AITF ANNUAL REPORT 2016

DR. RICHARD SUTTON

REINFORCEMENT LEARNING AND ARTIFICIAL INTELLIGENCE

AITF ANNUAL REPORT MARCH 31, 2016

1. EXECUTIVE SUMMARY

The RLAI research program pursues an approach to artificial intelligence and engineering problems in which they are formulated as large optimal-control problems and approximately solved by reinforcement-learning methods. Reinforcement learning is a new body of theory and techniques for optimal control that has been developed in the last thirty years primarily within the machine learning and operations research communities, and which has separately become important in psychology and neuroscience. Reinforcement learning researchers have developed novel methods to approximate solutions to optimal-control problems that are too large or too ill-defined for classical solution methods such as dynamic programming. For example, reinforcement-learning methods have obtained the best known solutions in such diverse automation applications as helicopter flying, elevator deployment, playing backgammon, and resource-constrained scheduling.

The objectives of the RLAI research program are to create new methods for reinforcement learning that remove some of the limitations on its widespread application and to develop reinforcement learning as a model of intelligence that could approach human abilities. These objectives are pursued through mathematical analysis, through computational experiments, through the development of robotic systems, and through the development and testing of computational models of natural learning processes.

The overall research team consists of about 73 members, 47 of whom are graduate students and, of those, 8 of which are recipients of major scholarships. The output of the research program remained strong, with 32 papers published or accepted for publication in refereed archival venues during the reporting period. Thirteen students were graduated this year (5 PhD and 9 MSc), more than in any prior year of the project.

Highlights of the research program this year include: 1) hosting a major multi-disciplinary conference attended by 200 researchers from around the world, 2) publishing a paper in *Science* on our essentially-perfect computer player of the most popular form of Poker, 3) developing new *n-step-delayed* algorithms for efficient multi-step reinforcement learning without eligibility traces, and 4) analyzing eligibility traces as a general mechanism enabling online prediction-learning algorithms whose computational complexity scales *independently of the span* of the predictions.

2. RESEARCH PROGRAM OVERVIEW

The most important development within the project this year was the hosting of a major conference attended by 200 researchers from around the world. The *Second Multidisciplinary Conference on Reinforcement Learning and Decision Making* was held at the University of Alberta on June 7-10, 2015, in the new Centennial Center for Interdisciplinary Science. It featured 36 speakers from diverse fields such as neuroscience, psychology, control theory, economics, operations research, medicine, and of course computer science and artificial intelligence. This series of meetings is special in bringing together people from different

disciplines all focused on the same topic—the learning challenges involved in decision making over time. Four tutorial sessions were held to overcome the disciplinary barriers. Over 120 papers were presented in the proceedings and poster sessions. Videos of all the talks are available at http://videolectures.net/rldm. About 30% of the conference's $150K budget was provided from RLAI funds, with the rest coming from registration fees and donations from Lockheed Martin, Google Deepmind, Osaro, and Microsoft Research. The weather cooperated, Edmonton and the campus were beautiful, and everybody seemed to have a good time and feel that the meeting was successful. On the last day, after the meeting was formally over, we gave informal tours of the RLAI labs in computing science and rehabilitative medicine to a subset of the conference participants. The conference made many more people aware of the RLAI research program in Alberta.

The past year was a big one for artificial intelligence and machine learning, with three landmark successes in the last 14 months, all with a significant connection to the RLAI project. It began in February 2015 when researchers from Google Deepmind published a paper in *Nature* announcing the attainment of human-level performance in computer playing of Atari video games, without game-specific knowledge, through a combination of reinforcement learning and deep learning. Four of the authors of the paper were alumni of the RLAI project, including two of the three main authors (Vlodymyr Mnih and David Silver). In addition, the Atari 2600 domain they used was developed within the project (see the 2011 annual report). The second landmark success was in May 2015 when RLAI PI Michael Bowling and his students published a paper in *Science* announcing the attainment of essentially perfect computer play in the most common form of two-person Poker (as covered in last year's annual report). The third landmark success was in January 2016 when Google Deepmind published another *Nature* paper announcing a large jump in performance of computer play of the ancient Chinese game of Go, and followed it up in March by convincingly defeating the world-champion human player 4-1 in a five-game match. Both of the two main authors of this work trained at the University of Alberta: David Silver was an RLAI graduate student and Aja Huang was a postdoctoral fellow with RLAI-associated faculty member Martin Mueller (see the 2008-10 annual reports). Again the Google success was due to a combination of reinforcement learning and deep learning, which has come to be known as *deep reinforcement learning*.

These successes, the prominence of their publication, and similar landmark successes with deep learning alone have created intense interest within the field. For example, attendance at the largest scientific meetings has more than doubled in the last two years. Industry interest is also way up. This year the RLAI project has developed substantial contacts with Samsung, Mitsubishi, Maluuba, and startups Osaro and Cogitai, and Google Deepmind donated $250K to support project research. Artificial intelligence has also come to be discussed more in the mainstream and business media. Several prominent people including Elon Musk, Stephen Hawking, and Bill Gates have expressed concerns about possible dangers should artificial intelligence succeed. RLAI personnel have been among those scientists trying to place both these successes and concerns within a larger context for the public.

3. RESEARCH PROJECTS

The overall RLAI project comprises many component research projects in three main areas: 1) designing improved reinforcement learning algorithms, 2) extending reinforcement learning to address the more ambitious goals of artificial intelligence, and 3) exploring applications of reinforcement learning algorithms. This section describes in more detail a few of these component research projects.

Learning to Predict Independent of Span

This year we have developed a new way of looking at one of the most common mechanistic components of reinforcement learning algorithms—the *eligibility trace*. Eligibility traces are fading records of stimuli or actions that are used to assign credit to them if there is a subsequent reward or temporal-difference error. In recent years we have become skilled at designing special eligibility traces, which we call *dutch traces*, for which one can establish equivalences to desirable forward-view algorithms. These mechanisms were seen as important for obtaining computationally inexpensive implementations, but the exact nature of the benefit and what it relied on was not clear.

In particular, it was generally believed that eligibility traces were a mechanism whose utility was in some way specific to temporal-difference (TD) learning (the key learning idea underlying most online reinforcement learning algorithms). As a result of our new analysis, we can clearly see that eligibility traces in general, and their new dutch forms in particular, have nothing to do with TD learning. They arise with the same utility even in classical Monte Carlo algorithms such as the least-mean-square (LMS) algorithm familiar from signal processing engineering. Eligibility traces and their computational advantages arise whenever there is a significant delay between the time a prediction is made and the time when the actual outcome becomes known, which we call the *span* of the prediction.

More precisely, the opportunity for productive use of eligibility traces arises whenever a sequence of predictions is made about a related event. The span is the largest number of predictions in the sequence elapsing between a prediction and the event predicted. For example, if on each day we predict what a stock market index will be at the end of the month, then the span is 30, whereas if we predict at each hour what the stock market index will be, then the span is $30 \times 24 = 720$. Clearly, if we update our predictions frequently, then the span can become quite large. Conventional LMS methods must keep an individual record of the features present at each time that a prediction was made; their memory requirements increase linearly with the span. Eligibility traces enable such algorithms to be implemented with per-time-step complexity that is independent of span. For LMS, the eligibility trace that achieves exactly the same update is the dutch trace. This is a new mechanism for the LMS update with potentially many applications of economic import. We can similarly derive a dutch-like eligibility trace for a many of other classical learning algorithms (but interestingly not all) that achieves them with independent of span computational complexity.

K-step Delayed Learning Algorithms

This year has brought renewed attention, and several new algorithms, for reinforcement learning methods based on 'forward-view' approximations that do not use eligibility traces. Many reinforcement learning algorithms can be viewed in both a *forward* way, in which one looks ahead to future rewards and predictions, and a *backward* way, in which one looks back to previous states in which predictions were made. Typically algorithms are analyzed first in the forward view, which is more straightforward and easier to understand, and then transformed into a backward view that is easier and computationally cheaper to implement even though its workings are more opaque. It is the backward view that uses eligibility traces and that we celebrated in the work described in the research project described above. However, not all forward-view algorithms of interest can be converted to exact backward-view implementations. In particular, it appears that this conversion can be done only to a limited extent for *nonlinear* learning systems such as multi-layer neural networks, also known as *deep learning*. This motivates looking at forward-view algorithms even though they are not independent of span.

Although forward-view algorithms are not independent of span, this computational weakness can be limited by using TD learning. In TD learning, the important sense of span is not that between a prediction and the outcome, but that between a prediction and the later prediction that it learns from. This special feature of TD learning—that it involves learning estimates from later estimates prior to knowing the actual outcome—is called *bootstrapping*. Ideal forward-view reinforcement learning algorithms like forward nonlinear Sarsa(λ), in which the span of bootstrapping is technically infinite, can be closely approximated by limited-span algorithms that delay their updates by the span limit. Let K be the limit on the span of bootstrapping. Then the K-step-delayed version of forward nonlinear Sarsa(λ) can be implemented with a delay of K steps and with per-step memory proportional to K. As K is increased, the approximation to exact forward nonlinear Sarsa(λ) becomes closer. In many cases the K-step delayed approximation is substantially better overall than the approximation of the corresponding backward-view algorithm, as shown in the figure below.

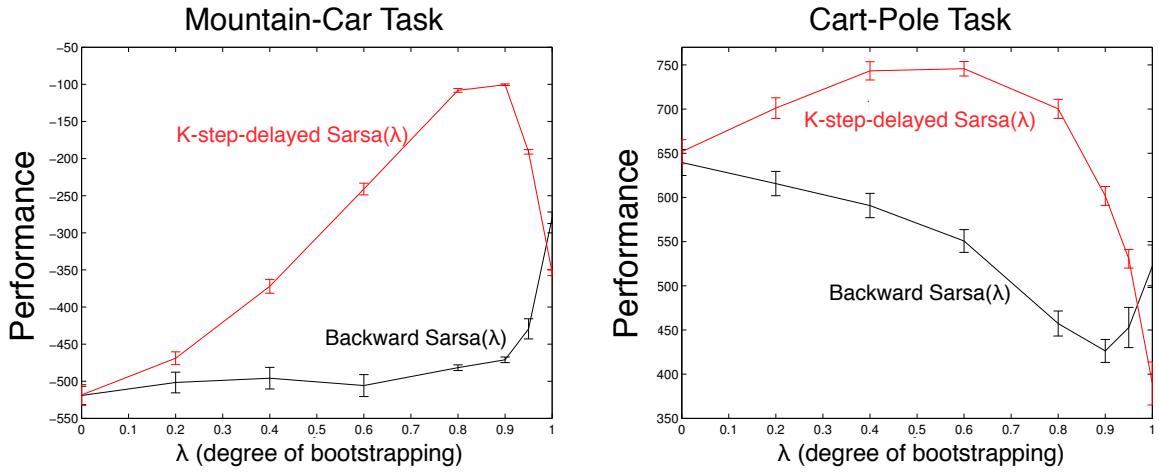

Figure 1. Improved performance of a K-step-delayed algorithm over the corresponding backward-view algorithm on two classic reinforcement learning tasks.

Horde on the Modular Prosthetic Limb

The RLAI project is pursuing a novel approach to knowledge representation based on the notion of value functions and on other ideas and algorithms from reinforcement learning. In our approach, knowledge is represented as a large number of approximate generalized value functions learned in parallel, each with its own policy, pseudo-reward function, and pseudo-termination function. Our architecture, called *Horde*, consists of a large number of independent reinforcement learning agents, which we call *demons*. The approach here is similar to that which we have taken in previous years with TD networks and options. Horde differs from TD networks in its more straightforward handling of state and function approximation (no predictive state representations) and in its use of GTD algorithms for off-policy learning, which are considerably more efficient than those used in prior work with TD networks. In past years we have deployed Horde on small mobile robots to learn a variety of predictions and behaviours off-policy and in real-time.

This year we have extended the Horde architecture to apply to an advanced prosthetic arm. The Modular Prosthetic Limb (MPL v3), purchased last year for $700,000 using CFI funds, is one of only two in the world and is the flagship of our research effort in adaptive prosthetics. We have demonstrated learning of 18,000 predictions on the MPL, and have shown their real-time adaptation in contact with a changing environment, including changes due to interacting with people.

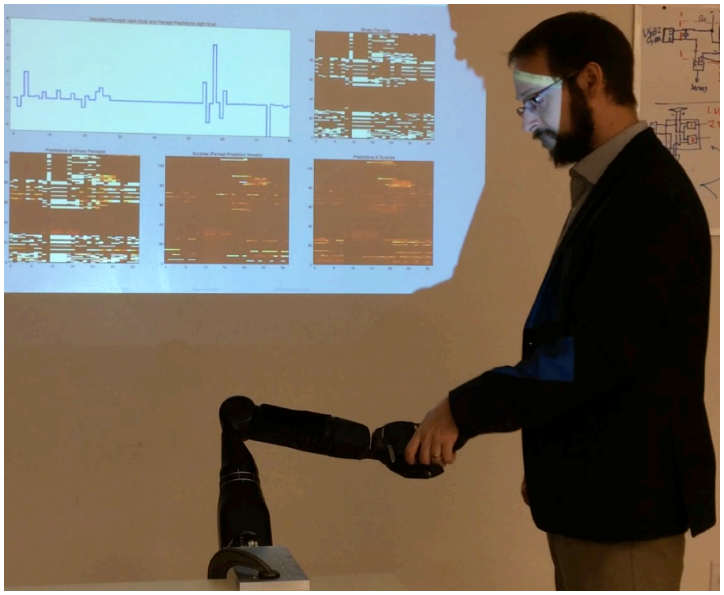

Figure 2. PI Patrick Pilarski shaking hands with the Modular Prosthetic Limb while it generates and learns 18,000 predictions, a subset of which are displayed in real-time.

4. OBJECTIVES FOR NEXT YEAR

With regard to core reinforcement learning algorithms, we envision taking several further steps over the next year. Much of this work is organized around producing a second edition of the textbook on reinforcement learning that Sutton authored together with Prof. Andrew Barto of the University of Massachusetts. The second edition will include K-step delayed forward algorithms, tree-backup algorithms, our new emphatic-TD and gradient-TD algorithms, and policy-gradient methods, as well as treatments of the relationships to psychology and neuroscience. The new edition will be finished and published next year, then published by MIT Press and on the Internet.

With regard to the larger ambitions of artificial intelligence, we plan another round of empirical system building with robot microworlds. In a simplified robot environment, we intend to explore how RLAI innovations like general value functions, nexting, and the Horde architecture can be used to learn substantial knowledge of the world that is both grounded in sensorimotor data and significantly abstract. Such knowledge can then form a model of the world suitable for reinforcement-learning based planning methods.

With regard to the application of the Horde architecture to adaptive prosthetics and the MPL, we plan to extend current efforts to utilize advanced hardware, notably GPUs, to scale the Horde architecture significantly farther. We expect to be able to increase the scaling by at least one order of magnitude.

5. RESEARCH TEAM MEMBERS AND CONTRIBUTIONS

a. Team Leader

Name	Role	Awards / Special Info
Richard Sutton	Principal Investigator	AICML, NSERC, Google Deepmind

b. Faculty Team Members

Michael Bowling	Faculty team member	AICML, NSERC
Dale Schuurmans	Faculty team member	AICML, NSERC, MITACS
Csaba Szepesvári	Faculty team member	AICML, NSERC
Patrick Pilarski	Faculty team member	AICML, NSERC
András György	Adjunct Faculty Member; Research	Now a Senior Lecturer at Imperial College London, Dept of Electrical and Electronic Engineering
Martin Müller	Associated faculty member	NSERC

c. Post Doctoral Fellows & Research Associates

Name	Role	Awards / Special Info
Joseph Modayil	Research Associate	Now a Research Scientist at Google/Deep Mind London, UK
Huizhen Yu	Research Associate	
Harm van Seijen	Post doctoral fellow	
Tor Lattimore	Post doctoral fellow	
Karim Abou-Moustafa	Post doctoral fellow	Now a financial machine learning specialist at AIMCo
Jae Young Lee	Post doctoral fellow	
Villiam Lisy	Post doctoral fellow	
Omid Namaki Shoushtari	Post doctoral fellow	

d. PhD Candidates

Name	Role	Scholarships / Awards / Special Info
Gabor Balazs	PhD candidate	
Nolan Bard	PhD candidate	Graduated. Now a postdoctoral fellow at the University of Alberta
Neil Burch	PhD candidate	Queen Elizabeth II Graduate Scholarship, Myer Horowitz Graduate Student's Association Scholarship
Katherine Chen	PhD candidate	
Kenneth Dwyer	PhD candidate	Now a research scientist at Two Hat Security, Kelowna, BC
Pooria Joulani	PhD candidate	
Ruitong Huang	PhD candidate	
Michael Johanson	PhD candidate	Graduated. Now a consultant in the Edmonton area
Anna Koop	PhD candidate	
Ashique Mahmood	PhD candidate	
Kory Mathewson	PhD candidate	NSERC, AITF, Walter H. Johns Graduate Fellowship, GRA Rice Graduate Scholarship in Communications
Ozlem Aslan	PhD candidate	
Bernardo Ávila Pires	PhD candidate	Alberta Innovates Graduate Student Scholarship
James Neufeld	PhD candidate	Graduated. Now Data Scientist at Twitter
Roshan Shariff	PhD candidate	
Craig Sherstan	PhD candidate	Alberta Innovates Graduate Student Scholarship, Vanier Scholarship, AITF
Sina Ghiassian	PhD candidate	
Adam White	PhD candidate	Now an Assistant Scientist at Indiana University
Jincheng Mei	PhD candidate	
Farzaneh Mirzazadeh	PhD candidate	Queen Elizabeth II Scholarship
Hengshuai Yao	PhD candidate	
Mohammad Ajallooeian	PhD candidate	

Junfeng Wen	PhD candidate	
Marlos Machado	PhD candidate	AITF Graduate Student Scholarship
Leah Hackman	PhD candidate	
Trevor David	PhD candidate	University of Alberta Doctoral Recruitment Scholarship
Dustin Morrill	PhD candidate	
Mahdi Karami	PhD candidate	
Kiarash Shaloudegi	PhD candidate	

e. MSc Candidates

Name	**Role**	**Scholarships / Awards / Special Info**
Trevor Davis	MSc candidate	Graduated. Now a PhD student at the University of Alberta
Kavosh Asadi Atui	MSc candidate	Now a PhD student at Brown University
Bing Xu	MSc candidate	Graduated. Now a PhD student at Univesité de Montreal
Robert Post	MSc candidate	
Roshan Shariff	MSc candidate	University of Alberta Doctoral Recruitment Scholarship. Graduated. Now a PhD student at the University of Alberta
Shun Jie Lau	MSc candidate	
Ann Edwards	MSc candidate	
Craig Sherstan	MSc candidate	Graduated. Now a PhD student at the University of Alberta
Brendan Bennett	MSc candidate	
Tian Tian	MSc candidate	Queen Elizabeth Scholarship II
Yifan Wu	MSc candidate	
Min Yang	MSc candidate (course-based)	Graduated. Now a consultant at CGI, Edmonton

Xiaowei Hu	MSc candidate	Dean's Excellence Award (MSc)
Tim Yee	MSc candidate	
Dustin Morril	MSc candidate	AITF, NSERC CGSM, Walter Johns Scholarship, NSERC PGSD. Graduated. Now a PhD student at the University of Alberta
Jaden Travnik	MSc candidate	
Vivek Veeriah	MSc candidate	
Banafshe Rafiee	MSc candidate	
Gautham Vasan	MSc candidate	
Nadia Ady	MSc candidate	
Weifeng Chen	MSc candidate	

f. Other Members

Name	**Role**
Beverly Balaski	Program administrator
Alexandra Kearney	Undergrad researcher
Adam Parker	Undergrad researcher
Dylan Brenneis	Undergrad researcher
Devin Bradburn	Undergrad researcher
Nikolaus Yasui	Undergrad researcher
Mark Sebestyen	Undergrad researcher
Touqir Sajed	Undergrad researcher
Zach Goldthorpe	High school summer student, July-August 2015
Parash Rahman	High school summer student, July-August 2015

g. Visitors

Name	Institution
Andrew Barto	University of Massachusetts, Amherst
Alborz Gerimafard	Amazon, Cambridge, Massachusetts
Hiroyuki Nakahara	RIKEN Brain Science Institute, Japan
G.Wu	Co-Founder and VP Business Operations; Maluuba, Waterloo, Ontario
Prashanth L.A.	University of Maryland, MD
Jan Leike	Australian National University
Pierre-Luc Bacon	McGill University, Montreal (two weeks)
Andras Antos	Senior Research Scientist from Hungary
Nader Bshouty	Technion - Israel Institute of Technology, Haifa, Israel
David Pal	Google, New York
Zoltan Szabo	Gatsby Unit, University College, London, UK
Shantanu Jain	Indiana University
Nathaniel Korda	University of Oxford, UK
Philip S. Thomas	Carnegie Mellon University, Pittsburgh, PA
Thomas Colin	Plymouth University, UK (two months)
Michael Littman	Brown University
Baharak Rastegari	University of Glasgow
Lucas Lehnert	McGill University
Sam Pasupalak	CEO, Maluuba, Waterloo
Amir-massoud Farimand, Masumoto, Tanaka, many others	Mitsubishi, Japan, MERL, Cambridge Massachusetts
Russell Kalmacoff	Rockmount
Eric Chalmers	University of Lethbridge

6. COLLABORATIONS

Provincial	
Participants	Nature of Collaboration
Alberta Ingenuity Center for Machine Learning (AICML)	R. Sutton, D. Schuurmans, Cs. Szepesvári, Patrick Pilarski, and M. Bowling are among the ten principle investigators for this center at the University of Alberta. Total annual funding for AICML is $2M/year
Patrick Pilarski and Richard Sutton (RLAI) and Dr. Kim Adams University of Alberta Rehab Med and Dr. Mahdi Tavakoli Department of Electrical and Computer Engineering University of Alberta	Collaborative work on the use of assistive robots to facilitate the development, skill acquisition, and cognitive assessment of young children and infants with severe motor impairments (e.g., children with cerebral palsy); work involves the use of machine intelligence to enable the automatic adaptation of a robot control system to match the needs and abilities of individual children. Pilarski and Adams co-supervise visiting graduate students from a partner institution in Mexico city. (CHRP $333K over three years)
Patrick Pilarski (RLAI) and Martin Ferguson-Pell (Faculty of Rehabilitation Medicine, University of Alberta	Collaboration between the RLAI and the Rehabilitation Robotics Sandbox Laboratory (Rehabilitation Medicine, U. Alberta); this collaboration focuses on using new machine learning methods to predict fatigue in wheelchair users and enable novel muscle stimulation paradigms.
Patrick Pilarski and Richard Sutton (RLAI), and CFI SMART Neuroprosthetics Team, University of Alberta	Preliminary work on the use of machine learning to adapt and optimize neural interfaces and assistive robotic devices for people with motor disabilities.
L.M. Pilarski (Dept. Oncology, University of Alberta), L.M. McMullen (Agriculture, University of Alberta), M. Gaenzle (Agriculture, University of Alberta), Patrick Pilarski (RLAI), X. Yang (Lacombe Research Centre)	The meat processing industry needs a portable and rapid platform to detect pathogens during meat processing. We work with end users and colleagues to develop inexpensive computer systems that enable the automation of the testing system. This work includes software and hardware automation, biomedical pattern analysis, and machine intelligence.

Dr. Ian Adatia, Director of the Pediatric Cardiac Critical and Intermediate Care Program, Stollery Children's Hospital and Roger Zemp, Department of Electrical and Computer Engineering University of Alberta	Joint research with Dale Schuurmans on adaptive signal processing techniques for heart disease diagnosis, focusing on predicting events with ECG and PCG signals.

National	
Participants	Nature of Collaboration
Serdar Yuksel, Queen's University, Kingston, ON	Joint research with Andras György on control with communication constraints (as a theoretical foundation of controlling a prosthetic arms)
Richard Sutton (RLAI) and Doina Precup (McGill University)	Joint research on reinforcement learning architectures for planning, learning, and acting. Precup student Pierre-luc Bacon has been an extended visitor.
Richard Sutton (RLAI) and Maluuba, a young company based in Waterloo and Montreal	Sutton has provided advice and guidance on reinforcement learning to this natural-language company. Sutton is expected to become an official advisor to the company.
Richard Sutton (RLAI), Ziad Shawwash (University of British Columbia), and BC Hydro	Joint research on applications of reinforcement learning ideas to forecasting, logistics, and management of the BC Hydro system.

International	
Participants	Nature of Collaboration
Zinhua Zhang, NICTA	Joint research with Dale Schuurmans on representation learning and optimization.
Yuhong Guo, Temple University	Joint research with Dale Schuurmans on machine learning in bioinformatics and machine learning methods for large scale multi-target prediction problems.
Yaoliang Yu, Carnegie Mellon University	Joint research with Dale Schuurmans on efficient optimization methods for machine learning, convex reformulations of representation learning problems, and algorithmic techniques for exploiting structured sparsity.
Andras Antos, MTA SZTAKI, Hungary	Joint research with Csaba Szepesvári on active learning.

Branislav Kveton, Adobe Research, San Jose	Joint research with Csaba Szepesvári on bandit combinatorial optimization.
Andras György, Imperial College, UK	Joint research with Csaba Szepesvári on online learning.
Mohammad Ghavamzadeh, Adobe Research, San Jose	Joint research with Csaba Szepesvári on classification calibration.
David Pal, Yahoo Research, New York	Joint research with Csaba Szepesvári on online learning.
Yasin Abbasi-Yadkori, Queensland University of Technology, Brisbane, Australia	Joint research with Csaba Szepesvári on online learning in control.
Erik Talvitie, Franklin and Marshall College	Joint research and supervision of graduate students with Michael Bowling, developing new algorithms for intelligent exploration in domains with sparse rewards.
Patrick M. Pilarski, J. Hebert, Craig Chapman (Phys Rec), Albert Vette (Mec ENG), and Cleveland Clinic (USA) and UNB (Canada)	DARPA HAPTIX program to measure effectiveness of bi-directional neuro-prostheses. (data science and multimodal sensorymodal data streams)
Richard Sutton and Patrick M. Pilarski (RLAI), Osmar Zaiane and Cameron Schuler (and others in AICML), and Mitsubishi Electric Co. (J.P.)	New reinforcement learning techniques for dealing with industrial problems with uncertain data. This is expected to become a substantial collaboration through our associated machine learning center (AICML) supported by Mitsubishi funds. Plans are being discussed for Mitsubishi researchers to come to Edmonton to learn about reinforcement learning.
Patrick M. Pilarski (RLAI) and Sophia Adamia (Dana Farber Cancer Institute, Harvard Medical School, USA).	Investigation of cancer-related genetic markers in patients. Collaboration involving machine learning and biomedical data mining methods.
Patrick Pilarski and Richard Sutton (RLAI), Jason P. Carey (Department of Mechanical Engineering, University of Alberta), Michael R. Dawson, Jacqueline S. Hebert, K. Ming Chan (Glenrose Rehabilitation Hospital).	"Bionic limbs for improved natural control" this collaboration investigates the use of reinforcement learning and real-time machine learning to enable adaptive, intuitive control of myoelectric prostheses and other assistive robotic devices.
Deniz Gunduz, Imperial College, London, UK	Joint research with Andras György on application of reinforcement learning in wireless communications.
Joel Veness, Google Deepmind, London, UK	Joint research with Andras György data compression
Richard Sutton (RLAI) and David Silver (Google Deepmind, London, UK)	Joint research on reinforcement learning. This collaboration received funding of CAD$250,000 in support of RLAI activities in 2015.
Richard Sutton (RLAI) and Itamar Arel and other principals at Osaro, a startup company in San Francisco, USA	Through skype meetings and visits in each direction, Sutton has provided advice and guidance on research directions for Osaro. Sutton is an official advisor to the company.

Richard Sutton and Michael Bowling (RLAI) and Cogitai, a new startup company founded by Peter Stone (University of Texas), Satinder Singh (University of Michigan), and Mark Ring	Sutton and Bowling are associated with Cogitai as advisors and as members of their "brain trust," researchers who can be drawn upon or can contribute as needed.
Richard Sutton (RLAI) and Andrew Barto (University of Massachusetts at Amherst)	Sutton and Barto are working together to produce a second edition of their successful textbook *Reinforcement Learning: An Introduction*.
Richard Sutton (RLAI) and Steven Eliuk and others at Samsung Research America	Sutton pitched RLAI research to Samsung Research America on September 21, 2015.
Richard Sutton (RLAI) and Samsung SDS, San Jose, USA	Samsung SDS has made a major effort toward delivering reinforcement learning as a service, and Sutton is advising them as a consultant
Richard Sutton (RLAI) and the international CIFAR research project in Neural Computation for Adaptive Perception (NCAP)	Sutton has joined this year as a new member of this long-standing CIFAR research program, with members around the world. The program is led by Yann LeCun (New York University and Facebook) and Yoshua Bengio (Universite de Montreal). It is now largely focused on deep learning and unsupervised learning.

7. GRADUATES

Name	**Degree**	**Research topic**	**Current Position**
Kavosh Asadi Atui	MSc	Strengths, Weaknesses, and Combinations of Model-based and Model-free Reinforcement	PhD student at Brown University
Trevor Davis	MSc	Using Response Functions for Strategy Training and Evaluation	PhD student at the University of Alberta
Roshan Shariff	MSc	Exploiting Symmetries to Construct Efficient MCMC Algorithms With an Application to SLAM	PhD student at the University of Alberta
Dustin Morrill	MSc	Using Regret Estimation to Solve Games Compactly	PhD student at the University of Alberta
Craig Sherstan	MSc	Towards Prosthetic Arms as Wearable Intelligent Robots	PhD student at the University of Alberta
Min Yang	Course-based MSc	A Large Margin Apprenticeship Learning Framework	CGI, Edmonton

Xu Bing	MSc	Deep Convolutional Networks for Image Classification	PhD student at Univesité de Montreal
Weifeng Chen	MSc	Motion Planning with Monte Carlo Random Walks	
Adam White	PhD	Developing a Predictive Approach to Knowledge	Assistant Scientist at Indiana University
Nolan Bard	PhD	Online Agent Modelling in Human-Scale Problems	Postdoctoral fellow at the University of Alberta
Michael Johanson	PhD	Robust Strategies and Counter-Strategies: From Superhuman to Optimal Play	Consultant in the Edmonton, Alberta area
Hengshuai Yao	PhD	Model-based Reinforcement Learning with State and Action Abstractions	
James Neufeld	PhD	Adaptive Monte Carlo Integration	Data Scientist at Twitter

8. INTELLECTUAL PROPERTY

Intellectual Property	Status	Short Description
PATENTS	none	
LICENSES		
Spinoff Companies	none	

9. PUBLICATIONS

REFEREED JOURNAL PUBLICATIONS

M. Elgendi, R. Fletcher, I. Norton, M. Brearley, D. Abbott, N. Lovell, and D. Schuurmans, "On time-domain analysis of photoplethysmogram signals for monitoring heat stress," *Sensors 15*(10):24716-24734, Sep. 2015.

M. Elgendi, R. Fletcher, I. Norton, M. Brearley, D. Abbott, N. Lovell, and D. Schuurmans, "Frequency analysis of photoplethysmogram and its derivatives," *Computer Methods and Programs in Biomedicine 122*(3):503-512, Oct. 2015.

M. Elgendi, P. Bobhate, S. Jain, L. Guo, J. Rutledge, Y. Coe, R. Zemp, D. Schuurmans, and I. Adatia, "The unique heart sound signature of children with pulmonary artery hypertension," *Pulmonary Circulation 5*(4):631-639, 2015.

J. Gunther, P.M. Pilarski, G. Helfrich, H. Shen, and K. Diepold, "Intelligent laser welding through representation, prediction, and control learning: An architecture with deep neural networks and reinforcement learning," *Mechatronics 34*:1-11, 2016.

R.S. Sutton, A. Mahmood, and M. White, "An Emphatic Approach to the Problem of Off-policy Temporal-Difference Learning," *Journal of Machine Learning Research*, accepted November 2015.

H. van Seijen, A. Mahmood, P.M. Pilarski, R.S. Sutton, and M. Machado, "True Online Temporal-Difference Learning," *Journal of Machine Learning Research*, accepted January 2016.

S.V. Albrecht, J. Christopher L. David, L. Buckeridge, A. Botea, C. Caragea, C. Chi, T. Damoulas, B.N. Dilkina, E. Eaton, P. Fazli, S. Ganzfried, M.T. Lindauer, Y. Malitsky, G. Marcus, S. Meijer, F. Rossi, A. Shaban-Nejad, S. Thiébaux, M. M. Veloso, T. Walsh, C. Wang, J. Zhang, Y. Zheng, and M. Machado, "Reports from the 2015 AAAI Workshop Program," *AI Magazine*, Jul. 2015.

HIGHLY REFEREED ARCHIVAL CONFERENCE PROCEEDINGS

Y. Abbasi-Yadkori, and Cs. Szepesvári, "Bayesian optimal control of smoothly parameterized systems," *Proc. Conf. on Uncertainty in Artificial Intelligence* (UAI 2015), May 2015.

D. Hsu, A. Kontorovich, and Cs. Szepesvári, "Mixing time estimation in reversible Markov chains from a single sample path," *Proc. Advances in Neural Information Processing Systems* (NIPS 2015), Dec. 2015, 21% acceptance.

X. Hu, Prashanth L.A., A. György, and Cs. Szepesvári, "Bandit Convex Optimization with Biased Noisy Gradient Oracles," *Proc. 19th Int'l Conf. on Artificial Intelligence and Statistics*, (AISTATS 2016), to appear.

R. Huang, A. György, and Cs. Szepesvári, "Deterministic independent component analysis," *Proc. Int'l Conf. on Machine Learning* (ICML 2015), Jul. 2015, pp. 2521–2530, 26% acceptance.

P. Joulani, A. György, and Cs. Szepesvári, "Fast Cross-Validation for Incremental Learning," *Proc. Int'l Joint Conf. on Artificial Intelligence* (IJCAI 2015), Jul. 2015, pp. 3597-3604, 28% acceptance.

V. Lisy, T. Davis, and M. Bowling, "Counterfactual Regret Minimization in Sequential Security Games," *Proc. AAAI Conf. on Artificial Intelligence* (AAAI 2016), Feb. 2016, 26% acceptance.

P. Joulani, A. György, and Cs. Szepesvári, "Delay-Tolerant Online Convex Optimization: Unified Analysis and Adaptive-Gradient Algorithms," *Proc. AAAI Conf. on Artificial Intelligence* (AAAI 2016), Feb. 2016, 26% acceptance.

C. Xiao, and M. Müller, "Factorization Ranking Model for Move Prediction in the Game of Go," *Proc. AAAI Conf. on Artificial Intelligence* (AAAI 2016), Feb. 2016, 26% acceptance.

B. Kveton, Z. Wen, A. Ashkin, and Cs. Szepesvári, "Tight Regret Bounds for Stochastic Combinatorial Semi-Bandits," *Proc. 18th Int'l Conf. on Artificial Intelligence and Statistics* (AISTATS 2015), May 2015.

B. Kveton, Z. Wen, A. Ashkan, and Cs. Szepesvári, "Combinatorial cascading bandits," *Proc. Advances in Neural Information Processing Systems* (NIPS 2015), Dec. 2015, 21% acceptance.

B. Kveton, Z. Wen, A. Ashkan, and Cs. Szepesvári, "Cascading bandits: Learning to rank in the cascade model," *Proc. Int'l Conf. on Machine Learning* (ICML 2015), Jul. 2015, 26% acceptance.

T. Lattimore, Cs. Szepesvári, and K. Crammer, "Linear Multi-Resource Allocation with Semi-Bandit Feedback," *Proc. Advances in Neural Information Processing Systems* (NIPS 2015), Dec. 2015, 21% acceptance.

T. Lattimore, "The Pareto Regret Frontier for Bandits," *Proc. Advances in Neural Information Processing Systems* (NIPS 2015), Dec. 2015, 21% acceptance.

G. Lever, R. Shawe-Taylor, R. Stafford, and Cs. Szepesvári, "Compressed conditional mean embeddings for model-based reinforcement learning," *Proc. AAAI Conf. on Artificial Intelligence* (AAAI 2016), Feb. 2016, 26% acceptance.

L. Li, R. Munos, and Cs. Szepesvári, "Towards Minimax Off-Policy Value Estimation," *Proc. 18th Int'l Conf. on Artificial Intelligence and Statistics* (AISTATS 2015), May 2015.

V. Lisy, T. Davis, and M. Bowling, "Counterfactual Regret Minimization in Sequential Security Games," *Proc. 30th AAAI Conf. on Artificial Intelligence* (AAAI 2016), Feb. 2016, 26% acceptance.

J. Mei, H. Zhang, and B.L. Lu, "On the Reducibility of Submodular Functions," *Proc. 18th Int'l Conf. on Artificial Intelligence and Statistics* (AISTATS 2015), May 2015.

F. Mirzazadeh, M. White, A. György, and D. Schuurmans, "Scalable Metric Learning for Co-Embedding," *Proc. European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, (ECML PKDD 2015), Aug. 2015, pp. 625-642, 23% acceptance.

F. Mirzazadeh, S. Ravanbakhsh, N. Ding, and D. Schuurmans, "Embedding Inference for Structured Multilabel Prediction," *Proc. Advances in Neural Information Processing Systems* (NIPS 2015), Dec. 2015, 21% acceptance.

H. van Seijen, and R.S. Sutton, "A Deeper Look at Planning as Learning from Replay," *Proc. Int'l Conf. on Machine Learning* (ICML 2015), Jul. 2015, 26% acceptance.

J. Wen, D. Schuurmans, and R. Greiner, "Correcting Covariate Shift with the Frank-Wolfe Algorithm," *Proc. Int'l Joint Conf. on Artificial Intelligence* (IJCAI 2015), Jul. 2015.

Y. Wu, A. György, and Cs. Szepesvári, "On Identifying Good Options under Combinatorially Structured Feedback in Finite Noisy Environments," *Proc. Int'l Conf. on Machine Learning* (ICML 2015), Jul. 2015, pp. 1283–1291, 26% acceptance.

Y. Wu, A. György, and Cs. Szepesvári, "Online Learning with Gaussian Payoffs and Side Observations," *Proc. Advances in Neural Information Processing Systems* (NIPS 2015), Dec. 2015, pp. 1360-1368, 21% acceptance.

Y. Wu, R. Shariff, T. Lattimore, and Cs. Szepesvári, "Conservative Bandits," *Proc. Int'l Conf. on Machine Learning* (ICML 2016), Feb. 2016, 26% acceptance.

Y. Liang, M. Machado, E. Talvitie, and M. Bowling, "State of the Art Control of Atari Games Using Shallow Reinforcement Learning," *Proc. 13th Int'l Conf. on Autonomous Agents and Multiagent Systems* (AAMAS 2014), accepted to appear May 2016.

OTHER CONFERENCE AND WORKSHOP PROCEEDINGS

K. Durkota, V. Lisy, B. Bosansky, and C. Kiekintveld, "Approximate Solutions for Attack Graph Games with Imperfect Information," *Proc. Int'l Conf. On Decision and Game Theory for Security* (GameSec 2015), Nov. 2015, pp. 228-249.

R. Klima, V. Lisy, and C. Kiekintveld, "Combining online learning and equilibrium computation in security games," *Proc. Int'l Conf. On Decision and Game Theory for Security* (GameSec 2015), Nov. 2015, pp. 130-149.

C. Sherstan, J, Modayil, and P.M. Pilarski, "A Collaborative Approach to the Simultaneous Multi-joint Control of a Prosthetic Arm," *Proc. IEEE Int'l Conf. on Rehabilitation Robotics* (ICORR 2015), Aug. 2015, pp. 13-18.

H. van Seijen, A. Mahmood, P.M. Pilarski, and R.S. Sutton, "An Empirical Evaluation of True Online TD(λ)," *Proc. European Workshop on Reinforcement Learning* (EWRL 2015), Workshop Paper, Jul. 2015.

B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," Proc. Deep Learning Workshop (ICML 2015), Workshop Paper, Jun. 2015.

M. Machado, S. Sriram, and M. Bowling, "Domain-Independent Optimistic Initialization for Reinforcement Learning," *Proc. AAAI Workshop on Learning for General Competency in Video Games*, (AAAI 2015), Workshop Paper, Jul. 2015.

A. Mahmood, H. Yu, M. White, Richard Sutton, "Emphatic Temporal Difference Learning," *Proc. European Workshop on Reinforcement Learning* (EWRL 2015), Workshop Paper, Jun. 2015.

P. Joulani, A. György, and Cs. Szepesvári, "Classification with Margin Constraints: A Unification with Applications to Optimization," *Proc. NIPS Workshop on Optimization* (OPT 2015), Workshop Paper, Dec. 2015.

W. Chen, and M. Müller, "Continuous Arvand: Motion Planning with Monte Carlo Random Walks," *Proc. 3rd ICAPS Workshop on Planning and Robotics (PlanRob 2015)*, pages 23-34, Workshop paper, June 2015.

R. Huang, A. György, and Cs. Szepesvári, "Easy data for independent component analysis," *Proc. NIPS Workshop on Learning faster from easy data II* (NIPS 2015), Workshop Paper, Dec. 2015.

H. van Hasselt, and R. S. Sutton, "Learning to Predict Independent of Span," arXiv:1508.04582, 2015.

SPECIAL/INVITED PRESENTATIONS

Person	Title	Venue
R. Sutton	Deep Questions	Team meeting of the CIFAR project on Neural Computation for Adaptive Perception, Montreal
R. Sutton	Introduction to Reinforcement Learning with Function Approximation	Tutorial at the Conference on Neural Information Processing Systems (audience of 1200-1500), Montreal
R. Sutton	The Future of AI	Distinguished Lecture Series, University of British Columbia (4600 youtube views)
R. Sutton	The Future of Artificial Intelligence	Invited lecture in LABMP 590, University of Alberta (520 youtube views)
R. Sutton	Reinforcement Learning: Learning without Labels	Invited presentation to Samsung Research conference in Mountain View, USA
D. Schuurmans	Convex Methods for Latent Representation Learning	University of Michigan
Cs. Szepesvári	Conservative Bandits	University of South California
Cs. Szepesvári	Conservative Bandits	Adobe Research
Cs. Szepesvári	Reinforcement Learning	Kyoto, Japan
Cs. Szepesvári	Lazy Posterior Sampling for Parametric Nonlinear Control	European Workshop on Reinforcement Learning, Lille, France
Cs. Szepesvári	Online learning and prediction on a budget	ICML Workshop on Resource-Efficient Machine Learning, Lille,
Cs. Szepesvári	Fast Cross-Validation for Incremental Learning	Deepmind, London, UK
Cs. Szepesvári	When to stop, and how to allocate resources with small regret: Two stories on the power of optimism	Indian Institute of Science, Bangalore
Cs. Szepesvári	Adaptive Monte-Carlo via Bandit Allocation	IBM Research Lab, Bangalore
Cs. Szepesvári	Optimistic Algorithms for Online Learning in Structured Decision Problems	UCL, London
M. Bowling	von Neumann's Dream	Canadian Science Deans
M. Bowling	Games and Intelligence of the Artificial Kind	Canadian Association of Science Centres Conference

M. Bowling	Games and Intelligence of the Artificial Kind	UNIty in diVERSITY
M. Bowling	Dr. Jekyll and Mr. Hyde	NIPS Workshop on Deep Reinforcement Learning
M. Bowling	von Neumann's Dream	NIPS Workshop on Inference and Control of Multi-agent systems
M. Bowling	Adventures in Implicit Agent Modelling	AAAI Workshop Multiagent interaction without prior coordination
A. György	Adaptive Monte Carlo via Bandit Allocation	CNRS, Linear Accelerator Laboratory (France)
A. György	Online Learning in Adversarial Markov Decision Processes	Department of Mathematics & Statistics, Queen's University,
P. M. Pilarski	Solving AI with help from Intelligence Amplification	Barbados Workshop in Reinforcement Learning
P. M. Pilarski	Prosthetics showcase	Breakfast TV, Media Coverage
P. M. Pilarski	Combined expertise leads to new avenues for PMI technology	O&P News
P. M. Pilarski	Autonomy and Intelligence in Rehabilitation Technology	Science Expo, Edmonton
P. M. Pilarski	Autonomy and Intelligence in Rehabilitation Technology	Fujita Health University, Nagoya, Japan
P. M. Pilarski	Reinforcement Learning: Algorithms for acquiring and using predictive knowledge	Information Technology R&D Centre, Ofuna, Kamakura, Japan
P. M. Pilarski	Machine Learning and why you should care	Startup Edmonton, Student Developer Conference, Edmonton
H. van Hasselt	Computational Game Theory for Security Applications	University of Calgary
M. Mueller	Computer Go Research - the Challenges Ahead	Keynote, 2015 IEEE Conference on Computational Intelligence and Games (CIG), Tainan, Taiwan
M. Mueller	Using Domain-specific Knowledge for Monte Carlo Tree Search in Go	National Chiao Tung University, Hsinchu, Taiwan
M. Mueller	Mastering the Game of Go - Can a Computer Program Beat a Human Champion?	Public talk, University of Alberta
Sutton, Bowling, Szepesvari, Mueller	AlphaGo and the Future of Computer Games: A Conversation at the University of Alberta	YouTube video created by Paul Lu with >3700 views

AWARDS

Richard Sutton was ranked as the 16th most influential author in computer science by "Semantic Scholar", a tool released by the Allen Institute for Artificial Intelligence and announced in *Science* on April 20, 2016 (technically just outside the reporting period).

P.M. Pilarski and C. Sherstan: ICORR 2015 Singapore: Best Student Paper Finalist Award, A Collaborative Approach to the Simultaneous Multi-joint Control of a Prosthetic Arm.

A. György: ICML 2015 Reviewer Award

THESES

Kavosh Asadi Atui, MSc, "Strengths, Weaknesses, and Combinations of Model-based and Model-free Reinforcement Learning," September 2015.

Roshan Shariff, MSc, "Exploiting Symmetries to Construct Efficient MCMC Algorithms With an Application to SLAM," June 2015.

Craig Sherstan, MSc, "Towards Prosthetic Arms as Wearable Intelligent Robots," September 2015.

Dustin Morril, MSc, "Using Regret Estimation to Solve Games Compact," February 2016.

Trevor Davis, MSc, "Using Response Functions for Strategy Training and Evaluation," June 2015.

Xu Bing, MSc, "Deep Convolutional Networks for Image Classification," March 2016.

Weifeng Chen, "Motion Planning with Monte Carlo Random Walks," November 2015.

Nolan Bard, PhD, "Online Agent Modelling in Human-Scale Problems," March 31, 2016.

Adam White, PhD, "Developing a Predictive Approach to Knowledge," June 2015.

Hengshuai Yao, PhD, "Model-based Reinforcement Learning with State and Action Abstraction," December 2015.

Michael Johanson, PhD, "Robust Strategies and Counter-Strategies: From Superhuman to Optimal Play," January 2016.

James Neufeld, PhD, "Adaptive Monte Carlo Integration," March 2016.

10. OUTREACH

Richard Sutton gave an interview with Marina Krakovsky of the *Communications of the ACM* on February 12, 2016.

Richard Sutton gave an interview with Tom Simonite of *MIT Technology Review* on March 4, 2016.

Richard Sutton gave an interview with Tanya Lewis of *Business Insider* on March 11, 2016, and was featured in her story at http://www.businessinsider.com/what-does-googles-deepmind-victory-mean-for-ai-2016-3.

Richard Sutton gave an interview with Katja Grace and John Salvatier for *AI Impacts*, a website supported by the Future of Life Institute.

Richard Sutton gave an interview with Omar Mouallem of *New Trail*, the University of Alberta alumni magazine on September 14, 2015, and was one of those featured in the resulting article at https://newtrail.ualberta.ca/winter-2015/features-dept/life-after-the-singularity.

Richard Sutton gave a six-minute radio interview about autonomous weapons with Dan Delmar of *CJAD* on July 27, 2015.

Richard Sutton gave an interview with Raffi Nhatchadourian of *The New Yorker* on July 23, 2015, and was quoted in his article in the magazine on November 23 (http://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom).

Richard Sutton gave an interview with Luke Dormehl, a journalist writing a book for Penguin Random House, on October 22, 2015.

Richard Sutton answered questions by email from Jacy Li, a reporter with the Beijing-based newspaper *Mirror Evening* on March 12, 2016.

Martin Müller and Richard Sutton gave interviews with Ivan Semeniuk, science reporter for *The Globe and Mail,* on March 15, 2016, and were quoted in his article (http://www.theglobeandmail.com/technology/science/crossing-ai-threshold-computer-program-conquers-worlds-most-complex-game/article29255280/).

Csaba Szepesvári supervised a high school student in July and August as part of the High School Internship Program and one WISEST student at the Department of Computing Science.

Patrick M. Pilarski presented robot demonstrations and talks at Two Hills Mennonite School.

Dustin R. Morrill gave a presentation at *Telus World of Science*. He showed a demonstration of Cepheus, our essentially-perfect Poker-playing program, at a *Dark Matters* event. The public was encouraged to play against the program, and Dustin discussed game solving and algorithmic game theory research.

Leah Hackman gave a Python Tutorial for WISEST, Heritage Youth Researcher Summer Program, and High School Internship students. 16 students attended

Timothy Yee gave a presentation on artificial intelligence research to Edmonton high school kids as a part of the university's Iverson day.

Harm van Seijen mentored a High School Internship student at the Department of Computing Science.

Martin Müller, Richard Sutton, Csaba Szepesvári, and Michael Bowling were interviewed in a video created by Paul Lu of the University of Alberta on March 11, 2016. The video has been viewed on YouTube more than 3700 times.