AI Matters

Annotated Table of Contents

Links

SIGAI website: http://sigai.acm.org/
Newsletter: http://sigai.acm.org/aimatters/
Blog: http://sigai.acm.org/ai-matters/
Twitter: http://twitter.com/acm_sigai/
Edition DOI: 10.1145/3203247

Join SIGAI

Students $11, others $25
For details, see http://sigai.acm.org/
Benefits: regular, student

Also consider joining ACM.

Our mailing list is open to all.

Notice to Contributing Authors to SIG Newsletters

Submit to AI Matters!

We're accepting articles and announcements now for the Spring 2018 issue. Details on the submission process are available at http://sigai.acm.org/aimatters.

AI Matters Editorial Board

Contact us: aimatters@sigai.acm.org

Contents Legend

Book Announcement
Ph.D. Dissertation Briefing
AI Education
Event Report
Hot Topics
Humor
AI Impact
AI News
Opinion
Paper Précis
Spotlight
Video or Image

Details at http://sigai.acm.org/aimatters

Welcome to AI Matters 4(1)

Amy McGovern, Co-Editor (University of Oklahoma; aimatters@sigai.acm.org)
Eric Eaton, Co-Editor (University of Pennsylvania; aimatters@sigai.acm.org)
DOI: 10.1145/3203247.3203248

Issue overview

Welcome to the first issue in our fourth year of *AI Matters*. This issue has lots of great new ways for you to catch up on the latest in AI News, beginning with reports on conferences and events. The first is a report on the first AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, which SIGAI helped to start. We also have a discussion on the 2018 AAAI job fair. We also have a forward looking vision of what AI could be like in 2017, from the perspective of four prominent AI researchers who discussed this issue at IJCAI 2017.

In this issue's **AI Education** column, Todd Neller discusses teaching Hidden Markov Models. He provides a number of resources, including online lectures and readings, to help AI educators on this important topic.

Next, we have an obituary for Professor Jon Oberlander from the University of Edinburgh, who passed away suddenly.

The AI Matters blog (http://sigai.acm.org/ai-matters/) contains regular postings on AI and current policy by Larry Medsker, our ACM SIGAI Public Policy Officer. He writes a summary article for each issue on **AI Policy**, but the blog contains the most up-to-date and full information.

We have a fascinating contribution from John Havens summarizing the work at IEEE on creating ethical standards for AI systems. SIGAI is represented on the executive committee for this work and we suggest that you read his article and also get involved.

Finally, we have a great humor contribution from Michael Genesereth written in the form of a newspaper article from the future when a computer is elected governor of California! It could happen!

You may notice that our **AI Interviews** column is missing in this issue! We asked a number of prominent AI researchers but were unable to get a response from anyone in a timely manner because they are so prominent. If you have suggestions on who we could interview next, we would really appreciate it!

New AI Matters co-editors needed!

Are you interested in learning about lots of great AI related news and information? Do you have a passion for a particular part of AI news that we should be covering?

AI Matters really needs new YOU!

We are in need of new co-editors as well as leaders for individual sections. We have ideas for new sections but need people to help take the lead. This is a fun and rewarding job and would be great for visibility for more junior researchers. If you are interested, please email us at `mailto:aimatters@sigai.acm.org`.

Submit to AI Matters!

Thanks for reading! Don't forget to send your ideas and future submissions to *AI Matters*! We're accepting articles and announcements now for the Spring 2018 issue. Details on the submission process are available at http://sigai.acm.org/aimatters.

Amy McGovern is a Co-Editor of AI Matters. She is an Associate Professor of computer science at the University of Oklahoma and an adjunct associate professor of meteorology. She directs the Interaction, Discovery, Exploration and Adaptation (IDEA) lab. Her research focuses on machine learning and data mining with applications to high-impact weather.

Eric Eaton is a Co-Editor of AI Matters. He is a faculty member at the University of Pennsylvania in the Department of Computer and Information Science, and in the General Robotics, Automation, Sensing, and Perception (GRASP) lab. His research is in machine learning and AI, with applications to robotics, sustainability, and medicine.

AAAI/ACM SIGAI Job Fair 2018: A Retrospective

John P Dickerson (University of Maryland; john@cs.umd.edu)
Nicholas Mattei (IBM Research; n.mattei@ibm.com)
DOI: 10.1145/3203247.3203249

Introduction

The 2018 AAAI/ACM SIGAI Job fair marked the third installment of the popular event for both recruiters and students. As we have seen steady growth in the attendance at AAAI over the last several years we have seen similar growth in the number of companies and students who are actively participating in the job fair. This year, twenty-one companies attended—typically with a team of recruiters and other representatives—and hundreds of students and other job seekers either uploaded their resumes to a public book before the event or came through the event space during the event. Those resumes were then shared with participating companies.

Participating Companies

- Adobe Research
- Alibaba
- Amazon
- ASAPP
- Baidu
- BBN Technologies
- DiDi Chuxing
- Georgian Partners
- HRL Laboratories
- Inferlink
- IBM Research
- JD.com
- Lionbridge
- Microsoft
- Nissan
- Prowler
- Tencent
- Bosch Research and Technology Center
- The Information Sciences Institute (ISI)
- Lawrence Livermore National Labs
- Samsung Research

In a change of pace, this year's job fair saw each company pitch themselves with a lightning talk to a room packed with students, postdocs, and other AAAI conference attendees. Each of the participating companies made a single compelling slide and a two-minute pitch for the types of talent they were looking to recruit and the opportunities available at their company. Participating companies were also allocated booth space, either in the main exhibition hall that was set up for the majority of the main conference, or in a specific large conference hall allocated specifically to the job fair. Many of the companies were looking for both theoretical and applied machine learning skills but nearly as many needed help with more classical symbolic AI techniques including logic programming, planning, and reasoning. In short the need for AI talent is large and AAAI is a great place to recruit that talent.

Figure 1: Attendees mingle with each other and recruiters before the formal start of the job fair.

More information can be found at the 2018 AAAI/ACM SIGAI Job Fair website: http://www.aaaijob-2018.preflib.org/. We hope that all the participating companies and students had a great time and that we can grow the attendance and reach of the job fair next year. If you have any feedback on things that could be improved or are interested in organizing the next installment of the job fair please get in touch with us!

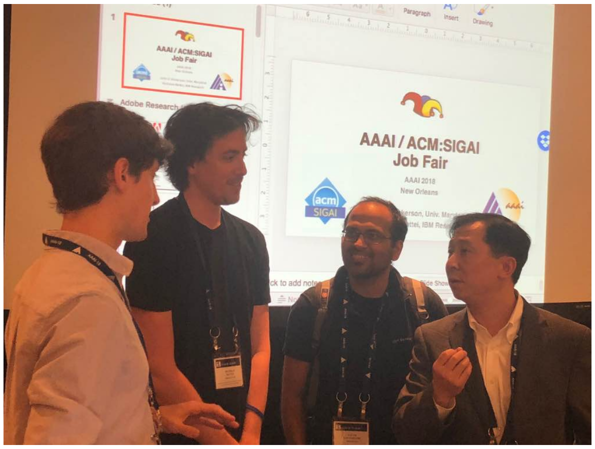
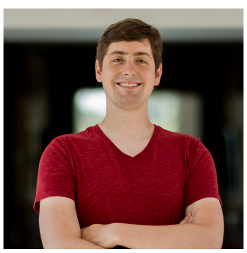


Figure 2: Representatives from each of the participating firms gave single-slide, two-minute pitches to attract job fair attendees to their respective booths.



John P Dickerson is an Assistant Professor of Computer Science at the University of Maryland. His research centers on solving practical economic problems using techniques from computer science, stochastic optimization, and machine learning.

Nicholas Mattei is a Research Staff Member in the IBM Research AI group at the IBM TJ Watson Research Laboratory. His research focuses on the theory and practice of AI, developing systems and algorithms to support decision making.

1st AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society: A Retrospective

Benjamin Kuipers (University of Michigan; kuipers@umich.edu)
Nicholas Mattei (IBM Research; n.mattei@ibm.com)
DOI: 10.1145/3203247.3203250

Introduction

The 1st AAAI/ACM Conference on AI, Ethics, and Society (AIES-18) was held February 1–3, 2018 at the Hilton New Orleans Riverside, in New Orleans, Louisiana. The event was held just before AAAI in order to highlight the overlap of the two conference and logistical support for the conference was provided by AAAI. By attendance measures the conference was a resounding success with a sold out registration of over 300 people. The conference brought together program chairs from the four major focal areas: AI and jobs: Jason Furman (Harvard University); AI and law: Gary Marchant (Arizona State University); AI and philosophy: Huw Price (Cambridge University); and AI: Francesca Rossi (IBM and University of Padova). All the paper from the conference are available for download at the conference website: http://www.aies-conference.com/

Student Program

ACM:SIGAI had a large part in funding and organizing the student program for the conference. With funding from all the conference sponsors each student received a $1,000 travel grant and complimentary registration. The student program was highly competitive with over 70 applicants competing for just 20 spots. The accepted students ran the gambit of conference areas with students from computer science, law, and philosophy represented. Each of the students participated in a special student lunch with all the invited speakers, had a poster during the student poster session, and had the opportunity to publish an abstract of their thesis work in the conference program.

Thursday February 1

AIES began with an evening reception and panel at Tulane University. The panel title was What will Artificial Intelligence bring? Discussing the advent and consequences of superhuman intelligence, and the panelists were Paula Boddington (Oxford), Wendell Wallach (Yale), Jason Furman (Harvard), and Peter Stone (UT Austin). The panel debated in a filled to capacity auditorium at Tulane touching on all the major topics of the conference including AI and Society and the future of AI.

Friday, February 2

The conference opened at the Hilton with a welcome from Francesca Rossi and the rest of the Program Chairs, and the announcement of the two Best Paper Awards.

The first invited speakers were Iyad Rahwan and Edmond Awad from MIT, describing The Moral Machine Experiment: 40 Million Decisions and the Path to Universal Machine Ethics. This very well-known crowd-sourced experiment asks volunteers on the Web to answer questions about a series of scenarios where a speeding self-driving car with failed brakes must choose which of two sets of people will be killed. The sets of people vary over many dimensions (number, gender, age, innocence, passenger vs pedestrian, etc.), and the respondents characteristics (age, gender, nationality, etc.) are also recorded. Many conclusions can be drawn from the collected data. The experimenters emphasize that the purpose of the experiment is descriptive, rather than prescriptive, providing insights into peoples attitudes, rather than determining the right answers to moral questions. Nonetheless, readers persist in interpreting the results prescriptively, and critics raise concerns about the extreme and unrealistic abstraction of the scenarios, and whether participants responses about hypothetical scenarios have meaningful interpretations. The presentation was very stimulating and led to vigorous discussion, a theme that persisted throughout the conference.

Each hour-long session for oral paper presentations gave each of four presenters 10 minutes both to present the paper and respond to a few direct questions, followed by 20 minutes for general questions or comments from the audience directed at any or all of the papers in the session. The first paper session focused on social norm learning and value alignment (including papers by the two authors of this report). The second morning session focused on bias and fairness, especially in machine learning/big data applications. The two afternoon sessions were on the topic of AI and Law, the first focused on Responsibility, and the second focused on Governance.

Then Carol Rose, from the ACLU of Massachusetts, gave a very compelling talk on the current impacts of AI/ML technology on the criminal justice system, and the need for technologists with interest and awareness of ethics, not to mention traditional American values of Liberty and Justice for All, to get directly involved in campaigning, consulting, and advising legislators on how to use these technologies wisely and appropriately.

The final invited panel of the first day was on the important role played by standards, and standard-making bodies, in shepherding social decisions about technology policy. The panel was chaired by Simson Garfinkel (USACM) for organizer John Havens (IEEE) who was unable to attend, and also included Takashi Egawa (NEC), Dan Palmer (British Standard Institute), and Annette Reilly (IEEE).

Each of these events prompted vigorous discussion, but there was plenty of discussion energy left over for the Conference Reception, sponsored by Deep Mind Ethics & Society.

Saturday, February 3

The second day began with an invited talk by Richard Freeman, an economist at Harvard, who reminded us that concerns about AI, robots, and the elimination of jobs are not new, illustrated with a quote by Herbert Simon from 1966. Previous scary predictions have not come true, but of course things are different now, in terms of the comparative advantage of automation over human workers across a wider range of tasks. Furthermore, it is worth observing that economic inequality has been increasing before AI and Robots have had significant economic impacts. At least, whatever happens, its not entirely our fault!

The first morning paper session included a paper from Georgia Tech describing their experience with a “Virtual TA” answering student questions for a Knowledge-Based AI class. One question is whether it is ethically required to tell students which TA is the virtual one, since most interaction is via web pages. Another paper, from McGill, described the potential for bias in data-driven dialog systems, and ways that bias and hate speech can be detected and avoided. In the second morning session, one paper described methods for generating explanations from deep neural networks, and another described “Purple Feed”, an approach to select high-consensus items for a news feed that cuts across traditional political “silos”. The two afternoon paper sessions included papers on regulating autonomous vehicles, the rights of service robots, trust in healthcare AI, non-intuition-based machine ethics, and a survey of when people want AI systems to make decisions (primarily, when the person asked has previous exposure to machines making decisions).

There were many more interesting papers than can be mentioned here. The papers are available on the conference website http://www.aies-conference.com/, so see for yourself.

The day, and the conference, concluded with two keynote presentations. The first was by Patrick Lin, philosopher from Cal Poly State University, who gave an overview of robot ethics. The second was by Tenzin Priyadarshi, the Buddhist chaplain at MIT. He described the “Because we can!” attitude that is often seen in developers of new technologies. He encouraged us to take moral responsibility for the design of intelligent systems, taking the well-being of humans very much into account.

Final Reflection

This conference reminded Ben in some ways of his first IJCAI in 1973. Both were very inspiring about the importance, the breadth, and the promise of the problems and methods being presented. In both cases, the papers were all over the place, presenting interesting results from many perspectives on many different problems, often with little connection to each other. Since then, AI has grown into a much larger intellectual and industrial enterprise, with great impact on society. That very impact suggests that the focus of this conference on AI, ethics, and society will also become increasingly important.

Benjamin Kuipers is a Professor of Computer Science and Engineering at the University of Michigan. His research focuses the representation, learning, and use of foundational domains of knowledge, including knowledge of space, dynamical change in physical systems, objects, actions, and now, ethics.

Nicholas Mattei is a Research Staff Member in the IBM Research AI group at the IBM TJ Watson Research Laboratory. His research focuses on the theory and practice of AI, developing systems and algorithms to support decision making.

Artificial Intelligence in 2027

Maria Gini (University of Minnesota; gini@umn.edu)
Noa Agmon (Bar-Ilan University; agmon@cs.biu.ac.il)
Fausto Giunchiglia (University of Trento; fausto.giunchiglia@unitn.it)
Sven Koenig (University of Southern California; skoenig@usc.edu)
Kevin Leyton-Brown (University of British Columbia; kevinlb@cs.ubc.ca)
DOI: 10.1145/3203247.3203251

Introduction

Every day we read in the scientific and popular press about advances in AI and how AI is changing our lives. Things are moving at a fast pace, with no obvious end in sight.

What will AI be ten years from now? A technology so pervasive in our daily lives that we will no longer think about it? A dream that has failed to materialize? A mix of successes and failures still far from achieving its promises?

At the 2017 International Joint Conference on Artificial Intelligence (IJCAI), Maria Gini chaired a panel to discuss "AI in 2027." There were four panelists: Noa Agmon (Bar-Ilan University, Israel), Fausto Giunchiglia (University of Trento, Italy), Sven Koenig (University of Southern California, US), and Kevin Leyton-Brown (University of British Columbia, Canada). Each of the panelists specializes in a different part of AI, so their visions span the field, providing an exploration of possible futures.

The panelists were asked to present their views on possible futures, specifically addressing what AI technologies they expected would be in widespread use in 2027, what they thought would still show potential but not have become widely accepted, and what they expected the AI research landscape to look like ten years from now.

This article summarizes the main points that each panelist made and their reflections on the topics. The focus in each contribution is not much on predicting the future but on bringing up specific open problems in each subarea and discuss how the current AI technologies could be steered to address them.

Noa Agmon, Bar-Ilan University[1]

The discussion about the fourth industrial revolution, and the part of AI and robotics within it, is wide. In the context of this revolution, autonomous cars and other types of robots are expected to gain popularity and, among other things, to take over human labor. While surveys like "When will AI exceed human performance?" (GSD+17) report that some researchers expect robots to be capable of performing human tasks, such as running five kilometers, within ten years, this will probably take much longer given the current state of robotic development.

Today, the use of robots is generally limited to three categories: non-critical tasks; settings in which robots are semi-autonomous, teleoperated, or remote-controlled (namely, not fully autonomous); and highly structured settings in which uncertainties are minimal. Examples of such settings include the Amazon Robotics warehouse robots, which work autonomously in a structured environment (the warehouse), semi-autonomous drones operated in military settings (usually follow a specified route autonomously, though operative decisions are made by human operators), robots that perform cleaning tasks, which are considered non-critical, Mars rovers, which operate semi-autonomously in unstructured environments, and more. When robots are required to operate fully autonomously in unstructured settings requiring them to handle unbounded uncertainties or completely unpredictable events, they tend to fail. One of many examples is the Knightscope robot, which drove into a fountain on its first day of deployment as a security guard in Washington, D.C.

Rather than arguing about the ability of robots

[1]Acknowledgments: I would like to thank Gal Kaminka and David Sarne from Bar-Ilan University for their helpful comments.

to outperform humans, and when this might happen, the following discussion examines the challenges and opportunities that will influence the development of intelligent robotics in the next ten years.

Dependence on hardware. As opposed to the progress of AI, which relies mainly on algorithmic development and benefits from processing improvements, progress in robotics is also intimately tied to the capabilities of electromechanics, physical sensors, and energy storage and management. Whatever apocalyptic or euphoric visions we have for working with robots, their realization is much more dependent on physical components than we, AI researchers and practitioners, tend to consider. For example, most quad-copters, which are considered to be a basis for breakthrough applications (such as home deliveries and emergency services), can only fly for 30 minutes or so. Likewise, vacuum cleaners are limited in the total area that they can cover before they have to be recharged. These energy concerns radically impact the usefulness of robots in applications which are otherwise within reach from a pure software perspective.

The good news is that the intimate connection between software and hardware works both ways. Just as modern SLAM algorithms (e.g., (DNC+01)) were able to overcome intrinsic sensor limitations to create reliable and accurate maps for navigation, advances in software can overcome some of the limitations posed by hardware.

AI influences on robotics. AI algorithms influence robotics not only in compensating for and improving the utilization of existing hardware capabilities, but also in enabling new tasks. Progress in natural language processing (NLP) and machine learning (used for chatbots, personal digital assistants, and surveillance, for instance) enables more natural forms of human-robot interaction with physical robots, and autonomous cars. However, such positive influences are somewhat asymmetric: AI will influence robotics more than robotics will influence AI. A personal robot benefits from NLP more than NLP can benefit from the consideration of multi-modal interactions (as in “talking with your hands.”)

Growing role for multi-robot systems (MRS). The academic research on MRS dates back to the early 1980’s, when robots were scarce and not autonomous. Research has progressed far beyond the deployment of such systems outside of labs. Improvements in the reliability of robots will make it easier to deploy MRS in various applications, continuing and accelerating current successful trends (e.g., in warehouses and hospitals). This, in turn, will accelerate research on fully distributed, fully autonomous systems, which are beyond current capabilities. It is obvious that human-robot interactions will be a major focus of research in the next ten years, as robots enter a greater number of unstructured environments in which humans operate. However, given the foreseen growth in the role of multi-robot systems, human-MRS and multiple operator-single robot collaborations will likely see increased efforts.

Increasing ties with other disciplines. A good example of large-scale fully distributed, fully autonomous systems also raises an additional trendthat of increasing ties with other disciplines. Swarms of molecular robots (nanobots), the size of which is measured in nanometers, are becoming a reality in medical applications (for example, targeted drug delivery). Trillions of such robots will be let loose in a patient’s body - the largest-scale MRS in robotics history. The computation of interactions between different types of nanobots has an immense impact on the ease and duration of development of new treatments (WKKH+16; KSSA+17). The capability to plan and reason about the interactions of these robots with each other and with the body requires deep collaboration between AI experts, biologists, and chemists.

Another example is reconfigurable robots, which can transform themselves into different shapes, depending on the environment and task. Research on such systems will benefit from close coordination with chemistry, physics, and biology, to take new findings into account.

Increasing accessibility, lower entry barrier, greater impact potential. A positive trend which I believe will continue to grow is the lowering of the entry barrier into robotics practice and research, at multiple levels. Research-grade robots for labs have seen dramatic decreases in cost, and the common availability of 3D printing, cheap embedded computers (Arduino, Raspberry Pi), as well as continued push on STEM education will make development of robots cheaper and easier than ever. The availability of common robot software middleware, such as ROS, make it easier for researchers to focus their attention on bringing their expertise to bear on specific components.

Bottom line: More of the same (which is good!) Within the next ten years and beyond, we will not see general-purpose robots. That is, robots will still be dedicated to one task, for example delivery, cleaning, or surveillance. Progress in the development of intelligent robotic systems will continue to focus on excellence in the performance of specific tasks, and on the introduction of new tasks to new types of robots. To some extent, this will make robot use more popular.

Fausto Giunchiglia, University of Trento[2]

Providing machines with knowledge, e.g., common sense or domain knowledge, has always been one of the core AI issues. Two are the main approaches to this problem. The first *deductive* approach, usually categorized under the general heading of "Knowledge Representation and Reasoning (KRR)", dates back to John McCarthy's *advice taker* proposal (McC60), and it is based on the idea of telling machines what is the case, for instance in the form of facts codified as logical axioms. The second *inductive* approach, usually categorized under the general heading of "Machine Learning (ML)", consists of providing machines with a set of examples from which to learn general statements, usually with a certain level of confidence.

A lot of relevant research has been be done in KRR, not least the work on the Semantic Web (BLHL01), and much more will be done. However, the success of ML mainly, but not only, because of the work in Deep Learning (see, e.g., (LBH15)), has been so overwhelming that, thinking of what the research in KRR could be in the next ten years and where it could lead, a relevant question is the extent to which these two lines of work should integrate in an effort to jointly produce results that either of them alone could not produce.

A lot of successful work in this area has been done, see for instance (GT07; RKN16). However, the extent to which the KRR research could be improved by exploiting the research developed in ML, or dually, the extent to which ML would *really* need to exploit any of the results developed in KRR is still unclear, at least for two reasons. The first is that this integration is far from being trivial. It is a fact that these two approaches start from somewhat opposite assumptions, the first assuming that knowledge consists of a set of facts which are either true or false, with nothing in between, the second having to deal with the issue that any fact learned via ML will hardly ever be guaranteed to be true or false with an infinite number of intermediate levels. Furthermore, the need for such an integration is far from being clear, at least from an ML point of view. Among other things, it is a fact that the current ML techniques have proven so powerful that, whenever applicable, they seem to be able to learn virtually unbound amounts of knowledge, far more knowledge than could be codified by any knowledge engineer.

At the same time both the KRR and ML techniques have their own weaknesses. Thus, on one side, KRR presents a main difficulty in how to express the inherent complexity and variability of the world, in particular but not only, when perception is involved. On the other side, instead, ML presents a main difficulty in making sense, in human terms, of the knowledge which is learned. In other words, the knowledge generated via ML does not often fit the people "intended semantics", namely how they would describe what is the case, for instance as perceived or as learned from a large amount of text messages.

This difficulty of ML techniques, and data driven approaches in general, has been known for many years. An explicit reference to this problem comes from the field of Computer Vision, where it is named the *Semantic Gap Problem (SGP)*. The SGP was originally defined in (SWS+00) as follows: "... The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation. ..." It can be noticed how this notion is completely gen-

[2]Acknowledgments: I would like to thank Kobi Gal, Daniel Gatica-Perez, Loizos Michael, Daniele Miorandi, Andrea Passerini and Carles Sierra for our many useful discussions on this topic.

eral and can be taken to refer to the human-machine misalignment which may arise with any type of information that a machine can extract, i.e., learn, from any type of data. The novelty of these last years is that many more instances of the SGP are showing up and many of them are also discussed in the news, the main reason being the increased use, increased power, and increased popularity of ML systems and AI in general. Thus, we have read of cases when a system learns biased opinions, or when it learns a language that it is not human, or when an autonomous car does not track another car thus causing an accident. And this, in turn, is the cause of a lot of public discussions about the interaction between AI and humans, about AI and ethics and also of an increased fear of AI.

A convincing explanation of how to deal with the SGP seems necessary for AI to be used beyond a set of niche (possibly very large) application areas and to be adopted by the general public. I believe it will be very hard to convince people to use machines that they do not feel they fully control, in high value application domains, e.g., health, mobility, energy, retail. But a convincing explanation will not be enough. A solution of the SGP is also needed for AI to be used in practice. There are at least three mainstream application scenarios where some solution to the SGP seems crucial. The first is the *anytime anywhere delivery of personalized services*, as enabled by personal digital devices, e.g., smart phones or smart watches. But for this to happen, people will have to be able to make sense of why certain decisions have been taken by the machine, and to agree with them. The second is the *empowerment of social relations*, exactly for the same reasons mentioned above. Facebook, Whatsapp, or Snapchat are just the beginning and I foresee the rise of a new generation of social networks empowering more specialized, more personalized, more diversity-aware interactions among people. The third, and maybe the most important, again because of the pervasiveness of digital devices, is that we are more moving towards *open world* application scenarios. By this I mean application scenarios where, at design time, it is impossible to anticipate the system functional and non-functional requirements. In this type of applications the effects of the SGP can be devastating, as the divergence between people and machines can only get worse in time.

In my opinion, a general solution of the SGP problem, and in particular a solution which is viable in open world application scenarios, can only be achieved via a tight integration of knowledge-based approaches and machine learning. The knocking down argument is that the only way to avoid the SGP is to make sure that machines learn representations of the world which are the same as their reference users. But, the fact that knowledge should be presented in human-like terms is exactly the assumption underlying all the work in KRR and also logic. More specifically, whatever knowledge will be learned via a data-driven approach, it will have be compared and ultimately aligned to the human knowledge. Someone could argue that this is exactly what supervised learning does. But this is not the case, as also witnessed by the fact that even the human supervision, how it is implemented up to now, does not make the SGP disappear. The problem is far more complex and it will require major advances in both KRR and ML, and in AI in general, many of which, I believe, will be disruptive. A list of four open issues is provided below, with the understanding that this list is not meant to be complete nor correct. This list reflects only my current personal understanding of some of the problems which will have to be dealt with when trying to solve the SGP.

1. Since the early days of AI a fundamental issue has been that of building machines which would *exceed human-level intelligence*. This goal has been reached in many domains, e.g. chess or GO playing, while it is very far from being reached in other domains, e.g., robotics, as mentioned above. A solution of the SGP will require building machines which will show *human-like intelligence*, representation and reasoning, as the basis for the mutual human-machine understanding. In this context, exceeding human-level intelligence seems a desired property but not strictly necessary.

2. A fundamental property of life, and of humans in particular, is their ability to *adapt to unpredicted events and evolve*. Both the research in KRR and in ML seems very far from achieving this goal. It is however interesting to notice how a particular instance

of this inability to adapt was recognized by John McCarthy and named *the problem of lack of generality* of the current representation formalisms (McC87).

3. The net result of the ability to adapt and evolve as a function of the local context will be that the resulting knowledge will be highly diversified. In turn, the *diversity of knowledge* will generate the need for further adaptation in an infinite loop with will result in the process of *knowledge evolution*, somewhat analogously to the kind of evolution we see in life. Notice how the proposed approach is quite different from that taken by the Semantic Web for the solution of the problem of semantic heterogeneity. The focus is not on representation tools, e.g., ontologies, or formalisms, e.g., Description Logics, but on the process by which knowledge gets generated, stored, manipulated, and used. In this perspective the standard logics, e.g., monotonic non-monotonic logics, seem to solve, at most, only part of the problem, and for sure not the most important.

4. A specific, but core, subproblem of the problem of managing knowledge diversity, is the integration of the knowledge obtained via perception, e.g., via computer vision, and the knowledge obtained via reasoning or by being told. An implicit assumption which has been made so far is that the linguistic representation of an object we talk about, e.g., the word "cat", and the representation of what we perceive as a cat is one-to-one. As discussed in detail in (GF16) this in general is not the case and there is a many-to-many mapping between linguistic representations and perceptual representations. On top of this, these mappings are highly dependent on the culture and on the single person and, even for the same person, change in time, as a function of the person current interests. A full understanding of how these mappings are built, of how linguistic representations influence the construction of perceptual representations, and vice versa, is a largely unexplored research area. Still, some form of solution to this problem will be needed in order to guarantee that the machine will describe what it will perceive coherently with what humans do.

Sven Koenig, University of Southern California[3]

AlphaGo (SHM+16) shows that it can be very difficult to judge technical progress, as also noticed by Stuart Russell in his invited IJCAI-17 talk. When it beat Lee Sedol in 2016, many experts thought that such a win was still at least a decade away. The AI techniques behind it already existed in principle. The ingenuity was in figuring out how to put them together in the right way. Progress on AI technology is often steadier than it appears, yet such engineering breakthroughs happen only from time to time, are difficult to predict, and often make AI technology visible in the public eye - creating the perception of waves of progress.

Various recent studies shed light on the expected progress of AI by 2027, such as the study on "AI and Life in 2030" as part of the One Hundred Year Study on AI (ai100.stanford.edu) and a recent survey of all ICML-15 and NIPS-15 authors (GSD+17). This survey, for example, predicts that AI will outperform humans around 2027 on tasks such as writing high school essays, explaining actions in games, generating top 40 pop songs, and driving trucks. Furthermore, humanoid robots will soon afterward beat humans in 5k races. Interestingly, North American researchers predicted that it will take about 74 years to reach high-level machine intelligence across human tasks, while Asian researchers thought it would take only 30 years. Indeed, there is currently lots of excitement and optimism, for example, in China about the potential of AI with large investments into application-oriented AI research by both the government and private sector.

In the following, I view AI as the study of agents to structure the discussion which kinds of research topics will be popular in 2027. I distinguish rational agents (that make good decisions), believable agents (that interact like humans), and cognitive agents (that think like humans). A large amount of AI research currently focuses on building rational agents on the task level - by studying single AI techniques in isolation and applying them to single tasks, resulting in narrowly intelligent agents. The current excitement about AI is often based on

[3]Acknowledgments: I would like to thank Paul Rosenbloom and Wolfgang Hönig from the University of Southern California for helpful comments.

the power of a small set of AI techniques combined with the availability of large amounts of data (due to progress on sensor technologies and the ubiquity of both smart phones and the internet) as well as progress in robotics for the embodiment of AI. For example, the term “big data” is typically used to characterize the current AI era, driven by the capability of machine learning techniques. In fact, 49 percent of submissions to the International Joint Conference on AI (IJCAI) in 2017 used as first keyword “machine learning,” which is about acquiring good models of the world. However, these models need to serve a bigger purpose, for example, to make good decisions. While machine learning can sometimes acquire evaluation functions that help with making good decisions (as AlphaZero (SHS+17), the successor of AlphaGo shows), it often requires lots of data, has limited capability for transfer, and has difficulty integrating prior knowledge (Mar16) and thus is of limited help for making decisions in novel or dynamic environments. Perhaps the term “big decisions” will be used to characterize the AI era around 2027, driven also by the capability of AI planning and similar AI techniques. Current faculty hiring in the US lags in research areas such as AI planning although the research community is already heading in that direction. For example, the popular textbook by Stuart Russell and Peter Norvig (RN09) views AI as the study of rational agents and thus essentially as a science of making good decisions with respect to given objectives. But many other disciplines could be characterized similarly, including operations research, decision theory, economics, and control theory (Koe12). AI researchers make use of techniques from some of these disciplines already. For example, the textbook by Stuart Russell and Peter Norvig discusses utility theory (from decision theory), game theory and auctions (from economics), and Markov decision processes (from operations research), yet research collaborations across these and other disciplines are still developing, which is why we should reach out more to researchers in other decision-making disciplines. There already exist some good but narrow interfaces, such as the Conference on the Integration of Constraint Programming, AI, and Operations Research (CPAIOR) or the ACM Conference on Economics and Computation (EC). There also exists an attempt to put a broader interface in place, namely the International Conference on Algorithmic Decision Theory (ADT), which ”seeks to bring together researchers and practitioners coming from diverse areas such as AI, Database Systems, Operations Research, Discrete Mathematics, Theoretical Computer Science, Decision Theory, Game Theory, Multiagent Systems, Computational Social Choice, Argumentation Theory, and Multiple Criteria Decision Aiding in order to improve the theory and practice of modern decision support” (sma.uni.lu/adt2017). Such interdisciplinary integration can result in economic success. CPAIOR, for example, started in 2004 (preceded by five workshops) and still thrives. In parallel, ILOG successfully integrated software for constraint programming and linear optimization and was acquired by IBM in 2009. My hope is that we will have a thriving conference on intelligent decision making by 2027 that will be attended by researchers from all decision-making disciplines, including AI. Of course, the different decision-making techniques also need to be integrated into systems. AI can lead the way by developing agent architectures with good theoretical foundations for how different parts should interact, resulting in more broadly intelligent agents on the job level (that is, across tasks). This is no simple feat as the restricted applications of current robot architectures show. Integrating decision-making techniques from different disciplines is even more difficult, for example, because of their different assumptions (often due to different application areas studied by different disciplines) and different ideas about what constitutes a good solution (due to disciplinary training), which is why we should start to give students multidisciplinary training in decision making.

While rational agents will continue to be important, human-aware agents will become more and more important and, with them, also believable agents that allow for interaction with gestures, speech, and other human-like modalities, understand human conventions and emotions, predict human behavior, and - in general - appear to be human-like. We already use intelligent assistants on a variety of platforms (such as Apple’s Siri or Amazon’s Echo) and will soon routinely have conversations - including negotiations - with all kinds of apparatus, perhaps including our elevators and toilets ☺.

The progress on cognitive agents, one of the early dreams of AI, is more difficult to judge. The research community currently works on hybrid approaches that combine ideas from symbolic, statistical, and/or neural processing and on a community-wide "Common Model of Cognition" (KT16).

Finally, AI researchers and practitioners slowly gain an understanding that they should not just develop AI techniques but also have some say in how they are being used. We need to ask ourselves questions such as:

> Do we need to worry about the reliability, robustness, and safety of AI systems and, if so, what to do about it? How do we guarantee that their behavior is consistent with social norms and human values? Who is liable for incorrect AI decisions? How to ensure that AI technology impacts the standard of living, distribution and quality of work, and other social and economic aspects in the best possible way? (BGK+17)

AAAI and ACM recently co-founded the AAAI/ACM Conference on AI, Ethics, and Society (AIES, www.aies-conference.com) to come up with answers to these questions. AIES was filled to capacity. I hope that AIES and its topics will be even more popular in 2027.

Kevin Leyton-Brown, University of British Columbia[4]

It is a daunting task to predict the direction AI research will take a decade from now, particularly given the checkered history of such prognostication in the past. In an attempt to go beyond idle speculation, I have therefore structured this reflection around three different *approaches* a forecaster might use in making such predictions. Despite recognizing the likelihood that some of what follows will appear foolish in retrospect, I strive to make bold claims about what the future will hold. I hope that AI researchers of 2027 will forgive me!

I. Forecasting via prototypes. Ten years sounds like a long time, but in fact it takes about that long for technologies to move from the lab to widespread practice: the transformative technologies of today existed in prototype form a decade ago. One approach to AI forecasting is thus to look at today's prototypes and to imagine their more widespread deployment.

Broadly speaking, today's AI prototypes offer tailored solutions for specific tasks rather than general intelligence. Some AI research topics that I expect to see making considerably broader social impact by 2027 include:

- Non-text input modalities (vision; speech)
- Consumer modeling (recommendation; marketing)
- Cloud services (translation; question answering; AI-mediated outsourcing)
- Transportation (automated trucking; some self-driving cars)
- Industrial robotics (factories; some drone applications)
- AI knowledge work (logistics planning; radiology; legal research; call centers)
- Policing & security (electronic fraud; cameras; predictive policing)

By considering where today's prototypes have achieved less traction, it is also possible to forecast sectors in which AI technologies are less likely to take off quickly. Overall, these are often areas in which major entrenched regulatory regimes need to be navigated; where there exist substantial social or cultural barriers to the adoption of new technologies; and/or where broad impact would depend on nontrivial hardware breakthroughs. Many such sectors are the focus of concerted research today and are likely to remain important in the research landscape in 2027; however, I believe that they are less poised for short-term practical impact. Some key examples are childcare, healthcare, and eldercare; education; consumer robots beyond niche applications; and semantically rich language understanding.

II. Forecasting via consumer desires. A second strategy is to assume that investment, entrepreneurial energy, and industrial R&D will focus on meeting consumer needs that are already apparent today, and hence that these areas will see future breakthroughs.

[4]Acknowledgments: I would like to thank the members of the AI100 2015–16 Study Panel for helping to shape my thinking about the future of AI: P. Stone, R. Brooks, E. Brynjolfsson, R. Calo, O. Etzioni, G. Hager, J. Hirschberg, S. Kalyanakrishnan, E. Kamar, S. Kraus, D. Parkes, W. Press, A. Saxenian, J. Shah, M. Tambe, A. Teller (SBB+16).

Labor automation. A fundamental consumer need is for someone else to perform unpleasant, routine tasks. The promise of automating such tasks has been part of the AI story from the beginning (e.g., Shakey the robot delivering coffee in an office setting (Nil84)) and is increasingly becoming a reality (e.g., robot vacuum cleaners in the home; ordering books via Amazon Alexa). However, there is much scope for additional innovation in this space, centering on currently unaddressed tasks to which large numbers of people currently devote considerable time. Some potential examples are household cleaning, yard work, pet care, shopping, and food preparation. Some needs may be met by directly replacing human with robotic labor; others may be met via "gig economy" platforms that use AI on the back end to more efficiently allocate human labor; and still others may be met in entirely new ways, such as by combining AI-driven logistics platforms with centralized industrial processes (e.g., replacing supermarkets with apps, warehouses, and courier services).

Social connection. We are highly social creatures, and are willing to pay handsomely for technologies that help us to make and strengthen connections with others. Current instantiations of such technologies (e.g., social networks; remote work platforms; online dating) are highly valuable, but relatively primitive from an AI perspective, relying mainly on micro-blogging, direct messaging, user modeling, and newsfeed curation. There is scope for more AI mediation of social connection, reducing the frictions that prevent people from easily finding others to interact with in the moment and making those interactions richer.

Entertainment. Our research community's focus on solving industrially or socially important problems sometimes may cause us to pay insufficient attention to AI's potential for transforming the entertainment industry, which addresses another fundamental consumer need. Gaming is already bigger than Hollywood (Che17), but the future of AI in entertainment will go far beyond what we now see as computer games. Future AI entertainments will increasingly be interactive and multimodal, and will intersect with sectors we now see as distinct, such as fitness, learning, performing useful tasks, and spending quality time with friends. AI will also play an increasingly critical role in the creation, delivery, and personalization of traditional, broadcast entertainment such as TV.

Education. Education is poised to grow as a consumer sector (ROO18), both as workers respond to the need to reskill and as individuals with extra leisure time follow their passions. I argued above that education will not be transformed by AI in a decade; however, particularly because of the dual role many AI researchers hold as educators, there is nevertheless considerable scope for AI technology to make incremental progress in improving the content and delivery of educational materials. Some examples include tailoring lessons to a student's skill level, making exercises more interactive, facilitating communication between both peers and instructors outside classroom settings, and reducing the drudgery currently entailed by grading student work. Such innovations could improve student outcomes, lower the cost of education, and broaden its reach.

III. Forecasting via extrapolation. A final strategy is to ask what we will be concerned with if current progress in AI continues.

AI beyond ML. Much recent progress in AI has arisen from improved techniques for learning to make predictions: finding a model that is currently built by hand and replacing it by a model that is learned from data (LBH15). We might therefore ask what problems would remain or become important if our capacity to build black-box models from data were to become arbitrarily effective. It is clear that even in such a world, we would be far from having achieved strong AI. Some problems that would remain open are still close to machine learning: explaining why a model made the prediction it did, or certifying fairness or compliance with legal requirements. Others, such as making counterfactual predictions (how would a system perform under a perturbation of the generating distribution?), go beyond the assumptions inherent in most supervised learning methods, requiring instead new, structural assumptions about an underlying setting. Still other problems extend beyond prediction to decision making, both in single-actor settings (e.g., optimization; planning) and multi-agent domains (weighing competing objectives via preference aggregation or mechanism design).

Increasing regulation. AI is touching the lives of individuals, the economy, and the political system in ever increasing ways. Many in society find specific instantiations of AI frightening; many special interests are threatened by new technologies. It is thus inevitable that politicians will increasingly see a need to respond, and that AI technologies will face increasing regulation. This is something we should welcome; any mature technology must be accountable to the society in which it operates. But the details will matter enormously. A major focus of AI research in 2027 will be helping to shape regulations before they become law and designing systems within the constraints implied by these regulations afterwards.

Superhuman intelligence. AI systems will increasingly become capable of reaching human-level performance in a variety of application domains. There is nothing special about this threshold, and so we should expect the advent of AI systems exhibiting superhuman intelligence in a growing set of domains. This is often cast as a frightening prospect, but I argue that we will quickly become comfortable with it. After all, superhuman intelligences are already commonplace: governments, corporations, and NGOs are all autonomous agents that exhibit behavior much more sophisticated and complex than that of any human. We are typically unconcerned that no one person can even fully understand decisions made by the French government, by General Motors, or by the Red Cross. Instead, we aim to manage and to gain high-level understanding about such actors via reporting requirements, specifications of the interests that they must act to advance, and laws that forbid bad behavior. Society encourages the creation of such superhuman intelligences today for the same reason it will welcome superhuman AI tomorrow: many important problems are beyond the reach of individual people. Some key examples are improved collective decision making; more efficient allocation and use of scarce resources; addressing under-served communities; and limiting and responding to climate change.

References

E. Burton, J. Goldsmith, S. Koenig, B. Kuipers, N. Mattei, and T. Walsh. Ethical considerations in Artificial Intelligence courses. *Artificial Intelligence Magazine*, 38(2):22–34, 2017.

Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.

Chet Chetterson. How games overtook the movie industry, 2017. http://www.geekadelphia.com/2017/06/29/how-games-overtook-the-movie-industry, Accessed March 22, 2018.

M.W.M. Gamini Dissanayake, Paul Newman, Steve Clark, Hugh F Durrant-Whyte, and Michael Csorba. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, 17(3):229–241, 2001.

Fausto Giunchiglia and Mattia Fumagalli. Concepts as (recognition) abilities. In *Formal Ontology in Information Systems: Proc. 9th Int. Conference (FOIS 2016)*, pages 153–166, 2016.

K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans. When will AI exceed human performance? Evidence from AI experts. ArXiv Preprint arXiv:1705.08807v2, 2017.

Lise Getoor and Ben Taskar. *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2007.

S. Koenig. Making good decisions quickly. *The IEEE Intelligent Informatics Bulletin*, 13(1):14–20, 2012.

Gal A. Kaminka, Rachel Spokoini-Stern, Yaniv Amir, Noa Agmon, and Ido Bachelet. Molecular robots obeying Asimov's three laws of robotics. *Artificial Life*, 23(3):343–350, 2017.

I. Kotseruba and J. Tsotsos. A review of 40 years of cognitive architecture research: Core cognitive abilities and practical applications. ArXiv Preprint arXiv:1610.08602, 2016.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

G. Marcus. Deep learning: A critical appraisal. ArXiv Preprint arXiv:1801.00631, 2016.

John McCarthy. *Programs with common sense*. RLE and MIT Computation Center, 1960.

John McCarthy. Generality in Artificial Intelligence. *Communications of the ACM*, 30(12):1030–1035, 1987.

Nils J Nilsson. Shakey the robot. Technical report, SRI International, Menlo Park, CA, 1984.

Luc De Raedt, Kristian Kersting, and Sriraam Natarajan. *Statistical Relational Artificial Intel-*

ligence: Logic, Probability, and Computation. Morgan & Claypool Publishers, 2016.

S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, 2009.

Max Roser and Esteban Ortiz-Ospina. Financing education, 2018. https://ourworldindata.org/financing-education, Accessed March 22, 2018.

Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller. Artificial intelligence and life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, 2016.

D. Silver, A. Huang, C. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

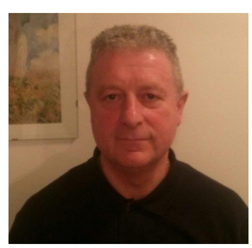
D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis. Mastering Chess and Shogi by self-play with a general reinforcement learning algorithm. ArXiv Preprint arXiv:1712.01815, 2017.

Arnold WM Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

Inbal Wiesel-Kapah, Gal A. Kaminka, Guy Hachmon, Noa Agmon, and Ido Bachelet. Rule-based programming of molecular robot swarms for biomedical applications. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3505–3512, 2016.

Maria Gini is a Professor in the Department of Computer Science and Engineering at the University of Minnesota. She studies decision making for autonomous agents for distributed task allocation for robots, robotic exploration, teamwork for search and rescue, and navigation in dense crowds. She is a AAAI and a IEEE Fellow, and current president of IFAAMAS. She is Editor-in-Chief of Robotics and Autonomous Systems, and is on the editorial board of numerous journals, including Artificial Intelligence, Autonomous Agents and Multi-Agent Systems, and Integrated Computer-Aided Engineering.

Noa Agmon is an Assistant Professor in the Department of Computer Science at Bar-Ilan University, Israel. Her research focuses on various aspects of single and multi robot systems, including multi-robot patrolling, formation, coverage and navigation, with specific interest in robotic behavior in adversarial environments. She is a board member of IFAAMAS, and on the editorial board of JAIR and IEEE Transactions on Robotics.



Fausto Giunchiglia is a Professor in the Department of Computer Science and Information Engineering (DISI) at the University of Trento, Italy. His current research is focused on the integration of perception and knowledge representation as the basis for understanding how this generates the diversity of knowledge. His ultimate goal is to build machines which reason like humans. HE is an ECCAI Fellow. He was a President and a Trustee of IJCAI, an Associate Editor of JAIR, and a member of the Editorial Board of many journals, including the Journal of Data Semantics and the Journal of Autonomous Agents and Multi-Agent systems.

Sven Koenig is a professor in computer science at the University of Southern California. Most of his research centers around techniques for decision making that enable single situated agents (such as robots or decision-support systems) and teams of agents to act intelligently in their environments and exhibit goal-directed behavior in real-time, even if they have only incomplete knowledge of their environment, imperfect abilities to manipulate it, limited or noisy perception or insufficient reasoning speed. He is a AAAI and AAAS fellow and chair of ACM SIGAI. Additional information about Sven can be found on his webpages: idm-lab.org.

Kevin Leyton-Brown is a Professor of Computer Science at the University of British Columbia in Vancouver, Canada. He studies the intersection of computer science and microeconomics, addressing computational problems in economic contexts and incentive issues in multiagent systems. He also applies machine learning to various problems in artificial intelligence, notably the automated design and analysis of algorithms for solving hard computational problems. He is a AAAI Fellow, chair of ACM SIGecom, an IFAAMAS board member, and has served as an Associate Editor for AIJ, JAIR, and ACM-TEAC.

AI Education Matters: Teaching Hidden Markov Models

Todd W. Neller (Gettysburg College; tneller@gettysburg.edu)
DOI: 10.1145/3203247.3203252

Introduction

In this column, we share resources for learning about and teaching Hidden Markov Models (HMMs). HMMs find many important applications in temporal pattern recognition tasks such as speech/handwriting/gesture recognition and robot localization. In such domains, we may have a finite state machine model with known state transition probabilities, state output probabilities, and state outputs, but lack knowledge of the states generating such outputs. HMMs are useful in framing problems where external sequential evidence is used to derive underlying state information (e.g. intended words and gestures).

Video Introductions

While there are many videos online dedicated to the topic of HMMs, I'll highlight two here. Daphne Koller's 12-minute video"Template Models: Hidden Markov Models - Stanford University"[1] provides a brief application-focused overview of HMMs and can set a basic context and expectation for the value of further learning in this area.

A full 52-minute UBC lecture by Nando de Freitas, "undergraduate machine learning 9: Hidden Markov models - HMM"[2], is a much-recommended classroom introduction to the general topic of inference in probabilistic graphical models, with focus on HMM representation, prediction, and filtering.

Texts and Articles

Russell and Norvig's *Artificial Intelligence: a modern approach* (Russell & Norvig, 2009) gives a brief introduction to HMMs in §15.3, and learning HMMs as an application of the expectation-maximization (EM) algorithm is described in §20.3.3.

[1]https://youtu.be/mNSQ-prhgsw

[2]https://youtu.be/jY2E6ExLxaw

Of the many options to give focused attention to HMMs, consider *Speech and Language Processing* by Daniel Jurafsky and James H. Martin (Jurafsky & Martin, 2009). The draft third edition chapter 9 on HMMs is currently freely available from Jurafsky's website[3].

Also often recommended is Lawrence Rabiner's tutorial (Rabiner, 1990)[4].

Christopher Bishop's *Pattern Recognition and Machine Learning* (Bishop, 2006) §13.2 covers HMMs, maximum likelihood parameter estimation, the forward-backward algorithm, the sum-product algorithm, the Viterbi algorithm, and extensions to HMMs.

More text and article resources have been recommended in the StackExchange thread "Resources for learning Markov chain and hidden Markov models"[5].

Other Resources

Numerous MOOCs touch on HMMs. Udacity's "Intro to Artificial Intelligence" course by Peter Norvig and Sebastian Thrun has coverage of HMMs[6], as does "Artificial Intelligence - Probabalistic Models"[7]. In Coursera's Peking Univeristy Bioinformatics course, Ge Gao lectures on HMMs[8]. There is no shortage of online course materials for HMMs.

[3]https://web.stanford.edu/~jurafsky/slp3/9.pdf

[4]http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/tutorial%20on%20hmm%20and%20applications.pdf

[5]https://stats.stackexchange.com/questions/3294/resources-for-learning-markov-chain-and-hidden-markov-models

[6]https://www.udacity.com/course/intro-to-artificial-intelligence--cs271

[7]https://www.udacity.com/course/probabalistic-models--cx27

[8]https://www.coursera.org/learn/bioinformatics-pku/lecture/7pbUo/hidden-markov-model

Jason Eisner has written about his spreadsheet[9] for teaching the forward-backward algorithm (Eisner, 2002). More diverse resources, including toolkits and open-source algorithm implementations, may be sampled via the Quora question "What are some good resources for learning about Hidden Markov Models?"[10].

Model AI Assignments

Model AI Assignments are free, peer-reviewed assignment materials made available in order to advance AI education. Two Model AI Assignments to date experientially teach HMMs:

Sravana Reddy's recent Model AI Assignment "Implementing a Hidden Markov Model Toolkit"[11] is targeted to advanced undergraduates and beginning graduate students seeking an introduction to Natural Language Processing (NLP). In the assignment, students "implement a toolkit for Hidden Markov Models with discrete outputs, including (1) random sequence generation, (2) computing the marginal probability of a sequence with the forward and backward algorithms, (3) computing the best state sequence for an observation with the Viterbi algorithm, and (4) supervised and unsupervised maximum likelihood estimation of the model parameters from observations, using the Baum Welch algorithm for unsupervised learning."

John DeNero and Dan Klein's popular Model AI Assignment "The Pac-Man Projects" has a probabilistic tracking project, "Project #4: Ghostbusters"[12] in which "probabilistic inference in a hidden Markov model tracks the movement of hidden ghosts in the Pac-Man world. Students implement exact inference using the forward algorithm and approximate inference via particle filters."

[9] http://cs.jhu.edu/~jason/papers/eisner.hmm.xls

[10] https://www.quora.com/What-are-some-good-resources-for-learning-about-Hidden-Markov-Models

[11] http://modelai.gettysburg.edu/2017/hmm

[12] http://modelai.gettysburg.edu/2010/pacman/projects/tracking/busters.html

Your Favorite Resources?

If there are other resources you would recommend, we invite you to register with our wiki and add them to our collection at http://cs.gettysburg.edu/ai-matters/index.php/Resources.

References

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Eisner, J. (2002). An interactive spreadsheet for teaching the forward-backward algorithm. In *Proceedings of the acl-02 workshop on effective tools and methodologies for teaching natural language processing and computational linguistics - volume 1* (pp. 10–18). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from https://cs.jhu.edu/~jason/papers/eisner.tnlp02.pdf

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing* (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall.

Rabiner, L. R. (1990). A tutorial on hidden markov models and selected applications in speech recognition. In A. Waibel & K.-F. Lee (Eds.), *Readings in speech recognition* (pp. 267–296). San Francisco, CA, USA: Morgan Kaufmann.

Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (3rd ed.). Upper Saddle River, NJ, USA: Prentice Hall.

Todd W. Neller is a Professor of Computer Science at Gettysburg College. A game enthusiast, Neller researches game AI techniques and their uses in undergraduate education.

Obituary: Jon Oberlander

Aaron Quigley (University of St Andrews; aquigley@st-andrews.ac.uk)
DOI: 10.1145/3203247.3203253

Introduction

Professor Jon Oberlander (University of Edinburgh) passed away suddenly on the 19th of December 2017. Jon was remembered by friends, family and colleagues, at a deeply moving ceremony in the Playfair Library at the University of Edinburgh this January. Stories from his life as a friend and father attest to the deep impact Jon had on all those around him, and the lasting legacy he now leaves behind.

Jon was born in Edinburgh, to which he returned after his undergraduate studies in Pembroke College, Cambridge in 1983. He remained in the University of Edinburgh for his PhD studies and then as a postdoctoral fellow, research associate, lecturer, and reader, before finally being promoted to Professor in 2005. His Personal Chair in Epistemics speaks to his grounding in Philosophy and his life long interest in the philosophical theory of knowledge and its scientific study.

In Cognitive Science, Jon was well known and regarded for his work in linguistic representational systems. He described his work as "getting computers to talk (or write) like individual people". This resulted in research efforts to both understand how people express themselves, while also developing systems that might adapt to people. This view of the world resonated throughout his work with collaborators around Scotland and across the world in languages and social science, cultural heritage, psychology, linguistics and computer science.

Jon co-founded the Scottish Informatics and Computer Science Alliance (SICSA) in 2008 and served as its first director and later as its graduate academy director. Jon helped lay the foundations for the international excellence in University-led research, education, and knowledge exchange activity in Computer Science and Informatics in SICSA today. Jon was instrumental in driving SICSA forward in a time of great change in the University sector in Scotland. His resolve and good humour helped draw together this community of scholars in a spirt of cooperation and collaboration.

In the past few years Jon had academic responsibility for the development of the University of Edinburgh's new Bayes Centre for Data Science and Technology. In time, he became the assistant principal for Data Technology and the Director of the Bayes Centre.

Working with Jon first on the Smart Tourism and later Palimpsest: Literary Edinburgh projects was a great opportunity for me. His curiosity and creativity brought a sense of deep wonder and inspiration to our interdisciplinary team. He helped build bridges and acted to translate the aspirations and ideas of researchers from many fields of inquiry. These projects, and countless more, are all the better for having had a touch of Jon's intellectual energy, rigour and sense of fun.

"The most truly generous persons are those who give silently without hope of praise or reward", Carol Brink said. In my short time knowing Jon he exemplified this. He gave generously of his time, energy and ideas to all those who came to know him. Those who knew him for a long time spoke of his lifelong generosity of spirit and how he enjoyed reaching out to help lift others up. He often helped new people who had moved to Scotland and students in their careers and life.

Jon was a true renaissance man and proud Scot, who was equally happy in the countryside and cities of Scotland, his home. His sense of optimism and adventure has left a lasting impression on everyone who knew him.

His impact on academia in Scotland and beyond our borders will continue to be felt for many years to come. Jon's approach to understanding knowledge has contributed to the new Scottish Enlightenment we see today underpinned by Computing.

Spending time with Jon was always a great pleasure and we will remember him fondly.

Jon admired the work of The Woodland Trust, so donations in his memory are welcome. https://www.justgiving.com/fundraising/jon-oberlander

Professor Aaron Quigley is the Chair of Human Computer Interaction in the School of Computer Science at the University of St Andrews. He is co-founder of SACHI and is its current director. His appointment is part of SICSA, the Scottish Informatics and Computer Science Alliance. Aaron's research interests include surface and multi-display computing, body worn interaction, human computer interaction, pervasive and ubiquitous computing and information visualisation. He has published over 175 internationally peer-reviewed publications including edited volumes, journal papers, book chapters, conference and workshop papers and holds 3 patents. In SACHI he currently supervises four PhD students Iain Carson, Daniel Rough, Evan Brown, Guilherme Carneiro and Hui-Shyong Yeo along with acting as second supervisor for Max Nicosia in the University of Cambridge. Aaron is the ACM SIGCHI Vice President for Conferences and a Distinguished Speaker of the ACM.

AI Policy

Larry Medsker (George Washington University; lrm@gwu.edu)
DOI: 10.1145/3203247.3203254

Abstract

AI Policy is a regular column in *AI Matters* featuring summaries and commentary based on postings that appear twice a month in the AI Matters blog (https://sigai.acm.org/aimatters/blog/).

Introduction

The SIGAI Public Policy goals are to:

- promote discussion of policies related to AI through posts in the AI Matters blog on the 1st and 15th of each month,
- help identify external groups with common interests in AI Public Policy,
- encourage SIGAI members to partner in policy initiatives with these organizations, and
- disseminate public policy ideas to the SIGAI membership through articles in the newsletter.

I welcome everyone to make blog comments to enrich our knowledge base of facts and ideas that represent SIGAI members.

AIES Conference

The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18) was on February 27, 2018, at the Hilton New Orleans Riverside. The AAAI/ACM Conference on AI, Ethics, and Society (AIES) was held at the beginning of AAAI-18. Developers and participants included members of SIGAI and USACM.

The AIES conference description follows: "As AI is becoming more pervasive in our life, its impact on society is more significant and concerns and issues are raised regarding aspects such as value alignment, data handling and bias, regulations, and workforce displacement. Only a multi-disciplinary and multi-stakeholder effort can find the best ways to address these concerns, including experts of various disciplines, such as ethics, philosophy, economics, sociology, psychology, law, history, and politics. In order to address these issues in a scientific context, AAAI and ACM have joined forces to start this new conference."

The full schedule for the AIES 2018 Conference is available at www.aies-conference.com. A panel relevant to our policy blog discussions "Prioritizing Ethical Considerations in Intelligent and Autonomous Systems - Who Sets the Standards" was designed by our IEEE/ACM committee and will be covered in a future post.

Educational Policy for AI and an Uncertain Labor Market

In the next few blog posts, we will present information and generate discussion on policy issues at the intersection of AI, the future of the workforce, and educational systems. Because AI technology and applications are developing at such a rapid pace, current policies will likely not be able to impact sufficiently the workforce needs even in 2024 - the time frame for middle school students to prepare for low skill jobs and for beginning college students to prepare for higher skilled work. Transparency in educational policies requires goal setting based on better data and insights into emerging technologies, likely changes in the labor market, and corresponding challenges to our educational systems. The topics and resources below will be the focus of future AI Policy posts.

Technology

IBM's Jim Spohrer has an outstanding set of slides[1] "A Look Toward the Future", incorporating his rich experience and current

[1] https://www.slideshare.net/spohrer/future-20171110-v14

work on anticipated impacts of new technology with milestones every ten years through 2045. Radical developments in technology would challenge public policy in ways that are difficult to imagine, but current policymakers and the AI community need to try. Currently, AI systems are superior to human capabilities in calculating and game playing, and near human level performance for data-driven speech and image recognition and for driverless vehicles. By 2024, large advances are likely in video understanding, episodic memory, and reasoning.

The roles of future workers will involve increasing collaboration with AI systems in the government and public sector, particularly with autonomous systems but also in traditional areas of healthcare and education. Advances in human-technology collaboration also lead to issues relevant to public policy, including privacy and algorithmic transparency. The increasing mix of AI with humans in ubiquitous public and private systems puts a new emphasis on education for understanding and anticipating challenges in communication and collaboration.

Workforce

Patterns for the future workforce in the age of autonomous systems and cognitive assistance are emerging. Please take a look at Andrew McAfee's presentation at the recent Computing Research Summit. Also, see the latest McKinsey Report[2] "Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation." Among other things, this quote from page 20 catches attention: "Automation represents both hope and challenge. The global economy needs the boost to productivity and growth that it will bring, especially at a time when aging populations are acting as a drag on GDP growth. Machines can take on work that is routine, dangerous, or dirty, and may allow us all to use our intrinsically human talents more fully. But to capture these benefits, societies will need to prepare for complex workforce transitions ahead. For policy makers, business leaders, and individual workers the world over, the task at hand is to prepare for a more automated future by emphasizing new skills, scaling up training, especially for midcareer workers, and ensuring robust economic growth."

Education for the Future

An article in Education Week "The Future of Work Is Uncertain, Schools Should Worry Now"[3] addresses the issue of automation and artificial intelligence disrupting the labor market and what K-12 educators and policymakers need to know. A recent report by the Bureau of Labor Statistics "STEM Occupations: Past, Present, And Future"[4] is consistent with the idea that even in STEM professions workforce needs will be less at programming levels and more in ways to collaborate with cognitive assistance systems and in human-computer teams. Demands for STEM professionals will be for verifying, interpreting, and acting on machine outputs; designing and assembling larger systems with robotic and cognitive components; and dealing with ethics issues such as bias in systems and algorithmic transparency.

Upcoming

Some themes planned for the SIGAI Public Policy posts for 2018 include algorithmic accountability and the impacts of AI and Data Science on the future of education and the labor market. We will look at potential policies for today that could mitigate impacts of AI on individuals and society. Policy areas include innovative educational systems, ideas for alternate economic systems, and regulatory changes to promote safe and fair technological innovation. We welcome your input and discussion at the AI Matters blog!

[2] https://goo.gl/rviGDC

[3] https://www.edweek.org/ew/articles/2017/09/27/the-future-of-work-is-uncertain-schools.html

[4] https://www.bls.gov/spotlight/2017/science-technology-engineering-and-mathematics-stem-occupations-pas
pdf/science-technology-engineering-and-mathematics-stem-occupations-
pdf

Larry Medsker is a Research Professor of Physics and Director of the Data Science graduate program at The George Washington University. Dr. Medsker is a former Dean of the Siena College School of Science, and a Professor in Computer Science and in Physics, where he was a co-founder of the Siena Institute for Artificial Intelligence. His research and teaching continues at GW on the nature of humans and machines and the impacts of AI on society and policy[a]. Professor Medsker's research in AI includes work on artificial neural networks and hybrid intelligent systems. He is the Public Policy Officer for the ACM SIGAI.

[a] http://www.humai.org/humai/ and http://humac-web.org/

Creating the Human Standard for Ethical Autonomous and Intelligent Systems (A/IS)

John C. Havens (IEEE; john.havens.us@ieee.org)
DOI: 10.1145/3203247.3203255

Introduction

Like most people, I first encountered Artificial Intelligence through movies - The Terminator, Blade Runner, 2001. As a rule, the future in these stories was always dystopian which I found irritating. If humanity was able to create such amazing technology, wouldn't they also have created ethical codes or standards to keep things from going awry? Granted, a film about a code of ethics isn't as sexy as killer robots, but picturing utopian futures powered or assisted by AI seemed like something not enough people were doing in my estimation.

I've been writing about technology since around 2011 for publications like The Guardian, Slate, and HuffPo. But it was an ongoing series on Artificial Intelligence I wrote for Mashable that led me to my current work with IEEE. In 2014 I wrote an article called "Coming to Terms With Humanity?s Inevitable Union With Machines[1]" as a way to genuinely confront fears I was facing about the nature not of killer robots but algorithms that might make choices for me to the point where I'd lose myself. Not being an engineer or programmer by training, I realize now this was an uninformed perspective, but it's one I believe the general public often shares when not fully understanding how AI functions under the hood. The article led to my writing my book, "Heartificial Intelligence: Embracing Our Humanity to Maximize Machines[2]", which I spoke about at the SXSW conference in Austin as a guest of IEEE.

When I spoke I had already done a great deal of research to identify any existing Codes of Ethics for AI (this was in 2014). Everyone I interviewed kept quoting Asimov's Laws of Robotics which as a newbie to AI I found to be quite alarming. Didn't people realize these came from his short story, "Runaround" from 1942[3]? While I appreciated the nature of the story was to demonstrate the conundrum of trying to have one simple set of laws apply to any robot / system (eg, "do no harm" doesn't make sense if you're creating a surgical robot) I also didn't understand why nobody had created a more formal and updated set of Principles.

Fortunately during my talk at SXSW there were two people from IEEE staff leadership in the audience who agreed with my assessment that there was a need to identify a set of global principles for AI. They recommended I present my ideas to IEEE on how they could create these Principles which I did a few months after SXSW.

The Council and The Chairs

When I presented my ideas in 2015 for members of IEEE's Management Council, it was the first time I met Konstantinos Karachalios. He's the Managing Director for the IEEE Standards Association and also sits on IEEE's Management Council. Konstantinos is the person who helped shape my initial ideas into what has now become The IEEE Global Initiative that also inspired the P7000 Standards Working Group series.

The Chair is Raja Chatila. After the initial core structure for The Initiative was in place, Konstantinos approached Raja who was at that time completing his tenure as President of the IEEEE Robotics and Automation Society to talk to him about The Initiative. Thankfully Raja was interested and began further shaping and developing the structure and makeup of The Initiative.

Our Vice-Chair is Kay Firth-Butterfield. I first

[1] https://mashable.com/2014/04/11/digital-humanity/

[2] https://www.amazon.com/Heartificial-Intelligence-Embracing-Humanity-Maximize/dp/0399171711

[3] https://en.wikipedia.org/wiki/Three_Laws_of_Robotics

met Kay while she was at Lucid AI[4], serving as Chief Officer of their Ethics Advisory Panel. Beyond being a barrister by trade and a gifted speaker, Kay is one of the most connected and respected people in the AI Ethics world. She's now serving as the Head of Artificial Intelligence and Machine Learning at the World Economic Forum.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (A/IS) was launched in April of 2016 to move beyond the paranoia and the uncritical admiration regarding autonomous and intelligent technologies and to illustrate that aligning technology development and use with ethical values will help advance innovation while diminishing fear in the process.

The goal of The IEEE Global Initiative is to incorporate ethical aspects of human well-being that may not automatically be considered in the current design and manufacture of A/IS technologies and to reframe the notion of success so human progress can include the intentional prioritization of individual, community, and societal ethical values.

The IEEE Global Initiative has two primary outputs. First, the creation and iteration of a body of work known as "Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems[5]". Second, the identification and recommendation of ideas for Standards Projects focused on prioritizing ethical considerations in A/IS.

Version 1 of Ethically Aligned Design (EAD) was released in December of 2016 as a Creative Commons document so any organization could utilize it as an immediate and pragmatic resource. Launched as a Request for Input (RFI) to solicit response from the public in a globally consensus-building manner, the document received over two hundred pages of feedback at the time of the RFI's deadline.

[4] https://mashable.com/2015/10/03/ethics-artificial-intelligence/

[5] http://standards.ieee.org/news/2016/ethically_aligned_design.html

Ethically Aligned Design, Version 2 features five new sections in addition to updated iterations of the original eight sections of EADv1. The IEEE Global Initiative has now increased from 100 AI/Ethics experts to more than 250 individuals including new members from China, Japan, South Korea, India, and Brazil and EADv2 now contains over 120 key Issues and Candidate Recommendations. Version 2 was also launched as a Request for Input. (You can download Ethically Aligned Design, Version 2 at this link: http://standards.ieee.org/develop/indconn/ec/auto_sys_form.html)

The Mission of The IEEE Global Initiative is to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity. By identifying A/IS oriented Principles and creating Standards directly relating to the challenges brought about by the widespread use of A/IS, The Initiative hopes to complement and evolve how engineers create technology in the algorithmic age.

The IEEE P7000™ series of Approved Standardization Projects

Along with creating and evolving Ethically Aligned Design, members of The IEEE Global Initiative are encouraged to recommend Standards Projects to IEEE based on their work. Below are titles and descriptions for each of these approved IEEE Standards Projects, and more information is available via the links included:

The IEEE P7000™ series of standards projects under development represent a unique addition to the collection of over 1300 global IEEE standards and projects. Whereas more traditional standards have a focus on technology interoperability, safety and trade facilitation, the P7000 series address specific issues at the intersection of technological and ethical considerations. Like their technical standards counterparts, the P7000 series empower innovation across borders and enable societal benefit.

There are currently thirteen approved Standards in the Series, incorporating key issues within the Autonomous/Intelligent and ethical realm including transparency, data access and control, algorithmic bias, robotic nudging, well-being, and more:

- IEEE P7000™ - Model Process for Addressing Ethical Concerns During System Design
- IEEE P7001™ - Transparency of Autonomous Systems
- IEEE P7002™ - Data Privacy Process
- IEEE P7003™ - Algorithmic Bias Considerations
- IEEE P7004™ - Standard on Child and Student Data Governance
- IEEE P7005™ - Standard on Employer Data Governance
- IEEE P7006™ - Standard on Personal Data AI Agent Working Group
- IEEE P7007™ - Ontological Standard for Ethically driven Robotics and Automation Systems
- IEEE P7008™ - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
- IEEE P7009™ - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
- IEEE P7010™ - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems
- IEEE P7011™ - Standard for the Process of Identifying and Rating the Trustworthiness of News Sources
- IEEE P7012™ - Standard for Machine Readable Personal Privacy Terms

For further information, please see: https://ethicsinaction.ieee.org

The Future

We are deeply fortunate to have a fantastic Executive Committee made up of representatives from UNESCO, The Partnership on AI, private sector industry and many more. Sven Koenig of SIGAI (ACM's Special Interest Group on Artificial Intelligence) is also member of our Executive Committee. Along with benefitting from Sven's deep expertise in AI, it has been fantastic to see ACM's efforts in the AI space, including their groundbreaking "Statement on Algorithmic Transparency and Accountability[6]".

It is with these thought leaders that we?ve only recently completed our plans for how ww'll complete the final version of Ethically Aligned Design. When looking at the document, you'll note that each of the thirteen committees has listed "Issues" and "Candidate Recommendations." Initially we were using the term, "concerns" instead of "issues" but Francesca Rossi (who's on our Executive Committee) made the excellent point that we didn't want an entire paper comprised of only "concerns." (We made that change before publishing version 1 of Ethically Aligned Design).

The idea of "Candidate" Recommendations (if memory serves) came from Richard Mallah of FLI. Rather than have EADv1 make it seem like we had finalized our thoughts on any particular subject, this process let us release EADv1 and EADv2 in a public Request For Input process. We received over two hundred pages of feedback for EADv1 (which you can see here: http://standards.ieee.org/develop/indconn/ec/rfi_responses_document.pdf) and are currently getting feedback for version 2. This is a unique process for IEEE which mirrors their consensus-building processes found in their Standards creation and other processes. For us, we wanted to make sure to not infer that a group of largely Western A/IS experts could define ethics in one fell swoop.

A lot of the feedback we received for Version 1 was people outside of the US and the EU letting us know it was important to include non-Western ethical ideas in future versions. We agreed, and we ended up inviting the people providing feebback along with a number of other global thought leaders from China, Japan, South Korea, Brazil, India, Mexico, Thailand and Africa to our work. A number of people from those countries even translated the introduction of EADv1 into their own languages (which you can see here: http://standards.ieee.org/develop/indconn/ec/ead_v1.html).

[6] https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

For our final push as we move forward, we'll be working to add more members from regions and societal constituencies we may have missed or from which we need more representatives. We are excited to have recently added a High School Committee made up of about twenty students from the great organization AI4ALL, plus we?ll be working with the IEEE Young Professionals to increase diversity of age and orientation as well as regionalization.

All Members from this point on, along with updating content for Ethically Aligned Design, will be asked to vote at various points to finalize our "Candidate Recommendations." We'll also be refining our list of General Principles to use as the criteria to help Committees decide what "Issues" align with those Principles so our final document will be a cohesive whole united by our overall philosophy of "Advancing Technology for Humanity" (that's IEEE's tagline) and Prioritizing Human Wellbeing with Autonomous and Intelligent Systems (the subtitle of Ethically Aligned Design).

Our goal is to publish the final version of EAD around Q2 of 2019. We'll also be releasing a number of white papers focusing on Committee content over the next few months, along with videos from members and a few big surprises planned for when we launch the final version.

So, stay tuned, and consider joining our ranks as a Member of The Initiative or in one of the IEEE P7000 Working Groups. We would greatly welcome any ACM Members to help us shape the future of A/IS ethics principles and standards.

John C. Havens is the Executive Director of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html To get involved or learn more, please email John.

AI Amusements: Computer Elected Governor of California
Corpus Legis trounces human opponents in state election

Michael Genesereth (Stanford University; genesereth@stanford.edu)
DOI: 10.1145/3203247.3203256

California today became the first state in the Union to elect a computer as governor. The independent candidate, **Corpus Legis**, handily defeated its human opponents in a special election following the retirement of former governor **Jerry (Moonbeam) Brown**.

The election was seen by some as a referendum on Brown. Many voters had criticized Brown for his ill-considered promotion of pet projects of dubious merit, such as High Speed Rail. Others were disturbed by his tendency to ignore laws he did not like, an increasingly common trend in Californian politics.

Exit polls suggest that Legis voters were swayed by the comparative potential for rationality and fairness offered by machines. For some, there was a strong belief in the need for a new approach to governance in a nation divided by the pettiness and partisan bickering of human politicians. Also, as one voter pointed out, the new governor, being a machine, will be able to work for the state 24/7.

Legis comes to the office with significant bona fides. It was the first chairman of the computer science department at Stanford. It was the first machine to sit on the Palo Alto City Council, before becoming the city's first electronic mayor. And it was the first machine to pass the state bar exam.

That said, the road to the governor's mansion was not an easy one. Legis, a rule-based system, first had to survive a brutal primary election in which it was pitted against another computer candidate **Alpha (Google) Watson**, a machine learning system. Early in the season, Watson enjoyed a comfortable lead in the polls based on its many scientific and commercial successes and its flashy marketing. During the election, Watson pummeled Legis with numerous hypotheses derived from its analysis of big data.

However, the tide turned against Watson when it could not explain its positions beyond citing statistics about how things had been done in the past. It stumbled further when, due to an absence of relevant data, it was unable to say how it would implement a new law. Finally, it lost credibility when, on the basis of its statistical analysis, it theorized that Legis was actually Antonin Scalia, ignoring the fact that Scalia had died several years before. Evidently, Watson was unaware that Scalia's death invalidated its theory, most likely due to its lack of background knowledge.

The results of the election are not without controversy. Multiple court challenges have already been filed by citizens alarmed at the prospects of an AI governor. **Steven Hawking** suggests that it is another example of AI beginning to dominate the world. **Bill Gates** argues that, since the system was developed on an Apple Computer, it is just a toy. At the same time, various luminaries have expressed their support. The **Academy of Motion Picture Arts and Sciences** lauds the election as an example of increasing diversity. And, in an unexpected turn of events, **Elon Musk** says it is a welcome turn of events, suggesting that we need to be a ”multi-technology species” to deal with the possibility of biological extinction.

Corpus Legis will be sworn in next month. Provided that the election survives the court challenges. And provided that the government can figure out to how to swear in a machine.

Michael Genesereth is a professor in the Computer Science Department at Stanford University. He is most know for his academic work on Computational Logic. However, he also writes the occasional news article to keep the general public informed about significant developments related to that work.