

# RESEARCH IN ACTION

Research Highlights

## ARTIFICIAL INTELLIGENCE

### AI: Computational Linguistics and Natural Language Processing

1. Adapting Acoustic and Lexical Models to Dysarthric Speech
2. Learning Mixed Acoustic/Articulatory Models for Disabled Speech
3. Speech Retrieval Engine
4. A Computational Study of Late Talking in Word-Meaning Acquisition

### AI: Knowledge Representation and Cognitive Robotics

5. Eliciting Reward Functions for Sequential Decision Making
6. Group Decision Making and Elicitation with Partial Preferences
7. Generalized Median Mechanisms for Group Decision Making in Multi-Dimensional Settings
8. Modeling, Customizing and Optimizing Business Processes
9. Monitoring the Execution of Partial-Order Plans via Regression

### AI: Machine Learning and Neural Networks

10. Leaksplorer - Interactive Visualization of Large Document Collections
11. Learning to Control with Recurrent Neural Networks
12. Machine Learning for Aerial Image Interpretation
13. Automatic Real-Time Music Generation
14. Unsupervised Feature Learning using Higher-Order Neural Networks
15. HOP Context
16. Opportunity Cost in Bayesian Optimization
17. Preferences, Learning and Matching: Algorithmic Models for Academic and Marital Bliss

### AI: Computational Vision

18. Fast Large-Scale Similarity Search by Multi-Index Hashing on Binary Codes
19. Detecting Reduplication in Videos of American Sign Language

## MACHINE LEARNING AND COMPUTATIONAL BIOLOGY

20. JBASE: Joint Bayesian Analysis of Sub-Phenotypes and Epistasis

## COMPUTATIONAL BIOLOGY

21. Towards Better and Faster Diagnostics of Genetic Diseases

## COMPUTER GRAPHICS

22. Slices: A Shape-Proxy Based on Planar Sections

## COMPUTER SYSTEMS

23. Speeding up Spatial Database Query Execution using GPUs
24. Recon: Declarative Invariants for Runtime Consistency Checking
25. Protecting Kernel Integrity from Untrusted Extensions Using Dynamic Binary Instrumentation
26. Jettison: Efficient Idle Desktop Consolidation with Partial VM Migration
27. Repeat After Me, "I am a Human": Verifying Human Users in Speech-Based Interactions
28. Understanding the Nature of DRAM Errors and the Implications for System Design

## DATABASES

29. BibBase and Linked Open Data on the Web

## HUMAN-COMPUTER INTERACTION

30. TAGLab - Technologies for Aging Gracefully Lab
31. Improv Remix
32. Reinventing the (Mouse) Wheel
33. Experimentation with a 1ms Latency Direct-Touch Device
34. Using Gestures and Speech for K-2 Education

## NUMERICAL ANALYSIS

35. Parameter Determination in Mathematical Modelling by ODEs
36. Approximations to Loss Probabilities of Credit Portfolios

## SOFTWARE ENGINEERING

37. Partial Models: Towards Modeling and Reasoning with Uncertainty
38. Who's Flying this Ship? Systems Thinking for Planet Earth
39. Business Intelligence Modeling and Reasoning
40. MarkUs
41. Exploring Contextual Complexities of Agile Adoption
42. Iterative, Interactive Analysis of Agent-Goal Models for Early Requirements Engineering

## THEORY OF COMPUTATION

43. Algorithms in Trading
44. Equivalence of Memory Consistency Models with Relaxed Program Order
45. Encryption Systems for the Modern World

## MESSAGE from the chair

The Department of Computer Science (DCS) at the University of Toronto, the oldest CS department in the country, was established in 1964. Our faculty and graduate students engage in ground-breaking research that shapes the future of computing and beyond, earning DCS a reputation as one of the top ten Computer Science departments in the world\*. With a multitude of interdisciplinary collaborations, the impact of DCS-incubated research is widespread; extending from biotechnology to business and from education to entertainment.

Within the last couple of years, it's become clear that things are changing on the academic landscape. Here at the department, we are sensing a significant change in direction; industrial relations are, more than ever, an important part of our future. In order to maintain the strength and impact of our research program, we need to reinforce and cultivate our ties with industry. Our Associate Chair for Research and Industrial Relations, Professor Gerald Penn, provides the critical liaison between companies and our faculty.

To this end, we are also pleased to offer our industrial partners multiple avenues for engagement: a more recent example would be the Master of Science in Applied Computing, our burgeoning professional master's (PM) program. Started in fall 2010, the MScAC includes an 8-month industrial internship, and has brought a new group of top-calibre graduate students to the department. We have been thrilled with the feedback we have received from our industry partners thus far, and we invite you to learn more about the program [[www.cs.toronto.edu/mscac](http://www.cs.toronto.edu/mscac)]. Professor Eugene Fiume, the Director of our PM program, would be happy to discuss with you how the program can meet your needs.

The Research in Action showcase also continues to be a great opportunity to establish and bolster these relationships, giving guests a sense of the limitless potential for collaboration across our 11 research groups, not to mention our ongoing work with the world-renowned divisions across the University of the Toronto, the hospitals located a block away, and our partners at peer institutions and companies across the world.

Much of our innovative research is shaped by collaborations with research partners in industry and in other academic disciplines. Our Research in Action event is designed to facilitate the expansion of our collaborative efforts and I invite prospective partners to explore the breadth and depth of the world-class research in DCS and to share with us their own ideas for new collaborations.

Sven Dickinson  
PROFESSOR AND CHAIR



\*[www.cs.toronto.edu/dcs/2011rankings](http://www.cs.toronto.edu/dcs/2011rankings)

## 1 Adapting Acoustic and Lexical Models to Dysarthric Speech

Dysarthric individuals are able to comprehend natural language, but lack the articulatory control and coordination to produce intelligible speech. Their speech is, in part, characterized by pronunciation errors that include deletions, substitutions, insertions, and distortions of phonemes. These errors follow certain consistent intra-speaker patterns that we exploit through acoustic and lexical model adaptation to improve automatic speech recognition (ASR) for dysarthric speech. We show that acoustic model adaptation yields an average relative word error rate (WER) reduction of 36.99% and that pronunciation lexicon adaptation (PLA) further reduces the relative WER by an average of 8.29% on a relatively large vocabulary task of over 1500 words for six speakers with severe to moderate dysarthria.

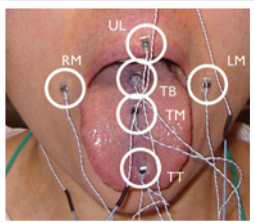
PROJECT TEAM

**Kinfe Tadesse Mengistu**, Postdoctoral Fellow  
**Graeme Hirst**, Faculty  
**Gerald Penn**, Faculty

## 2 Learning Mixed Acoustic/Articulatory Models for Disabled Speech

This poster shows that automatic speech recognition should accommodate speakers with disabilities by incorporating knowledge of the characteristics of their speech production. We describe the development of a new database of disabled speech that includes aligned acoustic and articulatory data obtained by electromagnetic articulography. This database is used to compare theoretical and empirical models of the vocal tract against discriminative models

such as neural networks, support vector machines, and conditional random fields on the task of phoneme recognition. Results show significant improvements in accuracy over the baseline through the application of articulatory data.



PROJECT TEAM

**Frank Rudzicz**, Postdoctoral fellow  
**Gerald Penn**, Faculty

## 3 Speech Retrieval Engine

With an increase in the general availability of resources such as streaming video/audio and inexpensive data storage, there is an apparent shift towards creation and dissemination of multimedia data. This consistent and considerable increase in multimedia data available call for information retrieval engines that can search through these data just as traditional text search engines have been searching through text-based data. We would like to present our speech retrieval system. Realizing that a significant portion of the information lies in the spoken part of an audio or video/audio document, this information retrieval engine, unlike current commercial search engines such as Google, Bing, or Yahoo, indexes all the spoken content. The system also provides an intuitive interface for the information user to search and browse through the multimedia document repository.

PROJECT TEAM

**Siavash Kazemian**, Graduate Student  
**Gerald Penn**, Faculty

## 4 A Computational Study of Late Talking in Word-Meaning Acquisition

Late talkers (LTs)-children who show a marked delay in vocabulary learning-are at risk for Specific Language Impairment (SLI), and much research has focused on identifying factors contributing to this phenomenon. We use a computational model of word learning to further shed light on these factors. In particular, we show that variations in the attentional abilities of the computational learner can be used to model various identified differences in LTs compared to normally-developing children: delayed and slower vocabulary growth, greater difficulty in novel word learning, and decreased semantic connectedness among learned words.

PROJECT TEAM

**Aida Nematzadeh**, Graduate Student  
**Afsaneh Fazly**, Postdoctoral Researcher  
**Suzanne Stevenson**, Faculty

# AI: KNOWLEDGE REPRESENTATION AND COGNITIVE ROBOTICS

## 5 Eliciting Reward Functions for Sequential Decision Making

Markov decision processes (MDPs) have proven useful modeling and finding optimal sequences of decisions under uncertainty, yet they require the specification of a large number of parameters to capture both system dynamics and a reward function. While dynamics can be learned by observation of the environment, the reward function reflects the subjective preferences of some user and can require sophisticated human judgment to assess relevant tradeoffs. Furthermore, the time-consuming process of specifying reward may need to be repeated to capture the varying preferences of different users. This work develops a framework for the incremental elicitation of reward functions for MDPs that reduces the burden on users. The framework begins with a loose specification of reward that replaces the certain reward function with a set of possible reward functions. We then compute policies of action that are robust with respect to reward uncertainty using the minimax regret criterion. The framework then reduces “regret” of the optimal robust policy by actively engaging with the user to incrementally reduce uncertainty over the reward function using simple queries. We have applied our approach to a variety of sequential decision making domains from “autonomic” self-managing computing servers to devices for the cognitive assistance of persons with dementia. We have observed that in many cases relatively few queries are necessary to produce provably optimal policies, significantly reducing the amount of information needed to specify MDP reward functions and addressing one of the fundamental bottlenecks in the specification of MDPs.

PROJECT TEAM

**Kevin Regan**, Graduate Student  
**Craig Boutillier**, Faculty

## 6 Group Decision Making and Elicitation with Partial Preferences

Our research is focused on making high quality preference-based group decisions while minimizing burdens on the end users. We pioneer practical and theoretically sound algorithms that can (1) make quantifiable near optimal decisions from only partial user preferences and (2) can query the most useful preferences to reduce user cognitive and communication burdens. Experiments

on real sushi preference data as well as real voting data demonstrate that our algorithms only need to query a fraction of the full preference information to make good enough group decisions.



PROJECT TEAM  
**Tyler Lu**, Graduate Student  
**Craig Boutilier**, Faculty

## 7 Generalized Median Mechanisms for Group Decision Making in Multi-Dimensional Settings

Effective group decision making requires assessment of the preferences of group members (users) to ensure maximal group satisfaction. In many settings, a user's preferences are dictated by their single "ideal" choice or outcome, with other choices being more or less preferred based on their "distance" from this ideal point. Such problems arise in industrial applications (e.g., warehouse location), public policy (placement of public facilities such as parks, libraries and highways), political choice, customer segmentation and product design, among others. In these settings, users often have an incentive to misrepresent their preferences. We consider the design of mechanisms for selecting multiple alternatives in multi-dimensional settings that ensure users will report their preferences truthfully, while at the same time maximizing various measures of social welfare. Our percentile mechanisms (a type of generalized median mechanisms) also exploit probabilistic information about the preferences of a user population to improve social welfare. Theoretical and empirical results show that these mechanisms have great promise.

PROJECT TEAM  
**Xin Sui**, Graduate Student  
**Craig Boutilier**, Faculty

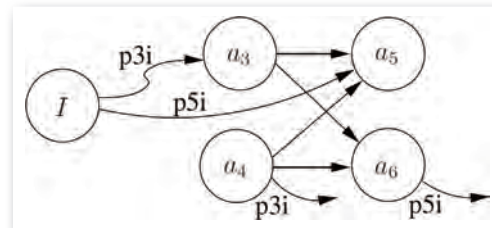
## 8 Modeling, Customizing and Optimizing Business Processes

Business processes are a natural vehicle for the creation and delivery of business intelligence; but one size does not fit all. Such business processes need to be customized and optimized to the context imposed by the specifics of the data, the users and stakeholders, and the delivery mechanism. We model and manipulate business processes in a manner that treats data and processes as first-class citizens, and that supports dynamic customization with respect to stakeholder requirements, including corporate and jurisdictional policies and regulations and optimization with respect to stakeholder preferences and priorities. A component of this work is funded by NSERC BIN (Business Intelligence Network).

PROJECT TEAM  
**Shirin Sohrabi**, Graduate Student  
**Sheila McIlraith**, Faculty

## 9 Monitoring the Execution of Partial-Order Plans via Regression

Partial-order plans are a flexible form of solution to an automated planning problem. In this work, we address the problem of monitoring the execution of a partial order plan so that an agent may operate in a dynamic environment and react to changes in the world. The approach we present generalizes the partial order plan into a policy for online execution, and allows the agent to be far more robust when executing the plan.



PROJECT TEAM  
**Christian Muise**, Graduate Student  
**Sheila McIlraith**, Faculty  
**Chris Beck**, Faculty (Mechanical & Industrial Engineering and CS)

NOTES

---

---

---

---

---

---

---

## 10 Leaksplorer - Interactive Visualization of Large Document Collections

Leaksplorer is a Machine Learning approach to quickly exploring enormously large document collections, such as the Wikileaks Iraq War Diaries. The algorithm extracts word features from the documents and embeds them into a 2D map while approximately preserving distances between the documents. In this map, similar documents will cluster together, revealing the high-level structure of the corpus. The map is presented in a fully native web application and can be panned and zoomed to explore it in more detail. Clicking the points displays the full documents, allowing the user to interactively explore their relationships. The application includes collaborative features, such as the possibility to comment on individual documents, bookmark them, and sort them into lists which can be shared via permalinks, Facebook and Twitter to other users. This way, the discovery of interesting information in large document collections can be effectively crowdsourced to thousands of contributors.



can be shared via permalinks, Facebook and Twitter to other users. This way, the discovery of interesting information in large document collections can be effectively crowdsourced to thousands of contributors.

PROJECT TEAM

**Hannes Bretschneider**, Graduate Student

**Brendan Frey**, Faculty (Electrical & Computer Engineering, Banting and Best Department of Medical Research, and CS)

## 11 Learning to Control with Recurrent Neural Networks

In this work we show how to train a neural network to control a simulated robot arm. The network learns to output a sequence of commands that bring the robotic arm to a any desired state. It can rapidly notice and counteract spontaneous external forces that are applied to the arm, a feat not achievable by any other control system. Our method relies on a new variant of the Hessian-Free optimizer that was shown to be very effective at optimizing recurrent neural networks.

PROJECT TEAM

**Ilya Sutskever**, Graduate Student

**Geoffrey Hinton**, Faculty

## 12 Machine Learning for Aerial Image Interpretation

We present a system that can automatically detect roads in aerial images. The system works by learning from hundreds of square kilometers of aligned road maps and aerial images. Our learning procedure can deal with maps that are often misaligned as well as maps that are sometimes inaccurate. Since spatial coherence is the

only property of roads that was built into our system, it can be applied to other types of objects, such as buildings and trees.



PROJECT TEAM

**Volodymyr Mnih**, Graduate Student

**Geoffrey Hinton**, Faculty

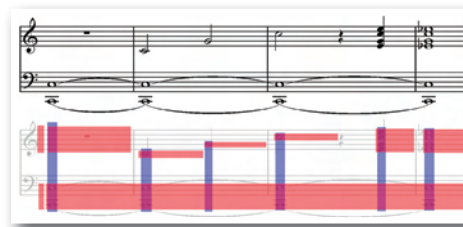
## 13 Automatic Real-Time Music Generation

This project implements an approach to continuous music generation, where original background music is composed in real-time, and is inexpensive to create, both financially and computationally. The basis for this work is an advanced application of Markov models that train on music of a given style, and combines the generated model with some basic constraints of music structure to produce an endless composition of musically coherent note of the training piece. As an added bonus, when a player moves from one region to another, these compositions would segue seamlessly and sensibly from one soundtrack to the next.

PROJECT TEAM

**Daniel Eisner**, Undergraduate

**Steve Engels**, Faculty



## 14 Unsupervised Feature Learning using Higher-Order Neural Networks

Unsupervised feature learning is a branch of machine learning devoted to automatically finding structure, also called features, in vast amounts of unlabeled data. These features are basic ingredients such that each piece of data is composed of one or more of these ingredients. For example, a table consists of four legs and a flat surface, while a human face consists of two eyes, a nose, and a mouth. These features can be useful for interpreting data, and for improving the performance of pattern recognition systems on previously unseen data. Many current methods use complex probabilistic models that result in a difficult learning problem. Using newly developed techniques from statistical estimation theory, we establish a link between these probabilistic approaches and a new class of non-probabilistic neural networks. We show that these neural network models can be significantly easier to train, while yielding similar empirical performance.

PROJECT TEAM

**Kevin Swersky**, Graduate Student

**Richard Zemel**, Faculty

## 15 HOP Context

When reading, listening, or observing, people make use of context to understand meaning, particularly when a sight or sound is ambiguous. Can computers be made to do the same? Our aim in this work is to try to get computers to use contextual cues in the task of image segmentation; that is, labeling each pixel in an image with the type of object that the pixel is a part of. Suppose we are interested in whether an image contains an airplane, and if so, which pixels are a part of the airplane(s). To see why context might be important, one can imagine that an image that resembles an airport will be more likely to contain an airplane. Also the sizes of structures, texture and other (usually aggregate) features of the image will give us an idea as to how large the airplane might be. The latter is a less obvious, yet very significant role of context. We teach our intelligent system to make use of contextual information, by supplying it with examples where the correct answer is known. We show that it performs better than a previous version of the same intelligent system which did not make use of contextual information.

PROJECT TEAM

**Nikola Karamanov**, Graduate Student  
**Danny Tarlow**, Graduate Student  
**Richard Zemel**, Faculty

## 16 Opportunity Cost in Bayesian Optimization

Bayesian optimization is a principled methodology for optimizing expensive 'black-box' functions. This project explores how to tune parameters of machine learning algorithms automatically through the Bayesian paradigm and in particular how to incorporate additional costs. As an example, we will demonstrate how to find the parameters required to cook the perfect soft-boiled egg using Bayesian optimization.

PROJECT TEAM

**Jasper Snoek**, Graduate Student  
**Richard Zemel**, Faculty

## 17 Preferences, Learning and Matching: Algorithmic Models for Academic and Marital Bliss

Recommendation systems, systems that make recommendations of interest to users, have been a recent hit both in the academic and industrial communities. Current recommendation systems usually assume that there is an infinite supply of each item, e.g., books or movies, being recommended. Recommending users to other users in online dating systems is different, because users may only be recommended to a single user at a time, and therefore require a new breed of recommendation systems. With this in mind we introduce recommendation systems that find a constrained match between users and other users (or items). We show that even when the system has limited information about the users it we can come up with high-quality matches. We also propose several active learning methods, which tailor the system based on customized interaction with the users.

PROJECT TEAM

**Laurent Charlin**, Graduate Student  
**Richard Zemel**, Faculty  
**Craig Boutilier**, Faculty

## 18 Fast Large-Scale Similarity Search by Multi-Index Hashing on Binary Codes

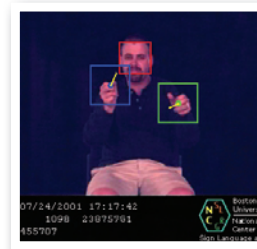
There has been growing interest in mapping high-dimensional data (images, videos, and documents) onto compact binary codes to support fast search for similar items in large data-sets. Binary codes are motivated by their use as direct indices (addresses) into a hash table. This facilitates fast retrieval with a few address look-ups as long as similar items have codes that differ from the query code by a few bits. However, codes longer than 32 bits are not being used in this context because the memory requirement for a big hash table becomes prohibitive, and the expected Hamming distance between the neighboring codes grows quickly with the code length. We present a novel way to build multiple hash tables on binary code substrings that enables exact K-nearest neighbor search in Hamming space, even for long codes. Our algorithm has sub-linear run-time behavior for uniformly distributed codes. It is storage efficient, and straightforward to implement. Empirical results show dramatic speedups over a linear scan baseline, both on 64- and 128-bit codes, for data-sets with up to one billion codes.

PROJECT TEAM

**Mohammad Norouzi**, Graduate Student  
**David Fleet**, Faculty

## 19 Detecting Reduplication in Videos of American Sign Language

Research in the automatic recognition of American Sign Language (ASL) lies at the interface of computer vision and natural language processing. The field has a long way to go, and many open problems to solve. We take a first stab at detecting the particular linguistic phenomena of reduplication by solving a string motif finding problem. In ASL, reduplication is used for a variety of linguistic purposes, including overt marking of plurality on nouns, aspectual inflection on verbs, and nominalization of verbal forms. Reduplication involves the repetition, often partial, of the articulation of a sign. If we could locate reduplication, we would be on our way to stripping an inflected sign down to its base form, for efficient indexing, matching, and retrieval in ASL databases. In this project, the apriori algorithm for mining frequent patterns in data streams is adapted for finding reduplication in videos of ASL. The apriori algorithm is extended to allow for inexact matching of similar hand motion subsequences and to provide robustness to noise. We demonstrate how this algorithm can be extended to automatically detect recurring signed sequences, given a discretization of the state space of hand configurations and a distance metric.

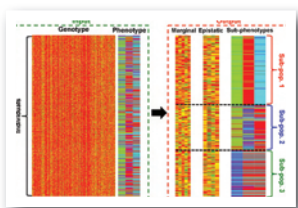


PROJECT TEAM

**Zoya Gavrilov**, Undergraduate  
**Sven Dickinson**, Faculty

## 20 JBASE: Joint Bayesian Analysis of Sub-Phenotypes and Epistasis

The rapid advances in genotyping and DNA sequencing technologies have enabled Genome-Wide Association Studies (GWAS) and thereby the discovery of many new associations. However, these associations explain only a small proportion of the theoretically estimated heritability of most diseases. This problem is referred to as 'missing heritability' and is a substantial obstacle on the way to personalized medicine. We propose the JBASE algorithm that simultaneously accounts for two of the main problems of missing heritability: epistasis (interaction between genetic variants) and disease heterogeneity. JBASE extends the probabilistic model of BEAM by explicitly modeling the hidden phenotypic differences within a given population and the associated causal variants. We have tested our system on Type 2 Diabetes finding some of the novel and potentially interesting sub-populations. It is our hope that our method can lead to the refinement of disease diagnosis and thus development of more personalized treatment ultimately resulting in better healthcare.



It is our hope that our method can lead to the refinement of disease diagnosis and thus development of more personalized treatment ultimately resulting in better healthcare.

PROJECT TEAM

**Recep Colak**, Graduate Student  
**Anna Goldenberg**, Scientist (Genetics and Genome Biology program, SickKids)

## 21 Towards Better and Faster Diagnostics of Genetic Diseases

Diagnosing genetic disorders is a complex and time consuming task for geneticists, requiring one to sift through a large number of variants per sequenced individual in search for the ones responsible for their symptoms. Our group is designing and developing tools that assist the geneticists in their work and help accelerate the identification of disease-causing genetic variants. From user friendly GUIs that make gathering and visualizing patient information easier, to clever algorithms that are able to do a reliable prioritization of genetic variants and direct the geneticist's attention to a small subset more likely to be harmful, these tools aim at improving significantly the time and effort specialists need to put in quality genetic diagnostics.

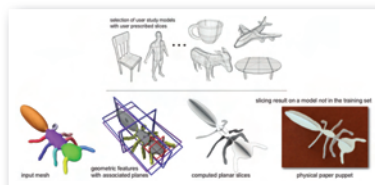


PROJECT TEAM

**Marc Fiume**, Graduate Student  
**Michael Brudno**, Faculty

## 22 Slices: A Shape-Proxy Based on Planar Sections

Based on a user study, where participants abstracted meshes of common objects using a small collection of planar slices, we develop an automatic algorithm to create planar slice based abstractions of (untrained) models. Starting from a set of planar slices, approximating the object's geometric features, the algorithm picks a subset of planes based on feature importance learned from the user study. A second user study verifies that the planar slice abstractions are as easily recognizable as the original models.

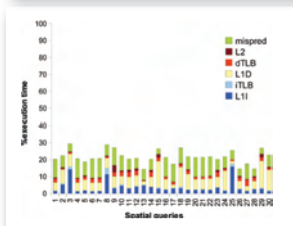
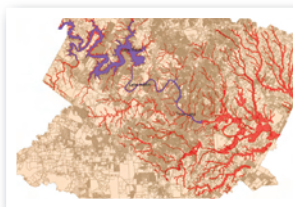


PROJECT TEAM

**James McCrae**, Graduate Student  
**Karan Singh**, Faculty

## 23 Speeding up Spatial Database Query Execution using GPUs

Spatial databases are used in a wide variety of real-world applications, such as land surveying, urban planning, and environmental assessments, as well as geospatial Web services. As uses of spatial databases become more widespread, there is a growing need for good performance of spatial applications. However, spatial database workload properties are not well-understood. We analyzed a set of typical spatial queries and characterized them in terms of their resource usage (CPU utilization and I/O activity trends). Our preliminary work suggests that although disk access latency is a significant component of spatial processing, as database buffer cache sizes increase and the workload complexity grows, the main performance bottleneck becomes the CPU. Spatial workloads tend to be computationally-intensive because of the highly-complex geometric processing involved. Furthermore, memory access stalls (due to the ever-increasing processor-memory speed gap) contribute significantly to spatial query execution time. With the advent of massively-parallel graphics-processing hardware (GPUs) and frameworks like CUDA, opportunities for speeding up spatial processing have emerged. GPUs not only benefit from massive parallelism, but they can also better hide the memory latency. We aim to speedup spatial query execution using CUDA and recent generation GPUs. Our GPU implementations of 6 representative queries ran from 62 to 318 times faster compared to CPU-equivalent code. These results show clear potential for huge performance gains, although further work is needed to create a full-fledged parallelized spatial DBMS prototype.



We aim to speedup spatial query execution using CUDA and recent generation GPUs. Our GPU implementations of 6 representative queries ran from 62 to 318 times faster compared to CPU-equivalent code. These results show clear potential for huge performance gains, although further work is needed to create a full-fledged parallelized spatial DBMS prototype.

PROJECT TEAM

**Bogdan Simion**, Graduate Student  
**Angela Demke Brown**, Faculty

## 24 Recon: Declarative Invariants for Runtime Consistency Checking

We have been developing a framework, called Recon, that uses runtime checking to protect the integrity of file-system metadata on disk. Recon performs consistency checks at commit points in transaction-based file systems. We define declarative statements called consistency invariants for a file system, which must be satisfied by each transaction being committed to disk. By checking each transaction before it commits, we prevent any corruption to file-system metadata from reaching the disk. Our first prototype required writing the consistency invariants in C, however using a declarative language to express and check these invariants improves the clarity of the rules, making them easier to reason about, verify, and port to new file systems. We describe how file system invariants can be written and checked using the Datalog declarative language in the Recon framework. Additionally, we show how we can separate Recon entirely from an untrusted OS and monitor its file system activity via a hypervisor.

### PROJECT TEAM

**Jack Sun**, Graduate Student  
**Daniel Fryer**, Graduate Student  
**Angela Demke Brown**, Faculty

## 25 Protecting Kernel Integrity from Untrusted Extensions Using Dynamic Binary Instrumentation

Device drivers are the major source of concern for maintaining security and reliability of an operating system. Many of these device drivers, developed by third parties, get installed in the kernel address space as extensions. These extensions are implicitly trusted and are allowed to interact with each other and the kernel through well-defined interfaces and by sharing data in an uncontrolled manner. Unfortunately, the assumed trust leaves commodity OSes vulnerable to misbehaving and malicious kernel extensions. Our approach uses dynamic binary instrumentation (DBI) to execute arbitrary kernel extensions securely. Our current proof-of-concept aims to enforce control-flow integrity (CFI) constraints in the Linux kernel using our DynamoRIO Kernel (DRK) DBI framework. DRK is an appealing host platform for this work due to its ability to provide fine-grained control over instrumentation. Our prototype is loaded as a kernel module and transparently monitors all control-flow transfers while the kernel is running. Our system is able to detect and report unauthorized control transfers by potentially malicious kernel modules. Building on this, we intend to develop a security system which will protect the kernel from multiple types of malicious activities by untrusted extensions.

### PROJECT TEAM

**Peter Goodman**, Graduate Student  
**Akshay Kumar**, Graduate Student  
(Electrical & Computer Engineering)  
**Angela Demke Brown**, Faculty  
**Ashvin Goel**, Faculty (Electrical & Computer Engineering and CS)

## 26 Jettison: Efficient Idle Desktop Consolidation with Partial VM Migration

The success of the Internet has led to the proliferation of applications that expect always-on semantics, such as instant messengers, Voice over IP clients, and remote desktop access servers. While these applications provide real value to users, researchers have observed that an unintended side effect is that computers (particularly in office environments) are left continuously running even when idle. Unfortunately an idle PC consumes close to 60% of the power of a fully utilized system. Our work introduces Jettison, a software system that encapsulates desktop sessions inside virtual machines (VMs) and efficiently migrates idle VMs to consolidation servers, allowing applications to continue to run while idle desktops are powered off. Jettison employs partial VM migration, a novel technique that migrates only the working set of the idle VM to the server, and on user activity, migrates only modified state back to the desktop. The working set constitutes a small fraction of the VM's memory and disk state. Partial VM migration has the benefit that both, the network and server infrastructure, can scale well with the number of users, while providing migration times that are very small. Our experimental deployment demonstrates Jettison's ability to reduce energy used by idle desktop systems by up to 90%, while maintaining low migration latencies under five seconds over a 1Gbps link, even in the presence of hundreds of users in the network.

### PROJECT TEAM

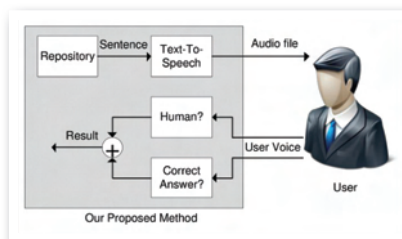
**Nilton Bila**, Graduate Student  
**Eyal de Lara**, Faculty

## 27 Repeat After Me, "I am a Human": Verifying Human Users in Speech-Based Interactions

Verifying that a live human is interacting with an automated speech based system is needed in some applications such as biometric authentication and detection of spam callers which is a major threat for VOIP (Voice Over IP) networks. We designed and implemented a system to verify that the user is human. Simply stated, our method asks the user to repeat a sentence. The reply is analyzed to verify that it is the requested sentence and said by a human, not a speech synthesis system. Our method is taking advantage of both speech synthesizer and speech recognizer limitations to detect computer programs, which is new, and potentially more accessible, way to develop CAPTCHA systems. Using an acoustic model trained on voices of over 1000 users, our system can verify the user's answer with 98% accuracy and with 80% success in distinguishing humans from computers.

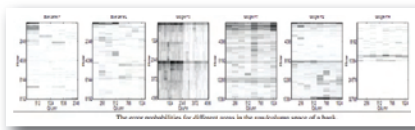
### PROJECT TEAM

**Sajad Shirali-Shahreza**, Graduate Student  
**Yashar Ganjali**, Faculty



## 28 Understanding the Nature of DRAM Errors and the Implications for System Design

Main memory is one of the leading hardware causes for machine crashes in today's datacenters. Designing, evaluating and modeling systems that are resilient against memory errors requires a good understanding of the underlying characteristics of DRAM errors in the field. While there have recently been a few studies on DRAM errors in production systems, these have been too limited in either the size of the dataset or the granularity to conclusively answer many of the open questions. Such questions include, for example, the prevalence of soft errors compared to hard errors, or the analysis of typical patterns of hard errors. In this work, we study data on DRAM errors collected on a diverse range of production systems in total covering nearly 300 terabyte-years of main memory. As a first contribution, we provide a detailed analytical study of DRAM error characteristics, including both hard and soft errors. We find that a large fraction of DRAM errors in the field can be attributed to hard errors and we provide a detailed analytical study of their characteristics. As a second contribution, our work uses the results from the measurement study to identify a number of promising directions for designing more resilient systems and evaluates the potential of different protection mechanisms in light of realistic error patterns. One of our findings is that simple page retirement policies might be able to mask a large number of DRAM errors in production systems, while sacrificing only a negligible fraction of the total memory in the system.



PROJECT TEAM  
**Andy Hwang**, Graduate Student  
**Ioan Stefanovici**, Graduate Student  
**Bianca Schroeder**, Faculty

## DATABASES

### 29 BibBase and Linked Open Data on the Web

The currently-live web application BibBase makes it easy for scientists to maintain their publications pages and facilitates the dissemination of scientific publications over the Internet as part of a linked open data cloud. As a scientist, you simply maintain a BibTeX-file of your publications, including links to the papers, and BibBase does the rest. When a web user visits your publications page, BibBase dynamically generates an always up-to-date HTML page from the BibTeX file, and allows the user to sort the publications by other than the default ordering (e.g. year, author, keywords, research area, and publication type). The exponential growth of digital information is a big challenge faced by large enterprises. Integration of large amounts of data from many heterogeneous sources is currently a complicated task that often requires expert users to set up complex software or manually analyze the data. Cornerstone to more efficient and optimal solutions to the data mining and data cleaning problems we face today is availability of linked data. Linked open data is structured data published on the web to enable structured querying and integration of web data into rapid business decision-making. We automate easy publishing of rich, semantically-linked, de-duplicated data. We also provide tools to find high quality links between internal BibBase entities, as well as discovering links to external data sources.

PROJECT TEAM  
**Nataliya Prokoshyna**, Graduate Student  
**Renée J. Miller**, Faculty

## 30 TAGLab - Technologies for Aging Gracefully Lab

TAGLab is comprised of talented individuals with backgrounds in computer science, engineering, human-computer interaction, graphic and interface design, and psychology. We work with researchers and clinicians to find ways that digital media can help people remain vigorous and independent, strengthen ties to family and community, and preserve their identity as they age. Our mission is R&D in support of aging throughout the life course. We identify "sweet spots" where technology seems relevant to human need, envision ways in which the technology could address those needs, then design and test prototypes.

Our researchers will be present to discuss:  
**VocabNomad**: A mobile application to improve the vocabulary acquisition process of second language learners  
**ALLT**: An accessible, large-print, listening, and talking e-book reading system



PROJECT TEAM  
**Mike Massimi**, Graduate Student  
**Jessica David**, Graduate Student  
**Carrie Demmans Epp**, Graduate Student  
**Abbas Attarwala**, Graduate Student  
**Velian Pandeliev**, Graduate Student  
**Ronald Baecker**, Faculty

## 31 Improv Remix

Musicians can play multiple instruments simultaneously by playing a single instrument, and then looping that recording as they play other instruments. In this project, we give this ability to theatrical improvisors by creating an on-stage gestural video editing interface. Improv Remix investigates gestural interfaces that can be used while in-character, as well as a new form of creative expression. The interface uses a Kinect and a projector.



PROJECT TEAM  
**Dustin Freeman**, Graduate Student  
**Ravin Balakrishnan**, Faculty

## 32 Reinventing the (Mouse) Wheel

Why is the default middle-mouse button behaviour a rate-controlled scroll? Is this really the best tool to help users navigate large documents? How can we make the middle-mouse button not suck? These are the questions that led us to explore the design space of desktop document navigation. We have developed a novel interaction technique that allows users to navigate a document spatially, through its physical structure, but also abstractly, through its logical structure, providing a much more useful middle-mouse button experience.

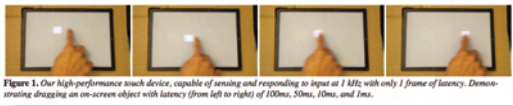
PROJECT TEAM  
**Michael Glueck**, Graduate Student  
**Fanny Chevalier**, Postdoctoral Fellow  
**Daniel Wigdor**, Faculty

### 33 Experimentation with a 1ms Latency Direct-Touch Device

Software designed for direct-touch interfaces often utilize a metaphor of direct physical manipulation of pseudo “real-world” objects. However, current touch systems typically take 50- 200ms to update the display in response to a physical touch action; our system, built by colleagues at Microsoft, operates at latency levels as low as 1ms. We are exploring the design and experimental space around high-performance touch, in terms of both human perception and the design questions that arise from a system of this nature.

PROJECT TEAM

**Julian Lepinski**, Graduate Student  
**Daniel Wigdor**, Faculty



### 34 Using Gestures and Speech for K-2 Education

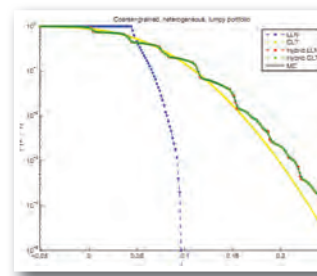
Using gestures and speech for K-2 education The HCI course project was implemented as a proof-of-concept, for demonstrating the combined use of gestures and speech to simplify complex learning and make classroom education interactive and fun. The multimodal, interactive, game-like interface is ideal for smaller children in a learning environment.

PROJECT TEAM

**Uzma Khan**, Graduate Student  
**Daniel Wigdor**, Faculty

### 36 Approximations to Loss Probabilities of Credit Portfolios

Credit risk analysis and management at the portfolio level are challenging problems for financial institutions due to their portfolios' large size, heterogeneity and complex correlation structure. The conditional independence framework is widely used to calculate loss probabilities of credit portfolios. The existing computational approaches within this framework fall into two categories: (1) simulation-based approximations and (2) asymptotic approximations. The simulation-based approximations often involve a two-level Monte Carlo method, which is extremely time-consuming, while the asymptotic approximations, which are typically based on the Law of Large Number (LLN), are not accurate enough for tail probabilities, especially for heterogeneous portfolios. We give a more accurate asymptotic approximation based on the Central Limit Theorem (CLT), and we discuss when it can be applied. To further increase accuracy, we also propose a hybrid approximation, which combines the simulation-based approximation and the asymptotic approximation. We test our approximations with some artificial and real portfolios. Numerical examples show that, for a similar computational cost, the CLT approximation is more accurate than the LLN approximation for both homogeneous and heterogeneous portfolios, while the hybrid approximation is even more accurate than the CLT



approximation. Moreover, the hybrid approximation significantly reduces the computing time for comparable accuracy compared to simulation-based approximations.

PROJECT TEAM

**Meng Han**, Graduate Student  
**Ken Jackson**, Faculty

## NUMERICAL ANALYSIS

### 35 Parameter Determination in Mathematical

In investigating the behaviour, over time, of a complex systems one often introduces a mathematical model which represents the system by a parametrised set of ordinary differential equations. A key associated task is then to determine the specific parameter values that best describe a particular system under observation. For example when modelling the behaviour of an emerging epidemic, one must rapidly determine key parameters such as infection rate(s), incubation period, immunity period, and initial population of infected individuals, in order to develop an effective vaccination and management strategy (to control the outbreak). We will discuss how accurate and reliable simulations of the underlying mathematical model can be combined with an optimization technique from machine learning to produce a very effective and robust approach for determining the optimal parameters.

PROJECT TEAM

**Bo Wang**, Graduate Student  
**Wayne Enright**, Faculty

## SOFTWARE ENGINEERING

### 37 Partial Models: Towards Modeling and Reasoning with Uncertainty

Models are good at expressing information about software but not as good at expressing the modeler's uncertainty about the information in the model. Yet the highly incremental and iterative nature of software development requires the ability to express uncertainty and reason with models containing it. We build on our earlier work on expressing uncertainty using partial models, by elaborating an approach to reasoning with such models. We evaluate our approach by experimentally comparing it to traditional strategies for dealing with uncertainty as well as by conducting a case study using open source software. We conclude that we are able to reap the benefits of well-managed uncertainty while incurring minimal additional cost.

PROJECT TEAM

**Michalis Famelis**, Graduate Student  
**Marsha Chechik**, Faculty

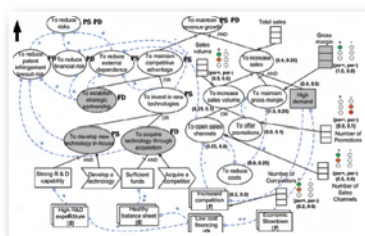
## 38 Who's Flying this Ship? Systems Thinking for Planet Earth

Given our human population has reached 7 billion people, the impact of humanity on the planet is vast: we use nearly 40% of the earth's land surface to grow food, we're driving other species to extinction at a rate not seen since the last ice age, and we've altered the planet's energy balance by changing the atmosphere. In short, we've entered a new geological age, the Anthropocene, in which our collective actions will dramatically alter the inhabitability of the planet. We face an urgent task: we have to learn how to manage the earth as a giant system of systems, before we do irreparable damage. Through this poster, we will describe some of the key systems that are relevant to this task, including climate change, agriculture, trade, energy production, and the global financial system. We will explore some of the interactions between these systems, and characterize the feedback cycles that alter their dynamics and affect their stability. We will discuss a framework for thinking about the leverage points that may allow us to manage these systems.

PROJECT TEAM  
**Alicia M. Grubb**, Graduate Student  
**Steve Easterbrook**, Faculty

## 39 Business Intelligence Modeling and Reasoning

Business intelligence offers tremendous potential for gaining insights into day-to-day business operations, as well as longer term opportunities and threats. However, much of today's BI technologies and tools are based on models that are too IT-oriented from the point of view of business decision makers. We propose an enterprise modeling approach to bridge the business level understanding of the enterprise with its IT representations in databases and data warehouses. The Business Intelligence Model (BIM) offers concepts familiar to business decision making - such as goals, strategies, processes, situations, influences, and indicators. Unlike many enterprise models which are meant to be used to derive, manage, or align with IT system implementations, the BIM aims to help business users organize and make sense of the vast amounts of data about the enterprise and its external environment. In this work, we focus especially on reasoning about situations, influences, and indicators. Reasoning with such concepts supports analysis of business objectives in light of current enterprise data, allowing analysts to explore scenarios and find alternative strategies. We describe how goal reasoning techniques from conceptual modeling and requirements engineering can be applied to BIM. Techniques are provided to allow reasoning with indicators linked to business metrics, including cases where specification of business metrics and indicators are incomplete. Evaluation of the proposed modelling and reasoning framework includes a prototype implementation, as well as an on-going case study.



PROJECT TEAM  
**Jennifer Horkoff**, Postdoctoral Fellow  
**John Mylopoulos**, Faculty

## 40 MarkUs

MarkUs is an open-source tool which recreates the ease and flexibility of grading assignments with pen on paper, within a web application. It also allows students and instructors to form groups, and collaborate on assignments. As students submit their work, MarkUs keeps track of the versions they submit. Graders annotate students' code and assign grades directly in the web application. Instructors can monitor the progress of the graders and can easily release the results to the students. MarkUs is built and maintained entirely by undergraduate students under the supervision of faculty members and former MarkUs developers. It is used by thousands of students in three universities in Canada and France.



PROJECT TEAM  
**Sean Budning**, Undergraduate Student  
**Danesh Dadachanji**, Undergraduate Student  
**Severin Gehwolf**, Undergraduate Student  
**Tobi Ogunbiyi**, Undergraduate Student  
**Jay Parekh**, Undergraduate Student  
**Karen Reid**, Faculty

## 41 Exploring Contextual Complexities of Agile Adoption

Agile methods are often proposed as a set of practices, from which development teams pick up their selected subset. One of the key challenges in the process of deciding whether to select an agile practice is to know in advance the potential impacts of that practice within the organization. In earlier work we introduced an evidence-based repository of agile practices. In this work we explain an experience of using the repository to guide and support the process of decision making over the enactment of a particular agile practice in a medium-scale software company.

PROJECT TEAM  
**Hesam Chiniforooshan**, Graduate Student  
**Eric Yu**, Faculty of Information and CS

NOTES

---



---



---



---



---



---

## 42 Iterative, Interactive Analysis of Agent-Goal Models for Early Requirements Engineering

Conceptual modeling allows abstraction, communication and consensus building in system development. It is challenging to expand and improve the accuracy of models in an iterative process, producing models able to facilitate analysis. Modeling and analysis can be especially challenging in early Requirements Engineering (RE), where high-level system requirements are discovered. In this stage, hard-to-measure non-functional requirements are critical; understanding the interactions between systems and stakeholders is a key to system success. Goal models have been introduced as a means to ensure stakeholder needs are met in early RE. Because of the high-level, social nature of early RE models, it is important to provide procedures which prompt stakeholder involvement (interaction) and model improvement (iteration). Most current approaches to goal model analysis require quantitative or formal information that is hard to gather in early RE, or produce analysis results automatically over models. Approaches are needed which balance automated analysis over complex models with the need for interaction and iteration. This work develops a framework for iterative, interactive analysis for early RE using agent-goal models. We survey existing approaches for goal model analysis, providing guidelines using domain characteristics to advise on procedure selection. We define requirements for an agent-goal model framework specific to early RE analysis, using these requirements to evaluate the appropriateness of existing work and to motivate and evaluate the components of our analysis framework. We provide a detailed review of forward satisfaction procedures, exploring how different model interpretations affect analysis results. A survey of agent-goal variations in practice is used to create a formal definition of the  $i^*$  modeling framework which supports sensible syntax variations. This definition is used to precisely define analysis procedures and concepts throughout the work. The framework consists of analysis procedures, implemented in the OpenOME requirements modeling tool, which allow users to ask “What if?” and “Is this goal achievable, and how?” questions. Visualization techniques are introduced to aid analysis understanding. Consistency checks are defined over the interactive portion of the framework. Implementation, performance and potential optimizations are described. Group and individual case studies help to validate framework effectiveness in practice. Contributions are summarized in light of the requirements for early RE analysis. Finally, limitations and future work are described.

PROJECT TEAM

**Jennifer Horkoff**, Postdoctoral Fellow  
**Eric Yu**, Faculty of Information and CS



## 43 Algorithms in Trading

The electronic exchange is the largest “market” in the modern world, where you can trade stocks, options, bonds, futures and forex; that means opportunities every day. Trading by human is error-prone, tiresome and time consuming. Algorithmic trading is a developing young field that so far is the privilege of large institutions and hedge funds. However, with the development of various trading platform and APIs, algorithmic trading is becoming more accessible to retailer investors and small firms. We are in the process of building an algorithmic trading platform that utilizes various algorithmic ideas and machine learning techniques to do automatic trading. This involves signal detections, trading strategies, order management, data organization and risk management. As you can imagine this requires a lot of new ideas and research in various areas of computer science such as theory, AI, database, networking and software engineering. The ideas being presented come from a course project, and are still at a very early stage.

PROJECT TEAM

**Yuli Ye**, Graduate Student  
**Allan Borodin**, Faculty

## 44 Equivalence of Memory Consistency Models with Relaxed Program Order

Many different memory consistency models have been proposed for multiprocessor systems. We consider models that can be classified according to how reads and writes are reordered. We provide the first formal proof that programs that are correct for a model that does not allow reorderings of reads or writes during execution are also correct for a model that allows reads and writes to be executed after reads to different memory locations that follow them in program order. We also show that this is not necessarily true for models that allow reads or writes to be executed after writes to different memory locations that follow them in program order.

PROJECT TEAM

**Oles Zhulyn**, Graduate Student  
**Faith Ellen**, Faculty

## 45 Encryption Systems for the Modern World

For a very long time, the term “encryption” used to be synonymous with the problem of Secure Communication – how to transmit a piece of data privately over a public channel. Not any more. Not in the modern world of Cloud Computing, Smartphones and Social Networks – a world where sharing of personal information is so ubiquitous that access to data will be governed by complex privacy policies. This new digital reality calls for nuanced notions of encrypting data that enable access to different parts of data to different parties, depending on a potentially intricate privacy policy. We highlight our work on a modern notion of encryption called “Functional Encryption” that enables such fine-grained privacy controls, and more.

PROJECT TEAM

**Sergey Gorbunov**, Graduate Student  
**Vinod Vaikuntanathan**, Faculty

# CONTACT LIST

LAST NAME	FIRST NAME	EMAIL	PROJECT NUMBER	LAST NAME	FIRST NAME	EMAIL	PROJECT NUMBER
Attarwala	Abbas	attarwal@cs.toronto.edu	30	Hwang	Andy	hwang@cs.toronto.edu	28
Baecker	Ronald M.	ron@taglab.ca	30	Jackson	Ken	krj@cs.toronto.edu	36
Balakrishnan	Ravin	ravin@dgp.toronto.edu	31	Karamanov	Nikola	karamano@cs.toronto.edu	15
Beck	Chris	jcb@mie.utoronto.ca	9	Kazemian	Siavash	kazemian@cs.toronto.edu	3
Bila	Nilton	nilton@cs.toronto.edu	26	Khan	Uzma	uzmakhan@cs.toronto.edu	34
Borodin	Allan	bor@cs.toronto.edu	43	Kumar	Akshay	iit.akshay@gmail.com	25
Boutillier	Craig	celby@cs.toronto.edu	5, 6, 7, 17	Lepinski	Julian	lepinski@dgp.toronto.edu	33
Bretschneider	Hannes	hannes@psi.utoronto.ca	10	Lu	Tyler	tl@cs.toronto.edu	6
Brudno	Michael	brudno@cs.toronto.edu	21	Massimi	Mike	mike@taglab.ca	30
Budning	Sean	sbudning@gmail.com	40	McCrae	James	mccrae@dgp.toronto.edu	22
Charlin	Laurent	lcharlin@cs.toronto.edu	17	McIlraith	Sheila	sheila@cs.toronto.edu	8, 9
Chechik	Marsha	chechik@cs.toronto.edu	37	Mengistu	Kinfe Tadesse	kinfe@cs.toronto.edu	1
Chevalier	Fanny	fanny@dgp.toronto.edu	32	Miller	Renée	miller@cs.toronto.edu	29
Chiniforooshan	Hesam	hesam@cs.toronto.edu	41	Mnih	Volodymyr	vmnih@cs.toronto.edu	12
Colak	Recep	rcolak@cs.toronto.edu	20	Muise	Christian	cjmuise@cs.toronto.edu	9
David	Jessica	j david@cs.toronto.edu	30	Mylopoulos	John	jm@cs.toronto.edu	39
Dadachanji	Danesh	g9dadach@cdf.toronto.edu	40	Nematzadeh	Aida	aida@cs.toronto.edu	4
de Lara	Eyal	delara@cs.toronto.edu	26	Norouzi	Mohammad	norouzi@cs.toronto.edu	18
Demke Brown	Angela	demke@cs.toronto.edu	23, 24, 25	Ogunbiyi	Tobi	t.ogunbiyi@gmail.com	40
Demmans Epp	Carrie	c demmans@cs.toronto.edu	30	Parekh	Jay	j.parekh@utoronto.ca	40
Dickinson	Sven	sven@cs.toronto.edu	19	Pandeliev	Velian	vpandeli@cs.toronto.edu	30
Easterbrook	Steve	sme@cs.toronto.edu	38	Penn	Gerald	gpenn@cs.toronto.edu	1, 2, 3
Eisner	Daniel	lordzuon@gmail.com	13	Prokoshyna	Nataliya	nataliya@cs.toronto.edu	29
Ellen	Faith	faith@cs.toronto.edu	44	Regan	Kevin	kmregan@cs.toronto.edu	5
Engels	Steve	sengels@cs.toronto.edu	13	Reid	Karen	reid@cs.toronto.edu	40
Enright	Wayne	enright@cs.toronto.edu	35	Rudzicz	Frank	frank@cs.toronto.edu	2
Famelis	Michalis	famelis@cs.toronto.edu	37	Schroeder	Bianca	bianca@cs.toronto.edu	28
Fazly	Afsaneh	afsaneh@cs.toronto.edu	4	Shirali-Shahreza	Sajad	shirali@cs.toronto.edu	27
Fiume	Marc	mfiume@cs.toronto.edu	21	Simion	Bogdan	bogdan@cs.toronto.edu	23
Fleet	David	fleet@cs.toronto.edu	18	Singh	Karan	karan@dgp.toronto.edu	22
Freeman	Dustin	dustin@cs.toronto.edu	31	Snoek	Jasper	jasper@cs.toronto.edu	16
Frey	Brendan	frey@psi.utoronto.ca	10	Sohrabi	Shirin	shirin@cs.toronto.edu	8
Fryer	Daniel	dfryer@cs.toronto.edu	24	Stevenson	Suzanne	suzanne@cs.toronto.edu	4
Ganjali	Yashar	yganjali@cs.toronto.edu	27	Sui	Xin	xsui@cs.toronto.edu	7
Gavrilov	Zoya	zoya.gavrilov@utoronto.ca	19	Sun	Jack	kuei.sun@utoronto.ca	24
Gehwolf	Severin	severin.gehwolf@utoronto.ca	40	Stefanovici	Ioan	ioan_stefanovici@cs.toronto.edu	28
Glueck	Michael	mgglueck@dgp.toronto.edu	32	Sutskever	Ilya	ilya@cs.toronto.edu	11
Goel	Ashvin	ashvin@eecg.toronto.edu	25	Swersky	Kevin	kswersky@cs.toronto.edu	14
Goldenberg	Anna	anna.goldenberg@utoronto.ca	20	Tarlow	Danny	dtarlow@cs.toronto.edu	15
Goodman	Peter	pag@cs.toronto.edu	25	Vaikuntanathan	Vinod	vinodv@cs.toronto.edu	45
Gorbunov	Sergey	sgorbunov@cs.toronto.edu	45	Wang	Bo	bevis@cs.toronto.edu	35
Grubb	Alicia	amgrubb@cs.toronto.edu	38	Wigdor	Daniel	dwigdor@dgp.toronto.edu	32, 33, 34
Han	Meng	m7han@cs.toronto.edu	36	Ye	Yuli	y3ye@cs.toronto.edu	43
Hinton	Geoffrey	hinton@cs.toronto.edu	11, 12	Yu	Eric	eric@cs.toronto.edu	41, 42
Hirst	Graeme	gh@cs.toronto.edu	1	Zemel	Richard	zemel@cs.toronto.edu	14, 15, 16, 17
Horkoff	Jennifer	jenhork@cs.toronto.edu	39, 42	Zhulyan	Oles	oles@cs.toronto.edu	44

## contact us

For more information on  
any of these projects:

**SARA FRANCA**

Senior Communications &  
External Relations Officer

Department of Computer Science  
10 King's College Road, Room 3302

Toronto, Ontario M5S 3G4

Telephone: 416-978-3619

E-mail: [sfranca@cs.toronto.edu](mailto:sfranca@cs.toronto.edu)

Website: [www.cs.toronto.edu](http://www.cs.toronto.edu)

In partnership with



**NSERC  
CRSNG**

