AI*IA Workshop on Deep Understanding and Reasoning: A Challenge for Next-generation Intelligent Agents – URANIA Some Concluding Remarks

Luigia Carlucci Aiello

Department of Computer, Control, and Management Engineering Antonio Ruberti
University of Rome La Sapienza, Rome, Italy
aiello@dis.uniroma1.it

When the organizers of URANIA contacted me with the invitation to join the program committee of this Workshop I gladly accepted. I was in fact intrigued by the title, the aims and scope, and even more by the list of research questions they posed in the call, as all of them brought the attention to research on the Science of AI more than on the AI-Technology.

The message conveyed by **the title** sounded to me as: when (almost) everybody speaks about "Deep Learning" let us reflect on "Deep Understanding and Reasoning". Does it mean that we do not believe in deep learning and statistics as the solution to all the problems of AI, even nowadays when Big Data are available? Yes, of course, this is the case. Even though the results achieved with Big Data and Deep Learning are in some cases astonishing, we are sure we still need symbol crunching and reasoning to build artifacts that show a human level intelligence.

It is not a surprise that research in AI (and research in general) goes through periods of fascinations and dismissals of topics, methods or approaches, passing through severe "religious" wars. At present statistical reasoning seems to be the answer to all the AI problems and is expected to significantly contribute to the advancement of technology in the near future (see [5]). But remarks about the limits and fears that it may even bring us into a new AI winter are already popping up.

Hence, while the world (both scientific journals and newspapers) resounds with the slogan "Big Data and Deep learning" it is good that somebody raises the flag of "Deep Understanding and Reasoning", not so much to have two parties in a "competition", but because we do believe that the limits of Big Data and Deep Learning can only be overcome with better Knowledge Representation and Reasoning. Hector Levesque with his IJCAI-13 talk and successive AI journal paper [3] points strongly out the fact that there are patterns of human reasoning not captured by statistical reasoning, and that, conversely, need reasoning based on knowledge, sometimes of the kind John McCarthy used to call common sense. Others [1, 2] note that despite the unprecedented successes reached with Deep Learning, nothing we see resembles reasoning at the human level or conclusions that can be explained to a human user. We all know that, in order to accept the interaction and the "suggestions" that come from our AI-artifact we want to

understand what her/his/its conclusions are and why she/he/it came to them. This requirement is more stringent as the applications become critical.

The vision of the proposers presented in **the aim and scope** of the Workshop is for completely autonomous systems that should go from the formulation to the solution of problems without human intervention. Hence, quoting the call for papers:

> In a long-term vision, next-generation artificial cognitive systems and robots will be autonomous end-to-end solvers that perform the whole problem-solving process without any human intervention. Starting from a (possibly multi-modal) problem description, an end-to-end-problem solver should automatically understand the problem, identify its components, devise a model, select a solving technique, and find a solution. Such autonomous intelligent agents should be pro-active and problem-solving driven; deep understanding and deep reasoning, not necessarily based on big-data, will be a crucial ingredient for their design.

Even though the full autonomy is not necessarily the ultimate goal, as the vision of a symbiosis between humans and intelligent artifacts is a more realistic scenario, at least to me, the challenges raised by the above vision are not to be underestimated. The machine, in order to be able to symbiotically cooperate with humans, has to build models of the word, reason, and autonomously find solutions to problems in a way that can be communicated to, and understood by, humans.

Even more intriguing, the aims and scope paragraph concludes with the following:

> In this context, it would be important to identify specific challenges, to assess the level of autonomy achieved, the effectiveness of end-to-end solvers, and to ease the dissemination of AI results to a general audience. This ambitious goal requires an unprecedented integration of AI area and techniques such as Natural Language Processing, Machine Learning, Constraint-based reasoning, Logic, Planning, Case-based reasoning, Human-Machine Interaction, and Cognitive Science, and could represent an important step forward reducing the fragmentation of modern AI.

Reconciliation of different approaches and reduction of the fragmentation of AI are indeed a valuable perspective.

As for the **list of questions** posed by the organizers, it is well conceived, precise and exhaustive. From the long list, I choose the following:

- How far is a next generation of AI engines going beyond the pure question answering by embedding deep reasoning and understanding?
- How to address reasoning aspects typical of human problem solving such as metaphorical reasoning, abstraction, creativity, and intuition?

- On the road to autonomy, which is the role played by human-computer interaction in the problem solving activity along the line of AI collaborators?
- Is this the time to go "Beyond the Turing Test"? What about the risks for humanity related to these next-generation intelligent agents?

I am not a fan of the Turing test, so I read this last question as: We need metrics to evaluate the performances of the AI-artifacts. This entire item is very relevant and should be taken into account by the researchers in AI. We know that the risks of the research in AI are not at all at the level depicted by some of the media. It is our moral duty to develop robust and reliable AI systems, it is our moral duty to communicate this to the non-specialists. I see a risk: ignorance about AI can induce fear, hence become an obstacle to future research. We have to be aware and contrast this risk.

As for the other points I selected above, I consider them the "core questions" of the Workshop. All of them have been addressed in various ways, and at various levels, by the contributors, with their papers, presentations and participation in the final discussion.

Occasionally during the presentations at the Workshop, something brought back to my mind some piece of "old literature" or some old, famous and even abused example, that I believe are still relevant and should be proposed to the young researcher, in particular the paper [4] published in 1969 by John McCarthy and Patrick Hayes. The solutions presented there may be too simplistic and now obsolete (if not in the spirit certainly in the details, as they have been immensely refined by successive research), but the problems and the issues raised are still up in their full glory. In particular two things come to my mind: the three criteria set forth by McCarthy and Hayes, and the problem of the mutilated chessboard.

Paper [4] provides us with a definition of intelligence:

> On this basis we shall say that an entity is intelligent if it has an adequate model of the world (including the intellectual world of mathematics, understanding of its own goals and other mental processes), if it is clever enough to answer a wide variety of questions on the basis of this model, if it can get additional information from the external world when required, and can perform such tasks in the external world as its goals demand and its physical abilities permit.

In paper [4] McCarthy and Hayes speak about three criteria to evaluate the adequacy of a representation: metaphysical, epistemological and heuristic. I here report only the definition of these criteria as exposed in [4], and point to that paper for their illustration:

> A representation is called metaphysically adequate if the world could have that form without contradicting the facts of the aspect of reality that interests us.

> A representation is called epistemologically adequate for a person or machine if it can be used practically to express the facts that one actually has about the aspect of the world.

A representation is called heuristically adequate if the reasoning processes actually gone through in solving a problem are expressible in the language.

When designing a representation we should ask ourselves these questions, in particular the heuristic adequacy. To this end, I'll use the same example John McCarthy often used, namely the "mutilated chessboard". This is a tiling puzzle introduced by the philosopher Max Black in 1946 that goes as follows:

"Take a standard 8x8 chessboard and remove two diagonally opposite corners. Is it possible to cover it with standard 2x1 domino tiles?"

The problem might appear very complex, but it becomes trivially simple if we consider that:

– a chessboard has 32 white and 32 black squares;
– by removing two opposite corners we remove two squares of the same color, say black.

So, there is no way of covering 32 white and 30 black squares with tiles, as each tile can only cover two adjacent squares, hence one black and one white square.

Of course a heuristically adequate knowledge representation for this problem must allow us/a machine to reason about numbers, oddness and evenness, equality, etc., rather than relying on colored pixels or Cartesian coordinates of the squares.

When the organizers of the Workshop invited me to draw some conclusions, chair the final discussion and then write this note, I accepted with some hesitation, knowing that it was going to be a challenging job. I have enjoyed the Workshop, the reader of these proceedings can easily see why, and I am looking forward to its continuation because it gathers researchers that are asking themselves fundamental questions for the advancement of AI.

References

1. Castelvecchi, D.: Can we open the black box of AI? Nature 538(7623), 20–23 (10 2016)
2. Kaplan, J.: Artificial Intelligence: Think again. Commun. ACM 60(1), 36–38 (2017), http://dl.acm.org/citation.cfm?id=2950039
3. Levesque, H.J.: On our best behaviour. Artif. Intell. 212, 27–35 (2014), http://dx.doi.org/10.1016/j.artint.2014.03.007
4. McCarthy, J., Hayes, P.J.: Some Philosophical Problems from the Standpoint of Artificial Intelligence. In: Machine Intelligence. pp. 463–502. Edinburgh University Press (1969)
5. Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., , Teller, A.: Artificial Intelligence and Life in 2030. Tech. rep., One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA (September 2016), http://ai100.stanford.edu/2016-report