MMPA

Master of Management & Professional Accounting

Institute for Management & Innovation
UNIVERSITY OF TORONTO
MISSISSAUGA

PROCEEDINGS OF THE 2019 MMPA CONFERENCE

ARTIFICIAL INTELLIGENCE: TECHNOLOGY, GOVERNANCE & SOCIAL INNOVATION

Held on November 15, 2019
Institute for Management & Innovation
Innovation Complex, Kaneff Centre
University of Toronto Mississauga
3359 Mississauga Road

Table of Contents

Foreword — CPA Canada

Michael Wong and Davinder Valeri

Artificial intelligence (AI) and machine learning are poised to transform the business world. Companies across the globe are all trying to innovate and redefine their business models with AI. They want to use AI to provide differentiated customer experiences, improve productivity, and create new product categories. Much like these companies, the CPA profession needs to innovative and adapt to this new AI-enabled world.

As professional accountants, we have all heard how AI may threaten to eliminate or replace many traditional accounting functions. Or how AI is making historical data less relevant as decisions become driven by real-time information. Yet those who understand the benefits and limitations of AI will instead see opportunities. AI, for example, enables entire data sets to be analyzed in minutes. This is positive progress for the accountant as it creates opportunity for CPAs to harness this technology to provide more efficient and effective services to clients.

Through CPA Canada initiatives like *Foresight: Reimagine the Profession* and our existing thought leadership resources on AI[1], CPA Canada is dedicated to developing resources that prepare the profession for tomorrow. So, we are pleased to support the *Artificial Intelligence: Technology, Governance, and Social Innovation* conference and the associated conference papers gathered here by the Master of Management & Professional Accounting (MMPA) Program at the Institute for Management & Innovation, University of Toronto.

In this publication, thought leaders:

- in **technology** explain the relationships between key AI terms and how the black box of deep learning impacts bias and fairness of algorithms
- in **governance** show us the risks and opportunities with AI, the importance of establishing proper governance, and the need to use it responsibly
- in **social innovation** introduce us to the Centre for Advancing Responsible & Ethical Artificial Intelligence (CARE-AI) initiative and how it can help address the issues around ethical AI use and unintended consequences.

This conference has shown that AI can improve financial forecasts, automate thinking, and make data-driven recommendations. However, the black box nature of machine learning algorithms does not inspire trust from society. Their ability to find the most optimal solution to a problem, irrespective of ethical considerations, is also worrisome. Thankfully, trust and ethics are core competencies for CPAs.

---

[1] See CPA Canada, *How could automation and AI change the CPA's role? Resources for CPAs* (https://cpacanada.ca/aiandautomation)

As technology adoption accelerates, the role of professional accountants will become even more important. We can help define ethical frameworks to guide AI design and development. We can establish proper governance to ensure that human judgement is not lost in the use of AI. As business leaders, advisors, board and oversight members, CPAs are in a unique position to drive AI innovation in a safe and responsible way that helps and protects the public interest and builds trust in this data-driven new world.

On behalf of CPA Canada, we invite you to read on and join the conversation to help shape the future of the profession.[2]

Michael Wong, CPA, CA
Principal
Research, Guidance and Support
277 Wellington Street West
Toronto ON M5V 3H2
michaelwong@cpacanada.ca

Davinder Valeri, CPA, CA
Director
Research, Guidance and Support
277 Wellington Street West
Toronto ON M5V 3H2
dvaleri@cpacanada.ca

---

[2] Join the *Foresight: Reimagining the Profession* conversation (https://foresight.cpacanada.ca/)

Artificial Intelligence: Technology, Governance, and Social Innovation — an Introduction

***Irene M. Wiecek**, Professor of Accounting, Teaching Stream, University of Toronto*

We need more innovative professional accountants. We need them to be more flexible; we need them to be more future focussed…Be flexible, be nimble, be adaptive. Most importantly, be innovative.

Davinder Valeri, CPA Canada, Opening Remarks

This year the MMPA conference is on Artificial Intelligence (AI).

Increasingly, AI is touching more and more aspects of our lives. And while AI is creating many opportunities, it is also presenting challenges for individuals, companies and, more broadly, for society.

AI and big data are two sides of the same coin. AI needs big data to fuel the machine-learning arena, and we need AI to make sense of the increasingly exponential amounts of big data. That relationship, as well as the increasing digitization trend, is making AI an unstoppable force.

Embracing the inter- and trans- disciplinary mandate of the Institute of Management & Innovation, the conference is organized along three pillars:

1) **Technology.** We need to increase our understanding of AI technology.
   - **Fakhri Karray** provides us with an overview of our *fourth industrial revolution* and helps us understand the relationships between the many associated AI topics that other speakers will cover in detail. His paper provides definitions of AI, machine learning, and deep learning that other papers expand upon.
   - **Richard Zemel** elucidates the black box of deep learning and neural networks and tells us why the layer below the output is so important in reducing bias and increasing fairness.
   - **Parvin Mousavi** explains applications of deep learning in medical diagnosis of cancer and heart disease and how *augmented AI* can democratize healthcare. He also introduces G7 and Canadian Chamber of Commerce views on AI and society.
2) **Governance.** We need to know what to do with AI, how to govern it, and how to use it safely and responsibly.
   - **Miklos Vasarhelyi** provides a synopsis of the AI evolution in accounting and auditing, from continuous audit and *exogenous data* as audit evidence, to an application of neural networks in financial audit. He tempers prognostications of job loss and introduces students to free courses to improve their AI competencies.

- **Anand Rao** provides a big-picture view of why AI is closer than we imagine, but farther away than we think. We learn how businesses can take advantage of AI's opportunities and deal with its risks by employing *responsible* AI.
- **Eric Santor** takes us into the Bank of Canada to give us an economic view of the digital economy, in terms of jobs created — and destroyed — and show how machine learning lowers the cost of *prediction*. He describes how the Bank of Canada leads in applying AI to understand AI, focusing on five key governance issues.

3) **Social Innovation.** We need to know how we can use AI to improve society.

- **Graham Taylor** introduces us to the culture of AI research and how its openness is threatened by malicious deployment of dual-use AI technologies. With the launch of the Centre for Advancing Responsible & Ethical Artificial Intelligence (CARE-AI), we see how a multidisciplinary approach to AI research can advance social innovation.

Increasingly at the heart of big data discussions are issues such as transparency, confidentiality, privacy, and ethics. Ethics is a big piece of this. As accountants, we have traditionally excelled in these areas.

What do transparency, confidentiality, privacy, and ethics look like in this evolving AI/big data world? This important question is one that we're really starting to take seriously.

Irene Wiecek

Our speakers show us the interconnectedness of the three pillars through repeated themes. In the curated summaries that follow, repetition increases understanding by allowing us to see different views.

The objective of the conference is to open our eyes to the tremendous advancements and opportunities presented by AI, but we also want to encourage everyone to start thinking critically about the downsides of AI. How might AI be affecting society in a negative way? What are the steps we can take to steer it back on course?

Increasingly at the heart of big data discussions are issues such as transparency, confidentiality, privacy, and ethics. Ethics is a big piece of this. As accountants, we have traditionally excelled in these areas. What do transparency, confidentiality, privacy, and ethics look like in this evolving AI/big data world? This important question is one that we're really starting to take seriously.

As you read through the summary of the day, we encourage you to think about AI, not only in terms of the opportunities that it presents, but also some of the perils. Our speakers show us that many professions are starting to think about these sorts of things. How is AI changing everything that we do?

Are we on the sidelines looking in, or are we at the heart of things, helping to drive change and facilitate advancement? I think the latter is the more exciting place.

Irene Wiecek

The CPA profession is taking a fresh look at what competencies define professional accountants through a mandate called ***Foresight: Reimagining the Profession***. It is a good time for us to reflect on technological innovations and, specifically, AI. We would like readers to think about how AI is changing our lives and how AI is changing business — the way we do business, the different value propositions, and the different business models.

Then think about how we might redefine the role of the accounting profession in this AI environment. Ask yourselves where accountants fit in: what are the opportunities for us as a profession to contribute to business and, more broadly, to society? Watch for ***Links to Foresight***[3] boxes within the summary that link potential new directions for the profession to the topics discussed.

Are we on the sidelines looking in, or are we at the heart of things, helping to drive change and facilitate advancement? I think the latter is the more exciting place.

We are pleased and honoured to have leading academics, practitioners, and innovators share their thoughts with you and provoke discussion.

COVID-19 Update

Just six weeks after this conference, Wuhan, China reported its first clustered cases of pneumonia of unknown cause. On March 11, 2020, the World Health Organization declared the spread of the causal agent — novel coronavirus, COVID-19 — a global pandemic. By late-March 2020, the *Economist* opined in an editorial,

In just a few weeks a virus a ten-thousandth of a millimetre in diameter has transformed Western democracies. States have shut down businesses and sealed people indoors. They have promised trillions of dollars to keep the economy on life support. If South Korea and Singapore are a guide, medical and electronic privacy are about to be cast aside. It is the most dramatic extension of state power since the Second World War.[4]

Mobile technology was deployed quickly in China, Hong Kong, South Korea and Singapore. Mandatory apps or customer data from telcos were used to track people and track the disease.

[3] Excerpts within the ***Links to Foresight*** boxes come from CPA Canada Foresight, *The Way Forward: Transforming Insights into Action* and its two appendixes: *The Way Forward, Appendix 1: Reimagining the Accounting Profession* (April 2019), and *The Way Forward, Appendix 2: Reimagining the Accounting Profession, Digital Engagement Report* (April 2019).

[4] Economist."The State in the Time of Covid-19: Big Government is Needed to Fight the Pandemic. What Matters is How it Shrinks Back Again Afterwards." *The Economist* (https://www.economist.com/leaders/2020/03/26/the-state-in-the-time-of-covid-19, March 26, 2020).

Canadian AI played an early role in notification of the disease, and its contributions include:

- **Notification by Canadian AI company, BlueDot.** Announcements in January 2020 by the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) about a novel coronavirus followed independent notification on New Year's Eve by Canadian AI company, BlueDot. BlueDot, since the SARS outbreak in 2003, has used natural language processing and machine learning to survey the Internet to identify disease outbreaks. In 2020, the Canadian government partnered with BlueDot to use the health data startup's platform "…to support modelling and monitoring of COVID-19, [and] guide government decision-making."[5]
- **AI Against COVID-19 Canada.** Canada's AI hubs — Quebec Artificial Research Institute/Montreal Institute for Learning Algorithms (Mila), the Vector Institute, and Alberta Machine Intelligence Institute (Amii) — and the Canadian global charitable organization, Canadian Institute for Advanced Research (CIFAR), collaborated to "map and coordinate AI projects in Canada that can contribute to solve the COVID-19 outbreak."[6]
- **Contact Tracing.** The Vector Institute and the University of Toronto developed **MyTrace** for contact tracing to reduce spread of the disease by identifying people who have been infected and notifying others of contact with infected persons, privately.[7] Mila, under "the strictest ethical rules" is developing non-profit tracing software *COVI*, whose collected data will be destroyed when the pandemic is over.[8] Mila is also using deep learning in projects to help discover antiviral drugs and to better model immune response.[9]

Watch for ***COVID-19 Update*** boxes within the summary.

Read on!

---

[5] Geoffrey Vendeville, "U of T infectious disease expert's AI firm now part of Canada's COVID-19 arsenal," *U of T News* (https://www.utoronto.ca/news/u-t-infectious-disease-expert-s-ai-firm-now-part-canada-s-covid-19-arsenal, March 27, 2020).

[6] AI Against COVID-19 Canada [website] (https://ai-against-covid.ca/, 2020).

[7] MyTrace [website] (https://www.mytrace.ca/about, 2020).

[8] Valérie Pisano and Yoshua Bengio, "COVI: Protecting the health and privacy of Canadians," *Mila* (https://mila.quebec/en/covi-protecting-the-health-and-privacy-of-canadians/, 2020).

[9] "Mila COVID-19 Related Projects," *Mila* (https://mila.quebec/en/covid-19/, 2020).

TECHNOLOGY PILLAR

AI and Industrial Innovations

Fakhri Karray,[10] *University Research Chair; Director, Centre for Pattern Analysis and Machine Intelligence; Co-director, Waterloo Artificial Intelligence Institute; and Professor, Department of Electrical and Computer Engineering, University of Waterloo*

"The impacts of AI and machine learning on business, industry and society are predicted to surpass the impact of the various industrial revolutions, combined. Welcome to the fourth industrial revolution fuelled by big data, AI, and machine learning."

Fakhri Karray

Introduction

Global industry has undergone three revolutions. The fourth is here, thanks to a convergence of technological advances:

- Powerful computing processors and huge storage facilities (cloud computing, edge computing)
- Tremendous amounts of often free, widely available data (text, images, video), some of which are labelled
- Scalable algorithms able to make use of the data and powerful cloud computing facilities
- Open-source code and open-source AI platforms

Integrated over the past 10 to 12 years, this "perfect storm" of technologies has been able to tackle a wide range of problems that seemed unsolvable or intractable only few years ago.

This paper provides a broad overview of AI topics upon which other speakers will expand. It outlines global industry's revolutions, how and why AI is truly transforming the world, and the relationships between areas associated with AI, along with some general definitions. Based on the University of Waterloo's AI research, steps are given on how to implement AI programs at the micro- and macro-level, and the positive sides and risks of AI are summarized.

Global Industry Revolutions

The past 200 years have been characterized by major milestones and now four industrial revolutions:

[10] For more information about current research work, see the Waterloo AI Institute at https://uwaterloo.ca/artificial-intelligence-institute, email karray@uwaterloo.ca, or go to the Centre for Pattern Analysis & Machine Intelligence at the University of Waterloo: https://uwaterloo.ca/centre-pattern-analysis-machine-intelligence. All published articles can be accessed through Google Scholar, Fakhri Karray: karray@uwaterloo.ca.

1) **First Industrial Revolution.** Mechanized production with the tools of *water* and *steam* power

2) **Second Industrial Revolution**. Mass production with the tool of *electricity*

3) **Third Industrial Revolution**. Automated production with the tools of *electronics* and *information technology*

4) **Fourth Industrial Revolution.** Integration of breakthrough technologies (**Figure 1)** — robotics, 5G, Internet of things (IoT), computer vision, etc. — fuelled by the tools of big data, AI, and machine learning (ML)

Happening now, the impact of the fourth industrial revolution is predicted to surpass the impact of all three preceding industrial revolutions, *combined*.

The reason for the hugely growing role of AI is the associated tremendous opportunity for economic development. According to Deloitte and PwC, growth in global GDP by 2030 will be up by 14% ($15.7 trillion or higher, some say). [11] This is something almost unimaginable only two decades ago; unthinkable only five to six years ago.

Of this growth of $15.7 trillion, $7 trillion will be in China; $3.7 trillion in North America; $1.8 trillion in northern Europe; $1.2 trillion in Africa and Oceania; $0.9 trillion in the rest of Asia outside China.

Many technologies developed over the last 15 years use the latest AI and machine learning tools (see **Figure 1**). Almost every industry has applied AI to solve a wide range of problems that only a few years ago were unsolvable. AI is transforming all major sectors of industry and business that have used data — or not used data — to digitize information and gain knowledge and extract insights never seen before.

ARTIFICIAL INTELLIGENCE (AI)

Artificial intelligence is the effort to automate intellectual tasks normally performed by humans.

Narrow AI (also called **weak AI** or **shallow AI**), is made up of narrowly intelligent systems that can exceed humans in specific tasks, such as playing chess or making medical diagnoses. These narrow capabilities are not transferrable (i.e., an AI chess player cannot be used to perform another task such as a medical diagnosis).[12]

General AI or **strong AI** refers to human-level intelligence that can transfer knowledge between domains. While narrow AI is all around us in language and vision recognition systems and recommendation engines, general AI may be the stuff of science fiction and movies — for now.[13]

---

[11] See Anand S. Rao and Gerard Verweij, *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* (PwC: 2017, https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf).

[12] CPA Canada. *A CPA's Introduction to AI: From Algorithms to Deep Learning, What You Need to Know*. (https://www.cpacanada.ca/en/business-and-accounting-resources/other-general-business-topics/information-management-and-technology/publications/a-cpa-introduction-to-ai, 2019), p. 10.

[13] Ibid., p. 9.

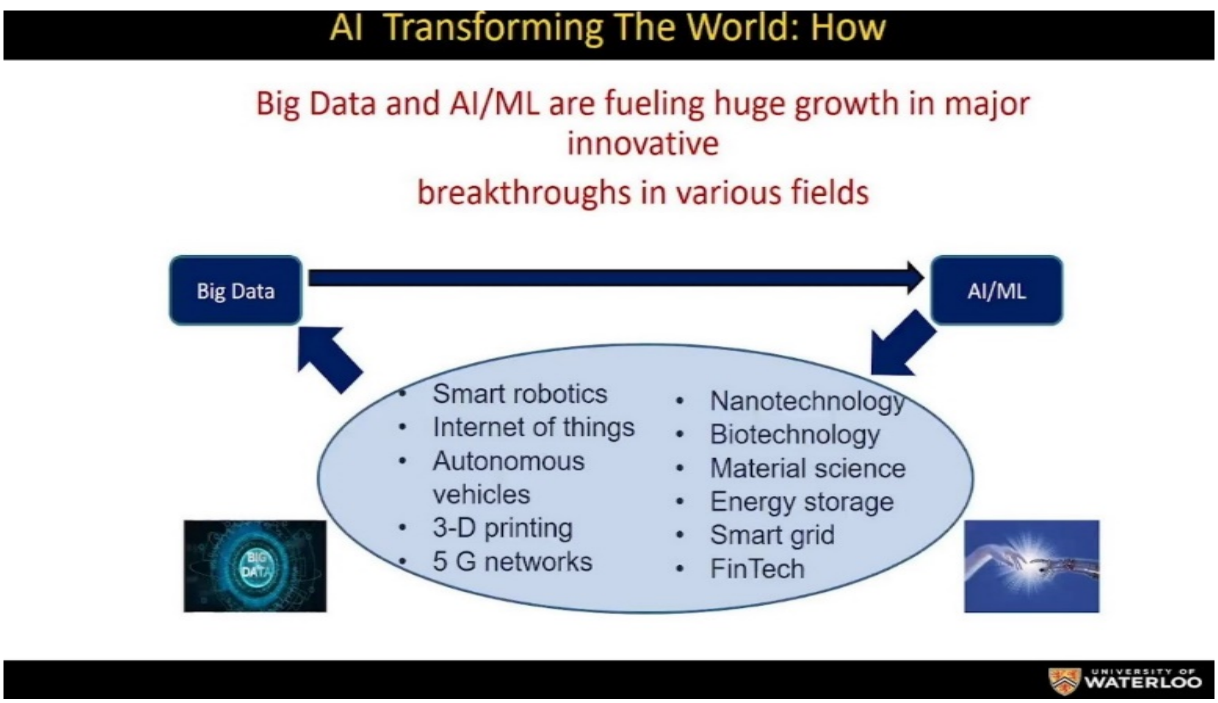
MACHINE LEARNING

Machine learning is used in email spam filters, voice recognition systems, spell checkers, and many more applications. The more experience — or "learning" — the application gains, the better it becomes at **predicting** spam or words from sounds or incorrect spelling, etc.

Machine learning applications are algorithms that learn patterns from the data they receive and then generate computational models. We call this their **training**. When presented with new (novel) data, the models guess correctly or incorrectly and learn from their mistakes. More data means more examples, so the models continuously learn and become better at their predictions over time.

Supervised learning is a method to train models by example. **Unsupervised learning** is the result of the system identifying relationships or insights through analysis of data that may not have trained datasets. As an example of the latter, Netflix may ask its system to find other customers whose viewing habits are similar to our own. It can then use those results to recommend movies that similar customers liked. Detecting **fraud** and **anomaly in transactions** often falls in this realm, because it involves looking for transactions that are "not typical."[14]

FIGURE 1: HOW AI IS TRANSFORMING THE WORLD

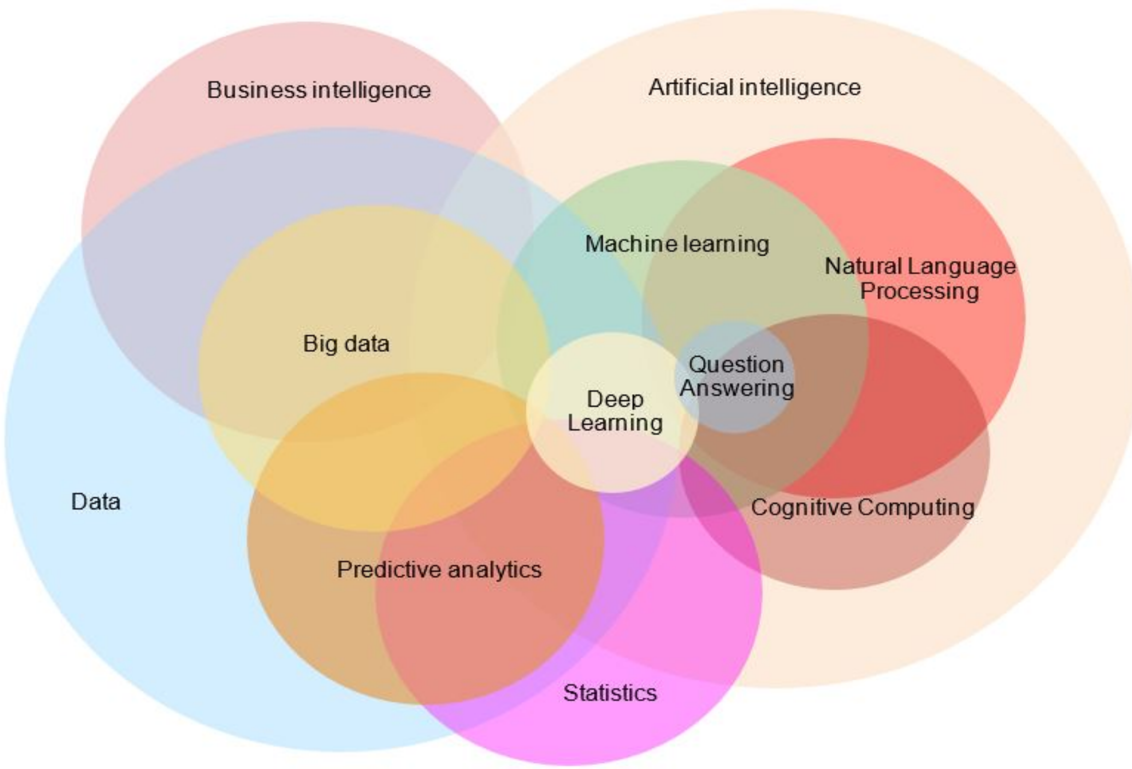


[14] Ibid., p. 11.

Relationships among AI, Machine Learning, and Big Data

Figure 2 shows the relationships between AI-associated areas. AI, the biggest area, encompasses a number of topics at this conference, including machine learning, natural language processing, cognitive computing, deep learning, and question-answering technologies like Siri and Alexa that allow the querying of knowledge bases in natural language. Data, the second largest area, comprises big data and predictive analytics, the major fuel for making these algorithms work.

FIGURE 2: APPROXIMATE RELATIONSHIP OF POPULAR TERMS[15]



[15] Source: Combrinck, Rikus. "Big Data Dictionary Data Word Soup — What Does It All Mean?" *South African Statistical Association* (https://sastat.org.za/sasa2017/big-data-dictionary, 2017).

DEEP LEARNING AND NEURAL NETWORKS[16]

Deep learning is a type of machine learning. It uses stacked algorithms (both supervised and unsupervised learning models) in multiple layers. Because this network mimics the brain, the structure is also called a **neural network**.

Neural networks have interconnected layers of nodes or neurons (depicted as circles) each doing computations using weights (depicted as lines) between each node to accomplish a specific type of task.

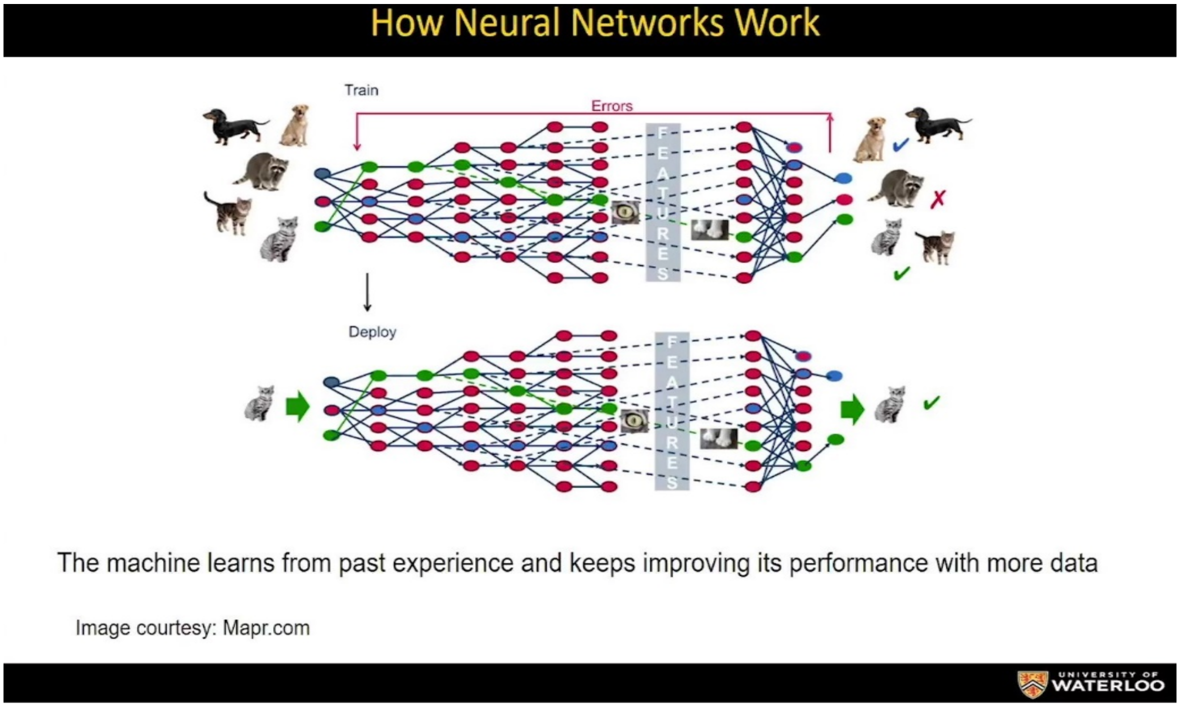
In **Figure 3**, the neural network is trained to recognize input images of animals. In training or **supervised learning**, the model is provided with tens or sometimes millions of **labelled** images. The model deconstructs the images to pixels (the **nodes**) and learns, in successive layers, the relationship (**weights**) between the pixels until it is able to identify **features** (nose, paw, tail, etc.). The output is the network's guess or **prediction** or **classification** of the input it was given. The **input** and **output** are visible layers; the other layers are hidden.

The **weights** are numbers that reflect the importance of the input data (e.g., the pixel) to carrying out the desired task. A weight close to zero will not significantly affect the output, for example.

When the model in Figure 3 is **deployed**, it is given a never-before-seen image of one of the types of animals in the training set (e.g., in a different posture), and the system should be able to recognize the animal and **label** or **classify** it.

[16] Based on Fakhri Karray's presentation and slides.

FIGURE 3: HOW NEURAL NETWORKS WORK[17]



Most of the discussions we on the news are actually breakthroughs in **deep learning**. The algorithms were discovered in the mid-1990s. They were improved over the years, but only now are finding application, because huge computational systems are available and able to store and quickly process huge loads of text, image, and speech data.

Tied to particular industries, the growth and competitiveness in AI technologies, in turn, generates growth and competitiveness in those industries. Examples include, but are not limited to:

- **Advanced traffic management**
    - Licence plate recognition
    - Traffic assessment
- **Security**
    - Fingerprint recognition
    - Image, face, object recognition
- **Virtual Assistants** (e.g., Siri, Alexa)
    - Speech recognition
    - Machine translation
- **Medicine**
    - Cancer detection

[17] Image courtesy of MapR Data Technologies, 2009-2019. The company was sold to Hewlett Packard Enterprise in August 2019. (Source: "MapR," *Wikipedia* (https://en.wikipedia.org/wiki/MapR, June 19, 2020).

Deep learning is only at the start of its growth and has not yet matured. Many deep learning applications are yet to be discovered.

Waterloo AI Institute

Waterloo Artificial Intelligence Institute (**Waterloo.ai**) aims at fostering cooperation between academic research and industry with support from government. It is one of Canada's four major research hubs in AI/ML, along with the Vector Institute in Toronto, Quebec Institute for Learning Algorithms (Mila) in Montréal, and the Alberta Machine Intelligence Institute (Amii) in Edmonton.

Large corporations now have the ability to lead as consumers *and* providers of AI technologies. Through their cooperation with academic institutions, like Waterloo.ai, research has accelerated to develop the latest AI technologies for several types of industries. Prototyping AI products no longer requires large equipment and expensive computational infrastructure. Prototype AI developed in a university research setting for software development, autonomous vehicles, or the service, insurance, health care or banking industries, can be taken to a higher level of **productization** when an industry partner is involved.

Waterloo.ai industry partners include start-ups, small to medium companies, and large corporations. Given the free availability of massive amounts of data, small companies can now push the envelope of AI technologies, something that was not possible even six years ago. Through existing open-data repositories and open-source code and software libraries, start-up companies develop software to solve particular problems and sell it to bigger companies. Corporations are purchasing start-ups in various fields to accelerate AI development and to increase their competitiveness.

***COVID-19 Update:* Open-source software will improve COVID-19 screening.**

Founding member of Waterloo.ai, Professor Alexander Wong, through the Vision and Image Processing (VIP) Lab and AI start-up Darwin AI, publicly released open-source AI software to detect COVID-19 infections from chest-x-rays on March 24, 2020. The "AI-assisted, x-ray screening method is meant to augment the…swab tests now in short supply in many areas of the world" by using a modality that is relatively low-cost and widely available. Using an open-source approach to the software and dataset of 6,000 images from 3,000 patients, the team's goals are to elicit input from "clinicians, researchers, and citizen data scientists" and to "build trust in the system" by providing an "explainability tool."[18]

[18] "Researchers use open-source software to improve COVID-19 screening with AI," *Waterloo News* (https://uwaterloo.ca/news/news/researchers-use-open-source-software-improve-covid-19, March 24, 2020).

Foundational AI versus Operational AI: Research at Waterloo.ai

Operational AI pertains to the *application* of AI for business and government. Research at Waterloo.ai focuses on partnering with industry on deep learning using big data in areas such as medical diagnostics, self-driving vehicles, manufacturing, consumer finance, agriculture, speech recognition and living architecture. Projects aim to be responsible and accessible by relying on minimal computing power and energy requirements.

At the same time, these applied projects contribute to AI's theoretical underpinnings or **foundational AI** by identifying and researching shortcomings in some AI approaches. Foundational research areas include machine learning, data mining, optimization and decision making, data science and analytics, cooperative agents, and game theory.[19]

Today, the two areas are converging, because companies requiring operational AI are investing in developing solutions to particular problems. This funding helps to support research in foundational AI.[20]

How to Build Capacity in AI

At the Micro (Company) Level

Businesses come to Waterloo.ai asking how they can utilize their huge volume of data to build capacity and to benefit from AI. At the company level, we advise them to:

- Anticipate trends and strategize for the future
- Increase business leaders' knowledge of advanced analytics to create demand
- Build a high-performance team in the field of AI and machine learning
- Acquire human resources early, because talent is becoming scarce[21]
- Build AI/machine language clusters with expertise in core areas of the business
- Build data storage capabilities; ingest analytically relevant data
- Embed, early on, data and analytics in transformational and strategic initiatives

At the Macro (Country) Level

- Government, public, and private institutions must invest *now* and *big* in educating the masses and corporations on the benefits of AI/machine learning.
    - In North America, this is happening, but not at the speed of other countries, like China, which are spending more money in this area.
- Massively train students and professionals in state-of-the-art tools of AI and machine learning.
    - At the University of Waterloo, we have almost 25 courses related to AI and machine learning, 10 more than two years ago. 10 courses are core courses.

---

[19] For more information on foundational and operational AI, see Anand Rao's paper in this volume.
[20] For more information on foundational and operational AI research at the Waterloo Artificial Intelligence Institute (Waterloo.ai), see https://uwaterloo.ca/artificial-intelligence-institute/research.
[21] For more information on the competition for AI talent, see Graham Taylor's paper in this volume.

- Promote tight cooperation between industry and research institutions to transfer knowledge on the most recent findings in AI and machine learning. This is of paramount importance.
    - From the University of Waterloo's interconnection with industry, we know businesses are eager to cooperate and to provide funding.
- Initiate workshops and seminars in various industries and businesses to educate employees on the benefits of AI and to form next-generation leaders.
    - Through the University of Waterloo, the AI Institute provided nearly 25 workshops worldwide in 2019, from basic to advanced courses in machine learning and AI.

AI: The Positive Side

"The AI revolution has the potential of positively affecting almost every sector of industry and a major portion of humanity."

Fakhri Karray

The tools of AI and machine learning have major potential benefits for business and society:

- Innovative technologies have the potential of improving the quality of life of individuals with access to the digital world. AI in medicine is growing exponentially, for example, in spin-off companies, diagnosis and labelling of diseases, and interaction with patients. Not without controversy, Apple and Google are competing to gain access to the huge repository of health unit data in the United States.[22]
- Spin-off technologies increase the efficiency of our personal lives. Many tedious tasks of the past (e.g., product comparison, shopping, library research, film selection) are simply done remotely now.
- Long-term gains in efficiency and productivity for small, medium, and large businesses are becoming more and more apparent.
- Cost of trade will diminish, thus opening new markets and driving economic growth globally.
- Transportation will become safer and friendlier for the environment; communication costs will drop; logistics and global supply chains will become more efficient. Examples include government investment to design smart cities, and industry development of *smart mobility*: connected vehicles, driverless/autonomous cars, smart highways, and advanced driver-assistance systems (ADAS), such as blind-spot monitors, lane minders, and proximity warnings.

SMART MOBILITY AND SMART CITIES

Smart mobility and its enabling technologies have the potential to manage the billion-plus vehicles on world roads today. "They allow for the design of more intelligent vehicles, permit safer journeys, and enable the design of more effective and smarter transportation networks, while significantly reducing

[22] For more information on Google and Apple's competition and different uses of health data, see Charlotte Cowles, "The Week in Business: Google Is Coming for Your Health Records," *New York Times* (https://www.nytimes.com/2019/11/17/business/this-week-in-business-google-health-records.html, November 17, 2019).

traffic congestion, road fatalities and injuries, fuel consumption and pollution."[23] Smart mobility "represents a cornerstone and an integral part of the **smart city** concept."

Smart cities, "the cities of the future…are built on data…Street sensors that reduce collisions and congestion, heated sidewalks that melt snow, buildings that predict when to heat, cool, and illuminate rooms, emergency and social services that know when and where people need help — These are the kind of urban improvements made possible when data about the behavior of people and objects is collected."[24]

Designers of AI systems in manufacturing and auto manufacturing say that without the current AI revolution, they would never have dreamt of being able to implement these systems in the real world. They used to talk about 10, 15, or 20 years for accomplishing some of the current or prospective capabilities. Now they are talking in terms of just a few years.

Similarly, smart cities are burgeoning around the world and are taking the technology to the next level. Much of the work around smart cities would not be possible without the revolutions in communication and AI systems.

This is the bright side of the AI and machine learning, and I hope we keep talking about the bright side. Of course, AI comes with challenges and issues, like any other type of technology.

AI Challenges and Risks

AI presents many opportunities, but also many challenges:

- Larger businesses have generally benefitted more from AI than smaller ones.
    - This discrepancy is changing with the availability of large amounts of free data, open-source codes, and cloud computing that allow smaller companies to participate.
- All that we have developed in AI and machine learning is still in its infancy; many problems remain to be solved. We have not yet reached the level of human intelligence. Artificial general intelligence (the AI of science fiction) is the ultimate goal, but is it a good thing?
- Trust in AI, and fairness and bias in machine learning, are major issues. Is human-AI interaction safe?
    - Technologies to be developed in the future will need humans at the centre to avoid the perception of robots or AI taking over the world. The latter will not happen, at least in the next few years, by good users of AI, but progress in that direction is being made. Currently, we are seeing what is known as **augmented AI**, not general AI.

[23] Fakhreddine Karray, *Advances in Smart Mobility: State of the Art and Challenges* [lecture], Electrical and Computer Engineering Program, Texas A&M, University of Qatar (https://www.qatar.tamu.edu/assets/img/images/programs/electrical%20and%20computer%20engineering/seminarpdfs/Karray%20-%20Nov%202016.pdf, November 21, 2016).

[24] Alex Ryan, "Can smart cities help their residents without hurting their privacy?" *World Economic Forum* (https://www.weforum.org/agenda/2019/12/smart-cities-help-residents-privacy-concerns/, December 9, 2019).

SMART CITY PRIVACY CONCERNS — UPDATE, TORONTO

"The debate around **ethical use of personal data** in civic development is affecting local communities and global markets, and [was] at the heart of [Google's Alphabet Inc.'s] **Sidewalk Labs**' smart city experiment in **Toronto**…"[25]

In May 2020, Sidewalk Labs pulled out of the project, citing economic uncertainty. However, Research in Motion's co-founder, Jim Balsillie, declared victory over the controversial project, because Alphabet Inc. would manage the data and be the only beneficiary. Involving local companies that have developed smart-city technologies would be better for Canada's innovation economy, he asserted.[26]

Balsillie said, "Sidewalk Toronto will go down in history as one of the more disturbing planned experiments in **surveillance capitalism**,[27] and I hope Canadian policy-makers will reconsider how we build Canada in the 21st Century's knowledge-based and data-driven economy."[28]

- Deep learning algorithms are not transparent and lack explainability.
    - Despite major advances in image processing and speech recognition, the black-box issue is still a major problem. Researchers are tackling this issue with some success. In the next few years, more transparent deep learning algorithms that are more explainable to users, when integrated with other tools of AI, could get us closer to achieving more powerful AI.
- More funding is required for AI for social good (long-term benefit for humanity). Waterloo.ai has partnered with for-profit, non-profit, and public sector companies to tackle problems of social benefit, including specific problems relating to the environment and sustainability, aging and health, support of citizen science, fairness and bias in machine learning, trust in AI systems, and human-AI collaboration. Waterloo.ai is also interested in creating an endowment fund which will support partnerships and research on problems of social benefit.[29] Two major global software companies have provided Waterloo.ai with some funding for current projects for research in the area of AI for social good.

[25] Alex Ryan, "Can smart cities help their residents without hurting their privacy?" *World Economic Forum* (https://www.weforum.org/agenda/2019/12/smart-cities-help-residents-privacy-concerns/, December 9, 2019).

[26] Megan Devlin, "Time to 'Start Over' on Sidewalk Labs Development, Balsillie says," *Globe and Mail* (https://www.theglobeandmail.com/business/article-time-to-start-over-on-sidewalk-labs-development-balsillie-says/, June 26, 2019).

[27] For more information on surveillance capitalism, See Miklos Vasarhelyi's paper in this volume.

[28] News staff, "Sidewalk Labs Pulling Out of Quayside Waterfront Project," *CityNews* (https://toronto.citynews.ca/2020/05/07/sidewalk-labs-pulling-out-of-quayside-waterfront-project/, May 7, 2020).

[29] *Waterloo Artificial Intelligence Institute — Waterloo.ai* (https://uwaterloo.ca/artificial-intelligence-institute, 2020)

Conclusion

"AI systems must integrate seamlessly with other tools and existing systems…like a harmonious symphony to provide the outcome envisaged for this technology: to avoid harm to humanity and provide the best possible outcome."

Fakhri Karray

AI presents many opportunities, many challenges, and many ethical, security, and privacy issues. Tools could come into the hands of people who do not utilize them for social good.

Deep learning by itself may solve many problems, but to move to general AI or to gain benefits from AI, systems must integrate seamlessly with other tools and existing systems. All of these have to come together like a harmonious symphony to provide the outcome envisaged for this type of technology: to avoid harm to humanity and provide the best possible outcome.

Keynote Speech: Controlling the Black Box: Learning Manipulable and Fair Representations

***Richard Zemel**, Research Director, Vector Institute; Professor, University of Toronto, Department of Computer Science, University of Toronto; Senior Fellow, Canadian Institute for Advanced Research*

*"The most noteworthy advances in deep learning are not the things that make it into the popular press, but instead, what is happening one level below the output; that is, not the performance of individual tasks, but the development of **intermediate representations**."*

Richard Zemel

Introduction

Supervised learning has produced some of deep learning's biggest success stories:

- Object recognition in images
- Speech recognition
- Machine translation (e.g., English sentences to French sentences)
- Sensor input mapping to driving controls for autonomous vehicles. [Mapping is the process of going from input *x* to output *y* using an algorithm(s). Examples of the sensor inputs are images, LIDAR; how to brake; how to steer)]

The most noteworthy advances are not the things that make it into the popular press, but instead, what is happening one level below the output; that is, not the performance of individual tasks, but the development of **intermediate representations**.

To get to that topic, we will pull up the lid of the black box to allow everyone to peer inside, get an idea of what is going on, and talk about the Vector Institute's current research that makes the black box a little more controllable and understandable. Peering inside means seeing the technical workings of the black box (simplified here).

This paper walks through progress in machine learning, from supervised to semi-supervised to unsupervised deep learning. Starting with a one-slide introduction to deep learning, depicting a neural network, it will show how the system, trained on an input of animal image, could provide as output the classification of a new image if it had been part of its training. From there, it shows how "one level before the output" is critical in *separating* characteristics of the input. In turn, this information separation is key to incorporating fairness in automated decision making. Further, the development of

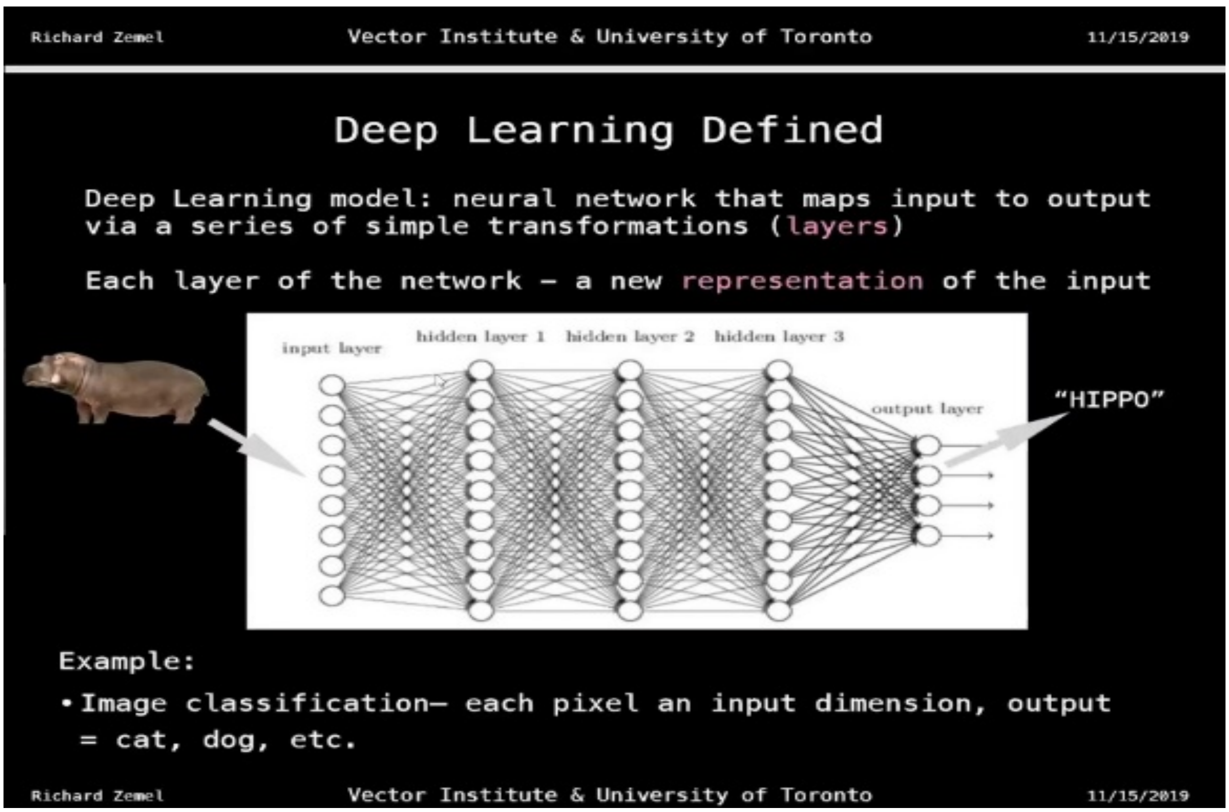
invertible networks both proves the integrity of the separated information and sheds light on the workings of the "box." Invertible networks have allowed progress in correcting two ways in which systems go badly wrong and in making results more predictable: a "white box" from a "black box."

Progress in Machine Learning

Deep Learning Defined: Supervised Deep Learning

Figure 4 depicts a deep learning classification model, where the model is trained on an image (a hippo) that it will learn to deconstruct and put back together—and classify—as a hippo. Because the inputs and outputs (the images and labels) are controlled or corrected by the operator/programmer/data scientist, this is an example of **supervised** deep learning.

FIGURE 4: DEEP LEARNING DEFINED



A deep learning model or system is also called a neural network because of its superficial similarity to signal transmission of neurons in the brain. An image is introduced to the model; in this case, a hippo. The hippo is the **input** image, comprised of pixels. Think of each pixel as an input *dimension*, depicted as a circle or *unit* (sometimes also called a neuron). Dimensionality comes from the position of each pixel relative to every other.

In Figure 4, four different outputs — or **classification decisions** (labels) — that can be made for a given input are shown. These might represent four animals the system has been trained to classify (e.g., cat, dog, turtle, hippo). What happens in deep learning is that a set of units goes between the input and the output. Each is a very simple computation that adds up the inputs from the unit in the layer below and produces some output value that is passed to the next unit. Each layer of the network is a new **representation** of the input. Key to the network are the values along the edges (the connecting lines) that go from one unit to the other, called the **weights** of the system. **Learning** inside the network means that we are going to adapt or change the values of the weights in the system.

WEIGHTS

The connections between these virtual neurons are represented by a number or **weight**. These weights reflect the importance attributed to the input data; the greater the weight the more important the input is to carrying out the desired task. [30]

Analogy: Think of weights as relationships, e.g., the pixels that together make the curve of a nose; the series of pixels that reconstruct the position of the nose relative to the eye. In very simple terms, the model will learn that the eye is close to the nose and further away from the foot. It will also learn that those spatial relationships differ between animals. The algorithms in the layers are learning these relationships.

The ultimate output of this network is to say what kind of animal is in the image: in this case, the network should say it is a hippo (see again Figure 4). The network achieves this by constructing a series of **new representations** of the input with each layer until, ultimately, the network can classify the image as a hippo or a cat or a dog. The neural network is often called a **classifier**.

Its learning is a **prediction** based on statistical probabilities. If an input image includes a long, hairy tail, the model will predict that it is not likely a hippo because the distribution of variability in hippos' tails does not include long, hairy ones. The model may be able to classify/label the image as a cat and distinguish it from a dog, based on the distribution of variability in dogs' and cats' tails that it has "seen," as well as the distributions of the relationships between tails and other features of those animals that it has learned. So, that is what image classification is all about.

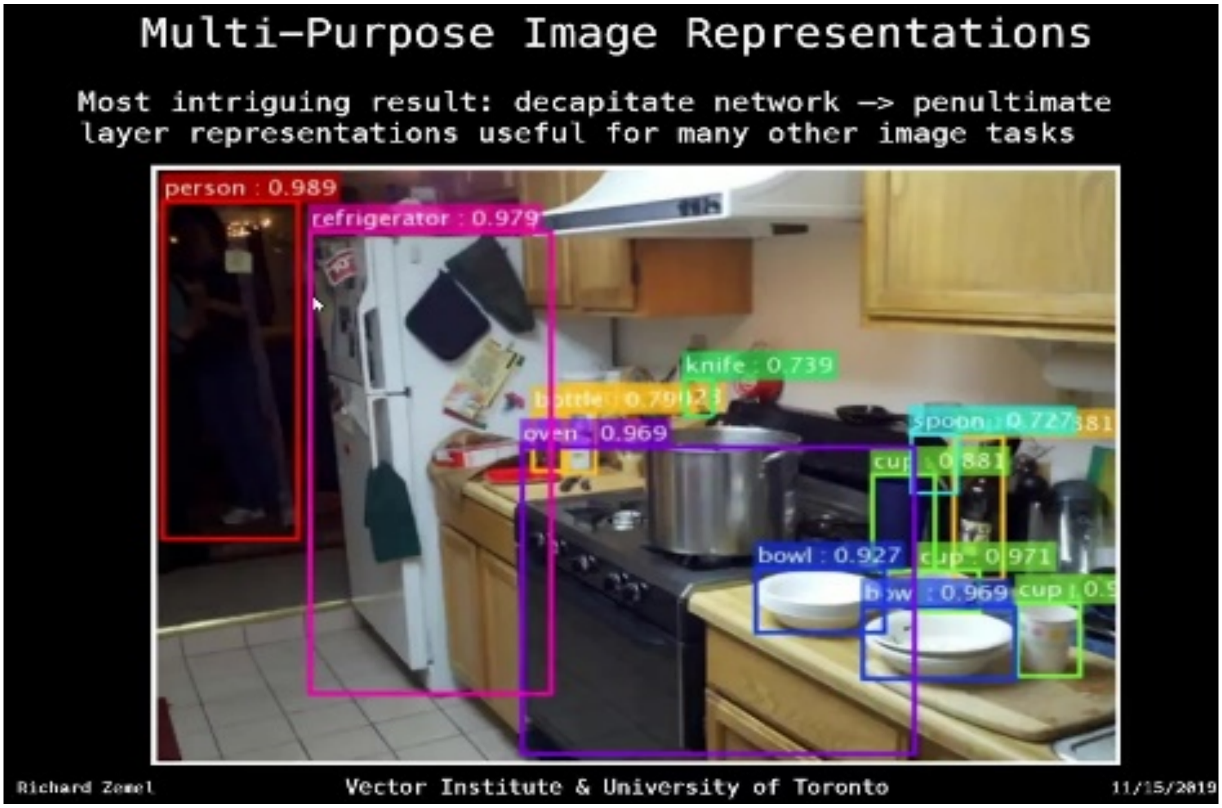
Developing Strong Representations: The Layer Below the Output

Imagine that we take the inputs that are mapping to the output, but we block out the last layer: we "decapitate" the network. The remarkable result is that the representations that are formed along the way from the input to the output are actually useful for a number of other tasks, not just for telling hippos from dogs and cats. For example, in Figure 5, a complex image of a typical North American kitchen, the system recognizes more than the dominant object (the person standing in the background). It also provides information about exactly what and where all the different objects are (fridge, stove,

[30] CPA Canada. *A CPA's Introduction to AI: From Algorithms to Deep Learning, What You Need to Know*. (https://www.cpacanada.ca/en/business-and-accounting-resources/other-general-business-topics/information-management-and-technology/publications/a-cpa-introduction-to-ai, 2019), p. 13.

pots, bowls, etc.), along with its estimate (confidence level) of its own accuracy in making those decisions.

FIGURE 5: MULTI-PURPOSE IMAGE REPRESENTATIONS



Generative Models: Systems That Can Produce Inputs (Semi-Supervised Learning)

To develop good representations that can be useful in other tasks, the key is to build a **generative model**: a network system that can generate **inputs**. A generative model turns the whole network around. Rather than thinking about mapping inputs (images) to output decisions (is it a hippo or a dog or a cat?), we want to turn it around and come up with something in the network that is actually going to *generate* inputs. A generative model is counterintuitive, but does several things:

- allows us to **test** that information about the input is not lost
- generates related but **novel** inputs: i.e., inputs similar to but different than the inputs on which the system was trained
- allows for **anomaly** detection. If the system is good at generating inputs, then it should be able to tell us whether an input comes from the same distribution or if it is very different (e.g., an input with a long, hairy tail does not fall within the distribution of hippos. The system could say if a temperature spike at a nuclear plant is anomalous or not when compared to a distribution of temperature-sensor readings).

The original formulation of a generative model was called an **autoencoder***.* Generating an input required a two-part solution:

- an **encoder** to map an input like an image to a **vector**, via a multi-layer network. The result of taking an input and putting it through a series of representations is called a vector.[31]
- a **decoder** that goes in the other direction. It takes a vector and maps it through a series of representations back to an image, the input.

So, we have an encoder, followed by a decoder.

Training the network

The main way of training this network is by defining an objective; that is, what is the system trying to achieve? What criteria determines if is it doing a good job? In this case, a good job is defined as the system *recovering the original input*. So, we start with the original input, we map it to a vector, and then that vector comes back as an **estimate** of the original input.

When that estimate matches the original input, we say that the network has encoded all of the information about that input. That is why it is called an **auto**encoder: it has encoded information on its own in a form of **unsupervised learning**.

Current research at the Vector Institute allows us to make the old generative model become more manipulable, more understandable, and allows us to address some issues about fairness and ethics. Here are three examples of how encoders have evolved so that we can address fairness and ethics in computations and, therefore, in computer models:

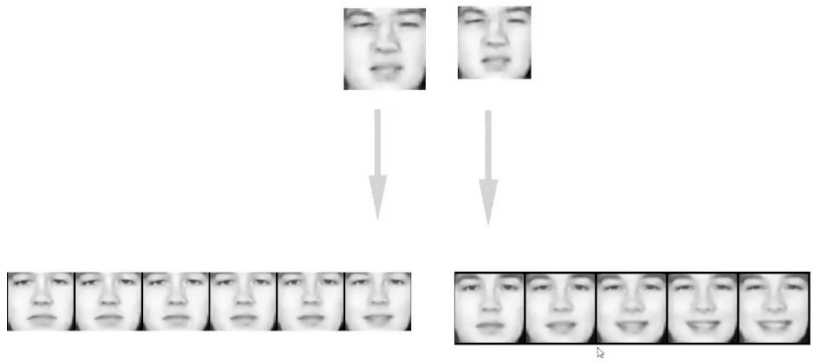
1. **Conditional Subspace Autoencoder**: a user-controlled fabrication or **generation** of novel items. Imagine a system trained on images of people's faces and trained to classify the emotion in the image. During training, we supply expression labels called **image tags** to classify expressions as happy, sad, disgusted, etc.

   We give the system an image of someone it has not seen before: a **novel** person with a particular emotional expression. The goal is not to answer, "What is the expression?" Instead, it is to *manipulate* the image so that the emotion of that person is *changed*, and a completely novel input is *generated*: the same person is given a different emotional expression.

   Figure 6 shows a novel person whose image is mapped to a neutral expression. Then, two series of images — generated by our system — show the person getting sadder and sadder or happier and happier. *The system was not trained to produce this completely novel result, yet it is able to do that.*

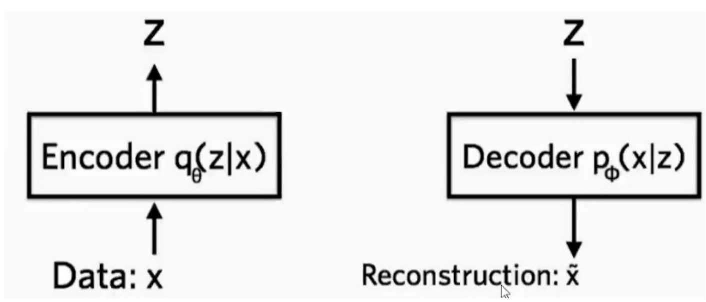
[31] The idea of having a vector to represent things was the genesis of the name for the Vector Institute. Vectors are really important in general and in neural networks, machine learning, and deep learning in particular.

FIGURE 6: SYSTEM-GENERATED SADDER (CENTRE TO LEFT) & HAPPIER (CENTRE TO RIGHT) IMAGES



2. **Variational Auto Encoder (VAE)**: a more modern instantiation of the autoencoder. Its key innovation is that both the encoder side and the decoder side are defined in terms of **probability distributions**. In **Figure 7**, the encoder takes the input data, *x*, and generates output, *Z*, which is a vector with a probability distribution. The decoder takes vector *Z* and generates a distribution over the reconstruction (or "estimate") of the input *x* called "*x-tilda*." We train the VAE so that these two systems (the encoder and the decoder) do a good job of reconstructing the data.

FIGURE 7: VARIATIONAL AUTOENCODER (VAE)



3. **Conditional Subspace Variational Auto Encoder (CSVAE):** This model *separates* the information associated with the input *x*: a key innovation.

FIGURE 8: CONDITIONAL SUBSPACE VARIATIONAL AUTO ENCODER (CSVAE) GRAPHICAL MODEL

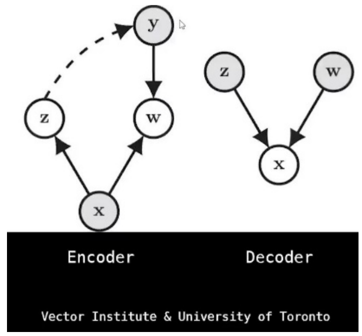


Figure 8 shows that from *z*, the encoder output for input *x*, we add a new variable, *y*, to represent, in this case, the facial expression. We then separate the information in *y*: we put all of the information about the expression into the new variable *w*, and let variable *z*, capture everything unrelated to the expression that is relevant to the input, e.g., the person's identity, the lighting, the angle of the image, etc. (This is called, "forming a rich **latent** or **hidden** representation of y, in z.") The system must learn to do this, so we have to train the system so that *z* contains no information about *y* — depicted as the dashed arrow — and all the information about *y* is put into *w*. This notion of coming up with a representation that *separates* information is a key point that will enable us to do other kinds of manipulations later.

Here are two examples of how the CSVAE works.

i) **Glasses or not? Discovering Latent Space.** A system is first trained on images and asked, "Is this person wearing glasses, or not?" Essentially, the answer is a binary variable, *w*: yes/no. Then, the system is given a new image of a person, and is asked, "What would this this person look like if she were wearing glasses?" The result is shown in Figure 9 (left side).

FIGURE 9: SYSTEM LEARNING

The result of the system's **unsupervised learning** is that it discovered two underlying dimensions for glasses. As a result, glasses frames get progressively thicker from left to right; they get darker from bottom to top. The system is creating variety all on its own. Without us telling it, it has discovered that underlying dimensions or **latent** (hidden) **subspaces** exist for the idea of what glasses look like.

> **Analogy:** In other words, the system is more observant than we could ever be!

So, now, we can control what kind of glasses this person is wearing by *exposing* this subspace and using this two-dimensional variable that was called *w* earlier.

Similarly, on the right-hand side of **Figure 9,** a system trained on a binary variable: yes/no for facial hair, generated its own variety in facial hair by discovering two underlying dimensions relevant for facial hair (i.e., thicker; more). The system learned that *on its own*.

ii) **Joint Attribute Transfer.** In another experiment, where we wanted the system to learn about both glasses and facial hair, we found an interesting result. When the system adds facial hair to the original image, the images become increasingly masculine. Conversely, images become more feminine as the system removes facial hair. This is not something that we wanted the system to learn; we just wanted to control facial hair and glasses. In this case, statistically, bearded women were not common in the training data. So, to add facial hair in a way that is consistent with the training data, the system is going to make the person look more masculine.

Developing Fairness in Automated Decision Making

"Algorithms make important, 'high stakes' decisions in all kinds of areas like healthcare, financial aid, university acceptance decisions, insurance rates, etc. We do not want a system that just tries to be as accurate as possible, because accuracy is based on historical data. If a system is accurate and is able to mimic decisions made in the past, it may perpetuate biases of the past."

Richard Zemel

Algorithms make important, "high stakes" decisions in all kinds of areas like healthcare, financial aid, university acceptance decisions, insurance rates, etc. We do not want a system that just tries to be as *accurate* as possible, because accuracy is based on historical data. If a system is accurate and is able to mimic decisions made in the past, it may perpetuate biases of the past. In other words, we now may not trust the answers or labels the system was given in the past. We can see this in the previous example, where people who identify as women and have facial hair would be considered anomalous because they were not part of what may have been a narrow training set, unrepresentative of the human population.

Therefore, we need a classifier that is accurate, but fair. How is "fair" defined? "Fairness" has exploded as a research area over the last five years. The most common setup of the learning algorithm is "fair classification," involving:

- some data, X (a data vector X), e.g., all the attributes of a given individual who might have applied for a loan.
- a label to predict, Y, which might be historical data, e.g., Did that person default on the loan or not?
- the model's prediction, Y^ (*Y-hat*); e.g., Is that person likely to default on the loan or not?
- a sensitive attribute, A, the new variable that we are adding, which is typically some sort of demographic variable, e.g., race, gender, age, socio-economic status.

The goal is for the algorithm to learn a classifier that is both accurate— meaning that Y^ matches Y, so it does good job of predicting whether the person is going to default on the loan or not—and also somehow fair with respect to the variable associated with A; e.g., fair with respect to gender.

The concept of fairness is complicated because different communities have their own definition. What we have been trying to ask is, "Can we build a kind of mathematical instantiation for each definition and use it to train a system?" Then we can look at the consequences and see to what extent people would think the result is fair, or perhaps, fairer than a system that does not take into account A at all.

The approach we have taken at the Vector Institute is something we call **fair representations**.

Fair Representations

The idea is that we do not necessarily just take an input X and come up with classification Y. Instead, we have an **intermediate representation**, much as I have described: we map X to a vector, then give that vector to a classifier. As an example, this construct is very relevant in the advertising world with targeted ads.

The way targeted ads work is:

1) A data owner categorizes (makes representations of) individuals. The data owner tracks individuals by looking at spending behaviour, browsing behaviour, etc., to come up with a group of similar individuals identified by a simple description, e.g., "White picket-fence females in their 30s." Those descriptions or **representations** are sold to vendors.

2) A vendor takes those representations and decides which ones are the right ones to target for directed ads.

Flexibly Fair VAE (FFVAE)

Typically, we are interested in being fair with respect to a **set** of attributes (e.g., race *and* gender *and* age, etc.), not just one attribute. We want to build a system that is very *flexible* that can be *fair* to any combination of the sensitive attributes. The goal in flexibly fair representation learning is to ensure that the data owner **removes** information about sensitive attributes, A (e.g., gender, race, etc.) in the representations of individuals, such that Z contains no information about A.

In other words, we want to preserve the information for classification so the vendor can still make good decisions, but we want to remove information about sensitive attributes.

Again, by using an autoencoder, we separate information in order to come up with a representation of an input of an individual, *Z*, that does not contain information about the sensitive attributes like race or age or gender. Another variable, *b*, will contain the sensitive information. Furthermore, we can then selectively **gate out** (mask) bits of *b* but retain all the other bits of *b* and *Z* about that individual by using a modified variational autoencoder.

We can demonstrate that we have achieved this separation in the representation, because when we just take *Z* into account, we do a very poor job of reconstructing the sensitive information in an input. Whereas, if we take the appropriate dimensions of *b*, we can actually predict the sensitive attributes — e.g., race and age — from *b*.

In addition, even if we have a malicious vendor with a complicated system, the sensitive information will remain hidden, because the system will be trained with a so-called **adversary** in the loop that must be overcome to ensure that representations are fair.

Information separation was useful when we wanted to manipulate facial expressions in images. Now, by separating pieces of information, we can make sure that advertisers are not taking into account sensitive attributes when they deliver ads.

Building more robust classifiers: Invertible Deep Networks

One of the more exciting developments in deep learning over the last few years is the idea that a deep network is actually **invertible**; i.e., all the information in the input is retained at every level of the network, and provably so. Comparing every representation to the input shows that all of the information is present. Recent work shows that this can be done with standard methods used in deep learning and, with just small changes, make them provably invertible. This means, now, that we can open up the lid a little more and find out how systems make **mistakes**.

Mistake 1: Discarded part of the signal dominates image content. The implication is that if we go back to the image of a hippo, we can separate all of the information about the hippo, *Zs*, from everything else in the image that is not relevant to the animal (the class), *Zn*; i.e., the background, the lighting, the angle, the day of the week, etc.

If we take *Zs* from the hippo image and combine it with the *Zn* of another image, then run the network backwards to see what kind of input results, we would expect a hippo with some elements of the background from the other image.

Surprisingly, the system does not do that at all. It does not retain any information from one image; instead, it gets almost all the information from the other image. So, the system does not do a very good job of separating information, even though there is enough information here to still be called a hippo, because we know that *Zs* is being used to classify it. To be fair, the system was not trained to **compose** (combine) the images, but the result shows a **failure** of the system. The failure is quite common, but not as popular to study as the next two.

Mistake 2: Perturbation-based adversarial images is a common type of system failure that has raised alarm about these systems. A **perturbed image** is not understood by the system to be essentially the *same* image, even though it has only been modified in a very *minor* way, by changing the background colour (adversarial noise), rotating it (adversarial rotation), or otherwise making minor manipulations to the image.[32] Outputs can be unpredictable and exhibit unintuitive failures. An example (not shown) is an image of a vulture which, when rotated, is classified by the system as an orangutan, even though to our eyes, it still looks like a vulture. This problem is called a **distribution shift**, meaning that the distribution of the introduced perturbed images differs from the distribution of the images used in training.

Mistake 3: Invariance-based adversarial images. This is a problem complementary to Mistake 2. Here, large changes are applied to images. They look very different to our eyes, but the system classifies them as the same thing. This problem is called **excessive invariance**, meaning that the classifier is — but should not be — invariant to all kinds of things about the input.

Because of the complementary nature of Mistakes 2 and 3, improving one may worsen the other. A somewhat successful improvement of both problems comes, again, from methods of separating information. For example, by separating part of the signal from an image — the class information — any perturbations, for example, the style of another image, may result in an image that retains the class identity, but gains a new style. An example (not shown) is an image of a numeral five of one script style, and numeral zero in another script style. We end up with something that still looks like a five, but it has a bit of the style from the zero.

Conclusion

We can open up the black box and make it a little bit more manipulable. We can create a representation in a deep network algorithm where parts of that representation can be manipulated in an interesting way. Instead of the system telling us if an image has eyeglasses or not, the system can manipulate the style of the glasses. In short, we expose the system to information and enable the system to manipulate it.

Our control over the information contained in the learned representations can be very important for things like fairness, where we want to be able to *take out* information about things that we do not want the system to pay attention to. In short, we are exposing the system to information and getting rid of it so that the system *cannot* make use of it.

By using these state-of-the-art classification models in a controllable way, we can generate new instances of that class. For example, our generative model takes a face it has never seen and changes its facial expression.

The underlying theme to advances in deep learning algorithms is the **separation** of information in representations.

[32] See the paper by Anand Rao in this volume. Illustrated is an error made by a deep learning algorithm in classifying a traffic stop sign that has small black and white stickers on its face. The deep learning algorithm erroneously — and potentially dangerously — classifies the stop sign is as a speed limit sign.

Current Research Directions

We are aiming to answer the following research questions:

- **Semi-supervised learning.** Can we extend these techniques to achieve this pinnacle of achievement in machine learning? In semi-supervised learning, we try to train a system with just a few labelled examples and many unlabelled examples.
- **Transfer learning.** If we control these representations in the way I have described, will that enable us to transfer learning to new tasks? That is, could we take the same representations used to train up a classifier and use them in other tasks? (Recall the example of being able to localize *all* the things inside the kitchen, not just the dominant object.) This is a step closer to general AI from narrow AI.
- **White-box models.** Incorporate known variables and relationships to try to achieve transparent models whose behaviour/predictions can be explained.
- **Uncertainty representations.** How can a model know what it knows and what it *does not* know? This is, again, a long-standing question in machine learning. At the Vector Institute, I think we are making some interesting progress along those lines.

Machine + MD: The Role of AI in Computer-Integrated Interventional Medicine

***Parvin Mousavi**, Professor and Director, Medical Informatics Lab, School of Computing, Queen's University*

"We have gone from hypothesis-driven to data-driven science."

Parvin Mousavi

Introduction: AI in Medicine

The use of AI in healthcare has seen tremendous growth, particularly in the imaging and diagnostics in the field of radiology. Radiologists[33] interpret and diagnose disease and injury through medical imaging techniques such as ultrasound and magnetic resonance imaging (MRI), the techniques discussed in this paper.

This talk summary explains a revolution in medical diagnosis, from failure of some early expert systems for cancer diagnosis to current physician-accepted **augmented intelligence** systems. Using examples of AI-improved diagnostic imaging for prostate cancer and heart disease, this paper outlines the benefits of the new technologies and underscore how they can socially and geographically democratize healthcare.

For students, recommendations from 2019 reports presented to the Summit of the G7 Academy of Sciences, and those commissioned by the Canadian Chamber of Commerce on the impacts of AI on business and the workforce, show three things: that only an interdisciplinary approach will maximize the societal benefits of AI; that upskilling is essential; and that **interfacers** — intermediaries between developers and businesses — will be highly necessary.

Hypothesis-Driven to Data-Driven Science

We have come from science that was heavily *hypothesis-driven* to science that is *data-driven*. In the past, we imagined reasons for things that we observed. We designed experiments; we collected data to test hypotheses. Now, we have many sources of data. Through machine learning, we look for features and associations in our data. We actually come up with hypotheses *from* our data. So, we have gone from hypothesis-driven to data-driven science.

[33] "Radiologists are medical doctors (MDs) who specialize in diagnosing and treating disease and injury by using medical imaging equipment such as x-rays, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), fusion imaging, and ultrasound." From: "What is a Radiologist?" *Ontario Association of Radiologists* [website] (https://oarinfo.ca/patient-info/what-radiologist, 2020).

Both in the popular press and scientific papers, we see headlines about algorithms (AI) that can make better diagnoses than humans. Despite these advances, the past has had some "busts," like rule-based expert systems called Computer Aided Diagnosis (CAD) developed in the late 1990s. One, presented to oncologists and surgeons to supposedly automate breast cancer diagnosis, had poor adoption, and false-negative results were too common.[34]

Another bust was the MD Anderson Watson Project.[35] The University of Texas MD Anderson Cancer Center partnered with IBM's AI division to use IBM Watson to create an Oncology Expert Advisor in 2013.

In both cases, physicians (MDs) — the ultimate users — were not integrated into the project and "did not demand better."[36] They were not involved in using the algorithms nor in providing input. IBM Watson understood statistics associated with the medical information it was fed, but it could not be trained to think like a physician.[37]

By contrast, the American College of Radiology (ACR) is currently developing a roadmap to move forward with AI. Open platform software will allow radiologists to develop AI algorithms and to be involved in the development of AI for radiology.[38]

"Artificial intelligence will not replace radiologists, but radiologists who embrace AI will replace those who don't." [39]

William T. Thorwarth Jr.

"So, AI will not necessarily take over their jobs; it will change *their jobs."*

Parvin Mousavi

Some of the ACR's recommendations for moving forward with AI include:

- Resisting the urge to over-hype AI capabilities

---

[34] For more information, see, for example, Manisha Bahl, "Detecting Breast Cancers with Mammography: Will AI Succeed Where Traditional CAD Failed?" *Radiology*, v. 290: 315–316 (https://pubs.rsna.org/doi/pdf/10.1148/radiol.2018182404, 2019).

[35] Eliza Strickland, "How IBM Watson Overpromised and Underdelivered on AI Health Care: After Its Triumph on Jeopardy!, IBM's AI Seemed Poised to Revolutionize Medicine. Doctors Are Still Waiting." *IEEE Spectrum* (https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care, April 2, 2019).

[36] Dr. Geraldine McGinty, Chair, Board of Chancellors, American College of Radiology.

[37] Ibid.

[38] Hagland, Mark. 2019. "The American College of Radiology's Transformational Work in AI in Radiology: Moving Forward: Mike Tilkin, CIO and EVP at the American College of Radiology, discusses recent advances being made in the collaborative development of AI-based algorithms for the practice of radiology." *Healthcare Innovation* [website] (https://www.hcinnovationgroup.com/imaging/radiology/article/21119057/the-american-college-of-radiologys-transformational-work-in-ai-in-radiology-moving-forward, December 21, 2019).

[39] Thorwarth Jr., William T. 2018. "Healthcare 2025: To Lead or Not to Lead, That Is the Question [Robert D. Moreton Lecture, Texas Radiological Society, March 24, 2018]." *American College of Radiology* (https://cdn.ymaws.com/www.txrad.org/resource/resmgr/docs/2018_Annual_Meeting_/DX_Post_Conf_Syllabus/7.3_Thorwarth_Moreton_Lectur.pdf, 2018), p. 95.

- Ensuring explainability
- Improving productivity and decision making
- Aiming for comprehensive integration with workflow across diagnostic processes, e.g., interpretation, measurement, and workflow standardization

In other words, make AI work as a **general-purpose technology**, not the AI of science fiction.

Essentially, in radiology over the short term, we are not looking at *artificial* intelligence, but **augmented intelligence**, in which the patient, patient-centric clinician, along with the technology, are *together* making the best decisions. With augmented intelligence, humans remain in the loop, and the system is learning continuously the background.[40]

In radiology and many other areas, we are not talking about MD *versus* machine, but MD *plus* machine.

Machine Learning for Computer-Aided Interventions

Heterogeneity: The Challenge for AI in Medicine

The uptake of technologies that actually change patient lives is not as fast as implied by the fantastic headlines in science and popular media. Uptake is slow because of three big challenges in developing AI-based algorithms:

1. **Noisy and imperfect data (technology/centre/operator).** The data, like many other areas of medicine, are imperfect and highly noisy. Why noisy? Because human bodies are heterogeneous, diseases are heterogeneous, tissues are heterogeneous, and data collection is heterogeneous. For example, compared to remote, small-community clinics, large urban hospitals have access to many technologies, more expensive equipment producing less noise, more patient flow resulting in more data and more experienced operators. The difference is huge, for instance, between the data collected by expensive, cart-model ultrasound equipment and inexpensive hand-held ultrasound devices.
2. **Noisy labels (due to heterogeneous label assignment).** In supervised machine learning, one of the ways that machines learn from data is through **labels**. [41] In medicine, those labels usually come from pathologists. When cancer is suspected, a clinician might conduct a biopsy. The biopsied tissue is studied by a pathologist who assigns a label related to the cancer being benign or malignant. Labels themselves are very heterogeneous because benign cells and malignant cells can each take many forms.
3. **Noisy labels (due to heterogeneity of disease).** The way the label is recorded is even more challenging. A biopsy sample from prostate cancer might be just two centimeters long, labelled by pathology as aggressive cancer; however, only three millimeters of the sample are cancerous. That is it. When I am imaging this, I have tens of thousands of data points, yet the whole sample is given one label. So, what do I do to train my model? Using this label for all those tens of thousands of

[40] For more information on the four ways to apply AI, see Anand Rao's paper in this volume.
[41] For more information on labels, see papers by Fakhri Karray and Richard Zemel in this volume.

points would be incorrect, because not all of those points are cancer. So, labels — our gold standard in medicine — are extremely noisy.

Innovation, Democratization and Data-Driven, Theory-Guided AI

At Med-*i* Lab (Medical Informatics Lab) at Queen's University, **innovation** means developing the latest hardware as well as software.[42] **Democratization** means making it (as well as what currently exists), available in some form to others. By taking advantage of the data from the hundreds of thousands of patients that might go through a Toronto centre, we discover information and use it for a remote area in the north.

In addition, *what* we study is inspired by the changing landscape of disease over the last 100 years. One hundred years ago, communicable disease was the leading cause of death; now that is cancer and heart disease.

***COVID-19 Update:* Has the COVID-19 pandemic returned communicable disease to the leading cause of death?**

For 2020, deaths in the United States from COVID-19, reached approximately 54,800 on April 26, 2020, and 93,606 by May 21, 2020.[43] In 2019, cancer caused 252,500 deaths; in 2017, heart disease caused 269,900 deaths in the United States.[44]

On April 17, 2020, the Centers for Disease Control and Prevention reported that "Hospitalization rates [for those with COVID-19] increase with age and are highest among older adults; the majority of hospitalized patients have underlying conditions,"[45] including heart disease.

Professor of social epidemiology at the Harvard T. H. Chan School of Public Health, Nancy Krieger says that although the "… amount of missing [racial or ethnic] data is horrific," the pandemic is "…also revealing the very different conditions in which we live because of social structures that are inequitable, both within the United States and between countries…[I]t's revealing patterns that have been long known in public health."[46]

---

[42] For more information on Medical Informatics (Med-*i*) Laboratory at the School of Computing, Queen's University, see *Med-i-Lab* at https://medi.cs.queensu.ca/.

[43] Johns Hopkins. "COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins" (https://coronavirus.jhu.edu/map.html, April 26, 2020).

[44] Sharon Begley Hyacinth Empinado, "It's difficult to grasp the projected deaths from Covid-19. Here's how they compare to other causes of death," *STAT* (https://www.statnews.com/2020/04/09/its-difficult-to-grasp-the-projected-deaths-from-covid-19-heres-how-they-compare-to-other-causes-of-death/, April 9, 2020).

[45] Centers for Disease Control and Prevention. "Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020," *Morbidity and Mortality Weekly Report* (https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e3.htm, April 17, 2020).

[46] Isaac Chotiner, "The Interwoven Threads of Inequality and Health," *The New Yorker* (https://www.newyorker.com/news/q-and-a/the-coronavirus-and-the-interwoven-threads-of-inequality-and-health, April 14, 2020).

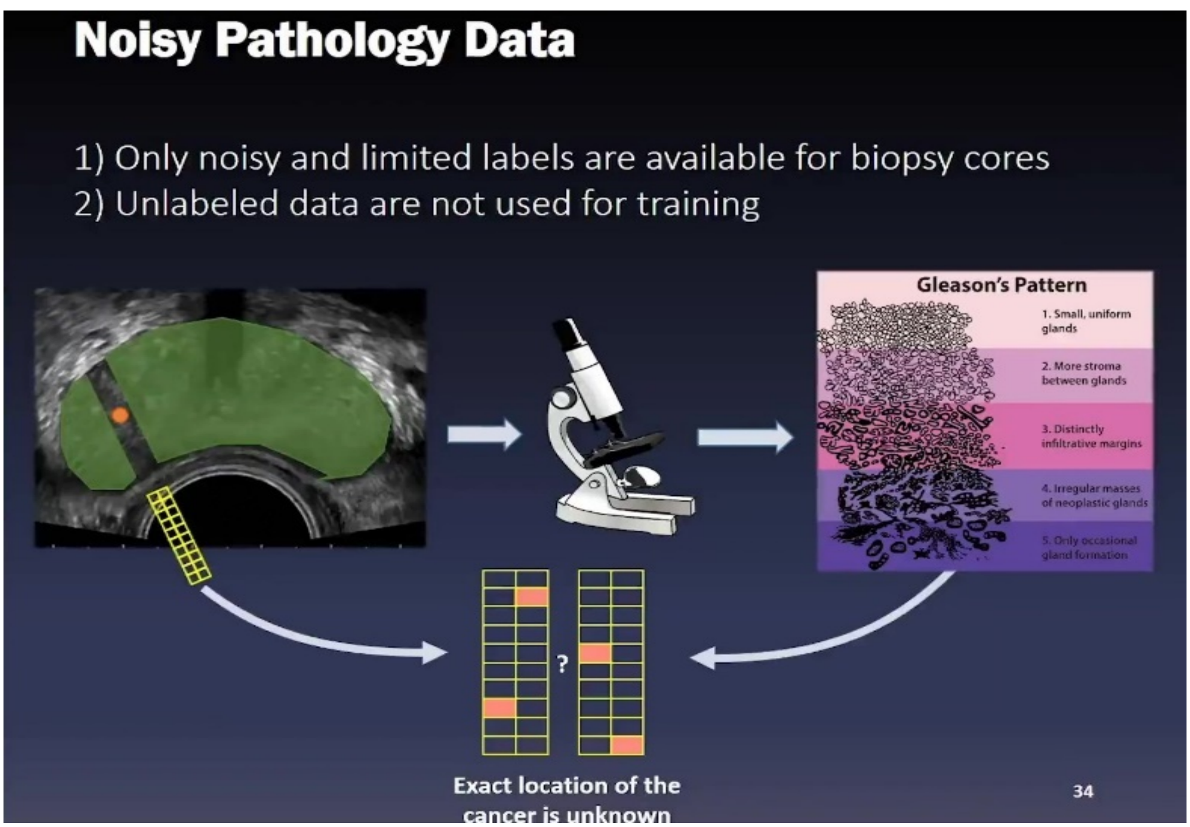
The two examples that follow show how we at Med-*i* Lab, and researchers at the University of British Columbia (UBC), are addressing the three challenges in developing AI-based algorithms for cancer and heart disease.

Example 1: Prostate Cancer Diagnosis and Intervention

The disease. One in seven men will be diagnosed with prostate cancer. If there is a positive side, this cancer is usually a disease of old age, and many types of it take a long time to develop to an aggressive level. So, if it is detected accurately and early, intervention is possible, and the disease can be managed.

Detection. Typically, diagnosis involves three things: a blood test to detect prostate-specific antigen; symptom assessment; and a physical examination of the prostate gland. When cancer is suspected, the clinician biopsies the gland. This involves the clinician using a needle to extract tissue sample cores while watching an ultrasound image. The image is delivered to a computer monitor through an ultrasound probe. The ultrasound shows and records the location of the samples. It acts as a map — like the GPS system of a car — but because cancer does not show itself in ultrasound images, a biopsy using ultrasound alone is still hit-and-miss when it comes to finding cancerous cells. Tissue sample cores, sent to pathologists, are sectioned, examined, and labelled under microscope. Digital images are generated. The biopsy cores are assigned a limited number of labels. The exact location of the cancer within the core is unknown, and the labels are noisy, because of the heterogeneity of cells under one label (Figure 10).

FIGURE 10: NOISY PATHOLOGY DATA



Pre-operatively — before a biopsy — the patient can be sent for MRI (magnetic resonance imaging). A different modality than ultrasound, MRI can detect areas of cancer, but an ultrasound image is also required to bring that information into the real-patient coordinate system. Recent software can fuse MRI and Trans-Rectal Ultrasound images (MR-TRUS) to generate a 3D view. The fusion of images locates cancer with greater certainty than with ultrasound guidance alone, but characterization of the severity of the cancer and its extent is still imprecise.

Through Med-*i* Lab, we use semi-supervised and unsupervised machine learning to create an **ultrasound probability map** in real time. While biopsying, the clinician refers to this *live* image to see *both* cancer location and severity in the prostate gland. We achieved this in stages:

1. First, we *took advantage of the noise* in ultrasound imaging. By combining MR-TRUS images with **Temporal Enhanced Ultrasound** images, which are ultrasound images of the same point many, many times over a very short period of time, we increased the signal-to-noise ratio.[47]
2. Next, by using machine learning, we were able to create, in real time, probability maps to tell the clinician the location of cancer in the prostate gland. The AI system learned to differentiate between noise and signal. It was trained to ignore noise. Clinicians loved these maps but said they needed more than the simple, binary yes/no label for the presence of cancer: they needed more definition. Clinicians needed to know the location of *aggressive* cancer.
3. The team went back to cancer pathology and asked, "Can we predict what is more aggressive than less aggressive cancer?" Working with the U.S. National Institutes of Health (NIH) for clinical trials, pathologists there compared *two* samples at each biopsy location on the prostate gland and labelled them using a cancer aggressiveness scale from 1 to 5 (least to most aggressive). Level 4 cancer requires immediate intervention. A patient with level 3 cancer would likely be sent home, to return for rechecking after six months. Missing a level 4 cancer with a normal biopsy (a false-negative result) is a risk for metastasis of the cancer within that six-month period.

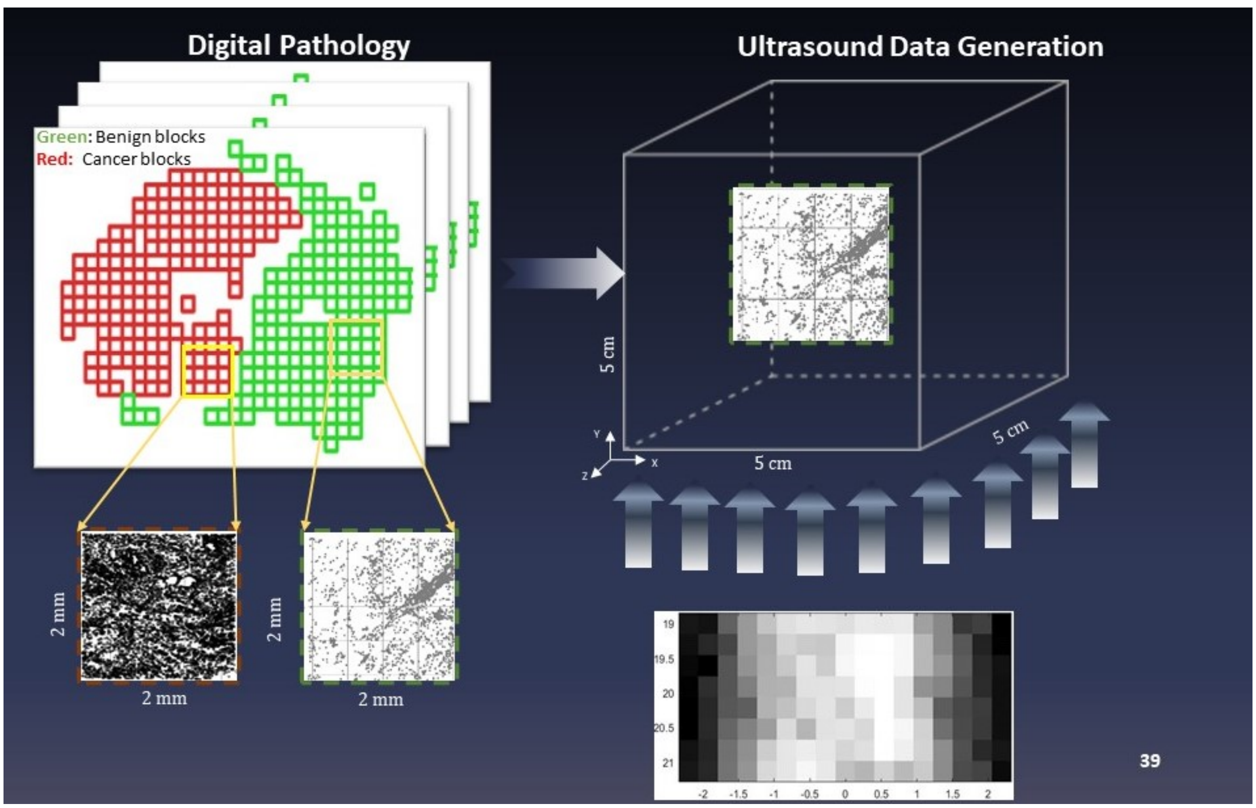
   The team combined pathology and technology to develop *real-time* colour ultrasound maps. Different colours told the clinician the likelihood of more aggressive versus less aggressive cancer. Still, the clinician could not be confident about all areas and saw heterogeneity in the live images. AI could not make the decision for the clinician on where or how many times to biopsy: it could only provide information to help the clinician make those decisions — MD *plus* machine.

   Recall, pathology provides only one label, whereas the reality is that in that little piece of tissue, there is a lot of information. The team took advantage of this.

[47] For more information on Temporal Enhanced Ultrasound (TeUS), see Shekoofeh Azizi, *et al.*, "Ultrasound-Based Detection of Prostate Cancer Using Automatic Feature Selection with Deep Belief Networks," in: N. Navab, J. Hornegger, W. Wells, A. Frangi (eds), *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015*), *Lecture Notes in Computer Science* [book series], v. 9350: 70-77 (https://doi.org/10.1007/978-3-319-24571-3_9, 2015).

4. Using some locations where we had confidence that we had just one type of cancer, we compared the rest of the tissue to those very high-probability cancer areas. Using machine learning — and the system's ability to use that training data and encode information *on its own* in a form of **unsupervised learning** —we predicted how confident we were that the information *outside the labelled area* is cancer, based on what we knew for sure.[48] This gave the clinician a probability map, in real time, of areas that are likely to be cancerous and of how severe the cancer is.
5. Finally, improving the model meant supplying it with more training data. We did not have hundreds of thousands of patients with ultrasound scans, but what we *did* have is hundreds of thousands of pathology slides of patients that had gone through Kingston General Hospital (KGH) or NIH over time. AI makes it possible for us to digitize these slides, come up with where the cells are, and then use the physics of ultrasound creation to *simulate* ultrasound images. So, from these pathology images, we get ultrasound images that we have *created*. By using these **simulated data**, we improve the real data for which we have fewer samples (Figure 11).

FIGURE 11: SIMULATED ULTRASOUND DATA



---

[48] For an explanation of unsupervised learning in this volume, see Richard Zemel's paper. For more information on unsupervised learning in detecting cancer, see Alireza Sedghi, *et al.*, "Deep Neural Maps for Unsupervised Visualization of High-grade Cancer in Prostate Biopsies," *International Journal of Computer Assisted Radiology and Surgery*, v. 14, 1009–1016 (https://doi.org/10.1007/s11548-019-01950-0, 2019).

The AI enables —

- **Enhanced biopsy and cancer detection.** Essentially, the clinician is given a **likelihood map**.
- **Virtual biopsy.** If our trials are successful, the best outcome would be that patients would undergo MRI and ultrasound but would have no need for biopsy. Recall Professor Zemel's eye-glasses example: by using a representation of someone with glasses, the system can then change the range of the glasses with machine learning, to get novel outputs ranging from reading glasses to sunglasses. [49] The same applies here. If we know the signatures of cancers, and we see this patient's particular signature today, machine leaning might enable us to predict what an aggressive cancer in that patient would look like in six months or a year. Ultimately, we could look for that altered signature through *virtual biopsy*, with no need for needles.
- **Potential cost savings.** Prostate biopsy is a US$4 billion business in North America. Of the biopsies aided by ultrasound alone, 70% are negative, but not because 70% of patients do not have prostate cancer. Patients still displaying symptoms of the disease are sent for MRI, which could cost $500 to $2,000. After MRI, 15% are considered negative for prostate cancer, and are sent home. Additional costs include a risk to patients of side effects (e.g., discomfort, infection, blood in urine, reaction to antibiotics) and associated costs for treatment.
    - If clinicians started with our **Temporal Enhanced Ultrasound (TeUS)** technology, only 40% of patients would be considered negative for prostate cancer, which means we would have 30% *fewer* false negatives. [50] That would come to about US$600 million in savings. This is value generation.
    - The 30% considered positive for cancer after ultrasound alone go for biopsies using MRI guidance. From that group, over half (62%) are considered negative for prostate cancer and go home.
    - If clinicians started with our ultrasound technology, instead of 62% of patients being sent home as negative for prostate cancer after MRI, we would see a 10% reduction in false negatives, for savings, ultimately, of around US$240 million. So, our technology has many cost benefits.

Example 2: Heart Disease

Heart disease is the leading cause of mortality in middle-aged adults.

Cardiac imaging modalities include MRI, CT (computed tomography), and fluoroscopy. Another is **echocardiography** ("**Echo**"), which is the most cost-effective of all. Most of us have heard of someone who has had an echocardiogram — an ultrasound of the heart.

---

[49] See the topic "Conditional Subspace Variational Auto Encoder (VAE)" in Richard Zemel's paper in this volume.

[50] Shekoofeh Azizi, *et al.*, "Toward a Real-time System for Temporal Enhanced Ultrasound-guided Prostate Biopsy," *International Journal of Computer Assisted Radiology and Surgery*, v. 13: 10071 (https://www.researchgate.net/publication/324042173_Toward_a_real-time_system_for_temporal_enhanced_ultrasound-guided_prostate_biopsy, March 2018).

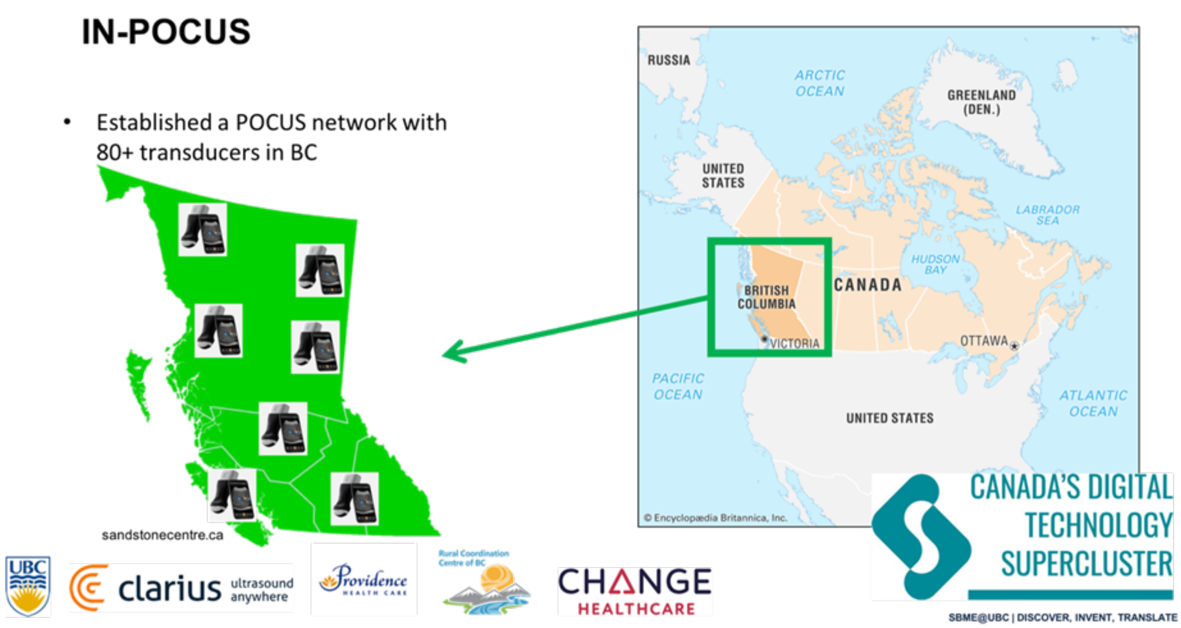
The advantages of echocardiography are that it is harmless, cost-effective, and generally accessible. It gives results in real time and is a standard of care.

Unfortunately, ultrasound delivers noisy, imperfect data. Less-complex imaging hardware and software produce lower quality images; less-experienced operators give rise to higher variability in measurements and outcomes; and patient age and body mass can affect images. Up until now, specialized facilities administered Echo, not small points of care (a general practitioner's (GP) office, or small centres).

UBC researchers, in partnership with Vancouver Coastal Health, Providence Health Care, and local industry, are trying to make Echo **ubiquitous**; that is, available at points-of-care, large or small, in cities or remote geographic places (Figure 12).[51]

Using AI, they raised the quality of point-of-care ultrasound (POCUS) using hand-held devices to that of expensive, cart-based equipment at large echo centres.[52]

FIGURE 12: IN-POCUS PROJECT, FUNDED BY CANADA'S DIGITAL TECHNOLOGY SUPERCLUSTER, AIMS TO MAKE ECHO UBIQUITOUS ACROSS B.C.



[51] Canada's Digital Technology Supercluster, "Intelligent Network for Point-of-Care Ultrasound: Using Technology to Deliver Equal Access to Life-saving Ultrasound Imaging," (https://www.digitalsupercluster.ca/programs/precision-health/intelligent-network-for-point-of-care-ultrasound/, 2018).

[52] Mohammad H. Jafari, *et al.*, "Cardiac Point-of-care to Cart-based Ultrasound Translation Using Constrained CycleGAN," *International Journal of Computer Assisted Radiology and Surgery*, v. 15: 877–886 (https://link.springer.com/article/10.1007%2Fs11548-020-02141-y, April 2020).

Fuelled with 10 years of Echo data from the entire city of Vancouver, they used a mix of algorithms and AI to develop a **cardiac diagnosis assistant** in the form of a mobile app. [53] [54]

While the operator scans the patient *in real time*, the app is telling them, based on AI, if the quality of their scan is good or not. How does it know quality?

For those 10 years of data, an expert cardiologist labelled the ultrasound image quality with a score between 0 (not acceptable) and 5 (excellent), and researchers trained the AI to predict image quality based on this diagnostic value.[55] Now, as a technician scans a patient with a hand-held transducer, the app displays the ultrasound image on a cell phone. A bar display shows the image quality in real time (green for high quality, red for low quality) so that the operator can change the position of the transducer to improve the scan quality. In the hands of a new or inexperienced technician at a smaller centre, this app improves patient outcomes by improving the consistency of data collection through improved image quality (Figure 13).

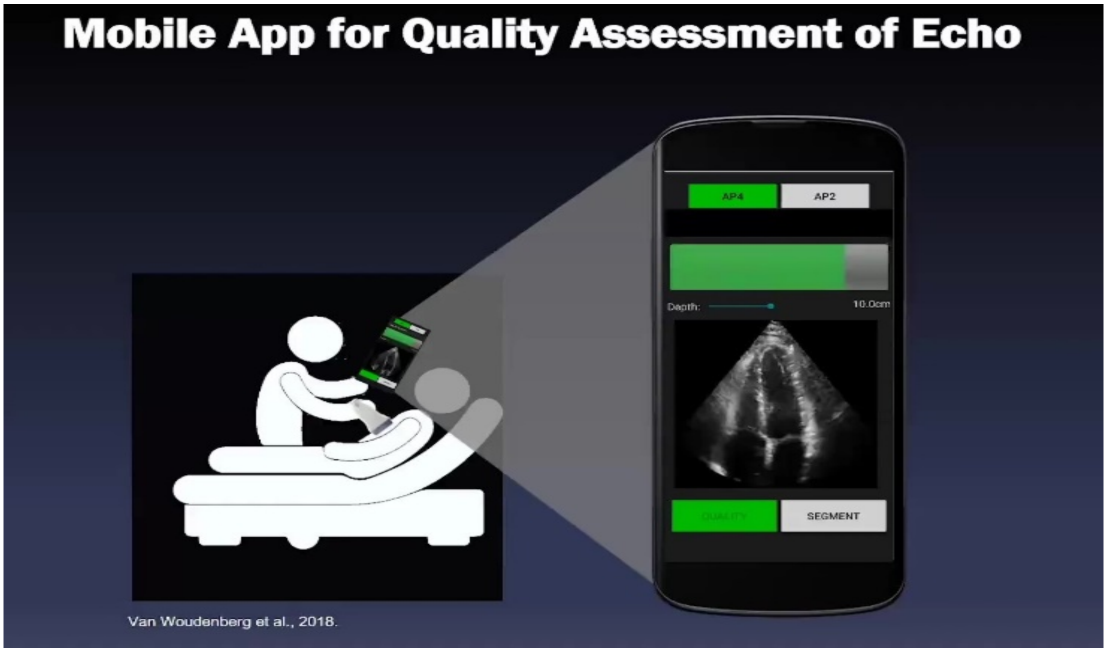
---

[53] See, for example, Andrea Fung, *et al*., "Artificial Intelligence Application for Assessing Point-of-Care Ultrasound Image Quality," *Journal of the American Society of Echocardiography*, v. 32 (6): 122 (https://www.researchgate.net/publication/334130584_Artificial_Intelligence_Application_for_Assessing_Point-of-Care_Ultrasound_Image_Quality , June 2019).

[54] Nathan Van Woudenberg, *et al*., "Quantitative Echocardiography: Real-Time Quality Estimation and View Classification Implemented on a Mobile Android Device," in: D. Stoyanov, *et al*. (eds), *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation. POCUS 2018, BIVPCS 2018, CuRIOUS 2018, CPM 2018, Lecture Notes in Computer Science* [book series], v. 11042: 74-81 (https://doi.org/10.1007/978-3-030-01045-4_9, 2018).

[55] Amir H. Abdi, *et al.*, "Automatic Quality Assessment of Echocardiograms Using Convolutional Neural Networks: Feasibility on the Apical Four-Chamber View," *IEEE Transactions on Medical Imaging*, v. 36 (6): 1221-1230 (https://doi.org/10.1109/TMI.2017.2690836, June 2017).

FIGURE 13: MOBILE APP FOR QUALITY ASSESSMENT OF ECHO



The G7: AI and Society

In 2019, the G7 Science Academies produced three public, joint statements with recommendations on priority areas for science, one of which was AI: [56]

- Science and Trust
- Citizen Science
- AI and Society[57]

As a committee member for *AI and Society*, I have reproduced the committee's recommendations here, which "policy makers should encourage, and scientists should commit to": [58]

- **Careful stewardship is necessary to help share the benefits of AI across society.** This will require close attention to the impact of AI on employment which will be in turn shaped by a range of factors including political, economic, and cultural elements, as well as progress in AI technologies.
- **AI systems and data should be trustworthy.** This should be facilitated through measures addressing the quality, lack of bias and traceability of data. While this can be further aided by making the data more accessible, personal data should not be made available to unauthorized parties.

---

[56] Institut de France, Académie des sciences, *G7 2019: Statements of the science academies* (https://www.academie-sciences.fr/en/Rapports-ouvrages-avis-et-recommandations-de-l-Academie/g7-2019-statement-academies-sciences.html, 2019).

[57] *Artificial Intelligence and Society* [Joint statement of G7 member countries produced at the Summit of the G7 science academies, March 25-26, 2019] (https://www.academie-sciences.fr/pdf/rapport/AI_G7_2019_EN.pdf, 2019)

[58] Ibid., p. 1.

- **AI systems and data should be safe and secure.** This is essential in the case of applications that involve human vulnerability and may require provably correct systems.
- **Further research is needed to help develop explainable AI systems.** When important decisions are suggested by AI impacting people, those concerned should be given sufficient information and be allowed to challenge the decisions (e.g., refuse a treatment or appeal a decision).
- **Insights from many fields are needed in order to maximize the societal benefits of AI.** Interdisciplinary research should involve diverse fields such as natural, life and medical sciences, engineering, robotics, humanities, economic and social sciences, ethics, computer science and AI itself.
- **Citizens need to be AI-ready.** A range of AI educational opportunities and information should be available and a well-founded dialogue with citizens is required to demystify this field.
- **Public policy debate on the destructive/military usage of AI should be promoted.** International undertakings limiting the risks of autonomous weapons should be considered by the relevant UN body.
- **Talent exchanges and cooperation between public research and the private sector should be encouraged.** This would facilitate safe and rapid deployment of applications in areas of great human benefit. Collaboration is important for large-scale collection of data that are crucial for developing AI systems.

The Canadian Perspective: Automation Not Domination

The Canadian Chamber of Commerce and McCarthy Tétrault LLP, with others, sponsored three AI roundtables in 2019 to bring together experts and business leaders across multiple industries and academia to discuss issues and opportunities in relation to AI in Canada.[59] The Chamber of Commerce published summaries of the discussions under a main title, *Automation Not Domination*, with three subtitles: *AI and the Workforce*;[60] *AI and Inclusion*;[61] and *Legal and Regulatory Frameworks for AI*.[62]

At a more specific level than the G7 report, these summaries provide statistics and recommended actions from participants for questions posed at the roundtables. For example, *AI and the Workforce*, says (p. 2):

- The OECD estimates that 14% of jobs in OECD countries are already highly automatable; 32% will be radically transformed.

---

[59] Including aws (Amazon Web Services), Google, Microsoft, IBM, University of Waterloo, and others.

[60] Goldenberg, Adam, and Michael Scherman, *Automation Not Domination: AI and the Workforce.* (Canadian Chamber of Commerce: 2019, http://www.chamber.ca/media/blog/190627-automation-not-domination-building-ai-policies-that-win/AI-Workforce.pdf).

[61] Goldenberg, Adam, and Michael Scherman, *Automation Not Domination: AI and Inclusion*. (Canadian Chamber of Commerce: 2019, http://www.chamber.ca/media/blog/190627-automation-not-domination-building-ai-policies-that-win/AI-Inclusion.pdf).

[62] Goldenberg, Adam, and Michael Scherman, *Automation Not Domination: Legal and Regulatory Frameworks for AI*. (Canadian Chamber of Commerce: 2019, http://www.chamber.ca/media/blog/190627-automation-not-domination-building-ai-policies-that-win/AI-Legal.pdf).

- By 2022, approximately 54% of employees will require significant re-skilling and upskilling.[63] In Canada, some 394,000 workers will move up the skills ladder by 2026.
- Disruption will not be easy. However, it is increasingly likely that the effect on Canada's labour market — which will evolve and adapt — will be net positive overall.
- Labour productivity improvements will drive 55% of global GDP gains from AI by 2030.[64]

Together, some of the recommendations from the summaries pertinent to students are these:

- **Cultivate Canada's talent pool.** Concentrating AI skills and knowledge will make the Canadian workforce more resilient. Canada lags behind other countries in deploying AI (*AI and Inclusion*, p. 4).
- **Build a resilient workforce.** Companies and governments should encourage career-long skills development (*AI and the Workforce*, p. 5).
- **Encourage AI "interfacers."** The go-betweens who see AI solutions, drive AI adoption, and act as the interface between developers and businesses — in the same way as the drivers of the projects I have described — should be encouraged (*AI and the Workforce*, p. 5).

LINK TO *FORESIGHT*: Defining opportunities for the profession

Part of the new direction for the profession is a focus on continuous learning, which will require CPAs to ensure their skills and competencies are current with changes in the business environment. To be seen as strategic leaders by users and society, CPAs must be technology-savvy, effective communicators, and agile learners who embrace and promote innovative thinking. (p. 37)

From *AI and Inclusion*:

- **Close the AI knowledge gap.** As AI tools have become more accessible, the knowledge gap about the technology has grown.
- **Remove barriers to procurement and implementation.** Canadians' cultural aversion to risk is currently compounded by a lack of data strategy and the slow pace at which AI adoption is occurring.
- **Address the needs and characteristics of Canadian businesses, society, and customers.**
- **Enable 5G access across Canada.**

From *Legal and Regulatory Frameworks for AI*:

[63] World Economic Forum, Centre for the New Economy and Society, "Future of Jobs Report 2018" (http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf, 2018).
[64] Anand S. Rao, and Gerard Verweij, *Global Artificial Intelligence Study: Exploiting the AI Revolution* (PwC: 2018, https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf).

- **Enable innovation and growth while protecting privacy.** The government should revise privacy laws in a way that enables innovation and the use of datasets while protecting privacy rights. A risk-based and balanced regime for the de-identification of personal information, and subsequent use thereof by technology companies, is one example.

MEDICAL RECORDS AND PRIVACY

DeepMind,[65] together with London, U.K.'s Moorfield Eye Hospital, reported *The Economist*,[66] are using machine learning to attempt "…to use the eye as a window through which to detect signals about the health of other organs" through machine learning.

The project uses a dataset of 300,000 eye scans, "whether [the patients] know it or not." Approvals to build the dataset, "while respecting privacy and confidentiality," took more than two years. "Names and other easily identifiable information are not preserved." Although controversy arose because patients were not informed directly, notices issued about the project and the use of patient data also explained the potential benefits. "Not one person complained."

"The project will…act as a model for linking disparate health data together in a useful way while respecting patients' rights."

Conclusion

We are moving to an era of AI as a general-purpose tool. It is not in my hands or other researchers' hands yet, but it is moving toward being the modality that everyone will take advantage of. Many books discuss this idea. One, written by cardiologist Eric Topol, explains simply how AI might make him a better physician.[67]

In my opinion, it is going to be *augmented intelligence* that will have tools that will pass the regulatory hurdles. Medicine is making progress. It has a bright future. It has a role in any area, and certainly for students there is a role.

---

[65] For more information about DeepMind, see the papers by Anand Rao and Graham Taylor in this volume.

[66] Anonymous, "A System Based on AI Will Scan the Retina for Signs of Alzheimer's: and After That, of Stroke Susceptibility and Heart Disease," *The Economist* (https://www.economist.com/science-and-technology/2019/12/18/a-system-based-on-ai-will-scan-the-retina-for-signs-of-alzheimers, December 18, 2019).

[67] Eric Topol, *Deep Medicine: How artificial intelligence can make health care human again.* (New York: Basic Books, 2019).

GOVERNANCE PILLAR

AI and the Future of the Accounting Profession: Déjà Vu All Over Again

***Miklos Vasarhelyi**, KPMG Distinguished Professor; Director, Rutgers Accounting Research Center and Continuous Auditing & Reporting Lab, Rutgers Business School, Rutgers University*

"People overestimate the impact of technology in the short term and underestimate the impact of technology in the long term."[68]

Roy Amara, "Amara's Law"

Introduction

With a 30-year perspective on researching and teaching AI, Miklos Vasarhelyi has seen an evolution in the still-maturing field of AI.

In this paper, particularly for accounting and auditing students, he outlines —

- Opportunities for continuous learning
- Auditing research at Rutgers University Continuous Audit and Reporting Laboratory (CarLab).
- How AI research on continuous auditing, intelligent process automation, and deep learning may change the auditing/accounting field
- Areas that will disappear and others that will emerge

New Skills and Competencies

"We are preparing students for jobs that do not exist, using technologies that have not been invented to solve problems that have not been identified."

Ellen Glazerman, Executive Director of the EY Foundation, and Director of University Relations, EY Americas

Students, this is your world, and I think Ellen Glazerman is reasonably correct. Professional accounting MBA degrees have not changed much in 30 years. We need agile processes, and we need agile learning strategies, and I do not know if traditional degrees make sense anymore.

[68] Paraphrasing of Stanford Research Institute computer scientist and futurist, Roy Amara (1925-2007), "Amara's Law," attributed in 2006. For more information, see "Roy Amara," *Wikipedia* (https://en.wikipedia.org/wiki/Roy_Amara, July 13, 2020) and "Roy Amara," *Oxford Essential Quotations* (https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00018679, 2016).

In Silicon Valley, students take a Python course and join Google at $80,000 a year. This will happen in accounting, too, but might take longer. We are much slower to change.

Until that time, the **Rutgers Accounting Digital Library** (RAW) has made public (and free) on YouTube every course we teach: over 200 course videos are available in accounting and audit analytics.[69]

AI Evolution

"Never mistake a clear view for a short distance."

Paul Saffo, Silicon Valley technology forecaster

Even after 30 years, AI can be thought of as having the ability to resolve only **narrow**-domain problems. The idea that robots can act like humans, that AI can interact and work on problems independently, is unrealistic for now.

In other words, I do not believe that **singularity** in AI is near, the time when machines will become self-sufficient and smarter than humans.[70] Tests for that are too simple. For example, the Turing Test says if a question is asked of a human and machine and the answers cannot be differentiated, then the machine is intelligent. That, in my opinion, is not a good test. Just think about people seeing today's smart phone 10 years ago. They would think it more intelligent than people.

In 1964, John McCarthy established the Stanford Artificial Intelligence Laboratory. He was one of the fathers of AI and its namer. He believed that construction of an artificially intelligent machine would only take about 10 years. Nearly 60 years later, this work is ongoing.

In 1965, Ed Feigenbaum joined Stanford and became one of the founders of its computer science department. Called the "father of expert systems," he interviewed human experts, extracted the knowledge from experts, and created systems which were called expert systems. This research led to Expert Systems (ES) becoming one of the five main areas of AI.

For 20 years AI did not progress too much, and by 1984 that original optimism had hit a rough patch known as **AI winter**. AI was taking longer than anyone expected. After the failure of a crop of start-up companies in Silicon Valley, veteran Silicon Valley technology forecaster Paul Saffo proclaimed, "Never mistake a clear view for a short distance."

By the 1990s the expert system field was the major area of AI research, with 60% to 70% of applications directed toward it. Expert Systems became the most popular area of AI and, eventually, the basis of many commercial, semi-commercial, and prototype systems.

By the 1990s, the five main areas of AI were:

1. Expert systems
2. Natural languages

[69] "Rutgers Accounting Digital Library (RAW)," *Rutgers Accounting Web* (http://raw.rutgers.edu/RADL.html).
[70] Singularity is a notion from the 1950s, advanced in 2005 by futurist Ray Kurzweil, author of *The Singularity is Near*. (New York: Viking, 2005).

3. Cognition and learning
4. Computer vision
5. Automatic deduction

By 2016, the list had not changed much, but we could add one more area: deep learning/cognitive computing.

In Accounting and Auditing

From 1985 to 1990, I was at Bell Laboratories, where we developed a continuous process auditing system (CPAS) and published a paper about it in 1991. This was the first **continuous auditing system** and the first paper on continuous auditing. It was a rule-based system for AT&T's billing system and was still in use when I left the labs five years ago. A number of us also edited six books on applications of AI in accounting and auditing (including expert systems) between 1989 to 2002.[71]

In 1999, the Canadian Institute of Chartered Accountants (CICA, now CPA Canada) and the American Institute of Public Accountants (AICPA) published a research report called *Continuous Audit*, also called *Red Book*. It was the basis of additional guidance published by the Institute of Internal Auditors (IIA) in 2005 and was published by the Information Systems Audit and Control Association (ISACA) in 2010.

In 2015, the newly updated *Red Book* was much more advanced than earlier versions, but in 2020, continuous auditing is still not ubiquitous.[72] This tells us how slowly AI or any other new technologies come into common use. I think companies really should be using continuous auditing. I think the Public Company Accounting Oversight Board (PCAOB), AICPA Auditing Standards Board (ASB), and International Auditing and Assurance Standards Board (IAASB) auditing standards are having trouble keeping up.

In 2013, Glen Gray at California State University and two of my Rutgers PhD students decided to study what happened with expert systems and accounting.[73] They found some rule systems are still in use today, built into mainstream systems. For example, American Express evaluates credit using Falcon, which started as an expert system. Expert systems that have become mainstream are not even called AI anymore. Can you imagine what an offense that would be to the original developers?

---

[71] M.A. Vasarhelyi, *et al.*, *Artificial Intelligence in Accounting and Auditing*, volumes 1 to 6 (New York: Markus Wiener Publishing Inc., 1989 to 2002). For more information, see http://www.markuswiener.com/table_of_contents/artificial-intelligence-in-accounting-and-auditing/.

[72] *Audit Analytics and Continuous Audit: Looking Toward the Future* (New York: American Institute of Certified Public Accountants, Inc. (AICPA), 2015, https://www.aicpa.org/InterestAreas/FRC/AssuranceAdvisoryServices/DownloadableDocuments/AuditAnalytics_LookingTowardFuture.pdf).

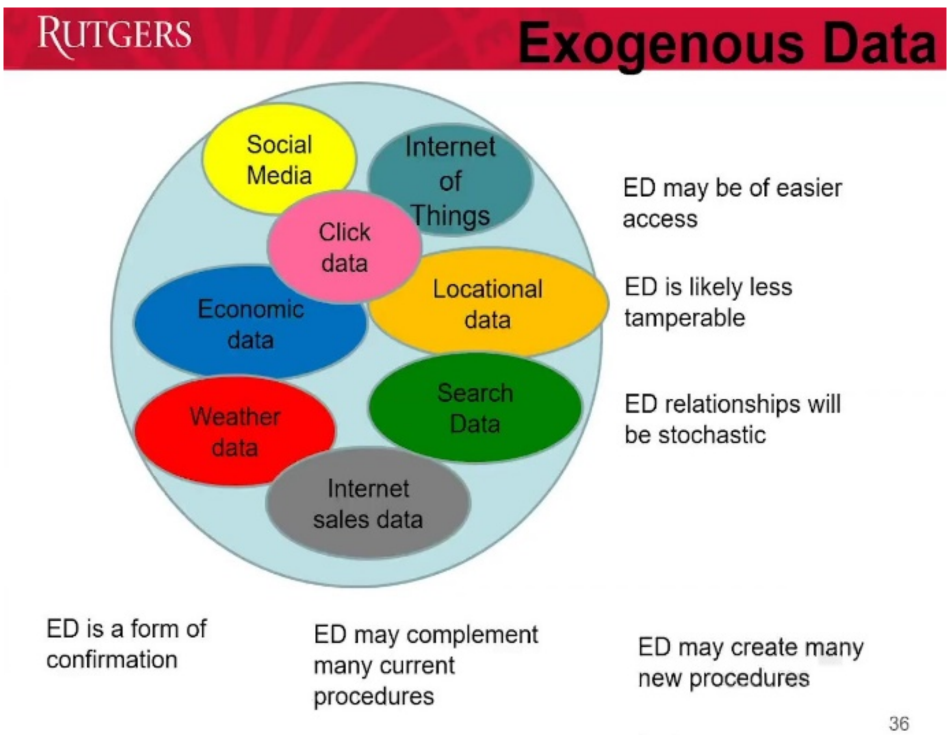
[73] Glen L. Gray, Victoria Chiu, Qi Li, and Pei Li, "The Expert Systems Life Cycle in AIS research: What Does it Mean for Future AIS Research?" *International Journal of Accounting Information Systems*, v. 15 (4), p. 423-451 (https://doi.org/10.1016/j.accinf.2014.06.001, December 2014).

Exogenous Data

Exogenous data are *external* to the company. They are a form of big data, usually unstructured, with **stochastic** (probabilistic) relationships. They differ from the usual company data (endogenous or structured data) that have deterministic (predictable) relationships.

Where do exogenous data come from? From our computers, from mobile devices, and, increasingly, the Internet of Things (IoT). [74] Examples of exogenous data include apps used, calls made, and those in Figure 14. Click data in the figure could include Google searches, advertisement sites visited, site history, Amazon sales. Social media examples could include facial recognition pictures, text analysis, content sharing, connections between users. Other big data could be photographs of parking lots or car licence plates.

FIGURE 14: EXOGENOUS DATA EXAMPLES



No Rules for Use

Exogenous data have no rules for use. Making money from these data has been called **surveillance capitalism**, a phrase introduced by Harvard professor, Shoshana Zuboff. Her 2018 book of the same

---

[74] Internet of Things: By 2035, one prediction* estimates that a trillion IoT devices will be distributed around the world (outnumbering people by 100:1). Today, the Internet is a person-to-person communication network. The Internet of the future will be a device-to-device communication network with few people involved. *See, for example, Philip Sparks, *The Route to a Trillion Devices: The Outlook for IoT Investment to 2035* (White Paper) (Arm Limited: 2017, https://learn.arm.com/route-to-trillion-devices.html).

name "…describes how global tech companies such as Google and Facebook [persuade] us to give up our privacy for the sake of convenience; how personal information ("data") gathered by these companies has been used by others not only to predict our behaviour but also to influence and modify it; and how this has had disastrous consequences for democracy and freedom [e.g., the Cambridge Analytica scandal]."[75] Users trade the loss of their privacy for the benefits of using electronic devices.

"Surveillance capitalism is an assault on human autonomy."[76]

Shoshana Zuboff

LINK TO *FORESIGHT*: Integrity, trust and ethics: CPAs as stewards of the public trust
Successful economies and societies rely on trust, but there is increasing evidence that people are more mistrusting of institutions and professions than ever before. This is a problem for accountants, and they must work to remedy it. The product of accountants' work is trust, so every incentive exists to ensure that integrity and ethical behavior is fundamental to the profession. Accountants must explore how to keep ethics at the forefront of the CPA program, recognizing the difficulty of translating codes and standards through to actual behavior. (p. 26)
LINK TO *FORESIGHT*: Partner on data governance and become the stewards of data integrity
In the digital age, data is more widespread and has become more valuable in and of itself. But at a time when data is so integral to business success and societal progress, there are few standards and frameworks to govern data integrity, security, and application. Given its legacy in assurance and standard setting, the accounting profession is well placed to address key governance issues such as data location, format and extraction; intellectual property (IP) strategy; and privacy. (Appendix 1, p. 25)

Exogenous Data Analytics for Auditing

For auditors, exogenous data may complement existing procedures and create others, with no need to visit a client company. Consider the possibilities of using:

- security recordings of arrivals and departures of trucks from parking lots to assure inventory changes
- photos of store parking lots as a proxy for the number of shoppers

[75] Joanna Kavenna, "[Interview] Shoshana Zuboff: "Surveillance capitalism is an assault on human autonomy,'" *The Guardian* (https://www.theguardian.com/books/2019/oct/04/shoshana-zuboff-surveillance-capitalism-assault-human-automomy-digital-privacy, October 4, 2019). [Note: The Cambridge Analytica scandal is an example of surveillance for behaviour modification for political, instead of commercial, gain.]

[76] Zuboff, Shoshana, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* [book] (New York: PublicAffairs, 2019).

- text mining of telephone records, email, contracts, etc., to validate sales and orders, or to identify discrepancies
- sources of unstructured data and structured data to bring into a continuous auditing process
- video streams on network TV to confirm advertisement placement. If linked to variations in orders/sales, the ad efficiency promised by ad agencies and marketing strategies could be validated.
- social media to determine the popularity of products — and their problems

I believe standard setters need to understand that these data *are* forms of evidence, even if they take a different form than traditional evidence. In addition, exogenous data may be more tamper-resistant than traditional company data and easier to access.

LINK TO *FORESIGHT*: Assuring information

Financial information is only one of the many types of information that will be important in future decision-making. To provide the value of assurance in the future, accountants must deal with a much broader range of information sources. (Appendix 1, p. 24)

CarLab Audit Research

At the Continuous Audit and Reporting Laboratory (CarLab), PhD students tackle problems ranging from AI, blockchain, to government accounting issues. Four examples show the extent of the topic range:

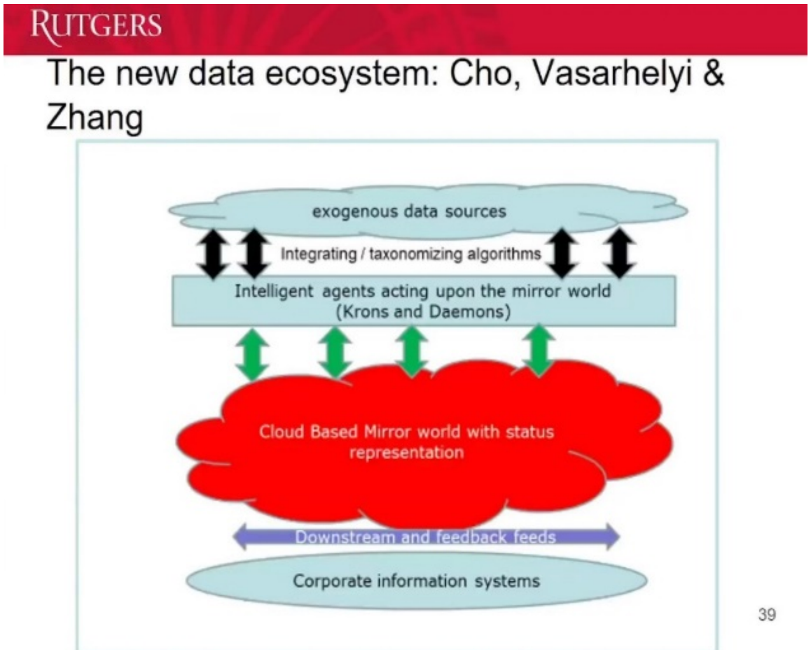
1. **New Data Ecosystem for Auditing.**

We predict that linking unstructured data and structured data in a continuous auditing process will cause big changes. What we call the new data ecosystem for auditing (Figure 15),

"…is a combination of relevant data and the technologies used to integrate information feeding a particular process or a set of functionalities. With the advent of substantially improved computing and storage capabilities, sophisticated algorithms, and easier access to exogenous data, the data ecosystem is facilitating increasingly automated and continuous business measurement and assurance processes that reflect the rhythm and nature of business events and data sources."[77]

[77] Soohyun Cho, Miklos A. Vasarhelyi, and Chanyuan (Abigail) Zhang, "The Forthcoming Data Ecosystem for Business Measurement and Assurance," *Journal of Emerging Technologies in Accounting*, v. 16 (2): 1–21 (https://doi.org/10.2308/jeta-10699, 2019).

FIGURE 15: THE NEW DATA ECOSYSTEM FOR AUDITING



2. **Cognitive Assistant for Auditing: Luca.**

The lab is running experiments to develop a **cognitive assistant** or **intelligent agent** called "Luca,"[78] the "Siri" or "Alexa" for auditing.[79]

It examines the same things as audit tests do today but accumulates auditor knowledge and responds and advises like a cognitive assistant (e.g., Alexa).

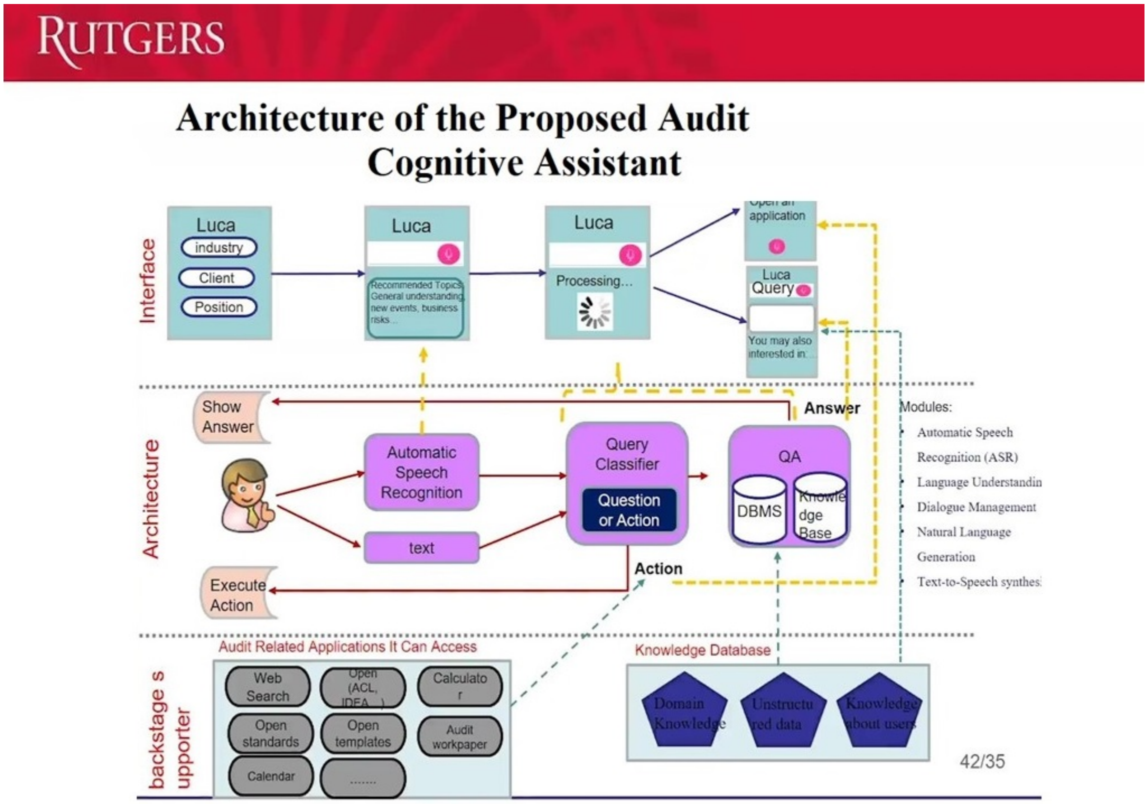
To develop Luca, we simulated the audit brainstorming exercise done at the audit planning stage. A Big Four senior partner explained the audit engagement to a manager who had not been on an engagement, and they discussed the audit plan. We used open-source speech-recognition software to record the session and to collect and organize the words (Figure 16). This "talk-aloud" method of data collection, first proposed by cognitive psychologists, had been used in behavioural research, but not in the development of a cognitive assistant.[80]

[78] Named after 15th century Luca Pacioli, the "father of accounting and bookkeeping."

[79] Li, Qiao and Vasarhelyi, Miklos, "Developing a Cognitive Assistant for the Audit Plan Brainstorming Session," *The International Journal of Digital Accounting Research*, v. 18. 119-140 (https://doi.org/10.4192/1577-8517-v18_5, 2018).

[80] See, for example, K. Anders Ericsson and Herbert A. Simon, "Verbal Reports as Data," *Psychological Review*, v. 87 (3): 215-251 (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.697.3088&rep=rep1&type=pdf, May 1980).

FIGURE 16: ARCHITECTURE OF THE PROPOSED AUDIT COGNITIVE ASSISTANT, "LUCA"



We divided the process into 11 steps to create the device, which, basically, accumulates the experience of partners and engagement managers over different engagements. It does not violate the auditor's client-confidentiality rule, which says data from one engagement cannot be used in another, because it is *experience* that is transferred, not data. The device could be a great experience-gatherer to help the quality of engagements.

Creating a cognitive aid in this way, using open-source software and a little programming, is much cheaper than using IBM Watson, collecting huge amounts of data, and trying to do machine learning. It allows us to break down the data into little pieces so that we can do a lot of interesting things.

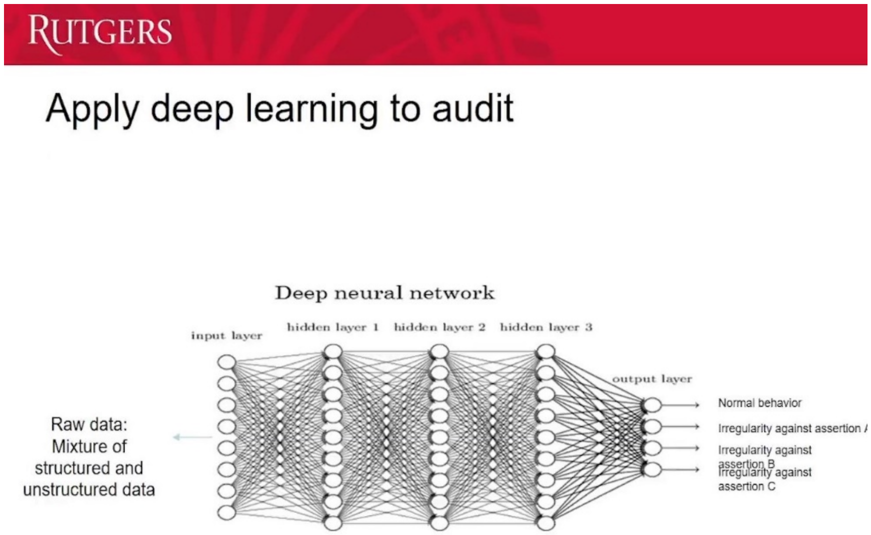
3. **Deep Learning in Auditing.**

a) Another PhD project applies deep learning to analyze big data for predictive auditing. Collecting data and creating a machine learning device for it would be prohibitively expensive. Instead, we found a situation where text data had been collected from analyst calls. We then used IBM Watson to classify the nature of the calls' sentiment.[81]

[81] Sun, Ting and Miklos A. Vasarhelyi. "Embracing Textual Data Analytics in Auditing with Deep Learning." *The International Journal of Digital Accounting Research* 18 (2018): 49-67.

b) Applying deep learning to financial audit. For a given assertion, the auditor's objective is to detect irregularities for this assertion. To do so, an **artificial neural network (ANN)** (Figure 17) can be used.

FIGURE 17: APPLYING DEEP LEARNING TO A FINANCIAL AUDIT



Simplified, the steps are these:

1. Train a huge volume of data (past data), including regular numerical data, semi-structured data (e.g., email, CSV files, data exchange formats), and unstructured data (e.g., video, audio, text).
2. Machine uses **deep learning** to extract features from the raw data (within the black box) and finds patterns (characteristic types).
3. Machine generates a model (Classifier A).
4. Auditors use model A to predict characteristic types from big data.
5. Auditors combine information identified in the last step with regular structured data (e.g., financial data) as **audit evidence**.
6. Past evidence can be used as training data to develop model B (using supervised shallow learning), and model B can be used to predict **frauds** (irregularities).
7. As new data are collected, the machine uses a **reinforced learning technique** to improve the accuracy of model A in a continuous self-correction process.

In 2004, I directed a dissertation in Finland that used machine learning of this type. At that time, only one or two hidden layers were common. Now, 1,000 layers are possible, though 100s are more common. What permits these tremendous competencies is not really algorithmic. Instead, we can store much more information, access it, and process a million times more cycles.

4. **Intelligent Process Automation (IPA) in Auditing.**

"IPA 'takes the robot out of the human.' At its core, IPA is an emerging set of new technologies that combines fundamental process redesign with robotic process automation (RPA) and machine learning. It is a suite of business-process improvements and next-generation tools that assists the knowledge worker by removing repetitive, replicable, and routine tasks."[82]

Federico Berruti, *et al*.

Applying intelligent process automation (IPA) to audit, we divide audit into three parts:

1) **Repeatable tasks.** This is the part I call the "Henry Ford" part: tasks that are "repetitive, standardized, **structured** and rule-based."[83] This part is automatable, and **Robotic process automation (RPA)** can be applied to these tasks.[84] The auditor is "outside the loop."

2) **Judgmental repetitive tasks.** These are **semi-structured** audit tasks that can also be automated. The auditor is "outside the loop." This is **assisted intelligence**.[85]

3) **Judgmental, non-repetitive tasks.** These are **unstructured** tasks that can be partially automated, but the auditor must be "in the loop." This is **augmented intelligence**.

Rutgers PhD student, Abigail Zhang, explains:[86]

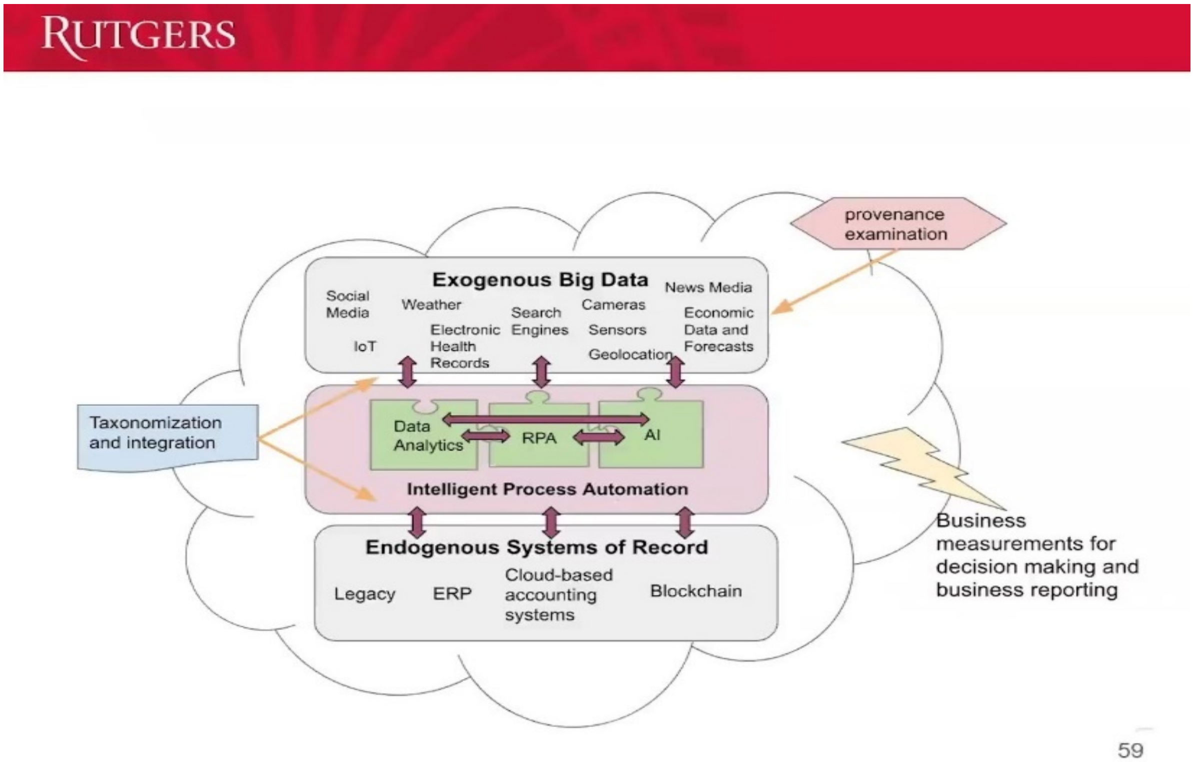
"The **structured** and **semi-structured** tasks (e.g., reconciliations, recalculations, and confirmations) can be automated in an unattended way, almost without auditors' intervention. The **unstructured** tasks (e.g., internal control risk assessment) can be partially automated (i.e., *attended* automation) in the sense that AI/cognitive computing will *assist* auditors. In the IPA ecosystem, auditors and bots will augment each other's abilities and cooperate to complete audit engagements.

"Also, there could be a major benefit from IPA in the audit: increased audit quality. First, the auditors can focus more on the tasks that need professional judgment, and they can make better decisions with the assistance of AI and cognitive computing. Second, automation can almost eliminate human errors in repetitive tasks. Third, IPA allows auditors to perform full population testing, which can provide more

[82] Federico Berruti, Graeme Nixon, Giambattista Taglioni, and Rob Whiteman, "Intelligent process automation: The engine at the core of the next-generation operating model," *McKinsey Digital* [website] (https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/intelligent-process-automation-the-engine-at-the-core-of-the-next-generation-operating-model, March 2017).
[83] [Abigail Zhang], "Beyond Robotics: How AI Can Help Improve the Audit Process," *AICPA [Guest Blogger]* (https://blog.aicpa.org/2018/08/beyond-robotics-how-ai-can-help-improve-the-audit-process.html#sthash.60CmoTjh.dpbs, August 1, 2018).
[84] For more on RPA and tasks to which it can be applied, see Eric Santor's paper in this volume.
[85] For more information on the four ways to apply AI, see Anand Rao's paper in this volume.
[86] [Abigail Zhang], "Beyond Robotics: How AI Can Help Improve the Audit Process," *AICPA [Guest Blogger]* (https://blog.aicpa.org/2018/08/beyond-robotics-how-ai-can-help-improve-the-audit-process.html#sthash.60CmoTjh.dpbs, August 1, 2018).

convictive support for the audit opinion than traditional sampling methods."Figure 18 depicts the IPA ecosystem.

FIGURE 18: THE IPA ECOSYSTEM



Conclusion

Computerization of Occupations

Researchers at the University of Oxford looked at the future of employment and how susceptible jobs are to computerization.[87] Their analysis used "detailed work activities," as defined by the U.S. Department of Labor, Employment, and Training Administration. They gave these probabilities for automation:

- Accountants and auditors — 94%
- Bookkeeping, accounting, and auditing clerks — 98%
- Tax preparers — 99%

Does this mean that CPAs will not have any jobs? I do not think so; not immediately. What will the degree of automation be in these jobs? New technology takes time to implement — remember,

[87] Carl Benedikt Frey and Michael Osborne, *The Future of Employment: How Susceptible are jobs to computerisation?* [Working Paper] (Oxford Martin School, University of Oxford: September 1, 2013, https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf).

continuous auditing has taken 30 years. In addition, new jobs appear. How many people developed code for cell phones 30 years ago?

Predicting *what* is going to happen is reasonably easy but predicting *when* it will happen is very difficult. People overestimate the impact of technology in the short term and underestimate its impact in the long term.

Risks of AI and Responsible AI

***Anand Rao**, Global & U.S. Artificial Intelligence and U.S. Data & Analytics Leader, PwC U.S.*

"…AI is both closer and farther off than we imagine…"[88]

Mark Zuckerberg

Introduction

Involved since 1985, and having seen AI's ups and downs, I say we can be cautiously optimistic about AI. We can be optimistic because of new advances in deep learning and so on, but cautious, given AI's 64-year history. If adoption is dependent upon *trust*, then addressing its risks, especially in critical areas like healthcare and finance, is imperative. This paper examines the future of AI adoption from a business perspective: its capabilities and risks and how a responsible AI framework can help businesses and nations deal with those risks.

AI is closer than we imagine…

AI is closer than we imagine, because of breakthroughs in deep learning and gaming. For example:

- In 1997, IBM's computer Deep Blue beat the chess champion Gary Kasparov at the ACM Chess Challenge. At that time, critics said that while the win was impressive, chess has a huge, but finite number of moves.
- In 2011, IBM Watson beat the Jeopardy! champion, ending a myth that AI cannot handle open-ended questions.
- In 2016, AI company DeepMind Technologies' program AlphaGo beat the Go champion, South Korean Lee Sedol. Go is an ancient game where the number of choices is more than the number of atoms in the universe, so the win ended the myth that AI could only handle finite choices. This victory caught Asia's attention.
- In 2017 AlphaGo's victory over Chinese teenager Ke Jie was so important that it "…became China's 'Sputnik moment': the incident that catalyzed Beijing's feverish rush to match American capabilities in AI."[89] To build AlphaGo, DeepMind researchers took data from human Go players. They created neural networks on two computers that played against each other. Unlike humans, computers need no breaks, no food, no water—just computer power. Run billions of times, one AI system would win, the other would lose, but both would *learn* from the losses and victories.

---

[88] Mark Zuckerberg, "Building Jarvis," *Facebook* (https://www.facebook.com/notes/mark-zuckerberg/building-jarvis/10154361492931634, December 19, 2016).

[89] Matt Schiavenza. "China's 'Sputnik Moment' and the Sino-American Battle for AI Supremacy," *Asia Society* (https://asiasociety.org/blog/asia/chinas-sputnik-moment-and-sino-american-battle-ai-supremacy, September 25, 2018).

- In 2017, DeepMind's AlphaGo Zero, a neural network starting with the rules of the Go game, but no game data (human or computer), beat all past versions of AlphaGo within 40 days of its deployment. It used reinforcement learning to become its own teacher as it played games against itself.[90]

I use these examples to show AI's rapid, recent progress and to say that, in business, we can learn from them.

In business, we often change strategy. Strategy is nothing but a company going into a new business, product, or geography and then trying to win the game of market share, competition, or margins. The examples are useful because we can use some AI techniques to gamefy strategy. That said, headlines stemming from events like those in the examples lead us to believe that AI will soon rule and will take away all of our jobs.

But AI is much farther off than we think…

AI is much farther off than we think, however, because **deep learning** is *not equal* to **deep understanding**. Consider the following very simple example. I use Alexa all the time. If I ask Alexa about the weather, she answers me for my location. If I then ask, "Alexa, what was my previous question?" Alexa gives me a *definition* of "previous question." It just does not understand my meaning; it does not understand *context* or language. While this problem might be fixable, it also ties into the maintenance and monitoring of AI performance raised by Eric Santor in his talk.[91]

The Digital Revolution

*"When we say digital revolution, it is all about **digitization**: taking anything in the physical- or online world and converting it into numbers. Once it is in numbers, we can do all kinds of things with it."*

Anand Rao

Two paths to AI

Enterprises are realizing value along two distinct paths from digitization to AI:

- **Automation Path.** Big data is standardized ➔ simplified ➔ a process is automated; e.g., **robotic process automation**[92] (e.g., robo-calls, email broadcasting) and diligent process automation ➔ productivity is improved.
    - Forty per cent of the $15.7 trillion increase to global GDP, quoted by Fakhri Karray from a PwC report, comes from productivity improvement.[93]

[90] For more information, see DeepMind, *AlphaGo Zero: Starting From Scratch* (https://deepmind.com/blog/article/alphago-zero-starting-scratch, October 18, 2017)
[91] See also Eric Santor's paper in this volume.
[92] For a detailed definition of robotic process automation, see Eric Santor's paper in this volume.
[93] See Fakhri Karray, in this volume. From Anand S. Rao and Gerard Verweij, *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* (PwC: 2017, https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf).

- **Analytics Path.** Data ➔ Personalization (e.g., Netflix, Amazon, Google customize search results personalized for each user) ➔ Cognification ➔ Analytics ➔ AI.
    - **Cognifying** some of the collective thinking in any industry, from healthcare to finance, leads to better decisions, and that is the **analytics** process. Cognifying means making things smart. In healthcare that may mean capturing the expertise of clinicians.

Both paths lead to AI. The automation path could be called the *top line*, whereas the analytics path could be called the *bottom line*, because it will improve revenues, profits, etc.

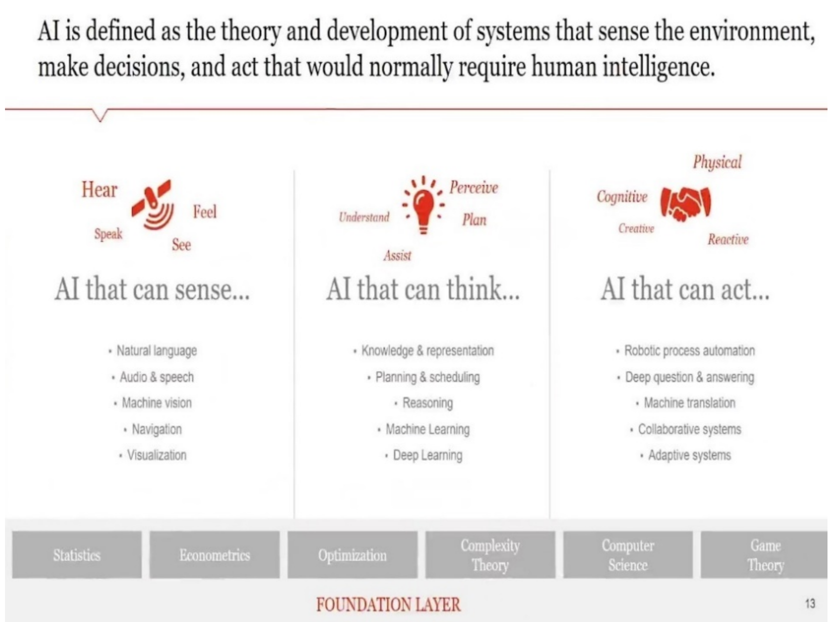
AI definition

A classic definition of AI is any computer system that can *sense*, *think*, and *act* in an environment, made up of other humans or other machines, to solve a specific objective (see Figure 19). By "specific," we mean **narrow AI**.

Sensing, thinking, or acting is **operational AI** (the top layer of Figure 19). Its theoretical foundation is called **foundational AI** (the bottom layer of Figure 19), which includes statistics, computer science, econometrics, and other areas.[94]

- **Sensing** includes natural language (Alexa) and image recognition (medical images, car images).
- **Thinking** includes machine learning and deep learning.
- **Acting**, either in the robotic sense or in the digital sense, includes robo-calling and deep questioning and answering (that IBM Watson can do).

FIGURE 19: AI THAT CAN SENSE, FEEL, AND ACT



[94] For definitions of operational and foundational AI, see also Fakhri Karray's paper in this volume.

Four Ways to Apply AI

In finance, accounting, and other professions, AI is being applied in four distinct ways (Figure 20):

- **Automated intelligence** automates repetitive tasks with fixed rules that humans do not need to perform.
- **Assisted intelligence** is a rules-based system that might require human judgment. For example, companies using rule-based underwriting systems for loans or insurance might automate the process and require human judgment before final approval. In assisted intelligence, the machine is not learning. Instead, it is a **tool**.
- **Augmented intelligence** is human-machine collaboration, where the system is learning continuously the background, and humans in the loop are also learning. The human might learn from the system's suggestions; the system might learn the human's preferences. People are still needed, and the system is adaptive. Examples include medical diagnosis, underwriting or Netflix recommendation.[95]
- **Autonomous intelligence** has not yet been reached, technically or socially. The dividing line between augmented and autonomous intelligence is *trust*, and that takes time to build. Autonomous driving would remove humans from the loop. The system would be completely adaptive and, no matter what the driving conditions, the car would go from point A to point B. More than anything else, we as humans would need to trust the system to deliver us — or our loved ones — safely.

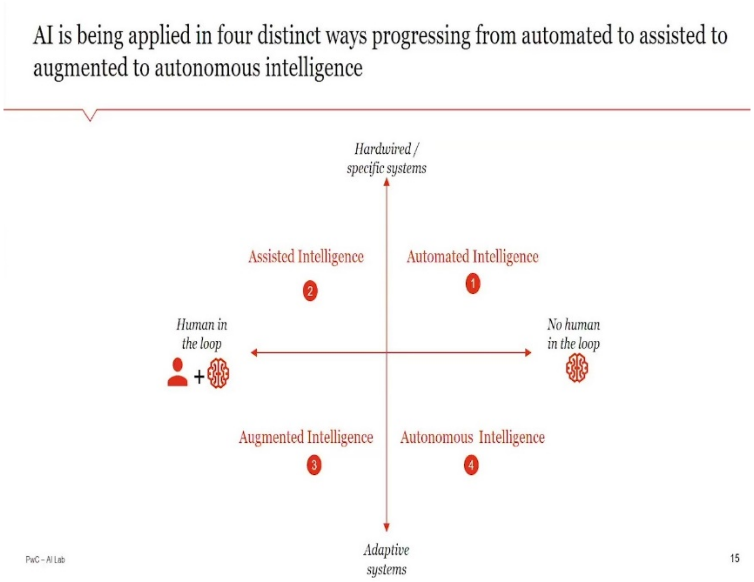
HOW WILL PEOPLE TRUST AI AND MACHINE LEARNING WHEN IT COMES TO LIFE-ALTERING PROCEDURES LIKE SURGERY?

PARVIN MOUSAVI: Over time that trust will be built based on precedents and statistics…and reliability and transparency. With augmented intelligence, the physician is in control—not an elephant, not a machine. That is another way of building trust.

ANAND RAO: Trust will build over time, and that is why I think it is very key to always have humans *in* the loop as opposed to humans *outside* the loop.

[95] See also Parvin Mousavi's paper in this volume.

FIGURE 20: AI IS APPLIED IN FOUR DISTINCT WAYS



AI in Finance

For finance, AI is going to:

- **Reduce costs and increase efficiency.** For example, at PwC, we do audit, tax, and consulting. On the audit side, in the past, most of the junior consultants were going through detailed 300-, 400- page documents then typing extracted information into an Excel spreadsheet. The work was boring and routine. Now, a machine will do all of that, but we still want people to check its work. When we gain confidence that the machine's work is correct, that is the point when we can just give it back to the machine.
- **Improve quality and control.** Rote, mechanical, no value-added work *will* be replaced. A key impact is that as the machine takes over the low value-added work, the humans had better start adding value. So one of the things that I tell all of our staff, whichever line they come from, is that if you are not adding value to whatever you are doing, then you had better rethink what exactly you are trying to do.
- **Improve decision making.** With 100% auditability replacing sampling, decision making can be more forward- than backward-looking. Looking forward allows *what-if analysis*. AI becomes the cockpit where CFOs can essentially move levers and ask, "I want greater market share, but I also want the profits. I only have so much investment to get that market share. How do I get there?" AI can simulate these scenarios and allow them to change course or make in-flight corrections.[96] So, a CFO is adding value in terms of their own thinking, augmented by AI, as opposed to just following the numbers.

[96] For more information, see PwC, *Stepping into the Cockpit: Redefining Finance's Role in the Digital Age* (https://www.pwchk.com/en/migration/pdf/finance-role-in-digital-age-jan2016.pdf, 2016).

What AI *Can* Do

AI can:

- **find patterns** in apparently random data or apply structure to unstructured data
- **plan** and think ahead
- **learn** through repeated exposure to particular problems
- **make sense** of human communication (speech or text); interpret and identify rich media (e.g., music and images)

What AI *Cannot* (Yet) Do

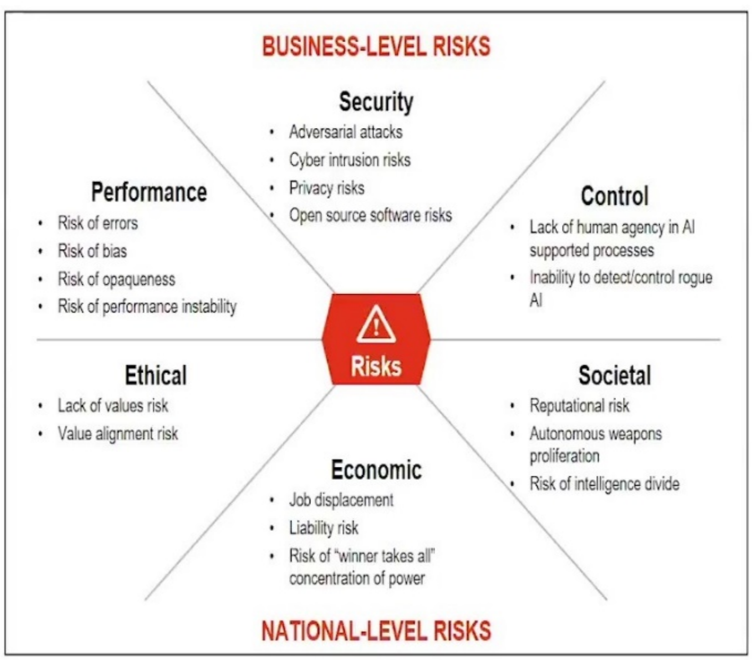
AI cannot:

- **place its work in context**, because it cannot see the big picture
- **self-improve**, because it is only as good as the data it has been trained on
- **reason**, because it has no common-sense and cannot reason from first principles
- **multi-task**, because its capability is limited to the purpose for which it was built

AI Risks

AI risks need to be assessed, mitigated, and managed. PwC has tried to categorize risks that impact consumers, businesses, societies, and nations into business-level risks and national risks (Figure 21).

FIGURE 21: AI'S BUSINESS- AND NATIONAL-LEVEL RISKS



Business-Level Risks

Business-level risks fall under three categories:

1) **Performance Risks**

- errors
- bias
- opaqueness
- performance instability

Example: If historical data are biased, then AI recommendations are biased. Bias can lead to unfairness, and the definition of fairness comes into play.

2) **Security Risks**

- adversarial attacks
- cyber-intrusion risks
- privacy risks
- open-source software risks

Example: A deep learning algorithm for an autonomous vehicle will recognize a stop sign. But after changing the sign with a few white labels and black labels, the system does not see a stop sign. Instead, it sees a speed limit sign (see Figure 22). A person might maliciously change the sign, but snow and dirt may have caused the change. Suddenly an autonomous car interprets "Stop" as "Go" and causes an accident. We are trying out the technology, but is it safe enough to deploy?

FIGURE 22: ALTERED STOP SIGN: VERY DIFFERENT MESSAGES

"Stop" is interpreted as 'Go":

3) **Control Risks**

- lack of human agency in ai-supported processes
- inability to detect/control rogue AI

Example: A telco provider introduced a machine-learning chat bot in December to reduce customer wait times by answering calls not requiring a customer service representative (CSR). In January, the performance was great; in February, performance deteriorated; in March, it deteriorated again. By April, it was passing almost all calls to a human CSR. Why? A human would know that iPhone 11 was so new that when documentation on accessories was missing, the Apple website was a resource. A human could retrieve that information, send it to the customer, and file a note saying the documentation is missing. The chat bot knew none of this, so it just said, "I don't know; I'm going to pass this on." From a governance perspective, how would the company find out about this? Did they know when performance dropped off? Do they now have procedures to check that?

National-Level Risks

National-level risks fall under three more categories:

4) **Societal Risks**

- reputational risk
- autonomous weapons proliferation
- risk of intelligence divide

Example: In 2018, a Barack Obama video, completely fake, circulated on the Internet.[97] Everything about it — voice, lip movement, and content — were fake. So, reputational damage can happen quickly. **Deepfake** is a huge threat.

5) **Economic Risks**

- job displacement/job loss
- liability risk
- risk of "winner takes all" and concentration of power

Example: If large corporations control AI, they may exclude small players from the game.[98]

6) **Ethical Risks**

- lack-of-values risk
- value-alignment risk

Example: From data collection and use to deepfake, AI used unethically or maliciously is a risk.[99]

LINK TO *FORESIGHT*: Shaping the world's data governance and becoming stewards of data integrity.

As a result of its legacy in assurance and standard setting, the CPA profession is well-placed to contribute to a much-needed set of initiatives that must ultimately develop around data governance…Issues such as data residency, consistency of data format, privacy and national intellectual property strategies, must ultimately be addressed at a policy level. In a world of fake news, there is no more valuable work than to ensure that decision makers can trust data. (p. 25)

[97] For more information, see "Tricked by the Fake Obama Video? Deepfake Technology, Explained [video]," *USA TODAY*, (https://www.youtube.com/watch?v=EtEPE859w94, April 30, 2019).

[98] For more information on small companies and AI, see also papers by Fakhri Karray and Graham Taylor in this volume.

[99] For more information on the use of AI for good or harm, see Graham Taylor's paper in this volume.

Responsible AI

PwC has developed a framework for implementing responsible AI in an organization (Figure 23) that looks at dimensions of responsible AI. [100]

Ethical and Societal

- **Ethical and legal.** Ensure AI development is in line with major local and global regulations, enacted and emerging. Allow the business to evaluate the ethics of an AI system and how to operationalize ethics in the organization. As an experiment, PwC used AI in the form of a natural language processing engine (NLP) and topic model to go through documents related to AI and ethics, to extract ethics principles. The resulting *Ethical Principles Traceability Matrix* can be used as a tool by global businesses to see, first, how principles differ from one organization to another and from country to country and, next, to further develop their own programs.

Performance and Security

- **Bias and fairness.** Uncover potential bias in the underlying data, model, and human interaction that could lead to unfairness of AI. Enable the business to understand what processes may lead to unfairness.
- **Interpretability.** AI users must increase their understanding of AI systems and tackle the black-box problem. **Interpretability** and **transparency** are important at the model level; **explainability** is understanding the reasoning behind a decision. Enable human users to understand, appropriately trust, and effectively manage the emerging generation of AI.
- **Robustness and security.** Assess the performance of AI over time to identify potential disruptions or challenges to long-term performance. Is the AI behaving as intended?

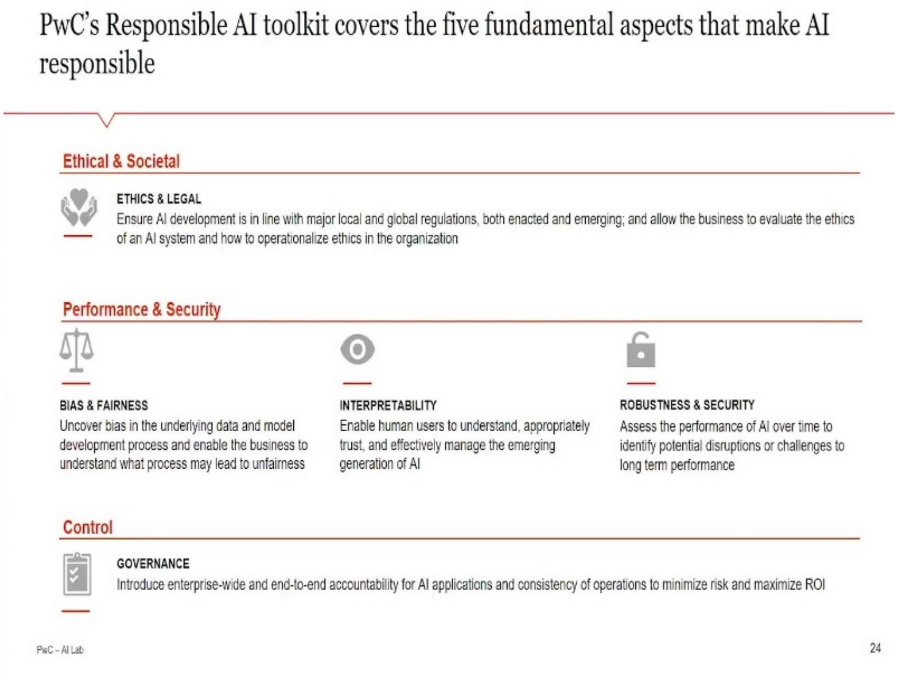
Governance

- **Introduce enterprise-wide governance.** Revamp existing governance structures across the entire organization to account for new AI-related issues. Ensure accountability and consistency of operations from the board level to the model level to minimize risk and maximize ROI.
- **Ensure end-to-end governance (across stakeholders).** Ensure auditability by allocating responsibility, accountability, and controls for AI. Stakeholders include the organization's ecosystem of third-party vendors (for example, AI startups). The organization must assume responsibility for third-party tools that are part of its processes.
- **Operate and monitor** the AI once implemented.

---

[100] For more information, see PwC, "PwC's Responsible AI Toolkit," *PwC's Responsible AI: AI you can trust* (https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html, 2017).

FIGURE 23: PWC'S RESPONSIBLE AI FRAMEWORK



Conclusion

"The most important skill to have as future CPAs is the ability to learn, learn…and learn fast."

Anand Rao

"Given the massive opportunities and potential risks associated with AI, companies, global bodies, nonprofit groups, citizens, and policymakers must come together to devise the right strategies that consider the various trade-offs in ways that make sense in their country. Not having a coherent, comprehensive national strategy could put future generations at a competitive disadvantage." [101]

More than 40 countries are working with PwC on national AI strategies under six important policy categories:

- Basic AI R&D: funding
- Specialized AI technology: e.g., drones, autonomous vehicles, service robots
- Business protection (and promotion) and algorithmic governance
- Workforce reskilling
- Consumer protection, e.g., data security, income security, digital anonymity
- Ethics and regulation, e.g., citizen monitoring, autonomous weapons, beneficial uses of AI

[101] Anand Rao, *Gaining National Competitive Advantage through Artificial Intelligence (AI): Policy Making & National AI Strategies* (PwC: 2019, https://www.pwc.lu/en/technology/docs/gaining-national-competitive-advantage-through-ai.pdf), p. 22.

To help businesses follow national developments, PwC has developed a tool called National AI Strategies (NAIS) Radar, "…that analyses all documents on national AI Strategies and performs topic modeling and summarises the key policy recommendations being made by different countries."[102]

Canada was the first country to have a national AI strategy, seeded by $125 million in 2017 by the Canadian government. The Canadian-based global charitable organization, **Canadian Institute for Advanced Research (CIFAR)**[103], was appointed to develop and lead the strategy, in partnership with three AI centres of excellence in Canada: the **Alberta Machine Intelligence Institute (Amii)**[104] in Edmonton, the **Vector Institute**[105] in Toronto, and the **Quebec Institute for Learning Algorithms (Mila)**[106] in Montréal.

For students, I say, the most important skill to have as future CPAs is the ability to learn, learn…and learn fast.

---

102 Anand Rao, *Gaining National Competitive Advantage through Artificial Intelligence (AI): Policy Making & National AI Strategies* (PwC: 2019, https://www.pwc.lu/en/technology/docs/gaining-national-competitive-advantage-through-ai.pdf), p. 14.

103 Canadian Institute for Advanced Research (CIFAR) (https://www.cifar.ca).

104 Alberta Machine Intelligence Institute (Amii) (https://www.amii.ca/).

105 Vector Institute (https://vectorinstitute.ai/).

106 Montreal Institute for Learning Algorithms (Mila) (https://mila.quebec/en/).

AI and Digitalizing the Financial System

***Eric Santor**, Advisor to the Governor of the Bank of Canada*

"…[T]he digital economy is pervasive — it is not so much a piece or sector or industry of the economy, rather it is transforming the entire economy." [107]

James Tebrake, former Director General of Macroeconomic Accounts, StatCan

Introduction

Employment growth in the digital economy in Canada is more than four times greater than in the rest of the economy. AI has transformed decision-making algorithms from rules-based to prediction-based. Because AI lowers the cost of prediction, problems that were not historically prediction problems have been *turned into* prediction problems, so AI's use as a general-purpose technology will grow.[108]

Along with AI's growth is the need to address its governance, from making output explainable and free from bias, to taking responsibility for model maintenance and ethical data sourcing.

"For the accounting profession (and others), many tasks, many jobs, will be replaced by this fourth industrial revolution. However, new jobs will arise, and the value of human judgment will be higher."

Eric Santor

At the Bank of Canada, partnering with Canada's AI community is transforming forecasting for economic policy-setting.

Canada's Digital Economy

We know that digitalization is affecting every aspect of our society and every aspect of the economy. It is affecting almost everything we do. There is a lot of hype around it, and we are seeing lots of use cases. One of the big difficulties we have when we are thinking about the digital economy, though, is that it is actually quite hard to see in National Accounts data.

---

[107] Loranger, André, Amanda Sinclair, and James Tebrake, "Measuring the Economy in an Increasingly Digitalized World: Are Statistics Up to the Task?" In *Data Governance in the Information Age: ACIGI Essay Series*. Waterloo, Canada: Centre for International Governance Innovation, 2018. [The full report: https://www.cigionline.org/sites/default/files/documents/Data Series Special Reportweb.pdf. The Essay: https://www.cigionline.org/articles/measuring-economy-increasingly-digitalized-world.]

[108] Ajay Agrawal, Joshua Gans, and Avi Goldfarb, "The Simple Economics of Machine Intelligence," *Harvard Business Review* (https://hbr.org/2016/11/the-simple-economics-of-machine-intelligence, November 17, 2016).

In 2019, Statistics Canada (StatCan) tried to measure the size of investments in data, data sciences, and databases in Canada. While not right now part of the National Accounts, this part of the digital economy could make up about $29 billion to $40 billion of Canada's $2 trillion economy, or up to 2% of GDP. [109] That investment is bigger than investment in machinery and equipment, bigger than the investment in transportation, and bigger than the investment in R&D. The upper limit of the net capital stock of data-related assets was estimated at $217 billion for 2018.[110] By comparison, that of the oil sands is estimated to be about $300 billion.[111] So, the amount of investment in data that has already occurred is huge and will only increase.

Employment growth in the digital sector in Canada was almost 40% between 2010 and 2017, or just over four times the employment growth in the rest of the economy, which sits between 8% or 9%.[112] Again, this percentage will only increase because younger generations are connected: their lives are digital.

Machine Learning as a General-Purpose Technology

"Machine intelligence is, in its essence, a prediction technology, so the economic shift will center around a drop in the cost of prediction." [113]

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb

AI and big data allow us to take a problem — a prediction problem — and lower the cost of making that prediction. So, more predictions, better predictions, and faster predictions will be sought, even where prediction was not historically an input. As the cost of prediction is lowered, the value of its complement will be raised. That complement is human judgment.[114]

Will some tasks of professional accountants, economists, lawyers, and others be replaced by AI? Yes, and AI will make them faster, cheaper, better. As this general-purpose technology filters through the

[109] Statistics Canada. 2019. "The Value of Data in Canada: Experimental Estimates," *Latest Developments in the Canadian Economic Accounts*. (Catalogue, no. 13-605-X, https://www150.statcan.gc.ca/n1/en/pub/13-605-x/2019001/article/00009-eng.pdf?st=Bouygwzw), p. 8. [The report "presents a preliminary set of statistical estimates of the amounts invested in Canadian data, databases and data science in recent years." (p.3)]

[110] Ibid., p 9. "[The Statistics Canada report] proposes that a significant amount of 'information-related' activity creates stores of value, from which firms draw in subsequent periods to produce goods and services. Given that own-account data, database and data science investments are being made by businesses, governments, and non-profit institutions each day, a stock of these assets is also being accumulated. This stock needs to be included on the balance sheet of the sector that owns it, at its market value."

[111] Natural Resources Canada, *Crude Oil Facts – Oil Sands* (https://www.nrcan.gc.ca/science-data/data-analysis/energy-data-analysis/energy-facts/crude-oil-facts/20064, 2020).

[112] Statistics Canada, "Measuring digital economic activities in Canada, 2010 to 2017," *The Daily* (https://www150.statcan.gc.ca/n1/daily-quotidien/190503/dq190503a-eng.htm, May 3, 2019).

[113] Ajay Agrawal, Joshua Gans, and Avi Goldfarb, "The Simple Economics of Machine Intelligence," *Harvard Business Review* (https://hbr.org/2016/11/the-simple-economics-of-machine-intelligence, November 17, 2016).

[114] Ibid.

economy, it will lower the cost of prediction for many, many decisions that professionals make. At the same time, human value — the value-added from judgment — will be much higher. That is important.

Turn Problems into Prediction Problems

AI's strength is in its predictive algorithms for prediction problems. Many things have prediction in them. Tasks can either *look* like prediction problems or they can be *turned into* prediction problems. For example:

- **Medical diagnostics** models can predict whether something is malignant or benign or is cancerous or not.[115]
- **Buying products** or watching movies, when treated as prediction problems, result in algorithms that make recommendations on what we should buy or view *next*. Should we be concerned about privacy implications when viewing something on one computer results in targeted suggestions on another computer?
- **Language translation** is not, intuitively, a prediction problem, but it was turned into one. The company DeepL "…trains artificial intelligence to understand and translate texts" from one language to another by using deep learning neural networks.[116]
- **Self-driving vehicle algorithms**, once approached as rule-based (i.e., if something happens, do this; if something happens, do that), are now *prediction*-based (i.e., what would a good driver do if faced with a particular problem?)

In the financial sector, AI as a general-purpose technology is applied to many prediction problems. Financial institutions need to predict behaviour of the economy and make predictions in every dimension of their business. They use machine learning in the following areas:

- **Loans.** Who is a good credit risk?
- **Investments.** What will portfolio returns look like? Robo-advisors, AI, machine learning, and big data are now used to inform portfolio decisions.
- **Trading.** Automatic algorithmic trading has been around for a long time; now it is getting even faster.
- **Insurance.** AI provides many opportunities to more precisely measure risk. Optimistically, some people suggest that the zone of insurability will expand, because insurance companies will be able to better calculate risk or insure against extreme events in a different-than-conventional way.
- **Robotic process automation.** RPA runs all of the formerly manual processes in a financial institution's back office, not just automating them, but making them smart.

[115] For more information, see Parvin Mousavi's paper in this volume.
[116] "AI Assistance for Language," *DeepL* [website] (https://www.deepl.com/en/home).

ROBOTIC PROCESS AUTOMATION (RPA)

Robotic Process Automation is a set of capabilities of software automation that can handle high-volume, repeatable tasks such as answering questions, making calculations, maintaining records and recording transactions. All are tasks that previously required humans to perform them. RPA can mimic tasks performed by humans and automate those tasks digitally. RPA systems are easily programmed and modified by non-technical workers…No additional decisions need be made, and no learning occurs as a result of performing the tasks. Therefore, it is important to note that RPA on its own is not AI.[117]

AI Governance Needs to Evolve

While rushing ahead to use AI and to include machine learning and big data in our work, we need to also think about how AI governance needs to evolve. If machine learning algorithms and big data are introduced into decision making, what governance is needed? What issues must the governance structure consider? We want to ensure that *people* are accountable for making decisions. How do we govern that?

In conventional banking, a *person* is responsible for extending or rejecting a loan and would have reasons for doing so. Machine learning and big data add another layer of complexity to that set of decisions, and that is what we need to worry about. We need to think about five key issues:

- **Explainability.** Setting up governance or accountability is difficult when the thing helping you make your decision is really difficult to explain. Even so, explainability implies a level of transparency that we must be mindful of. Imagine an economist at a bank making a forecast about GDP and sending it to the CEO's office. If the economist's justification for the forecast is, "The machine told me so," the CEO is not likely to be satisfied. Similarly, rejecting a loan applicant with the justification, "The machine said you shouldn't get a loan," is just not good enough.
- **Bias.** A big problem in machine learning algorithms is bias, either because the developer inadvertently or deliberately includes bias in the code, or the training data are inherently biased (e.g., culturally, geographically, etc.).[118] We have to be mindful of the fact that while bias occurs in machine learning, we also have our own biases. With machine learning, discovering and correcting bias may be possible in an objective way.
- **Model Maintenance.** Investing in human capital is necessary to make sure models are running well, and as intended, on an ongoing basis. Machine learning algorithms, in many senses, are alive.

[117] CPA Canada, *A CPA's Introduction to AI: From Algorithms to Deep Learning, What You Need to Know*. (https://www.cpacanada.ca/en/business-and-accounting-resources/other-general-business-topics/information-management-and-technology/publications/a-cpa-introduction-to-ai, 2019), p. 17.

[118] For more on dealing with bias and fairness in training data, see Richard Zemel's paper in this volume.

Reinforcement learning is ongoing. When an algorithm runs, its predictions and success/failure rate feed back into the dataset and into the algorithm. Without monitoring, results could drift over time and steer decision making in an unintended direction. So, models must be more actively managed than in the past.

- **Data Ownership.** We need to consider who owns and controls the data we use. Do we own data, or did we purchase data? Was it collected ethically? Risks from third party data are very important. Responsible and ethical use of data and data ownership are broad issues currently receiving a great deal of attention. Who owns and controls the artefacts resulting from data run through a machine-learning algorithm? What data are allowable to use? What data *should not* be used? How much collected information is *too much* information?
- **Accountability.** As a group of people who look at financial statements, how firms operate, and their compliance with rules, CPAs think about whether standards are being met. Right now, standards for machine learning are amorphous. Certainly, these are things CPAs could think about much more and find a role in defining.

LINK TO *FORESIGHT*: Mastering and shaping a data-driven economy.
Every CPA must ultimately become comfortable in a world that will be data-rich, data-intense, and data driven. In concert with this, CPAs should play a role in the development of standards for data governance and data integrity. (p. 10)
LINK TO *FORESIGHT*: Partner on data governance.
Given its legacy in assurance and standard setting, the accounting profession is well placed to address key governance issues such as data location, format and extraction; intellectual property (IP) strategy; and privacy. (Appendix 1, p. 25)

The Central Bank's Imperative

At the Bank of Canada, we look outside to the world. The world is changing. It is going digital. Digitalization is affecting every aspect of our economy.

The implication of this for setting economic policy was addressed by Bank of Canada governor Stephen Poloz in a speech and paper called *Technological Progress and Monetary Policy: Managing the Fourth Industrial Revolution*, in November 2019.[119] Poloz compares the widespread adoption of AI and machine

[119] Stephen S. Poloz, "Technological Progress and Monetary Policy : Managing the Fourth Industrial Revolution (Bank of Canada Staff Discussion Paper 2019-11)" (https://www.bankofcanada.ca/wp-content/uploads/2019/11/sdp2019-11.pdf, November 2019).

learning, the fourth industrial revolution to the previous three other industrial revolutions based on several shared characteristics.[120]

- New technology displaces workers
- Investor hype linked to the new technology leads to financial excesses
- New types of jobs are created
- Productivity and potential output rise
- Prices and inflation fall
- Real debt burdens increase, which can provoke crises when asset prices crash

Governor Poloz says these changes were described by the economist Joseph Schumpeter 80 years ago as "as a process of *creative destruction*."[121] With relevance for both students and seasoned professionals in the audience today, Poloz continues,[122]

"It is human nature to worry mainly about the destruction part of the story. Economic history is littered with tales of resistance to new technology because individuals could not imagine the final destination or what role they might play in it. In part, this is because that destination may be quite far off. Individuals are disrupted by the introduction of a new technology, and it may be a long time before they are able to find a new place in the labour market, if they ever do. Even so, history demonstrates that technological progress ultimately creates more jobs than it destroys, and the overall impact on economic growth warrants at least some of the optimism that tends to infect asset markets during such episodes."

Applying AI to Understand AI

To understand the world of digitalization, the central bank must use that technology to understand it. I argue that in order to be the premier authority on the Canadian economy and to fulfill our mandate of promoting the economic and financial well-being of Canadians, the central bank must also adopt those technologies in an appropriate way.

That means that we are using machine learning and big data to try and improve our forecasts, e.g., on house prices or GDP. To understand how sentiment is changing in the economy or a particular industry, we use natural language processing on datasets of text. Where it makes sense, we are introducing RPA and are leveraging the cloud.

---

[120] For more information on the on the fourth industrial revolution, see also Fakhri Karray's paper in this volume.
[121] J. A. Schumpeter, *Capitalism, Socialism, and Democracy* (New York: Harper and Brothers, 1942).
[122] Stephen S. Poloz, "Technological Progress and Monetary Policy: Managing the Fourth Industrial Revolution (Bank of Canada Staff Discussion Paper 2019-11)" (https://www.bankofcanada.ca/wp-content/uploads/2019/11/sdp2019-11.pdf, November 2019), p. 2.

Leading, But Not the Leader

That said, central bankers are economists, not data scientists. Just as many businesses do to make big advances, we are partnering with private sector and academic innovators. We ask them to solve particular challenges by experimenting with digital tools and technologies. They apply through the central bank's *Partnerships in Innovation and Technology (PIVOT) Program*. [123]

Conclusion

Canada is now on the leading edge of new technologies, with AI hubs in Toronto, Montréal, Vancouver, Edmonton and many other cities, including Waterloo and Guelph. The Bank of Canada has the advantage to leverage those partnerships. At the Bank of Canada, we need to be in sight of what is going on in terms of the economy, i.e.,

- how enterprises are using this technology, how it is affecting their businesses, and how they set their prices
- what it means for investment; what it means for jobs

We must also be aware of how we can use that technology ourselves, in the most efficient and responsible way. Whether with machine learning, big data, AI, or RPA, we are using technology to do our business better, so that we make good predictions, forecasts, monetary policy decisions, and good assessments of the state of stability or vulnerability of the financial system.

[123] For more information on the PIVOT program, see "PIVOT," *Bank of Canada* [website] (https://www.bankofcanada.ca/research/partnerships-in-innovation-and-technology-pivot-program/#Our-objectives).

SOCIAL INNOVATION PILLAR

Reproducibility and Responsibility in AI Research

Graham Taylor,[124] *Canada Research Chair in Machine Learning; Canada CIFAR AI Chair, School of Engineering, University of Guelph and Vector Institute*

"Without barriers to entry and without regulation, use of technology is outpacing the rate at which we are becoming collectively wise."[125]

Yoshua Bengio, Professor, Université de Montréal; founder and Scientific Director of the Quebec Institute for Learning Algorithms (Mila)

Introduction

From the perspective of a machine-learning practitioner, my goals for this paper are to show how rapidly AI is changing, take a high-level look at AI research and adoption by business in Canada, then to demystify some of the technology. As a general-purpose technology, machine learning offers great potential for social innovation, which ties into the machine-learning community's ethos. However, the field's low barriers to entry contribute to systemic issues that work against that ethos.

As developers, practitioners, and promoters of AI systems, we cannot simply be technologists. We need to value pluralism in tech practice, and we need to develop a capacity to anticipate ethical issues and unintended uses. To address these issues, I will introduce the **Centre for Advancing Responsible & Ethical Artificial Intelligence (CARE-AI)** at the University of Guelph (U of G), whose mandate is to foster multidisciplinary AI research to improve life.

[124] For more information about Graham Taylor's research, see https://www.gwtaylor.ca/, or the Machine Learning Research Group at the University of Guelph (U of G) (https://github.com/uoguelph-mlrg), or the Centre for Advancing Responsible & Ethical Artificial Intelligence (CARE-AI) at U of G (https://www.care-ai.ca/).

[125] Paraphrased from Yoshua Bengio, "On the Wisdom Race [keynote address]" at *NeurIPS'2019* Workshop on Fairness and Ethics, Vancouver, B.C. (http://www.iro.umontreal.ca/~bengioy/EthicsWkshpNIPS-13dec2019.pdf, December 13, 2019).

AI Adoption in Canada and the Competition for Talent

"Despite [AI's] potential benefits, and despite Canada's robust ecosystem of fundamental research on AI, ***Canadian firms lag global peers in using AI*** *to support and enhance their businesses."* [Emphasis added.]

Sarah Villeneuve, Brookfield report, 2019[126]

I started my PhD in at the University of Toronto in 2004. Like other graduate students in my group, I worked on neural networks. Perhaps by a stroke of luck during my five-year PhD, this community of "computer nerds" went from working on a niche research topic to working on something that was very commercially relevant. We were excited about what we were developing but we did not really understand its full potential. Looking back, I saw hints that the topic was getting big.

Not until I started a faculty job at the University of Guelph in 2012 did I notice that things had really started to change. In December 2013, at the conference on Neural Information Processing Systems, usually attended solely by university- or industry-based researchers, Mark Zuckerberg, CEO of Facebook, appeared. At a fairly small academic conference, he was a very strange sight. In 2014, Google acquired DeepMind, where a lot of my friends had gone to work, and two of my post-doctoral advisors were hired by Facebook that same year.

Over the past five years, many mega-companies' prominent products and services have resulted from acquisitions of AI companies: e.g., Amazon's Alexa; Apple's Siri; and Google's contributions in healthcare through the DeepMind team. The acquiring companies have amassed much larger and well-resourced teams than we have in academia right now, with the exception, in Canada, of the pan-Canadian AI institutes: Amii, Mila, and the Vector Institute. Those centres are important for amassing talent and uniting researchers across campuses. Since their establishment in 2017, they have been successful in stemming the flow southward of Canadian talent.

A recent report by the Brookfield Institute (the "Brookfield report") said that the Canadian AI institutes have encouraged Facebook, NVIDIA, Samsung, and Uber to open AI labs here.[127] They give Canada the third-largest concentration of AI experts in the world[128] and Toronto one of the highest concentrations of AI start-ups in the world.[129]

---

[126] Sarah Villeneuve, *Boosting Competitiveness of Canadian Businesses: Clearing a Path to Wide-scale AI Adoption: Literature Review.* (Toronto: Brookfield Institute, 2019, https://brookfieldinstitute.ca/wp-content/uploads/AI-Talent-Literature-Review-FINAL.pdf), p. 2.

[127] Brookfield Institute for Innovation + Entrepreneurship (https://brookfieldinstitute.ca/).

[128] Sarah Villeneuve, *Boosting Competitiveness of Canadian Businesses: Clearing a Path to Wide-scale AI Adoption: Literature Review* (Toronto: Brookfield Institute, 2019, https://brookfieldinstitute.ca/wp-content/uploads/AI-Talent-Literature-Review-FINAL.pdf), p. 3, based on information from J. F. Gagne, Fedor Karmanov, and Simon Hudson, *Global AI Talent Pool Report 2018* (https://jfgagne.ai/talent/, [2019?]), which based its estimation on experts with LinkedIn profiles.

[129] Sarah Villeneuve, *Boosting Competitiveness of Canadian Businesses: Clearing a Path to Wide-scale AI Adoption: Literature Review* (Toronto: Brookfield Institute, 2019, https://brookfieldinstitute.ca/wp-content/uploads/AI-Talent-Literature-Review-FINAL.pdf), p. 3.

Despite those distinctions, our Canadian research successes have not translated into similar commercial successes. For example, the majority of R&D opportunities for new graduates still mainly exist either outside the country or at the laboratories of foreign-controlled firms. Many Canadian AI start-ups still look outside the country for venture capital funding. Most alarmingly, Canadian businesses lag their global peers in AI adoption.

"Canadian firms lag global peers in using AI to support and enhance their businesses," and "…AI adoption by existing firms and diffusion across the economy is weak."[130]

Sarah Villeneuve, Brookfield report, 2019

While not every firm has a business case suitable to use AI — due to lack of data or no standout product-market fit — this rapidly developing technology is widely applicable. In earlier talks, we heard it called a general-purpose technology. Companies are under pressure to at least develop an adoption strategy.[131] The Brookfield report explains that a lack of access to talent is holding companies back. The author warns that "addressing this talent gap is essential if Canada wants to overcome barriers to AI adoption and see its benefits manifest throughout the economy."[132]

Demystifying the Technology

"The algorithm was developed to automate thinking and to remove difficult decisions from the hands of humans to solve contentious debates."

Franklin Foer, in *World Without Mind*[133]

Algorithms and Machine Learning's Learning Algorithms

Long before AI, there were algorithms.[134] As anyone knows who has undergone the very intense series of interviews typical at Google, Facebook, or Amazon, brushing up on the study of algorithms is almost a prerequisite to arriving. So why are algorithms so important?

Simply put, an algorithm is a step-by-step description of how to solve a problem. More philosophically, it is a way to encapsulate human knowledge, or a way "…to automate thinking and to remove difficult decisions from the hands of humans to solve contentious debates."[135]

The way conventional programming works is that a human programmer comes up with the logic and writes a computer program composed of human-designed algorithms. This mechanical thinking runs on

[130] Ibid., p. 2-3.

[131] See the papers by Parvin Mousavi and Eric Santor, also in this volume.

[132] Sarah Villeneuve, *Boosting Competitiveness of Canadian Businesses: Clearing a Path to Wide-scale AI Adoption: Literature Review.* (Toronto: Brookfield Institute, 2019, https://brookfieldinstitute.ca/wp-content/uploads/AI-Talent-Literature-Review-FINAL.pdf), p. 2.

[133] Franklin Foer, *World Without Mind: The Existential Threat of Big Tech.* (New York: Penguin Press, 2017).

[134] The word "algorithm" is named after ninth-century mathematician and philosopher Muhammad Ibn Musa Al-Khwarizmi, latinized as "Algorithmi." The term was not actually popularized in the area of computer science until the 1960s. From: "Algorithm." *Wikipedia* (https://en.wikipedia.org/wiki/Algorithm, 2020).

[135] Franklin Foer, *World Without Mind: The Existential Threat of Big Tech.* (New York: Penguin Press, 2017).

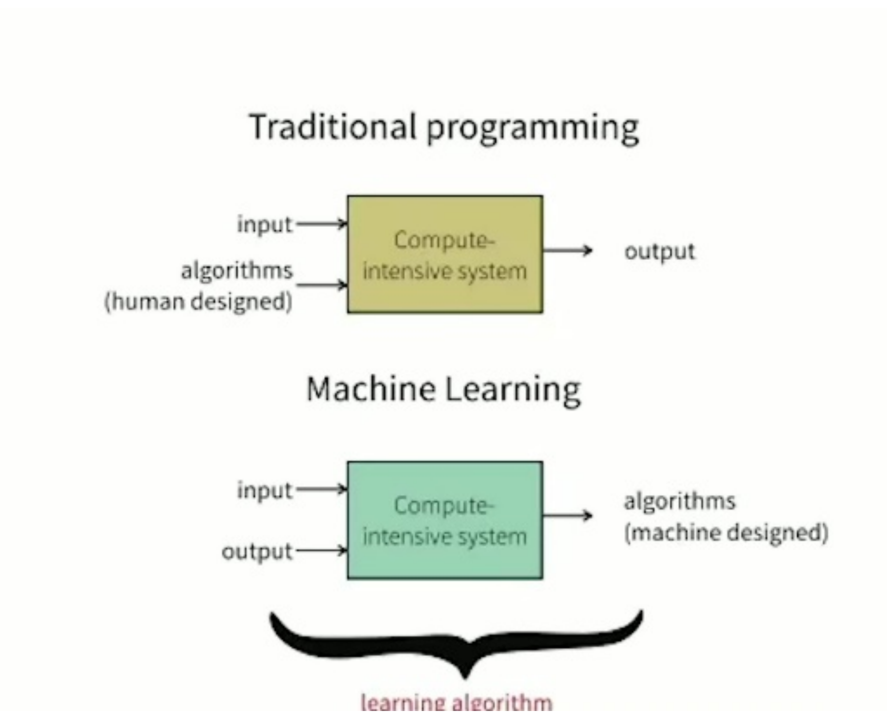
a **compute-intensive system** and, effectively, these algorithms transform input to output (top of Figure 24).[136]

Increasingly, the traditional algorithm is being replaced by something that is the product of machine learning. The resulting **machine-designed algorithms**, unlike the algorithms of traditional computing, are *not* just mechanical thinking, and this is what is really transforming computing.

These machine-learning systems, I would argue, are **highly general parameterized algorithms**, meaning that, at least in the case of neural networks, we rip out the rules and logic computed by a person and replace them with knobs — hundreds of millions of numbers. No longer tuned by a human, because that would be impossible, the knobs are tuned by a **learning algorithm** that takes data and transforms the settings of the knobs to make predictions (bottom of Figure 24).

Until now, these learning algorithms were designed by people like me: machine-learning practitioners. Even this is changing. At the cutting edge — and a bit scary — is something called **meta-learning**, which automates the learning algorithm. For now, however, human-designed learning algorithms automate the process of humans designing algorithms.

FIGURE 24: TRADITIONAL PROGRAMMING VERSUS MACHINE LEARNING



As algorithms are the building blocks of software, the revolution started with software. In 2011, Marc Andreessen, one of the world's software-engineering giants, famously said, "Software is eating the

[136] We say, "compute-intensive" system, rather than "computer," because, increasingly, algorithms are run on other devices, such as refrigerators, vehicles, etc.

world."[137] Now, as the AI revolution continues, Pete Warden, staff research engineer at Google, and prolific blogger, in a play on Andreessen's words, says, "Deep learning is eating software."[138]

Software 2.0 ["Two point Zero"]

"Thinking of an industry that has not been transformed by software is hard. By association, thinking of an industry that will not be transformed by machine learning is harder."

Graham Taylor

Thinking of an industry that has not been transformed by software is hard. By association, thinking of an industry that will *not* be transformed by machine learning is harder. We are seeing radical changes everywhere that software has been applied. Warden speculates "…that almost any data processing system with non-trivial logic can be improved significantly by applying modern machine learning."[139] Some people are calling this Software 2.0, a term popularized by Andrej Karpathy.[140]

This has big implications for how we teach these subjects. Warden predicts that our future developers are not going to be weavers of intricate logic, but teachers, curators of training data, and analysts. These do not require the same technical skills currently taught in software development. Instead, they require deep knowledge of the problem domain.[141]

Warden says that building software "should be far more accessible than traditional coding, once the tooling catches up…The essence of the process is providing a lot of examples of inputs, and what you expect for the outputs."[142]

This ties into the Brookfield report, which says,

*"**Firms will require individuals who can bridge the gap between business leaders and technologists** — often referred to as translators or explainers. Workers within this category will help to provide clarity on the potential and limitations of different AI applications and make recommendations regarding types of solutions and how to use them. Furthermore, businesses will also need to invest in individuals who will ensure a firm's AI solutions are **operating as intended**, flagging any unintended consequences that need to be addressed."* [143] [Emphasis added.]

---

[137] Marc Andreessen, "Why Software Is Eating the World," *Wall Street Journal* (https://www.wsj.com/articles/SB10001424053111903480904576512250915629460, August 20, 2011).
[138] Pete Warden, "Deep Learning is Eating Software," *Pete Warden's Blog* (https://petewarden.com/2017/11/13/deep-learning-is-eating-software/, November 13, 2017).
[139] Ibid.
[140] Andrej Karpathy, "Software 2.0," *Medium* (https://medium.com/@karpathy/software-2-0-a64152b37c35, November 11, 2017).
[141] Pete Warden, "Deep Learning is Eating Software," *Pete Warden's Blog* (https://petewarden.com/2017/11/13/deep-learning-is-eating-software/, November 13, 2017).
[142] Ibid.
[143] Sarah Villeneuve, *Boosting Competitiveness of Canadian Businesses: Clearing a Path to Wide-scale AI Adoption: Literature Review.* (Toronto: Brookfield Institute, 2019, https://brookfieldinstitute.ca/wp-content/uploads/AI-

LINK TO *FORESIGHT*: Build relevant skills and capabilities.

In addition to new strategy, business modelling, partnering, communications, and valuation of intangible skills, accountants must become digitally savvy. Digitization of transactions means that accountants will have to be much more familiar with data security, data analytics, structured and unstructured algorithms, and real-time reporting. Accountants, like all professionals, need to be agile, willing and able to learn new skills that are suitable for the fast-changing times…Accountants will need to work in concert with technology. (Appendix 1, p. 25)

Hardware: Google Tensor Processing Unit

What has happened to software is moving to hardware. Google's **tensor processing unit** is now in its third generation. It is a deep learning-specific series of circuits dedicated to running prediction algorithms rather than conventional computer programs. Low-powered versions are moving out of data centres and into our homes. These tiny AI chips "…will carry out machine-learning jobs in IoT devices."[144]

Machine Learning Tenets: Reproducibility and Open Source

Machine learning is a general-purpose, rapidly developing technology.[145] Fundamentally, it is a different class of algorithms and is very powerful. From the standpoint of developers, researchers, and students, machine-learning research has some unwritten tenets.

Reproducibility. The relationship between the machine-learning community and **open-source code** is very strong: we want other developers to be able to reproduce what we publish and recreate our experiments so that they can confirm that what we produce is not misleading or a mistake. At a simple level, **reproducibility** is making code public. At another level, Joëlle Pineau, professor at the School of Computer Science at McGill University and co-director of Facebook AI's research lab (FAIR), created a **reproducibility checklist**, which can be used to prepare papers for popular machine-learning conferences to ensure that submitted work is reproducible. [146]

Without freely available code, the field risks concentrating capabilities in the hands of a few massive companies and preventing others from profiting from the technology. So, reproducibility has an element of social good.

Talent-Literature-Review-FINAL.pdf), p. 10, based on work by H. Wilson, P. Daugherty, and N. Morini-Bianzino, "The Jobs that Artificial Intelligence Will Create," *MIT Sloan Management Review* (https://sloanreview.mit.edu/article/will-ai-create-as-many-jobs-as-it-eliminates/, Summer 2017).

[144] James Vincent, "Google Unveils Tiny New AI chips for On-device Machine Learning," *The Verge* (https://www.theverge.com/2018/7/26/17616140/google-edge-tpu-on-device-ai-machine-learning-devkit, July 26, 2018).

[145] See also papers by Parvin Mousavi and Eric Santor in this volume, who see machine learning becoming a general-purpose tool.

[146] [Joëlle Pineau], *The Machine Learning Reproducibility Checklist* (https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist-v2.0.pdf, April 7, 2020).

Worrisome Precedents

- **Closed Source?** In April 2019, I was the external examiner for Université de Montréal PhD student Sina Honari. Among many different contributions in his thesis was an algorithm that can take the face of an individual and re-target it to the pose of another individual.[147] Together, the student and his advisors, professors Christopher Pal (Polytechnique Montréal and Mila) and Pascal Vincent (Université de Montréal), decided not to release the source code for the algorithm, believing it could be used maliciously to produce fake or compromising images. In a field that really embraces reproducibility and the release of code, I found the decision surprising — but now, increasingly common.
- **Potential for malicious use?** In February 2019, media attention focussed on company OpenAI and its text generator, GPT-2, considered too dangerous to release because it was *too good.* "Trained on eight million web pages…GPT-2 is billed as the next generation of predictive text. The AI is said to write authentic-sounding prose that could fool humans, which has dangerous repercussions when it comes to the mass production of disinformation."[148] OpenAI worried that GPT-2 could be used to generate large-scale misinformation campaigns, polarize opinions in elections, or spread fake news. Was OpenAI purposely generating attention, in order to recruit talent in an increasingly competitive environment, or is GPT-2 actually dangerous? I had the opportunity to discuss GPT-2 with Jack Clark, policy director at OpenAI at an AI Ethics workshop in May 2019 hosted by CIFAR. Clark revealed that the U.S. government had conducted an export control review for AI technologies, likening them to dual-use technology, like rockets or nuclear power or night vision, which could be civilian or military, used for public good or public harm.[149]

Low Barriers to Entry

"Practitioners are racing to the field at a technical level to participate, but they might not have the maturity or training to participate in responsible deployment."

Graham Taylor

---

[147] Sina Honari, Feature Extraction on Faces : From Landmark Localization to Depth Estimation [thesis], *Papyrus* (Université de Montréal Institutional Repository) (https://papyrus.bib.umontreal.ca/xmlui/handle/1866/22658, December 2018).

[148] Ramona Pringle, "The Writing of this AI is so Human that its Creators are Scared to Release it: OpenAI's New System, Called GPT-2, is Described as 'Chameleon-like,' Matching both Subject and Style," *CBC.ca* (https://www.cbc.ca/news/technology/ai-writer-disinformation-1.5030305, February 25, 2019).

[149] Jack Clark of OpenAI contributed to a report that "…surveys the landscape of potential security threats from malicious uses of artificial intelligence technologies, and proposes ways to better forecast, prevent, and mitigate these threats" (p. 3). Further, the authors say, "Preparing for the potential malicious uses of AI…is an urgent task" (p. 65). See Miles Brundage, *et al.*, "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," *arXiv*:1802.07228 [cs.AI] (https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf, 2018).

So, is AI just another dual-use technology? Should it be managed the same way that other dual-use technologies have been managed in the past? I want to make the argument that *unlike* some other dual-use technologies, machine learning's barriers to entry are very low:

- Start-up costs are minimal, with no consumables other than energy and a computer.
- Open-source community- and industry-supported tools and frameworks are available, up-to-date, and free.
- Technical pre-requisites are not extreme, and a massive repository of learning resources exists (e.g., Massive Open Online Courses — MOOC.org; Coursera — coursera.org; Udacity — udacity.com, MITOpenCourseware — ocw.mit.edu; and Rutgers Accounting Digital Library (RAW) — raw.rutgers.edu).
- Rapid publishing of research results is typical and expected (e.g., through conferences, *arXiv.org* [commonly known as "the archive"], and *OpenReview.net*).

Practitioners are racing to the field at a technical level to participate, but they might not have the maturity or training to participate in *responsible* deployment. With low barriers to entry and without regulation, the technology, says Mila founder Yoshua Bengio, is outpacing the rate at which we are becoming collectively wise.[150] Jack Clark (Policy Director, OpenAI) and Gillian Hadfield (Vector Institute and OpenAI) say that existing regulatory systems are struggling to keep up.[151]

AI developers, practitioners, and users of the technology need to value pluralism in tech practice[152] and develop a capacity to anticipate ethical issues and unintended uses.

Barriers to Social Innovation

We can celebrate many things about the machine-learning community: we have openness towards releasing code, rapid publishing models, and reproducible research as core tenets in the discipline. Seeing AI as a general-purpose technology, we want to incentivize its development towards pressing social, economic, and political issues. But three *barriers* to social innovation work against us:

1. **Rate of development and dissemination.** The pressure on graduate students to publish research papers every four to six months discourages long-term research.[153] They also operate in simulated

---

[150] Paraphrased from Bengio, Yoshua. "On the Wisdom Race [keynote address]" at the *NeurIPS'2019* Workshop on Fairness and Ethics, Vancouver, B.C. (http://www.iro.umontreal.ca/~bengioy/EthicsWkshpNIPS-13dec2019.pdf, December 13, 2019).

[151] Jack Clark and Gillian K. Hadfield, "Regulatory Markets for AI Safety [conference paper]" at *SafeML ICLR 2019 Workshop* (https://sites.google.com/view/safeml-iclr2019/accepted-papershttps://sites.google.com/view/safeml-iclr2019/accepted-papers, May 6, 2019). [Safe Machine Learning Workshop at the eight annual International Conference on Learning Representations, 2019.]

[152] For more information on technocultural pluralism, see Osonde A. Osoba, "Technocultural Pluralism," *AI Pulse* (https://aipulse.org/technocultural-pluralism/, January 25, 2019).

[153] Yoshua Bengio, "Time to Rethink the Publication Process in Machine Learning," *Yoshua Bengio* [website] (https://yoshuabengio.org/2020/02/26/time-to-rethink-the-publication-process-in-machine-learning/, February 27, 2020).

time, rather than real-world time: i.e., algorithms might perform a task millions of times in a day. As a consequence, an agricultural dataset that might take several growing seasons to collect would seem too slow and time-consuming to study.

2. **Obsession with benchmark datasets.** This obsession stems from a desire for performance metrics, which are core to how we measure progress in our field.

3. **Machine-learning systems work for the population average.** In software engineering, the Pareto (80/20) rule is used in many ways for optimizing: e.g., 20% of the code might contain 80% of the bugs; 80% of users might use 20% of the code. In machine learning, we typically reject outliers, i.e., inputs that are unusual compared to the training dataset. In terms of **inclusive design**, this is a disaster:[154] if we carry the rule further, does this mean we make models that work well for 80% of the population, but exclude 20%? AI systems driven by machine learning are increasingly being used as part of algorithmic decision-making processes that impact peoples' opportunities and well-being. Lending, employment, education, and policing are areas which are increasingly — and alarmingly — becoming automated. Because machine-learning systems must be trained with data, they tend to work well for the "average person" and not well for marginalized groups, for which training data often do not exist. When we develop machine-learning models, we tend to focus on average accuracy, so "95% correct" may hide the fact that the system performs horribly for a minority group.

***COVID-19 Update:* The pandemic is affecting AI models.**

AI runs "…behind the scenes in inventory management, fraud detection, marketing, and more." Because of the preponderance of pandemic-related search terms entered into websites, some "[m]achine-learning models trained on normal human behavior are now finding that normal has changed, and some are no longer working as they should."[155]

Advancing Social Innovation

In response to rising uncertainty and apprehension around the long-term effects of AI on our health, on our economy, on our society at large, we recently launched the **Centre for Advancing Responsible & Ethical Artificial Intelligence (CARE-AI)** at University of Guelph.[156] Its mission is

[154] Idea presented at U of G, November 2019, by Jutta Treviranus, Director, Inclusive Design Research Centre (IDRC), OCAD University (https://idrc.ocadu.ca/about-the-idrc/staff-pages/112-jutta-treviranus-director).

[155] Will Douglas Heaven, "Our Weird Behavior During the Pandemic is Messing with AI Models: Machine-learning Models Trained on Normal Behavior are Showing Cracks —Forcing Humans to Step in to Set Them Straight," *MIT Technology Review* (https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/, May 2020).

[156] See Centre for Advancing Responsible & Ethical Artificial Intelligence (CARE-AI) at U of G, at https://www.care-ai.ca/.

…to advance multidisciplinary AI training and research, and its responsible application to improve life.

CARE-AI focusses on three AI-related themes: methodology, responsibility, and applications that improve life. As an example, an internship project at Facebook AI, by U of G PhD student Terrance DeVries, assessed computer vision models and the fact that they tend to work well for developed countries and not so well for the developing world.[157] The research shows that the data sets used to train models are very biased towards images that come from developed countries. For example, objects like soap and spices and toothpaste look very different in developing countries like Nepal or Burundi than they do in developed countries like Canada or the United Kingdom.

Collaboration is Important

Through CARE-AI, Guelph developed a collaborative specialization in AI. The Vector Institute-recognized Master's-degree program allows specialties in mathematics, statistics, engineering, bioinformatics, and computer science. The program includes a mandatory capstone course in AI Applications and Society. The case study-based course is co-taught by four departments, including philosophy.

Recall, the Brookfield report reminds us that collaboration is important for business, as well, when it says, that "…firms will require individuals who can bridge the gap between business leaders and technologists—often referred to as translators or explainers."

LINK TO *FORESIGHT*: Create an effective and flexible profession.

If CPAs are to become relevant to business and society beyond finance professionals, it is vital that they develop more effective and efficient ways of operating, while also allowing for experimentation. For example, the profession must explore how best to partner with non-members, other associated professions, and international bodies. It might also establish an innovation hub (sandbox) to experiment with new technologies and methodologies, to ensure their success before rolling them out to the members. (Appendix 1, p. 26)

Conclusion

In the first part of this paper, I described machine learning as a new way to build algorithms, not by careful tuning or programming but, automatically, from data. The idea of computers instead of humans writing software sounds scary, but it is the next evolution in computing. This view should make clear the notion of **general-purpose technology**.

[157] Terrance DeVries, et al., "Does Object Recognition Work for Everyone?" *arXiv*:1906.02659 [cs.CV] (https://arxiv.org/abs/1906.02659, June 6, 2019).

In a heavily empirical field, the machine-learning community has rallied around the idea of reproducible research. In its most basic form, this means releasing source code so that others can recreate published experimental results. Practitioners have proposed stronger notions of reproducibility, extending to both empirical and theoretical work, that are being integrated into submissions to our community's leading conferences. Recently, we have seen instances where researchers have held back code out of concern that its associated machine-learning systems would be misused by bad actors. I made the argument that saying AI is simply a dual-use technology is insufficient. Machine learning has extremely low barriers to entry, so developers and practitioners must reflect and act on the tension emerging between reproducibility and responsibility.

Finally, I described our efforts at U of G to train a generation of **tech pluralists**, who will harness this incredible technology and drive it to improve life.

About the Speakers & Moderators

Technology Pillar

Richard Zemel

Research Director, Vector Institute; Professor, University of Toronto, Department of Computer Science, University of Toronto; Senior Fellow, Canadian Institute for Advanced Research

Richard Zemel is a Professor of Computer Science and Industrial Research Chair in Machine Learning at the University of Toronto, and the co-founder and Research Director at the Vector Institute for Artificial Intelligence. Prior to that he was on the faculty at the University of Arizona, and a Postdoctoral Fellow at the Salk Institute and at CMU. He received a BSc in History & Science from Harvard, and a PhD in Computer Science from the University of Toronto. He is also the co-founder of SmartFinance, a financial technology startup specializing in data enrichment and natural language processing. His awards and honours include a Young Investigator Award from the ONR and a U.S. Presidential Scholar award. He is a Senior Fellow of the Canadian Institute for Advanced Research (CIFAR), a NVIDIA Pioneer of AI, and a member of the NeurIPS Advisory Board. His research is supported by grants from NSERC, CIFAR, Google, Samsung, Amazon, Microsoft, DARPA and iARPA.

Fakhri Karray

University Research Chair; Director, Centre for Pattern Analysis and Machine Intelligence; Co-director, Waterloo Artificial Intelligence Institute; and Professor, Department of Electrical and Computer Engineering, University of Waterloo

Dr. Fakhri Karray is the University Research Chair Professor in Electrical and Computer Engineering and the co-director of the Institute of Artificial Intelligence at the University of Waterloo. He is also the director of the University's Centre for Pattern Analysis and Machine Intelligence. He holds the Loblaw's Research Chair in Artificial Intelligence. His research interests are in the areas of intelligent systems design, augmented intelligence, concept mining, machine learning, and context-aware machines. His work has been applied to intelligent transportation systems, Internet of things, cognitive robotics, medical imaging and natural man-machine interaction. He is the co-author of two dozen U.S. patents and has served as the associate editor/guest editor for more than 14 research journals.

Dr. Karray's research work has been featured on Discovery Channel, CBC, *The Globe and Mail*, among others. He is the recipient of the University of Waterloo Best Performance Award. He also received national and international awards, including the Premier Research Excellence Award, the Pattern Recognition Society Best Paper Award, the World Automation Congress' Anderson Best Paper Award, and the IEEE Appreciation Certificate for Notable Services and Contributions to IEEE and the Engineering Profession. He has served as the University of Waterloo's Academic Advisor for Amazon's Alexa Fund Fellowship Program and is a Fellow of the Canadian Academy of Engineering. He serves as the president of the Association for Image and Machine Intelligence and is on the Advisory Board of a number of research journals and high-tech companies in North America.

Parvin Mousavi

Professor and Director, Medical Informatics Lab, School of Computing, Queen's University

Parvin Mousavi is a professor of Computer Science and Electrical and Computer Engineering at Queen's University, Canada, and a member of the Royal Society of Canada, College of New Scholars, Artists and Scientists. She received her PhD in Electrical and Computer Engineering from the University of British Columbia (UBC), Canada. Previously, she has held industrial positions with Molecular Mining Inc. and Biosystemix Inc., Canada. She has had visiting professorships at University of British Columbia and Harvard Medical School. Her research interests are in machine learning and artificial intelligence applied to oncology, computer-assisted interventions, and neurology. She has received over 30 international and national recognitions for research and has led numerous multi-national collaborative initiatives in these areas.

Professor Mousavi represented the Royal Society of Canada in the Summit of the G7 Academy of Sciences on the topic of AI and Society. She serves on and chairs several national and international granting committees and is an Associate Editor for *PLOS One* and *BMC Bioinformatics*. She is the General Co-chair of Information Processing in Computer Assisted Interventions (IPCAI), and an executive member of Medical Image Computing and Computer Assisted Interventions (MICCAI) in 2017 and 2020, and a founding member of Women in MICCAI. She is a senior member of IEEE, IEEE EMBS, and IEEE Women in Engineering.

Governance Pillar

Miklos Vasarhelyi

KPMG Distinguished Professor; Director, Rutgers Accounting Research Center and Continuous Auditing & Reporting Lab, Rutgers Business School, Rutgers University

Professor Miklos A. Vasarhelyi is the KPMG Distinguished Professor of Accounting Information Systems and Director of the Rutgers Accounting Research Center (RARC) & Continuous Auditing & Reporting Lab (CAR Lab) at Rutgers University. He is credited with developing the original continuous audit application and is the leading researcher in this field. He also leads the Rutgers AICPA Data Analytics Research Initiative supported by the eight leading CPA firms, AICPA, and CPA Canada.

Professor Vasarhelyi received his PhD in Management Information Systems from UCLA and his MBA from Massachusetts Institute of Technology. He has published more than 200 journal articles, 20 books, and directed over 40 Ph.D. theses. He is the editor of *the Artificial Intelligence in Accounting and Auditing* series and the *Journal of Emerging Technologies in Accounting*. Before joining Rutgers, he taught at USC, Columbia and worked at the Bell Laboratories. He was awarded the Outstanding Accounting Educator by the AAA in 2015 and received ISACA's Wasserman Award, among many distinctions.

Anand Rao

Global & U.S. Artificial Intelligence and U.S. Data & Analytics Leader, PwC U.S.

Dr. Anand S. Rao is a Partner in PwC's Advisory practice. He is the Global Artificial Intelligence Lead, Cross-vertical Analytics Champion, and the Co-sponsor for the AI Center of Enablement within PwC. With over 33 years of industry and consulting experience, Anand leads a team of practitioners who work with C-level executives at some of the world's largest organizations. As the global lead for AI, he is responsible for research and commercial relationships with academic institutions and startups, research, development and commercialization of innovative AI, big data and analytic techniques.

Prior to joining management consulting, Anand was the Chief Research Scientist at the Australian Artificial Intelligence Institute. He has received widespread recognition for his extraordinary contributions in the field of consulting and artificial intelligence research. He has received the Most Influential Paper Award for the Decade in 2007 from the Autonomous Agents & Multi-Agent Systems organization for his contribution on the Belief-Desire-Intention Architecture; MBA Award of Distinction from Melbourne Business School in 1997 and University Postgraduate Research Award (UPRA) from University of Sydney in 1985; Distinguished Alumnus Award from Birla Institute of Technology and Science, Pilani, India; He was recognized as one of Top 50 Data & Analytics professionals in the U.S. and Canada by Corinium; one of Top 50 professionals in InsureTech; and his recent paper on “A Strategist’s Guide to Artificial Intelligence” has won the National Gold Award by ASBPE for the Best Technical article in 2017 and the FOLIO editorial award.

Anand is on the Advisory Board of a number of educational and not-for-profit institutions focussed on AI including AI Global, Nordic AI Institute, and the International Congress for the Governance of AI. He has co-edited four books and published over 50 papers in refereed journals and conferences. He is a frequent speaker on artificial intelligence, behavioural economics, autonomous cars and their impact, analytics, and technology topics in academic and trade forums.

Eric Santor

Advisor to the Governor of the Bank of Canada

Dr. Santor was appointed Advisor to the Governor on Digitalization in March 2019. In this role, he leads the Bank’s digitalization work, including research into the impact of digitalization on the economy and financial system. Mr. Santor also leads the initiative to incorporate technologies such as artificial intelligence and machine learning, as well as big data, into the Bank’s operations. This involves leveraging programs such as Partnerships in Innovation and Technology (PIVOT) and the Bank’s relationship with the Creative Destruction Lab.

Dr. Santor joined the Bank in 2001 as an economist in the former Monetary and Financial Analysis Department. He moved to the International Economic Analysis Department in 2003, where he assumed increasing responsibilities until becoming Managing Director in 2013. Before his appointment as Advisor to the Governor on Digitalization, Mr. Santor served as Managing Director of the Bank’s Canadian Economic Analysis Department.

Dr. Santor’s research has focussd on issues relating to the incidence and effects of unconventional monetary policy, the international monetary system and global financial architecture, and the impact of ownership structure on Canadian firms. Dr. Santor was born in London, Ontario. He completed his BA in History and Political Science at Huron College, University of Western Ontario, and his PhD in Economics at the University of Toronto.

Social Innovation Pillar

Graham Taylor

Canada Research Chair in Machine Learning; Canada CIFAR AI Chair, School of Engineering, University of Guelph and Vector Institute

Graham Taylor is an Associate Professor at the University of Guelph where he leads the Machine Learning Research Group. He is the academic director of NextAI, non-profit initiative to establish Canada as the AI hub for research, venture creation, and technology commercialization, and is a member of the Vector Institute for Artificial Intelligence.

Professor Taylor received his PhD in Computer Science from the University of Toronto. His research focusss on statistical machine learning, with an emphasis on deep learning and sequential data. Much of his work has focussd on “seeing people” in images and video. He is especially interested in time series, having applied his work to better understand human and animal behaviour, environmental data (climate and agricultural), audio (music and speech) and financial time series.

Dr. Taylor was recently selected by the Canadian Institute for Advanced Research (CIFAR) as one of two Azrieli Global Scholars appointed to the Learning in Machines and Brains Program: an international competition recognizing excellence in research and leadership. He has received over $2M in external research funding, including a highly competitive NSERC-French National Research Agency Strategic Partnerships Grant, and trained 50 Highly Qualified Personnel since his appointment. Papers he has authored or co-authored have been cited over 3000 times.

Panel moderators and Conference organizers

Yue Li

Associate Professor of Accounting, University of Toronto

Yue Li is an Associate Professor at the University of Toronto with cross-appointments at the Institute for Management & Innovation and Joseph L. Rotman School of Management. He received his MBA from the University of Toronto and his PhD from Queen’s University. Yue is a CPA, CMA. His research focusses on the disclosure and valuation relevance of corporate environmental performance, corporate social responsibility, and sustainability practice. Yue has served as Ad Hoc Editor for Contemporary Accounting Research, as Associate Editor for the Asia-Pacific Journal of Accounting and Economics, Managerial Auditing Journal, and Accounting Forum, and as Guest Editor for the Journal of Management Accounting Research and Asia Review of Accounting. He has published his research in the leading accounting research journals, including The Accounting Review, Contemporary Accounting Research, Accounting, Organizations and Society, Journal of Accounting and Public Policy, European Accounting Review, Journal of Accounting, Auditing and Finance, among others.

Soo Min Toh

Associate Professor of Organizational Behaviour and HR Management, Director of the Institute for Management & Innovation, University of Toronto

Soo Min is an Associate Professor at the University of Toronto, is cross-appointed to the Institute for Management & Innovation and the Joseph L. Rotman School of Management, and is Professorial Fellow at the University of Edinburgh Business School. She received her PhD from Texas A&M University. Soo Min sits on several editorial review boards of significant international journals and was the Chair of the International Affairs Committee at the Society of Industrial & Organizational Psychology. At the University of Toronto, she teaches undergraduate and graduate students about leadership, bias, maximizing human potential, and conducting research. Her research interests include cross-cultural management, leadership, and cooperation. She has published in the *Academy of Management Journal, the Academy of Management Review*, the *Journal of Applied Psychology*, *Psychological Science*, and serves on the editorial board of the *Journal of International Business Studies*. In addition, her work on leadership, women leaders, and cross-cultural intergroup cooperation has been featured in international news and business media, including the *Financial Times*, *Fortune*, *The Globe & Mail*, and *Harvard Business Review*.

Irene M. Wiecek

Professor of Accounting, Teaching Stream, Director of the Master of Management & Professional Accounting Program and Director, BIGDataHUB, University of Toronto

Irene is a Professor at the University of Toronto, where she is cross-appointed to the Institute for Management & Innovation and Joseph L. Rotman School of Management. Irene has been involved in professional accounting education for over twenty-five years, sitting on various university and provincial / national professional accounting organization committees, as well as developing and directing the CPA Canada IFRS Immersion Programs for practising accountants and founding and co-directing the CPA/Rotman Centre for Innovation in Accounting Education. In the area of standard setting, she was a member of the Accounting Standards Board IFRS Discussion Group until 2020.

Irene has co-authored numerous books and publications including seven editions of the text *Intermediate Accounting* (by Kieso, et al.) for which she is one of two Canadian co-authors on the

Canadian edition. Irene’s interests lie in the area of International Financial Reporting Standards, integration in accounting education, and most currently, big data, artificial intelligence and emerging technologies, including blockchain. Irene is a member of the CPA Canada Foresight Working Group on Reimaging the Profession, as well as the group’s Data Governance Workstream. In 2019, she was appointed to the CPA Competency Map Task Force, which is taking a blank-sheet approach to creating the next CPA Competency Map. She is an FCPA, FCA and is the Director and founder of the BIGDataHUB at the University.