AI and Society: A vision for artificial intelligence research at University of Waterloo

Articial intelligence algorithms and systems are becoming increasingly integrated with our society, automating tasks and assisting humans with decisions. This deep integration of AI and humans have broad implications for both the well-being of individuals and the health of our society. Moving forward, the AI group at the David R. Cheriton School of Computer Science envision research, teaching and community building activities that tightly integrate fundamental scientific research, their application to immediate societal problems, and careful consideration of the ethical and societal implications of such innovation.

I. Research

Our research focuses on ensuring that AI systems have certain desired properties. AI systems should be

- **Safe.** Currently, the only performance guarantees one can provide for ML based decisions are statistical. For critical decisions, such as controlling an autonomous vehicle, robotic-assisted surgery or autonomous weapons, the high cost of an error makes such guarantees unsatisfactory. **Prof. Ben-David** is developing ML algorithms that would be able to ”raise a flag” when they encounter a situation that is outside their high confidence regions.
- **Fair.** As machine learning starts being applied to advice decisions relating to people (such as acceptance to university programs, judicial bail decisions, or granting mortgages) there are growing concerns about the tendency of such systems to use racial, gender or religion as features influencing their decisions. The issue of fairness of ML based decisions is complex and poorly understood. **Ben-David** studies how fairness should be defined and addresses and what utility tradeoffs are involved in taking it into consideration. **Larson** studies design and analyze voting systems in order to better understand how people’s lack of information and bounded rationality influence voting behaviour and outcomes. She explores concepts like ”fairness” in computational settings, designing systems which balance the incentives of the users with quality and fairness of the outcomes. Her work also explicitly try to reason with people’s preferences, while accounting for the fact that people often are not able to clearly articulate them.
- **Interpretable.** Making decisions and actions of machine learning algorithms transparent can allow people to detect sources of uncertainty, potential errors and biases of such systems. **Law** has some recent work involving human annotators deliberating ambiguous cases in text classification, and how one can mine such data arising from disagreement to generate more interpretable features for machine learning.
- **Ethical.** AI systems will increasingly make ethical decisions and moral tradeoffs. **Hoey** and **Cohen** are exploring how human-like moral decision-making can be built into a neural network system.
- **Trustworthy.** **Cohen** has done research on trust modeling in multi-agent systems, social media environments (e.g., for combating digital misinformation and addressing anti-social behaviour), and online marketplaces (e.g., collusion). She is also studying the dimensions that dene trusted AI, i.e., what are the concerns about AIs trustworthiness and how research has addressed these concerns. **Law** has explored the issue of trust and engagement in the context of a robot interacting with users to sort objects into trash versus recycling.
- **Involving Humans in the Loop.** There is now an explosion of data to be mined in business, in science, in law, and in health care at a scale that is beyond the analytic capabilities of a single person and at a level of complexity that challenges even the most sophisticated algorithms. At the same time, human intelligence is massively distributed and now readily accessible, yet untapped: there are millions of people online each day, performing computational tasks as a by-product of searching for information, playing games, organizing personal data collections, and interacting with communities. **Law**’s research in human computation explores how we can combine human and machine intelligence to tackle complex problems,

such as those requiring expertise or intricate dependencies. **Grossman**'s research explores how machine learning can be applied in the context of eDiscovery in legal proceedings, to identify relevant evidence with greater effectiveness and efficiency that lawyers working on their own.

Beyond these fundamental research questions, the AI group also has a track record of applying fundamental research to real-world problems, related to

- **Environment. Larson** have looked at applying computational and game-theoretic techniques to resource allocation problems arising in wildfire control in Canada. **Van Beek** has partnered with The City of Abbotsford, BC, to optimize the conservation of water resources. His system is designed to predict the water consumption at the hourly and daily level using demographic and weather information, and infer where water is consumed (e.g., indoor versus outdoor).
- **Healthcare. Cohen** has been a member of two different Strategic Networks with a focus on health-carehSITE (using technology for healthcare solutions) and CANet (focused on heart arrythmia research). She has done research on resource allocation in hospital and mass casualty scenarios, as well as ontologies for health care decision making. **Law** is leading a multi-year NSERC-CIHR Collaborative Health Research Project (CHRP) project to design a framework for hybrid machine and human computation to achieve accurate and scalable analysis of human clinical EEG recordings. Electroencephalography (EEG), i.e., signals of brain wave activities, is a key tool in the diagnosis of epilepsy and sleep disorders. **Grossman** has studied the use of machine learning for systematic reviews in evidence-based medicine. She is beginning to explore the use of machine learning to identify and decrease the impact of medical misinformation on the web; Cohen has explored the value of multiagent trust modeling for coping with healthcare discussion board misinformation as well.
- **Assistive Technology. Poupart** is most well known for his contributions to the development of approximate scalable algorithms for partially observable Markov decision processes (POMDPs) and their applications in real-world problems, including automated prompting for people with dementia for the task of handwashing. Other notable projects include stress detection based on wearable devices, and mobile assistive technology for voice, activity and location monitoring of dementia patients. **Hoey** works on building assistive technologies for persons with cognitive disabilities and in particularly for aging. He is a network investigator for the AGEWELL networks of centers of excellence to work on technologies for aging, and has funding from the American Alzheimers Association and the Canadian Consortium on Neurodegeneration and Aging (CCNA). **Cohen** is conducting research that leverages AI solutions (Bayesian reasoning and computer vision algorithms) to improve the online experience of users with assistive needs, e.g., decluttering or zooming, including one study on older adult users and another on improving screen-reader output for those with visual impairment.
- **Virtual Assistants. Poupart** has worked on chatbots for automated personalized conversations and spoken dialogue management. **Hoey**'s affective computing research focuses on understanding social and emotional factors in intelligence, building them into artificial agents, and investigating their interactions in social network and group settings. **Law** is leading a project on teachable robots, and their use in educational settings to enhance student curiosity. **Yu**'s research looks at facilitating individuals or groups to better catalogue their digital life, e.g., extract memorable moments from everyday video recordings.
- **Computational Social Science. Hoey** leads a multinational project funded by the Trans-Atlantic Partnership on social dynamics in online collaborative groups. The project is producing data-driven theoretical insights into what motivates self-organized collaborations and what determines their success, empirical validation of sociological theory and formal answers to important social science questions about collaboration.

- **Neuroscience. Orchard** develops neural learning algorithms that are biologically plausible, to inform our understanding of neuro-cognitive disorders (e.g., Schizophrenia, Autism). He develops neural network architectures that are robust against adversarial or ambiguous inputs.
- **International Development.** About six billion people live in developing countries, but many of the solutions appropriate for developed countries cannot be applied to the developing world. For example, in developed countries, major credit reporting agencies calculate the credit score of customers based on their credit history, but in many emerging markets, the cash economy is more prevalent and many potential customers have no credit history at all. (In fact, the World Bank estimates 61% of people in Latin America are outside the formal financial system.) **Ghodsi** is interested in how AI can be adapted for use in the developing world. To do so, the AI methods used to address problems in those regions (including problems to do with health, education, the environment, etc.) must take into account the specific limitations of those regions, especially in terms of the availability of computation power, internet, data, etc. Some examples of this kind of work would include the creation and computation of new economic indicators for emerging markets in the developing world and telemedicine in regions with limited resources.
- **Citizen Science.** Science is increasingly data-intensive; yet, many research tasks involving the collection, annotation and analysis of data are not yet fully automated by computers. The idea of citizen science is to engage massive number of people over the Web to help collect, annotate and analyze scientific artefacts. **Law** has developed a citizen science platform called CrowdCurio (http://crowdcurio.com). On this platform, each project addresses a real-world data collection or processing need of a scientist, while serving as a testbed for studying crowdsourcing questions related to incentive, task decomposition, expert-novice interactions and hybrid human-machine computation. Her efforts have led to multiple publications with scientists (e.g., in Ecology).

II. Teaching

The AI group offers a number of courses related to the theme of “AI and society”:

CS798: Human-Centric Machine Learning (Ben-David)
CS492: Social Implications of Computing (Cohen)
CS898: Technological Solutions for Social Problems of Computing (Cohen)
CS886: AI and Philosophy (Cohen)
CS886: Trust and Online Social Networks (Cohen)
CS489/798: Artificial Intelligence: Law, Ethics, and Policy (Grossman)
CS886: Affective Computing (Hoey)
CS889: Human-in-the-Loop Systems (Law)
CS889: Human-AI Interaction (Law)

Grossman is also teaching Ethics Module/Workshop in the MDSAI and Data Science Specialization programs, as of Winter 2020. The AI group has also discussed a few possible training initiatives to strengthen students understanding and involvement with AI and society research. One idea is to introduce a *Computing and Society Option* to the CS program. Another idea is to introduce an undergraduate research initiative, where students are given a mission (i.e., a societal problem to solve) and have to design an AI project to tackle the problem over two years.

III. Community Building

Workshop. To strengthen the research community around the “AI and society” theme, we propose to organize a workshop in April 2019 to bring together invited speakers to exchange research ideas and discuss challenges related to the integration of AI into everyday life.

Platforms. We propose that the Institute hosts high impact international research platforms. As an example, consider a platform for AI-powered accelerated health research. This platform would essentially be a virtual lab in which health researchers around the world could design an experiment, recruit thousands of remote participants and analyze the data collected with ML techniques effortlessly. Such platform can benefit from funding from the AI institute to hire technical staff to maintain the infrastructure and coordinate the on-boarding of health research partners. There can be other similar platforms related to citizen science, assistive technologies, environmental monitoring (e.g., firefighting), etc.