the emphasis on the seeking and not on what is sought: it is the Way that is the goal.

This is the reason why genuine scientific inquiry is essentially contemplative at its very core:

- it is vocational, permeated by an attitude of awe and wonder
- it relies on open communication, sharing and trust, thus creating community and building upon it; scientific inquiry is done in communion
- it requires silence, attention, patience, care and respect; it cannot be rushed into
- it is gratuitous and loving
- it is transformational, it "purifies" our experience, freeing it from mythical (μῦθος) understandings of reality
- and it is profoundly humble

The displacement by scientific inquiry of conventional forms of spirituality is actually a form of spiritual growth: scientific atheism has been for many people, including myself, an important step for going beyond conventional understandings of spirituality. The philosopher Richard Kearney I quoted before coined the expression "anathesit wager" to give a name to this spiritual journey that makes us move away from theism (when we are attached to some conceptualisation of God) to atheism (when we "get rid" of God alltogether) to anatheism (when we return to God "after God" – the prefix 'ana-' of 'anatheism' meaning "back to"– but it's a returning on God's terms, so to say, having actually let go of "him" completely). Hence, the AI research programme —as any other scientific inquiry– is to be situated in this larger context of the anatheistic journey, which is, ultimately, a journey of inquiry, of opening up to the unknown, to the stranger, and allowing oneself to be humbled and to be transformed.

Scientific Revolutions

About hundred years ago Sigmund Freud indentified three important scientific revolutions that have changed substantially the understanding of ourselves:

1. Nicolaus Copernicus revolutionised our understanding of the cosmos: he showed that we are not in the centre of the universe.
2. Charles Darwin revolutionised our understanding of life: he showed that we are not the pinnacle of the biological kingdom
3. And then Sigmund Freud himself claimed that he revolutionised our understanding of consciousness: he showed that we are not the masters of our mind, of our consciousness

These three revolutions have progressively dismantled the anthropocentric understanding we had of reality, displacing humanity from the centre to the periphery of the universe. And by humbling the human being, these revolutions have at the same time enlarged humanity's creative freedom, because it was only by exerting this creative freedom of inquiry that humanity was able to address and look for answers to the permanently open question of "Who am I?"

The Oxford philosopher of information Luciano Floridi has recently argued for a fourth revolution:

4. He claims that Alan Turing revolutionised our understanding of intelligence, of rationality: he showed that we do not have exclusivity of rational thought. He claimed that computing machinery could also think.

We probably lack the historical perspective to judge if we are indeed going through a new scientific revolution of the same significance and scope as the three previous ones. But there is some truth in the fact that AI research is indeed transforming our understanding of cognition. Our progress in AI research is making more evident which cognitive functions of our mind can or cannot be characterised in terms of information processing and thus can or cannot be mechanised. As a result, AI research is driving us to revise the dominant notion of "rationality" –rationality conceptualised metaphorically, according to the Rational-Actor model, as utility maximisation – and it is even driving us to revise the claim that it is "rationality" what distinguishes us as human beings… now we now that machines are better at being "rational" in this particular sense of the term.

Scientific revolutions, such as these four ones just mentioned, should be evaluated by their contribution to the creative freedom of all beings. And here lies the critical situation humanity is facing know, and which is indeed a matter of life and death. The current danger we face is to forget the metaphorical nature of our conceptualisations of "mind", "intelligence" and "rationality" and to take them literally, thus not seeing the metaphors anymore and endowing our scientific models with ontological reality. When this happens, our models cease to be only descriptive and predictive, and they threaten to become prescriptive: if intelligence *is* computation and to reason *is* to compute, then it is only a small step to attempt to take any process that requires to reason –e.g., medical diagnosis or sentencing in trials or personnel selection– and to standardise such processes in terms of an algorithmic step-by-step procedures whose outcome can be suitably measured and compared in order to achieve improved performance and to strive for excellence.

As we unconsciously "rationalise" our human activities in terms of algorithmic computation –due to our inability to see the metaphor anymore– we humans unnoticeably adjust our decisions and actions to these "rationalised" processes, without questioning the aptness of the metaphor. And since machines are better (and faster) at being "rational" (as conceptualised metaphorically), humans can be easily replaced by computing machinery, and it would even be considered unethical not to do so:

- if an AI-based diagnosis system performs medical diagnosis with a much lower error rate than a human physician, ¿shouldn't we not hand over the diagnosis task to it?
- if an AI-based autonomous vehicle transports goods and people with a much lower rate of accidents than a human lorry or taxi driver, ¿shouldn't we hand over the driving task to it?

Techno-Science and Freedom

Now, this creates the following paradox:

- during the last three centuries techno-science has become the prevailing instrument and symbol

of human liberation, providing more and more creative freedom to individuals and societies in deciding how they wanted to conduct their affairs in life: in the economy, government, law, healthcare, education, etc.

- but, by "rationalising" our processes due to the advancement of science and technology, and of computing technology in particular, we are removing the creative freedom originally granted to the decisions taken by traders, policy makers, judges, physicians, teachers, … as we straightjacket them within algorithmic procedures.

Science and technology's contributions should ultimately be evaluated, not in terms of utility only, not in terms of economic growth, not it terms of maximisation of some performance measure; they should be evaluated in relation to how they help and enlarge, and not impede or supplant, our creative freedom – and not only of people, but of all creation. And this also applies to how we approach AI research.

If we stay within the current framework inherited from industrial societies, centred on material production and utility, governed by a producer-consumer economy, with well-being conceptualised as wealth and measured in terms of monetary value; and if we stay within the conceptualisation of AI in terms of "autonomous, rational agents" that maximise expected utility, then AI-based technology may continue to provide benefit only to a minority of the human population and continue to strengthen current power distributions; and it may continue to nourish the producer-consumer economy which is not leading to liberation but to slavery. The AI research programme, and Luciano Floridi's fourth scientific revolution I mentioned earlier, can thus become the best ally of a societal model driven by Rational Choice Theory and by computational metaphors of rationality in which freedom is further and further curtailed. Are we not already starting to notice its effects?

An example is the trend towards "algocracy" –the rule of algorithms, i.e. the increasing reliance of decision-making processes on algorithms. Stock trading, and the financial market in general is almost exclusively governed by algorithms, and I don't need to remind you how the current financial system in our world has curtailed individual and social liberties as it led us into a global financial crisis. Another more local example would be algorithm-governed policing. In Chicago, police were using algorithm-generated 'heat maps' to identify people who were most likely to be involved in a shooting. However, a recent study indicated that the use of such maps did not reduce crime, but instead led to overpolicing of marginalized communities and to an increase of the likelihood that certain people would be targeted by the police and arrested after a shooting case. This is also curtailing our freedom.

Nourishing freedom is a sign of spiritual growth, as freedom itself is a spiritual dimension. We actually may see Freedom as a secular symbol of the Spirit. So, how can we make Freedom shine through AI research –as well as any other research endeavour in techno-science– as to make this research a true witness of humanity's journey of inquiry to fullness of life?

Information and Boundaries

Scientific inquiry is the sort of inquiry that leads us to a way of knowing that is based on information; it's a utilitarian form of knowledge. This is why scientific inquiry is also the driving force of our technology. Technology is a fairly new term for what until the 20th century was known as the "useful arts". Information itself arises from our capacity to establish distinctions on our experience. Notice that we live in a world of opposites. Opposites organise the physical space we move in:

- up / down
- inside / outside
- big / small
- here / there
- left / right

Opposites also provide content to our beliefs:

- true / false
- appearance / reality
- being / non-being

And they also structure our values:

- good / bad
- pleasure / pain
- freedom / bondage

Where do all these opposites come from? Contemporary cognitive scientists claim that our embodied experience with the world creates in our cognitive unconscious certain pre-linguistic structures called "image schemas" with which we construct our conceptual systems. For instance, when we were babies and we interacted with our environment putting things into others things, hiding them in boxes, grasping them again to take them out, etc. our cognitive unconscious constructed a schematic image of what containment in general is about: A boundary with its interior and its exterior. In subsequent stages of our cognitive development, we use the basic cognitive structures acquired from our sensorimotor experience –such as CONTAINMENT– and use them to structure more abstract concepts –such as 'categories'.

From our experience of observing that things that share common properties tend to be in the same bounded region (trees in a forest, grapes on vines, herds of animals, flocks of birds, …) we structure the abstract concept of 'category' using the CONTAINMENT schema, so we "put in" the category of 'grapes' what we identify to be particular grapes, while those things that we do not identify as grapes "lie outside" the category of 'grapes'. And so we classify things as being inside or outside the categories we create: 'horses', 'trees', 'birds', etc. And the further we grow in our cognitive abilities we start to create more and more abstract categories, but ultimately structured by our COINTAINMENT schema acquired as little infants, which establishes a boundary between interior and exterior. And so arises the world of opposites: big/small, true/false, good/bad, pleasure/pain,…

This binary logic of CONTAINMENT and the notion of classification into categories lies at the heart of the notion of 'information' and subsequently of the knowledge we gain by scientific inquiry. To inform is, literally, "to put into form", where the Latin word "forma" means a contour, a boundary. And the word "science" itself derives for the Latin "scindere", to cut, to divide.

Opposites and their Gestalt Structure

Image schemas, however, such as COINTAINMENT, form what is called a Gestalt structure:

- there is no interior if it isn't conceptualised with respect to an exterior and a boundary that divides interior and exterior;
- the same holds for the exterior, which does not exist without an interior and a boundary
- and whenever there is a boundary, we create also an interior and an exterior

This Gestalt structure transfers to the opposites we have conceptualised in terms of CONTAINMENT, and so there is no pleasure without pain and no pain without pleasure, as there is no goodness without badness and no badness without goodness. Now, as I said, an image schema such as CONTAINMENT is a pre-linguistic structure of our cognitive unconscious. When speech kicks in, then something very wonderful happens, which gives us our great power.

Speech enables us to give names to our categories and to the opposites we've created, and this allows us to evoke, with the simple utterance of a word such as 'pain' the entire universe of experiences we have classified as painful. But this capacity also alienates us of each individual particular unique moment of experience, the experience prior to its categorisation into the 'pain' category and outside the 'pleasure' category. And this our capacity of naming, in turn, gives us an illusion of separation, because we can now further "reason" with the 'word' that names a particular category, constructing arguments that only make use of the word 'pleasure', without ever explicitly mentioning the word 'pain'. And this, in turn, let's us think that we can get rid of one of the halves. This "manipulability" that speech provides us has been so successful for the flourishing of our species, and has granted us so much power, that we take the world as we came to know it *as real*, this world as it arises from this "in-formation" and subsequent "manipulability".

"What is real is informational and what is informational is real", as professor Floridi has put it.

Half of It

But this is only half of it. The power we have gained by categorising and naming comes with a loss of immediacy; the knowledge we gain by this "in-formation" comes at the loss of the "un-in-formed" immediate experience. Paradoxically, though, by virtue of the same capability we have to categorise and name we also gain the awareness that boundaries with its interiors and exteriors and the opposites they create are not intrinsically out there in the world; they are actually only in the eye of the beholder. And so is information. Our capacity of speech is thus a double-sided sword:

- on one hand it cuts the world in opposites, and makes us think the world is as we know it –full of dualities;
- but on the other hand it makes us aware that what we know is *relative* to our bodily experience shaping our cognitive unconscious, so as to make us realise that reality has also an "unknowable", an undivided, an *absolute*, nondual dimension.

But none of these two worlds –the relative and the absolute– is "more" real than the other: it is the same world, it is the same reality; not one, but not two either. It is not that there are aspects of the world we can know and others that we cannot know. No. All reality and every aspect of it can at the same time be known and be unknown.

The Two-fold Path of Inquiry

That's why scientific inquiry is itself only half of it. Scientific inquiry is intrinsically based on boundaries and categorisations. It creates dualistic knowledge. It is based on information. And it yields knowledge that can eventually be put into use through technology. Scientific inquiry is also the driving force leading to progress. For, 'progress' requires categorisation as it is basically moving away from what we categorise as 'bad' towards what we categorise as 'good'. So we progress by pursuing one half of the pair of opposites –health, pleasure, beauty, wealth– by attempting to eradicate the opposite half –sickness, pain, ugliness, poverty– driven by the system of opposites that constitute our value system. But it is this intrinsically dualistic nature of scientific inquiry that which renders it ill-suited for providing adequate orientation and motivation to the inquiry it is

conducting.

Because the system of opposites that constitutes the value system driving our scientific research, and the conceptual metaphors upon which our conceptualisations are founded, can only be themselves inquired into by "looking through" the illusion the opposites, by transcending them. This is ultimately what it means to be liberated; for freedom is the unmediated, direct apprehension of the absolute dimension of reality, the reality that is uninformational, undivided, unknowing, nondual, beyond opposites. Now, scientific inquiry will obviously continue yielding "advancements" in medicine, agriculture and technology; and these are going to be "advancements" in the sense that they are motivated and driven by the opposites that constitute the value system that make them be "advancements" from one half towards the other.

Human life needs opposites to find orientation. And we need to foster scientific inquiry to advance in this our life orientation. But scientific inquiry needs to go hand-in-hand with the sort of inquiry hat transcends the illusion of boundaries and the opposites they create, freeing us from our attachment to them, not basing our happiness and fullness of life on these boundaries. I shall call this kind of inquiry 'nondual' or 'advaitic' inquiry or, to use a term of the Christian tradition, 'kenotic' inquiry, stemming from 'kenosis', which literally means 'emptyness', i.e., kenotic inquiry as an "emptying" inquiry.

Kenotic Inquiry for Scientists

And how does such kenotic inquiry look like? Well, we don't need to "invent" this sort of inquiry from scratch. We can walk in the footsteps of those that preceded us in this journey of inquiry, in the same manner we do with our scientific inquiry. Any form of contemplative prayer is kenotic inquiry.

Meditation is probably the most obvious one. The silencing of our thoughts, images, and desires by the faithful repetition of a mantra, or the mindful attention to the breath, makes the contemplative, nondual dimension of reality shine through all our activities. My personal experience with my daily practice of meditation is that it has put me again in touch with the contemplative core of scientific inquiry:

- with the sense of awe and wonder driving it,
- with the need for silence, attention, care and respect
- with the love for the subject under investigation;
- with the pleasure of sharing ideas with colleagues;
- with the humility that scientific inquiry nourishes.

So, scientists and engineers would benefit greatly in the daily task of conducting scientific research and technological development if they integrate these tasks with a daily practice of meditation. It is ironic, that this sort of inquiry, kenotic inquiry, which some orthodox scientists consider "brainless" is actually that which is most free and creative, and which is an essential feature of our humanity. But not only scientists and engineers would benefit. Humanity would be enriched by the outcomes of contemplatively-driven science.

In addition to meditation here are a few other examples of contemplative practices that scientist and engineers could (and should) do in order to conjoin scientific inquiry with kenotic inquiry:

- C**ontemplative reading and writing**: The pressures for maximising number of publications and number of citations in science is so high, and the amount of papers that one needs to read and review is so large, that scientists have lost touch with the value gained by working through a text slowly, allowing times for reflection and silence in between – much in the same sense as a "lectio

divina". And also of writing slowly; journaling is a good contemplative practice: to be brief but to write daily, and to cultivate "taciturnitas", an attitude to keep silent and only to express that which is really necessary to communicate, and then to do it with a loving attitude. I like very much the Gaelic proverb that says "Abair ach beagen is abair gu math e" – say but little and say it well.

- **Deep listening and beholding**: My experience in scientific conferences is that there is not very much listening happening anymore. While the speaker goes through his or her slides, most of the audience is busy with their laptops, either answering email or working on their own slides for the talk they have to give a few minutes later. Scientists are forgetting to listen to each other attentively, respectfully. In our labs we should practice attentive listening: going regularly to the exercise of explaining each other our research and our open questions, and to practice to repeat as closely as possible what our colleague has said, until he or she feels truly heard. And also to learn to look at the "big picture". We scientists are often too worried with working out the small detail that we do not "let sink in" the big picture. To sit in front of our subject of research and to dissolve the subject-object boundary so as to let the immediate experience enrich the scientific investigation; in order to complement in this way our objective detachment required for carrying out the more analytical tasks.

- **Welcoming the stranger:** To cultivate intellectual hospitality is also an important contemplative practice in scientific research. St Benedict in his Rule says that "all guests who arrive be received like Christ" – scientist should be welcoming to new ideas that initially might look strange; Benedict further says that "in the reception of the poor and of pilgrims the greatest care and solicitude should be shown, … for as far as the rich are concerned, the very fear which they inspire wins respect for them." So it is easy to accept and praise the work of some famous researcher or of a team working at a famous university or research lab, but it is to the less known scientist to which we need to put more care and attention.

Hence humanity's journey of inquiry is, for it to be whole, a two-fold path of scientific and kenotic inquiry, of knowing and unknowing, itself a pair of opposites that has to be transcended as well –not two, not one.

Back to Ramon Llull

This brings us back to the Ramon Llull, the mysitc which is our "grandfather" of artificial intelligence. Many scholars of Ramon Lull create a divorce between Lull's emphasis on rationality and the subsequent development of his Art to mechanise reasoning on one hand; and his mystical experience and calling that he felt to apply his Art to the dialogue and argumentation with Muslims in order to convert them to Christianity on the other hand. So one group has focussed on Llull's logical system, independently of his mystical experience, rendering it unnecessary for evaluating his contributions to the mechanisation of thought; and others have focussed mainly on his significant contributions to Catalan literature and to Christian mysticism.

But Llull himself did not separate rationality –the domain of scientific inquiry and dualistic knowing– from the direct experience of the divine –the domain of kenotic inquiry and nondual unknowing. Because, in order to construct the whole edifice of rational reasoning aimed at the task of conversion, he started with that which was shared by Christians and Muslims alike, namely the direct experience of the divine, and which both traditions expressed by way of attributes associated to this direct experience: goodness, greateness, eternity, power, wisdom, will, virtue, truth, and glory. His entire Art would have been only abstract symbol manipulation, had he not grounded it on this direct apprehension of the Absolute. So reasoning and rationality only had a chance to convince and convert if they were ultimately linked with our inmost experience of reality.

And his inquiry was a two-fold inquiry of what he called “sciència” and “amància”, and he devised methods for both – the “ars inventiva” or “finding art” and the “ars amativa” or “loving art.” While one guides the intellect to the truth, the other gives orientation to it, by guiding the will towards what is good. Llull explains how these inquiries are intimately paired, warning us not to consider these methods as belonging to two different realms of endeavour.

As a mystic he was able to look through the boundaries into the nondual dimension of reality. And so I conclude with a quote taken from the most famous of his mystical writings, the *Book of the Lover and the Beloved*:

> “Say, O Fool, what is love?” He answered: “Love is that which throws the free into bondage, and to those that are in bonds gives liberty.” And who can say whether in love there is more of liberty or of bondage?