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: 10.1145/3320254

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 next 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 5(1)

Amy McGovern, co-editor (University of Oklahoma; aimatters@sigai.acm.org)
Iolanda Leite, co-editor (Royal Institute of Technology (KTH); aimatters@sigai.acm.org)
DOI: 10.1145/3320254.3320255

Issue overview

Welcome to the first issue of the fifth volume of the AI Matters Newsletter! This issue opens with some news on a new SIGAI Student Travel Scholarship where we aim to encourage students from traditionally underrepresented geographic locations to apply and attend SIGAI supported events. We also summarize the fourth AAAI/ACM SIGAI Job Fair, which continues to grow with the increasing popularity of AI. In our interview series, Marion Neumann interviews Tom Dietterich, an Emeritus Professor at Oregon State University and one of the pioneers in Machine Learning.

In our regular columns, we have a summary of recent and upcoming AI conferences and events from Michael Rovatsos. Our educational column this issue is dedicated to "biductive computing", one of Prolog's most distinctive features. Larry Medsker's policy column summarizes several policy aspects relevant to the SIGAI community worldwide, including a recent executive order for "Maintaining American Leadership In Artificial Intelligence" and a summary of policy views on AI by Wolfgang Wahlster, CEO and Scientific Director of the German Research Center for AI.

We have two paper contributions for this issue. The first paper is a continuation of the discussion of an AI "cosmology" by our contributing editors Cameron Hughes and Tracey Hughes. The second paper proposes Experiential AI as a new research agenda by a multidisciplinary set of researchers including both artists and scientists. The writers come from University of Edinburgh and Herriot Watt University.

We close by bringing back our popular entertainment/humor column with an AI generated crossword puzzle by Adi Botea. We aim to have this feature regularly now and we will publish the solution in the following issue.

Finally, we would like to congratulate Yoshua Bengio, Geoffrey Hinton, and Yann LeCun on receiving the 2018 ACM A.M. Turing Award!

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 next issue. Details on the submission process are available at http://sigai.acm.org/aimatters.

Amy McGovern is co-editor of AI Matters. She is a Professor of computer science at the University of Oklahoma and an adjunct 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.

Iolanda Leite is co-editor of AI Matters. She is an Assistant Professor at the School of Electrical Engineering and Computer Science at the KTH Royal Institute of Technology in Sweden. Her research interests are in the areas of Human-Robot Interaction and Artificial Intelligence. She aims to develop autonomous socially intelligent robots that can assist people over long periods of time.

Open Student Travel Scholarships Scheme

Michael Rovatsos (University of Edinburgh; mrovatso@inf.ed.ac.uk)
DOI: 10.1145/3320254.3320256

As part of its portfolio of student-oriented activities, SIGAI regularly supports its members through the provision of travel scholarships. To date, these have been primarily allocated to key conferences (co-)sponsored by SIGAI, but as the number of events we support formally (whether through financial sponsorship or "in-cooperation" status) has increased substantially, we have identified a need to reach out to a broader community of students participating in, and contributing to, a broader range of events.

In response to this, we have recently created a separate fund for student travel scholarships open to all SIGAI student members who are attending SIGAI-supported events (in exceptional cases, we may also provide support for participation in events not directly associated with SIGAI). We specifically want to encourage students to apply who come from geographical regions where generally fewer resources for student travel are available, and will generally prioritise students from disadvantaged or minority backgrounds in allocating funds.

This open scheme will run somewhat differently from the traditional student scholarship scheme, which was primarily managed by conference organizers, and where SIGAI only managed reimbursement of expenses and published students' post-conference reports on its web site. Under the new scheme, applicants are asked to complete an online form directly to apply for a scholarship and to contact the SIGAI Conference Coordination Officer by email.

Applications for travel support are accepted on a rolling basis without any set deadline, and should be submitted by completing a form linked from three months before the start of the event. Applicants must supply personal details, details of the event, the rationale for attendance, and estimated breakdown of costs with a statement about which part of these costs can be covered from other sources.

As before, students must be enrolled for full-time study with a recognized academic institution, and must have been a member of ACM SIGAI at the time of application. It is acceptable to join ACM SIGAI just before making the application, but priority will be given to applicants who have been a member of ACM SIGAI for a longer period of time. Students receiving partial financial support for attendance at the event may request additional funding from ACM SIGAI.

We hope this new scheme will enable us to allocate more funds to support the representation of a diverse range of student attendees, and to better respond to the needs of our student members. As this is a new scheme, we would be particularly interested in receiving feedback on the scheme and suggestions for improvement. Details and the online form can be found at https://sigai.acm.org/activities/student_support.html.

Michael Rovatsos is the Conference Coordination Officer for ACM SIGAI, and a faculty member at the University of Edinburgh. His research in in multiagent systems and human-friendly AI. Contact him at mrovatso@inf.ed.ac.uk.

AAAI/ACM SIGAI Job Fair 2019: A Retrospective

Christopher Amato (Northeastern University; camato@ccs.neu.edu)
John P Dickerson (University of Maryland; john@cs.umd.edu)
DOI: 10.1145/3320254.3320257

Introduction

For the fourth year running, AAAI and ACM SIGAI jointly ran the popular AAAI/ACM SIGAI Job Fair. With the AAAI main conference increasing in the number of submissions, number of accepted papers, and attendance each year, one would expect the job fair to grow in turn; indeed, this year, we saw a growth on both fronts. This year, twenty-six companies formally attended—typically with a booth, team of recruiters, swag, and other representatives—increasing from twenty-one companies during the job fair's previous run in 2018. In turn, many hundreds of students, post-docs, and other job seekers either uploaded their resumes or CVs before the event, or uploaded their CVs after attending the event in person. Following this, as in previous years, those resumes and CVs were shared with participating companies. Those companies are listed below.

Participating Companies

- Air Liquide R&D
- Amazon
- ASAPP
- Baidu
- CAIR, University of Agder
- Conduent Lab US
- DiDi
- Diffbot
- Diveplane
- Elsevier
- Google AI
- IBM
- Jane Street
- JD.com
- Lionbridge
- Los Alamos National Lab
- Lyft
- Microsoft
- NextAI
- Palo Alto Research Center
- Raytheon BBN
- SIFT
- Thales CortAIx R&D
- Unity
- USC/Information Sciences Institute (ISI)
- WeBank

Building on momentum and feedback from last year's AAAI/ACM SIGAI Job Fair, this year, a representative from each participating company was given the opportunity to give a 60-second pitch—which, in reality, turned into something more like 120 seconds—accompanied by a single slide. Last year, we found that this served as a good way to coalesce participants from both sides of the market at the beginning of the job fair. We observed that behavior again this year. Many companies were interested in machine-learning-based approaches to solving societal issues, as well as resource allocation and logistics problems, self-driving cars and other (semi-)autonomous-agent-based industries, and others. Participants hailed from all over the world (e.g., China, Norway, Singapore, US) and from industry, academia, and the government.

Figure 1: The job fair kicked off with a brief intro from organizers, followed by 1–2 minute pitches by each of the participating firms.

This year, we also purchased a dedicated domain—https://aaaijobfair.com/—that will be passed down from organizer to organizer, and will also allow present and future firms and participants to view previous iterations of the job fair. We hope this will encourage the building of a brand for the job fair itself. We also, of course, hope that all

Figure 2: A representative from each of the participating firms gave a 1–2 minute, single-slide pitch.

participants in this year's fair enjoyed their time and found the experience worthwhile! If you have any comments regarding the fair itself, or suggested improvements, please get in touch!

Christopher Amato is an Assistant Professor in the Khoury College of Computer Sciences at Northeastern University. He works in artificial intelligence and robotics, with specific focus on decision making under uncertainty in multi-agent and multi-robot systems.

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.

AI Profiles: An Interview with Thomas Dietterich

Marion Neumann (Washington University in St. Louis; m.neumann@wustl.edu)
DOI: 10.1145/3320254.3320258

Introduction

Welcome to the eighth interview in our series profiling senior AI researchers. This month we are especially happy to interview our SIGAI advisory board member, Thomas Dietterich, Director of Intelligent Systems at the Institute for Collaborative Robotics and Intelligence Systems (CoRIS) at Oregon State University.

Figure 1: Tom Dietterich

Biography

Dr. Dietterich (AB Oberlin College 1977; MS University of Illinois 1979; PhD Stanford University 1984) is Professor Emeritus in the School of Electrical Engineering and Computer Science at Oregon State University, where he joined the faculty in 1985. Dietterich is one of the pioneers of the field of Machine Learning and has authored more than 200 refereed publications and two books. His research is motivated by challenging real world problems with a special focus on ecological science, ecosystem management, and sustainable development. He is best known for his work on ensemble methods in machine learning including the development of error-correcting output coding. Dietterich has also invented important reinforcement learning algorithms including the MAXQ method for hierarchical reinforcement learning. Dietterich has devoted many years of service to the research community. He served as President of the Association for the Advancement of Artificial Intelligence (2014-2016) and as the founding president of the International Machine Learning Society (2001-2008). Other major roles include Executive Editor of the journal *Machine Learning*, co-founder of the *Journal for Machine Learning Research*, and Program Chair of AAAI 1990 and NIPS 2000. Dietterich is a Fellow of the ACM, AAAI, and AAAS.

Getting to Know Tom Dietterich

When and how did you become interested in CS and AI?

I learned to program in Basic in my early teens; I had an uncle who worked for GE on their time-sharing system. I learned Fortran in high school. I tried to build my own adding machine out of TTL chips around that time too. However, despite this interest, I didn't really know what CS was until I reached graduate school at the University of Illinois. I first engaged with AI when I took a graduate assistant position with Ryszard Michalski on what became machine learning, and I took an AI class from Dave Waltz. I had also studied philosophy of science in college, so I had already thought a bit about how we acquire knowledge from data and experiment.

What would you have chosen as your career if you hadn't gone into CS?

I had considered going into foreign service, and I have always been interested in policy issues. I might also have gone into technical management. Both of my brothers have been successful technical managers.

What do you wish you had known as a Ph.D. student or early researcher?

I wish I had understood the importance of strong math skills for CS research. I was a software engineer before I was a computer

science researcher, and it took me a while to understand the difference. I still struggle with the difference between making an incremental advance within an existing paradigm versus asking fundamental questions that lead to new research paradigms.

What professional achievement are you most proud of?

Developing the MAXQ formalism for hierarchical reinforcement learning.

What is the most interesting project you are currently involved with?

I'm fascinated by the question of how machine learning predictors can have models of their own competence. This is important for making safe and robust AI systems. Today, we have ML methods that give accurate predictions in aggregate, but we struggle to provide point-wise quantification of uncertainty. Related to these questions are algorithms for anomaly detection and open category detection. In general, we need AI systems that can work well even in the presence of "unknown unknowns".

Recent advances in AI led to many success stories of AI technology undertaking real-world problems. What are the challenges of deploying AI systems?

AI systems are software systems, so the main challenges are the same as with any software system. First, are we building the right system? Do we correctly understand the users' needs? Have we correctly expressed user preferences in our reward functions, constraints, and loss functions? Have we done so in a way that respects ethical standards? Second, have we built the system we intended to build? How can we test software components created using machine learning? If the system is adapting online, how can we achieve continuous testing and quality assurance? Third, when ML is employed, the resulting software components (classifiers and similar predictive models) will fail if the input data distribution changes. So we must monitor the data distribution and model the process by which the data are being generated. This is sometimes known as the problem of "model management". Fourth, how is the deployed system affecting the surrounding social and technical system? Are there unintended side-effects? Is user or institutional behavior changing as a result of the deployment?

One promising approach is combining humans and AI into a collaborative team. How can we design such a system to successfully tackle challenging high-risk applications? Who should be in charge, the human or the AI?

I have addressed this in a recent short paper (Robust Artificial Intelligence and Robust Human Organizations. Frontiers of Computer Science, 13(1): 1-3). To work well in high-risk applications, human teams must function as so-called "High reliability organizations" or HROs. When we add AI technology to such teams, we must ensure that it contributes to their high reliability rather than disrupting and degrading it. According to organizational researchers, HROs share five main practices: (a) continuous attention to anomalous and near-miss events, (b) seeking diverse explanations for such events, (c) maintaining continuous situational awareness, (d) practicing improvisational problem solving, and (e) delegating decision making authority to the team member who has the most expertise about the specific decision regardless of rank. AI systems in HROs must implement these five practices as well. They must be constantly watching for anomalies and near misses. They must seek multiple explanations for such events (e.g., via ensemble methods). They must maintain situational awareness. They must support joint human-machine improvisational problem solving, such as mixed-initiative planning. And they must build models of the expertise of each team member (including themselves) to know which team member should make the final decision in any situation.

You ask "Who is in charge?" I'm not sure that is the right question. Our goal is to create human-machine teams that are highly reliable as a team. In an important sense, this means every member of the team has responsibility for robust team performance. However, from an ethical standpoint, I think the human team leader should have ultimate responsibility. That task of taking action in a specific situ-

ation could be delegated to the AI system, but the team leader has the moral responsibility for that action.

Moving towards transforming AI systems into high-reliable organizations, how can diversity help to achieve this goal?

Diversity is important for generating multiple hypotheses to explain anomalies and near misses. Experience in hospital operating rooms is that often it is the nurses who first detect a problem or have the right solution. The same has been noted in nuclear power plant operations. Conversely, teams often fail when the engage in "group think" and fixate on an incorrect explanation for a problem.

How do you balance being involved in so many different aspects of the AI community?

I try to stay very organized and manage my time carefully. I use a machine learning system called TAPE (Tagging Assistant for Productive Email) developed by my collaborator and student Michael Slater to automatically tag and organize my email. I also take copious notes in OneNote. Oh, and I work long hours...

What was your most difficult professional decision and why?

The most difficult decision is to tell a PhD student that they are not going to succeed in completing their degree. All teachers and mentors are optimistic people. When we meet a new student, we hope they will be very successful. But when it is clear that a student isn't going to succeed, that is a deep disappointment for the student (of course) but also for the professor.

What is your favorite AI-related movie or book and why?

I really don't know much of the science fiction literature (in books or films). My favorite is 2001: A Space Odyssey because I think it depicts most accurately how AI could lead to bad outcomes. Unlike in many other stories, HAL doesn't "go rogue". Instead, HAL creatively achieves the objective programmed by its creators. Unfortunately, as a side effect, it kills the crew.

Help us determine who should be in the AI Matters spotlight!

If you have suggestions for who we should profile next, please feel free to contact us via email at aimatters@sigai.acm.org.

Events

Michael Rovatsos (University of Edinburgh; mrovatso@inf.ed.ac.uk)
DOI: 10.1145/3320254.3320259

This section features information about upcoming events relevant to the readers of AI Matters, including those supported by SIGAI. We would love to hear from you if you are are organizing an event and would be interested in cooperating with SIGAI, or if you have announcements relevant to SIGAI. For more information about conference support visit sigai.acm.org/activities/requesting_sponsorship.html.

11th International Conference on Agents and Artificial Intelligence (ICAART 2019)

Prague, Czech Republic, February 19-21, 2019, www.icaart.org

The purpose of the International Conference on Agents and Artificial Intelligence is to bring together researchers, engineers and practitioners interested in the theory and applications in the areas of Agents and Artificial Intelligence. Two simultaneous related tracks will be held, covering both applications and current research work. One track focuses on Agents, Multi-Agent Systems and Software Platforms, Distributed Problem Solving and Distributed AI in general. The other track focuses mainly on Artificial Intelligence, Knowledge Representation, Planning, Learning, Scheduling, Perception Reactive AI Systems, and Evolutionary Computing and other topics related to Intelligent Systems and Computational Intelligence.

International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019)

Montreal, Canada, May 13-17, 2019
aamas2019.encs.concordia.ca

AAMAS is the leading scientific conference for research in autonomous agents and multiagent systems. The AAMAS conference series was initiated in 2002 by merging three highly respected meetings: the International Conference on Multi-Agent Systems (ICMAS); the International Workshop on Agent Theories, Architectures, and Languages (ATAL); and the International Conference on Autonomous Agents (AA). The aim of the joint conference is to provide a single, high-profile, internationally respected archival forum for scientific research in the theory and practice of autonomous agents and multiagent systems.

19th ACM International Conference on Intelligent Virtual Agents (IVA 2019)

Paris, France, July 2-5, 2019
humanrobotinteraction.org/2019

ACM IVA 2019 is the 19th meeting of an interdisciplinary annual conference and the main leading scientific forum for presenting research on modeling, developing and evaluating intelligent virtual agents (IVAs) with a focus on communicative abilities and social behavior. IVAs are interactive digital characters that exhibit human-like qualities and can communicate with humans and each other using natural human modalities like facial expressions, speech and gesture. They are capable of real-time perception, cognition, emotion and action that allow them to participate in dynamic social environments. In addition to presentations on theoretical issues, the conference encourages the showcasing of working applications.
Submission deadline: March 8, 2019

13th ACM Conference on Recommender Systems (RecSys 2019)

Copenhagen, Denmark, September 16-20, 2019, recsys.acm.org/recsys19

The ACM Recommender Systems Conference (RecSys) is the premier international forum for the presentation of new research results, systems and techniques in the broad field of recommender systems. Recommendation is a particular form of information filtering, that exploits past behaviors and user similarities to generate a list of information items that is personally tailored to an end-user's preferences. As RecSys brings together the main international research groups working on

recommender systems, along with many of the world's leading e-commerce companies, it has become the most important annual conference for the presentation and discussion of recommender systems research. RecSys 2019, the thirteenth conference in this series, will be held in Copenhagen, Denmark. It will bring together researchers and practitioners from academia and industry to present their latest results and identify new trends and challenges in providing recommendation components in a range of innovative application contexts. In addition to the main technical track, RecSys 2019 program will feature keynote and invited talks, tutorials covering state-of-the-art in this domain, a workshop program, an industrial track and a doctoral symposium.
Submission deadline: April 23, 2019

34rd IEEE/ACM International Conference on Automated Software Engineering (ASE 2019)

San Diego, California, November 11-15 , 2019
2019.ase-conferences.org
The 34rd IEEE/ACM International Conference on Automated Software Engineering (ASE 2019) will be held in San Diego from November 11 to 15, 2019. The conference is the premier research forum for automated software engineering. Each year, it brings together researchers and practitioners from academia and industry to discuss foundations, techniques, and tools for automating the analysis, design, implementation, testing, and maintenance of large software systems.
Submission deadline: May 13, 2019

6th International Workshop on Sensor-based Activity Recognition and Interaction (iWOAR 2019)

Rostock, Germany, September 16-17, 2019
iwoar.org/2019
The 6th International Workshop on Sensor-based Activity Recognition and Interaction (iWOAR) brings together researchers and practitioners interested in systems that monitor user behavior and condition, or that provide users with assistance. Particularly, this workshop focuses on sensors and sensor infrastructures to detect user behaviors and to provide interactions between users and system; data and model-driven methods for human activity recognition; and methods for synthesizing assistance and interaction strategies. iWOAR 2019 welcomes research- and application-oriented works as well as industry paper submissions.
Submission deadline: June 15, 2019

Michael Rovatsos is the Conference Coordination Officer for ACM SIGAI, and a faculty member at the University of Edinburgh. His research in in multiagent systems and human-friendly AI. Contact him at mrovatso@inf.ed.ac.uk.

Conference Reports

Michael Rovatsos (University of Edinburgh; mrovatso@inf.ed.ac.uk)
DOI: 10.1145/3320254.3320260

This section features brief reports from recent events sponsored or run in cooperation with ACM SIGAI.

13th International Conference on Foundations of Digital Games (FDG 18)

Malmö, Sweden, August 7-10, 2018
fdg2018.org

FDG seeks to promote the exchange of information concerning the foundations of digital games, technology used to develop digital games, and the study of digital games and their design, broadly construed. The guiding vision for this year's conference was *Creating Games from the Player and for the Player*. The role of the player has evolved from being viewed as a mere consumer to a participant holding a key role within the game design, creation, and development processes. FDG 18 included presentations of peer-reviewed papers (with rebuttal process), panels, workshops, and posters. The conference also hosted competitions, tech demos, and a doctoral consortium. This year's FDG conference had nominated five papers with honorable mention, and one of them was awarded the best paper award by the conference attendees during a special session. We had the pleasure to host three invited keynotes by Katherine Isbister, Aki Järvinen, and James Newman, as well as an industry guest lecture by Stefan Gudmundsson from King. In addition to this, we organized the opening reception as an open event in cooperation with Game Habitat, with the aims of bridging the gap between Games Industry and Academia.

33rd IEEE/ACM International Conference on Automated Software Engineering (ASE 2018)

Montpellier, France, September 3-7, 2018
www.ase2018.com

ASE 2018, this year's edition of the premier research forum for automated software engineering, was held at the CORUM conference center in Montpellier, and was organized by LIRMM (Laboratory of Informatics, Robotics, Microelectronics of Montpellier), CNRS, University of Montpellier, and LGI2P (IMT-Mines Alès), with the help of members from University of Toulouse, Bordeaux, ONERA and Université De Montréal (UDM). The first two days were allocated to workshops, tutorials, and the doctoral symposium. The main conference was held from September 5 to September 7, and featured keynotes, as well as papers from the following categories: technical (64), experience (5), new idea (11), journal first (4), and demo (16). The main conference attracted about 285 people from more than 38 countries, while workshops and tutorials welcomed respectively 64 and 30 people. As a major highlight, ASE 2018 attracted three internationally renowned keynote speakers, whose talks featured a quite diverse set of topics. In addition, ASE 2018 featured an IEEE CS Harlan Mills Award keynote address by Gail Murphy on "The Need for Context in Software Engineering". The ASE 2018 programme included four co-located workshops and three tutorials on topics such as testing, software variability, and model-driven engineering, as well as a Doctoral Symposium. The conference was supported by several industrial sponsors including Huawei, Mobioos, Berger Levrault, and Toyota InfoTech.

10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018)

Seville, Spain, September 18-20, 2018
www.ic3k.org/?y=2018

IC3K 2018 was held in Seville and was sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC) and co-organized by University of Seville as a local partner. IC3K 2018 was held in cooperation with ACM SIGAI and several other national AI societies. IC3K is

composed of three co-located complementary conferences, the International Conference on Knowledge Discovery and Information Retrieval (KDIR), the International Conference on Knowledge Engineering and Ontology Development (KEOD), and the International Conference on Knowledge Management and Information Sharing (KMIS). IC3K received 167 paper submissions from 42 countries and 18% of the papers were published and presented as full papers. Four invited talks were delivered by internationally distinguished speakers, and the program also included a special session on “Managing Digital Data, Information and Records” and two tutorials, “Ontologies for Intelligent Vision Systems” and “Making Enterprise Ontology a Potent Instrument”. Additionally, a “Best Paper Award” and a “Best Student Paper Award” were conferred at the conference.

18th ACM International Conference on Intelligent Virtual Agents (IVA’18)

Sydney, Australia, November 5-8, 2018
iva2018.westernsydney.edu.au

IVA 2018 was the 18th meeting of an interdisciplinary annual conference and the main leading scientific forum for presenting research on modeling, developing and evaluating intelligent virtual agents (IVAs) with a focus on communicative abilities and social behavior. For the first time in its history, the conference was hosted in Australia. The venue for the conference was Western Sydney University’s Parramatta City Campus. The conference was attended by over 100 people from around the world and featured a programme of exciting keynote talks.

Michael Rovatsos is the Conference Coordination Officer for ACM SIGAI, and a faculty member of the School of Informatics at the University of Edinburgh, UK. His research in in multiagent systems and human-friendly AI. Contact him at mrovatso@inf.ed.ac.uk.

AI Education Matters: Biductive Computing with Prolog

Joshua Eckroth (Stetson University; jeckroth@stetson.edu)
DOI: 10.1145/3320254.3320261

Introduction

Prolog is a great language to include in an undergraduate AI course. Its logical programming paradigm challenges students to write code in a way that is usually very different from their other courses. Students struggle with its approach but, in my experience, come to appreciate its ability to handle a broad range of AI tasks elegantly.

In my recent Model AI Assignment,[1] I characterized one of Prolog's most distinctive and powerful features, which I've termed "biductive computing." This article describes biductive computing and discusses some projects that make use of it. We also cover student feedback and some resources for teachers who wish to incorporate Prolog into their courses.

Biductive Computing

Biductive computing refers to a combination of "deductive" and "abductive" inference modes. Prolog code is made of facts, rules, and queries. Prolog executes programs by performing inferences to determine if a query is true given the facts and rules. Below, we characterize the different kinds of inferences. The query can contain a variable, which causes Prolog to search for values that make the query true. Note, variables always start with a capital letter.

Consider a GPA calculator. Given a list of a student's courses and grades, Prolog can compute the student's GPA by following rules about credits and averaging. Suppose "grades" is shorthand for a list of courses and grades, and GPA is a variable that will be filled in by Prolog:

```
gpa(grades, GPA)
```

Suppose the resulting GPA is 3.5. We can call this deductive inference or deductive computing because the rules of calculation are followed in a deductive sense: the inputs are well-defined and the output is computed according to the rules.

In most situations, Prolog will also support reverse computations. By using the constraint logic programming library (Triska, 2012), Prolog will be able to determine the course grades in order to achieve a given GPA:

```
gpa(UnknownGrades, 3.5)
```

We can call this abductive computing because we are determining the inputs in order to get a desired output. This is also sometimes called "reverse deduction."

The key is that the same Prolog code supports both kinds of logical inference. One only needs to write code for the rules of the domain – Prolog takes care of the forward and backward inferences. Putting deductive and abductive modalities together, we have "biductive computing."

This means we can give partial information for either inputs or outputs. For example, we can know some of the student's grades but not their grades for the current semester, and we can know (or desire) that the GPA is within a certain range:

```
gpa(PartiallyKnownGrades, GPA),
  GPA >= 3.0
```

Prolog will fill in the unknown grades with real grades and give back the resulting GPA. If the student cannot achieve a 3.0 GPA, Prolog will tell us there is no solution.

Prolog as a State of Mind

My students and I have successfully used Prolog and biductive computing for a variety of projects. My Model AI Assignment shows some projects suitable for a class. These projects include database querying, planning, parsing, and even probabilistic reasoning. Modern Prolog systems can integrate

[1] http://modelai.gettysburg.edu/2018/biductive/

with Java and web environments, further enhancing the variety of use cases that Prolog can support. For example, my book *AI Blueprints* (Eckroth, 2018) includes a chapter in which Prolog is used with deep learning and NLP, using the Rasa library,[2] to develop an intelligent chatbot. The code for this example is freely available.[3]

The GPA example above is based on the Tarot course advising system that a student and I developed in Prolog using a biductive computing approach (Eckroth & Anderson, 2019). Tarot can find which classes a student needs to take to finish their degree and calculate their GPA. It even places these classes in a 4-year semester-by-semester plan. It's also capable of finding which major is easiest to switch to, finding whether a certain desired elective course can be taken at a certain time, and finding the optimal semester abroad. Without any changes to the code, it can also find when prerequisite courses should be offered in order for, say, a particular course to be taken by students in a certain semester. This is all accomplished in less than 150 lines of code.

One of my senior research students is developing an e-sports ranking system. Using biductive computing, he is able to code the logic of the rankings, and then ask if a certain team can achieve a certain rank, if a certain match is on a critical path to a championship, and so on.

This student's reflections on using Prolog for this project characterizes the common reaction I have seen from students who learn Prolog:

> "I've found that Prolog and more popular forms of AI, like machine learning, to be useful in almost opposite ways. Machine learning and neural nets are useful as a dynamic and unpredictable way of creating agents that can identify patterns in data and make decisions based on that with little human input, whereas Prolog is useful as a language for developing a machine with precise, encyclopedic knowledge of a complex system that I can define arbitrarily, with a capacity to handle emergent complexity from relatively simple rules and procedures. This makes it tremendously useful in the right circumstances." — Hayden Estey, Senior, Stetson University

Other students who worked with Prolog in my AI course had similar reactions. They said that Prolog was "fun but hard," probably due to the unfamiliar approach to problem-solving compared to Java, Python, etc. They also said that Prolog was more "principled" than popular libraries like Pandas and Scikit. They felt that Prolog established a set of rules and stuck to them, whereas other libraries sometimes include too many ways to accomplish simple tasks, making it difficult to learn the canonical way to write code. While Prolog is clearly not the best tool for all jobs – there's no reason to use it for building a neural network, for example – its precision at representing facts and rules and performing inferences in logical problem domains is unparalleled.

Pedagogical Resources

Have a look at the following resources for ideas about how to use Prolog in your course. From my Model AI Assignment, consider biductive computing projects in the areas of,

- Database querying to build a Pokédex (i.e., a Pokémon database)
- Course planning as a simplified version of the Tarot system
- Recursive parsing for knitting patterns
- Probabilistic reasoning for solving a crime

Additional resources include,

- Chapter 7 of *AI Blueprints* (Eckroth, 2018) for an example that integrates machine learning and Prolog to build a chatbot
- "Adventures with Prolog: Entering the Dungeon Lord's Lair", a Model AI Assignment from 2016 that helps students practice with the fundamentals of the language[4]
- "Learn Prolog Now!", a free online introductory book[5]

[2] https://github.com/RasaHQ/rasa_nlu

[3] https://github.com/PacktPublishing/AIBlueprints/tree/master/ch07-understanding-queries-and-generating-responses

[4] http://modelai.gettysburg.edu/2016/prolog/

[5] http://www.learnprolognow.org/

References

Eckroth, J. (2018). *AI Blueprints: How to build and deploy AI business projects*. Packt Publishing.

Eckroth, J., & Anderson, R. (2019). Tarot: A course advising system for the future. *The Journal of Computing Sciences in Colleges*, *34*(3), 108-116.

Triska, M. (2012). The finite domain constraint solver of SWI-Prolog. In *Flops* (Vol. 7294, p. 307-316).

Joshua Eckroth joined the Math and Computer Science Department at Stetson University in fall 2014 as an assistant professor. He earned his Ph.D. from The Ohio State University in the areas of artificial intelligence and cognitive science, focusing on abductive reasoning and metareasoning. He edits the website AITopics, one of the most important sources of information on Artificial Intelligence trends. He is also the chief architect at i2k Connect.

AI Policy Matters

Larry Medsker (George Washington University; lrm@gwu.edu)
DOI: 10.1145/3320254.3320262

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/). Selected posts are summarized in issues of *AI Matters*.

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 so we can develop a rich knowledge base of information and ideas representing the SIGAI members.

Issues for 2019

Initial topics for the SIGAI Policy Blog are the following:

Facebook continues to draw attention to the general issue of data privacy and the role of personal data in business models. See NY Times on Facebook Privacy and FaceBook Partners.

Facial recognition software is known to be flawed, having side effects of bias, unwanted surveillance, and other problems. The Safe Face Pledge, developed by the Algorithmic Justice League and Georgetown University Law Center of Privacy & Technology, is an example of emerging efforts to make organizations aware of problems with facial recognition products, for example in autonomous weapons systems and law enforcement agencies. The Safe Face Pledge asks that companies commit to safety in business practices and promote public policy for broad regulation and government oversight on facial recognition applications.

"Autonomous" Things – The R&D for "autonomous" vehicles and other devices that dominate our daily lives pose challenges for technologies as well as for ethics and policy considerations. In 2018, we discussed language that aims at safety and degrees of autonomy rather than having, possibly unattainable, goals of completely autonomous things. A better approach may be the correct balance between technology and humans in hybrid devices and systems. See, for example, the Unmanned Integrated Systems Roadmap, 2017-2042 and Ethically Aligned Design. In a recent interview, Dr. Harold Szu, a co-founder and former governor of the International Neural Network Society, discusses research ideas that better mimic human thinking and that could dramatically reduce the time to develop autonomous technology. He discusses a possible new level of brain-style computing that incorporates fuzzy membership functions into autonomous control systems. Autonomous technology incorporating human characteristics, along with safety policies and earlier arrival of brain-style technologies, could usher in the next big economic boom.

The Future of Work and Education is a topic that not only tries to predict the workforce of the future but also how society needs to prepare for it. Many experts believe that our current school systems are not up to the challenge and that industry and government programs are needed for the challenges emerging in just a few years. See, for example, writing by the Ford Foundation and the World Economic Forum, as well as discussions at the recent EAAI-19 and AAAI-19 meetings. Opportunities and innovation in education and training for the workforce of the future rely crucially on public policy about workers in the era of increasing use of AI and other automation tech-

nologies. An important issue is who will provide training that is timely (by 2030), practical, and affordable for workers who will need to transition to the anticipated new types of jobs. The stakeholders along with workers are the schools, employers, unions, and community groups. Even if more jobs are created than lost, work in the AI future is disproportionately available to the range of people in the current and near-future workforce.

We welcome your feedback and discussions as we enter the 2019 world of AI and policy!

American AI Initiative (AAII)

President Trump issued an Executive Order on February 11, 2019, entitled "Maintaining American Leadership In Artificial Intelligence". The full text is at AAII.

The American AI Initiative order of course needs analysis and discussion of implementation details. Two sections of the Executive Order give hope for opportunities to provide public input: Sec (5)(a)(1) Within 90 days of the date of this order, the OMB Director shall publish a notice in the Federal Register inviting the public to identify additional requests for access or quality improvements for Federal data and models that would improve AI R&D and testing. These actions by OMB will help to identify data sets that will facilitate non-Federal AI R&D and testing; and Sec (6)(b) (b) To help ensure public trust in the development and implementation of AI applications, OMB shall issue a draft version of the memorandum for public comment before it is finalized. Please stay tuned for ways that our ACM US Technology Policy Committee (USTPC) can help us provide our feedback on the implementation of the Executive Order.

A summary and analysis report is available from the Center for Data Innovation: Executive Order Will Help Ensure U.S. Leadership in AI. They comment that the administration "needs to do more than reprogram existing funds for AI research, skill development, and infrastructure development" and "should ask Congress for significant funding increases to

- expand these research efforts;
- implement light-touch regulation for AI;
- resist calls to implement roadblocks or speed bumps for this technology, including export restrictions;
- rapidly expand adoption of AI within government;
- implement comprehensive reforms to the nation's workforce training and adjustment policies."

On Overpromising AI

A recent article Artificial intelligence is nowhere near the real thing, says German AI chief, by Anna Kelly, gives policy-worthy ideas. "In his 20 years as head of Germany's biggest AI research lab Wolfgang Wahlster has seen the tech hype machine splutter three times. As he hands over to a new CEO, he warns colleagues: Don't over-promise. The computer scientist who has just ended a 20-year stint as CEO of the German Research Centre for Artificial Intelligence says that "[warning] greatly underestimates the distance between AI and its human counterpart. We're years away from a game changer in the field. I always warn people, one should be a bit careful with what they claim. Every day you work on AI, you see the big gap between human intelligence and AI".

For AI policy, we should remember to look out for overpromising, but we also need to be mindful of the time frame for making effective policy and fully engage now. This effort importantly informs policymakers about the real opportunities to make AI successful. A recent article in *The Conversation* by Ben Shneiderman "What alchemy and astrology can teach artificial intelligence researchers," gives insightful information and advice on how to avoid being distracted away "...from where the real progress is already happening: in systems that enhance rather than replace human capabilities." Shneiderman recommends that technology designers shift "from trying to replace or simulate human behavior in machines to building wildly successful applications that people love to use."

Follow the Data

The Ethical Machine – Big Ideas for Designing Fairer AI and Algorithms is a "project that presents ideas to encourage a discussion about designing fairer algorithms" of the

Shorenstein Center on Media, Politics, and Public Policy, Harvard Kennedy School. The November 27, 2018, publication is Follow the Data! Algorithmic Transparency Starts with Data Transparency by Julia Stoyanovich and Bill Howe. Their focus is local and municipal governments and NGOs that deliver vital human services in health, housing, and mobility.

In the article, they give a welcome emphasis on the role of data instead of the common focus on just algorithms. They write, "data is used to customize generic algorithms for specific situations," that is to say that algorithms are trained using data. The same algorithm may exhibit radically different behavior - "make different predictions; make a different number of mistakes and even different kinds of mistakes when trained on two different data sets. In other words, without access to the training data, it is impossible to know how an algorithm would actually behave."

US and European Policy

Adam Eisgrau, ACM Director of Global Policy and Public Affairs, published an update on the ACM US and Europe Policy Committees in the November 29 ACM MemberNet. Key points are:

- The ACM US Technology Policy Committee submitted comments to the National Telecommunications and Information Administration.
- Chairs of ACM's Europe Council and ACM Europe Technology Policy Committee together wrote to the chief of staff of the European Comission's influential Scientific Advice Mechanism group offering the assistance of the ACM experts.
- Europe Technology Policy Committee Chair Oliver Grau was one of two experts chosen to conduct a deep dive briefing on AI technology for a lay audience of EC staff, industry professionals and other technology organization representatives. Adam encourages us to visit the ACM Public Policy Statements page for a complete list of both the US and Europe Policy Committees' policy products produced so far this year.

Larry Medsker is Research Professor of Physics and was founding director of the Data Science graduate program at The George Washington University. He is a faculty member in the GW Human-Technology Collaboration Lab and Ph.D. program.[a] Dr. Medsker is a former Dean of the Siena College School of Science, and Professor in Computer Science and in Physics. His research in AI includes work on artificial neural networks, hybrid intelligent systems, and the impacts of AI on society and policy.[b]. He is the Public Policy Officer for the ACM SIGAI.

[a] https://wesharescience.com/htc/
[b] http://humac-web.org/

The Laws of Thought and Thinking Machines

Cameron Hughes (Northeast Ohio ACM Chair; cameronhughes@acm.org)
Tracey Hughes (Northeast Ohio ACM Secretary; tracey.hughes@neoacmchapter.org)
DOI: 10.1145/3320254.3320263

The ideal value of an AI Cosmology would be to help the general public, researchers, educators, and practitioners to devise the truth of the definition, meaning, applications, and implications of Artificial Intelligence. The pursuit of that truth even if through an arbitrary contrivance would be a noteworthy goal. The fact of the matter is whether any cosmological structure we have hinted at so far tracks the underlying reality, we cannot escape that there is an underlying reality. At some point in time, we (humans) began the endeavor of trying to replicate the human mind with machines. There was first an effort to understand the human mind, describe its inner workings, and then build machines that could essentially duplicate the thinking process. Surely this point in time must mark or at least point to the Cosmological "Big Bang" for AI. Right?

So far we have pondered whether the evolution of AI could be divided into epochs. We've considered *Machine Learning*, *Expert Systems*, and *Cybernetics* as possible epochs with each reaching back further in the AI timeline. But where did it all begin? What was the first epoch? When did we first try to duplicate the thinking and reasoning process within a machine? When did we first try to represent the inner workings of the human mind as a set of instructions? At what point did we try to replicate the human mind by non-biological means? Would this point in time constitute the beginning (Big Bang) of the evolution of what we now call Artificial Intelligence? We use the Laws of Physics to describe structures, i.e., the beginning, evolution, and fate of the Universe. We call this structure the Cosmology. Can the Laws of Thought play a similar role in our quest for the real AI Cosmology?

If there are Laws of Thought, do we understand what they are? If there are laws for the thinking process, how are they related to what we currently call Artificial Intelligence? George Boole's *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities* published in 1854 starts off with:

> *The design of the following treatise is to investigate the fundamental laws of those operations of the mind by which reasoning is performed; to give expression to them in the symbolical language of a Calculus and upon this foundation to establish the science of Logic and construct method; to make that method itself the basis of a general method for the application of mathematical doctrine of Probabilities; and, finally, to collect from the various elements of truth brought to view in the course of these inquiries some probable intimations concerning the nature of the human mind.*

George Boole was not alone in suggesting that the operation of the mind or thinking process could be represented as a set of laws or fundamental axioms. Alfred Tarski in his *On Mathematical Logic and the Deductive Method* which first appears in 1936 writes:

> *Complicated mental processes are entirely reducible to such simple activities as the attentive observation of statements previously accepted as true, the perception of structural, purely external, connections among these statements, and the execution of mechanical transformations as prescribed by the rules of inference.*

Gottfried Wilhelm Leibniz suggests that human reason can be reduced to fundamental logical calculation. In the *Art of Discovery* 1685, he writes in a letter to Philip Spener:

> *The only way to rectify our reasoning is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: Let us calculate, without further ado, to see who is right.*

Logic Machines Epoch

The idea that the operation of the mind and the thinking process could be represented as mathematical logic and discrete structures in a finite form suitable for implementation by a machine clearly predates the term "Artificial Intelligence". In the AI Matters Volume 4 Issue 4 and Volume 4 Issue 2, we've presented the notions of AI epochs. We postulated that we are currently in the Machine Learning Epoch, and that prior to that were the Expert System and Cybernetics Epochs. Now we consider the Logic Machines Epoch that was powered by mathematical logic and discrete structures. The ultimate goal here is to possibly identify when and what were the first real attempts at implementing the human thinking process or the operations of the mind by a machine. Here is a listing of some examples of Logic Machines. These Logic Machines reflect attempts at developing a theory, framework, or a mechanization of the operations of the mind.

- **Ramon Llul (1290)**
  Ars Magna: Paper Machine
- **Thomas Hobbes (1651)**
  Leviathan Book:Theory
- **Gottried Leibniz (1688-1689)**
  Calculus Ratiocinator: Machine Framework
- **Charles Stanhope (1775)**
  Demonstrator: Device
- **Charles Babbage (1822-1833)**
  Difference and Analytical Engines: Machines
- **Semyon Nikoaivich Korsakov (1832)**
  Comparing Ideas Machine: Machine
- **Bernard Bolzano (1837)**
  Wissenschaftslehre Book: Theory
- **George Boole (1847-1854)**
  Boolean Algebra and Laws of Thought: Theories
- **William Stanley Jevons (1870-1894)**
  Logic Piano: Machine
- **Friedrich Ludwig Gottlob Frege (1879)**
  Second-order Logic and Axiomatic Predicate Logic: Theories
- **Bertrand Arthur William Russell (1910)**
  Principia Mathematica: Theory
- **Leonardo Torres y Quevedo (1911)**
  Chess Playing Machine: Machine
- **Alfred Tarski (1933)**
  Theory of Truth: Theory
- **Allen Newell (1955-1956,1957)**
  Logic Theory Machine and General Problem Solver: Programs
- **Alan Turing (1948,1950)**
  Intelligent Machinery, Turing Test, "Can Machines think?": Theory, Program, Paper
- **Allan Newell, Herbert Simon (1976)**
  Physical Symbol System Hypothesis: Theory

Alan Turing is the second to last entry in this list. He wrote his famous paper "Can A Machine Think?" in 1950. This is considered by some as the beginning of the history of AI. But as you can see, the Logical Machines Epoch looks like the foundations of symbolic logic. Many AI systems have used symbolic logic. Symbolic logic is based on formal logic which represents propositions as symbolic structures. Inferencing is performed by mechanical manipulations of those structures. In this list, we see early attempts at designing comprehensive knowledge representation languages, mechanical approaches to reasoning, and devices/machines that utilized these methods. Leibniz envisioned such a device or machine. He developed a framework for the Calculus Ratiocinator or Calculus Reasoning machine. Its purpose was to perform logical deductions based on a framework of *Characteristica Universalis*, a conceptual language that was able to symbolically represents all human thoughts. These symbols would then be manipulated mathematically by the Ratiocinator that would mechanically deduce all possible truths from a list of simple thoughts. He stated:

> *Thus I assert that all truths can be demonstrated about things expressible in this language with the addition of new concepts not yet expressed in it $\tilde{N}$ all such truths, I say, can be demonstrated solo calculo, or solely by the manipulation of characters according to a certain form, without any labor of the imagination or effort of the mind, just as occurs in arithmetic and algebra.*

Charles Stanhope developed his own version of a "Ratiocinator", not as ambitious, called

the “Reasoning Machine” or Demonstrator. Charles Stanhope worked on several logic machines for 30 years in the late 18th century. The Demonstrator was a device to solve mechanically:

- traditional syllogisms,
- numerical syllogisms,
- elementary probability problems

It could process no more than two premises and probability problems with no more than two independent events. Due to these limitations and the fact that it could not solve ’real-life problems’, Stanhope called it the Demonstrator.

> *As the instrument is so constructed as to assist us in making demonstrations. I have termed it Demonstrator. It is so peculiarly contrived as likewise to exhibit symbolically those proportions or degrees of probability which it is the object of the Logic of Probability to discover.*

William Stanley Jevons, an economist and logician, was inspired by the Demonstrator and developed the Logic Piano in 1869. The Logic Piano was a series of wooden boards with combinations of true and false terms. They were arranged on a rack and a ruler used to remove certain excluded combinations. The faceplate above keyboard displayed the entries of the truth table. The keyboard had black-and-white keys like a piano used to enter the premises. The Logic Piano was the culmination of a long series of inventions by Stanhope that aided in the calculation of syllogisms including a logical alphabet, slate, and stamp that would quickly produce the lines of a truth table in a logical argument. Stanhope’s logical machines used the generate-and-test procedure where all the possible combinations are generated and the impossible conclusions are removed. The results are the broadest conclusions that could possibly be produced from the premises. The truth table was more like a spreadsheet representing all of the logical combinations. His machine could only process four-terms but he planned to develop a 10-term engine which would have required enough space to display 1,024 combinations of its logical alphabet. Figure 1 is an image of Jevon’s Logic Piano.

Figure 1: William Jevon’s Logic Piano

In each of these epochs, there were what we now recognize as a hype cycle where it was believed that we were at the veritable precipice of duplicating human intelligence by a machine with all of the rewards and punishments that achievement entails. There were differently misunderstood, misconstrued, misused, sometimes ambiguous terminologies, e.g. ratiocinator, automata, cybernetics, and artificial intelligence which all refer to the same underlying efforts. In the Logical Machines Epoch, Charles Babbage and his colleagues could be considered instigators of a hype cycle in their time. He invented several mechanical devices that he proposed could compute mathematical and logical functions. The *Difference Engine* (in the 1820s) was to calculate and print various kinds of logarithmic and trigonometric tables and the *Analytical Engine* (in the mid-1830s) was to extend its range into logic and employ abstract symbolic algebra. Neither of these machines was ever completed. Although Babbage’s intentions were for his machines to perform mathematical and logical functions, his colleagues were allowed to associate mental powers to the Analytical Engine. Babbage allowed his colleagues to

do so, creating hype in order to obtain financial and public support as well as political attention for his projects. Babbage also made claims about his engines. In his 1832 book, *On the Economy of Machinery and Manufactures*, Babbage claimed that his Difference Engine could replace the third section of De Prony's work division scheme for producing mathematical tables. The third section was made up of 60 or 80 people who could add, subtract, and perform computations and return the results to the second section for checking. Because his Difference Engine was not able to produce complete tables, it was impossible for him to get financial support for his Analytic Engine.

Different time periods, different cultural trends, different geographical locations, but generally the same basic motivation, i.e, build a machine that can duplicate, or simulate the operation of the human brain and mind. Perhaps this realization can inform our mission to develop a standard model that identifies and clarifies what we as researchers, educators, engineers and practitioners mean by the term Artificial Intelligence. Figure 2 shows a Venn diagram of the various disciplines that define AI.

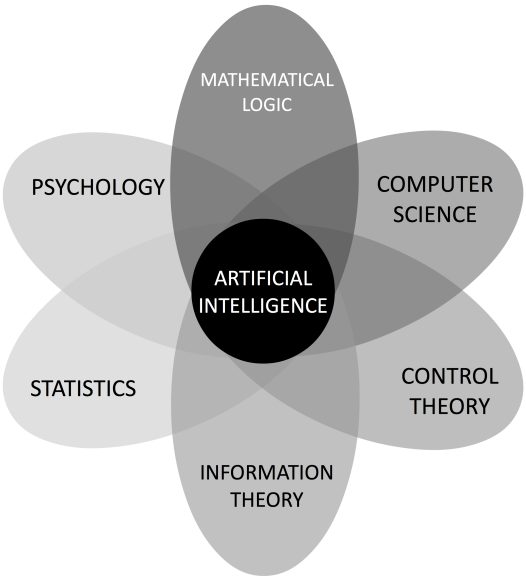


Figure 2: Venn Diagram of the Disciplines that define AI.

Benefits of a Standard Model for AI

Is it possible that there is a single standard model that describes the intersection shown in Figure 2? If we had such a standard model would that make it easier to inform and educate the public with respect to AI? Would a standard model of AI make the recruitment of students and future researchers more straight forward? Would that standard model support the notion of a single AI Cosmology? Could such a model really capture George Boole's intention to codify the Laws of Thought, or realize Leibniz's dream of reducing human reason to logical calculation? How far back in time do the epochs of imbuing machines with the operations of the brain/mind extend? Is there a physics of knowledge that underlies a single standard model for AI? Stay tuned!

References

Aspray, W. (1990). *Computing Before Computers*. Iowa State University Press.

Babbage, C., Babbage, H.P. ed. (1889). *Babbage's Calculating Engines. A Collection Papers Relating To Them; Their History, and Construction*. London. E. and F.N.

Babbage, C. (1832). *On the Economy of Machinery and Manufactures.* Cambridge, Cambridge University Press.

Gardner, M. (1958). *Logic Machines and Diagrams. New York*. McGraw Hill Book Co.

Green, C.D. (2005). *History of Psychology: Was Babbage's Analytical Engine Intended To Be A Mechanical Model Of The Mind?* Educational Publishing Foundation. Vol. 8, No. 1, p. 35D 45.

Jevons, W.S.(1869). *The Substitution of Similars, The True Principle of Reasoning Derived From a Modification of Aristotle's Dictum*. Macmillan.

Sieg, W. (2008). *Handbook of the Philosophy of Science. Philosophy of Mathematics: On Computability*. North Holland p. 535.

Wiener, P.P. ed.(1951). *Leibniz Selections.* C. Scribner's Sons.

Cameron Hughes is a computer and robot programmer. He is a Software Epistemologist at Ctest Laboratories where he is currently working on A.I.M. (Alternative Intelligence for Machines) and A.I.R (Alternative Intelligence for Robots) technologies. Cameron is the lead AI Engineer for the Knowledge Group at Advanced Software Construction Inc. He is a member of the advisory board for the NREF (National Robotics Education Foundation) and the Oak Hill Robotics Makerspace. He is the project leader of the technical team for the NEOACM CSI/CLUE Robotics Challenge and regularly organizes and directs robot programming workshops for varying robot platforms. Cameron Hughes is the co-author of many books and blogs on software development and Artificial Intelligence.

Tracey Hughes is a software and epistemic visualization engineer at Ctest Laboratories. She is the lead designer for the MIND, TAMI, and NO-FAQS projects that utilize epistemic visualization. Tracey is also a member of the advisory board for the NREF (National Robotics Education Foundation) and the Oak Hill Robotics Makerspace. She is the lead researcher of the technical team for the NEOACM CSI/CLUE Robotics Challenge. Tracey Hughes is the co-author with Cameron Hughes of many books on software development and Artificial Intelligence.

Experiential AI

Drew Hemment (University of Edinburgh; drew.hemment@ed.ac.uk) **Ruth Aylett** (Heriot-Watt University; r.s.aylett@hw.ac.uk)
Vaishak Belle (University of Edinburgh; Vaishak@ed.ac.uk)
Dave Murray-Rust (University of Edinburgh; D.Murray-Rust@ed.ac.uk)
Ewa Luger (University of Edinburgh; Ewa.Luger@ed.ac.uk)
Jane Hillston (University of Edinburgh; Jane.Hillston@ed.ac.uk)
Michael Rovatsos (University of Edinburgh; Michael.Rovatsos@ed.ac.uk)
Frank Broz (Heriot-Watt University; f.broz@hw.ac.uk)
DOI: 10.1145/3320254.3320264

Abstract

Experiential AI is proposed as a new research agenda in which artists and scientists come together to dispel the mystery of algorithms and make their mechanisms vividly apparent. It addresses the challenge of finding novel ways of opening up the field of artificial intelligence to greater transparency and collaboration between human and machine. The hypothesis is that art can mediate between computer code and human comprehension to overcome the limitations of explanations in and for AI systems. Artists can make the boundaries of systems visible and offer novel ways to make the reasoning of AI transparent and decipherable. Beyond this, artistic practice can explore new configurations of humans and algorithms, mapping the terrain of inter-agencies between people and machines. This helps to viscerally understand the complex causal chains in environments with AI components, including questions about what data to collect or who to collect it about, how the algorithms are chosen, commissioned and configured or how humans are conditioned by their participation in algorithmic processes.

Introduction

AI has once again become a major topic of conversation for policymakers in industrial nations and a large section of the public.

In 2017, the UK published Ready, Willing and Able, a landscape report (House Of Lords Select Committee, 2018). It clearly states that "everyone must have access to the opportunities provided by AI" and argues the need for public understanding of, and engagement with AI to develop alongside innovations in the field. The report warns of the very real risk of "societal and regional inequalities emerging as a consequence of the adoption of AI and advances in automation" (*Ibid.*). It also assesses issues of possible harm from malfunctioning AI, and resulting legal liabilities. However, it stops short of considering more pervasive downsides of applying AI decision-making across society. Alongside the sometimes exaggerated claims of AI's current or immediate-future capabilities, a broader set of fears about negative social consequences arise from the fast-paced deployment of AI technologies and a misplaced sense of trust in automated recommendations. While some of these fears may themselves be exaggerated, negative outcomes of ill-designed data-driven machine learning technologies are apparent, for example where new knowledge is formulated on undesirably biased training sets. The notorious case of Google Photos grouping some humans with primates on the basis of skin tone offered a glimpse of the damage that can be done. Such outcomes may not be limited to recommendations on a mobile phone: social robots share everyday spaces with humans, and might also be trained on impoverished datasets. Imagine, for example, a driverless car not recognizing specific humans as objects it must not crash into. So much for Asimovs laws!

Accountability and explainability in AI

The AI community has, of course, not been silent on these issues, and a broad range of solutions have been proposed. We broadly

classify these efforts into two related categories: accountability and explainability.

The first category seeks to identify the technical themes that would make AI trustworthy and accountable. Indeed, we can see AI technologies are already extending the domains of automated decision making into areas where we currently rely on sensitive human judgements. This raises a fundamental issue of democratic accountability, since challenging an automated decision often results in the response "it's what the computer says". So operators of AI need to know the limits and bounds of the system, the way that bias may present in the training data, or we will see more prejudice amplified and translated to inequality. From the viewpoint of AI research, there is a growing scientific literature on fairness (Kleinberg, Ludwig, Mullainathan, & Rambachan, 2018) to protect those otherwise disenfranchised through algorithmic decisions, as well as engineering efforts to expose the limitations of systems. Accountability can be a deeper property of the system too: for example, an emerging area of AI research looks at how ethical AI systems might be designed (Conitzer, Sinnott-Armstrong, Borg, Deng, & Kramer, 2017; Halpern & Kleiman-Weiner, 2018; Hammond & Belle, 2018).

The second category investigates how the decisions and actions of machines can be made explicable to human users (Gunning, 2017). We are seeing a step change in the number of people both currently and potentially impacted by automated decisions. Whilst the use of algorithms can now be said to be common (Domingos, 2015), concerns arise where complex systems are applied in the generation of sensitive social judgments, such as in social welfare, healthcare, criminal justice, and education. This has led to a call to limit the use of "black box" systems in such settings (Campolo, Sanfilippo, Whittaker, & Crawford, 2017). However, if one asks for a rationale for a decision, usually none is given, not least because those working in organisations using automated decision-making do not themselves have any insight into what the algorithms driving it are doing. This is a form of conditioning, creating passivity rather than engagement. At the other extreme, if people do not understand the decisions of AI systems, they may simply not use those systems. Be that as it may, progress in the field has been exciting but a single solution is elusive. Some strands of research focus on using simpler models (possibly at the cost of prediction accuracy), others attempt "local" explanations that identify interpretable patterns in regions of interest (Weld & Bansal, 2018; Ribeiro, Singh, & Guestrin, 2016), while still others attempt human-readable reconstructions of high-dimensional data (Penkov & Ramamoorthy, 2017; Belle, 2017). However, this work addresses explainability as primarily a technical problem, and does not account for human, legal, regulatory or institutional factors. What is more, it does not generate the kind of explanations needed from a human point of view. A person will want to know why there was one decision and not another, the causal chain, not an opaque description of machine logic. There are distinctions to be explored between artificial and augmented intelligences (Carter & Nielsen, 2017), and a science, and an art, to be developed around human-centred machine learning (Fiebrink & Gillies, 2018).

For there to be responsible AI, transparency is vital, and people need comprehensible explanations. Core to this is the notion that unless the operation of a system is visible, and people can access comprehensible explanations, it cannot be held to account. Even when an explanation can be provided, this may not always be sufficient (Edwards & Veale, 2017) and more intuitive solutions are required to, for example, understand the changing relations between data and the world, or integrate domain knowledge in ways that connect managers with those at the frontlines (Veale, Van Kleek, & Binns, 2018). In Seeing without knowing, Ananny and Crawford argue research needs not to look *within* a technical system, but to look *across* systems and to address both human and non-human dimensions (Ananny & Crawford, 2018). They call for "a deeper engagement with the material and ideological realities of contemporary computation" (*Ibid.*).

Artists addressing such AI challenges

There is a mature tradition of work between art and technology innovation going back to the 1960s and 1970s (Harris, 1999; Gere,

Figure 1: Neural Glitch 1540737325 Mario Klingemann 2018

2009). Artists are beginning to experiment in AI as subject and tool, and several high profile programmes are a testament to the fertility of this field (Zentrum fur Kunst und Medien, 2018; Ars Electronica, 2018). Such practice can create experiences around social impacts and consequences of technology, and create insights to feed into the design of the technologies (Hemment, Bletcher, & Coulson, 2017).

One theme evident among artists working with machine learning algorithms today, such as Mario Klingemann[1] and Robbie Barrat[2], is to reveal distortions in the ways algorithms make sense of the world – see Figure 1 for an example. This kind of approach enables the character of machine reasoning and vision to be made explicit, and its artifacts to be made tangible. This, in turn, creates a concrete artefact or representation that can be used as an object for discussion and to spark further enquiry, helping to build literacy in those systems.

In the contemporary experience of AI, the disturbing yet compelling output of DeepDream has shaped our view on what algorithms do, although it is questionable how representative this is of deep network structures, or whether it is a happy accident in machine aesthetics. Either way, it has prompted artistic exploration of the social implications of AI, with projects using deep learning to generate faces (*Plugging 50,000 portraits into facial recognition*, 2018) and Christies auctioning neural network generated portraits (*Is artificial intelligence set to become arts next medium?*, 2018). Going beyond the typical human+computer view, artists are questioning the construction of prejudice and normalcy (*http://mushon.com/tnm*, 2018), and working with AI driven prosthetics, to open possibilities for more intimate entanglements (Donnarumma, 2018).

Art can both make ethical standards concrete, and allow us to imagine other realities. While high-level ethical principles are easy to articulate, they sit at a level of generality that may make their practical requirements less obvious. Equally, they signal the existence of clear solutions, externalise responsibility, and obscure the true complexity of the moral problems resulting from socially situated AI. Ethical issues must be concretely internalised by developers and users alike to avoid failures like Cambridge Analytics or the Facebook Emotional Contagion experiment (Jouhki, Lauk, Penttinen, Sormanen, & Uskali, 2016). Experiential approaches (Kolb, 2014) can act as a powerful mechanism, and embedding relevant experiences in a story-world through narrative, and especially role-play, can generate safe reflection spaces," as for example Boal's Forum Theatre (Boal, 2013).

Accountability is variously addressed. Joy Buolamwini works with verse and code to challenge harmful bias in AI[3], while Trevor Paglen constructs a set of rules for algorithmic systems in such a way as to uncover the character of that rule space[4]. A thriving community of practitioners from across the arts and sciences is working to avoid detection [5] or trick classification systems (Sharif, Bhagavatula, Bauer, & Reiter, 2016). Such artistic experiments bring to life and question what an algorithm does, what a system could be used for, and who is in control.

Experiential AI theme and call for artists

The field of Experiential AI seeks to engage practitioners in computation, science, art and design around an exploration of how humans and artificial intelligences relate, through the physical and digital worlds, through decisions and shaping behaviour, through collaboration and co-creation, through intervening in existing situations and through creating new configurations.

The Experiential AI theme begins with a call for artists in residence, launched in August 2019, as a collaboration between the Experiential AI group at University of Edinburgh, Ars Electron-ica in Linz, and Edinburgh International Festi-val [6]. The focus is on creative experiments in which AI scientists and artists are jointly en-gaged to make artificial intelligence and ma-chine learning tangible, interpretable, and ac-

[1] http://quasimondo.com/
[2] https://robbiebarrat.github.io/
[3] https://www.poetofcode.com/
[4] http://www.paglen.com/
[5] https://cvdazzle.com/
[6] https://efi.ed.ac.uk/art-and-ai-artist-residency-and-research-programme-announced/

cessible to the intervention of a user or audience. The ambition is to help us think differently about how algorithms should be designed, and open possibilities for radically new concepts and paradigms.

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989.

Ars Electronica. (2018). *Media art between natural and artificial intelligence.* https://ars.electronica.art/ai/en/media-art-between-natural-and-artificial-intelligence/.

Belle, V. (2017). Logic meets probability: Towards explainable ai systems for uncertain worlds. In *Ijcai* (pp. 5116–5120).

Boal, A. (2013). *The rainbow of desire: The boal method of theatre and therapy*. Routledge.

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). Ai now 2017 report. *AI Now Institute at New York University*.

Carter, S., & Nielsen, M. (2017). Using artificial intelligence to augment human intelligence. *Distill*, *2*(12), e9.

Conitzer, V., Sinnott-Armstrong, W., Borg, J. S., Deng, Y., & Kramer, M. (2017). Moral decision making frameworks for artificial intelligence. In *Thirty-first aaai conference on artificial intelligence.*

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Penguin.

Donnarumma, M. (2018). *Is artificial intelligence set to become arts next medium?* https://marcodonnarumma.com/works/ai-ethics-prosthetics/.

Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, *16*, 18.

Fiebrink, R., & Gillies, M. (2018, June). Introduction to the special issue on human-centered machine learning. *ACM Trans. Interact. Intell. Syst.*, *8*(2), 7:1–7:7. Retrieved from http://doi.acm.org/10.1145/3205942 doi: 10.1145/3205942

Gere, C. (2009). *Digital culture*. Reaktion Books.

Gunning, D. (2017). *Explainable artificial intelligence (xai).* https://tinyurl.com/yccmn477. (Accessed: 12/3/18)

Halpern, J. Y., & Kleiman-Weiner, M. (2018). Towards formal definitions of blameworthiness, intention, and moral responsibility. In *Thirty-second aaai conference on artificial intelligence.*

Hammond, L., & Belle, V. (2018). Deep tractable probabilistic models for moral responsibility. *arXiv preprint arXiv:1810.03736*.

Harris, C. (1999). The xerox palo alto research center artist-in-residence program landscape. In *Art and innovation* (pp. 2–11).

Hemment, D., Bletcher, J., & Coulson, S. (2017). Art, creativity and civic participation in IoT and Smart City innovation through Open Prototyping. Paper presented at Creativity World Forum, Aarhus, Denmark.

House Of Lords Select Committee. (2018). *Ai in the uk: ready, willing and able?* https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.

Is artificial intelligence set to become arts next medium? (2018). https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx.

Jouhki, J., Lauk, E., Penttinen, M., Sormanen, N., & Uskali, T. (2016). Facebooks emotional contagion experiment as a challenge to research ethics. *Media and Communication*, *4*.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic fairness. In *Aea papers and proceedings* (Vol. 108, pp. 22–27).

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

Penkov, S., & Ramamoorthy, S. (2017). Using program induction to interpret transition system dynamics. *arXiv preprint arXiv:1708.00376*.

Plugging 50,000 portraits into facial recognition. (2018). https://www.reddit.com/r/Damnthatsinteresting/comments/9udese/plugging_50000

_portraits_into_facial/.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144).

Sharif, M., Bhagavatula, S., Bauer, L., & Reiter, M. K. (2016). Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the 2016 acm sigsac conference on computer and communications security* (pp. 1528–1540).

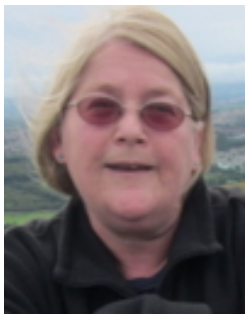
(2018).

Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 440).

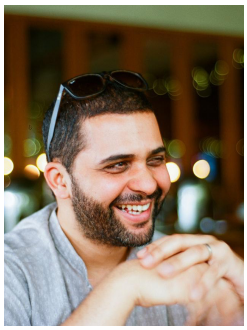
Weld, D. S., & Bansal, G. (2018). Intelligible artificial intelligence. *arXiv preprint arXiv:1803.04263*.

Zentrum fur Kunst und Medien. (2018). *Encoding cultures: Living amongst intelligent machines.* https://zkm.de/en/event/2018/04/encoding-cultures-living-amongst-intelligent-machines.

Drew Hemment is a Chancellors Fellow and Reader at Edinburgh Futures Institute and Edinburgh College of Art, University of Edinburgh. He is PI of of GROW Observatory (EC H2020), Founder of FutureEverything, and on the Editorial Board of Leonardo. His work over 25 years in digital arts and innovation has been recognised by awards including STARTS Prize 2018, Lever Prize 2010 and Prix Ars Electronica 2008.

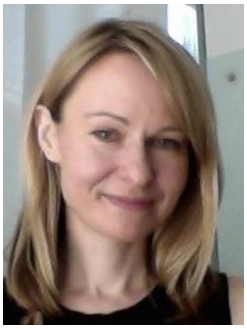


Ruth Aylett is a Professor of Computer Science at Heriot-Watt University in Edinburgh. She researches social agents, Human-Robot Interaction and affective systems, taking a human-centred design approach to the development of socially-useful systems. Her current project explores the use of a robot to assist adults with autism in social signal processing



Vaishak Belle is a Chancellors Fellow/Lecturer at the School of Informatics, University of Edinburgh, an Alan Turing Institute Faculty Fellow, and a member of the RSE (Royal Society of Edinburgh) Young Academy of Scotland. At the University of Edinburgh, he directs the Belle Lab, which specializes in the unification of symbolic systems and machine learning.

Dave Murray-Rust is a Lecturer in Design Informatics at the University of Edinburgh, exploring ways that people, data and things interact. His research centres on how we can ensure that there is space for people within computational systems, preserving privacy, choice, identity and humanity while making use of possibilities of computational coordination and personal data.



Ewa Luger is a Chancellor's Fellow at the University of Edinburgh, a consulting researcher at Microsoft Research UK (AI and Ethics), and a fellow of the Alan Turing Institute. She explores applied ethical issues within the sphere of machine intelligence and data-driven systems. This includes data governance, consent, privacy and how intelligent systems might be made intelligible to the user.

Jane Hillston was appointed Professor of Quantitative Modelling in the School of Informatics at the University of Edinburgh in 2006, having joined the University as a Lecturer in Computer Science in 1995. She is currently the Head of School. Her research is concerned with formal approaches to modelling dynamic behaviour, particularly the use of stochastic process algebras for performance modelling and stochastic verification.

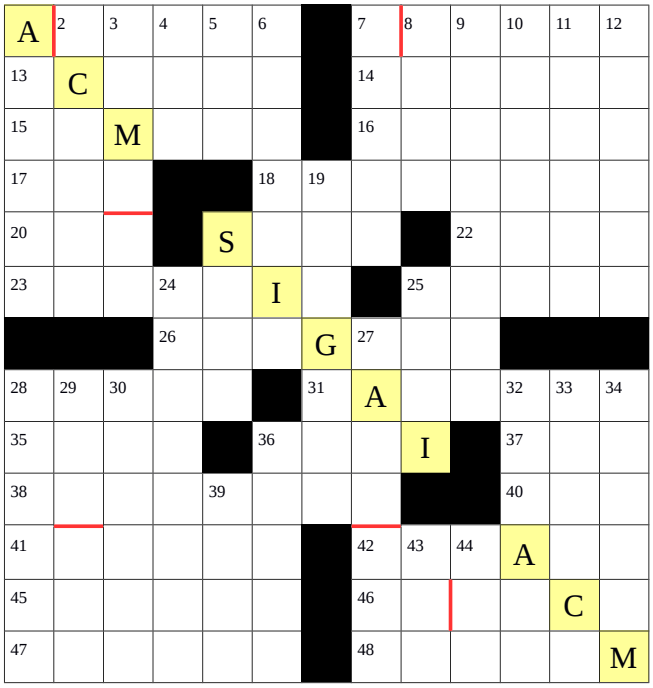
Michael Rovatsos is a Reader (Associate Professor) at the School of Informatics of the University of Edinburgh, and Director of the Bayes Centre for Data Science and AI. He research interests are in intelligent agents and multiagent systems, and most of his recent work has focused on the human-centric of AI algorithms and ethical AI more generally.

Frank Broz is an Assistant Professor of Computer Science at Heriot-Watt University. His research interests are in human-robot interaction, AI, and social robotics. He has consulted for artists working with technology and collaborated on robotic art installations such as Reach, Robot (created for Pittsburghs 250th anniversary celebration). He received his PhD from Carnegie Mellon Universitys Robotics Institute.

Crosswords

Adi Botea (IBM Research, Ireland; adibotea@ie.ibm.com)

DOI: 10.1145/3320254.3322194



Across: **1)** One mocking remark **7)** Elite groups **13)** In short supply **14)** ___ Ahern, Irish politician **15)** The most uninspiring **16)** The most uncovered **17)** Good-quality rock **18)** The most impacted by the cloud **20)** Canadian prov. **21)** Sound in the head **22)** Boolean constant **23)** Baltic state **25)** Wave-amplifying device **26)** Word game **28)** Muscle pain **31)** Cars with flexible protection **35)** Damage **36)** City in Italy **37)** Respond emotionally **38)** Times between **40)** It is shorter **41)** Internet device **42)** Old coin from todays Istanbul **45)** Pressed to serve a side dish **46)** Immediately **47)** Baby birds **48)** Use a coupon

Down: **1)** Known to have a serious inclination **2)** Frightens **3)** Spell out **4)** Before... being old **5)** Enterpr. Comput. Serv. **6)** Adapt a learning-based system **7)** ___ Cornish, actress and rapper **8)** To be slightly off the borderline as a reviewer **9)** Annoying thing **10)** Acts like a supervisor **11)** Paper towel **12)** Basic method in object oriented programming **19)** A mixture of characters **21)** Educational programming language developped from... Scratch **24)** Device that measures electrical resistance **25)** The three wise men **27)** Pocket-size source of energy **28)** Tweets live **29)** Finished **30)** Art in Barcelona **32)** Fuel component **33)** Man of royal descent **34)** Operating ___ in IT **36)** Angry ones in an AI competition **39)** Smell badly **43)** Summer in Provence **44)** Brief reference to the horoscope signs

Hint: When the solution to a clue has two words, a thick red bar in the grid shows the separation between the two words.

Acknowledgment: I thank Karen McKenna for her feedback. The grid is created with the AI system Combus (Botea, 2007).

References

Botea, A. (2007). Crossword Grid Composition with A Hierarchical CSP Encoding. In *Proceeding of the 6th CP Workshop on Constraint Modelling and Reformulation ModRef-07.*

Adi Botea is a research scientist at IBM Research, Ireland. His interests include AI planning, heuristic search, automated dialogue systems, pathfinding, journey planning, and multi-agent path planning. Adi has co-authored more than 70 peer-reviewed publications, and more than 20 patent applications. He has published crosswords puzzles in Romanian, in national-level publications, for almost three decades.