Editor-in-Chief's Introduction to the Special Issue on "Superintelligence", AI and an Overview of IJCAI 2017

This editorial consists of two parts: first, an introduction to the Superintelligence special issue and, second, the traditional AI and IJCAI overview, which this year has been a little delayed.

1 Superintelligence special issue

Being Editor-In-Chief means many tedious hours of rapid proof-reading for all the papers, several times, as well as choosing interesting special issues and writing editorials. One of the great joys of the editorial work is to see an excellent special issue delivered, like this one on superintelligence. Let me shed some light on the editorial part of the special issue.

First of all, with help from colleagues at the Future of Life Institute we came across Roman Yampolskiy, born in Latvia and a graduate of the University of Buffalo. His book "Artificial Superintelligence: A Futuristic Approach" (Figure 1) represents a more technically oriented viewpoint than Bostrom's philosophical "Superintelligence: Paths, Dangers, Strategies". He is an associate at the Global Catastrophic Risk Institute, a think tank that analyses global risks to the survival of human civilization. Whatever the case, Roman gave a major boost to the superintelligence special issue.

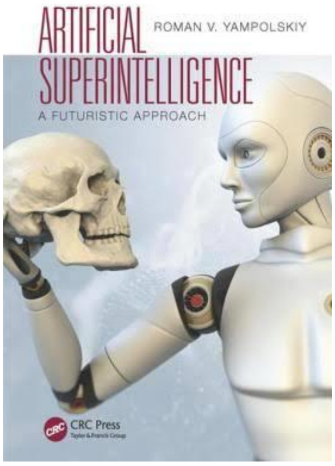


Figure 1: Yampolskiy's book on Artificial Superintelligence.

Three more special editors joined in the following weeks: Nell Watson, Matthijs Maas and Ryan Carrey.

Nell Watson is an engineer and futurist thinker, a kind of Northern Irish Ray Kurzweil, the latest probably the best-known singularity enthusiast and forecaster. She also advises The Lifeboat Foundation, helping to pinpoint the trajectories of human progress. She works at the Singularity University and the United Nations University.

Matthijs Maas is a PhD Fellow in Law and Policy on Global Catastrophic and Existential Threats at the University of Copenhagen's Centre for International Law, Conflict and Crisis (CILCC). He holds an M.Sc. in International Relations from the University of Edinburgh. His research interests include the safe governance of artificial intelligence and the effects of emerging technologies on strategic stability, amongst others. He is also a Junior Associate of the Global Catastrophic Risk Institute.

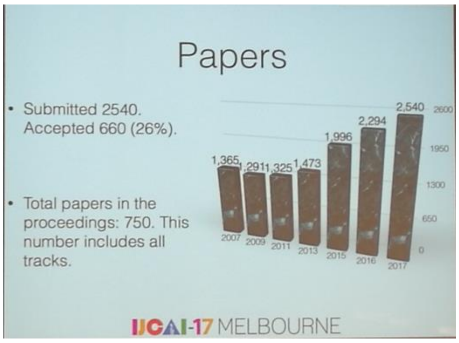
Ryan Carey is a research intern in AI safety at Ought Inc and a research affiliate at the Centre for Study of Existential Risk. He edited the Effective Altruism Handbook, a compilation of essays about how to do more good with limited resources. He also founded the Effective Altruism Forum and cofounded Effective Altruism Melbourne.

While the quality of any journal issue comes down to the authors of the papers, the editors of the special issue deserve particular attention as well.

Last but not least, let me thank Tine Kolenik, a student that helped me with this special issue, and Drago Torkar, technical editor of Informatica, for special efforts with the editorial system.

2 AI and IJCAI 2017

The progress of artificial intelligence (AI) is certainly fast and furious from the technical point of view. Each year there are scores of new achievements in academia, gaming, industry, and real life, having implications for the way we live and work. For example, autonomous vehicles are improving constantly, and are being introduced into more and more countries. Recently, IJCAI presented its annual general overview of the AI SOTA and progress.


Papers

- Submitted 2540. Accepted 660 (26%).
- Total papers in the proceedings: 750. This number includes all tracks.

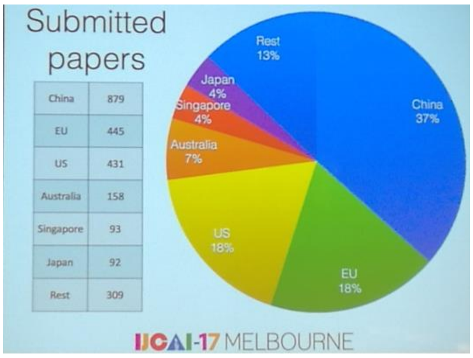
IJCAI-17 MELBOURNE

Figure 2: Increase in the number of IJCAI papers in recent years.

The 26th International Joint Conference on Artificial Intelligence was held in Melbourne, Australia in August 2017 [6]. Melbourne has been judged the world's most liveable city for the seventh year running and indeed it is safe, clean, uncrowded, full of green nature and architectural wonders. It is a prosperous city that hosted a prosperous scientific event!

The growth in AI is indicated by the number of papers submitted to the IJCAI conference (Figure 2). In

2016 in New York there were 2294 papers, while in 2017 in Melbourne, 2540 papers were reviewed. The growth has been steady since 2009.



Submitted papers

China	879
EU	445
US	431
Australia	158
Singapore	93
Japan	92
Rest	309

Figure 3: Papers per country at IJCAI 2017.

A study of the papers submitted per country (Figure 3) at IJCAI 2017 indicates that the majority was from China (37%), with the EU second (18%) and the US third (18%).

Although there was a lack of Eastern-European papers and papers from Russia, on September 1 Vladimir Putin, speaking with students, warned that whoever cracks AI will 'rule the world' [9]. Will that be China, as it already submits the most AI papers? Among others, at least 600 top Chinese ministers and military officials use quantum-encrypted links for all confidential communication. But the current leader is still USA with most of the awards given to the USA researchers at IJCAI 2017.

While the number of AI papers from China might come as a surprise, their industry achievements are astonishing as well. One might not be as familiar with the Chinese solutions as with Google or Amazon AI systems, but the Chinese systems are close to the top. For example, in 2017 China's Alibaba Group Holding Ltd introduced a cut-price voice-assistant speaker, similar to Amazon.com Inc's "Echo". It is called "Tmall Genie" and costs $73, significantly less than the western counterparts by Amazon and Alphabet Inc's Google, which cost around $150. Similarly, Baidu, China's top search engine, recently launched a device based on its own Siri-like "Duer OS" system. Alibaba and China's top tech firms have ambitions to become world leaders in AI.

In 2017, two games stood out as another example of AI beating the best human counterparts: unlimited Texas hold'em poker (10 on 160 possibilities) and Dota 2. Both games were slightly limited – in poker, there are only two players instead of more, and Dota 2 was also reduced to only two players instead of 10. Nevertheless, both games are the most-played human games with award funds going into the tens of millions. Both games are quite different from formal games like chess or Go. For example, poker includes human-bluffing interactions and hidden cards. Dota 2 is another surprise since it resembles fighting in the real world, although everything is more of a fantasy story. The key components were strategic plans with global and local decision making, and adapting to the adversary. From Wikipedia: "Dota 2 is originally played in matches between two teams of five players, with each team occupying and defending their own separate base on the map. Each of the ten players independently controls a powerful character, known as a "hero", who all have unique abilities and differing styles of play. During a match, the player collects experience points and items for their heroes in order to successfully fight the opposing team's heroes, who are doing the same. A team wins by being the first to destroy a large structure located in the opposing team's base, called the "Ancient", which is guarded by defensive towers."

Regarding the methods, reinforcement learning and deep neural networks were the most commonly applied; however, the AI field at IJCAI 2017 was presented for more than 10 major areas.

Various types of deep neural networks (DNNs) continue their excellence in visual recognition tasks and in real-life diagnostics, such as diagnosing which tissue contains malignant cancer cells. When fed with huge numbers of examples and with fine-tuned parameters, DNNs regularly beat the best human experts in increasing numbers of artificial and real-life tasks, like diagnosing tissue in several diseases. There are other everyday tasks, e.g., the recognition of faces from a picture, where DNNs recognized hundreds of faces in seconds, a result no human can match. Figure 4 demonstrates the progress of DNNs in visual tasks: around 2015 the visual recognition in specific domains was comparable to humans; now, it has surpassed humans quite significantly – again, in particular visual tests. BTW, DNNs are currently breaking the CAPTCHA test – the simplest way so far to differentiate between SW agents and humans.

The effects of only visual superiority are astonishing on their own, but several services emerge from visual analyses. For example, eye analyses make it possible to detect certain diseases like cancer or Alzheimer's [3]. Furthermore, DNN studies of facial properties can reveal sexual orientation, IQ, and political orientation. When shown five photos of a man, a recent system was able to correctly select the man's sexuality 91 per cent of the time, while humans were able to perform the same task with less than 70% accuracy [7]. This Stanford University study alone confirmed that homosexuality is very probably of genetic origin. The consequences of a single study can be profound. Will job applications also be assessed by a DNN study of facial properties? Will dictatorships prosecuting homosexuality punish their citizens on the basis of their faces?

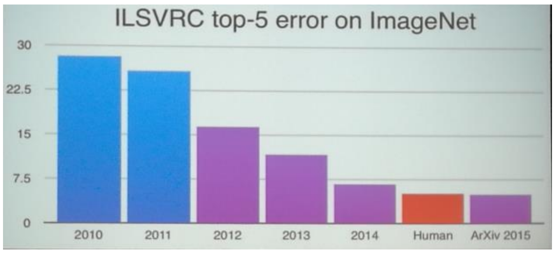


Figure 4: Error of DNNs on ImageNet over time.

There were several demonstrations and competitions at IJCAI 2017, including the traditional Angry Birds' competition. Most attractive, however, were soccer competitions with off-line Nao robots that were not trained or advised as a team, but performed on their own in a group decided on the spot. Unfortunately, the local computing powers are at the level of a mobile phone, which means they are insufficient for good play. The robots were often wandering around, searching for the ball. Still, they demonstrated some quite cunning skills, e.g., accurately kicking the ball into the goal using a specific angle relative to the foot.

Next year will be of particular interest. ICML with 3500 attendees, IJCAI+ECAI with 2500, AAMAS with 700, ICCBR with 250 and SOCS with 50 attendees will be hosted in Stockholm in a 2-week event in July 2018. Optimistic jokes are emerging that the critical mass of 6–7000 attendees will be sufficient to ignite general intelligence or even a path to superintelligence [2, 8, 10].

As part of the tradition, the IJCAI overview avoids mentioning people's names; however, there is one person that deserves special attention – Toby Walsh, Local Arrangements Committee Co-Chair, a key organiser of the letter "Killer robots: World's top AI and robotics companies urge United Nations to ban lethal autonomous weapons" and organizer of Melbourne's public Festival of Artificial Intelligence. Congratulations!

There are two additional matters worth mentioning: the ban on autonomous weapons and the Asilomar principles.

2.1 Ban on autonomous weapons

There are two major reasons for the proposed ban:

- Fully autonomous weapons will likely make war inhumane, whereas humans – if war cannot be avoided – need some rule of engagement to preserve some level of humanity and prevent human suffering being too extreme.
- This is one of the preconditions on the road to prevent superintelligence from going viral and malignant [2, 8, 10].

There is good reason for celebrating the first successes of the pro-ban efforts – the movement is spreading through social media since it started years ago by scientists like Toby Walsh or Stuart Russel and is currently coordinated by Mary Wareham.

Slovenia is involved in the ban at the European and national levels, where four societies (SLAIS for artificial intelligence, DKZ for cognitive science, Informatica for informatics, ACM Slovenia for computer science) drew up a letter and sent it to the UN and Slovenian government, and recently the Slovenian AI society SLAIS wrote a letter to the European national communities to join activities in this direction. Our initiative was also debated at the European AI society EurAI meeting at IJCAI 2017.

Second, Elon Musk and the CEOs of 155 robotic companies signed a letter in which they say "Once developed, lethal autonomous weapons will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend. These can be weapons of terror, weapons that despots and terrorists use against innocent populations, and weapons hacked to behave in undesirable ways."

"We do not have long to act. Once this Pandora's Box is opened, it will be hard to close."

On the other hand, the world's superpowers are rapidly not only developing, but also applying autonomous weapons, from drones to tanks or submarines. Some even argue that it is already too late to stop these autonomous weapons.

Another example: the EU parliament accepted new legislation giving artificial systems some of the rights of living beings. This is exactly one of the rules of thumb that should not be done to avoid potentially negative AI progress. So why did EU politicians accept such a law? It is not dangerous yet, but clearly worrisome.

2.2 The 23 Asilomar principles

The Future of Life Institute's [4] second conference on the future of AI was organized in January 2017. The purpose of this section is to introduce, in the rather original way, the 23 Asilomar AI principles [1] defined at the BAI 2017 conference.

The opinion of the BAI 2017 attendees and the world-wide AI community is widely held: "a major change is coming, over unknown timescales but across every segment of society, and the people playing a part in that transition have a huge responsibility and opportunity to shape it for the best." Therefore, a list of Asilomar principles was designed to provide directions for future AI research.

The first task of the organizers was to compile a list of scores of opinions about what society should do to best manage AI in the coming decades. From this list, the organizers distilled as much as they could into a core set of principles that expressed some level of consensus. The coordinating effort dominated the event, resulting in a significantly revised version for use at the meeting. There, small breakout groups discussed subsets of the principles, giving detailed refinements and commentaries on them. This process generated improved versions of the principles. Finally, they surveyed the full set of attendees to determine the level of support for each version of each principle.

After this time-consuming and meticulous process, a high level of consensus emerged around many of the statements during the final survey. The final list retained principles that at least 90% of the attendees agreed on. The 23 principles were grouped into research strategies, data rights and future issues including potential superintelligence, signed by those wishing to associate their name with the list. The principles with additional interviews can be obtained from the web pages of the event at the Future of Life Institute [4]. The principles will hopefully provide some guidelines as to how the power of AI can be used to improve everyone's lives in future years.

AI has already provided useful tools that are employed every day by people all around the world. Its continued development, guided by the following principles, will offer amazing opportunities to help and empower people in the decades and centuries ahead.

3 Conclusion

AI's progress is both fascinating and accelerating. An increasing awareness of AI-related changes in human society is being recognised by the scientific, academic and general public. Dozens of major reports have emerged from academia (e.g., the Stanford 100-year report), government (e.g., two major reports from the White House), industry (e.g., materials from the Partnership on AI), and the non-profit sector (e.g., a major IEEE report). The special issue on superintelligence will hopefully spur discussion and awareness among the public, media and government, helping them to understand that the times are changing rapidly, and that new approaches and methods are needed for humans to successfully cope with the future.

On the other hand, AI stubbornly lacks general intelligence and other human properties like consciousness. There are specific claims that current computers do not provide the kind of computing that can emulate the best human intellectual properties [5]. The Turing test remains as a mission impossible for even the most advanced systems. It may be that we do not need to worry so much about overall global superintelligence bypassing humans in every category, but to focus on the technical progress of AI and its applications.

Scientific understandings about AI, its influence on everyday life, and the future of human civilization are stacking up. Scientists are able to provide some guidelines about which direction we humans should develop AI to avoid the dangers of the negative effects of the rising power of AI. While AI often frightens the general public, I, and several AI researchers, find its rapid progress a necessity to prevent the decline or even the self-destruction of human civilization. These potential dangers are real, not fictitious, primarily because of the simple fact that any major power can be easily misused to cause harm to humans, and second, there are some strong indications that civilizations tend to destroy themselves (the Fermi paradox). By raising awareness, we increase the chances to reap the positive aspects of an amazing future AI and avoid the negative ones.

References

[1] Asilomar principles. 2017, (https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/).

[2] Bostrom, N. 2014. Superintelligence – Paths, Dangers, Strategies. Oxford University Press, Oxford, UK.

[3] Eye Scans to Detect Cancer and Alzheimer's Disease, https://spectrum.ieee.org/the-human-os/biomedical/diagnostics/eye-scans-to-detect-cancer-and-alzheimers-disease

[4] Future of life institute, https://futureoflife.org/

[5] Gams, M. 2001. Weak intelligence: through the principle and paradox of multiple knowledge. Nova Science.

[6] IJCAI conference, 2017, https://ijcai-17.org

[7] Kosinski, M., Wang. Y. 2017. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. https://osf.io/zn79k/

[8] Kurzweil, R. 2006. The Singularity Is Near: When Humans Transcend Biology, Sep 26, Penguin Books.

[9] Mail online, Science and technology, Vladimir Putin warns whoever cracks artificial intelligence will 'rule the world', http://www.dailymail.co.uk/sciencetech/article-4844322/Putin-Leader-artificial-intelligence-rule-world.html

[10] Yampolskiy, R.V. 2016. Artificial Superintelligence. CRC Press.

Matjaž Gams