

# The Association for the Advancement of Artificial Intelligence 2020 Workshop Program

*Grace Bang, Guy Barash,  
Ryan Bea, Jacques Cali, Mauricio  
Castillo-Effen, Xin Cynthia Chen, Niyati Chhaya,  
Rohan Dhoopar, Sebastijan Dumancic, Huáscar Espinoza,  
Eitan Farchi, Ferdinando Fioretto, Raquel Fuentetaja, Michel  
Galley, Christopher Geib, José Hernández-Orallo, Xiaowei Huang,  
Sarah Keren, Seokhwan Kim, Kokil Jaidka, Xiaomo Liu, Tyler Lu,  
Zhiqiang Ma, Richard Mallah, John McDermid, Martin Michalowski,  
Reuth Mirsky, Seán Ó Héigeartaigh, Deepak Ramachandran, Javier  
Segovia-Aguas, Arash Shaban-Nejad, Onn Shehory, Vered Shwartz,  
Siddharth Srivastava, Kartik Talamadupula, Jian Tang, Dell Zhang, Jian Zhang*

■ *The Association for the Advancement of Artificial Intelligence 2020 Workshop Program included twenty-three workshops covering a wide range of topics in artificial intelligence. This report contains the required reports, which were submitted by most, but not all, of the workshop chairs.*

**T**he Association for the Advancement of Artificial Intelligence (AAAI) 2020 Workshop Program was held February 7–8, 2020 in New York, New York. Biplav Srivastava and Min-Ling Zhang served as cochairs of the event, which included the following twenty-three workshops: Affective Content Analysis: Interactive Affective Response (W1), Artificial Intelligence for Cyber Security (W2), Artificial Intelligence for Education (W3), Artificial Intelligence

in Team Sports (W4), Artificial Intelligence of Things (W5), Artificial Intelligence Safety (W6), Cloud Intelligence: Artificial Intelligence (AI)/Machine Learning (ML) for Efficient and Manageable Cloud Services (W7), Deep Learning on Graphs: Methodologies and Applications (W8), Dialog System Technology Challenge (W9), Engineering Dependable and Secure Machine Learning Systems (W10), Evaluating Evaluation of AI Systems (W11), Generalization in Planning (W12), Health Intelligence (W13), Intelligent Process Automation — Robotic Process Automation (RPA) Meets AI (W14), Interactive and Conversational Recommendation Systems (W15), Knowledge Discovery from Unstructured Data in Financial Services (W16), Plan, Activity, and Intent Recognition (W17), Privacy-Preserving Artificial Intelligence (W18), Reasoning and Learning for Human-Machine Dialogues (W19), Reasoning for Complex Question Answering (W20), Reinforcement Learning in Games (W21), Reproducibility in AI (RAI 2020) — Future Direction and Reproducibility Challenge (W22), and Statistical Relational AI (W23).<sup>1</sup>

## Affective Content Analysis: Interactive Affective Response (W1)

The third Workshop on Affective Content Analysis was conducted to stimulate interdisciplinary discussions for affect in content, and engage the AI and ML community about the open problems in affective content analysis and understanding, with a special focus on affect in language and text. The theme of the workshop was interactive affective responses. It also hosted an AI shared task — Computational Linguistics Affect: Get It #OffMyChest; to encourage the development of new models and approaches for modeling disclosure and supportiveness in text conversations.

This year, the focus of the workshop was on considering subjectivity and context when modeling *affect* as a response to an interaction. A few key challenges include standardizing the measurement of interactive affect, and the cross-media, cross-domain, and cross-platform analysis of affect.

The workshop program focused on the analysis of emotions, sentiments, and attitudes in textual, visual, and multimodal content for applications in psychology, natural language processing (NLP), computer vision, and marketing science. The shared task system submissions of the Computational Linguistics Affect: Get It #OffMyChest, along with the results, were discussed. Besides original research presentations and posters, the workshop also hosted a range of keynote speakers that highlighted the state of the art in affective computing in a range of fields:

Highlights of the workshop included Louis-Philippe Morency from Carnegie Mellon University, who shared some exciting work on multimodal affect analysis, focusing particularly on understanding human communication dynamics. Daniel McDuff from Microsoft Research presented work on building

intelligent and visceral machines. Natasha Jaques, from Google Brain, presented an approach leveraging reinforcement learning for affective content generation in dialogs. Robert Kraut from Carnegie Mellon University provided the humanities angle at looking at conversations in social groups. The study discussed self-disclosure and supportiveness in social group dynamics. Finally, Donna Hoffman and Tom Novak, from George Washington University, discussed ML approaches for discovering triggers and patterns in a real-world Internet-of-Things application. The marketing science perspective here provided insights beyond the computer-science, AI community.

The workshop concluded with a panel discussion among the keynote speakers that was moderated by the organizers, on the potential directions for future events, and the scope of interdisciplinary research. The workshop papers have been published through CEUR Workshop Proceedings.<sup>2</sup>

Niyati Chhaya, Kokil Jaidka, Jennifer Healey, Lyle Ungar, and Atanu Sinha cochaired the workshop. This report was written by Niyati Chhaya and Kokil Jaidka.

## Artificial Intelligence for Cyber Security (W2)

No report from the Artificial Intelligence for Cyber Security workshop organizers was submitted.

## Artificial Intelligence for Education (W3)

No report from the Artificial Intelligence for Education workshop organizers was submitted.

## Artificial Intelligence in Team Sports Workshop (W4)

The goal of the AAAI-20 workshop on Artificial Intelligence in Team Sports was to build a community of AI researchers in team sports and showcase recent works.

Sports is a domain that has grown significantly over the last 20 years, and now has become a key driver of many economies. As the market has grown so has the amount of data that is collected. This means that there are a number of challenging problems to predict and optimize performance but, so far, such problems have largely been dealt with by domain experts (such as coaches, managers, scouts, and sports health experts) with basic analytics.

The growing availability of datasets in sports presents a unique opportunity for the AI and ML communities to develop, validate, and apply new techniques in the real world. In team sports, real-world data are available over long periods of time, about the same individuals and teams, in a variety of environmental contexts, thereby creating a unique live test-bed for AI and ML techniques. Hence, this

workshop brought together leading researchers in the AI and sports-analytics communities and fueled discussion on new ideas in the field. We hope that this will encourage new research that will benefit both communities and industry.

The workshop attracted speakers and attendees from across a number of different disciplines, across many countries, and with backgrounds in a wide variety of sports. There were many high-quality submissions to the workshop. On the day, seven of these were presented orally and three were presented in the poster session. We had focus on sports such as football (soccer), ice hockey, basketball, and curling. One theme that we saw at the event were papers focused on evaluating the value of individual players in teams. Yudong Luo (Simon Fraser University, Canada) discussed how reinforcement learning can be used to do this in ice hockey, and Maaïke Van Roy (KU Leuven, Belgium) presented a critical comparison of some key approaches to this in football. We also saw a theme of papers focused on strategy in team sports. Leonardo Lamas (University of Brasilia, Brazil) presented research focused on a simulation approach to basketball strategy, and Kotaro Ataka (Hokkaido University, Japan) presented his work on curling strategy and predicting scoring. In the remaining presentations, Tom Decroos (Katholieke Universiteit Leuven, Belgium) discussed his work on interpretable prediction of goals in soccer, Kanav Vats (University of Waterloo, Canada) discussed using computer vision for puck tracking in ice hockey, and finally, Guiliang Liu (Simon Fraser University, Canada) presented his work on learning contextualized player representations with a variational hierarchical encoder.

Accompanying the research presentations were three prestigious keynote speakers who presented their work. Laurie Shaw from Harvard University discussed his recent award-winning (FC Barcelona 2019)<sup>3</sup> work on dynamic analysis of team strategy in professional soccer. Patrick Lucey from STATSPerform (one of the leading sports analytics companies) presented a talk on interactive sports analytics, which showed some of his company's work on the ability to evaluate the performance of a player or a team in a given situation and compare it against another player in exactly the same position. Finally, Mehrsan Javan, cofounder and Chief Technology Officer of Sportlogiq, a leading sports computer vision company, demonstrated some of the real-time tools that they have developed over the last few years using broadcast footage in ice hockey. The workshop concluded with a panel session consisting of the keynote speakers (and chaired by Tim Swartz), during which attendees were able to ask questions.

The organizing committee for the workshop included Ryan Beal, Arvapali Ramchurn, Georgios Chalkiadakis, Onn Shehory, and Tim Swartz. This report was written by Ryan Beal. We would like to thank the program committee for their help reviewing papers. All papers are available on the workshop website.<sup>4</sup>

## Artificial Intelligence of Things Workshop (W5)

The Artificial Intelligence of Things workshop served as a forum for researchers and practitioners from multiple areas such as devices, sensor networks, AI, and ML, to share and learn AI-powered internet-of-things solutions. It received twenty-four submissions, twelve of which were accepted.

Jian Zhang (Microsoft), program cochair, represented the organizing committee to make opening remarks, after which three distinguished researchers delivered keynote addresses. In the opening keynote, FarmBeats: Empowering Farmers with Affordable Digital Agriculture, Ranveer Chandra, the chief scientist at Microsoft Azure Global, introduced FarmBeats,<sup>5</sup> a system that Microsoft developed through integration of cloud, internet of things, and AI innovations for agriculture. The system enables seamless collection and analysis of data across various sensors, cameras, drones, and satellites to boost agricultural productivity by increasing yields, reducing losses, and cutting down input costs. He also outlined some of the AI challenges they are currently addressing for agriculture.

Diana Marculescu, department chair in electrical and computer engineering at the University of Texas at Austin, presented the keynote Putting the *Machine* Back in Machine Learning: The Case for Hardware-ML Model Codesign. She uncovered the need for building accurate, platform-specific power and latency models for convolutional neural networks and efficient hardware-aware convolutional neural networks design methodologies, which enables machine learners and hardware designers to identify a neural network configuration that offers not just the best accuracy but also satisfies given hardware constraints. She introduced a modeling, analysis, and optimization framework that enables the codesign of hardware and ML model orders-of-magnitude faster than state of the art, while satisfying both accuracy and latency or energy constraints.

In the closing keynote, The Opportunities and Challenges of Artificial Intelligence and the Internet of Things, Jie Liu, dean of AI research at the Harbin Institute of Technology China, outlined challenges in AI internet of things through use cases in ambient intelligent environments. He introduced research that tackles these challenges such as how multimodality sensor fusion can overcome the limitation of single sensor types; how to use simulation to generate useful training data; and how to use neural architecture search to optimize the models for embedded platforms. He concluded by sharing a few future research directions.

Invited speaker Xiang Sheng, a research scientist at Facebook research, introduced the sensing as a service concept and how it is used to support Facebook's mission to give people the power to build community and bring the world closer together.

In addition, authors of accepted submissions presented their work in the project showcase and technical paper sessions. There were two major themes in the technical paper sessions — model optimization to support running neural network on resource limited devices and image; and video processing techniques to support AI-of-things scenarios such as autonomous driving.

In the project showcase session, Microsoft showcased their in-car cognition solutions with edge AI accelerators, and HCL Technologies introduced their radar tracking system and analytic solutions for retail business management.

Additional information about the workshop is located at the Third International Workshop on Artificial Intelligence of Things on the web.<sup>6</sup>

Jian Zhang (Microsoft) and Jian Tang (DiDi ChuXing) served as program coauthors of the workshop. Yiran Chen (Duke University) served as general chair of the workshop. The papers of the workshop were published on the workshop website. This report was written by Jian Zhang and Jian Tang.

## Artificial Intelligence Safety (W6)

The AAAI-20 Workshop on Artificial Intelligence Safety aimed to explore new ideas at the intersection of AI and safety, as well as broader strategic, ethical, and policy-oriented aspects of safety and AI as a whole.

Safety in AI is increasingly becoming a substantial aspect of AI research, deeply intertwined with the ethical, legal, and societal issues associated with AI systems. Even if AI safety is considered a design principle, there are varying levels of safety, diverse sets of ethical standards and values, and varying degrees of liability, for which we need to deal with trade-offs or alternative solutions. The Artificial Intelligence Safety workshop sought to explore these issues by covering a wide range of AI paradigms, considering systems that are application-specific, and also those that are general. The main interest of the proposed workshop is to look holistically at AI and safety engineering, jointly with the ethical and legal issues, to build trustable intelligent autonomous machines. In addition, the workshop supports the AI safety landscape initiative.<sup>7</sup> This initiative aims at defining an AI safety landscape providing a view of the current needs, challenges, and state of the art and the practice of this field.

The workshop received forty-five submissions and accepted thirteen full papers, two talks, and fifteen posters, resulting in a full-paper acceptance rate of twenty-nine percent and an overall acceptance rate of sixty-seven percent. The workshop program was organized in five thematic sessions, which followed a highly interactive format. Each session included were short pitches, along with a panel to discuss both individual paper contributions and common issues.

Session 1 discussed adversarial ML. A novel bio-inspired approach was presented to deal with adversarial attacks. In addition, two adversarial ML approaches presented solutions for the problem of bias in

applications such as face-recognition systems. Session 2 explored solutions of assurance cases for AI-based systems, including a proposal of hazard-contribution modes of ML components, and a research work on assurance argument patterns and processes for ML in safety-related systems. Session 3 focused on considerations for building an AI safety landscape. This session included a proposal for founding the domain of AI forensics, and a work that explored AI safety in degrees — generality, capability, and control. Session 4 covered AI fairness and bias, by considering budgeted-constrained decision-making systems and multimodal and interactive explanation systems. Additionally, an approach suggested that models could be learned to conceal unfairness-from-explanation methods. Finally, Session 5 discussed uncertainty in safety-critical systems. The session discussed works for robust deep learning, safety guarantees, benchmarking uncertainty-estimation methods, and perceptual uncertainty.

A keynote opened the morning sessions. Ece Kamar (Microsoft Research AI) talked about AI in the open world, discovering blind spots of AI. Francois Terrier (Commissariat a l'Energie Atomique) talked about evolutionary qualification of AI-based systems. Finally, Sameer Singh (University of California Irving) presented a talk on evaluating and testing nature language processing systems.

Eight coauthors served the workshop — Huáscar Espinoza, José Hernández-Orallo, Xin Cynthia Chen, Seán Ó Héigeartaigh, Xiaowei Huang, Mauricio Castillo-Effen, Richard Mallah, and John McDermid. The papers were published as volume 2560 in the Central Europe workshop series.<sup>8</sup>

## Cloud Intelligence: AI and Machine Learning for Efficient and Manageable Cloud Services (W7)

The first Cloud Intelligence workshop brought together researchers and practitioners from a variety of fields such as systems, software analytics, AI, and data science to share and learn AI and ML solutions for efficient and manageable cloud services.

The workshop was launched with support from Microsoft along with representatives from both industry (including Google, Facebook, Salesforce, and Alibaba) and academe (including the Massachusetts Institute of Technology, Columbia University, Peking University, Queen's University, the University of California San Diego, and John Hopkins University, along with others).

In the opening keynote, Marcus Fontoura, a technical fellow of Microsoft Azure, shared the company's vision of infusing AI into the Azure platform and DevOps process along with a case study of Resource Central,<sup>9</sup> a novel ML and prediction-serving system for improving cloud resource management. In addition, Murali Chintalapati, a partner group engineering manager, and Yingnong Dang, a principal data

science manager, at Azure introduced the company's AI Ops effort, which leverages AI and ML solutions to improve system availability, engineering efficiency, and customer satisfaction.

In invited talk sessions, representative researchers, data scientists, and engineers shared their experience and view on cloud intelligence. Tim Kraska, an associate professor of electrical engineering and computer science and codirector of the Data System and AI Lab at the Massachusetts Institute of Technology, presented the invited talk *The Case for Learned Data Structures and Algorithms*. In his talk, Kraska shared that recent research result of his group shows that ML has the potential to significantly alter the way in which algorithms and data structures are implemented and the performance they can provide.

Wendy Zhao, a principal engineer and senior director of engineering at Alibaba Cloud Intelligence Business Group delivered an invited talk on intelligent cloud infrastructure management at Alibaba. In the talk, she overviewed some of the top issues that cloud infrastructure operation is facing, and shared some recent progress on specific topics such as resource capacity planning, datacenter anomaly detection, hardware failure prediction, and cluster-level self-healing.

Elena Novakovskaia, a principal data scientist at Salesforce, shared Salesforce's data science approach to augment and enhance the efficiency of data center operations with an interpretable ML model in her talk *Building an Interpretable ML Model for Proactive Data Center Management*.

In the panel interview, Igal Figlin, a general manager of Azure, interviewed Andy Li of Tongdun Company, Asaf Cidon of Columbia University, Sid Sen of Microsoft research, and Wendy Zhao of Alibaba on their views about the challenges and opportunities of cloud intelligence. The panelists pointed out the importance of community and open data sets and called for collaboration in this space.

In addition to the invited talks and panel, authors of accepted submissions also presented their work during a technical paper and a project showcase session. The topics presented covered a variety of techniques used for multiple aspects of cloud service management, such as capacity, performance, incidents, and security.<sup>10</sup>

We thank the authors for publishing and presenting their papers and projects in the Cloud Intelligence workshop, the program committee for their professional evaluation and help during the review process, and the steering committee for their support in launching this workshop.

Jian Zhang (Microsoft Azure) served as program chair of the workshop and wrote this report.

## Deep Learning on Graphs: Methodologies and Applications (W8)

No report from the Deep Learning on Graphs workshop organizers was submitted.

## Dialog System Technology Challenge (W9)

The Dialog System Technology Challenge (DSTC) has been a premier research competition for dialog systems since its inception in 2013. This workshop marks the eighth time the challenge has been held. Like its predecessors, it focused on end-to-end dialog tasks, to explore the issue of applying end-to-end technologies to dialog systems in a pragmatic way.

To accelerate the development of new dialog technologies, the DSTC competitions have provided common testbeds for various research problems. The earlier challenges focused on developing a single component for dialog-state tracking on goal-oriented human-machine conversations. The fourth and fifth challenges introduced human-human conversations and started to offer multiple tasks not only for dialog-state tracking, but also for other components in dialog. After it rebranded itself as the *DSTC*, the sixth and seventh events organized multiple main tracks in parallel to address a wider variety of dialog-related problems.

For this workshop challenge, we received seven track proposals and went through a formal peer review process focusing on each task's potential for broad interest from the research community, practical impact of the task outcomes, and continuity from the previous challenges. Finally, we ended up with four main tracks including two newly introduced tasks and two follow-up tasks from the seventh challenge.

The workshop challenge consisted of four tracks: Multidomain Task Completion, NOESIS II<sup>11</sup>: Predicting Responses, Audio Visual Scene-Aware Dialog, and Schema-Guided State Tracking.

The Multidomain Task Completion challenge addressed the end-to-end response generation problems in multidomain task completion and cross-domain adaptation scenarios. The NOESIS II: Predicting Responses challenge explores a response selection task extending the previous iteration of the NOESIS track, and offered two additional sub-tasks for identifying task success and disentangling conversations. The Audio-Visual Scene-Aware Dialog track also followed up a previous track and aimed to generate dialog responses using multimodal information given in an input video. Finally, the Schema-Guided State Tracking challenge revisited dialog-state tracking problems in a practical setting associated with a large number of services and application programming interfaces required to build virtual assistants in practice.

A total of 288 participants registered for DSTC8,<sup>12</sup> and seventy teams participated in the final challenge. We had a one-day wrap-up workshop to review the state-of-the-art systems, share novel approaches to the Challenge tasks, and discuss future directions for dialog technology. We had about seventy-five pre-registrations for the workshop and more participants joined on-site. We accepted thirty-four system papers

reporting the systems submitted to the DSTC8, as well as four papers describing the different tracks. To initiate DSTC9, we had a session to introduce the nine track proposals, for which up to four will be selected for next year's challenge. We had three sponsors to offer travel grants for student papers selected for oral presentations.

The DSTC organizing committee included Seokhwan Kim (general chair), Michel Galley (workshop chair), Chulaka Gunasekara (publication chair), and Sungjin Lee (publicity chair). This report was written by Seokhwan Kim and Michel Galley.

## Engineering Dependable and Secure Machine Learning Systems (W10)

The AAAI-2020 Workshop on Engineering Dependable and Secure Machine Learning Systems is the third in a series of workshops on this topic. Continuing the trend from previous workshops the event was well attended and inspired lively discussions, addressing adversarial ML, reliable ML, secure ML, and the relationships among them. The program included eleven presentations, of which one was an invited keynote presentation and ten were of peer-reviewed, accepted papers.

Contemporary software systems increasingly encompass ML components. In similarity to other software systems, ML-based systems must meet dependability, security, and quality requirements. Standard notions of software quality and reliability such as deterministic functional correctness, black-box testing, code coverage, and traditional software debugging may become irrelevant for ML systems. This is due to their nondeterministic nature, the reuse of high-quality implementations of ML algorithms, and the lack of understanding of the semantics of learned models such as when deep learning methods are applied. This calls for novel methods and new methodologies and tools to address quality and reliability challenges of ML systems.

Broad deployment of ML software in networked systems inevitably exposes the ML software to attacks. While classic security vulnerabilities are relevant, ML techniques have additional weaknesses, some already known (for example, sensitivity to data manipulation), and some yet to be discovered. Hence, there is a need for research as well as practical solutions to ML adversarial attacks.

The Engineering Dependable and Secure Machine Learning Systems workshop focused on such topics. It included original contributions exposing problems and offering solutions related to dependability, and quality assurance of, and adversarial attacks on, ML systems. It combined disciplines such as adversarial ML and software engineering (with emphasis on quality assurance). It also promoted a discourse between academia and industry in a quest for well-founded practical solutions.

The workshop was a well-attended, lively meeting of researchers from academia and industry. The presentations, and discussions that followed, were very fertile and inspiring, leading to new future research on adversarial, reliable, and secure ML.

The workshop was organized by Eitan Farchi (IBM Research), Onn Shehory (Bar Ilan University), and Guy Barash (Western Digital), who also wrote this report.

## Evaluating Evaluation of AI Systems (W11)

No report from the Evaluating Evaluation of AI Systems workshop organizers was submitted.

## Generalization in Planning (W12)

The goal of the AAAI-20 workshop on Generalization in Planning was to bring planning and learning communities together toward more autonomous systems that can generalize over sequential decision-making problems and domains.

The problem of finding an approach to imitate the human thinking process by coming up with general solutions to problems has been of significant interest to the AI community since its early days. There are two possible interpretations of generalization for sequential decision-making problems: generalization over a set of problems from a given a domain; and generalization across domains where current solutions on particular instances might help on solving tasks in new domains.

Automated planning is a special case of the principle of computing general solutions, where solutions are plans computed independently for each problem and domain. Although this technique can solve problems in a domain-independent way, the solutions do not generalize in the two interpretations just defined. Generalized planning is an extension of automated planning that aims to generalize over a set of problems of a domain, and where solutions are algorithm-like structures instead of plans. Also, other techniques such as transfer and reinforcement learning have been shown to succeed when generalizing across domains for sequential decision problems. This workshop placed a special emphasis on bringing both communities together.

The workshop program included two invited keynote talks. Leslie Pack Kaelbling (Massachusetts Institute of Technology) presented Policies and Value Functions, and Models, Oh My!, where she discussed different strategies for combining planning with policy, model, and value-function learning such that systems can learn incrementally, showing generalization capabilities. Hector Geffner (Universitat Pompeu Fabra and the Catalan Institution for Research and Advanced Studies) presented Representation, Learning, and Planning, where he defended that the right representations to compute general policies in the form of abstractions are easier to be learned than derived.

Then, planners can compute the general policies on the learned representation. He also proposed future challenges, like following a model-based approach to deal with deep reinforcement learning benchmarks such as BabyAI, but by learning and using first-order symbolic representations.

There were also four survey talks. Peter Stone (University of Texas at Austin) presented his work in task and motion planning where general policies are learned in the space of plans. Sheila McIlraith (University of Toronto) showed her recent publications where a new reinforcement learning algorithm exploits automata-based structures named *reward machines*, which can be used to represent reward functions, and deals with partial observability. Giuseppe De Giacomo (Sapienza Università di Roma) presented his work on reactive synthesis and world modeling with some key takeaways such as the model of the world and the task specification should be separated; that we need to be realistic in the size of both specifications; and that planning should be applied for reaching goals instead of synthesis. Finally, George Konidaris (Brown University) presented an adaptive system that consists of a bottom-up learning approach that can learn probabilistic symbols from motor controllers such that planning can be applied at the high-level while policies are applied at the low-level.

In addition, there were sixteen paper presentations. The topics ranged from heuristics and search for generalization, to adaptive systems through model learning, general policies, and general reinforcement learning algorithms. The workshop concluded with a panel where all six invited speakers discussed the role of learning and planning for generalization and the relevance of the representation. Some interesting comments were that trends in AI have swung between symbolic and nonsymbolic, and recently the pendulum has moved too much toward the latter; that general solutions should guarantee some properties such as soundness as a starting point; that symbols in the form of abstractions could boost the learning and planning process for generally adaptive systems; and that we need to connect both the planning and learning communities, because they are looking at similar problems.

This workshop was the fourth in a recurring series. The papers and invited talks are publicly available on the workshop website.<sup>13</sup> The workshop was organized by Javier Segovia-Aguas, Siddharth Srivastava, Raquel Fuentetaja, Aviv Tamar, and Anders Jonsson. This report was written by Javier Segovia-Aguas, Siddharth Srivastava, and Raquel Fuentetaja.

## Health Intelligence (W13)

The AAAI-20 Workshop on Health Intelligence aimed to bring together a wide range of computer scientists, clinical and health informaticians, researchers, students, industry professionals, national and international health and public health agencies, and non-governmental organizations interested in the theory

and practice of computational models of population health intelligence and personalized healthcare to highlight the latest achievements in the field.

Population health intelligence includes a set of activities to extract, capture, and analyze multidimensional socioeconomic, behavioral, environmental, and health data to support decision-making to improve the health of different populations. Advances in AI tools and techniques and internet technologies are dramatically changing the ways that scientists collect data and how people interact with each other, and with their environment. The Internet is also increasingly used to collect, analyze, and monitor health-related reports and activities and to facilitate health-promotion programs and preventive interventions. In addition, to tackle and overcome several issues in personalized healthcare, information technology will need to evolve to improve communication, collaboration, and teamwork among patients, their families, healthcare communities, and care teams involving practitioners from different fields and specialties.

This workshop follows the success of previous health-related AAAI workshops including the ones focused on personalized (2013)<sup>14</sup> and population (2014)<sup>15</sup> healthcare, and the three subsequent joint workshops held at AAAI-17,<sup>16</sup> AAAI-18,<sup>17</sup> and AAAI-19.<sup>18</sup> This year's workshop brought together a wide range of participants (roughly 100 registrants) from the multidisciplinary field of medical and health informatics. Participants were interested in the theory and practice of computational models of web-based public health intelligence as well as personalized healthcare delivery.

The papers (full and short) and the posters presented at the workshop covered a broad range of disciplines within AI, including knowledge representation, ML, NLP, prediction, mobile technology, inference, and dialog systems. From an application perspective, presentations addressed topics in epidemiology, environmental and public health informatics, disease surveillance and diagnosis, medication dosing, health behavior monitoring, and human-computer interaction.

The workshop included an invited talk from Ram Sriram (chief of the Software and Systems Division, The National Institute of Standards and Technology), who gave a presentation on transforming healthcare through AI revolutions. With a total of twenty paper and fifteen poster presentations, the workshop participants engaged in discussions around many cutting-edge topics affecting the way evidence is produced for and delivered in healthcare to improve patient outcomes. To include participants unable to attend the workshop, we successfully incorporated virtual and prerecorded presentations into the program.

Martin Michalowski and Arash Shaban-Nejad served as cochairs of this workshop and authored this report. The workshop papers were published by Springer in their *Studies in Computational Intelligence* series.<sup>19</sup>

## Intelligent Process Automation — Robotic Process Automation Meets AI (W14)

The goal of the AAI-20 Workshop on Intelligent Process Automation was to exchange ideas and foster discussions on the integration of advanced AI techniques with the fast-growing robotic process automation (RPA) software applications. The invited talks and paper presentations centered around the learning of structured and executable processes (programs) from human demonstrations, natural language instructions, or interactions with a software environment.

RPA aims to provide software robots (softbots) that can mimic human users to perform all kinds of mundane and repetitive tasks on their computers, with higher accuracy and speed. Since the term *RPA* was coined by Blue Prism in 2012, it has quickly spread through the business world. RPA is now the fastest-growing category of the enterprise software market (Gartner 2019),<sup>20</sup> and it is expected to reach \$10.7 billion by 2027 (Grand View Research 2020).<sup>21</sup>

AI is promising to take RPA to new heights by not only equipping RPA softbots with various intelligent data processing skills (in computer vision and NLP) but also offering novel approaches to the creation and training of RPA softbots. Although RPA is very popular in the software industry today, it has not yet received sufficient attention from the AI research community. This workshop is, to our knowledge, the first-ever workshop dedicated to RPA in a major AI conference. It successfully brought together researchers working in a variety of fields, most notably process mining, program synthesis, and interactive task learning. The participants came from leading academic institutes (such as Carnegie Mellon, Columbia, Tokyo) and industrial laboratories (such as Microsoft, IBM, Samsung).

The workshop received twenty-nine new submissions and accepted thirteen of them after double-blind peer-reviewing (yielding an acceptance rate of forty-five percent). The eight accepted papers about improving RPA in general with AI were presented orally in three sessions, while the other five each addressing a specific problem in RPA (such as handling invoice forms) were presented as posters. Among the oral paper presentations, one important theme was about the different aspects of automating business processes including discovery, comparison, evaluation, and optimization; another major theme was about the possible ways to build and enhance RPA softbots via natural language conversation. The concrete techniques ranged from symbolic logical reasoning to neural network-based deep learning and reinforcement learning. (In particular, we express our thanks to Yara Rizk from IBM Research, who presented not only her coauthored paper about a unified conversational assistant framework for business process automation but also another paper about the

discovery of business process structure using long short-term memory networks, on behalf of her Chinese colleagues, who were prevented from attending this workshop due to the coronavirus travel restriction.)

The first keynote speech was given by Sumit Gulwani from Microsoft Research, the father of the popular flash-fill feature in Microsoft's Excel software. He led a ground tour of recent developments in program synthesis (which enables nonprogrammers to create programs easily from natural forms of intent expression such as input/output examples), and also illustrated its wide potential applications in RPA including automated data extraction (from various sources such as web pages, text files, or portable document formats), data transformations (over strings, Javascript Object Notation), and data querying, as well as generalization of repetitive actions in traces. This topic was echoed by a paper from the University of Melbourne and the University of Tartu, which described how state-of-the-art program synthesis methods were adapted and optimized to analyze user interaction logs and identify repetitive data transformation routines that could then be automated by RPA. Their contribution spanned across process mining and program synthesis and received the Best Paper Award along with a \$1,000 honorarium.

The second keynote speech was given by Joyce Chai from the University of Michigan, who is well known for her research in recent years on interactive task learning where humans can teach robots new tasks through natural language communication and action demonstration. She laid out the key challenges in grounding language to perception and action as well as grounding communication between humans and robots. The discussion of those research issues continued in another invited talk given by Toby Li from Carnegie Mellon University. He explained PUMICE<sup>22</sup> in detail, Carnegie Mellon University's newly developed end-user programmable softbot based on a multimodal domain-independent approach, which seamlessly combines natural language programming and programming by demonstration. A highlight of PUMICE is that users can define new procedures and concepts by demonstrating and referring to contents within the graphical user interfaces of existing mobile apps.

The runner-up for the Best Paper Award came from the University of Munich, which described tasks for customer support in a real-world industrial setting where the lack of structured training data was mitigated by the utilization of transfer learning.

The RPA company Blue Prism Group PLC demonstrated its commitment to advancing the research in this field by co-organizing the workshop, contributing a coauthored long paper, and fully sponsoring the two keynote speeches, the lunch for workshop participants, and the Best Paper Award.

Dell Zhang (Blue Prism), Andre Freitas (University of Manchester), Dacheng Tao (University of Sydney), and Dawn Song (University of California, Berkeley)



served as cochairs of the workshop. This report was written by Dell Zhang and Jacques Cali. The papers of this workshop have been made available in a non-archival proceedings.<sup>23</sup>

## Interactive and Conversational Recommendation Systems (W15)

Participants at the Interactive and Conversational Recommendation Systems workshop<sup>24</sup> presented novel research and open questions regarding such next-generation recommendation systems and their constituent AI technologies, with a focus on explaining recommendations, improving preference modeling through interaction with users, and training multimodal dialog agents capable of deep domain understanding. A moderated panel discussion at the end of the workshop covered future research directions that participants expected to be explored, and how to sustain the budding research community around this topic.

Recent years have seen an increase in work on interactive and sequential (for example, session-based) interactions with recommender systems. Furthermore, the rise of conversational, AI-based assistants such as Apple's Siri, Amazon's Alexa, and the Google Assistant have all invigorated interest in dialog-driven recommendations, often with a limited degree of personalization. This workshop was a forum to present and discuss novel research directions in interactive and conversational recommender systems, as well as the constituent AI technologies that represent the next generation of recommender systems and personalized conversational assistants. It drew participants from a wide range of interested parties, including researchers from universities and technology companies, and practitioners from consumer retailers, financial services firms, and more.

The first major theme of papers presented was explainable recommendations. Maarten de Rijke (University of Amsterdam) identified two benefits to explainability — for model developers, facilitating error analysis and debugging; and for end users, understanding why items were recommended to them so that they can make decisions to choose between them more effectively. A talk given by Sixun Ouyang (University College Dublin) suggested using an explainer neural network to transform user and item inputs to weights that can both be mapped to a human-interpretable attribute vocabulary, and then used as input to a recommender model of choice, training the explainer and recommender together. Scott Sanner (University of Toronto) presented work that combines deep critiquing (where a model predicts the latent attribute that makes a user prefer an item) as another training objective alongside learning a user representation from item preferences (using variational autoencoders). Finally, Fabio Cozman (University of São Paulo) outlined a way to explain recommendations by finding information relevant

to items in a knowledge graph, then augmenting the graph using entity and relation embeddings to predict missing relation edges between entity vertices. For future work, the panel would like to see explainable recommendations being leveraged for preference elicitation.

Another major theme of the Interactive and Conversational Recommendation Systems workshop was better preference modeling, especially via interaction and conversation with users, to improve the accuracy of recommendations. Craig Boutilier (Google Research) discussed the challenges of eliciting preferences, including the cold-start problem of getting users to initially indicate preferences, the complexity and context-dependence (for example, time, place) of preferences, and the potentially noisy signals of preferences derived from limited feedback to a recommendation. One approach to making better recommendations was presented by Yuheng Bu (Massachusetts Institute of Technology), who used active learning in a collective matrix factorization framework to model how an information-seeking conversational agent would learn to sequentially ask a user the questions eliciting the most signal about their preferences. Another approach came from Alessandro Antonucci (Dalle Molle Institute for Artificial Intelligence Research) — combining Bayesian modeling with conversational preference elicitation to build a recommendation system that improved user engagement for a website pairing performers looking for gigs with venues that have openings. Finally, Nikola Milojkovic (Oracle Labs) gave a talk on a broadly applicable algorithmic improvement: extending multigradient descent to multiobjective recommender systems.

The third major theme was training multimodal dialog agents capable of deep domain understanding. Michelle Zhou (Juji, Inc.) demonstrated her company's work on making chatbots with strong natural language understanding capabilities, such as inferring a user's personality traits (and tailoring responses accordingly) just from their utterances in a conversation with a bot. For training agents capable of multimodal understanding, Noriaki Kawamae (NTT Comware Corporation) presented a system that utilizes recent models for speech to text (convolution neural networks), image classification (very deep convolution network; VGG),<sup>25</sup> and machine translation (seq2seq with attention) to generate combined representations that are usable for recommendation. In contrast to the potential unpredictability of modern neural models, Christian Muise (IBM Research AI) discussed recent advances in nondeterministic AI planners, a classic approach that ensures accountability through predictable bot responses. Zhou Yu (University of California, Davis) concluded the workshop with a talk demonstrating through her group's Amazon Alexa Prize-winning work how dialog systems can be used for recommendations in many different verticals and the societal considerations that come with them.

Following the individual presentations, Scott Sanner moderated a panel discussion among the invited speakers. A key question was about the role of academic versus industry research; the panelists discussed the potential role that user simulators can play in evaluating new methods, and as a way to bridge the data accessibility gap in academe. At the end of the workshop, participants expressed interest in continuing the discussions online (so the public interactive-recommendations-research Google Group was set up to do so), and in reconvening for another in-person workshop in the future.

Scott Sanner (University of Toronto), Tyler Lu (Google Research), Deepak Ramachandran (Google Research), and Joyce Chai (University of Michigan) cochaired the workshop. This report was written by Rohan Dhoopar, Deepak Ramachandran, and Tyler Lu.

## Knowledge Discovery from Unstructured Data in Financial Services (W16)

Knowledge discovery from unstructured data including news, web, and social media has gained the attention of many practitioners in AI research over the past decades. However, its application to data in professional settings such as legal documents and financial filings had been relatively limited. In the financial services industry, vast analysis work requires knowledge discovery from numerous data sources, such as U.S. Securities and Exchange Commission filings, loan documents, and industry reports. This manual knowledge-discovery and extraction process is one of the key bottlenecks for financial services companies in improving their operating productivity. Furthermore, while alternative data such as social media feeds and news are gaining traction as promising new knowledge sources for financial institutions, the valuable knowledge is comingled with immense noise. Thus, the precision and recall requirements for such extracted knowledge to be used in the business process are fastidious.

The goal of the workshop is to bring together academic researchers and industry practitioners to share their insights and ideas on solving knowledge discovery problems for the financial services industry. The workshop involved research paper presentation, keynotes, poster presentations, and an industry panel to foster and enhance the interactions between the participants.

The workshop received thirty manuscripts submitted globally and accepted fourteen of them with a balanced mixture of industry (eight) and academic (six) work. The submissions covered a wide range of critical topics in the finance industry such as knowledge extraction from financial reports, knowledge representation specialized for the financial domain, graph neural network for transactions data, and environmental, social, and governance knowledge and data extraction from sustainability reports. Eight of the accepted papers were orally presented in the workshop and

others were poster presentations. The workshop was extremely well attended, with registration beyond seventy participants across academia and industry.

The workshop also included four keynotes. The first keynote was addressed by Charles Elkan (Goldman Sachs), whose talk focused on the recent advances in ML, especially in deep learning and the limitations of shallow understanding in comparison with deep understanding. He also discussed how ML fits into the landscape of traditional quantitative methods widely used in finance. The second keynote talk came from Dan Roth (University of Pennsylvania), who presented Learning from Incidental Supervision Signals. Roth stated that it is fairly difficult to learn models for natural language understanding and information extraction tasks to support business decisions because signals gathered from these tasks are usually very sparse, and thus the generation of supervision signals is not scalable. His point was particularly relevant to the financial industry, where labeled data are often expensive to obtain. Roth also presented some of his research in identifying and using incidental supervision signals.

The afternoon session featured two keynotes, one from Danqi Chen (Princeton University) and the other from Alfio M. Gliozzo (IBM Research). Chen's talk focused on text question answering. She introduced the use of the Bidirectional Encoder Representations from Transformers model in improving question-answering efficiency by leveraging the span information of the text. She envisioned that the full dense embeddings of passage retrieval and structured graph of text passages could further move forward the state-of-the-art. Gliozzo's keynote, titled Minimally Supervised Knowledge Graph Induction from Text, discussed his research on knowledge induction, which enables customers to build their knowledge graph assets with minimal domain adaptation effort from subject matter experts. Entity types and relations are induced from texts by using distant supervision, transfer learning, and knowledge-base completion and validation. Gliozzo showcased the applications of the technology applied to financial data and tasks.

After the keynotes, an industry panel was held to discuss the AI opportunities and challenges in finance by five industry experts — Sameena Shah (JP Morgan), Charles Elkan (Goldman Sachs), Eric Brown (Capital One), Raghav Madhavan (Alto Meta), and moderator Eren Kurshan (Bank of America).

The Knowledge Discovery from Unstructured Data in Financial Services workshop was co-organized by Xiaomo Liu, Sameena Shah, Manuela M. Veloso, Quanzhi Li, and Le Song. This report was written by Xiaomo Liu, Zhiqiang Ma, and Grace Bang.

## Plan, Activity, and Intent Recognition (W17)

Plan recognition, activity recognition, and intent recognition all involve making inferences about other actors from observations of their behavior, that is,

their interaction with the environment and with each other. The observed actors may be software agents, robots, or humans. This synergistic area of research has become ever-more critical as AI systems take on ever-greater roles within society and we demand that they be explainable. Research in the plan, activity, and intent-recognition community embraces this challenge, combining and unifying techniques from user modeling, machine vision, intelligent user interfaces, human and computer interaction, autonomous and multiagent systems, natural language understanding, and ML.

The Plan, Activity, and Intent Recognition 2020 workshop included three invited talks, sixteen papers, and a demo and poster session that was chaired by Mor Vered from Monash University and Ramon Fraga Pereira from the Pontifical Catholic University of Rio Grande do Sul, Brazil.

During the demo and poster session, attendees of the workshop had an opportunity to see a number of working recognition tools applied to a variety of applications.

The workshop was extremely successful, with many new participants and lively discussions throughout the day. This year, with the objective of extending our community to include researchers from the ML community, the workshop was centered around the relationship between data-driven and model-based approaches to recognition, and the need to bridge the gap between the two approaches. During this workshop, we had the pleasure of hosting three great speakers.

Sheila McIlraith from the University of Toronto discussed how our ability to make sense of other's behavior is informed by our expectations, and how heavily the quality of the conclusions we draw relies on both the observation and the expectation. She also argued for a larger role for actively sensing and manipulating the world to aid our recognition. Hector Geffner from the Catalan Institution for Research and Advanced Studies, and the Universitat Pompeu Fabra, provided a review of prior work on goal recognition using classic planning models, Markov decision processes, partially observable Markov decision processes, and the implicit assumptions in using each of them. Finally, Gal A. Kaminka from Bar Ilan University discussed challenges that are specific to recognition in dynamic and continuous environments. He specifically advocated a *be lazy* approach that saves work by delaying the generation and selection of recognition hypotheses until required. The abstracts and slides of the talks, along with the full schedule and papers, can be found on the workshop website.<sup>26</sup>

Sarah Keren, Reuth Mirsky, and Christopher Geib served as cochairs of the workshop and wrote this report.

## Privacy-Preserving Artificial Intelligence (W18)

The goal of the AI Privacy-Preserving Artificial Intelligence workshop was to provide a platform for

researchers to discuss problems and present solutions related to privacy issues arising within AI applications.

The availability of massive amounts of data, coupled with high-performance cloud computing platforms, has driven significant progress in AI and, in particular, ML and optimization. Indeed, much scientific and technological growth in recent years, including in computer vision, NLP, transportation, and health, has been driven by large-scale data sets that provide a strong basis to improve existing algorithms and develop new ones. However, due to their large-scale and longitudinal collection, archiving these data sets raise significant privacy concerns. They often reveal sensitive personal information that can be exploited, without the knowledge or consent of the involved individuals, for various purposes including monitoring, discrimination, and illegal activities.

The goal of the workshop was to provide a platform for researchers to discuss privacy problems arising when implementing optimization and learning systems and present solutions related to these privacy issues. The workshop brought together researchers from a variety of subfields of AI and security and privacy, including optimization, ML, differential privacy, and multiparty computation.

Two major themes of papers and posters presented at the workshop were the development of techniques that illustrate the brittleness of ML model to privacy attacks, and the development of effective differential privacy protocols for ML and optimization. The workshop included two invited talks on these research themes. A talk given by Catuscia Palamidessi (National Institute for Research in Digital Science and Technology) showed the risks of privacy connected to the use of ML and explored how to use ML to estimate the leakage of private information from black-box models and how to use ML to construct strong privacy-preserving mechanisms that offer good trade-offs between privacy and utility. The talk by Aleksandar Nikolov (University of Toronto) focused on how to use optimization to design optimal differential privacy algorithms to answer important classes of queries.

Another theme looked at how to achieve privacy when data are distributed across a set of agents. The talk of Boi Faltings (the École Polytechnique Fédérale de Lausanne) showed that preferences and constraints involved in the optimization of a multiagent problem may leak private information and presented solutions based on homomorphic encryption and multiparty computation to address such challenges. The papers related to this topic discussed the importance of defining appropriate learning paradigms when data are distributed among multiple agents.

The workshop panel, served by Boi Faltings (the École Polytechnique Fédérale de Lausanne), Antonis Papadimitriou (Duality Technologies), and Helen Toner (Georgetown's Center for Security and Emerging Technology), focused on the theme: Grand Challenges in Privacy 2020: What Are They and What Are

We Missing? The panelists discussed the importance to raise awareness on the privacy risks associated with various computational models, on the pressure that companies are facing to use privacy-preserving technologies, and on the need to focus on protecting information in sophisticated multimedia data, such as images, video, and audio.

Ferdinando Fioretto, Pascal Van Hentenryck, and Rachel Cummings served as cochairs of this workshop. The report was written by Ferdinando Fioretto.

## Reasoning and Learning for Human-Machine Dialogues (W19)

No report from the Reasoning and Learning for Human-Machine Dialogues workshop organizers was submitted.

## Reasoning for Complex Question Answering (W20)

The Reasoning for Complex Question Answering 2020 workshop was held to a packed audience at the Hilton Midtown in New York City on February 8, 2020. The goal of the workshop was to bring together research and researchers from the fields of reasoning, ML, and NLP as pertains to the problem of question answering in natural language systems and techniques.

The workshop followed up on a very successful inaugural workshop in 2019 at AAI-19<sup>27</sup> in Hawaii, and brought together the top researchers from the fields of AI, ML, and NLP once again. The 2020 workshop was designed around six invited talks by leading researchers in the field, along with oral and spotlight paper presentations and two poster sessions for deeper discussions among participants.

The workshop was kicked off by a very engaging invited talk by Ray Mooney, who continued to stand behind his famous saying — “you can’t cram the meaning of a whole sentence into a #\$\$# vector.” However, he clarified that it was beneficial to combine distribution representations with symbolic reasoning to answer complex questions. This was quite similar to the point made by Nobel laureate Daniel Kahneman — that humans are capable of thinking fast (pattern matching, as in distributional representations) and thinking slow (symbolic reasoning and inference). Such approaches have been applied successfully to relation extraction and work on compositionality in neural networks by Jacob Andreas. Yet another interesting approach outlined by Mooney was plan-based understanding, which provides for deeper comprehension of narrative text that supports answering *why* questions. This was introduced in the plan applier mechanism (PAM) system in the 1980s, but we need new statistical learning and probabilistic inference methods to make this process robust.

Nasrin Mostafazadeh talked about the two bottlenecks in AI research: implicit knowledge acquisition,

and incorporating common-sense knowledge into downstream applications. She introduced a crowdsourcing project for collecting multiple knowledge dimensions (cause, effect) for sentences in a narrative story.

Bishan Yang talked about incorporating user feedback into the steps performed by question answering models: understanding the question, gathering relevant information, and synthesizing them to produce a coherent answer.

Dan Roth discussed several shortcomings of current question-answering approaches, and suggested challenges that might require symbolic reasoning to address. One such example that was presented was answering questions for which answers are not in the text, but may be inferred from it. A question like *Did Aristotle have a laptop?* should be answered *no*, based on background knowledge of his death year and the time the laptop was invented. Even in the easy cases that question-answering systems solve today, reasoning is not often used. Moreover, current systems manage to answer the question correctly without the imperative context, but are easily confused with the right choice of distractors.

Robyn Speer talked about ConceptNet,<sup>28</sup> a knowledge graph that is widely used in AI and NLP systems to complement distributional representations and language models, which often fail to state the obvious. Speer discussed the various ways in which ConceptNet is currently integrated with ML systems, and suggested best practices for using it.

Sameer Singh discussed the shortcomings and lack of robustness in question-answering systems, including sensitivity to changes in the question or context where the answer shouldn’t change; higher error rate on questions that involve complex phenomena such as temporal inference, negation, and coreference; inconsistency in the answer for semantically co-reference questions (for example, *which type* and *what type*); and no global consistency with respect to inferred questions (for example, answering the question *How many birds are in the picture?* with a number larger than 1 requires answering *Are there birds in the photo?* with *yes*).

In addition to the invited speaker program, the workshop featured nine accepted papers. Two of these were selected for presentation as twenty-minute oral talks, while the other seven were presented as shorter ten-minute spotlights. The workshop also featured the unveiling of two new datasets and leaderboards for the community: the open-reading bookmark, or ORB,<sup>29</sup> dataset from the Allen Institute for AI (AI2), and the TechQA<sup>30</sup> dataset from IBM.

The audience was united in their feedback that they found the invited talk program at the workshop quite engaging, and that they wished to see more such cross-pollination efforts across multiple mainstream communities. The invited talks summarized in this report — as well as copies of all the accepted papers — are available via the workshop website.<sup>31</sup>

Kartik Talamadupula, Vered Shwartz, Jay Pujara, Rachel Rudinger, Mausam, Nanyun Peng, and Pavan Kapanipathi served as co-organizers for this workshop. The report was written by Kartik Talamadupula and Vered Shwartz.

## Reinforcement Learning in Games (W21)

No report from the Reinforcement Learning in Games workshop organizers was submitted.

## Reproducibility in AI — Future Direction and Reproducibility Challenge (W22)

Even research that is completely conducted on a computer is not necessarily reproducible. Variations in hardware, operating systems, and compiler settings can affect the reproducibility of experiments conducted fully on computers. AI research is not spared. The same problems haunt AI research, but as the field is young, there are other problems too. Some of these are related to research methodology, such as methods of evaluation and performance metrics, while others relate to the stochasticity of both models and the environment they are applied to. The second AAAI Workshop on Reproducible AI is one among many proofs that these problems have been acknowledged recently by the AI research community.

Yolanda Gil, professor at the University of Southern California and President of AAAI, gave the first keynote of the day. She presented her longtime work on open and transparent research with a focus on making science reproducible. She advocated for the scientific paper of the future where not only data and code are shared in a structured manner, but also the workflows of the experiments. Yan Liu, assistant professor at the University of California Santa Cruz, presented the paper Replication Markets: Results, Lessons, Challenges and Opportunities in AI Replication, in which it is suggested that replication markets could be used to reduce the shortage of good peer reviewers. Instead of doing peer reviews, the wisdom of the crowd could be used to make quick predictions on whether claims are reproducible. Michael Gordon, a PhD student at Massey University, presented the paper Replication Markets in the Social and Behavioral Sciences, and gave further insights into replication markets and their effectiveness on predicting replicability.

Jakub Kowalski, assistant professor at the University of Wroclaw, presented Experimental Studies in General Game Playing: An Experience Report, where different game description languages were analyzed and discussed. Several recommendations were given. Among them were that the exact game definitions must be provided to ensure that game descriptions in different languages that are compared actually *are* the same. Sweitze Roffel and George Tsatsaronis from

Elsevier presented the paper Toward Reproducible Artificial Intelligence: The Roles of Researchers and Publishers, where they discussed the efforts made by Elsevier to support reproducibility along the three axes *Methods*, *Data*, and *Experiments*. Edward Raff presented his excellent 2019 Neural Information Processing Systems conference paper A Step Toward Quantifying Independently Reproducible Machine Learning Research<sup>32</sup> and some new results that are yet to be published.

Joelle Pineau, associate professor at McGill University, co-managing Director of Facebook AI Research, and the author of the *Machine Learning Reproducibility Checklist*,<sup>33</sup> gave the second keynote on Machine Learning Reproducibility: Lessons Learned from the NeurIPS Reproducibility Program. One of the very impressive results was the increase in papers sharing code after introducing the Machine Learning Reproducibility Checklist in 2018. For Neural Information Processing Systems conference 2018, just below fifty percent of the papers shared code. The numbers increased to sixty-seven percent for the International Conference on Machine Learning 2019 and seventy-five percent for the Neural Information Processing Systems conference 2019. This effect was gained only from introducing the checklist and a code submission policy. Following these was not mandatory. Finally, Vani Mandava from Microsoft Research presented the Microsoft Research Open Data Project, which is an initiative that combines features of a traditional data repository with easy access to compute resources. The main aim is to increase reproducibility of research outcomes by making datasets associated with research papers published by Microsoft researchers available broadly.

The workshop was organized by the cochairs Odd Erik Gundersen, Daniel Garijo, and David Aha. This report was written by Odd Erik Gundersen.

## Statistical Relational AI (W23)

The primary purpose of the series of workshops on Statistical Relational AI (StarAI)<sup>34</sup> is to provide a bridge toward the holy grail of AI — integrating learning and reasoning aspects of AI.

This workshop brings together researchers and practitioners from three fields: logical (or relational) AI/learning; probabilistic (or statistical) AI/learning; and neural approaches for learning/AI with knowledge graphs and other structured data. These fields share many key features and often solve similar problems and tasks. Until recently, however, research in them has progressed independently with little or no interaction. The fields often use different terminology for the same concepts and, as a result, keeping up and understanding the results in the other field(s) is cumbersome, thus impeding research. One of the growing applications at this confluence of relational AI and statistical approaches is the widespread use of embedding and representation learning that is cognizant of relational structure. Our long-term goal

is to provide a forum for removing the barriers across subfields and promoting the synergy among logical, statistical, and neural AI, with the StarAI workshops serving as stepping stones toward realizing this big-picture view on AI.

The workshop is currently provoking much new research and has tremendous theoretical and practical implications. The ongoing research mainly follows two directions. First, a substantial effort is invested into the long-standing dream of AI, combining logic and probability in a unified representation with solid theoretical foundations and practical tools. Successfully building general-purpose reasoning and learning tools in this context will enable new applications in several large, complex real-world domains including those involving Big data, heterogeneous data, and varied amounts of prior knowledge. Such domains are often characterized by rich relational structures and large amounts of uncertainty. Logic helps the application designer effectively handle the former while probability helps her effectively manage the latter. Second, an explosion of work on neural architectures for relational learning has added a new perspective on the logical and probabilistic hybrids. These approaches tackle problems that typically require logical or probabilistic reasoning, including knowledge graph completion and (visual) question answering.

The StarAI workshop series has traditionally been a fertile ground for establishing connections among different subfields of AI. This ninth meeting, which was the largest workshop to date, again brought together researchers that drive forward different AI subfields by using statistical relational techniques. Invited talks by Guy Van den Broeck (University of California Los Angeles), Lise Getoor (University of California, Santa Cruz), and Yejin Choi (University of Washington and Allen Institute for Artificial Intelligence) provided a synthesis of various active research directions in StarAI, which include querying complex probabilistic models that might not have crisp symbolic information; an overview of a decade-long progress in probabilistic soft logic; and the recent interest in common-sense reasoning with neural approaches and newly developed datasets. In two lively poster sessions, forty-two full technical papers and four short position papers and abstracts were presented. The selected papers covered a wide range of topics in StarAI, such as structure learning, lifted inference, probabilistic programming, deep relational models, the unification of neural and symbolic representations, (weighted) model counting, and constraint optimization, among others, and clearly show the promise of StarAI. The main themes of the workshop were novel prospects in logical and probabilistic reasoning inspired by deep learning techniques, lifted inference techniques for scalable inference for complex probabilistic models, and novel applications of StarAI techniques.

Sebastijan Dumancic, Angelika Kimmig, David Poole, and Jay Pujara served as cochairs of this workshop. This report was written by Sebastijan Dumancic.

## Notes

1. [aaai.org/Conferences/AAAI-20](http://aaai.org/Conferences/AAAI-20).
2. [CEUR.WS.org](http://CEUR.WS.org).
3. FC Barcelona award, <https://barcainnovationhub.com/event/barca-sports-analytics-summit-2019/>.
4. [aaai.org/Conferences/AAAI-20](http://aaai.org/Conferences/AAAI-20).
5. [www.microsoft.com/en-us/research/project/farmbeats-iot-agriculture/](http://www.microsoft.com/en-us/research/project/farmbeats-iot-agriculture/).
6. [Aiotworkshop.github.io](https://aiotworkshop.github.io).
7. [www.ai-safety.org](http://www.ai-safety.org).
8. [ceur-ws.org/Vol-2560/](http://ceur-ws.org/Vol-2560/).
9. <https://resourcecentral.org/>.
10. [cloudintelligenceworkshop.org](https://cloudintelligenceworkshop.org).
11. [www.noessolutions.com/](http://www.noessolutions.com/).
12. [sites.google.com/dstc.community/dstc8/home](https://sites.google.com/dstc.community/dstc8/home).
13. [aaai.org/Conferences/AAAI-20](http://aaai.org/Conferences/AAAI-20).
14. [www.aaai.org/Workshops/ws13.php](http://www.aaai.org/Workshops/ws13.php).
15. [www.aaai.org/Workshops/ws14.php](http://www.aaai.org/Workshops/ws14.php).
16. [www.aaai.org/Workshops/ws17.php](http://www.aaai.org/Workshops/ws17.php).
17. [aaai.org/Conferences/AAAI-18/ws18/](http://aaai.org/Conferences/AAAI-18/ws18/).
18. [aaai.org/Conferences/AAAI-19/ws19/](http://aaai.org/Conferences/AAAI-19/ws19/).
19. Springer 2020, <https://www.springer.com/series/7092>.
20. Gartner 2019, <https://www.gartner.com/en/newsroom/press-releases/2019-06-24-gartner-says-worldwide-robotic-process-automation-sof>.
21. Grand View Research 2020, <https://finance.yahoo.com/news/robotic-process-automation-market-size-093500133.html>.
22. [www.cs.cmu.edu/~NatProg/papers/p577-li-Pumice-UIST.pdf](http://www.cs.cmu.edu/~NatProg/papers/p577-li-Pumice-UIST.pdf).
23. [arXiv:2001.05214](https://arxiv.org/abs/2001.05214).
24. [sites.google.com/view/wicrs2020](https://sites.google.com/view/wicrs2020).
25. [www.robots.ox.ac.uk/~vgg/research/very\\_deep/](http://www.robots.ox.ac.uk/~vgg/research/very_deep/).
26. [aaai.org/Conferences/AAAI-20/ws20workshops/](http://aaai.org/Conferences/AAAI-20/ws20workshops/).
27. [aaai.org/Conferences/AAAI-19/ws19/](http://aaai.org/Conferences/AAAI-19/ws19/).
28. [conceptnet.io/](https://conceptnet.io/).
29. [leaderboard.allenai.org/orb/submissions/about](https://leaderboard.allenai.org/orb/submissions/about).
30. [leaderboard.techqa.us-east.containers.appdomain.cloud/](https://leaderboard.techqa.us-east.containers.appdomain.cloud/).
31. [rcqa-ws.github.io](https://rcqa-ws.github.io).
32. [papers.nips.cc/paper/8787-a-step-toward-quantifying-independently-reproducible-machine-learning-research.pdf](https://papers.nips.cc/paper/8787-a-step-toward-quantifying-independently-reproducible-machine-learning-research.pdf).
33. [www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf](http://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf).
34. [www.starai.org](http://www.starai.org).

**Grace Bang** is a data scientist at S&P Global.

**Guy Barash** is at Western Digital.

**Ryan Beal** ([ryan.beal@soton.ac.uk](mailto:ryan.beal@soton.ac.uk)) is a PhD research student at the University of Southampton in the Agents, Interaction and Complexity research group.

**Jacques Cali** is the director of Research at Blue Prism Group PLC.

**Mauricio Castillo-Effen** is a senior researcher and the technical leader of the Trustworthy Autonomous Systems focus area at Lockheed Martin's Advanced Technology Laboratories in Arlington, VA, US.

**Xin Cynthia Chen** is a researcher on AI safety at the University of Hong Kong, China.

**Niyati Chhaya** is at Adobe Research.

**Rachel Cummings** is an assistant professor at the School of Industrial and System Engineering, Georgia Institute of Technology.

**Rohan Dhoopar** is a software engineer at Google Research.

**Sebastijan Dumancic** is a post-doctoral researcher at Katholieke Universiteit Leuven, Belgium.

**Huáscar Espinoza** is principal researcher of AI Runtime Safety Monitoring and Enforcement at the Commissariat à l'Énergie Atomique, France.

**Eitan Farchi** is at IBM Research.

**Ferdinando Fioretto** is an assistant professor at the Department of Electrical Engineering and Computer Science, Syracuse University.

**Raquel Fuentetaja** is a lecturer in the Planning and Learning research group at Universidad Carlos III de Madrid.

**Michel Galley** is a principal researcher at Microsoft Research AI.

**Christopher Geib** is a principal researcher at SIFT LLC.

**Odd Erik Gundersen** is an associate professor at the Norwegian University of Science and Technology, Norway.

**José Hernández-Orallo** is a professor at the Universitat Politècnica de València, Spain and an associate fellow at the Leverhulme Centre for the Future of Intelligence, UK.

**Xiaowei Huang** is a lecturer at the Department of Computer Science, University of Liverpool, UK.

**Kokil Jaidka** is at the National University of Singapore.

**Sarah Keren** is a post-doctoral fellow at the Harvard School of Engineering and Applied Sciences in the fields of Multi-agent Environment Design and Goal Recognition Design.

**Seokhwan Kim** is a senior ML scientist at Amazon Alexa AI.

**Xiaomo Liu** is a director of Data Science at S&P Global.

**Tyler Lu** is a research scientist at Google Research.

**Zhiqiang Ma** is a senior data scientist at S&P Global.

**Richard Mallah** is director of AI Projects at the Future of Life Institute, US.

**John McDermid** is a professor of Software Engineering at the University of York and director of the Assuring Autonomy International Program, UK.

**Martin Michalowski** is an assistant professor at the University of Minnesota School of Nursing Minneapolis, MN, US.

**Reuth Mirsky** is a post-doctoral fellow at the University of Texas at Austin, Austin, TX, US.

**Seán Ó Héigeartaigh** is the executive director of Cambridge University's Centre for the Study of Existential Risk, and a program director at the Leverhulme Centre for the Future of Intelligence, UK.

**Deepak Ramachandran** is a staff software engineer at Google Research.

**Javier Segovia-Aguas** is a post-doctoral researcher in the Perception and Manipulation group at Institut de Robòtica



*The New York Hilton Midtown.*

i Informàtica Industrial, Spanish Council for Scientific Research, Technical University of Catalonia.

**Arash Shaban-Nejad** is an assistant professor in the Oak Ridge National Laboratory, Center for Biomedical Informatics at the Department of Pediatrics, the University of Tennessee Health Science Center.

**Onn Shehory** is at Bar Ilan University.

**Vered Shwartz** is a post-doctoral researcher at the Allen Institute for AI (AI2) and the University of Washington, Seattle, WA, US.

**Siddharth Srivastava** is an assistant professor in the Autonomous Agents and Intelligent Robots lab of the School of Computing, Informatics, and Decision Systems at Arizona State University.

**Kartik Talamadupula** is a research staff member at IBM Research.

**Jian Tang** is the chief scientist at DiDi AI Labs.

**Pascal Van Hentenryck** is a professor at the School of Industrial and System Engineering, Georgia Institute of Technology.

**Dell Zhang** is a staff research scientist at Blue Prism AI Labs and a reader (on leave) at Birkbeck, University of London.

**Jian Zhang** is a program manager at Microsoft Azure.