DEBUNKING SOME OF THE MYTHS SURROUNDING ARTIFICIAL INTELLIGENCE

BY STAN MATWIN[1]

Following a condensed history of Artificial intelligence the paper presents the personal views of the author about the common, somewhat pessimistic perspectives on Artificial Intelligence, encountered often in the media and supported by some visionaries.

INTRODUCTION

Artificial Intelligence (AI) is today on everybody's minds and lips. It has become part of the vocabulary of scientists, but also of engineers, physicians, business people, politicians, the media and visionaries. AI today attracts attention, money and people to an extent unprecedented in its short history. While I do remember meetings of neural networks researchers attended by some 200 participants in the early 2000s, the same meetings today sell 6000 registrations in a matter of minutes, or even offer a lottery to thousands of people willing to travel around the world to attend. Figure 1 illustrates this interest surge graphically by showing Google searches for the terms "Artificial Intelligence" and "Machine Learning" since 2012. Starting in 2015 the gradient of the latter is much sharper. A lot of this popularity is due to the obvious attractiveness of the idea of "automating intelligence". This current interest is due, at least in part, to a number of myths and misconceptions that have been created and supported by people from outside the field. This brief article attempts to de-bunk four such generally held and common beliefs present in the public sphere. The views presented below, while based on almost 40 years of experience in different areas of AI, represent a purely personal perspective many would not agree with. I believe, however, that in order to progress as a field, we need to be capable of introspective reflection.

SUMMARY

This brief article presents author's views about some of the claims concerning Artificial Intelligence, appearing often in the media and supported by certain visionaries. It debunks four such myths, based on misunderstanding and unwarranted extrapolation of the current technical developments in the field.

Stan Matwin
<stan@cs.dal.ca>
Faculty of Computer Science and Institute for Big Data Analytics, Dalhousie University, Halifax, Nova Scotia
B3H 4R2

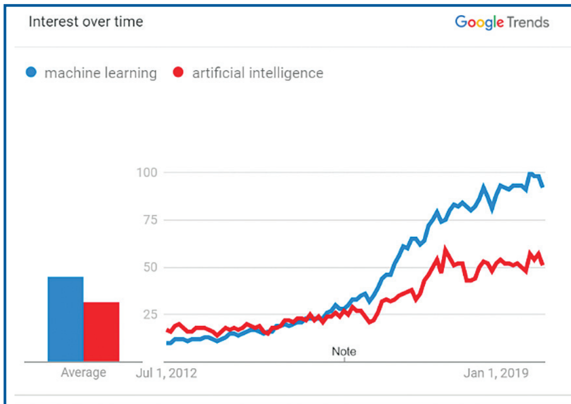

Fig. 1 Trends in Google searches for the terms "Machine Learning" and "Artificial Intelligence" since 2012. The y axis is the Google Search Volume Index, Google's proprietary measure for comparing relative popularities of search queries.

A VERY BRIEF HISTORY

AI, a child of the Cold War, was born in a DARPA workshop at Dartmouth College in 1956, where the late John McCarthy from MIT coined the term "Artificial Intelligence". To this day there is no general agreement on the definition of AI, but the one used operationally is that it is an area of science "researching and building systems capable of intelligent behavior". From its early days the field of AI consisted of a number of sub-fields. Knowledge representation and reasoning, natural language processing, machine learning (ML), computer vision (CV), and planning were the main sub-areas of AI. These fields were relatively disjoint and worked on different problems. It was quickly realized that almost all AI problems are intractable, or NP-complete: any algorithm to solve a particular problem would, for the hardest dataset for this problem, be no better than trial and error. The researchers therefore worked on heuristics that would not guarantee such optimal solutions but would nevertheless produce results close to optimal, and be sufficiently efficient to work on larger and larger data. Many of the successful solutions and progress milestones of the first twenty or

1. Also affiliated with the Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland.

Fig. 2 The wolf vs husky training set. From [4].

thirty years of AI were based on the community efforts that resulted in the development of specialized resources: large dictionaries summarizing a wealth of knowledge about languages (especially English), or benchmarking datasets allowing ML researchers to compare their algorithms in a methodological manner. Inadequate computational resources often stood in the way of progress, e.g., in artificial neural networks. When specialized "Lisp machines" and "Prolog machines" failed to solve AI in the mid-1980s, the second "AI winter" took place, lasting until the mid-1990s. Then, in the case of ML, robust ML methods slowly morphed into data mining, and advances in the theoretical foundations of ML (e.g., Support Vector Machines or boosting) put the field on a stronger footing. A breakthrough took place in 2012, when Geoff Hinton and colleagues from the University of Toronto showed [2] how their deep learning architecture could beat state-of-the art computer vision algorithms on the standard AlexNet image classification dataset by more than 5%, where standard incremental progress of the CV field over the years was on average 1% per year. People started paying attention to "deep learning" (many other elegant and powerful deep learning architectures, algorithms and representations ought to be at least mentioned here, e.g., autoencoders, embeddings, generative adversarial networks, etc. See [1] for an excellent, condensed introduction to Machine Learning with deep artificial neural networks). Moreover, advances in hardware democratized high-performance computing after 2010 through the availability of "for rent" cloud-based, parallel computing environments on the one hand, and off-the-shelf cheap GPUs processors on the other hand. Highly parallelizable deep learning computational tasks became solvable on generally available, inexpensive computing platforms. Ease of sharing data (and code) through the internet and managing it methodically with the use of open-source, high quality database software were other factors contributing to the Big Data revolution (more on the relationship between Big Data and AI below). AI researchers from all its subfields turned to Machine Learning for addressing specific tasks in their research. By 2018 the fields of NLP and CV became infused by ML, and planning has become by and large deep reinforcement learning. Machine Learning has taken over AI, and has attracted hundreds of thousands of young, creative minds from around the world. Investors are lining up to fund promising AI companies. Governments pour hundreds of millions of dollars, euros and yuans into AI research institutes[2]. Further progress in inevitable, but as argued below it might not be linear.

2. Canada, for once, is at the forefront of AI research. Among the three funding fathers of the Deep Neural Networks and Deep Learning: Y. Bengio, G. Hinton and Y. LeCun (recipients of the 2018 ACM Turing Award, generally believed to be the Nobel Prize of Computer Science), two (Yoshua Bengio at the Université de Montréal, and Geoff Hinton at the University of Toronto) are Canadian. Their research has survived all the AI winters in part due to the Canadian Institute for Advanced Research (CIFAR) that has steadily funded AI research in lean and fat years.

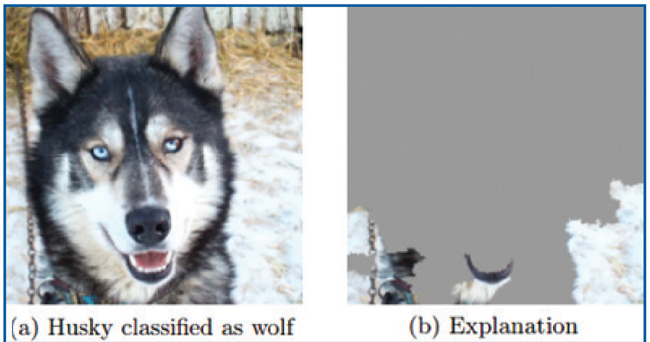

Fig. 3 (a) A new example (a husky) misclassified as a wolf and (b) the explanation (a "saliency map") for this classification. From [4].

MYTH I: AI CAN SOLVE ANYTHING

Hardly a week goes by without media headlines about new achievements of AI, how the discipline can solve problems until now tackled only by humans. "Intelligent Machines are Teaching Themselves Quantum Physics", "Artificial intelligence better than humans at spotting lung cancer" or "How AI understands passengers' emotions for in-car safety systems" are just a few examples. Such titles are often true only in a narrow sense: while the problem is general, the AI solution refers to a particular, simplified rendering or aspect of the problem, reducing it to, e.g., an image recognition or text classification task. Image classification and clustering is the area in which the advances of "new AI" are probably the strongest. However, as pointed by Melanie Mitchell in her recent book [3], the Convolutional Neural Network (CNN) algorithms at the basis of many such successful applications do not really understand the images the way humans do. Instead, they find hidden, often complex (non-linear) combinations of features whose presence or absence in a given image is characteristic of its "class" (e.g., presence vs absence of quantum phase transition in an image, or telling apart an x-ray of a healthy vs sick patient, or recognizing a dog's breed in a photo). The intrinsic lack of understandability of such "classification" is a major shortcoming of these solutions, especially if they were to be used to make decisions about humans. As an example, let us consider an experiment in which a CNN was trained on a dataset of images of dogs, in which each image contained either a wolf or a husky. The "training" images of wolves also contained, on purpose, snow in the background (Fig. 2). When the result of Machine Learning — a trained CNN — was asked to classify a husky image in the centre of Fig. 3a, it declared it to be a wolf. As for an explanation, the system pointed to the snow in the picture: in fact, in the training set there was an overwhelming evidence of snow in all the wolf pictures (Fig. 3b), and therefore the CNN targeting recognition of wolf images was optimized to detect large white areas, without learning anything about dogs. This is what we mean by the correlational nature of Machine Learning. As long as a task can be reduced to identifying patterns in data, preferably continuous, "smooth" data such as images or sound, and there is a very large set of annotated "instances" whose classes are known, we can obtain a good solution with modern Machine Learning. But trusting this solution, e.g., with a patient diagnosis, incarceration decisions or a school admission policy recommendation based on the student's expected academic performance is a different question. AI systems outperform humans in tasks which are often associated with a "high level of intelligence" (e.g., playing chess, or GO), but are not anywhere near human capacity in other tasks in which humans are very good without any training (e.g., telling jokes). It is because we all have an enormous knowledge base, known as "common sense", which we are still unable to circumscribe, let alone codify and feed into AI systems. Attempts in that direction have been made over many years, but are generally believed to come far short of the expected results (e.g., the CYC system, www.cyc.com). AI will continue solving some difficult problems better than humans, but is a long way from solving others at which even children excel.

MYTH II: AI WILL SURPASS HUMAN INTELLIGENCE BY 2045

It has been predicted by eminent contributors to AI that systems "surpassing humans" will take place before 2050. Ray Kurzweil's "singularity" prediction from 2005 said that in 2045 "machine intelligence will be infinitely more powerful than all human intelligence combined" [5]. Yet this prediction is based on the belief that the current growth of AI, characterized as exponential, will continue as such for the next 25 years. This is highly unlikely, as barriers will almost certainly arise on the way. One such barrier is the complexity of AI systems: some of the modern networks trained on very large data contain billions of parameters. The complexity of these systems, and a lack of understanding of the adverse interaction of their features, will make it very difficult to engineer larger systems from function-specific components. It is difficult to see how such components can be assembled and connected without understanding how to set them up for a given task. A lack of any serious results in the standardization of fundamental tools that advanced AI uses (e.g., data representation and description languages) is another likely barrier. Finally, a scarcity of the annotated ("labeled") data necessary to train the "supervised" (i.e., the most powerful) ML algorithms is another. While the Big Data movement brought focus to the questions of collection, management and analysis (using ML) of large, heterogenous and constantly growing data, the question of annotation of massive (order 10^6 or 10^7 instances) datasets is still open. A practical solution used today is the global crowdsourcing of data annotation, with its non-scaling cost, difficult quality assurance issues and the unsolvable questions of hidden cultural and demographic biases in annotating certain types of data. There is also the difficult question of data ownership: much of the promising data that, combined, could potentially lead to breakthroughs in a number of society-level issues belongs to several major players known as GAFA (Google, Amazon, Facebook, Apple) and is treated as a proprietary asset by these organizations. Given all these issues and the lack of a convincing perspective for addressing them in the fragmented, competitive data ecology, makes it almost certain that hurdles will appear in the growth path of AI. Even if some of the challenges outlined in Myth I were solved, others will no doubt appear. No technology has continued its progress forever without reaching a plateau at some point. The exponential growth of AI will not continue eternally, and as it is the central argument for surpassing human intelligence, singularity is not certain.

MYTH III: AI WILL HARM HUMANS

The fear of AI revolting against its human creators has long left the safe territory of science fiction and has been repeatedly brought up by thinkers and visionaries. On the one hand there are no technical or scientific arguments substantiating these beliefs. Some such doomsday scenarios have been presented by people who are thought leaders in science or human history — e.g., Stephen Hawking and Yuval Harari, but do not necessarily have in-depth, technical understanding of AI as it is today. On the other hand, strong technical arguments have started appearing that argue the impossibility of such a "robot revolt". In particular, a recent book [6] by Christof Koch, Chief Scientist and President of the Allen Institute for Brain Science argues that it will not be possible to construct artificial (computerized) conscience. Koch presents an analytical, scientific argument about the impossibility of "artificial consciousness". For robots to differentiate themselves from humans, let alone attack them, self-conscience would be necessary. This is not to say that we do not need to think about the ethics of AI — we do, because AI systems make decisions that concern practically all of us, and that will grow even further. But from seeing a path from transparency of specific decisions to autonomous intelligence that may evolve the goal of harming humans is a far fetched conclusion.

MYTH IV: AI WILL ELIMINATE JOBS AND MAKE HUMANS SECOND-CLASS CITIZENS

This myth has different versions. In some, it is similar to the luddite movements from the 19th century where people were destroying textile machines because they were threatening pre-industrial revolution jobs. In its much more refined version, discussed e.g., in Harari's 2017 book "Homo Deus" [7], humans will become inferior members of a society, with all the higher-level intellectual decision-making powers reserved for AI systems. I have a different view of the effects of AI on human work. It is clear that some very common jobs will most likely be eliminated by AI within the next 10 years. According to Statistics Canada, driving a truck for a living is the second most frequent occupation among men in Canada, and yet most of these jobs will be likely replaced by autonomous vehicles before the decade is out. Other jobs, e.g., even some positions in the legal profession, are also likely to either be eliminated or greatly scaled down. But does that mean massive unemployment and the universal need for a Guaranteed Basic Income? I believe that new jobs, which we cannot even imagine and articulate today, will appear, just like we could not predict in 1980 that hundreds of thousands of people will make a living from adding value to the internet (e.g., webpage and app design), because the internet concept or even the name itself did not exist at that time. Similarly, new technologies will appear and will create new kinds of jobs. It is often raised that these jobs will require much higher levels of math and science than those existing today. That is most likely true, but I believe that significantly raising the level of universal training in mathematics and science is not impossible. Most of the population was illiterate before the industrial revolution, but once this revolution happened and literacy of the workforce has become a must, societies have been to be able to build educational systems in one generation. This is the challenge we are facing today. And here is also a fascinating opportunity for AI. Artificial Intelligence may be a major tool and enabler in creating better ways of individualized, engaging and thorough ways of training our youth in mathematics, physics and science through simulation, visualization and interaction, and one-on-one conversations, with the student making the material relevant to their personal interests. In that way, the jobs that AI will take away will be replaced by new jobs that AI will train people for. There is no reason to believe that humans will be degraded to intellectual slaves.

CONCLUSION

The four myths discussed above are far from an exhaustive list. I do not pretend that my answers to these myths are complete, as any opinion based on a prediction with significant uncertainty, may turn out true, or perhaps not. Many others remain — e.g., are AI systems capable of true creativity, in the sense of inventing new concepts as described in Kuhn's seminal book "The Structure of Scientific Revolutions"? Our society needs to discuss these issues in a perhaps more thorough, systematic manner than is the case today. More participation by scientists, particularly from the field of AI, is needed. "Aucun n'est prophète dans son pays", but we need to try.

REFERENCES

1. Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, *Nature*, **28;521**(7553), 436-444 (2015).
2. A. Krizhevsky, I. Sutskever, G. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, NIPS 2012.
3. Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*, Farrar, Strauss and Giroux 2019.
4. M. Ribeiro, S. Singh, C. Guestrin, *Why should i trust you?: Explaining the predictions of any classifier*; Procs. Of KDD 2016, pp. 1135-1144.
5. R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, 2005.
6. Christof Koch, *Feeling of Life Itself": Why Consciousness is Widespread but Cannot Be Computed*, MIT Press, 2019.
7. Y. Harari, *Homo Deus: A Brief History of Tomorrow*, Random House, 2016.