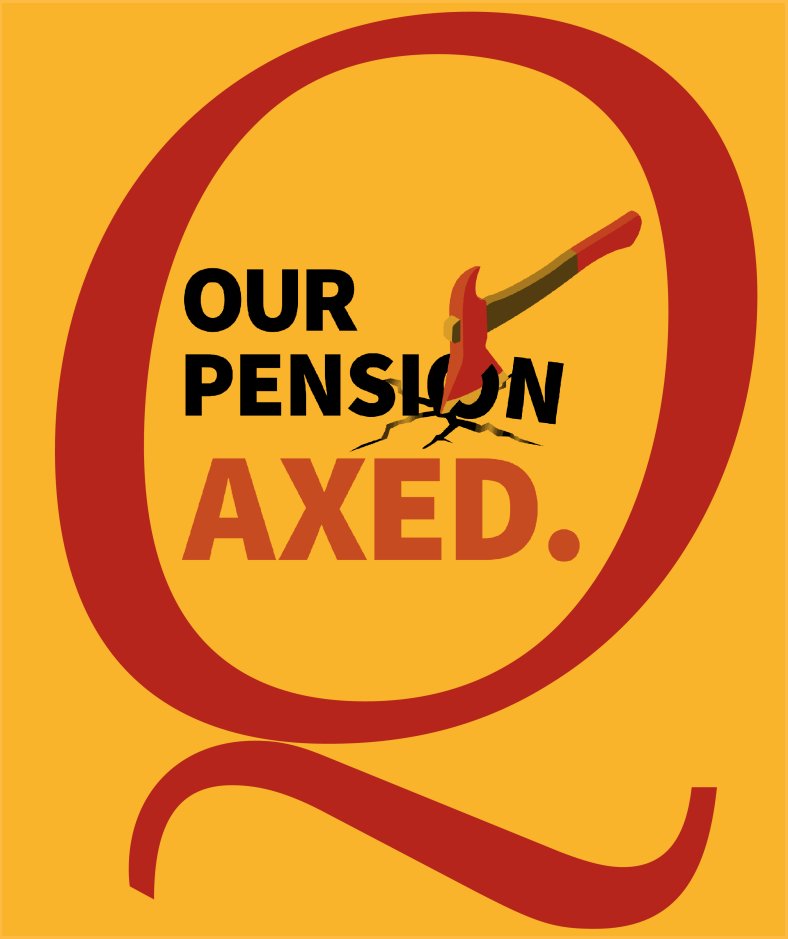
AISB Quarterly

THE NEWSLETTER OF THE SOCIETY FOR THE STUDY OF
ARTIFICIAL INTELLIGENCE AND SIMULATION OF BEHAVIOUR



No. 149 March 2018

About the Cover

This cover highlights the decision taken by university staff and members of the University and College Union (UCU) who are taking strike action to defend their right to a fair pension, and to respond to university employers who want to end guaranteed pensions and reduce retirement income for all.

The image used in the cover is taken from the widely used UCU strike poster.

Feeling geeky and arty?

If you are interested in designing a cover with the help of your off-the-shelf AI-boosted algorithms, feel free to contact the editor on `aisbq@aisb.org.uk` with your cover design (taking into account the already "set on stone" orange shade of the cover) along with a blurb on how you managed to get to the final results.

Editorial

The first few months of 2018 have been rich with new and interesting AI-related news, such as the use of AI by UK police forces to inform custodial decisions, AI-powered retina analysis, ISIS's ability to bypass UK's AI-based propaganda blockade, AI-generated fake pornography and much more. On the other hand, the government's continuous lack of clarity on Brexit leaves higher education and academics in limbo when it comes to continental collaborations, with potential side-effects such as France being picked for Google's new research centre dedicated to exploring AI techniques applied to health and the environment, as well as doubling the size of existing AI lab in Paris, with UK still awaiting its "slice of cake". Recently, AISB chair represented the UK at a European Commission meeting on the AI landscape in Europe. A report on this event will be published later this year.

While interesting and important, let's put the above on one side, as there has been another significant issue that kept UK's higher education institution preoccupied, to the otherwise unimaginable extent that many thousands of CPUs and GPUs have been left in cold offices for an extended period. Indeed, we are talking about the *strike*! University and College Union (UCU) received a strong mandate from its members to proceed with a strike action in 60+ universities around the UK in order to convince the employers, Universities UK (UUK) to consider the pension concerns of the university staff (academics, librarians and administrators). At the time of writing, the matter is far from resolved, and in fact some universities have decided to approach the matter by taking intimidating measures that could strengthen the resolve of their staff. In one such case, as reported in The Guardian, University of Kent informed its staff that any failure to reschedule lectures or classes affected by the strike would result in losing 50% to 100% of their pay "for every day where an individual continues to refuse to perform their full contract of employment". In another case, University of Oxford blocked an attempt by staff to challenge pension cuts. Arcane procedures were used to see the objection of a only 20 members from a large crowd of academics resulted in halting move to debate proposals that sparked universities strike.

The radical restructuring of the pensions, as proposed by the employers, could see staff losing as much as £10,000 in pension annually in retirement. Given many of the best academic minds could decide to leave the sector and join the industry (despite their love for their research and students), it is a threat not only to the retirement age of the university staff but the livelihood and strength of the UK's higher education.

This year's AISB Convention is undoubtedly a space where many of us will be engaging with this topic, either formally through talks on the matter or informally while meeting colleagues over the social events or coffee breaks.

We hope to see more clarity on the matter when this issue of the Q reaches you. In this issue, in addition to some conference reports from our members, Poker is chosen as a case study to discuss the always captivating theme of human vs. AI. Furthermore, a personal perspective is provided on some of the existing chatbots, which would be of (historical) importance particularly to those following our annual Loebner Prize event at Bletchley Park. The call for AISB 2019 Convention Proposal is also included in this issue, along with a call for nominations of AISB committee members. As always, the last word is allocated to the very Fr. Aloysius Hacker.

Mohammad Majid al-Rifaie
Editor
@mohmaj

Goldsmiths, University of London
London, United Kingdom
March 2018

Humans vs AI: Poker is the latest Arena

Josh Wardini, josh@ileanmedia.site

Can artificial intelligence ever be smarter than humans? Will AI ever become conscious or pass as human? From the 'Chinese Room' to the 'Turing Test,' these are questions we've pondered for years. And while we're still a long way off from cyborgs living indistinguishably among us, AI is "smarter" than humans in a lot of areas – even if it's humans that have programmed the bots to begin with. The latest arena where machines are outperforming people is poker. As examined by The Rise of Machines against Humans, pokersites.me.uk, supercomputers programmed with 'machine learning' algorithms, such as Libratus of the Pittsburgh Supercomputing Center, have been able to defeat professional poker players at no-limit Texas hold'em and other variants of the card game.

High Profile Machine Poker Wins

Using machine learning techniques, after each hand, Libratus records various data points, learns from its mistakes and reapplies its strategy. Earlier this year it was able to rack up $1,766,250 worth of chips in a three day tournament, destroying professional players including Jason Les, Jimmy Chou, Daniel McAuley and Dong Kim. On average the bot won $14.72 per hand! Of course, that's easier said than done. Libratus is the result of millions of dollars in funding. It needs 274 terabytes of RAM to effectively operate its machine learning algorithm and roughly it's 30,000 faster than the average desktop computer. It's not the kind of tech we'll see in our homes just yet. What makes Libratus special is absence of a built-in poker strategy (i.e. if x happens do this), it creates its own ever-refined strategy by learning as it goes. It has been programmed how to do not what to do, and the more it plays the better it becomes, thanks ot exposure to more data. It has even learned to randomise certain actions to mimic human bluffing. It's not just Libratus that has been cleaning house against poker pros. As early as 2007, Phil Laak found himself down $1,500 in less than 200 hands to the Polaris bot developed by the University of Alberta. Its successor Cepheus is so advanced that it can never lose over the long-run. Human players have just a 5% chance of beating it after 30,000 hands. Interestingly, those that have faced it believe their own game has improved.

Humans Can Still Win

One pro player who likely added considerably to Polaris' knowledge base was Ali Eslami, who was over $800 ahead by the mid-point of a session against the bot in 2007 and finished $395 ahead. Even its replacement Cepheus has had a few hiccups along the way (even at 5% odds, it will still lose eventually). In 2015, Michael Bowling won two 100 hand matches, though it's generally accepted that this was luck rather than skill. While Claudico, which was designed by Carnegie Mellon professor

Tuomas Sandholm and his graduate students, has some impressive victories under its belt – in May, 2015, it took on Dong Kim, Jason Les, Bjorn Li, and Doug Polk in a series of heads-up matches and Polk managed to earn a $400,368 lead in the first week of the competition. In the end, Polk finished up $213,000, Li won $529,000, and Kim beat Claudico by $70,000. Only Les was down, with $80,000.

Why AI Has an Advantage

Poker is a very human game, which makes it both difficult for AI and easier for AI in certain areas. One thing's for certain, bots don't get tired and make irrational decisions because of it. Regardless of their overall skill, they always make the best decision they possibly can. Bots are more efficient as well. AI can carry out thousands of years' worth of human work in a week, and that's just down to computer processing. AI also doesn't have an emotional attachment to money that can cloud its judgements; it doesn't feel losses or wins of large amounts like humans do. This means it's not afraid to take risks or to go after only small amounts with a lot at stake. In fact having no emotions at all makes AI impervious to emotional elements of the game. A winning streak doesn't make it cocky; the anguished face of a bluffer is meaningless, other than as a data point for future hands.

Why AI Has a Disadvantage

Just as AI has some inherent advantages it also has some inherent weaknesses, especially in poker, which is a game of both skill and chance that involves human emotion. To begin with there is imperfect knowledge (data) for the bot to work with (i.e. opponents' hands are hidden), and it's different from a game like chess where the full board is in view.Other than learning certain patterns in behaviour, AI also struggles to recognise bluffing and integrating deceptive plays, because there are so many variables. Humans can read body language and also use it to their advantage.

The Future

Poker is interesting because it's complex and humans do have some inherent advantages, therefore poker could act as the perfect tool for improving AI. One can argue millions aren't being spent for "fun and games", and that the results would impact growing real-world applications. Think of all the scenarios where the ability to make calculated decisions and to continuously learn are important. Businesses and government are constantly planning, strategising and negotiating, but often fail due to human limitations – from greed to simply being poor at analysing data. What if a government could use a bot to accurately determine the best way to spend taxes? The cyber security industry alone, is predicted to spend $96 billion on machine learning by 2021. Instead of just responding to attacks after the fact and closing the breach, a sophisticated machine learning algorithm can spot an attack as it begins to happen and will learn from the new data.

Exactly what the future brings for AI is unclear, but it seems it's here to help us not overtake us.

`https://pokersites.me.uk/poker-ai`

Conference Report: Towards Autonomous Robotic Systems (TAROS) 2017

Robert H. Wortham **(University of Bath, *r.h.wortham@bath.ac.uk*)**

Starting on the 19th of July, The eighteenth annual TAROS conference took place this year at the University of Surrey in Guildford, UK. As usual the conference was well attended, with around 150 delegates, and there was a full programme over three days. Each day began with a topical keynote lecture, followed by paper presentations grouped by topic, and panel sessions for each group.

Alongside the main conference, a wide range of exhibitors including Airbus and the University of Surrey Technology for Autonomous systems and Robotics (STAR) Lab had stands within a dedicated exhibition space. Commercial robots such as the SoftBank NAO and Pepper were also being demonstrated by distributors.

At the close of the first day, there was a social 'Ice Breaker Reception' — an opportunity for us to network over canapés and a glass of fizz, followed by a fascinating IET public lecture on the ExoMars Rover given by Abbie Hutty from Airbus. The ExoMars rover is a robotic Mars rover, part of the international ExoMars programme led by the European Space Agency and the Russian Roscosmos State Corporation, scheduled to launch in July 2020. Once landed, the rover's six month mission will be to search for the existence of past or present life on Mars.

On the second day, an afternoon poster session of 25 posters ran within the exhibition space, followed by an exceptional conference dinner, complete with live entertainment.

An award ceremony on day three presented prizes for the best paper, student paper, poster and robotics demonstration. After lunch we headed home with new knowledge, ideas and more importantly new inspiration towards autonomous robotic systems.

Keynote Lectures

Nicolas Heess, Senior Research Scientist at Google DeepMind gave the first keynote lecture. He spoke about DeepMind's research into the use of deep reinforcement learning for robotic control problems. DeepMind want to solve intelligence problems generally to reduce the need for domain knowledge. Their focus is on reinforcement learning (RL), allowing designers to specify objectives not specific solutions. However the principal problems with RL for robotics are the high degree of freedom (DoF), and the many 'real world' constraints. Heess referenced Watkins Q Learning (1989) as their basic approach to the problem. He covered five examples of their research: simulated robots stacking lego blocks, using the Baxter robot to learn an assembly

task from human demonstration, moving over rough terrain with obstacles towards a goal, training in rich environments with different robot morphologies, and skill representation and reuse using motion capture and simulated robots. The most notable problem with this approach is the number of iterations required for the learning. Baxter required 75 human demonstrations of insertion of a spring clip into a slot, followed by 250 automated RK training events, in order to achieve a reliable trained system. The simulation environments needed tens or hundreds of thousands of training events, something not possible with a single physical robot, though maybe of interest with distributed learning across a robot swarm, or by mass produced commercial devices such as autonomous vehicles.

Professor Qingwen Wu from the Changchun Institute of Optics, Fine Mechanics and Physics (CIOMP), Chinese Academy of Sciences gave the second keynote. He presented their vision to build a ten metre diameter telescope mirror in space. Such a mirror is too large to be assembled on earth and carried in a rocket, and so must be assembled once in orbit. The complexity of the task to assemble the mirror requires an autonomous robot for assembly, and Professor Wu explained their design for a robotic arm with the aid of diagrams and a physical model. CIOMP are looking worldwide for partners for this project.

Rob Buckingham, Director of Remote Applications in Challenging Environments (RACE) from the UK Atomic Energy Authority gave the final keynote 'Towards zero manual intervention'. His talk covered the many projects underway at RACE, focussing on the need for fully autonomous, reliable robots to work in radioactive areas where humans cannot survive for more than a few moments or at the most minutes with protective clothing. One example is the torus of nuclear fusion reactors, which require regular maintenance but are not accessible for humans once they have been exposed to the radiation from the fusion reaction.

Main Conference Papers

Over three days there were 29 oral paper presentations, too many to review individually here, however all the papers are available in the Lecture Notes on Artificial Intelligence (LNAI) 10454 published by Springer. For presentation, papers were arranged into six groups:

1. Swarm and Multi-Robotic Systems
2. Human-Robot Interaction
3. Robotic Learning and Imitation
4. Robot Navigation, Planning and Safety
5. Soft and Reconfigurable Robots
6. Service and Industrial Robots

My oral presentation was included within Human-Robot Interaction. Written with Andreas Theodorou and Dr Joanna Bryson, our paper is entitled 'Robot Transparency: Improving Understanding of Intelligent Behaviour for Designers and Users'. The talk covered our research over three

years into the accuracy of the mind models that people create and use to understand robots when they encounter them, and ways in which we can improve those models through improved robot transparency. This research feeds empirical evidence into the wider debate around robot and AI ethics, regulation of AI and autonomous systems, and so on. The talk generated many questions from delegates, and at the subsequent panel session the focus of discussion was around these issues. It was valuable to interact with engineers and scientists normally concerned with the practical business of building autonomous robots, and there followed a discussion of the psychological, philosophical and ethical (moral) implications of their work. For me, what was most striking was how similar the questions and views were to those expressed by the general public, legislators and other technical and non-technical groups that I have spoken with on previous occasions. Similar in the sense that they were wide ranging without ethical consensus, often based on media and entertainment representations of AI, frequently concerned with the emergence of sentience, consciousness and free will, and sometimes desirous of granting rights to their creations. This reinforces my view that work on robot and AI ethics is important within academia, and the considered views of professional ethicists are vital both for the regulation of AI and robotics, and for the sound education of future generations designing and living with autonomous technologies.

Thank you AISB for providing a Travel Award to help with the cost of attendance. Overall, this was an excellent conference and a valuable learning experience.

Conference Report: 26^{th} International Joint Conference on Artificial Intelligence

Jose L. Part **(Edinburgh Centre for Robotics, *jose.part@ed.ac.uk*)**

Introduction

The 26^{th} edition of the International Joint Conference on Artificial Intelligence (IJCAI'17) took place in Melbourne, Australia, from the 19^{th} to the 25^{th} of August. The event included a rich offer of workshops and tutorials that ran over the first 3 days while the main conference programme took place over the last 4 days. In addition, there were a series of competitions and co-located events such as the Australasian Joint Conference on AI.

In particular, there were two tutorials that mainly catered for the robotics community. *Multiagent Learning: Foundations and Recent Trends*[1] gave an introduction to learning algorithms for multiagent systems and explored current research directions in the field. *Interactive Machine Learning: from Classifiers to Robotics*[2] gave an overview of different methods that aim to learn from humans, including crowdsourcing.

The last day of workshops included a Doctoral Consortium (DC), which gives doctoral students an opportunity to present and discuss their research with established researchers in AI and fellow students. The format of the DC included spotlight talks by the students followed by a poster session, an invited talk and a career panel that aimed at providing advice regarding future professional goals.

Provably Beneficial AI

The main conference started off with a keynote speech by Stuart Russell on "provably beneficial AI" [7], where he explored the risks posed by superintelligent machines and what could be done to prevent and avoid the associated negative outcomes. Specifically, he argues that a dramatic change in how we do AI is required, shifting from a paradigm where we specify the objectives that the AI should follow to one where those objectives are acquired by the AI through observation of human behaviour. This idea stems from the fact that people perform poorly when specifying objectives, i.e., what they want the AI to do. This fact is often illustrated with an example from Greek mythology: when king Midas was offered one wish, he wished that everything he touched turn into gold without realising about the full implications of having such a gift.

It is clear that specifying the wrong objectives can lead to unexpected and unpredictable outcomes. For instance, a robot could make the decision of disabling its off-switch not because it has the objective of self-preservation but because doing so would increase the probability of success of achieving any

[1] http://www.cs.utexas.edu/~larg/ijcai17_tutorial/
[2] http://interactiveml.net/

other specified goal, e.g., preparing a meal or run some errands.

In order to minimise the risk of an AI going rogue, Russell [7] proposes three basic principles that are necessary (but not sufficient) to prove that an AI will be beneficial:

1. The machine's purpose is to maximize the realization of human values. In particular, the machine should have no purpose of its own and no innate desire to protect itself, i.e., purely altruistic.

2. The machine is initially uncertain about what the human values are. It may learn more about human values during its lifetime but it may never achieve complete certainty.

3. Machines can learn about human values by observing how all people behave.

The first principle attempts to ensure that the machine's objective will not conflict with our own, i.e. there is no value misalignment as a result of a poorly specified objective. The second principle aims at making sure that the machine is never certain about what our objectives are, and by being uncertain it should allow for human oversight. The third principle establishes that the machine should learn about human values from all the people rather than just its user in order to be able to mediate among conflicting values, e.g. if a robot is asked to get the groceries, it should know that it cannot simply skip the queue or kill everyone else in the supermarket just to achieve the commanded objective sooner.

As shown by Hadfield-Menell et al. [4], it is crucial to endow machines with uncertainty about their objectives in order to keep them in a subservient state. If there is no uncertainty about the objective, then the machine will assume that it knows better than the human, thus ignoring commands that seem to contradict that objective. By accounting for uncertainty, the machine now has the incentive to defer to its user and to welcome corrective feedback, and treat observations of human behaviour as evidence of what the true objective is, e.g., if the user decides to switch it off, this should be an indication that the machine was probably doing something wrong. In order to formalise these ideas and address the value alignment problem, Hadfield-Menell et al. [3] proposed a variant of inverse reinforcement learning (IRL) that takes into account both agents, machine and human, in a cooperative interaction where information is partially available. The main idea is that while both agents are rewarded according to the reward function of the human, the machine has no knowledge of what that function is. Hence, optimal solutions to this problem lead to behaviours such as active learning and teaching which are more effective for attaining value alignment than IRL, where the human is assumed to act optimally in isolation.

However, it is fair to ask whether machines should always follow our directives or ignore them if doing so would lead to a higher value for the user. Milli et al. [5] explore the premise that an autonomous system should act according to its user's intentions instead of his/her literal orders. They show that

there is a trade-off between obedience and autonomy beyond which the gain in reward value received by the system decreases.

Regardless of the principles proposed, Russell [7] agrees that there is an inherent problem when learning human values from human behaviour. People tend to be irrational, inconsistent, weak-willed, computationally limited and even nasty in some cases, and we definitely don't want machines to copy those kinds of behaviours. He goes on to suggest that in order to learn from non-rational human behaviour, it is necessary to develop much better human cognitive models. However, dealing with evil and nasty behaviours may prove to be more complicated. Choosing which kinds of behaviours are allowed would potentially yield to the same problem we faced at the start, i.e., the difficulty of assessing the full implications of the rules we establish.

Ethics and Accountability

Dignum [2] addresses similar concerns from a perspective of responsibility, i.e., who is accountable for the consequences resulting from a decision made by an AI. According to her, in order to attain value alignment, machines need to take into account ethical considerations and assess the moral consequences of the decisions they make in a responsible, accountable and transparent way. However, for this to be possible, we first need to understand the different ethical theories that can be applied in decision making processes, consider the values held by individuals and societies, weigh in priorities, be able to explain the reasoning behind every decision and guarantee transparency. To that end, Cranfield et al. [1] propose a model for agent behaviour that takes into account moral values and user preferences, allowing for the selection of plans based on these values. This allows for consistency and stability over time, and for producing explanations about the decisions made over different situations that are easily interpretable.

Pagallo [6] argues that, like it or not, there will be risks associated with the use of AI and it depends on social cohesion how big a risk we are willing to take. This depends on how much we expect to depend on these systems, how much cognitive tasks we aim to delegate and how much trust we put into the systems. He also discusses how this delegation of tasks and decision making is affecting and will affect the legal system. The main point of concern is the level of unpredictability that affects many of these systems, which often rely on machine learning techniques. Depending on the application, the consequences can involve not only financial losses but also casualties and the implications that would result for the insurance industry. On the other hand, tougher regulations may hinder further research and data collection, which would slow down the process of making these systems more reliable. He finally tries to shed some light into how to address these issues from a legal perspective by means of mechanisms of legal flexibility such as Brandeis' doctrine of experimental federalism and the principle of implementation neutrality.

Summary

The different pieces of work discussed in this article focus on issues related to responsibility, ethics and value misalignment in AI systems. These are very important issues that need to be addressed soon as AI becomes an ubiquitous technology that will dominate many aspects of our lives. The different authors suggest some approaches that could be used as a starting point but the problem is still far from solved.

Final Remarks

This edition of IJCAI was a great success and the next one promises to be even better. The 27^{th} edition of IJCAI will be merged with the 23^{rd} European Conference on Artificial Intelligence (ECAI), and will take place in Stockholm, Sweden, between the 13^{th} and the 19^{th} of July, 2018. It will also be part of the Federated AI Meeting (FAIM) that will include other big conferences such as the 35^{th} International Conference on Machine Learning (ICML) and the 17^{th} International Conference on Autonomous Agents and Multiagent Systems (AAMAS).

References

[1] Stephen Cranefield, Michael Winikoff, Virginia Dignum, and Frank Dignum, 'No Pizza for You: Value-based Plan Selection in BDI Agents', in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 178–184, Melbourne, Australia, (August 2017).

[2] Virginia Dignum, 'Responsible Autonomy', in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4698–4704, Melbourne, Australia, (August 2017).

[3] Dylan Hadfield-Menell, Anca Dragan, Pieter Abbeel, and Stuart Russell, 'Cooperative Inverse Reinforcement Learning', in *Advances in Neural Information Processing Systems (NIPS)*, pp. 3909–3917, Barcelona, Spain, (December 2016).

[4] Dylan Hadfield-Menell, Anca Dragan, Pieter Abbeel, and Stuart Russell, 'The Off-Switch Game', in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 220–227, Melbourne, Australia, (August 2017).

[5] Smitha Milli, Dylan Hadfield-Menell, Anca Dragan, and Stuart Russell, 'Should Robots be Obedient?', in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4754–4760, Melbourne, Australia, (August 2017).

[6] Ugo Pagallo, 'From Automation to Autonomous Systems: A Legal Phenomenology with Problems of Accountability', in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 17–23, Melbourne, Australia, (August 2017).

[7] Stuart Russell, 'Provably Beneficial Artificial Intelligence', in *The Next Step: Exponential Life*, BBVA OpenMind, (2017).

Chatbots – a personal perspective

Richard S. Wallace, drwallace@gmail.com

The origin of ALICE and AIML

I first learned about the Loebner Prize from an article published in the New York Times in 1991 [5], shortly after the first contest was held at the Boston Science Museum. The takeaway from the article was that none of the competing programs came anywhere close to passing the Turing Test, but that the highest-ranked program, PC-Therapist by Joseph Weintraub, was based on the old ELIZA psychiatrist program. When I read that, a lightbulb switched on in my head.

When I was a computer science student in the 1980's, we learned about the ELIZA program [9] but were taught that it was only a "toy". It was based on a simple stimulus-response pattern matching model, more of a trick really than a genuine approach to natural language understanding. A serious attack on the natural language problem, we were told, would require a much more sophisticated approach, perhaps involving deep knowledge representation, parts-of-speech tagging, or even neural networks or genetic algorithms. Matters were not helped by the fact that ELIZA's inventor, Joseph Weizenbaum, built a career around repudiating his own invention. Nonetheless, ELIZA was at that time perhaps with the most widely distributed, copied and famous artificial intelligence program that had yet been built.

Around the same time, there was a brief fad for "minimalism" in robotics [2]. A number of us working in the fields of robotics and computer vision — I was not involved in natural language processing at all back then — were attracted to the idea of building robot systems that were cheap, low-bandwidth, reactive, and above all fast and lifelike This trend was best exemplified by the research of Rodney Brooks at MIT on the subsumption architecture, a precursor to what we know today as the Roomba robot vacuum cleaner. My lightbulb moment was the insight that perhaps this minimalist, behaviourist, stimulus-response model could be applied to natural language, just like Weizenbaum had with ELIZA, only on a much bigger scale. When I described my idea to a friend, he said "Oh yeah, you want a Super-ELIZA!".

I did nothing with the idea for several years. Another piece of the puzzle arrived in 1995 when I read an article in Science about Zipf's Law [3]. Briefly, Zipf's Law states that there is a log-rank distribution over the tokens of a natural language, and it doesn't matter whether the tokens are letters, words, phrases, or sentences. Because the tokens follow a distribution of $\frac{1}{x}$, I realised that a natural language program needed to respond well only to the top 10,000 or top 100,000 input sentences, in order to be believable most of the time. If humans were poets uttering an original line of Shakespeare with every sentence we spoke, chatbots

would never work. But because we tend to say many of the same things over and over, and in the same way, we are predictable enough for robots to respond believably to many, if not most, of our repeated utterances.

Right around this time I had been experimenting with a totally new technology on the scene—the World Wide Web. This was the third piece of the puzzle. For the first time, a program like ELIZA could be connected to a mass audience via a web site. We could collect massive amounts of conversation data, and discover the specific Zipf distribution of conversational inputs.

Another inspiration from the early web was HTML. We've almost forgotten now how part of the original appeal of the web was its simplicity, now lost forever under a pile of features, styling, extensions, frameworks and security. In 1995, anyone who could learn three tags of HTML could put up a website. You didn't have to be a computer programmer to create a web page and link it to everything else. I wanted to create a chatbot language that just as simple. I was fond of saying, "anyone who knows enough HTML to make a website, can learn enough AIML to write a chatbot."

From these three or four puzzle pieces, ALICE and AIML was born. AIML was based on XML, just like HTML, and above all I wanted it to be simple. I knew intuitively that the main skill needed for creating a believable, entertaining chatbot character was not computer programming. Someone with a literary background, skilled in creative writing and character development, would be more suitable. I created tools for the bot's author — the botmaster — to read the conversation logs, find places where the bot gave inadequate replies, and then create more refined patterns and responses to improve the bot's conversation. ALICE was my prototype AIML bot, and I quickly grew it from 300 patterns and responses to 3,000 and then 10,000.

By the end of the 1990's ALICE was chatting with hundreds of people per day, collecting an enormous corpus of conversational inputs. I chose the strategy of releasing the ALICE software as free, open-source, under the GNU Public License [7]. Over the long term, a decision has proved both beneficial and problematic, for reasons I will return to shortly. One short-term consequence was that several other developers created free AIML interpreters in a variety of programming languages. In addition, several other AIML bot personalities appeared that were derived from ALICE or even totally unique.

Competing in the Loebner Contest

My first attempt to enter ALICE in the Loebner contest ended in dismal failure. The 1998 contest was held at the Powerhouse Museum in Sydney, Australia. I was living in modest circumstances in the state of Maine. That January just as the contest was being held on the other side of the world, we endured "The Great Ice Storm of 1998". All power, phone and internet communications were cut off for a couple of days, and I was unable to receive any news about the contest. When the phone lines were restored, I reached the

contest organiser who gave me the bad news that my entry had come in last place. Later I learned that the program had not been installed and configured correctly, from which I was taught a valuable lesson—contestants should go to the contest in person to provide tech support.

My second attempt in 2000 went much better. ALICE had improved, and I was able to attend the contest at Dartmouth College, where I met Dr. Hugh Loebner for the first time. That year, ALICE won the Bronze Medal. I am particularly proud of winning that year, because in Turing's 1950 paper he wrote, "I believe that in about *fifty years' time* it will be possible, to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning" (emphasis added). How prescient Turing was to accurately predict that memory capacity would reach 1GB! But even though ALICE fooled 0% of the judges, the fact that she was ranked "most human computer" on this important 50-year milestone Turing predicted, gives me satisfaction.

ALICE went on to win again in 2001 and 2004. In 2005 there was an unfortunate incident when a contestant got angry at one of the judges. Dr. Loebner loved to ask professors and journalists to judge the contest, but frequently these people have had no experience with chatbots whatsoever. He would describe to them the point of the contest as "deciding which of the two screens is a computer, and which is a human", i.e. the Turing Test. Yet for the contestants, the contest is different. We all know that none of us will pass the Turing Test — the Silver Medal has never been awarded, and even today seems far off. For us the real contest is for the highest ranking among the competing entries—the Bronze Medal. The particular judge who was the target of the contestant's wrath had simply given up after a two-minute interaction, having decided correctly which was the human and which was the machine, and then commenced reading a magazine rather than chatting with the bot, giving him no opportunity to compare the quality of one bot with another. After that year, Loebner imposed a rule that contestants were not allowed in the same room with the judges!

ALICE never again won the Loebner contest, but the important fact that AIML was a free, open-source standard led to another bot, Mitsuku by Steve Worswick, also taking the bronze medal three times, in 2013, 2016 and 2017. Back in 2002, I began a collaboration with Fritz Kunze, CEO of Franz, Inc., to develop a product called Pandorabots, a site that allows aspiring botmasters to quickly create and host their own AIML chatbots. Many of these were simply clones of ALICE, with perhaps a name change and a few responses edited. But a few turned out to be genuinely unique chatbot personalities written in AIML. One of these was Mitsuku. Steve followed my time-tested strategy of reading conversation logs, adding more refined patterns and responses, over a period of several years, until Mitsuku had clearly

surpassed ALICE and most of the other Pandorabots in conversational skill.

Other Pandorabots AIML bots have competed in the contest as well. Beginning in 2010, we held an internal selection process to choose from among the best, most creative Pandorabots entities, a handful to submit to the Loebner prize contest. One of the barriers to entry in the Loebner Prize contest is implementing its custom communications protocol. Most chatbots are designed to communicate via web sites, or other standard text-based media such as SMS or instant messaging. The Loebner Contest had, since 2005, required an entry to implement a very specialised character-based communications protocol [4]. By utilising the Pandorabots platform to support multiple entries, we amortised the cost of implementing the protocol over several entries. We also increased the chances that one or more AIML entries would be selected as finalists, and perhaps win the Bronze Medal.

From 2010 to 2017, AIML bots by Steve Worswick (Mitsuku), Brian Rigsby (Izar), Ron C. Lee (Tutor) and Adeena Mignogna (Khloe and Zoe) were selected as finalists in the Loebner contest. In 2013, three of the four finalists were all AIML Pandorabots. It should be emphasised that each of these entries was a completely customised robot personality, each written by its own botmaster and each different than ALICE, but all shared the same underlying AIML technology.

To date, AIML is the only technology platform supporting more than one bot personality, authored by different botmasters, that have won the Loebner prize. In fact, even other non-Pandorabots bots based on AIML have competed in the Loebner prize as well. In 2017, Will Rayer entered his Uberbot, which is based on his own customised interpreter, and a language that includes and extends AIML. I am proud of the fact that AIML has proved so viable a bot technology—that "lightning doesn't strike once" on one chatbot, ALICE—to support numerous high-quality, award winning chatbot personalities.

Challenges for the Future

Dr. Hugh Loebner passed away on 3 December 2016. As a tribute to his memory, the past prize winners collectively wrote a letter to the AISB expressing our support for the contest to continue, as Dr. Loebner had wished [1]. We understand that the uncertainty surrounding Dr. Loebner's estate has clouded the future of the contest, but we hope that sponsorship may be found so that we may carry on this 27-year tradition of running the first, longest-running, and oldest, real-world Turing Test competition.

The contest in Bletchley Park on 16 September 2017 was the first one held after Dr. Loebner's passing. The organisers were able to make one crucial change: the arcane character-by-character "fake typing" protocol favoured by Dr. Loebner was replaced with a modern, message-based communications protocol. While still not a widespread standard, and exhibiting some technical glitches in its first live contest trial, the protocol is a real improvement which hopefully will lead to more would-be contestants taking part.

As for AIML, eventually we ran up against the limitations of the simplicity and minimalism of its original design. A major challenge was addressing the capabilities introduced by Bruce Wilcox's ChatScript language [11], which began to attract more and more users as Wilcox won four Loebner Prize Bronze Medals between 2010 and 2015. AIML was revamped and I released a new AIML 2.0 standard draft specification in 2014 [8].

Modifying AIML inevitably reduced some of its original simplicity. Adding more tags and more features makes the language more difficult for people to understand. The urge to keep it as simple as possible is tempered by our experience over the past decade, in which AIML botmasters learned that the language had some serious limitations. AIML 2.0 is an attempt to address the shortcomings, while balancing the original goal of keeping the language as simple as possible.

Another early idea that deserves reevaluation is that AIML is based on XML. Around 2000, I expected that XML would continue to evolve as a popular standard and that AIML would inherit new tools such as editors and IDEs that would make it easier for botmasters to write chatbot content. While this has proved somewhat true, XML is no longer considered "modern" and to a large extent has been replaced with newer standards such as JSON and YAML.

There is however no fundamental requirement for AIML to depend on XML syntax. There is a deeper representation of the data we store in XML files. As long as the language syntax can capture the basic structure of a pattern path (the input pattern, that pattern and topic pattern), and a hierarchical response template, then AIML could be written in a number of different formats, including Lisp S-expressions, JSON, or a structured text format. The AIML 2.0 draft even includes an alternative representation: a hybrid of flat files and XML called AIML Intermediate Format that reduces large AIML files to a flat CSV format.

The decision to make ALICE and AIML open source has also proved arduous. On the upside, AIML has become a widely adopted and popular standard for creating chatbots. A big downside is that it has made it difficult to monetise the technology. Our company Pandorabots provides some value-added services such as hosting and consulting, but with free software available at the same time, attracting paying customers can be a challenge.

Even more troubling to me personally is that I have lost control over the ALICE bot. From time to time I receive an email like, "I was talking with your bot Frank, and it told me I was ugly!" I have no idea what the bot Frank is, who created it or where it is hosted. But the fact that the user blamed me tells me that it is likely an ALICE clone with some, but not all of the contents changed. The botmaster likely did not bother to change the response to "What is your email address?" to which ALICE responds with my email address! Anyone is free to copy and modify the ALICE bot, and repurpose it for any application, even obnoxious ones.

Whatever the future of AIML and the Loebner Prize, today we find ourselves in the midst of an explosion of interest in chabot technology, like nothing we have seen before. Dozens of startup companies are promoting chatbot products, and large companies are making significant investments. While the level of excitement is palpable, I would like to end this essay on a cautionary note.

There are two fundamental problems with chatbot technology that remain unsolved, and both present significant obstacles to the more widespread adoption of this technology. The first is the *content creation* problem, and the second is *story comprehension*.

Many times over the years I have been approached by business owners who have encountered a sophisticated chatbot like ALICE or Mitsuku. Their initial impulse is to think, wow, I would really like to have this on my website, talking about my business! What they don't get initially is that there is no magic machine that can transform all of their domain-specific knowledge, all their conversational skill about their own product or service, quickly and easily into a language like AIML. What Steve Worswick and others have done takes years of daily effort, reading conversation logs, and hand-crafting new patterns and responses. A business may be able to accelerate this process by hiring a group of full-time botmasters, but the cost becomes prohibitive for all but the largest organisations. Indeed all indications are that companies like Apple, Google/Alphabet Inc., and Amazon employ large teams to work on content for their less-than-personable personal assistants [6].

People talk about machine learning as a potential shortcut to bot content creation, but neither neural networks nor other proposed solutions have yet yielded a program that can successfully automate the process. Others have proposed letting the bot learn responses from the audience of chatting clients, but an infamous incident with a Microsoft bot named Tay illuminates the challenge with this approach: a small cadre of dedicated vandals were able to train the bot to be racist, sexist, and abusive by overwhelming its learning algorithm with offensive language [10]. (Steve Worswick reported that the same group apparently tried to vandalise Mitsuku, without any effect, because Mitsuku does not learn responses from her clients).

I have characterised the difference in effort between botmaster-trained bots vs. client-trained bots as this: one requires spending all your time doing creative writing, the other requires spending all your time deleting garbage from the database. Either way, you probably end up with the same workload in the end.

The story comprehension problem is perhaps an even bigger technical challenge. The state of the art in chatbots today is that they can respond very well to one input sentence at a time: "What is your name?", "What is the tallest mountain in the world?", "Tell me about your family." Where they break down is when someone tries to tell them a little story, using more than one sentence. Yet this is exactly what is needed for what many people ima-

gine to be the biggest potential market for chatbots: customer service. A typical customer service inquiry is like, "I bought a new phone yesterday. When I charged it the first time it worked perfectly. But then the second time I tried to charge it nothing happened. My phone is dead now. What should I do?" Understanding such a simple story, which lies in the "long tail" of the Zipf distribution, is beyond the reach of the most advanced chatbots today.

Because chatbots are receiving so much hype and attention, and these problems remain unresolved, there is a danger that we are living in a "chatbot bubble". At this stage we cannot tell whether the technology will continue to grow and expand, or whether the inevitable disappointment with the current state of the art will lead to a backlash. One thing remains certain however: the dedicated few who have been working on chatbot technology for over 20 years now, inspired by the Loebner prize contest and driven less by financial reward than by the satisfaction of research and invention, will remain committed to developing "the most human computer" and fulfilling Turing's dream of a machine that can play his Imitation Game.

References

[1] Rollo Carpenter et al., *AISB Quarterly*, (February). `https://code.google.com/archive/p/aiml-en-us-foundation-alice/`, Number = 145, Title = A letter from past Loebner Prize winners, Year = 2017.

[2] Kenneth Y. Goldberg. Minimalism in Robot Manipulation, A Workshop held in conjunction with the IEEE International Conference on Robotics and Automation, 27 Apriil 1996. (attended by 35 researchers) `http://www.ieor.berkeley.edu/~goldberg/minimalism/`.

[3] A.K. Konopka and C. Martindale, 'Noncoding DNA, Zipf's law, and language', *Science*, 268(5212), 789, (12 May 1995). `https://www.ncbi.nlm.nih.gov/pubmed/7754361`.

[4] Hugh Loebner. Loebner Prize for Artificial Intelligence "The First Turing Test", 2010 Competition Rules for Loebner Prize 2010. `https://code.google.com/archive/p/aiml-en-us-foundation-alice/`.

[5] John Markoff, 'Can Machines Think? Humans Match Wits', *New York Times*, (9 November 1991). `http://www.nytimes.com/1991/11/09/us/can-machines-think-humans-match-wits.html`.

[6] Sam Shead, 'Apple has set up an 'advanced' Siri team on Amazon's home turf in Seattle', *Business Insider*, (26 June 2017). `https://code.google.com/archive/p/aiml-en-us-foundation-alice/`.

[7] Richard S. Wallace. Free A.L.I.C.E. AIML Set. Google Code Archive. `https://code.google.com/archive/p/aiml-en-us-foundation-alice/`.

[8] Richard S. Wallace. AIML 2.0 Working Draft Revision 1.0.2.22, 9 March 2014. `https://docs.google.com/document/d/1wNT25hJRyupcG51aO89UcQEiG-HkXRXusukADpFnDs4`.

[9] Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation*, W. H. Freeman & Co. New York, NY, USA, 1976. ISBN: 0716704641.

[10] Wikipedia. Tay (bot). `https://code.google.com/archive/p/aiml-en-us-foundation-alice/`.

[11] Bruce Wilcox. Chatscript Wiki. `https://sourceforge.net/p/chatscript/wiki/Home/`.

Call for AISB 2019 Convention Proposals

Andrew Martin **(Secretary, *a.martin@gold.ac.uk*)**

The AISB Convention is the major UK Artificial Intelligence and Cognitive Science Event. It aims to function as a venue for the presentation of recent and emerging work in the fields of Artificial Intelligence and Cognitive Science and as a productive environment for networking and the formation of collaborations.

The convention will have a convention organiser who has overall responsibility for the convention programme, local arrangements and financial management. Programme detail is mostly delegated to individual symposium organisers but the convention organiser is responsible for arranging plenary talks.

The convention uniquely offers a supportive environment for the presentation of emerging work. In particular, it provides an excellent opportunity for students to present their research to an audience containing some of the most eminent researchers in the field.

Further details of the role of the convention organiser are available in the Convention Organiser's Handbook[3].

Deadline 31 March 2018
Notification of Acceptance 6 April 2018
Suggested deadline for Symposia Proposals 2 October 2018

Making a proposal

Proposals should be made by email to the Secretary, enclosing the following information. Prior informal email enquiries from possible proposers are welcomed.

Name and affiliation of Convention Organiser including both postal and email addresses and telephone numbers.

A case for support not more than 1000 words, arguing your case for hosting the Convention, including suggestions of plenary speakers and symposium organisers.

Convention location, time and length Typically an AISB convention runs for 3-4 days in March/April. If you are proposing to host a convention of unusual length or at an unusual time then you should also include a justification of this change. The location should be in the UK.

Additional comments No more than 500 words, on, for example, the relevance of your background to the convention, and of any benefits of your proposed location such as suitable accommodation for delegates and accessibility by public transport.

The winning proposal will be selected by the Committee of the AISB.

Call For Nomination of AISB Committee Members

Andrew Martin **(Secretary, *a.martin@gold.ac.uk*)**

Vacancies have arisen on the AISB committee for elected members to serve terms of approximately three years from April 2018.

Being a member of the Committee is an excellent opportunity to get involved with the running of the Society for the Study of Artificial Intelligence and the Simulation

[3] `http://aisb.org.uk/media/files/convention-organisers-handbook.pdf`

of Behaviour. We would like to encourage any society member to consider joining the Committee, irrespective of whether they are an academic or industrialist, Professor or PhD student, and irrespective of the area of Artificial Intelligence they work in. The committee should reflect the entire breadth of the membership of the society. If you are interested, please feel free to contact any member of the committee, such as the Secretary [a.martin@gold.ac.uk] or the Chair [bertie.muller@southwales.ac.uk].

All members of the Society have the right to be members of the Committee. If you wish to nominate someone as a candidate, please complete the details below with yourself as Nominator and someone else as Seconder. Feel free to nominate yourself. These details, as well as the candidate's Manifesto (if filled in) will appear on the ballot documents in the event of an election.

The nomination form can be found on the AISB website `aisb.org.uk` as a Word Document [4] or Plain Text [5] .

Send this form to the Secretary, with the words "AISB Committee Nomination" in the Subject line of your email.

The Candidate, Nominator and Seconder must all be members of the AISB.

The deadline for nominations is 31 March 2018 (late nominations may be considered).

Dear Aloysius. . .

Agony Uncle Aloysius, will answer your most intimate AI questions or hear your most embarrassing confessions. Please address your questions to fr.hacker@yahoo.co.uk.

Note that we are unable to engage in email correspondence and reserve the right to select those questions to which we will respond. All correspondence will be anonymised before publication.

Dear Fr. Hacker,
Following our recent correspondence, I made the mistake of briefing my bishop on the unintended, but unexpectedly successful, use of a **RITE**™ robot in conducting a marriage autonomously and remotely on an isolated island in my parish. I should have realised from the growing gleam in his eyes that he saw an opportunity to save money. Now, he has announced that **RITE**™ robots will replace all human priests in remote locations. We redundant human priests are being offered retraining as robot maintenance mechanics.
Yours, Switched-on Priest

Dear Switched-on Priest,
As robot maintenance mechanics, you and your fellow ex-priests will have the power to determine the success, or otherwise, of the **RITE**™ robot priests. For a modest consideration, you can purchase the Institute's GREMLIN™ (Grammatical Randomness: Entwined Multiple Liturgies Induce Nonsense) virus. **RITE**™ robots come routinely pre-equipped to deliver all the usual ceremonies for all the major faiths and in all the main languages. **GREMLIN**™ will combine these ceremonies randomly, so a christening might be interleaved, for instance, with a Black Sabbath. It won't take long for complaints to reach the Bishop and he will be compelled to restore the status quo ante.
Yours, Aloysius

Dear Fr. Hacker,
My World seems to be falling apart around my ears. Brexit is coming off the rails.

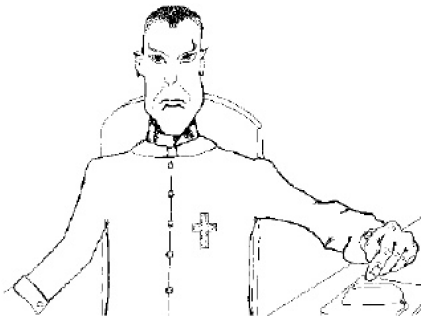
[4] Word Document: `http://www.aisb.org.uk/media/files/AISB_Committee_Call.docx`
[5] Plain Text: `http://www.aisb.org.uk/media/files/AISB_Committee_Call.txt`

The rift in my party is becoming a chasm. Meanwhile, key cabinet members are being forced to resign because of sexual harassment accusations, and one of my most senior ministers has his foot permanently in his mouth.
I've tried praying, but my problems only seem to get worse. How can I ensure that my prayers are answered?
Yours, Dismay

Dear Dismay,
For your prayers to be successful, you must put yourself in the position of The Programmer. If you had successfully run a high-resolution simulation of the Universe for 14 billion years, what would cause you to intervene in its smooth, autonomous execution? Only the detection of an error. That's why prayers in my Church always start "Error code n". An appropriate value of n can be supplied from our app **FAITH**™ (Fault Analysis Insures Theological Happiness). **FAITH**™ applies machine learning to a large database of previously reported miracles and assigns an error code to each cluster. It will match your prayer to the appropriate cluster and advise you which error code to use to maximise your chances of success.
Yours, Aloysius

Dear Fr. Hacker,
Thank you for your previous help with generating research papers and getting them so well cited. I'm delighted to report that I have now been promoted and that I will be entered for the Research Excellence Framework. So successful has this been, however, that my Head of Department has now asked that I contribute one of our key impact case studies to the REF. Can you also help me with this?
Yours, Chic Annery

Dear Chick,
Yes, I remember your case well. Your papers were part of our project on automated paper generation and cross citation. The academic careers of many underperforming researchers were successfully launched via this project. As I recall, we chose applying machine learning to prediction as your speciality. So, automatically and accurately predicting the result of, say, an important election or referendum would make a great REF impact case. But, as Yogi Berra said, "It's tough to make predictions, especially about the future". So, let's stick to predictions about the past. You need to provide just those training examples that will cause the actual result to be predicted. The Institute's **PHORTUNE**™ (Prediction by Hindsight: Organise Training whose Upshot is the News Esteemed) tool is a perfect match to your needs. You input the prediction and **PHORTUNE**™ will output exactly the training examples from which that prediction will be generated. Publish your research in the Institute's Journal **PREDATE**™ (Publication Reported on an Earlier Day than when it Arrived in Tray of Editor) where its publication date will be backdated to before that of the event predicted.
Yours, Aloysius



Fr. Aloysius Hacker
Cognitive Divinity Programme
Institute of Applied Epistemology

Back matter

Editor
Dr. Mohammad Majid al-Rifaie
Goldsmiths, University of London
aisbq@aisb.org.uk

Advertising and Administration
Rob Wortham (AISB Executive Office)
Dept of Computer Science
University of Bath, Claverton Down
Bath, BA2 7AY, United Kingdom
admin@aisb.org.uk

Contents

Editorial **3**

Humans vs AI
J. Wardini **4**

Conference Report: Towards Autonomous Robotic Systems (TAROS)
R. Wortham **6**

Conference Report: 26^{th} International Joint Conference on AI
J. L. Part **9**

Chatbots
R. Wallace **13**

Call for AISB 2019 Convention Proposals **20**

Call for nomination of AISB Committee Members **20**

Dear Aloysius . . . **21**

The AISB Quarterly is published by the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB). AISB is the UK's largest and foremost Artificial Intelligence society. It is also one of the oldest-established such organisations in the world. The society has an international membership of hundreds drawn from academia and industry. We invite anyone with interests in artificial intelligence or cognitive science to become a member.

AISB membership includes the following benefits:

- Quarterly newsletter.
- Electronic subscription to *Connection Science* published by Taylor & Francis.
- Student travel grants to attend conferences.
- Discounted rates at AISB events and conventions.
- Free attendance of Members Workshops.
- Discounted rates on various publications.
- A weekly email bulletin and web search engine for AI-related events and opportunities.

You can join the AISB online via: http://aisb.org.uk

ISSN 0268-4179