

BARCELONA DECLARATION FOR THE PROPER DEVELOPMENT AND USAGE OF ARTIFICIAL INTELLIGENCE IN EUROPE

VERSION 1
8 March 2017

This text is the outcome of a B·Debate on 8 March 2017 at CosmoCaixa, Barcelona.
B·Debate is a joint initiative of Biocat and “la Caixa” Foundation.
www.bdebate.org/ca/debat/artificial-intelligence-next-step-evolution

BARCELONA DECLARATION FOR THE PROPER DEVELOPMENT AND USAGE OF ARTIFICIAL INTELLIGENCE IN EUROPE¹

Artificial intelligence (AI) is a collection of computational components to build systems that emulate functions carried out by the human brain. The field started in the mid-nineteen fifties and has since gone through cycles of promise, enthusiasm, criticism and doubt. At this moment we see a strong wave of enthusiastic adoption of AI in many areas of human activity.

We distinguish between knowledge-based AI and data-driven AI.

- Knowledge-based AI, which has become applicable in the late seventies, attempts to model human knowledge in computational terms. It starts in a top-down fashion from human self-reporting of what concepts and knowledge individuals use to solve problems or answer queries in a domain of expertise, including common sense knowledge, and then formalizes and operationalizes this knowledge in terms of software. Knowledge-based AI emphasizes conceptual models, ontologies, common sense knowledge bases, reasoning and problem solving strategies, language processing, and insight learning. It rests primarily on highly sophisticated but now quite standard symbolic computing technologies and has already had a huge impact, for example, on the semantic web.
- Data-driven AI, also commonly known as machine learning, became applicable only the last decade. It starts in a bottom-up fashion from large amounts of data of human activity, which are processed with statistical machine learning algorithms, such as the deep learning algorithms, in order to abstract patterns that can then be used to make predictions, complete partial data, or emulate behavior based on human behavior in similar conditions in the past. Data-driven AI requires big data and very substantial computing power to reach adequate performance levels.

Knowledge-based AI has shown to be most successful in intellectual tasks, such as expert problem solving, whereas data-driven AI is most successful in tasks requiring intuition, perception, and robotic action. The full potential of AI will only be realized with a combination of these two approaches.

The current surge of interest and application of AI is without precedent. There is a growing consensus that AI is of huge importance for the future economy and functioning of human societies. AI is now understood to be a powerful, novel way to link producers and consumers, and a novel way to add value to products, build new ones, and improve production processes. Moreover, AI can help to introduce more efficiency and quality into bureaucratic procedures and give greater access to knowledge and creativity for all.

We believe that AI can be a force for the good of society, but that there is also concern for inappropriate, premature or malicious use so as to warrant the need for raising awareness of the limitations of AI and for collective action to ensure that AI is indeed used for the common good in safe, reliable, and accountable ways.

¹ Revision by Luc Steels after B-debate on 8/3/2017.

The present declaration is a step in this direction. It proposes elements towards a *code of conduct* for AI practitioners (both AI developers and users) in Europe.

1. PRUDENCE. The leap forward in AI has been caused by a maturation of AI technologies, vastly increased computing power and data storage, the availability of delivery platforms through the internet, and an increased willingness of many economic actors to try out the technology for their own application domain. But we must be aware of the limitations. Many fundamental problems of artificial intelligence are not yet solved and will require radical breakthroughs, and perhaps some issues will never be solved. The application of AI therefore demands that practitioners take very stringent prerequisites into account; otherwise the results could be disappointing and potentially very harmful. For example, the application of knowledge-based AI requires the availability of human expertise and sufficient resources to analyze and model it. The application of data-driven AI requires enough high quality data and careful choices of which algorithms and parameters are appropriate in each case. These application prerequisites need to be investigated and spelled out in much more detail so that those responsible for applying AI can exercise the necessary prudence.

2. RELIABILITY. All artificial systems that are used in our society have to undergo tests to determine their reliability and security. So it is normal that the same is done for AI systems, particularly in domains like medicine or autonomous robots. Although verification and validation procedures have been developed for knowledge-based systems in the nineteen-eighties and nineties, they are still lacking for data-driven AI. Sure, at the moment machine learning practices make a distinction between an example data set used for training and a test set used to gauge in how far a system has reached adequate levels of performance, but there is still a significant difference between a test set and actual testing in real world conditions. Moreover, once adequate verification and validation methodologies are available, we will need a network of agencies in European countries (or a central European agency) that use them. They should become the authority to certify AI applications before they are put into widespread usage. The European Parliament has recently decided to create an agency for robotics and AI which could potentially take up this task.

3. ACCOUNTABILITY. When an AI system makes a decision, humans affected by these decisions should be able to get an explanation why the decision is made in terms of language they can understand and they should be able to challenge the decision with reasoned arguments. This is particularly important in domains such as decisions on loans, legal decisions (for example about granting parole), insurance, taxation, etc. Many AI systems, particularly those based on data-driven approaches, are currently unable to provide this kind of explanation. Their decisions are based on a large set of statistically derived network parameters. Techniques to make sense of these parameters are in their infancy and will probably require a combination of knowledge-based and data-driven AI. But accountability is clearly a precondition for sound deployment.

4. RESPONSIBILITY. There is a growing worry about AI chat-bots and other kinds of automatic messaging systems operating on the Internet and in social media,

designed for the manipulation of political opinion, disinformation through the propagation of false facts, extortion, or other forms of malicious activity that is dangerous for individuals and destabilizing our society. These chat-bots pretend to be human and do not give away the identity of those behind them. The use of AI has made these chat-bots sufficiently realistic that unsuspecting users are not able to make a distinction and get misled. A possible solution to this issue could be to make it obligatory clear whether an interaction originates from a human or from an AI system, and that, in the case of an artificial system, those responsible for it can be traced and identified. This solution could possibly be implemented by a system of *water marking* and become mandatory in Europe.

5. CONSTRAINED AUTONOMY. AI systems have not only the capacity to make decisions. When they are embedded in physical systems, such as self-driving cars, they have the potential to act upon their decisions in the real world. This understandably raises questions about safety and about whether autonomous AI will not at some point overtake humans. Although some of these worries belong more in the domain of science fiction than reality, the proper circumscription of autonomous intelligent systems and their alignment to human values is an important challenge that must be addressed. It is necessary to have clear rules constraining the behavior of autonomous AI systems, so that developers can embed them in their applications. It is also necessary to clarify who is responsible for failure - as is indeed the case with all products.

6. HUMAN ROLE. The current undeniable enthusiasm for AI gives sometimes the impression that human intelligence is no longer needed and it has led some organizations to fire employees and replace them by AI systems. This is a very serious mistake. All AI systems critically depend on human intelligence. Knowledge-based systems model the knowledge and insight of human expertise and data-driven AI systems rely critically on data of human behavior. It follows that human expertise should continue to be taught, developed and exercised. Moreover, in almost any area, human expertise still far outstrips artificial intelligence, particularly for dealing with cases that have not appeared in the example data sets from which AI systems are learning, and often real benefit comes from the synergy between human and artificial intelligence.

Given the important benefit of AI and their growing importance in maintaining a competitive economy, we call upon European funding agencies and companies to invest in the development of AI at a scale which is adequate for the challenge, and in such a way that ALL European regions and citizens can profit. This investment should strive towards the creation of a complete ecosystem with a network of high end research labs with sufficient structural (as opposed to project-based) funding, diffusion of AI techniques to form a significant number of AI practitioners, and proper conditions and stimuli for successful AI entrepreneurship. Of particular importance is the development of open resources, such as corpora, ontologies and software frameworks, which should be available as a common infrastructure on which specific applications get built.